nature genetics

Article

https://doi.org/10.1038/s41588-025-02263-6

G-quadruplexes are promoter elements controlling nucleosome exclusion and RNA polymerase II pausing

Received: 7 March 2023

Accepted: 11 June 2025

Published online: 22 July 2025

Check for updates

Cyril Esnault[1] ✉, Amal Zine El Aabidine[1], Marie-Cécile Robert[1,2,3], Anne Cucchiarini[4], Talha Magat[1], Alexia Pigeot[1], Soumya Bouchouika[1], Encar Garcia-Oliver[1], Kevin Gawron[1], Eugénia Basyuk[1,5], Magdalena A. Karpinska[1,2,3], Alja Kozulic-Pirher[1,2], Yu Luo[4,6], Daniela Verga[6], Raphael Mourad[7], Ovidiu Radulescu[8], Jean-Louis Mergny[4], Edouard Bertrand[1,2,3] & Jean-Christophe Andrau[1,9] ✉

Despite their central role in transcription, it has been difficult to define universal sequences associated with eukaryotic promoters. Within the chromatin context, recruitment of transcriptional machinery requires promoter opening, but how DNA elements contribute to this process is unclear. Here we show that G-quadruplex (G4) secondary DNA structures are highly enriched at mammalian promoters. G4s are located at the deepest point of nucleosome exclusion at promoters and correlate with maximum promoter activity. We found that experimental G4s exclude nucleosomes in vivo and in vitro while favouring strong positioning. At model promoters, impairing G4s affected both transcriptional activity and chromatin opening. G4 destabilization also resulted in an inactive promoter state and affected the transition to effective RNA production. Finally, G4 stabilization resulted in global reduction of proximal promoter pausing. Altogether, our data introduce G4s as bona fide promoter elements allowing nucleosome exclusion and facilitating pause–release by RNA polymerase II.

Initially defined by analogy to bacterial transcription models[1–3] and based on in vitro transcription assays from the pregenomic era, eukaryotic core promoters were defined as "the minimal stretch of contiguous DNA sequence that is sufficient to direct accurate initiation of transcription by the RNA polymerase II (Pol II) machinery"[4]. However, and unlike in bacteria, eukaryotic promoters require nucleosome exclusion for transcriptional machinery recruitment. Beyond core promoters, eukaryotic promoters include broader sequence contexts that vary across species but commonly generate a well-positioned nucleosome array and open chromatin[5,6]. In mammals, CpG islands (CGIs) contribute to chromatin opening, independently of transcription, due to their high GC and CpG content[7,8].

Promoters are also highly enriched in potential DNA secondary structures, suggesting that these structures play a role in transcription regulation[9]. Among these, G4s are single-stranded and stable structures that consist of a planar arrangement of stabilized guanines that are

[1]Institut de Génétique Moléculaire de Montpellier, CNRS-UMR 5535, University of Montpellier, Montpellier, France. [2]Institut de Génétique Humaine, CNRS-UMR9002, University of Montpellier, Montpellier, France. [3]Equipe labélisée Ligue Nationale Contre le Cancer, Montpellier, France. [4]Laboratoire d'Optique et Biosciences, Ecole Polytechnique, CNRS, Inserm, Institut Polytechnique de Paris, Palaiseau, France. [5]Laboratoire de Microbiologie Fondamentale et Pathogénicité, CNRS-UMR 5234, Université de Bordeaux, Bordeaux, France. [6]CNRS UMR9187, INSERM U1196, Institut Curie, PSL Research University, Orsay, France. [7]LBCMCP, Centre de Biologie Intégrative, Université de Toulouse, CNRS, UPS, Toulouse, France. [8]Laboratory of Pathogen Host Interactions, UMR CNRS 5235, University of Montpellier, Montpellier, France. [9]Equipe labéllisée Fondation pour la Recherche Médicale, Montpellier, France. ✉e-mail: cyril.esnault@igmm.cnrs.fr; jean-christophe.andrau@igmm.cnrs.fr

overrepresented in regulatory regions. Their predicting sequences are present at over a million occurrence in the human genome and more specifically at promoters. However, and to date, it was unclear how they could contribute to transcription, as it was described that they either activate or repress transcription depending on promoter contexts[10–13]. Furthermore, while it was proposed that their formation in vivo could be dependent on a high level of transcriptional activity[14,15] and that mutating a specific G4 in the *MYC* promoter could result in a change of transcription start site (TSS)[16], the possibility that they could represent promoter elements on their own was not directly tested.

Here, we identify G4s as key mammalian core promoter elements. Using predicted and experimental G4 data, both in vitro and in vivo, we find G4s located at the deepest point of nucleosome exclusion at promoters, correlating with maximum promoter activity in the Survey of Regulatory Elements (SURE) assay. Furthermore, G4s exclude nucleosomes both in vivo and in vitro and exhibit strong nucleosome-positioning potential. Mutating G4s in model promoters affected transcription activity and chromatin accessibility. Destabilizing G4s increased the probability of promoters being in an inactive state (OFF times) and impaired the transition to effective RNA production in live imaging experiments. Conversely, stabilizing G4s using ligands globally reduced promoter-proximal Pol II pausing, aligning with live imaging results.

Altogether, our findings establish G4s as functional promoter elements that enable nucleosome exclusion and facilitate Pol II pause–release. These results provide a new perspective on promoter architecture, emphasizing G4s as critical components of transcription regulation.

Results

G4s are enriched at promoters and correlate with activity

Based on the knowledge that the TFIID general transcription factor binds naked DNA in vitro in a window frame of 40–50 bp around TSSs[17], sequence analyses of core promoters were often restricted to this window frame. These searches often yielded motifs poorly enriched, lowly conserved in evolution or highly degenerated[18,19]. Because the most open areas of chromatin extend on average up to 100 bp, we performed a motif search on core promoters associated with open chromatin upstream and downstream (−100 and +20) of experimental TSSs in three mammalian cell types (primary T cells, K562 and Raji cells; Fig. 1a and Extended Data Fig. 1; Supplementary Table 1 describes all datasets). This analysis revealed a prominent G-stretch containing-motif in the three cell types, highly compatible with the formation of G4s in vitro, using G4Hunter (G4H) predictions[20] at various stringencies. To consolidate this result, we investigated the frequency of predicted G4s (pG4s) or other motifs identified. We found overall that stringent G4 predictions (G4H1.5 and G4H2.0)[20] show very high frequency as well as a strong enrichment above control sequences (observed/expected)[8,21]. Enrichments of pG4s (20–45%) are also higher than for ETS and NF-Y and far above the TATA box motifs (Fig. 1a, Extended Data Fig. 1 and Supplementary Table 2). Their enrichment is similar to BRE and SP1 motifs, both compatible with G4 formation (Extended Data Fig. 2a). Finally, we note that pG4s containing promoters exhibit less of the other motifs (Extended Data Fig. 2b), suggesting that they have more propensity to function autonomously.

A large fraction of mammalian promoters were shown to be intrinsically bidirectional[22,23]. We thus thought to analyze the enrichment of pG4s upstream of sense and antisense experimental TSSs. This showed that pG4s peak on average at positions at −56 and −51 of the TSS, respectively (Fig. 1b). We note that the presence of a canonical or noncanonical TATA box influences far more directional and focused transcription (Extended Data Fig. 2c) as described previously[8,21].

To further investigate whether G4s contribute to transcription initiation and promoter activity, we took advantage of four orthogonal approaches for G4 formation assessment, including G4access, a technology we recently developed[24], which allows mapping G4s globally in the chromatin context (Fig. 1c). We also used G4 chromatin immunoprecipitation (ChIP)[14,25], mapping G4s in living cells, G4 sequencing (G4-seq)[26], globally assessing G4s in vitro, and single-stranded DNA (ssDNA) sequencing (ssDNA-seq)[27], mapping ssDNA genome-wide. We performed G4access in Raji and K562 cells and processed published datasets for the other methods whenever available (Extended Data Fig. 2d,e). In this analysis, we ranked the promoters containing pG4s (G4H > 2.0) by increasing G4 signal. G4-seq and ssDNA-seq in Raji cells were used to further validate folded G4s. G4access and G4 ChIP outputs of G4 measurement are very comparable. ssDNA-seq globally confirmed G4 formation in Raji cells, while G4-seq validated that selected sequences can form G4s in vitro. Because G4-seq monitors genomic G4s outside of the nuclear context, signals of the six groups originally defined from low to high G4 formation in living cells remained largely unchanged. This further indicates that G4access and ChIP do map G4s in the context of chromatin in living samples.

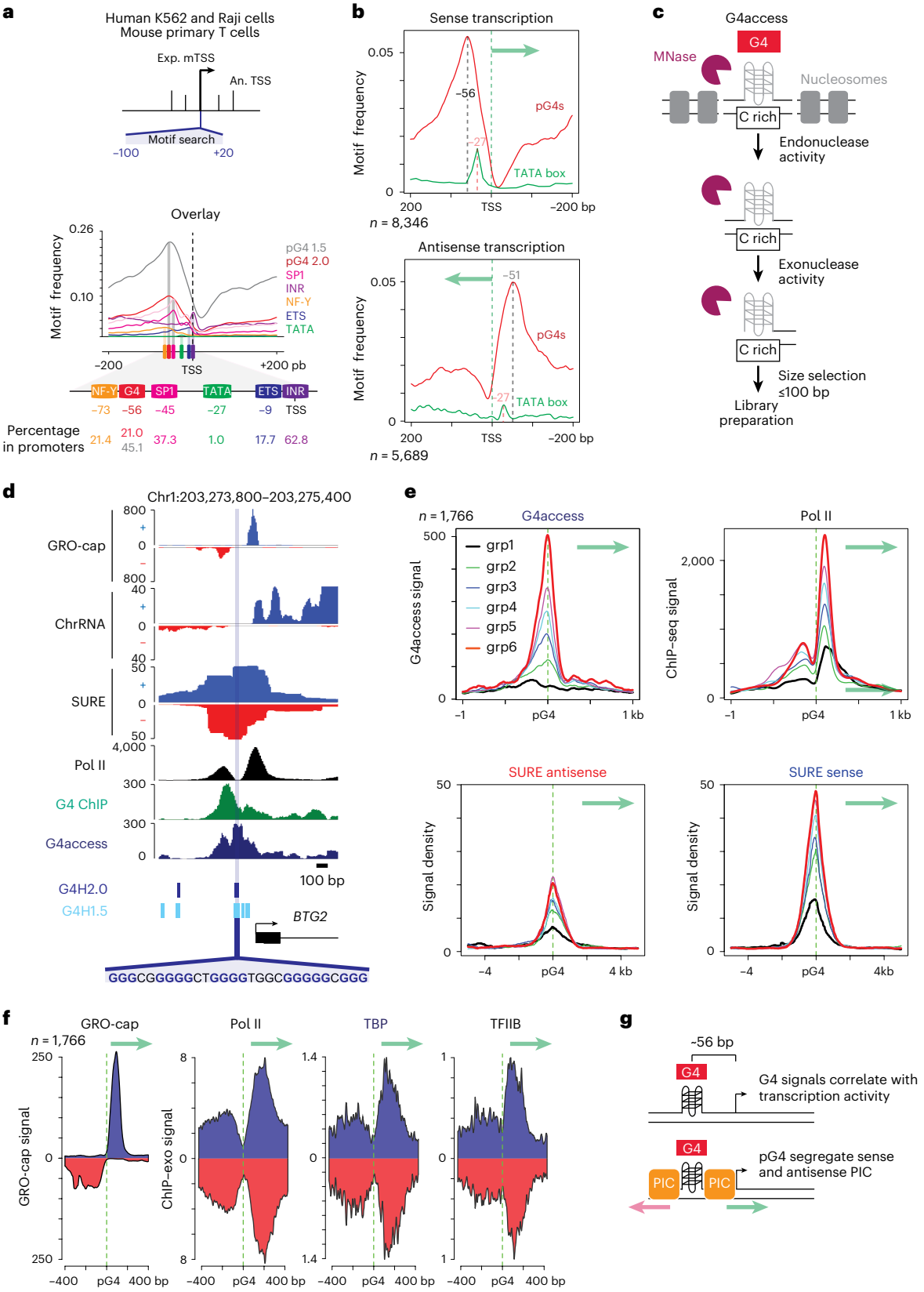
We then compared G4 formation to datasets (Supplementary Table 1) monitoring the nascent transcriptome and large-scale measurement of the promoter activity assay (SURE)[28] (Fig. 1d,e). As illustrated at the *BTG2* promoter, G4s are located at the midpoint of Pol II signals that reflect promoter divergent transcription. They also hallmark maximum promoter activity assessed by SURE (Fig. 1d). To establish this statement globally, we analyzed the correlation of G4access at pG4 promoter locations with that of SURE and Pol II and found that both increase with experimental G4 levels (Fig. 1e). In addition, we observed that pG4 locations overlapped with the midpoints between

Fig. 1 | pG4s are highly enriched at promoters and correlate with maximum mammalian promoter activity. a, Promoter motif search around experimental TSSs highlights pG4 motifs in various cells. Top: search area around experimental (exp.) TSSs in our three model cell types. Bottom: motif distances to experimental TSSs in K562 cells, overlay of motif densities, relative representation at promoters and enrichment over control sequences are shown. INR, initiator. The cartoon shows that often the main experimental TSS (Exp. mTSS) differs from the annotated TSS (An. TSS). See also Extended Data Fig. 1 and Supplementary Table 2 for details of motif discovery analyses in the three model cells. SP1, specificity protein 1; NF-Y, nuclear transcription factor Y; ETS, winged helix-turn-helix DNA binding domain ETS transcription factors; TATA, TATA box. **b**, pG4s mark genomic areas between sense and antisense promoters. Motif densities at K562 promoters show enrichment of pG4s (G4H2.0) 56 bp upstream of sense experimental TSSs (n = 8,346) and 51 bp of antisense experimental TSSs for genes for which divergent transcription initiation is detected (n = 5,689). See heatmaps in Extended Data Fig. 1 for Pol II and short chromatin RNA profiles. Green arrows indicate sense of transcription. **c**, G4access principle. Chromatin is digested at a moderate MNase level to which G4s are resistant and enriched in the subnucleosomal fractions. These fractions are purified and subjected to library preparation and high-throughput sequencing. Details of the method are described in ref. 24. **d**, Example of a pG4 sequence fitting the midpoint of upstream and downstream Pol II and nascent transcription (chrRNA and GRO-cap) at the *BTG2* promoter, experimental G4s (G4access and G4 ChIP) and maximum promoter activity (SURE assay) in K562 cells are displayed. Sequence of the pG4 (G4H2.0) is indicated below the gene. Chr, chromosome. **e**, Experimental G4 signals correlate with transcription and the SURE promoter activity assay. Groups (grp)1–6 correspond to increasing levels of G4access signals. Green arrows represent the sense of gene transcription. Metaprofiles are centered on the G4 motif upstream of the TSS as shown below the tracks. **f**, pG4s are located at the midpoints of GRO-cap, Pol II and general transcription factors using ChIP–exo datasets. See also Extended Data Fig. 2e for analyses in Raji and mouse T cells. **g**, Model of average pG4 locations as determined by the G4H algorithm upstream of TSSs and as a midpoint of upstream and downstream Pol II peaks. See Supplementary Table 1 for datasets used, references and Gene Expression Omnibus (GEO) accession numbers.

https://doi.org/10.1038/s41588-025-02263-6

sense and antisense transcription initiation mapped by ChIP with exonuclease (ChIP–exo) or ChIP with sequencing (ChIP–seq) of Pol II, TATA-binding protein (TBP), transcription factor II B (TFIIB)[29] and Global run-on associated to CAP sequencing (GRO-cap)[30] (Fig. 1f and Extended Data Fig. 2g). Finally, we also observed that pG4s located at the promoter are found upstream of R-loops (Extended Data Fig. 2f) that form where Pol II and GRO-cap signals increase, on the side of the pG4s (Fig. 1f).

Collectively, these analyses show that pG4s represent major motifs of extended core promoters, located on average at a relatively fixed position from TSSs. Moreover, G4s correlate with both promoter activity and midpoints of divergent transcription (Fig. 1g).



https://doi.org/10.1038/s41588-025-02263-6

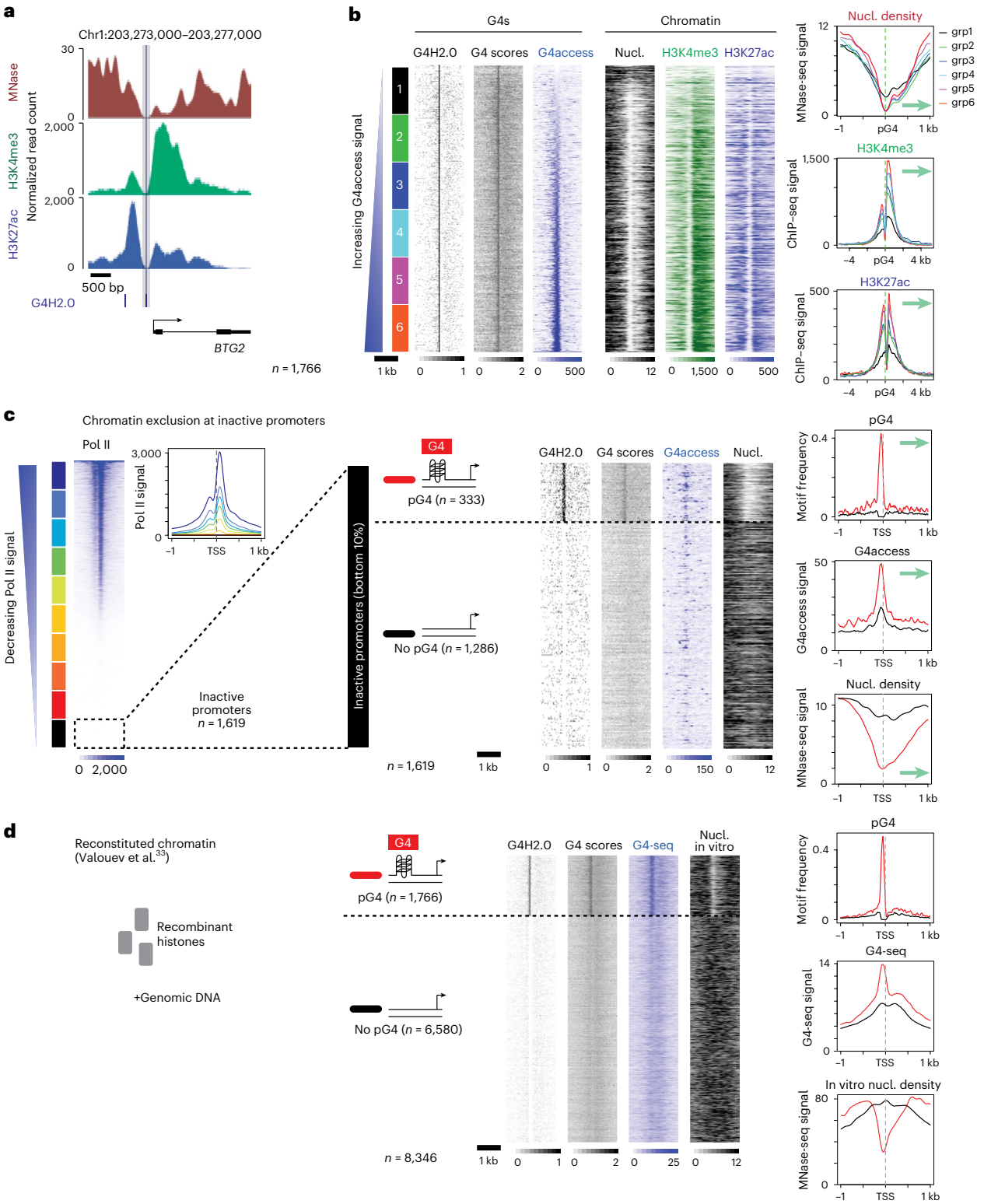


Fig. 2 | G4s promote nucleosome exclusion at active and inactive promoters in vivo and in vitro. a, Example of a pG4 sequence fitting the maximum of the NDRs at the *BTG2* promoter in K562 cells using MNase-seq and ChIP–seq of active chromatin marks (H3K4me3, H3K27ac). **b**, Experimental G4 signals correlate with levels of nucleosome depletion and with active histone modification marks. Promoters that harbor strong G4 predictions using the G4H algorithm G4H2.0 were ranked by G4access signal as depicted in the heatmap; corresponding heatmaps of G4 ChIP, G4H2.0, computed G4 scores, MNase-seq, H3K4me3 and H3K27ac are also shown (left). Promoters were split in six groups, as in Fig. 1e. Metaprofiles of nucleosome densities (nucleosomes (nucl.), MNase-seq) and of H3K4me3 and H3K27ac of all groups are displayed (right). See also Extended Data Fig. 2h for other cell types. Green arrows represent the sense of gene transcription. **c**, pG4s promote nucleosome exclusion at inactive promoters in K562 cells (Encyclopedia of DNA Elements (ENCODE)). Transcriptionally inactive promoters (bottom 10% of Pol II signals shown on the left) were split in two groups with or without G4H2.0 prediction. See also Fig. 3 for pG4 influence on nucleosome positioning at inactive IGRs and Extended Data Fig. 5a after transcription inhibition with triptolide. **d**, pG4-containing promoters have intrinsic nucleosome exclusion properties on in vitro reconstituted chromatin (analyzed from ref. 33). Promoter selections are based on K562 active promoters shown in Fig. 1. See also Extended Data Fig. 6a,b for the effect of pG4s on nucleosome exclusion properties in vitro at intergenic and intragenic regions.

G4s carry an intrinsic ability for nucleosome exclusion

Apparent nucleosome-depleted regions (NDRs) are hallmarks of core promoters in eukaryotic cells allowing space for preinitiation complex (PIC) recruitment[18,31]. To understand the link between G4s or pG4s and nucleosome positioning, we performed nucleosome mapping by micrococcal nuclease (MNase) sequencing (MNase-seq) in Raji cells and reanalyzed published datasets for nucleosomes and active epigenetic marks in our two other mammalian models (Supplementary Table 1). Strikingly, we found the center of NDRs overlapping with pG4s at a very large fraction of promoters (~91%), including at the *BTG2* promoter described above (Fig. 2a). We then investigated all active promoters that contain pG4s and confirmed that pG4s are found at the deepest points of NDRs globally (Fig. 2b and Extended Data Fig. 2h). By comparing increasing G4 signals to promoter opening, we observed less opening in the absence of G4access signal (group 1). Conversely, deeper NDRs were observed when G4access was present (groups 2–5), suggesting a threshold effect in the G4access signal. This effect could at least in part be related to the lower GC/CpG content of group 1 (but not the other groups that are more comparable; Extended Data Fig. 2d). Active histone marks increased together with G4 strength (group 1–6), indicating overall chromatin opening and modifications depending on the G4 formation (Fig. 2b). At inactive promoters, the presence of pG4s also hallmarked the Polycomb-deposited histone H3 lysine 27 (H3K27) trimethylated inactive chromatin mark (Extended Data Fig. 2i). This set of promoters carries the hallmark of CGIs, with strong GC and CpG content as expected for the Polycomb signal.

To further formally exclude that pG4-dependent nucleosome exclusion could depend on sequence or transcription factor binding site features in vivo rather than the G4 itself, we performed additional analyses. For this, we combined (1) G4 experimental signal and predictions and cell type-specific analyses, (2) G4 analyses in the presence or absence of NDRs and (3) the interplay between G4 formation and GC box motifs containing motifs, allowing us to reasonably conclude that G4s are involved in nucleosome exclusion (Extended Data Figs. 3 and 4 and Supplementary Note 1).

Next, we tested whether nucleosome exclusion at pG4s was dependent on transcriptional activity. For this, we analyzed nucleosome densities at inactive promoters that were separated into two groups, with or without pG4s. Interestingly, only pG4-containing promoters showed nucleosome exclusion (Fig. 2c). We confirmed that transcription is not required for nucleosome exclusion at pG4s of active promoters by inhibiting Pol II transcription with triptolide without substantial loss of NDRs (Extended Data Fig. 5a). To further assess whether transcription activation was influencing G4 formation, we performed a serum shock, which results in the rapid induction of tens of genes. In these conditions, serum-responsive genes already harbor G4access signals before activation that increase after 15 min of activation (Extended Data Fig. 5b–d). This result is consistent with the previously described chromatin-opening prior activation of constitutively bound serum response factor (SRF) sites[32].

To further infer the direct link between pG4 sequences and chromatin opening, we made use of nucleosomes assembled in vitro with human genomic DNA and recombinant histone[33]. As before, promoters were split in two groups, with or without strong pG4s (Fig. 2d and Extended Data Fig. 5e). This analysis revealed that only the pG4-containing group associates with nucleosome exclusion. Thus, pG4 DNA carries the intrinsic ability to exclude nucleosomes, as in vitro, in the absence of any other transcription factor or proteins, we could observe this property. It also indicates that nucleosome occupancy and G4 formation are mutually exclusive. This ability to exclude nucleosomes is similarly observed for SP1-unbound promoters (Extended Data Fig. 4). Together, our data show that, at both active and inactive promoters, G4s promote intrinsic nucleosome exclusion in vivo and in vitro.

G4-forming sequences are nucleosome organizers

Next, we investigated chromatin opening and organization around G4 predictions at nonpromoter locations, including intergenic regions (IGRs). These locations are mainly not transcribed, and yet pG4s are associated with nucleosome exclusion (Fig. 3a), indicating that these sequences carry this property over the whole genome and not only at promoters. Further investigation of in vitro reconstituted chromatin[33] at all genomic locations including promoters and intergenic and intragenic regions demonstrated their intrinsic abilities to evict nucleosomes globally (Extended Data Fig. 6a,b). Nucleosome exclusion was found at over 90% for promoter pG4s and over 80% for intragenic and intergenic pG4s (Extended Data Fig. 6b). Interestingly, the regions showing the largest nucleosome exclusion areas also showed the highest G4 prediction densities (upper part of the heatmaps), possibly because their presence in clusters increases the chances of G4 formation. These observations are consistent with the high stability of G4 ssDNA structures[34,35], which makes them incompatible with nucleosome formation. To test that in vitro nucleosome exclusion specifically relies on G4 formation and not just on the GC content, we used dimethyl sulfate sequencing and G4-seq data to monitor the ability of a given genomic sequence to form a G4 in vitro in either the double-stranded DNA or ssDNA context, respectively[26,36]. Intergenic pG4s (G4H2.0) and controlled regions of the exact same GC distributions but with low G4H scores (G4H < 1) were used in this comparison (Extended Data Fig. 6c). Strikingly, pG4s showed high dimethyl sulfate footprinting and G4-seq signals and were largely associated with nucleosome exclusion in vitro, while sites unable to fold into G4 were associated with nucleosome occupancy (Extended Data Fig. 6d).

To infer how nucleosomes behave around G4-forming sequences, we analyzed both nucleosome densities and midpoints around intergenic G4 predictions in our model cell lines. Midpoint analyses allowed us to better assess whether given sequence locations display positioning properties (Fig. 3a). This clearly revealed a high level of positioning associated with pG4s in the model cells. The periodicity of nucleosome positioning is highly similar to nucleosomal organization around specific pioneer transcription factors[37] or the CCCTC-binding factor (CTCF) insulator[38], showing an almost identical nucleosomal repeat length (Fig. 3b,c), although with a more modest amplitude, in line with previous observations[39]. Hence, our results highlight the association of pG4s with open chromatin regions at promoters and at IGRs where pG4s also associate with nucleosome array organization. Although experimental G4 signals correlate with transcription, pG4- and G4-driven nucleosome depletion appears independent of transcription. This observation is consistent with results described recently by us and others[24,40]. Our results support a new and previously unappreciated role for G4s as global chromatin organizers at transcribed and untranscribed regions.

Increased chromatin opening at CGIs containing pG4s

As CGIs are able to promote nucleosome depletion at promoters[8], we wondered what was the contribution of pG4s in this process. To address this question, we considered all human CGI annotations containing or not containing strong G4 predictions (Fig. 4). We analyzed genomic features associated with experimental G4s: nucleosome positioning, active chromatin marks, transcription and promoter activity. As expected, G4access and ChIP–seq show stronger signals at CGIs harboring pG4s. In addition, active histone marks (histone H3 lysine 4 (H3K4) trimethylation (H3K4me3) and H3K27 acetylation (H3K27ac)) and Pol II are also enhanced in this class (Fig. 4a,b). Furthermore, nucleosome occupancy exhibits wider and deeper chromatin opening at pG4-containing CGIs. Finally, analysis of promoter activity by the SURE assay confirmed that G4-containing CGIs have ~1.5 times higher promoter activity. To further validate our results, we confirmed our analyses on a set of CGIs of the same length and CG content (Extended Data Fig. 7a) with a more stringent selection of non-G4-forming sequences. G4-forming sequences

https://doi.org/10.1038/s41588-025-02263-6

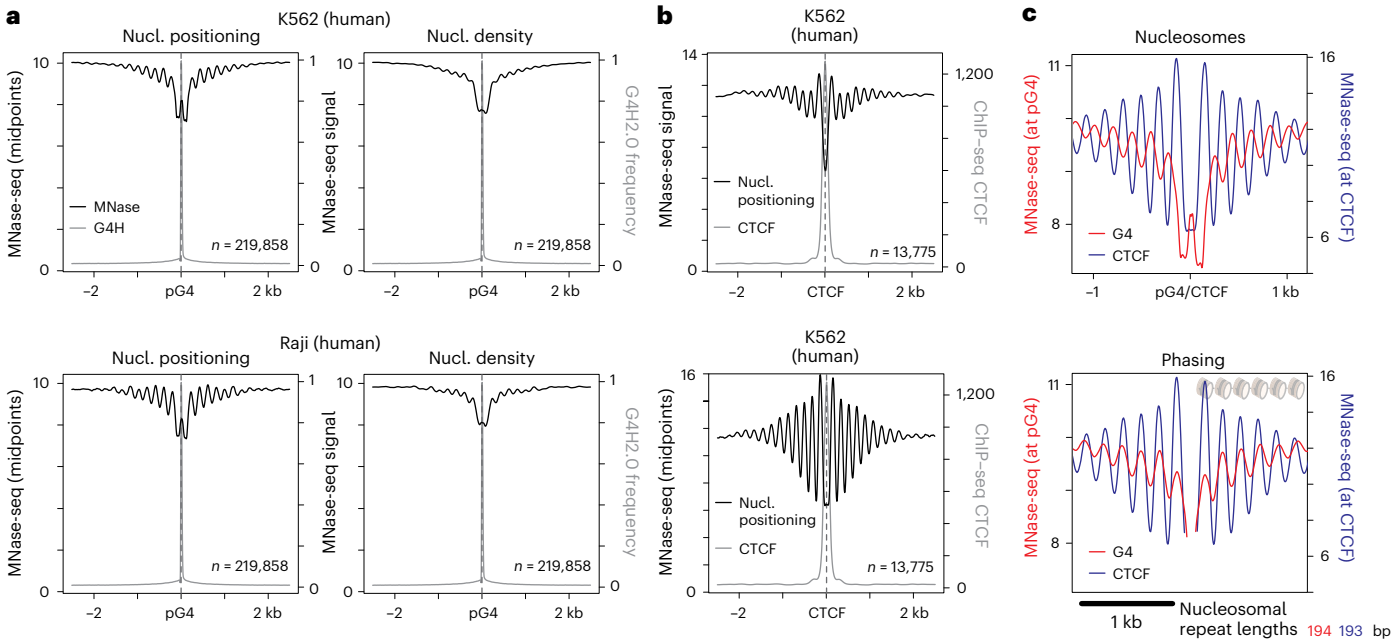


Fig. 3 | pG4s organize nucleosomes at IGRs. a, pG4s mark the center of organized nucleosome arrays at IGRs in K562 (ENCODE) and Raji (this study) cells. All G4H2.0 predictions from IGRs are shown. Heatmaps of nucleosome organization mapped by MNase-seq are displayed. Nucleosome positioning (left; Methods), MNase-seq density (right). **b**, Nucleosomal organization at CTCF sites in K562 cells. Metaprofiles of MNase-seq signal (top), positioning (bottom) and CTCF ChIP–seq (GSE30263) at the 13,775 identified binding sites in K562 cells. **c**, Nucleosomal organizations at CTCF and pG4 show similar nucleosome phasing at IGRs. Top: overlay of metaprofiles of MNase-seq at IGRs centered either on pG4s or CTCF ChIP–seq sites in K562 cells. Bottom: the two nucleosomes surrounding CTCF or pG4 sites (−1 and +1) were aligned to compare the similarity of nucleosome phasing (observed nucleosome repeat lengths of 193 and 194 nucleotides, respectively).

were considered with a G4H score > 1.5 and non-forming sequences with a G4H score < 1.2. This analysis confirmed the association of G4-forming sequences with more open chromatin and active transcription and epigenetic marking within CGIs (Extended Data Fig. 7b,c). Altogether, our observations support the notion that features characteristic of promoters are enhanced in the presence of both experimental G4s and pG4s at CGIs and that G4s might represent essential determinants of CGIs' ability to exclude nucleosomes.

G4 promoter mutations decrease transcription in single cells

To fully demonstrate that G4s are promoter elements that extend the concept of the core promoter, we mutated G4s in model promoters. We inserted G4-containing mouse promoters in human HeLa cells using the Flp-In system[41] and upstream of a reporter containing 256 × MS2 repeats allowing single-cell measurement of transcription by single-molecule fluorescence in situ hybridization (smRNA FISH) (Fig. 5a). We chose promoters with a strong G4 prediction that was verified experimentally using G4access[24] or G4 cleavage under targets and tagmentation[42] in mouse embryonic stem cells (Extended Data Fig. 8a,b). We also observed that most of these promoters displayed a signal in ssDNA-seq experiments in vitro and, for some of them, in vivo in mouse T cells[39]. For three of five models (*Taok1*, *Pkm* and *Klf6*), the sequences did not contain any TATA box, and, for three of them, there was no SP1 or GC box site (*Taok1*, *Pkm* and *Klf6*). We designed G4 mutations that minimally affected the promoter primary sequence while impairing the G4 potential (Fig. 5b, Supplementary Tables 3 and 4, Extended Data Fig. 8c and Supplementary Note 2). For *Eef1a1*, one mutant broke the G stretches with A (G4mut) and the other mutant with C (G4mut2), respecting the GC content of the sequence. To quantify the transcriptional output of the model promoters, we performed single-cell measurements using smRNA FISH over hundreds of cells. The data presented in Fig. 5c show G4 mutants' transcription as compared with that of their wild-type (WT) promoter counterparts. The level of nascent RNA reduction ranked from 23% to 80% in the mutants, indicating that the G4 mutations had a significant effect on transcription of the core promoter. Interestingly, inverting the G4 orientation in the promoter did not impact its activity in the case of the *Eef1a1* model (Eef1a1inv), suggesting that the G4 functions as a promoter element in an orientation-independent manner.

Together, our data show that G4 mutations impair transcription quantitatively and establish that G4-forming sequences most likely function as promoter elements in an orientation-independent manner.

Modeling transcription with G4 and TATA box elements

To decode the direct effects of G4s as promoter elements on transcription, as compared to those of the TATA box, we performed further mutational analyses in the context of the *Eef1a1* promoter, which contains both a canonical TATA box and a stable G4. We mutated the G4 (using G4mut), the TATA box or both elements (Fig. 6a). We first analyzed their ability to promote transcription, to recruit the PIC and to assemble chromatin. Our smRNA FISH of MS2 reporter results showed that both G4 and TATA elements are required for full activity (Fig. 6b). We found that pG4 mutation resulted in more pronounced effects (~60%), while less impact of the TATA box was observed, and an additive or synergic effect seemed at play in the double mutant.

To infer how the G4 and the TATA box influence promoter activity and transcription, we analyzed nucleosome organization and PIC assembly at these promoters in bulk assays. The MNase-PCR assay revealed that G4 mutation led to increased nucleosome density at the NDR location, which also corresponds to the G4 structure coordinate (Fig. 6c). Remarkably, this effect was not observed in the TATA box mutant. The double mutant showed synergic effects of the elements, consistent with nascent transcription data. These results suggest that the pG4 is the main element controlling *Eef1a1* promoter opening and that the TATA box influence could only be seen in the G4mutTATA-mut context. We next monitored PIC recruitment using Pol II and TBP

https://doi.org/10.1038/s41588-025-02263-6

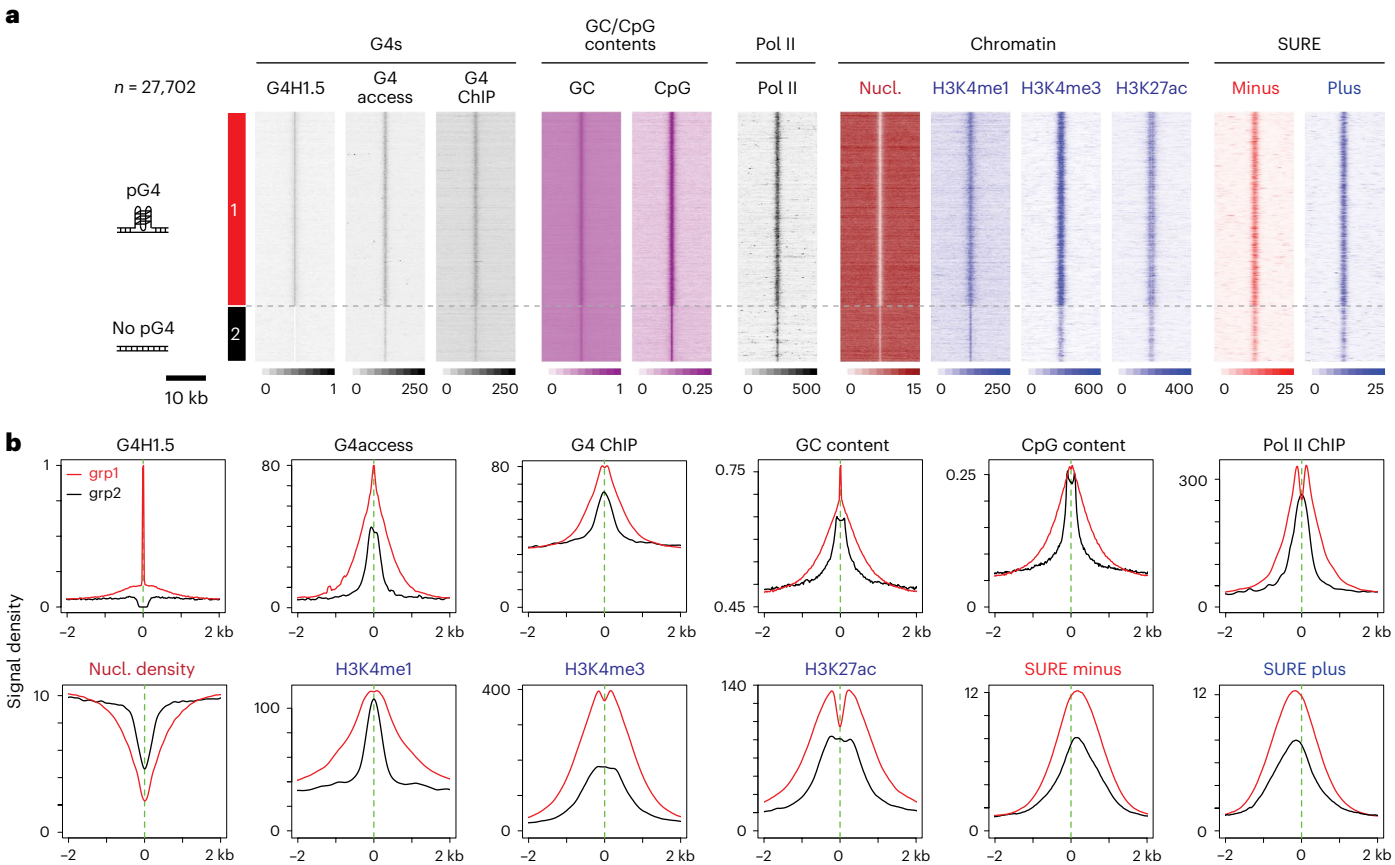


Fig. 4 | pG4s and experimental G4s hallmark nucleosome exclusion at CGIs. **a**, CGIs with pG4s have enhanced chromatin opening, histone modifications and transcription activity. Heatmaps of the 27,702 human CGIs were split into two groups with (n = 21,520) or without (n = 6,182) pG4 (G4H1.5) annotations. Corresponding signals for pG4, G4access, G4 ChIP, GC and CpG content, Pol II, nucleosomes (MNase-seq), chromatin marks and SURE promoter activity are shown as indicated. H3K4me1, H3K4 monomethylation. **b**, CGIs with pG4s have a deeper chromatin opening, increased active histone modifications and promoter activity. Metaprofiles of all marks at CGIs with or without pG4 displayed in the heatmaps from **a** are displayed in a 2-kb window. Further selection and controls for this analysis are presented in Extended Data Fig. 7.

ChIP–quantitative PCR (qPCR) assays. Interestingly, while Pol II recruitment was reduced in all mutants, TBP binding was only impaired when the TATA box was mutated but not in the G4mut (Fig. 6d). We conclude that the primary effect of the pG4 mutation is to restrict chromatin and Pol II accessibility but not TBP recruitment, while the TATA box mutation affects TBP and ultimately Pol II recruitment (Fig. 6a–d).

Next, we analyzed *Eef1a1* promoter dynamics and transcription in living single cells by imaging a green fluorescent (GFP)–MS2 coat protein (MCP) green fusion protein. We quantified transcription sites (TSs) in hundreds of cells including permissive (ON) and inactive (OFF) periods (Extended Data Fig. 9a, Fig. 6e and Supplementary Note 3). These analyses clearly show that both TATA and G4 mutants displayed shorter ON and longer OFF periods. Moreover, OFF periods were slightly longer for the G4 mutant than for TATAmut (Fig. 6e,f). To complete this picture, we also recorded short movies at high temporal resolution to model the entire dynamics of the *Eef1a1* promoter (Fig. 6g). By modeling the distribution of initiation events and interpolating the signal intensities (Extended Data Fig. 9b)[43], we were able to show that the WT promoter is described by a two-state promoter model (ON and OFF). By contrast, both G4 and TATA mutants require a three-state model to describe the experimental data, with the additional promoter state being a long-lived inactive state (~2-h lifetime; Fig. 6h). Given the results of the biochemical analysis of the mutant promoters, the additional state likely corresponds to a TBP-unbound state for the TATAmut, and a closed chromatin state for G4mut. Moreover, examination of transcription initiation rates derived from the models (k_3) indicates a slower transition into the processive elongation state for the G4 mutant (0.3 s versus 0.15 s). These live-cell kinetic data are consistent with the idea that G4 mutations severely limit chromatin opening. In addition, changes in k_3 (Fig. 6h and Extended Data Fig. 9b,c) suggest that G4 mutations also slow down promoter escape and/or Pol II pause–release (Fig. 6h). We also analyzed *Eef1a1* G4mut2 and observed a comparable increase in OFF times (Extended Data Fig. 9d,e).

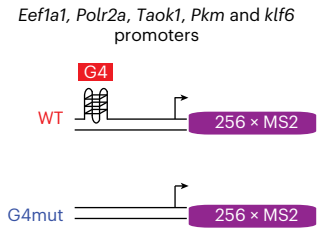
Our observations on the model *Eef1a1* promoter (summarized in Fig. 6h) lead us to propose that the role of G4s at promoters is to promote and maintain nucleosome exclusion as a prerequisite for stable PIC recruitment. By contrast, when the TATA box is mutated, formation of an active promoter state is also impacted due to defects in TBP and PIC recruitment. Collectively, our reporter experiments in bulk and single cells further support the role of G4s and pG4s as promoter elements, conditioning nucleosome exclusion and the rates of promoter transition toward an active state competent for transcription.

G4 stabilization results in Pol II promoter pause–release

To investigate the influence of G4 stabilization on transcription at the global level, we treated human Raji cells with pyridostatin (PDS), a well-known G4 ligand[44]. This ligand stabilizes G4 structures by limiting the transition from folded G4 to unstructured ssDNA or double-stranded B-DNA[44]. Therefore, because promoters are highly enriched in pG4s, their stabilization could either positively or negatively impact transcription. To address this question, we monitored PDS effects on Pol II densities and nascent transcripts, using ChIP–seq, chromatin-associated

https://doi.org/10.1038/s41588-025-02263-6

a



b

		Score	Length	IDS	TDS	CD	Conclusion
Eef1a1	WT	2.14	59	Yes	Yes	Yes	G4
	G4mut	1.05	59	No	No	NC	No G4
	G4mut2	0.69	59	ND	ND	ND	No G4*
	G4inv	**2.14**	**59**	Yes	Yes	Yes	G4
Polr2a	WT	1.62	37	Yes	Yes	Yes	G4
	G4mut	0.81	37	No	No	No	No G4
	G4mut2	0.43	37	No	No	No	No G4
Taok1	WT	2.22	37	Yes	Yes	Yes	G4
	G4mut	0.65	37	No	No	?	No G4
Pkm	WT	1.93	27	Yes	Yes	Yes	G4
	G4mut	0.44	27	No	No	?	No G4
Klf6	WT	2.29	41	Yes	Yes	Yes	G4
	G4mut	0.56	41	No	No	?	No G4

c

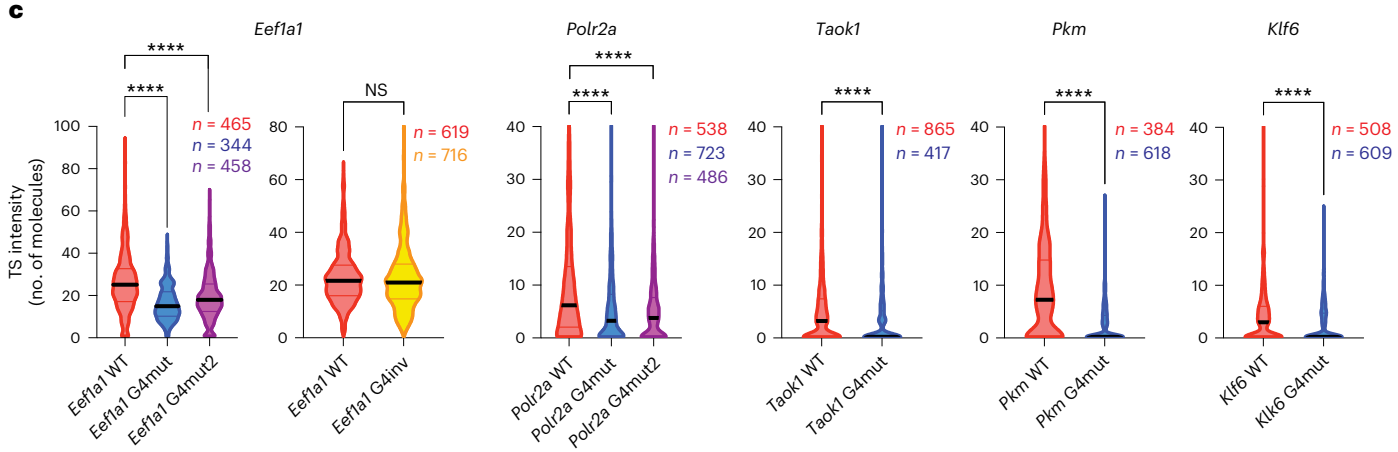


Fig. 5 | G4 mutations impair promoter activity in single cells. a, Scheme of integrated reporter constructs for the indicated mouse G4-containing promoters used for smRNA FISH. These constructs were integrated in the human genome (HeLa) using a Flp-In strategy. Experimental G4 signals for the mouse promoters in embryonic stem cells and in vitro are shown in Extended Data Fig. 8a,b. **b**, Sequence features and G4 structural assessment in the model promoters indicated in **a**. G4 scores determined by the G4H algorithm, reflecting stability and likelihood of formation, are indicated for WT and mutant sequences (see Supplementary Table 3 for the individual sequences). Three independent assays were performed to make conclusions about G4 formation in vitro on the oligonucleotide (last column) with the exception of *Eef1a1* (G4mut2) indicated by an asterisk, which was only assessed by the fluorescence resonance energy transfer melting competition assay (FRET-MC). Sequences of all oligonucleotides are provided in Supplementary Table 3. CD, circular dichroism; IDS, isothermal difference spectra; TDS, thermal difference spectra; NC, not conclusive; ND, not determined. **c**, Quantification of smFISH images of MS2 reporter activity of *Ee1fa1*, *Pkm*, *Klf6*, *Taok1* and *Polr2a* WT and mutant promoters. Data are shown as violin plots in which thick bars are medians and thin bars are quartiles. Mann–Whitney two-sided tests were used (not significant (NS), $P > 0.05$; ****$P \leq 0.0001$). For *Polr2a*, the two mutants have moderate (G4mut) or strong mutations (G4mut2), while *Eef1a1* (G4mut2) has a balanced GC content as compared with G4mut (**b**). Sequences of all promoters are provided in Supplementary Table 4.

RNA sequencing (chrRNA-seq) and transient transcriptome sequencing (TT-seq). We used short time points of treatment (10–60 min) to avoid indirect effects[44,45] resulting from the appearance of double-strand breaks at later time points (Extended Data Fig. 10a). In these assays, we observed that PDS led to changes in Pol II profiles after only 10 min of treatment, with increased signal in gene bodies while the signal at promoters was decreased. Examples in Fig. 7a illustrate this effect for model genes in which Pol II release is illustrated by either a decrease in promoter density, an increase in gene body or both. These effects most likely reflect a general promoter-proximal pause–release of Pol II. We confirmed a global decrease in Pol II pausing by computing apparent pausing scores at the genome-wide level (Fig. 7b and Extended Data Fig. 10b). The effect was found more pronounced for a subset of 602 genes, enriched in stress and RNA-processing functions (Extended Data Fig. 10c). Consistent with pause–release, Pol II average profiles at this subset showed a decrease around TSSs and an increase in Pol II density over gene bodies (Fig. 7c–e). This was further confirmed by nascent RNA-seq analyses (Extended Data Fig. 10d,e). Because PDS treatment was short and the average speed of Pol II did not allow it to cross all gene bodies during the first 10 min, we analyzed Pol II changes on long genes (Fig. 7f). This revealed a clear pause–release effect of PDS over the first 27 kb, consistent with an elongation rate around 2.5–3 kb min^{-1}, as previously described[46]. Furthermore, and strikingly, a clear correlation appeared between the different levels of pause–release and the G4 potential of the corresponding promoters (Extended Data Fig. 10g). At later time points, reduced pausing remained up to 30 min while the impact of PDS was reversed after 60 min over gene bodies (Extended Data Fig. 10f). Finally, G4access and nucleosome dynamics indicated a global G4 increase at all promoters, while nucleosome densities were essentially unaffected (Extended Data Fig. 10h,i). Overall, our observations suggest that G4 stabilization results in an increased ability of Pol II to escape from pause states. This is also consistent with our mathematical modeling of the *Eef1a1* model promoter where G4 mutations affected the efficiency of Pol II released from the promoter.

Discussion

Our study has shown that G4s are highly enriched in mammalian core promoters. We propose a new function for these structural elements, which is to intrinsically exclude nucleosomes. This function could define a previously unappreciated essential property of promoters in vivo. The mechanism of nucleosome exclusion by G4s could be simply explained by the incompatibility of stable ssDNA formation and its incorporation into stable nucleosomes. As G4s are not substantially present in all eukaryotic promoters[47], other secondary structures or sequence contexts could have the same role in other organisms, for example, AT stretches in yeast[48,49]. This G4 property is however in contrast to the previous observation proposing that it is transcription that stimulates G4 formation based on in vitro transcription assays[15] or transcription activation in vivo[14] but is in line with more recent observations[40], including from our laboratory[24]. In this work, we show that transcription inhibition results in maintenance but reduction of G4 signal, indicating that transcription does not precede G4 formation at the promoter but that it does increase or stabilize its structure. Here, using serum shock, we further show that activated genes often show preformed G4s at their promoters, which become further enhanced during activation. In sum, our data support a model in which G4s are formed prior from or in the absence of transcription because pG4s and G4s are detected at NDRs in vitro or at silent promoters and because transcription inhibition does not substantially alter NDRs at pG4 locations. The presence of frequent transcription factor binding site motifs in addition to pG4 elements also opens up the possibility that PIC recruitment in vivo does not directly rely on motifs that recruit general transcription factors but rather on various already bound transcription factors, possibly in combination with coactivator to allow further recruitment of the PIC. Also in support of this model is the observation that TFIID does not contact DNA on its own upstream of TSSs on TATA-less promoters[50,51], which represent the vast majority of promoters, but is rather recruited through direct chromatin contacts[52–54] or via interaction of transcription factors such as NF-Y[55]. Interestingly, recent work has shown that the ATRX–DAXX complex actively suppresses G4 formation and promotes heterochromatin formation, providing a nice duality opposing G4s and chromatin[56].

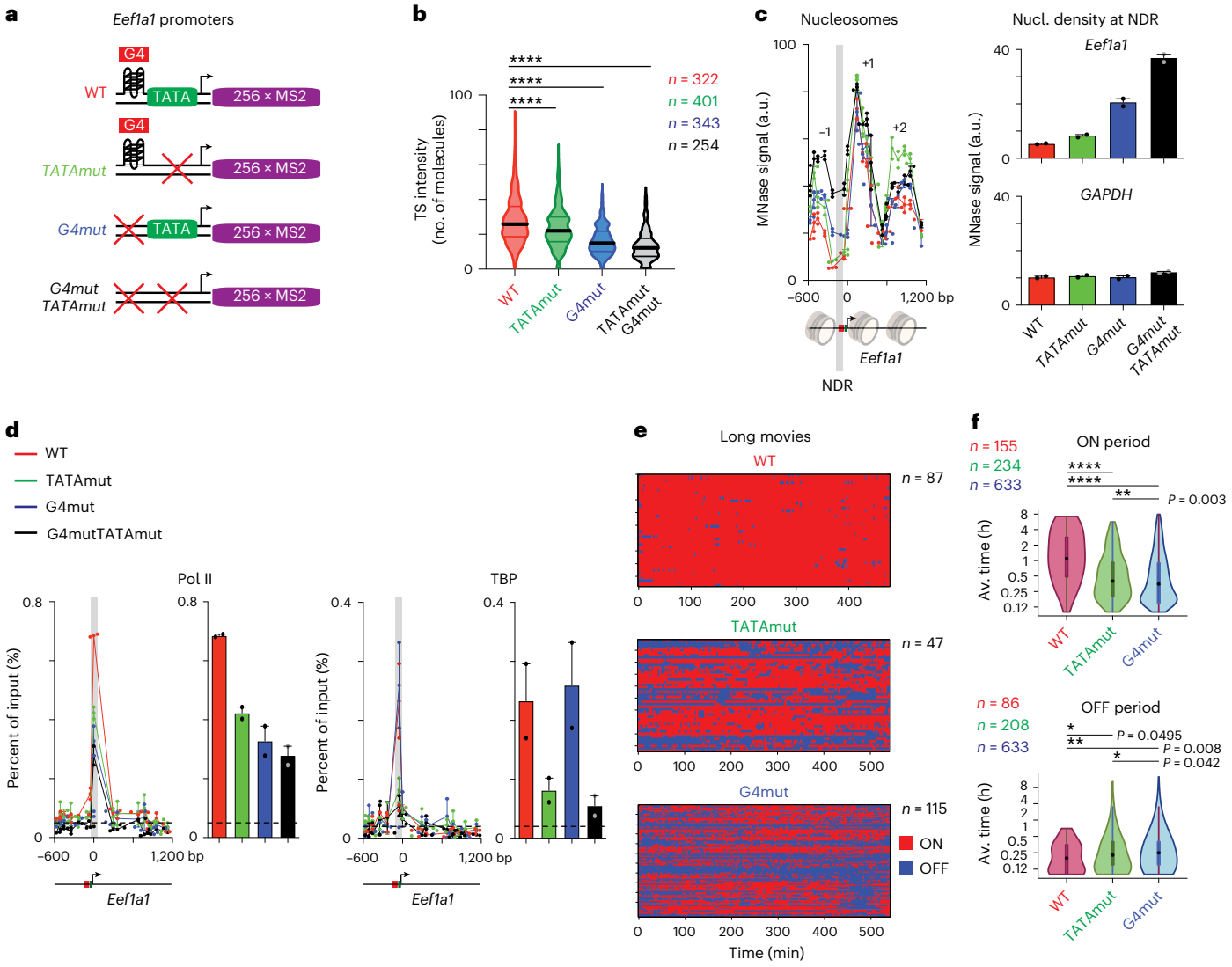
Our data are consistent with a role of G4s favouring pause–release by Pol II. First, modeling transcription at the *Eef1a1* G4 mutant promoter indicates one additional limiting step for transcription (OFF1) likely corresponding to chromatin opening in the absence of G4. Another limiting step includes the transition to productive elongation, which comprises pausing, and also shows a substantial time increase (k_3). Second, stabilizing G4s globally with PDS in vivo resulted in pause–release of genes containing stronger G4s at their promoter. These results are in apparent contrast with recent work indicating that treatment with another G4 ligand reduces transcription initiation[57]. However, the times of treatments performed in this study were much longer, opening up the possibility for more indirect effects, including accumulation of DNA breaks. Overall, our data are in agreement with previous work showing that G4-containing promoters tend to show less poised Pol II[58]. How could G4s facilitate pause–release? Because G4s at the promoter exclude nucleosomes, the presence of one or multiple G4s would favor not only open chromatin but also a premelted template for Pol II. Such structures would potentially facilitate the formation and the extension of the transcription bubble. As a consequence, crossing the +1 nucleosomal barrier would become easier for the Pol II complex, resulting in pause–release. The comparison of experimental G4 signal (G4access) to ssDNA scored by $KMnO_4$ footprinting supports this hypothesis because it indicates a correlation between G4 formation and open complex at active promoters. Finally, and given previous links made between the action of G4 stabilization by PDS, pausing and topoisomerase II (TopII)[59,60] that could relax supercoiling locally by promoting double-stranded breaks at the promoter level, one could also envision that TopII comes into play during that process, even though our data do not indicate accumulation of the histone variant γ-H2AX at early time points.

Another striking property of G4-forming sequences, also visible at experimental G4s determined by G4access, is their ability to position nucleosomes. The nucleosome repeat length observed is in the range of that described for CTCF but also of the strongly positioning nucleosome in vitro[33]. Because G4s tend to be present as clusters in promoters, at these locations, nucleosome positioning is less visible probably due to the presence of multiple G4s introducing fuzziness in the adjacent nucleosomes and to their additive effects. We do not know at this stage whether the positioning property of IGRs is inherent to G4 structure or could be explained by the recruitment of G4 binders. In any of these scenarios, the barrier constituted by the G4s would dock the surrounding nucleosomes. It will thus be interesting in future studies to investigate the precise interplay between the G4 and CTCF, as recent work suggests that they could be locally associated[61,62].

Our data point to a role of the G4 as a driver of CGI's properties, possibly because they yield a more robust and/or constitutive NDR. Over 70% of G4access- or G4 ChIP[25]-enriched areas are actually located in CGIs, and, consistently, CGIs also contain large G4 clusters, increasing the likelihood of their formation locally. Our analyses suggest that pG4s could be one essential determinant of the ability of CGIs to exclude nucleosomes, thus adding a new determinant, aside from GC and CpG content[7], of these essential areas of the genome. Importantly, while our study shows that G4-forming sequences behave as promoter elements by excluding nucleosomes, it does not demonstrate per se that G4 structures are formed in situ in the context of chromatin nor that other structures are not at play where G4s are observed (Supplementary Note 4). Nevertheless, the use of various orthogonal techniques to score for pG4s based on different principles and the pG4 ability to exclude nucleosomes intrinsically pleads for their direct

Fig. 6 | G4 mutations at *Eef1a1* model promoters decrease transcription and increase nucleosome density at NDRs while increasing promoter OFF times in single cells. a, *Eef1a1* mouse model promoters were inserted in HeLa cells with the indicated mutations (see Supplementary Table 3 for sequences). **b**, G-forming sequences and the TATA box regulate gene expression from the *Eef1a1* model promoter. Quantification of MS2 reporter activity by smFISH of WT and mutant promoters (Mann–Whitney two-sided tests were used, $n = 322, 401, 343$ and 254). **c**, *Eef1a1* G4 mutations result in increased nucleosome density at the NDR location. MNase signal of WT and mutant promoters was quantified by qPCR ($n = 2$ biological replicates independently processed: different cultures, times and experiments; mean ± s.e.m). NDR location is highlighted in gray. Quantifications of the nucleosome density signal at the NDR are shown on the right together with *a GAPDH* control (oligonucleotides are in Supplementary Table 5). a.u., arbitrary units. **d**, Pol II and TBP recruitment at the *Eef1a1* model promoter and mutants. ChIP was quantified by qPCR ($n = 2$ biological replicates independently processed: different cultures, times and experiments; mean ± s.e.m). Pol II is affected in all mutant contexts, while TBP is impaired specifically in the TATAmut. **e**, Heatmaps of permissive (red) and nonpermissive (blue) transcription periods for WT and mutant *Eef1a1* promoters in single-cell live imaging (long movies). Each line represents individual cell assessments. Representative movie images and quantification are shown in Extended Data Fig. 9a. **f**, Violin plots of ON and OFF period duration in WT or mutated *Eef1a1* promoters measured in long movies. Computed *P* values (two-sided Mann–Whitney test) are as follows: $^{****}P < 1 \times 10^{-4}$, $^{**}P < 1 \times 10^{-2}$, $^{*}P < 5 \times 10^{-2}$). Boxes represent the median and the first and last quartiles. $n = 155, 234, 633, 78, 208$ and 612. Av., average. **g**, Heatmaps of polymerase densities for WT and mutant *Eef1a1* promoters in single-cell live imaging (short movies). Each line represents individual cell assessments. **h**, Interpretative scheme of the signal deconvolution and time constant analysis derived from the long and short movies. While WT transcription can be described in two main steps, both TATA and G4 mutants require at least three steps. The G4 mutant has limiting OFF1 and ON states (Extended Data Fig. 9b,c). NA, not applicable.

https://doi.org/10.1038/s41588-025-02263-6



https://doi.org/10.1038/s41588-025-02263-6

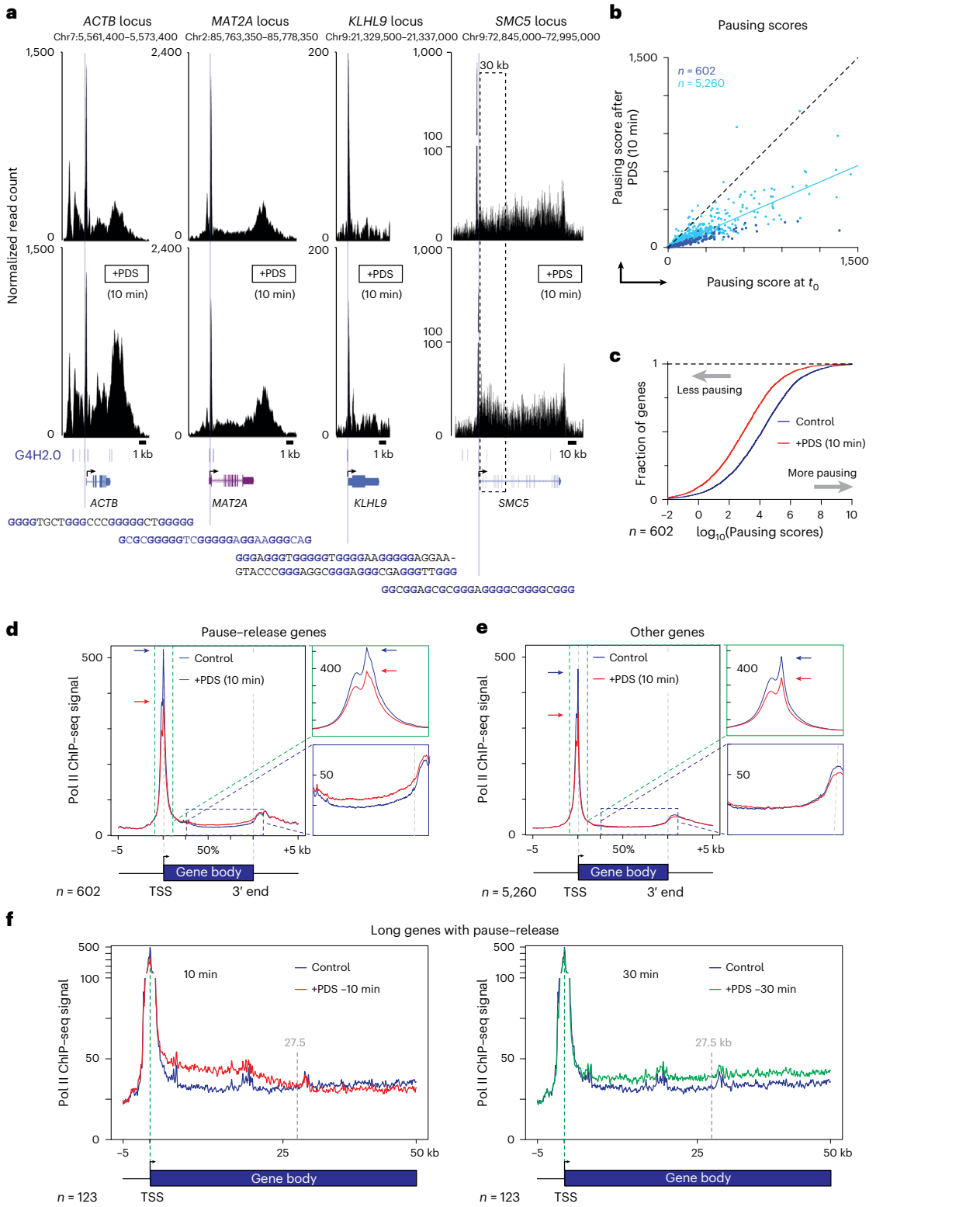


Fig. 7 | Stabilization of G4 by ligand results in global pause–release by Pol II. **a**, PDS treatment (10 min) results in Pol II release from the promoter area at the *ACTB*, *MAT2A*, *KLHL9* (short genes) and *SMC5* (long gene) loci. Tracks of Pol II ChIP–seq (magnified around the TSS) before and after PDS treatment are shown. G4 predictions and the sequence of the pG4 upstream of the TSSs associated with the NDRs are indicated and are shown below the tracks (G4H2.0). **b**, Pausing scores are globally reduced at the genome scale following PDS treatment. A scatterplot comparing pausing scores before and after PDS treatment (10 min) is shown. Genes with enhanced Pol II signal at gene bodies (DESeq, *P* value < 0.05) are highlighted in dark blue. The light blue line represents the linear regression curve of the other points. The whole kinetic analysis is presented in Extended Data Fig. 10b. **c**, Cumulative plots of pausing score dynamics following PDS treatment. **d**, Metaprofiles of Pol II ChIP–seq signal at selected genes with decreased pausing scores ($n = 602$) depict reduced promoter and increased gene body occupancies. **e**, Metaprofiles of Pol II ChIP–seq signal at control genes following PDS treatment ($n = 5{,}557$). **f**, Metaprofiles of Pol II ChIP–seq signal at long genes (10 and 30 min with PDS).

structural involvement as promoter elements rather than protein docking sites on DNA. However, and as suggested by cell-specific and serum induction analyses, G4-associated NDR formation is likely to be regulated by and possibly integrated with other processes. All in all, our work opens a new gate in our understanding and definition of a promoter in vivo and readjusts the existing paradigms. It will also support future work on targeting secondary structures to control their activity using specific ligands in cancer therapy.

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/s41588-025-02263-6.

References

1. Jacob, F., Ullman, A. & Monod, J. The promotor, a genetic element necessary to the expression of an operon. *C. R. Hebd. Seances Acad. Sci.* **258**, 3125–3128 (1964).
2. Pribnow, D. Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. *Proc. Natl Acad. Sci. USA* **72**, 784–788 (1975).
3. Rosenberg, M. & Court, D. Regulatory sequences involved in the promotion and termination of RNA transcription. *Annu. Rev. Genet.* **13**, 319–353 (1979).
4. Butler, J. E. & Kadonaga, J. T. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev.* **16**, 2583–2592 (2002).
5. Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.* **10**, 161–172 (2009).
6. Radman-Livaja, M. & Rando, O. J. Nucleosome positioning: how is it established, and why does it matter? *Dev. Biol.* **339**, 258–266 (2010).
7. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022 (2011).
8. 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).
9. Bansal, M., Kumar, A. & Yella, V. R. Role of DNA sequence based structural features of promoters in transcription initiation and gene expression. *Curr. Opin. Struct. Biol.* **25**, 77–85 (2014).
10. Agarwal, T., Roy, S., Kumar, S., Chakraborty, T. K. & Maiti, S. In the sense of transcription regulation by G-quadruplexes: asymmetric effects in sense and antisense strands. *Biochemistry* **53**, 3711–3718 (2014).
11. 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).
12. David, A. P. et al. G-quadruplexes as novel *cis*-elements controlling transcription during embryonic development. *Nucleic Acids Res.* **44**, 4163–4173 (2016).
13. Smestad, J. A. & Maher, L. J. 3rd Relationships between putative G-quadruplex-forming sequences, RecQ helicases, and transcription. *BMC Med. Genet.* **16**, 91 (2015).
14. Hansel-Hertsch, R. et al. G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* **48**, 1267–1272 (2016).
15. 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).
16. Esain-Garcia, I. et al. G-quadruplex DNA structure is a positive regulator of *MYC* transcription. *Proc. Natl Acad. Sci. USA* **121**, e2320240121 (2024).
17. Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**, 549–561 (1989).
18. Haberle, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell Biol.* **19**, 621–637 (2018).
19. Vo Ngoc, L., Wang, Y. L., Kassavetis, G. A. & Kadonaga, J. T. The punctilious RNA polymerase II core promoter. *Genes Dev.* **31**, 1289–1301 (2017).
20. Bedrat, A., Lacroix, L. & Mergny, J. L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **44**, 1746–1759 (2016).
21. Bagchi, D. N. & Iyer, V. R. The determinants of directionality in transcriptional initiation. *Trends Genet.* **32**, 322–333 (2016).
22. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008).
23. Seila, A. C. et al. Divergent transcription from active promoters. *Science* **322**, 1849–1851 (2008).
24. Esnault, C. et al. G4access identifies G-quadruplexes and their associations with open chromatin and imprinting control regions. *Nat. Genet.* **55**, 1359–1369 (2023).
25. Mao, S. Q. et al. DNA G-quadruplex structures mold the DNA methylome. *Nat. Struct. Mol. Biol.* **25**, 951–957 (2018).
26. Chambers, V. S. et al. High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* **33**, 877–881 (2015).
27. Kouzine, F. et al. Global regulation of promoter melting in naive lymphocytes. *Cell* **153**, 988–999 (2013).
28. van Arensbergen, J. et al. Genome-wide mapping of autonomous promoter activity in human cells. *Nat. Biotechnol.* **35**, 145–153 (2017).
29. Pugh, B. & Venters, B. Genomic organization of human transcription initiation complexes. *PLoS ONE* **11**, e0149339 (2016).
30. Core, L. J. et al. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat. Genet.* **46**, 1311–1320 (2014).
31. Andersson, R. & Sandelin, A. Determinants of enhancer and promoter activities of regulatory elements. *Nat. Rev. Genet.* **21**, 71–87 (2020).
32. Esnault, C. et al. Rho-actin signaling to the MRTF coactivators dominates the immediate transcriptional response to serum in fibroblasts. *Genes Dev.* **28**, 943–958 (2014).
33. Valouev, A. et al. Determinants of nucleosome organization in primary human cells. *Nature* **474**, 516–520 (2011).
34. Guedin, A., Gros, J., Alberti, P. & Mergny, J. L. How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res.* **38**, 7858–7868 (2010).
35. Sen, D. & Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **334**, 364–366 (1988).
36. Umeyama, T. & Ito, T. DMS-seq for in vivo genome-wide mapping of protein–DNA interactions and nucleosome centers. *Cell Rep.* **21**, 289–300 (2017).
37. Barozzi, I. et al. Coregulation of transcription factor binding and nucleosome occupancy through DNA features of mammalian enhancers. *Mol. Cell* **54**, 844–857 (2014).
38. 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).
39. Kouzine, F. et al. Permanganate/S1 nuclease footprinting reveals non-B DNA structures with regulatory potential across a mammalian genome. *Cell Syst.* **4**, 344–356 (2017).
40. Shen, J. et al. Promoter G-quadruplex folding precedes transcription and is controlled by chromatin. *Genome Biol.* **22**, 143 (2021).

41. Tantale, K. et al. A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. *Nat. Commun.* **7**, 12248 (2016).
42. Lyu, J., Shao, R., Kwong Yung, P. Y. & Elsässer, S. J. Genome-wide mapping of G-quadruplex structures with CUT&Tag. *Nucleic Acids Res.* **50**, e13 (2022).
43. Tantale, K. et al. Stochastic pausing at latent HIV-1 promoters generates transcriptional bursting. *Nat. Commun.* **12**, 4503 (2021).
44. Rodriguez, R. et al. A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. *J. Am. Chem. Soc.* **130**, 15758–15759 (2008).
45. Olivieri, M. et al. A genetic map of the response to DNA damage in human cells. *Cell* **182**, 481–496 (2020).
46. 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).
47. Marsico, G. et al. Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res.* **47**, 3862–3874 (2019).
48. Kaplan, N. et al. Nucleosome sequence preferences influence in vivo nucleosome organization. *Nat. Struct. Mol. Biol.* **17**, 918–920 (2010).
49. Segal, E. & Widom, J. What controls nucleosome positions? *Trends Genet.* **25**, 335–343 (2009).
50. Burke, T. W. & Kadonaga, J. T. *Drosophila* TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev.* **10**, 711–724 (1996).
51. Parry, T. J. et al. The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery. *Genes Dev.* **24**, 2013–2018 (2010).
52. Vermeulen, M. et al. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**, 58–69 (2007).
53. Bhuiyan, T. & Timmers, H. T. M. Promoter recognition: putting TFIID on the spot. *Trends Cell Biol.* **29**, 752–763 (2019).
54. Muller, F. & Tora, L. Chromatin and DNA sequences in defining promoters for transcription initiation. *Biochim. Biophys. Acta* **1839**, 118–128 (2014).
55. Frontini, M. et al. NF-Y recruitment of TFIID, multiple interactions with histone fold TAF_{II}s. *J. Biol. Chem.* **277**, 5841–5848 (2002).
56. Teng, Y. C. et al. ATRX promotes heterochromatin formation to protect cells from G-quadruplex DNA-mediated stress. *Nat. Commun.* **12**, 3887 (2021).
57. Li, C. et al. Ligand-induced native G-quadruplex stabilization impairs transcription initiation. *Genome Res.* **31**, 1546–1560 (2021).
58. Dao, P., Wojtowicz, D., Nelson, S., Levens, D. & Przytycka, T. M. Ups and downs of poised RNA polymerase II in B-cells. *PLoS Comput. Biol.* **12**, e1004821 (2016).
59. Bossaert, M. et al. Transcription-associated topoisomerase 2α (TOP2A) activity is a major effector of cytotoxicity induced by G-quadruplex ligands. *eLife* **10**, e65184 (2021).
60. Herrero-Ruiz, A. et al. Topoisomerase IIα represses transcription by enforcing promoter-proximal pausing. *Cell Rep.* **35**, 108977 (2021).
61. Tikhonova, P. et al. DNA G-quadruplexes contribute to CTCF recruitment. *Int. J. Mol. Sci.* **22**, 7090 (2021).
62. Wulfridge, P. et al. G-quadruplexes associated with R-loops promote CTCF binding. *Mol. Cell* **83**, 3064–3079 (2023).

Methods

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

Cell lines and culture

Data presented in this article were issued from the analysis of human cell lines (K562, Raji, HeLa) or mouse cell lines (primary thymocytes ($CD4^+CD8^+$ (DP) and NIH-3T3). Original data presented concern essentially Raji, HeLa and NIH-3T3 cells, but all cellular models are described in this section.

K562 is a pseudotriploid ENCODE Tier I erythroleukemia cell line derived from a female (aged 53 years) with chronic myelogenous leukemia. Raji cells are lymphoblast-like cells from a male (aged 11 years) with Burkitt's lymphoma. HeLa Flp-In (Invitrogen) H9 cells (a kind gift from S. Emiliani, Institut Cochin, Paris, France) is a cell line derived from the parent HeLa line. The HeLa line was derived from a female (aged 31 years) with adenocarcinoma. Mouse $CD4^+CD8^+$ DP cells were sorted from the thymuses of 5–6-week-old mice as described[8,63]. NIH-3T3 is a fibroblast cell line that was isolated from a mouse NIH Swiss embryo.

Raji cells were grown in RPMI 1640 medium supplemented with 10% fetal calf serum, penicillin–streptomycin (100 U l^{-1}) and glutamine (2 mg l^{-1}) at 37 °C with 5% CO_2. For triptolide transcription inhibition experiments (Extended Data Fig. 5a), cells were treated with 1 µM triptolide for 2 h as described[24]. For the PDS experiments (Fig. 7 and Extended Data Fig. 10), cells were treated with 10 µM PDS at the indicated times.

HeLa Flp-In H9 cells (a kind gift of S. Emiliani) used for reporter assays (Figs. 5 and 6 and Extended Data Fig. 9) were maintained in DMEM medium supplemented with 10% fetal calf serum, penicillin–streptomycin (100 U l^{-1}) and glutamine (2.9 mg l^{-1}) at 37 °C with 5% CO_2. HeLa cells with integrated constructs were transfected with plasmids using jetPRIME (Polyplus), following manufacturer recommendations.

NIH-3T3 cells were grown in DMEM supplemented with 10% fetal calf serum, penicillin–streptomycin (100 U l^{-1}) and glutamine (2.9 mg l^{-1}) at 37 °C with 5% CO_2. Cells at 50% confluence were serum starved overnight (0.3% fetal calf serum) and stimulated with 15% fetal calf serum for 15 min before cell collection for G4access and chrRNA-seq experiments (Extended Data Fig. 5b–d).

Genome-wide datasets

All datasets used in this study including published and original experiments are described in Supplementary Table 1. All GEO accession numbers are included. The GEO accessions for specific experiments related to this study are recorded under GSE159276.

MNase, MNase-seq and DNase-seq

For sequencing of nucleosomal DNA in Raji cells, 3.5×10^7 cells were resuspended in 350 µl solution I (150 mM sucrose, 80 mM KCl, 5 mM K_2HPO_4, 5 mM $MgCl_2$, 0.5 mM $CaCl_2$, 35 mM HEPES, pH 7.4), and NP-40 was added at a final concentration of 0.2%. Cell membranes were permeabilized for 5 min at 37 °C. MNase was prepared at 50, 25, 12, 6 or 3 U in 0.5 ml solution II (150 mM sucrose, 50 mM Tris, pH 8, 50 mM NaCl, 2 mM $CaCl_2$) and incubated with 50 µl of a cellular preparation, corresponding to 5×10^6 cells, for exactly 10 min at 37 °C. The reactions were stopped by adding EDTA at a final concentration of 10 mM. The cells were lysed using 1.45 ml SDS lysis buffer (1% SDS, 10 mM EDTA, pH 8, 50 mM Tris, pH 8), with a 10-min incubation at 4 °C. Aliquots (200 µl) were taken for purification, and the remaining extracts were stored at −80 °C. An equal volume of TE buffer (200 µl) was added to the aliquots, followed by subsequent 2-h treatments with 0.2 µg ml^{-1} RNase A and proteinase K at 37 °C and 55 °C, respectively. DNA was extracted by two subsequent phenol–chloroform–isoamyl alcohol (25:24:1) extractions and further purified using QIAquick PCR purification columns (Qiagen). Nucleosomal digestion was verified by running 500 ng DNA on a 1.5% agarose gel as well as on DNA high-sensitivity 2100 Bioanalyzer chips (Agilent). Digestions showing 75% of mononucleosomes (running at 150 bp) were selected for library preparation. Fragments below 250 bp were purified with AMPure XP Beads (Beckman Coulter) following manufacturer instructions. Libraries were prepared with the TruSeq ChIP Library Preparation Kit (Illumina) and sequenced on HiSeq 2000 or 4000 sequencers (Illumina).

Chromatin analysis by MNase treatment of HeLa cells (Fig. 6c) was performed as follows, as adherent cells collected with trypsin tend to clump using the method described above. HeLa cells were collected using trypsin and washed twice with ice-cold PBS. Cells were resuspended in 250 µl of ice-cold Nuclei Buffer I (15 mM Tris-HCl, pH 7.5, 300 mM sucrose, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EGTA, 0.5 mM DTT, 0.1 mM PMSF, 3.6 µg ml^{-1} aprotinin) before adding 250 µl of ice-cold Nuclei Buffer II (15 mM Tris-HCl, pH 7.5, 300 mM sucrose, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EGTA, 0.5 mM DTT, 0.1 mM PMSF, 3.6 µg ml^{-1} aprotinin, 0.4% IGEPAL CA-630). Extracts were incubated for 10 min on ice and layered on a 1-ml Nuclei Buffer III–sucrose cushion (15 mM Tris-HCl, pH 7.5, 1.2 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EGTA, 0.5 mM DTT, 0.1 mM PMSF, 3.6 µg ml^{-1} aprotinin). Nuclei were isolated by centrifugation at 10,000*g* for 20 min at 4 °C and were resuspended in 600 µl MNase digestion buffer (50 mM Tris-HCl, pH 7.5, 320 mM sucrose, 4 mM $MgCl_2$, 1 mM $CaCl_2$, 0.1 mM PMSF) and incubated on ice for 3 min. MNase was added for exactly 10 min at 37 °C, using 50, 25, 12, 6 or 3 U of the enzyme. The reactions were then stopped by adding EDTA at a final concentration of 10 mM. SDS lysis buffer (100 µl) was added (1% SDS, 10 mM EDTA, pH 8, 50 mM Tris, pH 8), and, after 10 min of incubation at 4 °C, samples were processed for DNA purification using phenol–chloroform extraction followed by ethanol precipitation. qPCR quantification was performed by using the primers described in Supplementary Table 5.

DNase I analysis was performed on Raji cells as previously described[64] with minor modifications. In brief, nuclei were prepared from 5×10^6 Raji cells per sample and incubated for 10 min at 37 °C with limiting concentrations of the DNA endonuclease DNase I (Sigma) supplemented with Ca^{2+} and Mg^{2+}. A dose–response experiment of 16 samples was performed with a 2× dilution of DNase I between samples in cascade. Digestion was stopped by adding EDTA, and the samples were treated with proteinase K. Small fragments (<500 bp in length) were recovered using an agarose gel from the sample in which a smear just started to appear using the Qiagen gel purification kit, and libraries were prepared using the Illumina TruSeq kit.

G4access

The complete G4access procedure is described in ref. 24, and the principle of the method is summarized in Fig. 1c. In short, K562 and NIH-3T3 cells were pelleted and rinsed twice with PBS. For each experiment, 5×10^6 cells per titration point were resuspended in 50 µl of prewarmed permeabilization buffer (150 mM sucrose, 80 mM KCl, 5 mM KH_2PO_4, 5 mM $MgCl_2$, 0.5 mM $CaCl_2$ and 35 mM HEPES, pH 7.4) supplemented with 0.2% (vol/vol) NP-40 and incubated for 5 min at 37 °C before digestion. MNase digestions were then performed by adding 500 µl of prewarmed MNase reaction buffer (150 mM sucrose, 50 mM Tris-HCl, pH 8, 50 mM NaCl and 2 mM $CaCl_2$) supplemented with either 3, 6, 12, 25 or 50 U MNase (Merck, 10107921001). Digestions were incubated at 37 °C for 10 min and stopped on ice and by adding 11 µl of 500 mM EDTA to each reaction. Samples were then incubated for 10 min on ice with 550 µl SDS lysis buffer (1% (vol/vol) SDS, 10 mM EDTA and 50 mM Tris-HCl, pH 8). Before DNA purification, 1 ml of water was added to dilute the SDS, and the samples were incubated with 5 µl RNase A (Thermo Fisher, EN0531) at 37 °C for 2 h and with 8 µl proteinase K (Euromedex, 09-0911) at 56 °C for 2 h to complete the lysis. To then perform quality control of the MNase digestions, 125 µl of each sample was cleaned up using the QIAquick PCR Purification Kit (Qiagen, 28106) and assessed with an agarose gel and the Bioanalyzer. At this step, for efficient G4access, samples should present ~30% (±5%)

of mononucleosomes. The remaining samples were then purified by phenol–chloroform extraction and ethanol precipitation for subsequent steps. The 0–100-bp size-selected fragments from MNase digestions that had ~30% of mononucleosomes were subjected to DNA library preparation. Paired-end libraries were constructed using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, E7645S) using a starting material of 50 ng. DNA fragments were end-repaired, A-tailed and ligated with Illumina-compatible adaptors. Cleanup of adaptor-ligated DNA was performed by using CleanNGS beads (CNGS-0050) with a bead/DNA ratio of 2/1. The purified products were amplified with eight cycles of PCR. Finally, samples were cleaned up with a bead/DNA ratio of 0.8/1 to remove the free sequencing adaptors. Libraries were sequenced on the Illumina NextSeq 500 sequencer using paired 50–30-bp reads. G4access data were deposited at the GEO database under GSE159276.

ChIP–seq and ChIP–qPCR

Fifty million cells were used to perform each Pol II ChIP–seq experiment. Cells were cross-linked for 10 min at 20 °C with cross-linking 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 2.5 ml of cold LB1 buffer (50 mM HEPES, pH 7.5, 140 mM NaCl, 1 mM EDTA, pH 8, 10% glycerol, 0.75% NP-40, 0.25% Triton X-100) at 4 °C for 20 min on a rotating wheel. Nuclei were pelleted by centrifuging at 1,350*g* in a refrigerated centrifuge and washed with 2.5 ml LB2 buffer (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 buffer (1 mM EDTA, pH 8, 0.5 mM EGTA, pH 8, 10 mM Tris, pH 8, 100 mM NaCl, 0.1% sodium deoxycholate, 0.5% *N*-lauroylsarcosine) and sonicated using the Bioruptor Pico device (Diagenode) in 15-ml tubes for 20 cycles of 30-s ON and 30-s OFF pulses in a bath at 4 °C. All buffers (LB1, LB2 and LB3) were supplemented with EDTA-free Protease Inhibitor Cocktail (Roche), 0.2 mM PMSF and 1 µg ml^{-1} pepstatin just before use. After sonication, Triton X-100 was added at a final concentration of 1% followed by centrifugation at 20,000*g* and 4 °C for 10 min to remove particulate matter. After setting aside a 50-µl aliquot to serve as input and to analyze fragmentation, chromatin was aliquoted, 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× elution buffer (100 mM Tris, pH 8.0, 20 mM EDTA, 2% SDS) and incubated at 65 °C for 12 h for reverse cross-linking. 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 µg ml^{-1}) at 37 °C for 1 h and with proteinase K (0.2 µg l^{-1}) 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 the 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 N20 (Santa Cruz, sc-899x, lot H3115)- and TBP N12 (Santa Cruz, sc-204, lot LO214)-specific antibodies (5 µg per immunoprecipitation). Sonicated chromatin (1 ml) was added to precoated beads (250 µl), 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% NP-40, 0.7% sodium deoxycholate, 1× protease inhibitor cocktail), followed by one wash with TE-NaCl buffer (10 mM Tris, pH 8 and 1 mM EDTA, pH 8, 50 mM NaCl) and a final wash with TE buffer (10 mM Tris, pH 8 and 1 mM EDTA, pH 8). Immunoprecipitated chromatin was eluted by two sequential incubations with 50 µl elution buffer (50 mM Tris, pH 8, 10 mM EDTA, pH 8, 1% SDS) at 65 °C for 15 min. The two eluates were pooled and incubated at 65 °C for 12 h to reverse cross-link the chromatin, followed by treatment with RNase A and proteinase K and purification of DNA, as described above for input samples. Both input and ChIP samples were subjected to Bioanalyzer analysis to check that the major bulk of isolated DNA was in the range of 250 bp.

Samples were analyzed by qPCR (Stratagene) in HeLa cells following the manufacturer's recommendations. Oligonucleotide pairs used for qPCR in this study are presented in Supplementary Table 5. For ChIP–seq experiments in Raji cells, purified DNA was quantified with the Qubit dsDNA HS Assay (Thermo Fisher Scientific). ChIP DNA (5 ng) was used to prepare sequencing libraries with the Illumina ChIP Sample Library Prep Kit (Illumina). After end repair and adaptor ligation, library fragments were amplified with 12 cycles of PCR. Barcoded libraries from different samples were pooled together and sequenced on the Illumina HiSeq 2000 platform in paired-end sequencing runs.

Chromatin RNA sequencing

Chromatin-associated RNAs (chrRNAs) were isolated from 2×10^7 Raji cells before and after 10 min of PDS treatment or from 2×10^7 NIH-3T3 cells before and after serum stimulation (15% fetal calf serum) for 15 min (Extended Data Figs. 5b,c and 10d,e) as described previously[65], followed by TURBO DNase treatment. In brief, 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 $MgCl_2$, 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 $MgCl_2$, 0.5% IGEPAL CA-630, 10% sucrose) before collecting the nuclear fraction at 1,400*g* for 5 min at 4 °C. Nuclei were resuspended with 125 µl NUN1 buffer (20 mM Tris-HCl (pH 8), 75 mM NaCl, 0.5 mM EDTA, 50% glycerol) followed by 1.2 ml NUN2 buffer (20 mM HEPES-KOH (pH 7.6), 300 mM NaCl, 0.2 mM EDTA, 7.5 mM $MgCl_2$, 1% IGEPAL CA-630, 1 M urea). After 15 min of vigorous vortexing, chromatin was centrifuged at 15,000*g* for 10 min at 4 °C and washed with 500 µl NUN2 buffer. After discarding the supernatant, chromatin was resuspended in 500 µl TRIzol. At this stage, the chromatin pellet is very tight and needs to be dissolved in TRIzol by repeated pipetting with tips of decreasing volume (1 ml–200 µl–10 µl) and then pushing through very small syringe needles. Next, 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 TURBO DNase (Invitrogen, AM2238) treatment. TRIzol RNA extraction and TURBO DNase treatment were performed two more times, and RNA was resuspended in 20 µl of nuclease-free water. Purified RNA was quantified by Qubit, and quality was assessed using the RNA Assay kit (Agilent RNA 6000 Pico reagents) with the Bioanalyzer 2100 (Agilent Technologies). ChrRNA was then subjected to library preparation using the TruSeq 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 ribodepletion). The datasets were submitted to the GEO database and with the manuscript[24] (Supplementary Table 1).

Transient transcriptome sequencing

TT-seq experiments were performed as previously described[66] with minor modifications. Briefly, experiments were performed in biological duplicates using 47×10^6 Raji cells treated with 10 µM PDS (Sigma, SML0678) or water (control) for 10 min. Newly synthesized RNA was labeled with 500 µM 4-thiouridine (4SU; Sigma, T4509) for 10 min at 37 °C at the time of treatment. In parallel, 60×10^6 *Drosophila* S2R+ cells were labeled for 10 min with 500 µM 4SU. Cells were then collected by centrifugation, and total RNA was isolated with the QIAzol reagent (Qiagen) according to the manufacturer's instructions. *Drosophila* RNA was used as a spike-in and added to Raji total RNA (1:40 ratio). Total RNA was fragmented (two times, 105 µg in 120 µl) using a Covaris M220 sonicator (7 °C; power, 75 W; duty factor,

1%; burst, 200) to generate RNA fragments of 200 bp to 10 kb. The 4SU-labeled RNA from 210 µg of fragmented RNA was biotinylated, pulled down using streptavidin beads and eluted with freshly prepared 100 mM DTT. Next, 4SU-labeled RNA (TT-seq) was treated with RNase-free DNase I, purified with a miRNeasy MinElute spin column and eluted in 15 µl nuclease-free water. Before library preparation, samples were quantified with the NanoDrop and quality was assessed with a Bioanalyzer Nano RNA ChIP. Ribosomal RNA was depleted from samples using the NEBNext rRNA Depletion Kit v2 (NEB, E7400), and sequencing libraries were prepared with the NEBNext Ultra II Directional kit (NEB, E7760) according to the manufacturer's instructions. RNA libraries were then sequenced in 2 × 50-bp paired-end mode on a NovaSeq 6000 instrument (Extended Data Fig. 10d,e).

Bioinformatics

Motif analysis. Canonical promoter elements and de novo and known motifs (TRANSFAC) were analyzed across all expressed genes in K562, Raji and DP cells (Fig. 1 and Extended Data Fig. 1). De novo motif discovery and known motif identification were performed using MEME and DREME[67,68], using fragments from −100 to +20 bp of experimental TSSs because this area encompasses not only the promoter but also the majority of the NDR. Enrichment of canonical promoter elements was tested using bedtools 'intersect' against all promoters of expressed genes (−100 to +20 bp of experimental TSSs) and against 10,000 permutations of random genomic areas of 121 bp. Random controls were generated using bedtools 'random' and using or not using GC constraints. GC thresholds were determined to fit exactly the GC bias observed at promoters of expressed genes. Motifs used for this analysis (Supplementary Table 2) were the BRE and the canonical or noncanonical TATA boxes as indicated in Extended Data Fig. 1. Additionally, we also used the quadparser QP1-7 ($(G_{n>2}N1\text{-}7) \times 4$) and bed files generated by the G4H algorithm G4H2.0 or G41.5 using a window of 25 bp as described[20,24].

G-quadruplex predictions, GC and CpG content. Most G4 predictions were performed using the G4H[20] algorithm, which allows attribution of a numerical value reflecting G4 propensity to a sequence of a defined window length. pG4s at stringencies of 1.52 and 2.0 were used throughout this study, reflecting strong and very strong likelihood of G4 formation in vitro. Previous experiments have shown that these G4H thresholds allow us to experimentally confirm 92% and 100% of the pG4 structures[20]. In some analyses (Fig. 2b–d and Extended Data Fig. 10g), G4 scores were extracted from the reference genome (hg19 with a window of 25 and a minimal score of 0.5) using the G4H algorithm[20], and 5-bp-bin-sized wiggle files were produced for whole-genome analyses (G4 score track). GC (G or C) and CpG (CG dinucleotides) content was defined across the genome, and wiggle files were generated. To compare G4 propensity and PDS effects on Pol II pause–release, genes were segregated in ten groups depending on the ratio of pausing indexes before and after PDS treatment. Regression curves and metaprofiling of G4H scores and predictions (at G4H1.5 and G4H2.0 thresholds) were then completed for all groups (Extended Data Fig. 10g).

ChIP–seq, ChIP–exo and MNase-seq analyses. All genomic experiments from this study or reanalyzed from available datasets were processed using our pipeline. Sequencing files were analyzed using Bowtie 2 (ref. 69) and Pasha[70]. Raw sequencing reads were aligned to human hg19 or mouse (mm9) genomes using Bowtie 2. Duplicate reads with identical coordinates (sequencing depth taken into account) were removed to avoid potential sequencing and alignment artifacts. For ChIP–seq and MNase-seq (nucleosome density) signal analyses, aligned reads were elongated in silico using the DNA fragment size inferred from paired reads or an estimated optimal fragment size for orphan reads using the Pasha R package. These elongated reads were then used to calculate the number of fragments that overlapped at a given nucleotide, thus representing an enrichment score for each bin in the genome. For nucleosome-positioning analyses (midpoints) presented in Fig. 3, to determine the average nucleosome positions, wiggle files representing the central nucleotides of DNA fragments were also generated. For ChIP–exo, the nucleotide located at the 5′ extremity of the DNA fragments was considered to generate wiggle files, because it represents the exact points where the nucleases stopped. Wiggle files representing the average enrichment score every 50 bp or 10 bp were generated. Sequencing data from input samples were treated in the same way to generate input wiggle files. All wiggle files were then rescaled to normalize the enrichment scores to reads per million. For ChIP–seq datasets, enrichment scores from input sample wiggle files were subtracted from ChIP sample wiggle files. This allows removing or reducing the overrepresentation of certain genomic regions due to biased sonication, local duplications and DNA sequencing. Finally, for MNase-seq, we smoothed the signal by replacing each 10-bp bin by the average of the five surrounding bins on each side.

RNA-seq analysis. All RNA-seq datasets reanalyzed in this study were processed using the following pipeline. Raw sequencing reads were aligned to the mouse genome (mm9) or the human genome (hg19) using TopHat2 (ref. 71). Alignment files were then treated using Pasha[70] to generate wiggle files. In Raji and DP cells, experimental TSSs were determined as the summit in short RNA-seq signals in a window of 300 bp of annotated TSSs.

Average binding profiles and heatmaps. To generate average binding profiles (Figs. 1–4 and 7 and Extended Data Figs. 2–7 and 10), R scripts previously developed were used to retrieve bin scores in defined regions from 10- or 50-bp-bin-sized wiggle files[70]. Heatmaps were generated from produced or published wiggle files, viewed and color-scaled according to sample read depth using Java Treeview[72] and dedicated R scripts[8]. Regions were defined either as centered on experimental TSSs (see above), on the center of the pG4 from G4H or the center of the area if no G4 was predicted. In addition, pG4s that were not located in annotated gene features or further than 200 bp from annotated TSSs were considered as intergenic. To analyze the relationships between G4 formation, gene expression and cell specificity, datasets were ranked by the ratio of G4access signals between Raji and K562 cells (Raji/K562). Three groups were defined (Raji-specific, common and K562-specific peaks) to perform metaprofiling of transcription and chromatin marks and to quantify expression using nascent chrRNA-seq datasets (Extended Data Fig. 3c–e). NDRs and deepest NDR points were defined using inversed MNase-seq signal, 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 from −300 to +100 of annotated TSSs in the sequencing signals). Next, peaks and peak summits corresponding to NDRs were determined. Peak calling was performed using wigpeakcaller[73], fixing a threshold based on the peak height and the gap between two adjacent signals (Extended Data Fig. 3f). Metaprofiling on the chromatin remodeler and modifiers (Supplementary Table 1), G4H annotations and G4access were centered at the peak summit of the inversed MNase-seq signal corresponding to the deepest points of NDRs.

To generate average binding profiles of Pol II, chrRNA and TT-seq (Fig. 7d–f and Extended Data Fig. 10d–g), hg19 RefSeq 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.

Identification of inactive promoters. To select inactive promoters, we selected the bottom 10% of Pol II signal over the defined areas (Fig. 2c and Extended Data Fig. 2i). hg19 RefSeq gene annotations were used to extract values from wiggle files associated with the selected genes, and bin scores in a region of 2 kb before and after the TSSs were determined.

Nucleosome arrays. We have performed a precise assessment of nucleosome repeat length and phasing, comparing pG4s and CTCF[74] in K562 cells (Fig. 3c), which yielded average nucleosome repeat lengths of 194 and 193 nucleotides calculated over five nucleosomes from the docking site and 215 and 216 nucleotides over ten nucleosomes, respectively.

Pausing scores. To analyze how the G4 ligand PDS impacts Pol II pausing (Fig. 7 and Extended Data Fig. 10), we have determined pausing scores based on the ratio of Pol II signals at promoters and in gene bodies[75]. Our approach for pausing score determination is comparable to the one previously described[8] with modifications. It takes into account Pol II density on either promoter regions (TSS) or gene bodies. Promoter regions were considered between −300 and +100 of the TSS to define paused Pol II density for calculations. Densities at genes bodies were analyzed in intervals of 50–100% of the length for all genes <25 kb. For longer genes (>25 kb), the interval from 10 kb to 25 kb was considered, assuming an elongation speed of at least 2.5 kb min^{-1} for Pol II. The use of these intervals avoids detecting signal originating from the promoters for short genes or genes with exceptionally large initiation areas and allows detecting more substantial signal of elongating Pol II. To avoid interferences between promoter and gene body read counts, only genes larger than 3 kb were considered. Read count was performed using HTSeq[76], normalized to the length of the genomic regions and expressed as reads per kb per million (RPKM). Only genes with sufficient read coverage were considered (>10 RPKM at promoters and >2 RPKM at gene bodies, $n = 6{,}159$). Pausing scores were expressed as the ratio of TSS/gene body. To define a high-confidence set of genes with pause–release effect and, because, in our datasets, PDS globally affected Pol II pausing (Fig. 7 and Extended Data Fig. 10, linear regression slope for $t = 0$ versus 10 min = 0.43, Wilcoxon test <0.00001, $n = 6{,}159$), we further selected genes with a significant Pol II signal increase in their gene bodies using DESeq (P value < 0.05, $n = 602$; P value < 0.05, $n = 6{,}159$). The distributions of cumulative frequencies as a function of pause indexes were established using GraphPad (Fig. 7c).

Plasmids and cloning

Sequences for the repeats of the 256 × MS2 binding sites were cloned from chemically synthesized oligonucleotides into pMK123 (ref. 77). Sequences for the MS2 stem loops are separated by a linker of only three nucleotides and were cloned in the pIntro-MS2x256 plasmid, which also contains an FRT-Hygro cassette for Flp-In recombination[41,78]. pUC57 plasmids containing the *Eef1a1* (1,513 bp) or *Polr2a* (711 bp) WT mouse promoters were purchased from GenScript; pGL4.17 plasmids containing the *Pkm* (200 bp), *Klf6* (400 bp), *Taok1* (300 bp) or mutagenized mouse promoters were purchased from GeneCust (Supplementary Table 4). All constructs were then subcloned into pIntro-MS2x256 between SnaBI and MluI sites. To introduce mutations into the largest promoters, *Polr2a* and *Eef1a1*, smaller fragments (206 and 219 bp, respectively) were purchased and subcloned into the full-length promoter between NotI and Nhe1 sites for *Eef1a1* and between NotI and MluI for *Polr2a*. Additionally, for a second version of *Eef1a1* and *Polr2a* WT promoters and for Eef1a1inv, *Polr2a* G4mut and *Polr2a* G4mut2 promoters, an additional luciferase reporter was added downstream of the MS2 reporter using the NEB assembly builder kit and the two following oligonucleotides: Gibintro-BsrG1-ires-fwd, GGTTTTCCAGTCACACCTCATGTACAGGCCCCTCTCCCTCCCCCCC and Gibluc-BsiW1-intro-Rev, TGTAAGTCATTGGTCTTAAACGTACGTCTAGAATTACACGGCGATC.

Stable expression of MCP–GFP was achieved by retroviral-mediated integration of a self-inactivating vector containing an internal ubiquitin promoter. The MCP used dimerizes in solution and contains the deltaFG deletion, the V29I mutation and an SV40 NLS24 (ref. 41). MCP–GFP-expressing cells were grown as a pool of clones and sorted by flow cytometry to select cells expressing low levels of fluorescence. Isogenic stable cell lines expressing the reporter genes were created using the Flp-In system, and HeLa Flp-In H9 integrants were selected with hygromycin (150 μg l^{-1}). For each construct, several individual clones were picked and analyzed by in situ hybridization. Clones usually looked similar, and two of them were further selected for the experiments after PCR and a sequencing check.

Circular dichroism spectroscopy

CD spectra were recorded on a Jasco J-815 spectropolarimeter equipped with a Peltier temperature control accessory (JASCO). Each spectrum was obtained by averaging three scans at a speed of 100 nm min^{-1}. A background CD spectrum of the corresponding buffer solution was subtracted from the average scan for each sample. The CD profile was monitored between 220 nm and 300 nm using quartz cells with a path length of 5 mm and a volume of 1,000 μl. Oligonucleotides were tested at a strand concentration of 3 μM.

Fluorescence resonance energy transfer melting competition assay with double-stranded DNA

FRET-MC was performed after hybridizing the two strands (the oligonucleotides with pG4s and their complementary strand) in a buffer containing potassium and therefore allowing G4 formation. To do so, each strand was diluted in FRET-MC buffer (10 mM lithium cacodylate, 10 mM KCl, 90 mM LiCl, pH 7.2) and annealed at 95 °C for 5 min and allowed to cool for 2–3 h until reaching room temperature before hybridization. In parallel, the fluorescent oligonucleotide F21T was folded in the same buffer and added to a 96-well microplate at a final concentration of 0.2 μM. Formed double-stranded DNA (3 μM) was added in the wells with or without 0.4 μM PhenDC3. The experiment was conducted in a CFX96 Real-Time PCR instrument (Bio-Rad), which was programmed to record FAM emission from 25 °C to 95 °C, with a temperature gradient of 1 °C per minute. Fluorescent emission curves were normalized using OriginPro 2019 software (OriginLab) to determine the melting temperature (T_m) corresponding to a normalized fluorescence of 0.5. The ΔT_m value is calculated by the difference in T_m of F21T with and without PhenDC3. The S factor was calculated as previously described for each sequence and plotted in a bar graph. A low S factor value indicates efficient competition and likely stable G4 formation. Experiments were performed in duplicate in two different plates.

Absorbance spectroscopy

All spectra were recorded on a Cary 300 (Agilent Technologies) spectrophotometer in 10 mM lithium cacodylate buffer (pH 7.2) at an oligonucleotide strand concentration of 3 or 4 μM in the presence or absence of 100 mM KCl.

TDS were obtained by taking the difference between the absorbance spectra of unfolded and folded oligonucleotides that were recorded at high (95 °C) and low (25 °C) temperatures, respectively, in a buffer containing 100 mM KCl. TDS provide specific signatures of different DNA structural conformations, provided that the structure is not too heat stable (a number of G4 structures do not melt at high temperatures).

IDS were obtained as described previously[79] by taking the difference between the absorbance spectra from unfolded and folded oligonucleotides. These spectra were recorded at 25 °C before and after adding potassium cation (100 mM KCl), respectively. IDS provide specific signatures of different DNA structural conformations.

Acquisition and analysis of smFISH images

smFISH was performed as previously described[41,80], with a mix of ten fluorescent oligonucleotides hybridizing against the MS2 × 32 repeat and each oligonucleotide containing four molecules of Cy3. Because each oligonucleotide bound eight times across the MS2 × 256 repeats, each molecule of pre-mRNA hybridized with 80 oligonucleotides, thereby providing excellent single-molecule detection and signal-to-noise ratios. In brief, smFISH images were recorded with an OMX Delta Vision microscope in SIM mode. Acquisition was performed in 3D with *a z* spacing of 0.125 µm, with a ×100 objective, a ×2 intermediate lens and an Evolve 512 × 512 EMCCD camera (Photometrics). To obtain the number of nascent pre-mRNAs per cell, the same images were used without SIM reconstruction and were analyzed using FISH-quant. TSs were first identified manually, and isolated pre-mRNA molecules located in the nucleoplasm were used to define the point spread function and the total light intensity of single molecules, which finally allowed us to determine the intensity of TSs. To obtain the number of nascent pre-mRNAs per cell, TSs were identified manually and isolated pre-mRNA molecules located in the nucleoplasm were used to define the point spread function and the total light intensity of single molecules, which finally allowed us to determine the intensity of TSs expressed in a number of full-length transcripts.

Live-cell image acquisition

Cells were plated on coverslips 25 mm in diameter (0.17 mm thick). After 24–48 h, the coverslips were mounted in GFP-imaging medium ($DMEM^{gfp}$-2, Evrogen) with rutin in a temperature-controlled chamber with CO_2 and imaged on an inverted OMX DeltaVision microscope in time-lapse mode. A ×100, 1.4-NA objective was used, with an intermediate ×2 lens and an Evolve 512 × 512 EMCCD camera (Photometrics). Stacks of 11 planes with *a z* spacing of 0.6 µm were acquired, with one stack collected every 3 min for 8 h.

Quantification of short movies

Short movies were analyzed as previously described[41,43] (Fig. 6g). In short, we manually defined the nuclear outline and the region within which the TS is visible, and stacks were corrected for photobleaching using a fitted curve with a sum of three exponentials. This curve was used to normalize each time point such that nuclear intensities were equal to the intensity of the first time point. We then filtered the image with a two-state Gaussian filter. First, the image was convolved with a larger kernel to obtain a background image, which was then subtracted from the original image before the quantification was performed. Second, the background-subtracted image was smoothened with a smaller kernel, which enhances the SNR of single particles to facilitate spot predetection. TS positions in each frame of the filtered images were determined as the brightest pixel above a user-defined threshold in the predetected region of the TS. When no pixel was above the threshold, the last known TS position was used. Next, the TS signal was fitted with a 3D Gaussian estimating its standard deviation σ_{xz} and σ_z, amplitude, background and position. We performed two rounds of fitting: in the first round, all fitting parameters were unconstrained. In the second round, the allowed range was restricted for some parameters to reduce large fluctuations in the estimates, especially for the frames with a dim TS or no detectable TS. More specifically, σ_{xz} and σ_z were restricted to the estimated median value ± s.d. from the frames where the TS could be predetected, and the background was restricted to the median value. The TS intensity was finally quantified by estimating the integrated intensity above background expressed in arbitrary intensity units. With the live-cell acquisition settings, the illumination power was low, and we could not reliably detect all individual molecules. We therefore collected, right after the end of the movies, one 3D stack with increased laser intensity (50% of maximum intensity, compared to 1% for the movie), which allowed reliable detection of individual RNA molecules. We also collected slices with a smaller *z* spacing for a better quantification accuracy (21 slices every 300 nm). Quantification of TS site intensity in the calibration stack was done with FISH-quant as follows: (1) when calculating the averaged image of single RNA molecules, we subtracted the estimated background from each cell to minimize the impact of the different backgrounds and (2) when quantifying the TS in a given cell, we rescaled the average image of single RNA molecules such that it had the same integrated intensity as the molecules detected in the analyzed cell. To calibrate the TS intensities in the entire movie, that is, to express the TS intensity as a number of equivalent full-length transcripts, we used the fact that the last movie frame was acquired at the same time as the calibration stack. We then normalized the extracted TS intensity in the movies, I_{MS2}, to obtain the nascent counts $N_{nasc;calib}$: $N_{nasc;calib}(t) = I_{MS2}(t) \times (N_{nasc,final}/I_{final})$, where $N_{nasc,final}$ stands for the estimated number of nascent transcripts in the calibration stack and I_{final} for the averaged intensity of the last four frames.

Analysis of long movies

To quantify long movies acquired at a low frame rate (one 3D image stack every 3 min), we used ON-quant, a rapid analysis tool that identifies TSs, measures their intensities and attributes the ON or OFF states of transcription, based on the defined intensity threshold under which a TS is considered to be silent and above which a TS is considered to be active. For the analysis of the *Eef1a1* G4mut2 and reanalysis of the *Eef1a1* lines, Big-FISH, a toolbox part of the FISH-quant package (https://fish-quant.github.io/) was used due to the upgrade of our analysis pipeline. The intensity threshold was defined based on the mean intensity of single molecules[41].

Mathematical modeling, short and long movie analysis

Intermittent transcriptional activity of the promoters is modeled using a Markov process with one active state and multiple inactive epigenetic states. The number of states and the transition rate parameters are obtained using the algorithms and pipeline first described in ref. 43.

For the sake of consistency, we provide a short description of the pipeline. The cells were imaged live for 30 min every 3 s (short movies) or for 8–9 h every 3 min (long movies).

Deconvolution and position of Pol II initiation events in short movies. The Pol II positions were found by combining a genetic algorithm with a local optimization procedure. Before initiation of the analysis algorithm, several key parameters were established. The Pol II elongation speed was fixed at 67 bp s^{-1} (refs. 41,43). The reporter construct transcript was divided into three sections consisting of the pre-MS2 fragment (PRE = 700 bp), 256 × MS2 loops (SEQ = 5,800 bp) and the post-MS2 fragment until the polyA site (POST = 1,600 bp). An extra time P_{poly} = 100 s was added to POST, corresponding to the polyadenylation signal (during this time the polymerase is past the polyA site and remains on the DNA[41]). The frame rate of short movies is sufficient to detect processes that occur on the order of seconds.

To find the positions of initiation events via the deconvolution pipeline, all the possible initiation times were discretized using a step size of 0.45 s (or 30 bp at 67 bp s^{-1}). This step was chosen as it is smaller than the minimum polymerase spacing and large enough to still accommodate a reasonable computation time. For a movie of 30 min, this choice corresponds to a maximum number of 4,020 positions. The deconvolution algorithm was implemented in MATLAB R2020a using Global Optimization and Parallel Computing Toolboxes for optimizing Pol II positions in parallel for all nuclei in a collection of movies. Waiting times were then computed from the position of each initiation event (Fig. 6g).

Long movie waiting time distribution. For long movies, the low resolution (3 min) does not allow a precise positioning of initiation events. In this case, we binarize the signal by considering that the TS is active or

inactive if the measured intensity is above or below a threshold level, respectively, which is set to be slightly higher than the similar intensity of a single polymerase. The inactive intervals then indicate long waiting times between successive polymerases. The active intervals are used to estimate the probability that waiting times are larger than the movie frame rate (3 min), which is one of the parameters needed for connecting long and short time distributions and obtaining a multiscale distribution[43].

Multi-exponential regression fitting of the survival function and model reverse engineering using the survival function. Waiting times were extracted as differences between successive Pol II initiation events from all the resulting traces, and the corresponding data were used to estimate the nonparametric cumulative short movie distribution function by the Kaplan–Meier method. Data from long movies were used to generate the nonparametric cumulative long movie distribution function. The two distribution functions were fitted together into a multiscale cumulative distribution function using the total probability theorem and estimates of two parameters p_l and p_s, representing the probabilities that waiting times are longer than the long movie frame rate and longer than the length of the short movie, respectively (see ref. 43 for details).

Next, multi-exponential regression fitting of the multiscale distribution function produced a set of $2N-1$ distribution parameters, where N is the number of exponentials in the regression procedure (three for $N=2$ and five for $N=3$). The regression procedure was initiated with multiple log-uniformly distributed initial guesses and followed by local gradient optimization. It resulted in a best-fit solution with additional suboptimal solutions (local optima with an objective function value larger than the best fit).

The $2N-1$ distribution parameters can be computed from the $2N-1$ kinetic parameters of an N-state transcriptional bursting model. Conversely, a symbolic solution for the inverse problem was obtained, allowing computation of kinetic parameters from distribution parameters and reverse engineering of the transcriptional bursting model. In particular, it is possible to know exactly when the inverse problem is well posed, that is, when there is a unique solution in terms of kinetic parameters for any given distribution parameters in a domain (Fig. 6h and Extended Data Fig. 9c).

Transcriptional bursting models. The transcriptional bursting models used in this paper are as follows:

For a promoter two-state model ($N=2$), the model corresponds to the well-known ON–OFF telegraph model. In this case, there are three distribution parameters and three transition rates parameters.

The distribution parameters are $A_1, \lambda_1, \lambda_2$, defining the survival function

$$S(t) = A_1 e^{\lambda_1 t} + (1 - A_1) e^{\lambda_2 t}.$$

These parameters are obtained by biexponential fit of the empirical survival function.

The transition rate parameters of the ON–OFF telegraph model can be obtained from the distribution parameters using the formulas

$$k_3 = -S_1,\ k_2' = S_1 - \frac{S_2}{S_1},\ k_2 = \frac{S_3 S_1 - S_2^2}{S_1 (S_1^2 - S_2)},$$

$$S_1 = A_1\lambda_1 + A_2\lambda_2,\ S_2 = A_1\lambda_1^2 + A_2\lambda_2^2,\ S_3 = A_1\lambda_1^3 + A_2\lambda_2^3,\ A_2 = 1 - A_1,$$

where k_3, k_2, k_2' are the initiation rate, the OFF-to-ON and ON-to-OFF transition rates, respectively.

For a promoter three-state model ($N=3$), there are five distribution parameters and five kinetic parameters.

The distribution parameters are A_1, A_2, λ_1, λ_2, λ_3, defining the survival function

$$S(t) = A_1 e^{\lambda_1 t} + A_2 e^{\lambda_2 t} + (1 - A_1 - A_2) e^{\lambda_3 t}.$$

The corresponding model represented Fig. 6h and Extended Data Fig. 9c has two OFF states and one ON state. The five transition rate parameters can be obtained from the distribution parameters:

$$k_3 = -S_1,\ k_2 = \frac{S_2^2 - S_1 S_3}{S_1 (-S_1^2 + S_2)}, k_2' = S_1 - \frac{S_2}{S_1},$$

$$k_1 = \frac{L_3(-S_1^2 + S_2)}{S_2^2 - S_1 S_3}, k_1' = \frac{A_1 A_2 A_3 S_1 (\lambda_1 - \lambda_2)^2 (\lambda_1 - \lambda_3)^2 (\lambda_2 - \lambda_3)^2}{(-S_1^2 + S_2)(S_2^2 - S_1 S_3)},$$

where

$$S_1 = A_1\lambda_1 + A_2\lambda_2 + A_3\lambda_3,\ S_2 = A_1\lambda_1^2 + A_2\lambda_2^2 + A_3\lambda_3^2,$$
$$S_3 = A_1\lambda_1^3 + A_2\lambda_2^3 + A_3\lambda_3^3,\ A_3 = 1 - A_1 - A_2,$$

$$L_1 = \lambda_1 + \lambda_2 + \lambda_3,\ L_2 = \lambda_1^2 + \lambda_2^2 + \lambda_3^2,\ L_3 = \lambda_1^3 + \lambda_2^3 + \lambda_3^3,$$

and $k_3, k_2, k_2', k_1, k_1'$ are the transcription initiation, OFF2-to-ON, ON-to-OFF2, OFF1-to-OFF2 and OFF2-to-OFF1 rates, respectively. S1, S2, S3, L1, L2, L3 are symmetric polynomials in the distribution parameters.

Error intervals. Distribution parameters result from multi-exponential regression fitting using gradient methods with multiple initial data. These optimization methods provide a best fit (global optimum) but also suboptimal parameter values. Using an overflow ratio (a number larger than one, in our case, 2) to restrict the number of suboptimal solutions, we define boundaries of the error interval as the minimum and maximum parameter values compatible with an objective function less than the best fit times the overflow.

Choice of the number of exponentials. The number of exponentials was determined by a parsimony principle: we have chosen the smallest N that fits well. More precisely, starting with $N=2$, we have increased N as long as the goodness of fit reduced without an increase in overfitting. We have used parametric uncertainty (error intervals) as a proxy for overfitting (Extended Data Fig. 9b).

Statistics and reproducibility

Sample size. No statistical methods were used to predetermine sample sizes. This study mainly used nonparametric tests such as Wilcoxon–Mann–Whitney with the exception of Extended Data Fig. 10a, for which data distribution was assumed to be normal, but this was not formally tested, and Student's t-test was applied.

Statistics and reproducibility. No statistical method was used to predetermine the sample size.

Data exclusion. In Extended Data Fig. 3e, genes that were not expressed (RPKM = 0) were not included.

Replication. Experiments were repeated in replicates as indicated in the paper (between two and three times). All replicates were successful.

Randomization. Randomization of genome sequences for motif enrichment analysis was repeated 10,000 times.

Blinding. The investigators were not blinded to allocation during experiments and outcome assessment.

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

Data availability
All datasets used in this study including published and original experiments are described in Supplementary Table 1. Original datasets produced in this study correspond to MNase-seq, ChIP–seq, DNase-seq, TT-seq, chrRNA-seq and G4access, which are recorded in the GEO database under GSE159276. G4 mapping and predictions are available under GSE187007. Source data are provided with this paper.

Code availability
Software and codes are all publicly available and have been previously described[8,70,73]. Codes used for this study have been deposited in the linked Zenodo repository: https://zenodo.org/record/15124924 (ref. 81).

References

63. Koch, F. et al. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat. Struct. Mol. Biol.* **18**, 956–963 (2011).
64. John, S. et al. Genome-scale mapping of DNase I hypersensitivity. *Curr. Protoc. Mol. Biol.* **103**, 21.27.1–21.27.20 (2013).
65. Nojima, T. et al. Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* **161**, 526–540 (2015).
66. Lidschreiber, K. et al. Transcriptionally active enhancers in human cancer cells. *Mol. Syst. Biol.* **17**, e9873 (2021).
67. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
68. Bailey, T. L. DREME: motif discovery in transcription factor ChIP–seq data. *Bioinformatics* **27**, 1653–1659 (2011).
69. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
70. Fenouil, R. et al. Pasha: a versatile R package for piling chromatin HTS data. *Bioinformatics* **32**, 2528–2530 (2016).
71. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
72. Saldanha, A. J. Java Treeview—extensible visualization of microarray data. *Bioinformatics* **20**, 3246–3248 (2004).
73. 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).
74. Maurano, M. T. et al. Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. *Nat. Genet.* **47**, 1393–1401 (2015).
75. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731 (2012).
76. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
77. Alexander, R. D. et al. RiboSys, a high-resolution, quantitative approach to measure the in vivo kinetics of pre-mRNA splicing and 3′-end processing in *Saccharomyces cerevisiae*. *RNA* **16**, 2570–2580 (2010).
78. Boireau, S. et al. The transcriptional cycle of HIV-1 in real-time and live cells. *J. Cell Biol.* **179**, 291–304 (2007).
79. Renaud de la Faverie, A., Guedin, A., Bedrat, A., Yatsunyk, L. A. & Mergny, J. L. Thioflavin T as a fluorescence light-up probe for G4 formation. *Nucleic Acids Res.* **42**, e65 (2014).
80. Mueller, F. et al. FISH-quant: automatic counting of transcripts in 3D FISH images. *Nat. Methods* **10**, 277–278 (2013).
81. Esnault, C. & Andrau, J. C. Scripts and codes for G-quadruplexes are promoter elements controlling nucleosome exclusion and RNA polymerase II pausing. *Zenodo* https://doi.org/10.5281/zenodo.15124924 (2025).

Acknowledgements
This work was supported by institutional grants from the CNRS including an 80prime2021 'Deciph G4' and a grant from the 'Agence Nationale de la Recherche' (G4access, ANR-20-CE12-0023), Fondation pour la Recherche Medicale FRM EQU202403018039 and INCA PLBIO20-225. C.E. was supported by a grant from ARC (retour postdoc); E.G.-O. and S.B. were supported by grants from EpiGenMed LabEx of Excellence and the 'Agence Nationale de la Recherche' (ANR-18-CE12-0019). J.-L.M. was supported by Inserm, the CNRS and Fondation de l'École Polytechnique. We are grateful to S. Spicuglia, D. Fisher, R. Feil and M. Lagha for critical reading of the paper and to L. Esnault, M. Esnault, F. Mueller and R. Topno for help with quantification of live-cell imaging data.

Author contributions
J.-C.A. and C.E. designed the experiments. C.E., E.G.-O., T.M., K.G., S.B. and A.P. performed the experiments. C.E. and A.Z.E.A. analyzed the genomic data. C.E., E. Basyuk, M.A.K., M.-C.R. and E. Bertrand performed and interpreted the microscopy experiments. Y.L., A.C., D.V. and J.-L.M. characterized and interpreted the G4s in vitro. O.R. performed the mathematical modeling. A.K.-P. and R.M. contributed to the early stage of the project. J.-C.A. and C.E. conceived of the project, suggested and interpreted experiments and wrote the article. All authors reviewed the paper.

Competing interests
The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41588-025-02263-6.

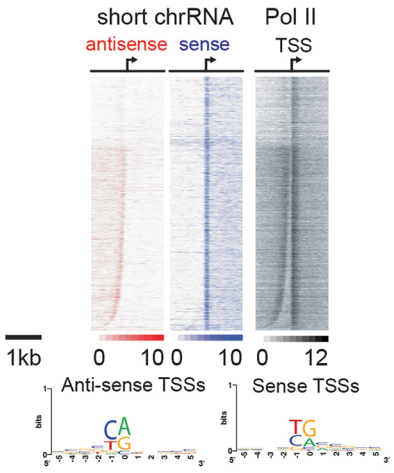
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-025-02263-6.

Correspondence and requests for materials should be addressed to Cyril Esnault or Jean-Christophe Andrau.

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

https://doi.org/10.1038/s41588-025-02263-6

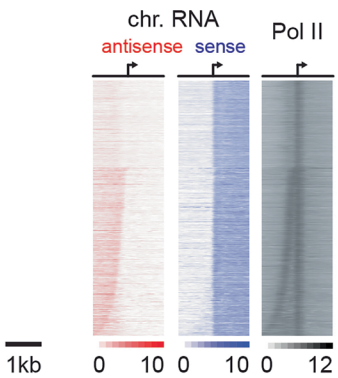
a



Motif	MEME Sites	MEME E-value	MEME Rank	MEME Seq. Logo	DREME E-value	DREME Rank	DREME Seq. Logo
G4/SP1	2846	$2.2e^{-155}$	1		$1.4e^{-298}$	1	
Ets	1308	$1.2e^{-32}$	2		$3.7e^{-194}$	2	
NFY	1071	$2.2e^{-20}$	3		$1.2e^{-104}$	3	
TATA		n.a			$6.6e^{-57}$	9	

	% in promoters	% (cor. for GC)	obs./exp.	Pvalue	Random cont.
BRE (SSRCGCC)	45.2% (3772)	6.3% (528)	7.1	$<1e^{-99}$	2.9% (242)
G4 (G4H2.0)	**21.1%** (1766)	3.1% (262)	**6.7**	$<1e^{-99}$	1.5% (121)
G4 (QP1-7)	**21.0%** (1757)	3.2% (274)	**6.4**	$<1e^{-99}$	1.5% (125)
G4 (G4H1.5)	**45.1%** (3763)	12.0% (1005)	3.7	$<1e^{-99}$	5.9% (491)
TATA (TATAWAAG)	1.0% (80)	0.6% (46)	1.7	$<1e^{-99}$	1.5% (127)
TATA (TATAW)	8.6% (721)	16.5% (1373)	0.5	ns	33.7% (2811)

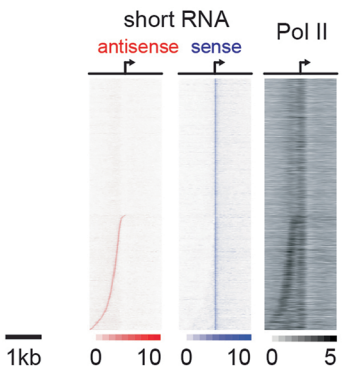
b



Motif	MEME Sites	MEME E-value	MEME Rank	MEME Seq. Logo	DREME E-value	DREME Rank	DREME Seq. Logo
G4/SP1	3230	$3.9e^{-238}$	1		$7.4e^{-238}$	2	
Ets	2187	$1.0e^{-96}$	2		$7.3e^{-260}$	1	
NFY	934	$6.9e^{-57}$	3		$8.0e^{-156}$	3	
TATA		n.a			$3.5e^{-28}$	10	

	% in promoters	% (cor. for GC)	obs./exp.	Pvalue
BRE (SSRCGCC)	44.0% (3680)	6.3% (529)	7.0	$<1e^{-99}$
G4 (G4H2.0)	**17.3%** (1444)	3.1% (262)	**5.5**	$<1e^{-99}$
G4 (QP1-7)	**17.6%** (1467)	3.2% (275)	**5.3**	$<1e^{-99}$
G4 (G4H1.5)	**40.7%** (3398)	12.0% (1007)	3.4	$<1e^{-99}$
TATA (TATAWAAG)	0.9% (73)	0.6% (47)	1.6	$<1e^{-99}$
TATA (TATAW)	6.3% (523)	16.5% (1375)	0.4	ns

c



Motif	MEME Sites	MEME E-value	MEME Rank	MEME Seq. Logo	DREME Motif	DREME E-value	DREME Rank	DREME Seq. Logo
G4/SP1	2249	$3.9e^{-238}$	1		G4/SP1	$1.5e^{-227}$	1	
A-stretch	210	$2.1e^{-47}$	2		Ets	$1.3e^{-135}$	2	
Ets or NFY		n.a			NFY	$1.8e^{-96}$	3	
TATA		n.a			TATA	$6.3e^{-46}$	7	

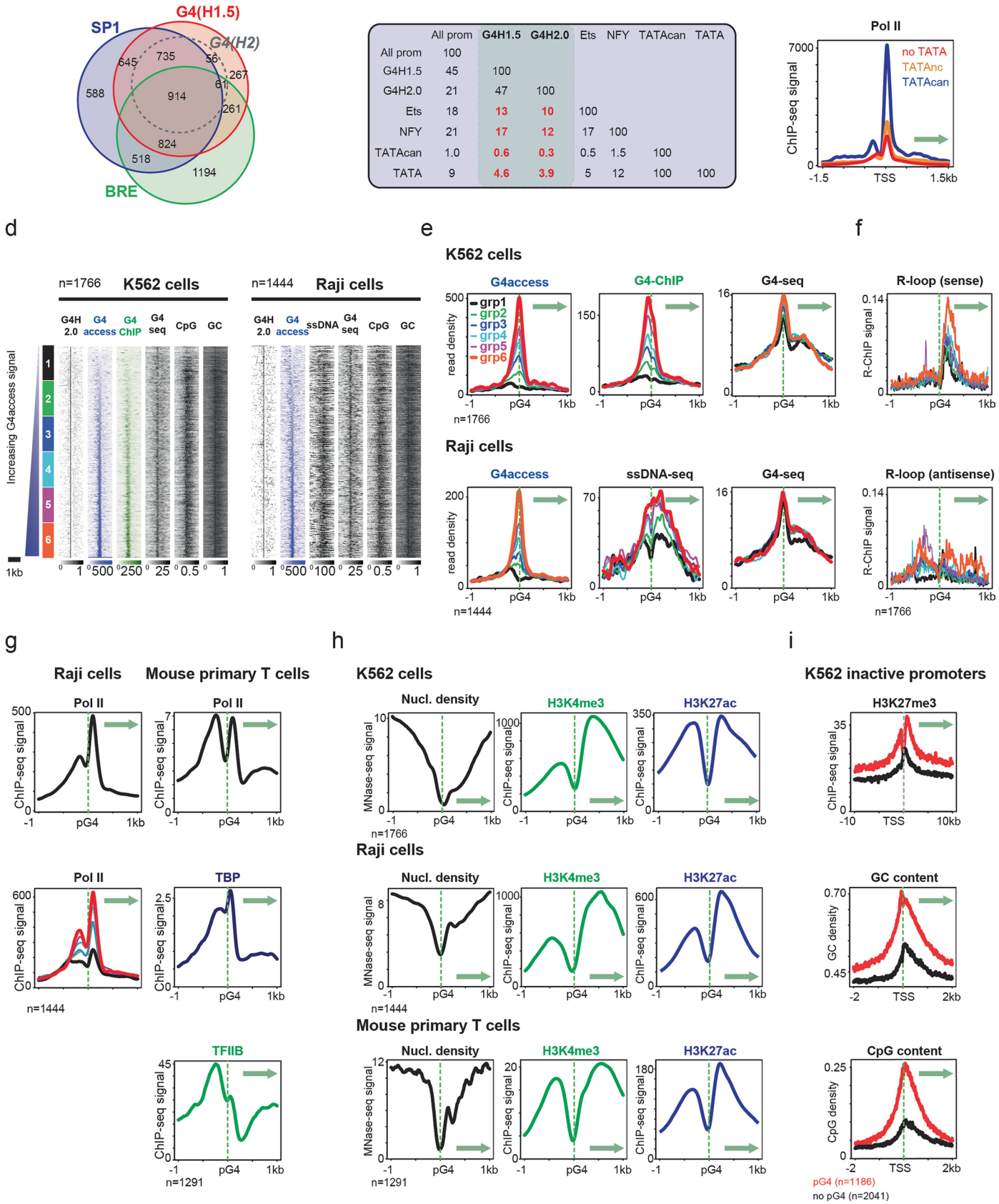
	% in promoters	% (cor. for GC)	obs./exp.	Pvalue
BRE (SSRCGCC)	42.3% (3364)	6.5% (512)	6.6	$<1e^{-99}$
G4 (G4H2.0)	**16.2%** (1291)	3.0% (235)	**5.5**	$<1e^{-99}$
G4 (QP1-7)	**16.3%** (1294)	3.0% (238)	**5.4**	$<1e^{-99}$
G4 (G4H1.5)	**40.1%** (3184)	11.0% (944)	3.4	$<1e^{-99}$
TATA (TATAWAAG)	1.0% (83)	0.6% (46)	1.8	$<1e^{-99}$
TATA (TATAW)	6.4% (509)	16.9% (1345)	0.4	ns

Extended Data Fig. 1 | See next page for caption.

https://doi.org/10.1038/s41588-025-02263-6

Extended Data Fig. 1 | Analysis of promoter elements and motifs in 3 independent cell lines. a- Definition of major TSS (mTSS), MEME and DREME promoter motif analyses in K562 cells. CAGE datasets (FANTOM consortium) were used to define mTSSs at the nucleotide resolution (n=8346). Heatmaps show short chromatin (nascent) RNA-seq (analysed from GSE52914) and Pol II ChIP-seq (ENCODE) docked on the main sense mTSS and ranked by increasing distance between sense and antisense short RNAs (top left panel). Sequence motif analysis of the transcription initiating nucleotides (INR) at the sense and antisense mTSSs are shown (bottom left). On the right panels are shown TFBS analyses using MEME or DREME and frequency analyses for all sequences features including Quadparser (QP1-7) and BRE motifs. The random control column depicts a search for the motif in 8346 random genomic sequences (with 10000 permutations). See also Supplementary Table 2 for detailed frequencies of all motifs. **b-** Definition of mTSSs and MEME and DREME motif analyses in Raji B cells. The analysis was performed as in (**a**) over 8356 human promoters and using chrRNA-seq in Raji for mTSS determination (analysed from GSE52914). **c-** Definition of mTSSs, MEME and DREME motif analyses in mouse primary T cells. The analysis was performed as in (**a**) over 7947 mouse promoters and using short-RNA datasets (size-selected below 50 bp, analysed from GSE38577) to define mTSSs at the nucleotide resolution. Short RNA-seq and Pol II ChIP-seq data sets are shown.

https://doi.org/10.1038/s41588-025-02263-6

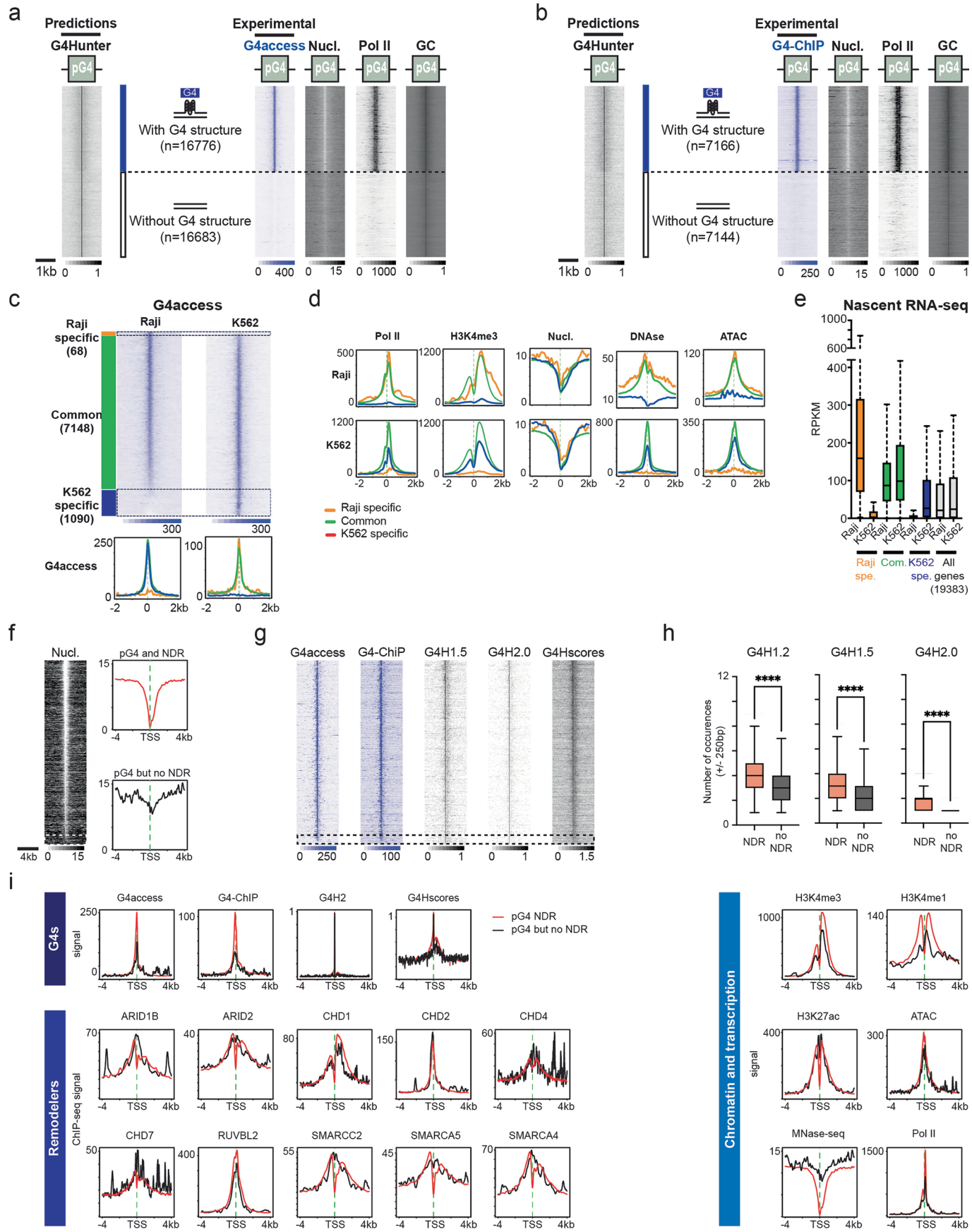


Extended Data Fig. 2 | See next page for caption.

https://doi.org/10.1038/s41588-025-02263-6

Extended Data Fig. 2 | Association of the TATAbox and G4 motifs with transcription initiation. a- pG4, BRE and SP1 motifs largely overlap at active promoters. Venn diagram of active promoters containing SP1, BRE motifs or pG4s. G4Hunter is displayed at two stringencies (1.5, red circle or 2.0, dotted inner circle). The principle of the G4Hunter algorithm is to score positively Gs and G stretches, while scoring negatively Cs and C stretches within a defined window (typically 25 nt). Overlapping G4s, above a defined threshold, are concatenated. A and T nucleotides score are fixed as null. G4Hunter scores >1.5 and 2.0 correspond to likelihood of G4 formation *in vitro* of >95 and 99%[20,24]. **b**- Promoters containing G4 predictions (G4H1.5 or 2.0) tend to harbour less other TFBS or promoter elements as compared to all promoters (All prom). This analysis was performed with the selection described in Fig. 1 for K562 cells. **c**- Promoters with TATA boxes show more focused and directional transcription. Pol II ChIP-seq profiles in K562 cells (ENCODE) at promoters of expressed genes that contain no TATA or a noncanonical (TATAW) or a canonical (TATAWAAG) TATA box. **d**- Experimental G4 signals definition. Groups 1–6 correspond to increasing level of G4access signals. Heatmaps in K562 and Raji of predicted G4 (G4Hunter 2.0) and G4 signals using G4access[24], G4 ChIP[14,25] G4-seq[26], GC and CpG contents and in Raji predicted G4, G4access, ssDNA-seq[27], G4-seq, GC and CpG contents. See also Figs. 1e and 2b. The heatmaps were centred on the G4 motif upstream of TSS as shown below the tracks. **e**- Experimental G4 signal metaprofiles. Metaprofiles of the heatmaps shown in d. The signals were divided in 6 groups of ascending G4 access signals. Metaprofiles were centred on the G4 motif upstream of TSS as shown below the tracks. **f**- Metaprofiles of R-loops at active promoters and docked on pG4s (G4H 2.0) upstream of experimental TSSs in K562 cells. The signals were divided in 6 groups of ascending G4 access signals **g**- Metaprofiles of ChIP-seq of Pol II, TBP and TFIIB. (Left) Pol II signals were centered on pG4s (G4H2.0) in 1444 pG4-containing promoters of active genes (−100,+20 bp) in Raji cells and in the 6 classes defined in Extended Data Fig. 2d. (Right) Metaprofiles of ChIP-seq of Pol II, TBP, TFIIB, centered on pG4s (G4H2.0) on 1291 pG4-containing promoters of active genes (−100,+20 bp) in mouse T cells (GSE38577). **h**- Metaprofiles of Nucleosome density, H3K4me3 and H3K27ac. Heatmaps of MNase-seq signals (Nucl. Density) and ChIP-seq signals of H3K4me3 and H3K27ac are shown on promoters containing G4 predictions in K562, Raji and mouse primary T cells. Signals are centered on pG4s (G4H2.0). **i**- Heatmaps and metaprofiles of Polycomb-deposited H3K27me3 mark, GC and CpG-content in groups 1 and 2 defined in Fig. 2c. A green arrow indicates the sense of transcription in each graph, All accession numbers are presented in the Supplementary Table 1.

https://doi.org/10.1038/s41588-025-02263-6

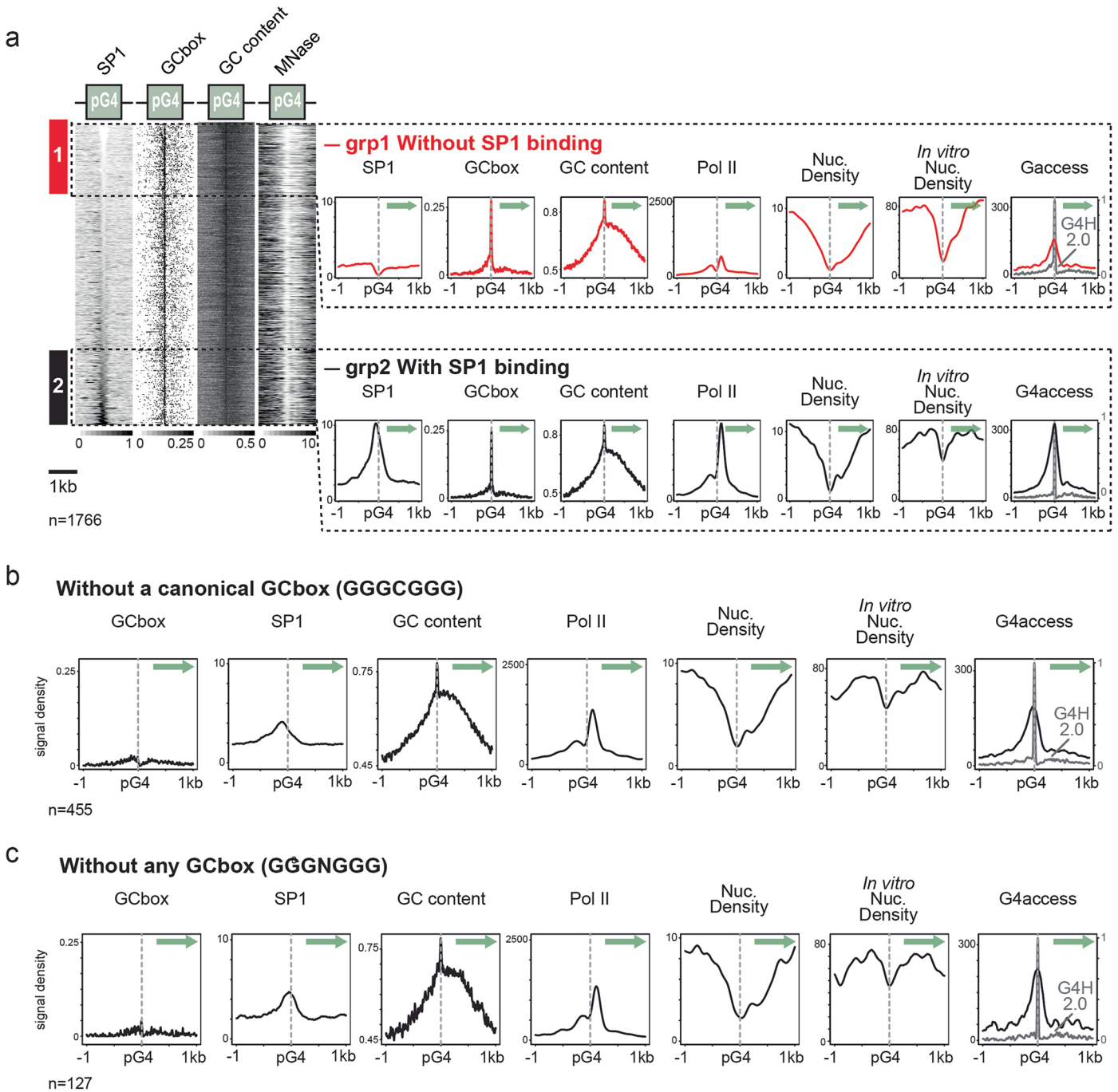


Extended Data Fig. 3 | See next page for caption.

https://doi.org/10.1038/s41588-025-02263-6

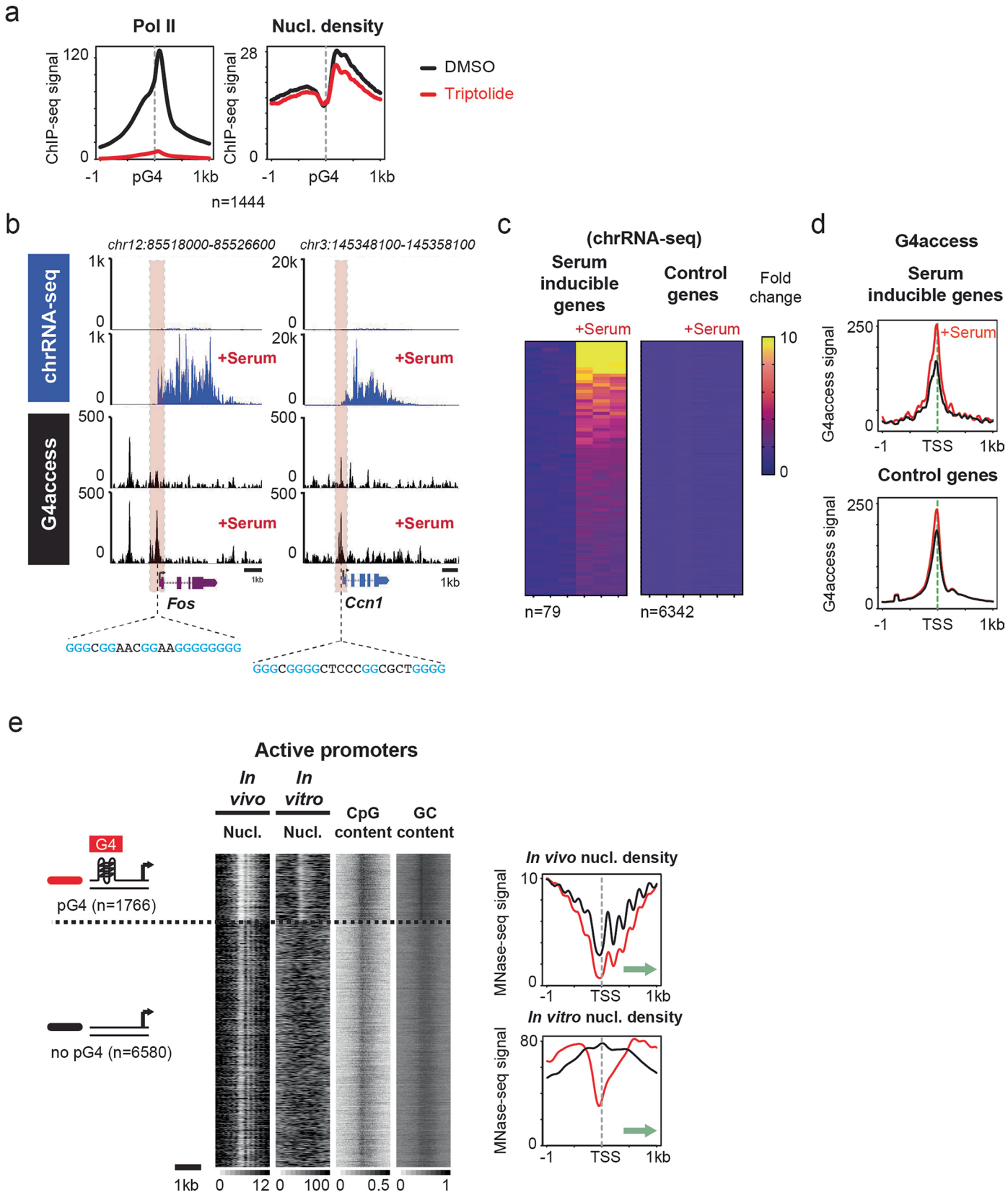
Extended Data Fig. 3 | G4s associate with nucleosome exclusion, Pol II binding and chromatin modifiers. a- G4 forming sequences mapped by G4access associates with nucleosome eviction and high Pol II signals. G4Hunter (G4H1.2) containing regions associated to G4access peaks were selected and compared to a control group of G4Hunter (G4H1.2) regions that do not yield G4 signals. Corresponding signals of G4Hunter (GH1.2), G4access, MNase-seq, Pol II signals and GC content track are displayed. **b-** G4 forming sequences mapped by G4 ChIP associates with nucleosome eviction and high Pol II signals. G4Hunter (G4H1.2) containing regions associated to G4 ChIP peaks were selected and compared to a control group of G4Hunter (GH1.2) regions that do not yield G4 signals. Corresponding signals of G4hunter (GH1.2), G4-ChIP, MNase-seq, Pol II signals and GC content track are displayed. **c-** Cell specific G4access signals segregate promoter activity. G4access signals at promoters in Raji and K562 cells were isolated and the union of peaks were analysed in each cell type and ranked by the ratio of signals in Raji vs K562, allowing to score for 3 classes, as indicated, of Raji, common and K562 specific regions. Signals for each class are indicated below the heat maps. **d-** Cell-specific G4access yields specific transcription and chromatin patterns in Raji and K562 cells. Transcription, active chromatin, nucleosome density and accessibility were analysed for the 3 classes in the 2 cell lines. **e-** Cell-specific G4access yields specific nascent RNA patterns in Raji and K562 cells. Box plot of chrRNA-seq at genes controlled by the 3 classes of promoters as indicated. Pval indicated by **** represents statistical values assessed by Mann-Whitney two-sided test and $<10^{-3}$. N=64; 6489; 975 and 19383. Box and whiskers with median, upper and lower quartiles and minimal and maximal ranges. **f-** Heatmap and signal plot of NDR and no NDR selections of promoters harboring G4 signal (separated by dashed line). **g-** Heatmaps of experimental and predicted G4s at both selections. **h-** Distribution of the number of moderate (G4H1.2), strong (G4H1.5) and very strong G4 predictions (G4H2.0) in the two selections. N=53 and 1634. Box and whiskers with median, upper and lower quartiles and +/- 1.5x inter quartile range. **i-** G4s, remodelers, chromatin and transcription signals at NDR and no NDR selections.

https://doi.org/10.1038/s41588-025-02263-6



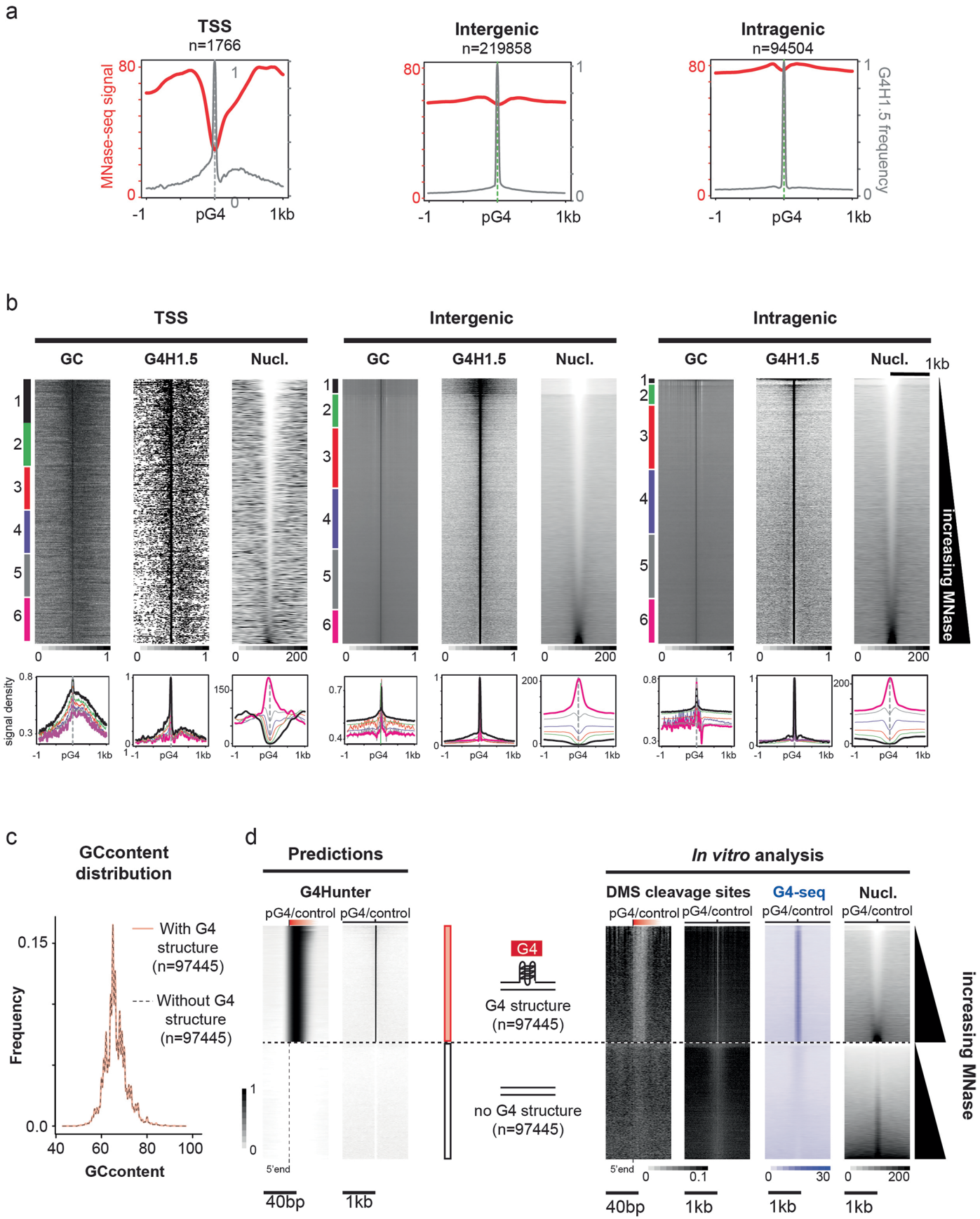
Extended Data Fig. 4 | G4, nucleosome and Pol II Patterns of promoters with or without SP1/GC boxes. a- Heatmaps ranked by increasing signals of SP1 binding and centred on a pG4 (ChIP-seq, ENCODE). All promoters of expressed genes that contain a pG4 in K562 cells are presented (n=1766). Group 1 and 2 are depleted or enriched for SP1, respectively. Metaprofiles derived from the heatmaps (groups 1 and 2) of SP1 density, GC content nucleosome densities and G4access are shown on the right. **b**- Metaprofiles of promoters not containing a canonical GC box/SP1 binding site. **c**- Metaprofiles of promoters not containing a non-canonical GC box/SP1 binding site. Green arrows indicate the sense of transcription in each graph.

https://doi.org/10.1038/s41588-025-02263-6



Extended Data Fig. 5 | Chromatin and G4 dynamics following transcription inhibition, activation and chromatin opening *in vivo* and *in vitro* at promoters. a- Persistence of NDRs at pG4 sites following transcription inhibition by triptolide. Raji cells were treated for 2h with 1 μM. Pol II clearance and nucleosome depletion mapped by MNase-seq from active genes containing pG4s (G4H2.0) are shown (GSE187007 and this study). **b-** Examples of G4 (G4access) and transcription (chrRNA) dynamics following serum shock at serum-inducible genes. NIH3T3 were treated 15 min with serum following overnight starvation (see methods). **c-** Serum-inducible genes as determined by DESeq (Padj=0.05) are plotted as a heatmap by decreasing fold change as indicated (left). Invariant control genes heat maps are also shown (right). **d-** Metaprofiles of G4access signals following serum induction at inducible and control genes. **e-** Heatmaps and metaprofiles of nucleosome densities *in vivo* and *in vitro*, GC and CpG-content in groups 1 and 2 defined in Fig. 2d.

https://doi.org/10.1038/s41588-025-02263-6

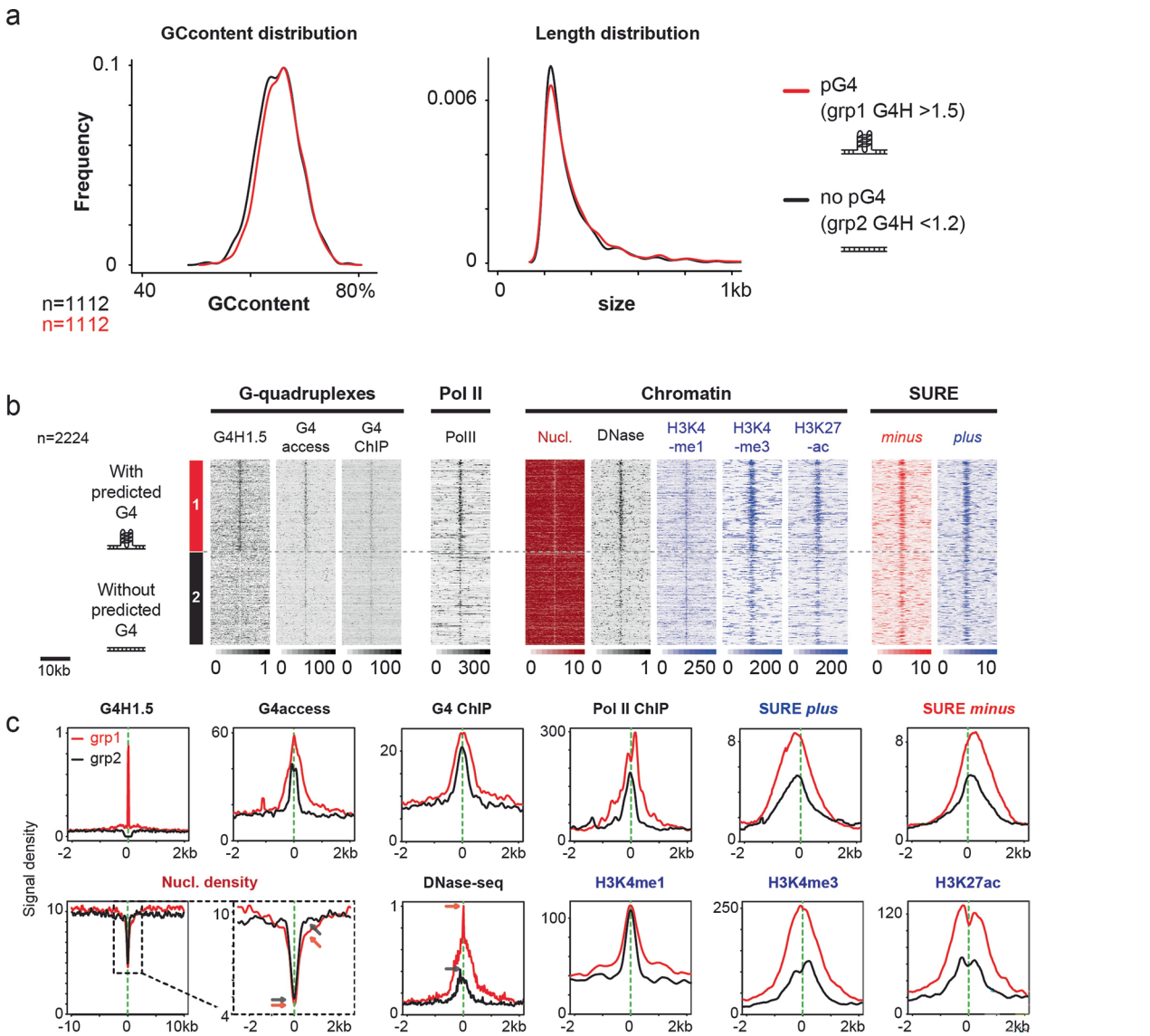


Extended Data Fig. 6 | See next page for caption.

https://doi.org/10.1038/s41588-025-02263-6

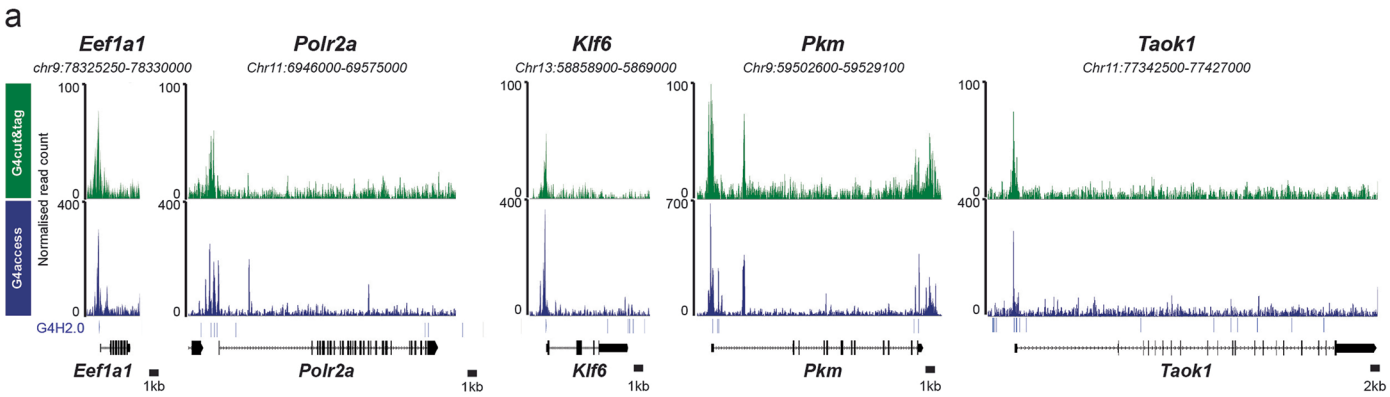
Extended Data Fig. 6 | Strong G4 predictions yield nucleosome depletion and footprint of both ss and dsDNA *in vitro*. **a**- Metaprofiles of MNase-seq from *in vitro* reconstituted chromatin on T cell genomic DNA (GSE25133) centred on pG4 at the active pG4-containing promoters (−100, +20 bp), intergenic or intragenic regions defined in K562 cells. **b**- Heatmaps ranked by increasing MNase signals showing nucleosome or G4H1.5 (docked on G4H2.0) signals. Six visually defined groups based on relative nucleosome densities are further plotted as graphs below the heatmaps. **c**- Two sets of sequences were randomly selected with (G4H>2) or without (G4H <1) G4 predictions but with identical GC content in intergenic areas of the human genome. - GC content distribution of the two selections were sampled to yield identical patterns. **d**- G4 forming potential but not GC content correlates with nucleosome exclusion. Heatmaps of the 2 selections for G4H2.0 annotation densities, DMS-seq, G4seq and *in vitro* nucleosome densities, ranked by increasing MNase signal. DMS-seq monitors G4's footprinting by quantifying cleavage sites in dsDNA human genomic context while G4seq monitors G4 formation *in vitro* at genomic sequence based on nucleotide misincorporation (ssDNA). This figure shows that for equivalent GC content, most areas able to form G4s *in vitro* do deplete nucleosome by default.

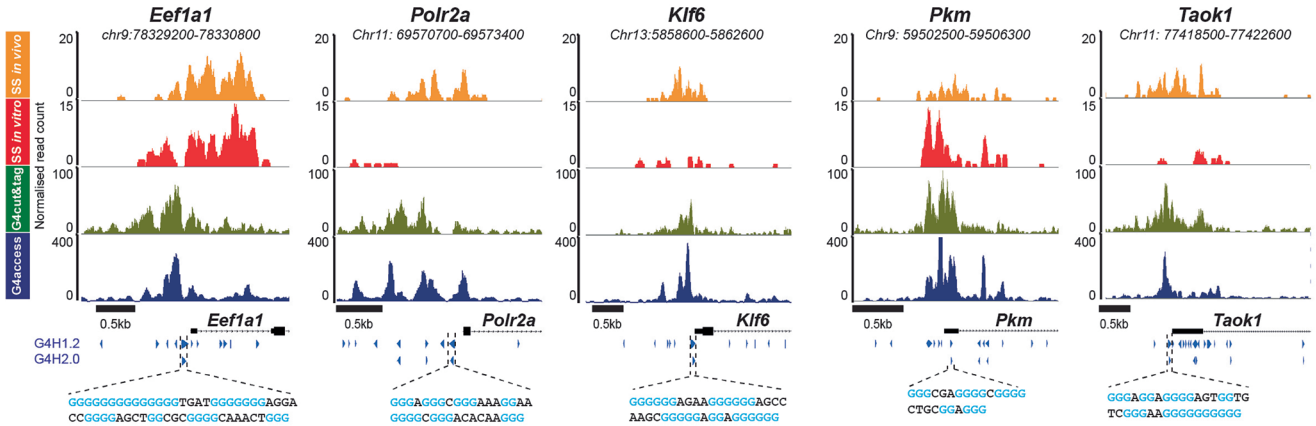
https://doi.org/10.1038/s41588-025-02263-6

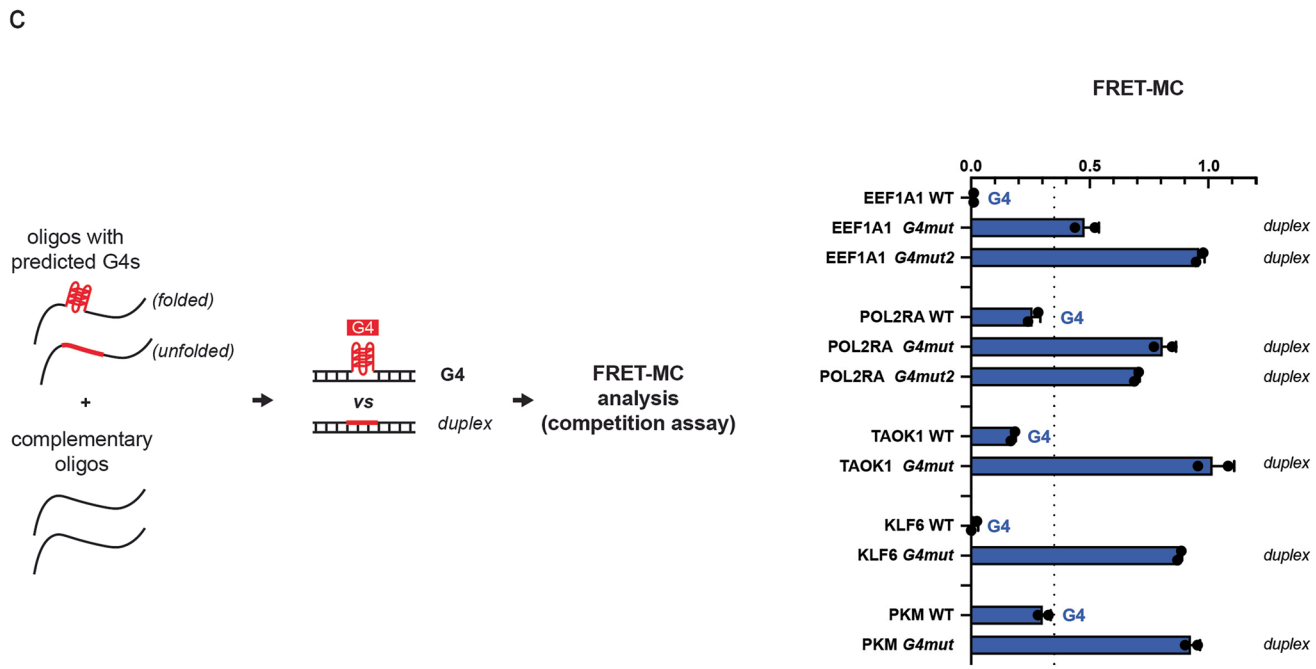


Extended Data Fig. 7 | G4s contribute to CpG islands openness and activity. **a**- Selection of G4-containg and G4-depleted CpG islands. The selections were performed on the same number of sequences (2 x 1112) with similar length and GC content, with the indicated G4Hunter thresholds. For the groups to be of equal size, equivalent length and GC content the initial populations of CGIs with G4H>1.5 (21536) or G4H<1.2 (2191) were randomized to end up with 2 groups of 1112 sequences. **b**- Heatmaps as in Fig. 4a for the selections presented in a. The DNAse data used here indicate more chromatin opening in the G4-containing group (ENCODE data for K562 cells, GSE32970). **c**- Average profiles of the group 1 and 2 shown in a and b, zoomed over 4 kb. For Nucleosome density, both a 20 kb and 4 kb windows are indicated to best show the differences in the 2 groups (observed essentially for the width of the NDRs).

https://doi.org/10.1038/s41588-025-02263-6





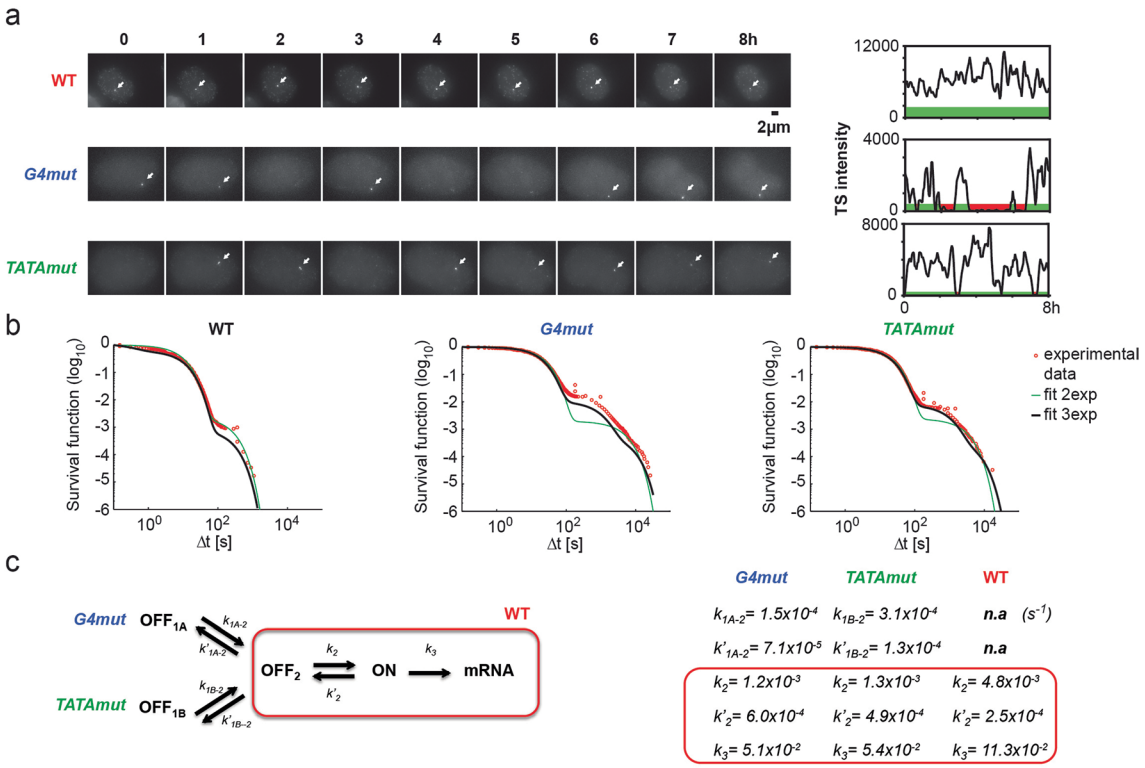


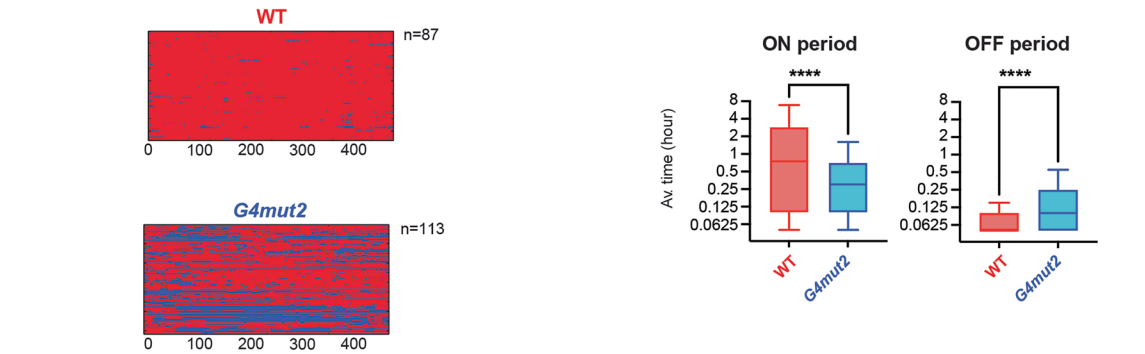
Extended Data Fig. 8 | See next page for caption.

https://doi.org/10.1038/s41588-025-02263-6

Extended Data Fig. 8 | Evidence for G4 formation at pG4s in model promoters *in vivo* and *in vitro*. **a**- G4access and CUT&Tag[42] at model promoters. Signals were extracted from data obtained in mouse ES cell lines at indicated model promoters. Predicted G4s scored by G4Hunter G4H2.0 are indicated. Their sequences, used for promoter assays in single cells, are shown below the tracks in the zoomed areas. Each of the G tracks (n>1) are indicated in light blue. Complete sequences of the model core promoters are indicated in Supplementary Table 3. **b**- G4access, G4 cut&Tag as in (**a**) and ssDNA *in vitro* (mouse T cells) and *in vivo* at model promoters and zoom around the gene starts. G4Hunter annotations and model G4 sequences used in FISH are shown. **c**- *In vitro* assessment of G4 structures of model G4s using a dsDNA FRET-MC assay. (left) Cartoon representing the principle of the assay. Template oligonucleotides are incubated in G4 folding conditions prior hybridization and completion of the FRET-MC competition assay. (right) Plot of the S Factor of dsDNA samples of wt and mutants as indicated. In this assay a score <0.35 indicates a G4 is formed. Sequences of all oligonucleotides are provided in Supplementary Table 3. (n=2 technical replicates, means ±s.e.m).

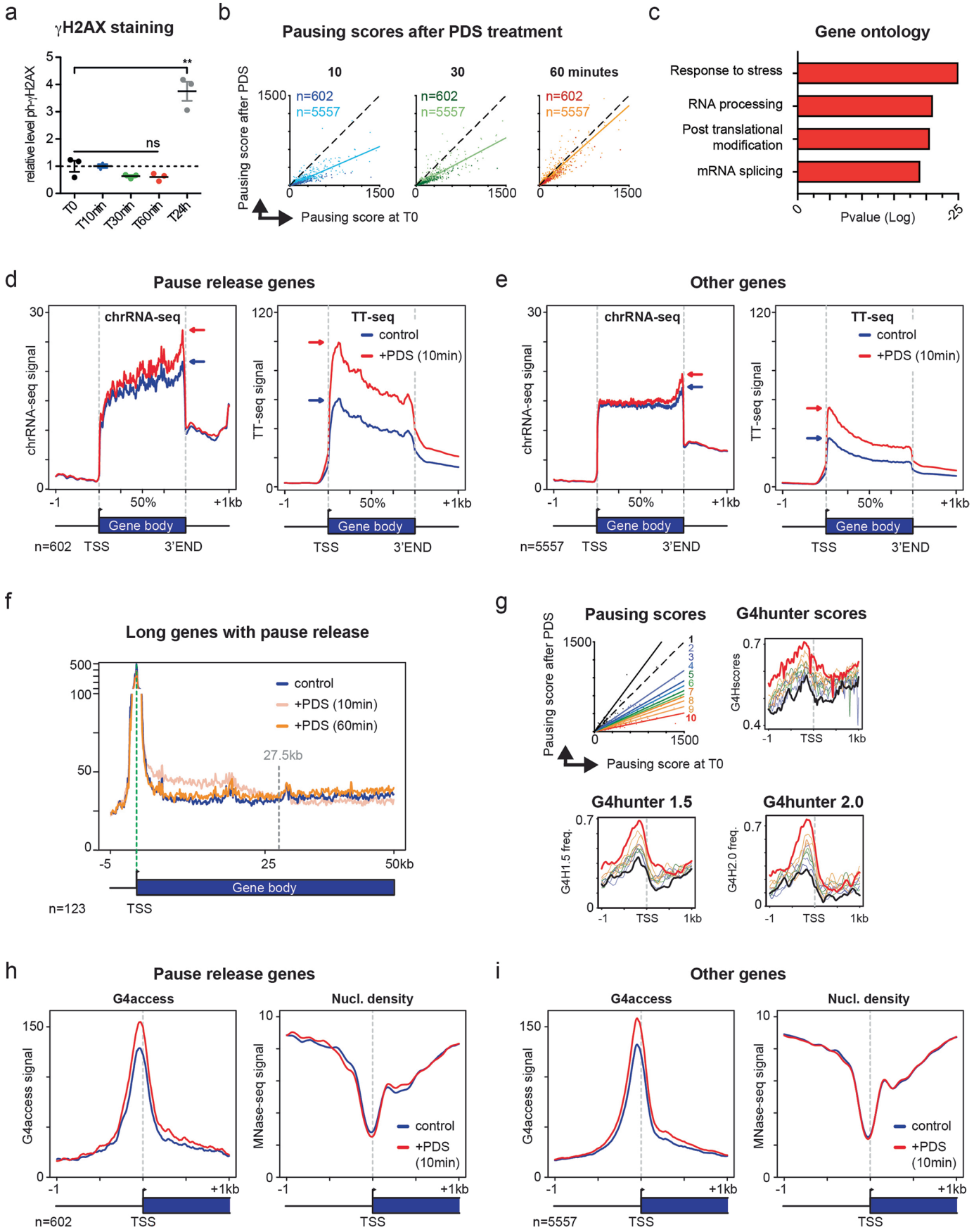
https://doi.org/10.1038/s41588-025-02263-6





Extended Data Fig. 9 | G4 forming potential is involved in Eef1a1 model promoter activity. a- Representative images of long movie analyses. Long movies of MS2 reporter activity in WT and mutant cell lines as indicated. (Left) maximum image projections of selected 3-D image stacks from the 8h movies. The arrows indicate transcription sites. (Right) graphs display the corresponding quantifications of the movies, with the bars representing ON (green) and OFF (red) states. **b-** Survival functions describing transcriptional bursting at the model promoters and deconvolution indicating 2 fits to describe optimal function in the WT situation and 3 fits for both TATA and G4 mutants. **c-** WT promoters require 2 main steps for transcription, and mutant promoters 3 main steps defined by 3- or 5-time constants, respectively. The values of the derived time constants (described in methods) are indicated in the right panel. **d-** Heatmaps of permissive (red) and non-permissive (blue) transcription periods for promoters of WT and mutant Eef1a1 G4mut2 in single cell live imaging (long movies, 8h, with stacks every 3 min). Each line represents an individual cell assessment. **e-** Box plots of ON and OFF period average duration in WT and G4mut2 promoters (measured in long movies). Computed P-values (two-sided Mann-Whitney test) are as follows: ****<1e^{-4}. N= 204;869;169 and 857. Box and whiskers with median, upper and lower quartiles and +/− 1.5x inter quartile range.

https://doi.org/10.1038/s41588-025-02263-6



Extended Data Fig. 10 | See next page for caption.

https://doi.org/10.1038/s41588-025-02263-6

Extended Data Fig. 10 | Time course analysis of G4s stabilization by PDS effects on nucleosome exclusion and positioning, and Pol II pausing. **a-** FACS analysis of γH2AX levels following PDS treatment at the indicated time (n=3 biological replicates - different cultures, time and experiments, means ±s.e.m). T24h is significantly different from t0 (Pval=0.0025) as assessed by a two-sided Student t-test. **b-** PDS influences Pol II pausing. Changes of Pol II pausing scores (see methods) in response to 10 µM PDS of all genes with detectable Pol II at promoters and gene bodies for 0,10, 30 and 60 min are displayed. Genes that have increased gene body signals with a P-value<0.05 (DESeq) at t=10 min are highlighted in dark blue (t=10 min), dark green (t=30 min) or dark orange (t=60 min), showing that pause release is relatively transient following treatment. **c-** Gene ontology analysis of the 602 selected pause release genes using DAVID webtool. **d-** Pause release profiles (nascent RNA) over gene bodies using chrRNA-seq and TT-seq read-outs. **e-** Other genes profiles over gene bodies. **f-** Pause release for long genes (Pol II profiles) and G4Hunter scores of the 2 classes following 10 and 60 min PDS. **g-** Pausing score's dynamics correlates with G4 strength/stability. (top left) ten groups of pausing scores changes were defined and their trend is shown. most pause release is found on the genes with higher G4score (top right) densities of G4H annotation (G4H1.5 or 2.0, bottom panels). **h-** G4access signal and nucleosome densities dynamics at pause release genes. **i-** G4access signal and nucleosome densities dynamics at other genes.

nature research

Corresponding author(s): Jean-Christophe Andrau

Last updated by author(s): Apr 17, 2025

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	SRA-Toolkit (version 2.9.0) was used for data collection.
Data analysis	bedtools (v2.21.0) bowtie2 (2.1.0) HTSeq (0.6.1p1) For MEME and DREME: MEME-ChIP (5.0.2) PASHA (0.99.21) FastQC (0.10.1) R (3.3.1) TopHat2 (v2.0.10) TreeView (1.2.0-osx) DESeq (version 1.26.0) G4hunter (DOI 10.5281/zenodo.7816663) FISH-quant (https://fish-quant.github.io/) Matlab Softwares for image analysis are deposited on https://zenodo.org/uploads/15124924 (Please also see Tantale et al. 2021; https://doi.org/10.1038/s41467-021-24462-5)

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

Data availability
Data availability
The GEO accessions for specific experiments related to this study are recorded under GSE159276. G4 mapping and predictions are available under GSE187007.
Genome used were hg19 and mm39 (For Homo sapiens and Mus musculus genomes).

All genomic raw datasets used in this study are available.
The extended table 1 lists all genomic datasets used in this study.
Genomic datasets produced in this study are deposited under the accession number: GSE159276.
UCSC links are:
https://genome.ucsc.edu/s/Cyril_Esnault/hg19_tracks_esnault_et_al_manuscript_PDS
https://genome.ucsc.edu/s/Cyril_Esnault/hg19_tracks_esnault_et_al_manuscript_alpha_amanitin

Figures with associated raw data:
Fig.3,7 and ED Fig.2,5 and 10

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

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	No sample size was predetermined in our study. Number of genomic loci used in Fig.1-4, 7 and Extended Data Fig. 1-7, 10 were biologically defined by their epigenetic or genomic features described in the manuscript. We used large number of cells for our microscopy analyses (Fig. 6e-g and Extended Data Fig. 9d-e) to derived meaningful statistic as in Tantale, K. et al. Nat Commun 7, 12248 (2016); Tantale, K. et al. Nat Commun 12, 4503 (2021) and Pimmett, V.L. et al. Nat Commun 12, 4504 (2021). Finally, our specific FRET-MC MNAse , ChIP and RNA-seq experiments Extended Data Fig. 5c, 8c) were performed in duplicates or triplates to draw conclusions as in Esnault, C. et al. Mol Cell 65, 1081-1095 e5 (2017); Gualdrini, F. et al. Mol Cell 64, 1048-1061 (2016) Policarpi, C., Nat Genet 56, 1168-1180 (2024), Butz, S. et al. Nat Genet 54, 1702-1710 (2022), Esnault, C. et al. Nat Genet 55, 1359-1369 (2023).
Data exclusions	In Extended Data Fig. 3e, genes that are not expressed (RPKM = 0) were not included.
Replication	Experiments were repeated in replicates as indicated in the manuscript, (Between 2 and 3 times). All replications were successful.
Randomization	Randomization of genome sequences were repeated 10000 times for motif search analyses. Genomic loci were defined by their epigenetic signatures or their properties as described in the manuscript. For ChIP, RNA, MNase, FRET-MC and microscopy analyses , samples were defined by the treatments or genetic features.
Blinding	Blinding wasn't used in this study, samples were defined by their treatments or genetic features.

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
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used: Total PolII (N20 - sc899 - lot H3115 - Santa Cruz Biotechnology), TBP (N12 - sc204 - lot LO214 - Santa Cruz Biotechnology) 5µg per immunoprecipitation

Validation: These antibodies are 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-blots 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
- 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.
- Mouse CD4+ CD8+ DP cells were sorted from thymuses of 5 to 6 weeks old mice as described (Reference:8)
- NIH-3T3 is a fibroblast cell line that was isolated from a mouse NIH/Swiss embryo. ATCC-CRL-1658

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.

GSE159276

Files in database submission

Raji_Mnase-seq_1
Raji_Mnase-seq_2
Raji_Mnase-seq_a-amanitin_18h
Raji_PolII_T0_PDS1
Raji_PolII_T0_PDS2
Raji_PolII_T10_PDS1
Raji_PolII_T10_PDS2
Raji_PolII_T30_PDS1
Raji_PolII_T30_PDS2
Raji_PolII_T60_PDS1
Raji_PolII_T60_PDS2
Raji_PolII-seq_a-amanitin_18h
Raji_chrRNA-seq_T0_PDS_rep1
Raji_chrRNA-seq_T0_PDS_rep2
Raji_chrRNA-seq_T10_PDS_rep1
Raji_chrRNA-seq_T10_PDS_rep2
Raji_Mnase-seq_DMSO_1

Raji_Mnase-seq_DMSO_2
Raji_Mnase-seq_Triptolide_1
Raji_Mnase-seq_Triptolide_2
Raji_Dnase-seq
Raji_TT-seq TT-seq_T0_PDS1
Raji_TT-seq TT-seq_T0_PDS2
Raji_TT-seq TT-seq_T10_PDS1
RajI_TT-seq TT-seq_T10_PDS2
NIH3T3_ChrRNA_Starved_1
NIH3T3_G4access_Starved_1
NIH3T3_ChrRNA_Starved_2
NIH3T3_G4access_Starved_2
NIH3T3_ChrRNA_Starved_3
NIH3T3_G4access_Starved_3
NIH3T3_ChrRNA_FCS_15min_1
NIH3T3_G4access_FCS_15min_1
NIH3T3_ChrRNA_FCS_15min_2
NIH3T3_G4access_FCS_15min_2
NIH3T3_ChrRNA_FCS_15min_3
NIH3T3_G4access_FCS_15min_3

Genome browser session
(e.g. UCSC)

https://genome.ucsc.edu/s/Cyril_Esnault/hg19_tracks_esnault_et_al_manuscript_PDS
https://genome.ucsc.edu/s/Cyril_Esnault/hg19_tracks_esnault_et_al_manuscript_alpha_amanitin

Methodology

Replicates

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

Sequencing depth

	Sequenced reads	% aligned hg19	%aligned dm6	Ends
Raji_PolII_T0_PDS1	76259482	92.97	2.85	Paired
Raji_PolII_T0_PDS2	55373464	92.55	3.02	Paired
Raji_PolII_T10_PDS1	41480600	92.18	3.14	Paired
Raji_PolII_T10_PDS2	72778747	91.11	2.75	Paired
Raji_PolII_T30_PDS1	63274863	93.07	2.80	Paired
Raji_PolII_T30_PDS2	63394482	92.63	2.82	Paired
Raji_PolII_T60_PDS1	39629398	92.27	3.02	Paired
Raji_PolII_T60_PDS2	50858067	92.55	2.52	Paired
Raji_PolII_a-amanitin_0	45718545	88.64	n/a	single
Raji_PolII_a-amanitin_18h	33089764	86.90	n/a	single
Raji_PolII_a-amanitin_12h	34887489	89.30	n/a	single
Raji_PolII_a-amanitin_24h	33100975	86.59	n/a	single
Raji_PolII_a-amanitin_36h	36181944	88.04	n/a	single
Raji_Mnase-seq_a-amanitin_36h	112039139	92.85	n/a	single
Raji_Mnase-seq_1	62261437	98.08	n/a	Paired
Raji_Mnase-seq_2	60791243	97.97	n/a	Paired
Raji_Mnase-seq_a-amanitin_0	104651872	97.38	n/a	single
Raji_Mnase-seq_a-amanitin_18h	74101850	98.22	n/a	single
Raji_Mnase-seq_a-amanitin_12h	68774551	98.17	n/a	single
Raji_Mnase-seq_a-amanitin_24h	77606227	92.00	n/a	single
RRaji_Mnase-seq_DMSO_1	447494863	98.19	n/a	Paired
Raji_Mnase-seq_DMSO_2	158497010	98.21	n/a	Paired
Raji_Mnase-seq_Triptolide_1	278624563	98.21	n/a	Paired
Raji_Mnase-seq_Triptolide_2	246732278	98,16	n/a	Paired
Raji_Dnase-seq	46857249	97,37	n/a	Paired
Raji_TT-seq TT-seq_T0_PDS1	101826817	88.41	1.42	Paired
Raji_TT-seq TT-seq_T0_PDS2	100194398	88.93	1.65	Paired
Raji_TT-seq TT-seq_T10_PDS1	75120618	89.30	0.90	Paired
RajI_TT-seq TT-seq_T10_PDS2	106917842	90.21	0.94	Paired

	Sequenced reads	% aligned mm39	%aligned dm6	Ends
NIH3T3_G4access_Starved_1	16989511	87.67	0.03	Paired
NIH3T3_G4access_Starved_2	24965224	83.57	0.04	Paired
NIH3T3_G4access_Starved_3	14825875	80.48	0.06	Paired
NIH3T3_G4access_FCS_15min_1	18196487	79.89	0.04	Paired
NIH3T3_G4access_FCS_15min_2	20196122	78.47	0.05	Paired
NIH3T3_G4access_FCS_15min_3	21172930	80.90	0.03	Paired

	Sequenced reads	% aligned hg19	Sequenced ERCC reads	Ends
Raji_chrRNA-seq_T0_PDS_rep1	56655887	93.60	487155	Paired
Raji_chrRNA-seq_T0_PDS_rep2	64989206	95.90	522964	Paired
Raji_chrRNA-seq_T10_PDS_rep1	60144820	96.60	449610	Paired
Raji_chrRNA-seq_T10_PDS_rep2	49780663	94.50	429856	Paired

	Sequenced reads	% aligned mm39	Sequenced ERCC reads	Ends
NIH3T3_ChrRNA_Starved_1	43988301	95.87	164323	Paired
NIH3T3_ChrRNA_Starved_2	66245805	94.09	217860	Paired
NIH3T3_ChrRNA_Starved_3	17284116	93.74	72013	Paired
NIH3T3_ChrRNA_FCS_15min_1	35165406	96.32	145349	Paired
NIH3T3_ChrRNA_FCS_15min_2	59158731	89.19	169049	Paired
NIH3T3_ChrRNA_FCS_15min_3	55059302	87.98	182529	Paired

Antibodies

Total PolII (N20 - sc899 - Santa Cruz Biotechnology)

Peak calling parameters

Peak calling methods and parameters are clearly stated in the methods section.

Data quality

Data quality was assessed with FastQC

Software

All sequencing reads were aligned to the hg19 genome assembly using Bowtie2 (allowing 2 mismatches, keeping uniquely aligned reads only) or TopHat2.
Wig files were generated using PASHA (Fenouil et al. 2016; https://doi.org/10.1093/bioinformatics/btw206)
For ChIP-seq experiments, all samples were input subtracted and signals were scaled to reads per million so that the same number of units appeared in each experiment."