Dimeric G-quadruplex motifs determine a large fraction of strong replication origins in vertebrates

Jérémy Poulet-Benedetti[1], Caroline Tonnerre-Doncarli[1], Anne-Laure Valton[1,2,3], Marc Laurent[1], Marie Gérard[1], Natalja Barinova[1], Nikolaos Parisis[1], Florian Massip[4,5,6], Franck Picard[7*] and Marie-Noëlle Prioleau[1*]

Affiliations

[1] Université Paris Cité, CNRS, Institut Jacques Monod, F-75013 Paris, France.

[2]Present address: Department of Systems Biology, University of Massachusetts Chan Medical School, Worcester, MA, USA

[3]Present address: Howard Hughes Medical Institute, Chevy Chase, MD, USA

[4] MINES ParisTech, PSL-Research University, CBIO-Centre for Computational Biology, 75006 Paris, France

[5] Institut Curie, Paris, Cedex, France

[6] INSERM, U900, Paris, Cedex, France

[7]Laboratory of Biology and Modelling of the Cell, Université de Lyon, Ecole Normale Supérieure de Lyon, CNRS, UMR5239, Université Claude Bernard Lyon 1, Lyon, France.

* Correspondence: franck.picard@ens-lyon.fr and marie-noelle.prioleau@ijm.fr

Abstract

Replication of vertebrate genomes is tightly regulated to ensure accurate duplication, but our understanding of the interplay between genetic and epigenetic factors in this regulation remains incomplete. Here, we investigated the involvement of three elements enriched at gene promoters and replication origins: guanine-rich motifs potentially forming G-quadruplexes (pG4s), nucleosome-free regions (NFRs), and the histone variant H2A.Z, in the firing of origins of replication. We show that two pG4s on the same DNA strand (dimeric pG4s) are sufficient to induce assembly of an efficient minimal replication origin without inducing transcription, and that dimeric pG4s in replication origins are located in an NFR next to well-positioned nucleosomes enriched in H2A.Z. Thus, our data suggest a crucial role for dimeric pG4s in the organization and duplication of vertebrate genomes.

Introduction

DNA replication is a tightly constrained and highly regulated process that allows cells to completely duplicate their genome during each cell cycle. Replication is initiated at sequences called origins of replication and occurs in units of about 0.2 to 2 Mb (replication timing domains) at early, mid, or late replication (*1–3*). The spatiotemporal regulation of this program is controlled by complex interactions between multiple genetic and epigenetic determinants that are bound by essential initiation proteins forming the pre-replication complex (pre-RC), at the origin of replication. In budding yeast, the genetic determinant is an 11-bp consensus sequence motif that defines an autonomous replication sequence, and additional neighboring cis-elements are needed to form a nucleosome-free region (NFR) and an effective origin (*4*). In vertebrates, genome-wide mapping of replication origins has revealed a strong association of efficient origins with CpG islands (CGIs) and promoters (*5*, *6*). Rather than a consensus sequence as seen in yeast, however, the only vertebrate determinant identified is contained in a cage-like local secondary DNA structure of guanine-rich sequences known as potentially forming G-quadruplexes (pG4s). These motifs are enriched at replication origins and are essential for origin activity (*7–10*). However, a single pG4 is insufficient to define an active origin, suggesting that a more complex assembly of elements could be necessary to trigger replication initiation (*8*, *11*). As in yeast, strong

replication origins in vertebrates co-localize with NFRs and are associated with epigenetic marks such as the histone variant H2A.Z, which has been shown to promote origin activity (*12–16*). Despite recent advances in understanding the key regulatory elements of replication initiation, the lack of a consensus operational *cis*-regulatory element(s) driving origin formation has thus far prevented a causal link to be established between *cis*-elements and specific chromatin organization at origins.

In the present study, we investigated the requirements for and interplay between pG4s, NFRs, and H2A.Z in the formation of origins of replication using the chicken vertebrate β^A-globin promoter/origin as a model system. Using a genetic approach, we show that same-strand dimeric pG4s are necessary and sufficient for the initiation of replication within the model origin, and that this activity is correlated with the formation of an NFR next to well-positioned nucleosomes. We also demonstrate that dimeric pG4s induce specific NFR formation patterns at replication start sites genome-wide. Thus, our results identify dimeric pG4 as a new functional genetic element in the formation and regulation of a large proportion of strong replication origins in vertebrates.

Results

The presence of dimeric pG4s is essential for the chicken β^A-globin origin activity

To delineate novel *cis*-motifs essential for metazoan replication origin activity within gene promoters, we employed a model origin construct that has the capacity to induce strong enrichment of short nascent strands (SNSs), a characteristic marker of replication origins, and to locally advance the replication timing (RT) of a middle-late replicated region in population-based assays in avian DT40 cell lines (Fig. 1A–C). We used the 1.1-kb chicken β^A-globin promoter/origin associated with the IL2R reporter gene fused to a poly A sequence and flanked by two upstream stimulatory factor (USF) binding sites (Fig. 1A). This wild-type construct and mutant versions with specific elements deleted were inserted at the same locus by homologous recombination. We previously identified the essential role in the activity of this model origin of a pG4 (pG4#1 in Fig. 1A) and the downstream 245-bp sequence containing an erythroid-specific factor binding site (CACC box) as well as CCAAT and TATA boxes, which are

binding sites for the ubiquitous transcription factors nuclear factor Y and TATA-binding protein, respectively (Fig. 1A) (*8*). We quantified the capacity of each mutated origin to initiate replication by measuring SNS enrichment in two independent clones. We found that single or combined deletion of the CCAAT and TATA boxes slightly decreased the relative SNS enrichment compared with the full β^A-globin origin, indicating a limited but detectable impact on origin activity (Fig. 1C). Focusing on the CACC box and the partially overlapping pG4#2 (located on the complementary strand of pG4#1), deletion of the box and the adjacent downstream 12-bp sequence of pG4#2 (ΔCACC+pG4#2 mutant) slightly increased origin activity (Fig. 1C). In contrast, deletion of the region containing a third pG4 (pG4#3) framed by the CCAAT and the TATA boxes (ΔCCAAT to TATA mutant) resulted in a dramatic loss of replication activity, comparable to that resulting from deletion of the entire 245-bp module (Fig. 1C) (*8*). To confirm the essential role of pG4#3, we deleted CCAAT and TATA boxes and replaced three guanines with adenines to eliminate the three triplets of Gs (Fig. 1A, asterisks). These mutations resulted in a degree of SNS enrichment similar to that observed when the entire CCAAT box to TATA box sequence was deleted (Fig. 1C, compare ΔCCAAT to TATA with ΔCCAAT+TATA mpG4). Thus, both pG4#1 and pG4#3 are critical for β^A-globin origin activity.

The TATA box plays a role in RT control

While quantification of SNS enrichment at the inserted β^A-globin promoter/origin assesses its capacity to induce replication initiation, measuring RT shifts can establish whether a change in initiation is occurring in the majority of a cell population, and also identifies an advance or delay in RT as positive or negative profile shifts, respectively (*8*, *17*, *18*). The RT shift observed when the full β^A-globin promoter/origin is inserted in the genome on both chromosomes points to the presence of *cis*-elements that ensure early initiation of this origin in the majority of cells (Fig. 1B). As shown in previous studies, the active β^A-globin origin (100% relative SNS enrichment) showed a significant median RT shift of +20.2, which compared with a median shift of +6.5 (~10% relative SNS enrichment) for inefficient origins (Fig. 1D) (*8*, *17*). Insertion of the mutant origins that were found to have little to no effect on origin activity (see Fig. 1C) resulted in a positive RT shift for ΔCCAAT and ΔCACC+pG4#2 and a

negative shift for ΔTATA and ΔCCAAT+TATA box origins (Fig. 1D and figs. S1, S2, and S3A). In contrast, insertion of the inactivate mutant origins ΔCCAAT to TATA and ΔCCAAT+TATA mpG4, in which pG4#3 is inactive, had no effect on RT (Fig. 1D and figs. S3B and S4). These results identify a prominent role of the TATA box on RT. Notably, the delay in RT observed when the TATA box was deleted in active (but not inactive) origins suggests that the origin affects the initiation of flanking origins and/or replication fork progression in the altered region; however, the mechanism by which this may occur is beyond the scope of this study and must await further investigation. Overall, these results confirm the results obtained with relative SNS enrichment and indicate that the TATA box in the β^A-globin promoter/origin DNA region is a key *cis*-acting element for RT control, together with the previously identified USF binding sites (*19*).

Identification of an active metazoan autonomous minimal replication origin

To determine whether particular combinations of pG4#1, pG4#3, CCAAT box, and/or TATA box are necessary and sufficient to induce the formation of a functional and efficient origin, we designed a minimal 90-bp β^A-globin origin containing pG4#1 followed by a 13-bp linker naturally present in the origin, and the CCAAT to TATA box sequence containing pG4#3 (Fig. 2A). This minimal origin resulted in a 2.5 increase in SNS relative enrichment compared with the active β^A-globin full origin (Fig. 2B). This increased activity may have resulted from the shift in primer pairs used for SNS quantification relative to the pG4s in the minimal origin compared with the full origin (132 bp *vs* 305 bp downstream of pG4#3). In agreement with this hypothesis, a new primer pair (primers 1'), located 317 bp downstream of pG4#3 in the minimal origin, resulted in an SNS enrichment of 125%, similar to that observed for the full origin (Fig. 2B). Additionally, the minimal origin gave a median RT shift of +15.2, which is also comparable to that of the active full origin (Fig. 2C and fig. S5). These results suggest that the 90-bp minimal β^A-globin origin is active and efficient and that these *cis*-elements are sufficient to form a functional origin. To probe this further, we deleted the CCAAT and TATA boxes from the minimal origin (Fig. 2A, middle) and examined the effects on SNS enrichment and RT shift. Notably, these results demonstrated that the CCAAT and TATA boxes are not essential for the replication initiation

activity of the minimal origin but deletion of the TATA box impairs the ability of the minimal origin to locally advance the RT (Fig. 2B and C and fig. S6).

The replication origin determinant is defined by two pG4s on the same DNA strand

We previously showed that the orientation of pG4#1 relative to the 245-bp ectopic β^A-globin module determined the location of replication initiation 3′ of the G tract found inside pG4#1 (Fig. 2D and (*8*)). Therefore, to refine our definition of a functional replication origin, we investigated the effect of switching pG4#1 to the complementary strand on the activity of the β^A-globin minimal origin (pG4#1 rev compl, Fig. 2A). Surprisingly, this switch eliminated SNS enrichment both downstream and upstream of pG4#1 (primer pairs 0 and 1, Fig. 2B), suggesting that the positioning of pG4#1 and pG4#3 on the same strand is essential for minimal origin activity. This raised the question of why pG4#1 could be inverted when associated with the 245-bp module (Fig. 2D). We hypothesized that pG4#2 may be able to cooperate with pG4#1 rev compl as part of a functional origin. To test this hypothesis, we employed a previously reported pG4#1 deletion β^A-globin origin (ΔpG4#1), which is devoid of origin activity (*8*), and attempted to rescue the wild-type origin activity by switching pG4#2 to the same strand as pG4#3 (ΔpG4#1-pG4#2 compl, Fig. 3A). Consistent with our hypothesis, we observed a diffuse enrichment of SNS (~30% at positions 1, 2, and 1′, Fig. 3B) and also detected a significant RT shift toward earlier replication (Fig. 3C and fig. S7) with the ΔpG4#1-pG4#2 compl origin. The latter result confirmed that this origin is active and also suggested that the pattern of SNS enrichment observed probably reflected a more diffuse replication initiation site. Taken together, these results indicate that firing of this model origin depends on the presence of two pG4s positioned on the same strand. We found that all three possible pairs of pG4 combinations led to a functional origin, as did the presence of linkers of various sizes, suggesting that there is considerable flexibility in the arrangement of pG4s within the model origin (Fig. 3D).

Chromatin organization at the minimal β^A-globin origin is similar to that observed at strong origins

In vertebrates, “strong” replication origins (i.e., highly efficient origins, experimentally defined as the top 25% most active origins in terms of SNS enrichment) are characterized by three important features. First, initiation occurs at a well-positioned but labile nucleosome; second, they are associated with an NFR encompassing the G-rich sequence*;* and third, the histone variant H2A.Z is associated with nearby nucleosomes (*12*, *14*, *20*). However, because most efficient origins are found at active transcription start sites (TSSs), it is difficult to dissociate the effects of transcription from those of the replication process on nucleosome organization. Our β^A-globin minimal origin circumvents this problem because it does not drive transcription and therefore allows us to examine only the causal link between nucleosome organization and replication origin function (table S1). We used micrococcal nuclease–deep sequencing (MNase-Seq) to map nucleosomes on the minimal origin and included the inactive pG4#1 rev compl minimal origin as a negative control. The minimal and mutant origins were inserted on both chromosomes in DT40 cells and MNase-Seq was performed in synchronized cells at the G1/S transition (pre-RC-bound), synchronized cells in G2 phase (not pre-RC-bound) or in asynchronous cells (figs. S8 and S9). We observed well-positioned nucleosomes (nucleosomes +1 to +5, Fig. 4A) flanking an NFR at the β^A-globin minimal origin. Nucleosomes +2 and +3 overlap and therefore cannot coexist on the same chromosome. Upstream of the minimal origin (and the large NFR), nucleosome organization is less well defined but at least four discrete positions can be delineated. Finally, we found that the large NFR included the 90-bp construct and about 100 additional bp upstream, leaving the 2X USF binding sites uncovered. In contrast to the functional β^A-globin minimal origin, no NFR was detected at the 90-bp minimal nonfunctional origin pG4#1 rev compl (Fig. 4B). Instead, a well-positioned nucleosome could be detected, suggesting that the presence of two pG4s alone is insufficient to exclude the nucleosome. In addition, a merged nucleosome (Fig. 4B, +2/3) replaced the discrete +2 and +3 nucleosomes detected in the active minimal origin. Overall, the positioning of nucleosomes observed on and around the active β^A-globin minimal origin resembled the pattern found genome-wide on aggregated origins containing a G tract (*14*). To determine whether H2A.Z was enriched at these origins, we performed H2A.Z chromatin immunoprecipitation (ChIP) assays and we found enrichment of the histone variant on both the active and inactive origins predominantly at nucleosomes +1 to +5 (Fig. 4), suggesting that pG4 density controls H2A.Z recruitment independently of pG4 orientation. Thus, the

precise nucleosome organization, including the NFR, depends on the presence of two same-strand pG4s (dimeric pG4s) and the results clearly demonstrate that a single pG4 is insufficient to organize a nucleosome pattern characteristic of strong origins. Moreover, we can conclude that the transcriptional machinery is not involved in establishing this nucleosome pattern, given that this minimal origin is not transcribed (table S1). Finally, we observed that the NFR is already formed in G2, a time when the pre-RCs are not loaded, suggesting that licensing is not responsible for its formation (Fig. 4A).

Replication origins are associated with dimeric pG4s genome-wide

To evaluate whether same-strand dimeric pG4s are required to trigger the formation of an NFR and to fire origins, we investigated their colocalization genome-wide. Current estimates suggest that the human and chicken genomes contain ~1.2 million pG4s and ~350,000 pG4s (1.41% and 1.36% of the genome sizes, respectively), among which about 85% are monomeric, ~10% are dimeric (defined here as two pG4s <100 bp apart), ~2% are trimeric, and about 1–2% are greater than trimeric (Fig. 5A) (*21*). Dimeric pG4s will refer to all pG4s grouped in pairs or more. In both human and chicken cell lines, we found that dimeric pG4s were more commonly associated with functional elements (CGIs, TSSs, origins), H2A.Z, and NFRs genome-wide compared with monomeric pG4s (Fig. 5B). In particular, about 50% of dimeric pG4s were associated with replication origins (46% in human, 50% in chicken) compared with ~30% of monomeric pG4s (27% in human, 31% in chicken). This trend was observed in multiple human cell lines (table S2). Although an association between origins and pG4s has previously been established (*7*, *10*), further analysis of our data revealed a much greater enrichment of dimeric pG4s compared with monomeric pG4s at origins, further supporting the putative functional role of dimeric pG4s in replication initiation genome-wide (fig. S10A, tables S3 and S4, and supplementary materials). To verify this, we examined correlations between the number of monomeric and dimeric pG4s at the origins and the strength of origin firing (SNS enrichment). While the strength of origins (CGI and non-CGI) was positively associated with the abundance of both monomeric and dimeric pG4s, the presence of dimeric pG4s had a much stronger influence on origin activity (fig. S10B). Indeed, the presence of

up to two dimeric pG4s at a non-CGI rendered its activity similar to that of the more active CGI origins (fig. S10B).

NFR formation by dimeric pG4s is associated with origin activity genome-wide

We next used MNase-Seq to investigate nucleosome coverage around pG4s across the chicken genome. We found that dimeric pG4s associated with origins were significantly depleted of nucleosomes at the 5′ side but showed a stronger nucleosome coverage on the 3′ side, and this trend was stronger around dimeric pG4s within CGIs (Fig. 5C, dimeric pG4 panels). Note that, as in our minimal origin, the NFR is located 5′ of the G tract (Fig. 5C). The difference in nucleosome coverage at replication origins around dimeric pG4s was even more pronounced outside of CGIs, as illustrated by the fact that non-origin-associated dimeric pG4s were enriched in nucleosomes whereas origin-associated dimeric pG4s marked a transition between an NFR and well-positioned nucleosomes (Fig. 5C, dimeric pG4 non-CGI panel). Moreover, all monomeric pG4s showed a clear enrichment in nucleosomes, irrespective of their proximity to origins, indicating that the depletion in nucleosomes is specific to dimeric origin-associated pG4s (Fig. 5C, monomeric pG4 panel). Similar results were obtained when the complementary strand was analyzed, and these data were further confirmed by NFR coverage profiles obtained by ATAC-Seq (Fig. 5D and fig. S11A–C). Consistent with the observations with our model origin (Fig. 4A), NFRs were already formed in G2, whereas pre-RCs were not assembled on chromatin (figs. S10C and S12). Collectively, these results show that the presence of dimeric pG4s correlates with formation of an NFR next to well-positioned nucleosomes and is also associated with strong replication initiation activity (Fig. 5E).

An investigation of the genome-wide association of NFRs with functional elements revealed strong colocalization of NFRs with origins (~40% total, 16% with strong origins) and with pG4s (30% with monomeric pG4s, 12% with dimeric pG4s) in chicken cells (fig. S10D). The distribution of pG4s around NFRs was strongly dependent on the proximity to CGIs. Within CGIs, NFRs showed a significant peak of dimeric pG4s regardless of the association with origins; in contrast, outside of CGIs, only origin-associated NFRs showed a significant enrichment in dimeric pG4s (Fig. 5F [dimeric pG4

panel] and fig. S11D). Notably, NFRs showed no particular accumulation of monomeric pG4s as compared with random sequences. These data indicate that dimeric pG4s are critical for the establishment of functional origins through NFR formation (Fig. 5F [monomeric pG4 panel] and fig. S11D).

The histone variant H2A.Z is associated with pG4s genome-wide but does not predict origin activity

Consistent with the established regulatory role of H2A.Z in origin selection (*12*), we found that H2A.Z was broadly associated with replication origins (~60%) and with pG4s (65% with monomeric and 79% with dimeric) in chicken cells, consistent with an as yet unknown role for pG4s in the recruitment of this histone variant (Figs. 5B and 6A and fig. S10A). In contrast to NFRs, H2A.Z did not exhibit a specific accumulation pattern around either monomeric or dimeric pG4s (Fig. 6B). Conversely, the presence of dimeric pG4s was not associated with H2A.Z accumulation around either origins or NFRs, which suggested that the presence of H2A.Z was insufficient to induce formation of an origin and NFR (Fig. 6C and fig. S11E). The lack of a specific H2A.Z pattern around dimeric pG4s genome-wide is in agreement with our genetic analyses showing that H2A.Z is recruited at pG4s independently of the formation of an efficient origin or NFR (Fig. 4). The high proportion of H2A.Z peaks not associated with a replication origin (74% in human, 71% in chicken) points to a role for additional regulatory elements, such as the formation of an NFR and the presence of dimeric pG4s, in defining an origin of replication (Fig. 6D). This is consistent with the strong enrichment of H2A.Z peaks at NFRs (73%, fig. S10D). Taken together, these genome-wide analyses suggest that pG4s may participate in H2A.Z recruitment but that this property is insufficient to define a functional origin.

Discussion

In the present study, we sought to investigate the requirements for pG4s, NFRs, and H2A.Z in the formation of replication origins in chicken and human cells. Using genetic dissection of a model origin, the chicken β^A-globin promoter, we showed that the presence of two pG4s on the same strand within

about 100 bp, which is the length of DNA wrapped around a single nucleosome, is sufficient to form an efficient origin. Analyses of chromatin structure showed a causal link between the formation of an NFR at the pG4 dimer and the formation of an active origin. Further, we demonstrated that an array of well-positioned nucleosomes carrying the histone variant H2A.Z, which is known to be a key regulatory factor for origin function, is present downstream of this NFR at the replication initiation site (*12*). We showed that this dimeric pG4 *cis*-signal correlates with a significant fraction of replication origins as well as with the presence of an NFR next to well-positioned nucleosomes containing the H2A.Z variant. Thus, the results of this study establish a direct link between the presence of *cis*-elements and the formation of a significant fraction of origins of replication (including ~30% of strong origins), in vertebrates, and identify the crucial role of chromatin organization in this association. Key replication factors, such as the origin recognition complex and Mdm2 binding protein, are known to more favorably recognize structured G4s, which reinforces the central role of dimeric pG4s in the formation and initiation of vertebrate replication origins (*16*, *22*, *23*). Collectively, these results highlight the essential nature of *cis*-elements that organize the replication of vertebrate genomes through the establishment of a canonical chromatin structure, and additionally provide a foundation for further studies aimed at understanding the mode of action of dimeric pG4s on the dynamics of vertebrate genomes.

Acknowledgments

The authors thank the members of the laboratory of M-N.P for useful insights and discussions. We thank S. Duharcourt, L. Duret, P. Gilardi-Hebenstreit, A. Piazza, V.Studitsky for critical reading of the manuscript. We thank the ImagoSeine core facility of the Institut Jacques Monod platform, notably Magali Fradet and Nicolas Valentin for performing cell sorting. Sequencing of replication timing data was performed by the GENOM'IC core facility at Institut Cochin-Paris.

Funding

This work was supported by grants from the Association pour la Recherche sur le Cancer (ARC-Equipe Labellisée), the Agence Nationale pour la Recherche (ANR-15-CE12-0004-01) and La ligue contre le cancer, comité d'Ile de France (2020-2022). JP-B was a recipient of a ARC doctoral fellowship. NB was

a recipient of an ARC post-doctoral fellowship and NP of a post-doctoral fellowship from the LABEX Who Am I? M-N.P is supported by Inserm.

Author contributions

J.P-B conducted most experiments. C.D performed MNase-seq and ChIP-seq experiments. A-L.V initiated the project, made and analyzed several mutants. M.L analyzed transcriptional activity and SNS enrichment of several clones. M.G analyzed nucleosome positioning along minimal origins. N.B performed cell synchronization. N.P made ATAC-seq experiments. F.M and F.P performed the statistical analyses. J.P-B, C.D, F.P and M-N.P designed the experiments. J.P-B, C.D, F.P and M-N.P wrote the paper with input from A-L.V. F.P and M-N.P supervised the project. F.P and M-N.P obtained funding.

Competing interests

The authors declare no competing interests.

Data and materials availability

All data are available in the main paper or the supplementary materials. DT40 cell lines are available from M-N.P upon request.

Figure Legends

Fig. 1. Identification of two pG4s essential for replication origin activity.

(**A**) Structure of the ectopic chicken β^A-globin full origin. pG4#1-, pG4#2-, and pG4#3-forming DNA sequences are bold underlined; CACC box is in blue; CCAAT box is in green; and TATA box is in brown. Boxed outlines represent the large deletions in the 'ΔCACC+pG4#2' and 'ΔCCAAT to TATA' mutant origins. Asterisks represent the three guanines mutated to alanines in ΔCCAAT+TATA mpG4. Black lines below and above Primer pairs 1 and 2 indicate the position of amplicons used for short nascent strand (SNS) quantification. (**B**) BrdU pulse-labelled cells were sorted into four S-phase fractions from early to late (S1 to S4) and the immune-precipitated newly synthesized strands (NS) were deep sequenced. RT profiles observed at the targeted middle-late insertion region of chromosome 1

(genomic position: chr1: 72,450,000–72,650,000 bp; 200 kb; galGal5). The insertion site is indicated with a red dotted line. Tracks of nascent strands (NS) enrichments in the four S-phase fractions are shown separately (S1–S4) for the WT and a cell line harboring the ectopic wild-type β^A-globin full origin at the two chromosomes 2 × (β^A-globin full origin) (right). NS-enriched and depleted regions for each fraction are represented in purple and blue, respectively. Initiation zones (IZ) and termination zones (TZ) are labeled. Single reads from SNS aligned, tracks of replication origins (Ori peaks) determined in (*24*), annotated (RefSeq) genes, and CpG islands are shown below. (**C**) Relative SNS enrichment for the β^A-globin full origin and the indicated mutants. Error bars indicate the standard deviation for two independent clones and qPCR duplicates, mean values are indicated above each bar. Primers 1 and 2 refer to the two amplicons at the initiation site of the ectopic origin and Bckg refers to the amplicon located 5 kb away from the site of integration (background signal), respectively. One amplicon within the endogenous ρ-origin was arbitrarily set at 100% to quantify the relative SNS abundance. (**D**) Distribution of RT shift values for clonal cell lines containing active or mutant origins. Boxes show the median (red line) and 0.25 and 0.75 quartiles (lower and upper box edges), and minimum and maximum RT shift values (whiskers). Filled circles represent individual clones (n = 2–10 as indicated). ***P < 0.001 by Wilcoxon's nonparametric two-tailed test.

Fig. 2. Two pG4s on the same DNA strand are sufficient to form an efficient origin

(**A**) Structure of the ectopic β^A-globin minimal origin. Mutant sequences with CCAAT and TATA boxes deleted (ΔCCAAT+TATA), and pG4#1 switched to the reverse complement strand (pG4#1 rev compl) are shown. See Fig. 1A for explanation of colored and underlined sequences. Positions of primers used for SNS enrichment are indicated at the top. (**B**) Relative SNS enrichment for the β^A-globin minimal origin and mutants. Error bars indicate the standard deviation for two independent clones and qPCR duplicates, mean values are indicated above each bar. (**C**) Distribution of RT shift values for the two mutant containing active origins. Boxes are as described for Fig. 1D. **P < 0.01, ***P < 0.001, ns = not significant by Wilcoxon's nonparametric two-tailed test. (**D**) Schematic showing the impact of pG4#1 orientation with respect to the 245 bp module (sequence shown below) on the location of the origin

previously observed in (*9*). The pG4s hypothetically involved in the initiation of replication are indicated on the bottom of each box.

Fig. 3. The combination of dimeric pG4s necessary to form an efficient origin is flexible

(**A**) Structure of the ectopic β^A-globin ΔpG4#1-pG4#2 compl origin. See Fig. 1A for explanation of shaded and bold sequences. Positions of primers used for SNS enrichment are indicated at the top. (**B**) Relative SNS enrichment for the ΔpG4#1-pG4#2 compl mutant. Error bars indicate the standard deviation for two independent clones and qPCR duplicates, mean values are indicated above each bar. (**C**) Distribution of RT shift values for the ΔpG4#1-pG4#2 compl clonal cell line. Boxes are as described for Fig. 1D. ***P < 0.001, ns = not significant by Wilcoxon's nonparametric two-tailed test. (**D**) List of functional β^A-globin origin constructs. The specific pG4s involved in the formation of the functional origins and the distance in bp between the end of the first pG4 and the beginning of the second pG4 for each construct are shown.

Fig. 4. Chromatin organization at the β^A-globin minimal origin is similar to that at strong origins

(**A**) Nucleosome occupancy profiles (by MNase-Seq) along the ectopic β^A-globin minimal origin and in flanking regions, measured in synchronized cells at the G1/S transition and G2 and in asynchronous cells (AS). The lower two profiles show ChIP results with anti-H2A.Z antibodies on AS samples and the Log2 ratio between ChIP H2A.Z and MNase-Seq data. (**B**) Analysis as in (A) for the inactive pG4#1 rev compl minimal origin. A schematic of the construct is shown between (A) and (B). Black dotted lines delineate the NFR, and red dotted lines indicate the position of the centers of nucleosomes +1 and +5.

Fig. 5. Dimeric pG4s are associated with strong replication origins and NFRs located next to well-positioned nucleosomes genome-wide

(**A**) Distribution of pG4 types in the human (upper) and chicken (lower) genomes. Dimeric, trimeric, and >trimeric refer to 2, 3, and >3 same-strand pG4s positioned <100 bp apart. (**B**) Associations of

monomeric and dimeric pG4s with specific genomic features in human and chicken cell lines. Numbers above bars indicate the log-odd-ratio of the logistic regression to test enrichment with respect to random segments. ***P < 0.001 by the Wald test. (**C**) Nucleosome coverage (number of reads in the input sample from MNase-Seq) around dimeric and monomeric pG4s on the plus strand with respect to their association with origins (Ori/non-Ori). Strong origins are defined as the top 25% most active origins in terms of SNS enrichment. Random origins are as defined in Materials and Methods. Vertical black line corresponds to the first position in pG4s. (**D**) NFR coverage (number of peaks) around dimeric and monomeric pG4s on the plus strand of cells in G1. (**E**) SNS coverage (number of reads) around dimeric and monomeric pG4s on the plus strand. (**F**) Dimeric and monomeric pG4 coverage around NFRs on the plus strand. Coverage plots represent the average number of genomic features in sliding windows of 20 bp, overlapping by 10 bp.

Fig. 6. H2A.Z is associated with replication origins and pG4s genome-wide

(**A**) Dimeric and monomeric pG4 coverage around H2A.Z peaks on the plus (top) and minus (bottom) strands. (**B**) H2A.Z coverage (number of reads) around dimeric and monomeric pG4s on the plus (top) and minus (bottom) strands. (**C**) H2A.Z coverage (number of reads) around NFRs containing dimeric and monomeric pG4s on the plus (top) and minus (bottom) strands. (**D**) Associations of H2A.Z peaks with genomic features in human and chicken cells. Numbers above bars indicate indicate the log-odd-ratio of the logistic regression to test enrichment with respect to random segments. ***P < 0.001 by the Wald test. Coverage plots represent the average number of genomic features in sliding windows of 20 bp, overlapping by 10 bp. See Fig. 5 for additional definitions.

References

1. I. Hiratani, T. Ryba, M. Itoh, T. Yokochi, M. Schwaiger, C.-W. Chang, Y. Lyou, T. M. Townes, D. Schübeler, D. M. Gilbert, Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol.* **6**, e245 (2008).

2. C. Marchal, J. Sima, D. M. Gilbert, Control of DNA replication timing in the 3D genome. *Nat. Rev. Mol. Cell Biol.* **20**, 721–737 (2019).

3. O. Hyrien, Peaks cloaked in the mist: the landscape of mammalian replication origins. *J Cell Biol*. **208**, 147–160 (2015).

4. M. L. Eaton, K. Galani, S. Kang, S. P. Bell, D. M. MacAlpine, Conserved nucleosome positioning defines replication origins. *Genes Dev.* **24**, 748–753 (2010).

5. J.-C. Cadoret, F. Meisch, V. Hassan-Zadeh, I. Luyten, C. Guillet, L. Duret, H. Quesneville, M.-N. Prioleau, Genome-wide studies highlight indirect links between human replication origins and gene regulation. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 15837–15842 (2008).

6. J. Sequeira-Mendes, R. Díaz-Uriarte, A. Apedaile, D. Huntley, N. Brockdorff, M. Gómez, Transcription initiation activity sets replication origin efficiency in mammalian cells. *PLoS Genet.* **5**, e1000446 (2009).

7. C. Cayrou, P. Coulombe, A. Puy, S. Rialle, N. Kaplan, E. Segal, M. Méchali, New insights into replication origin characteristics in metazoans. *Cell Cycle*. **11**, 658–667 (2012).

8. A.-L. Valton, V. Hassan-Zadeh, I. Lema, N. Boggetto, P. Alberti, C. Saintomé, J.-F. Riou, M.-N. Prioleau, G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *EMBO J.* **33**, 732–746 (2014).

9. P. Prorok, M. Artufel, A. Aze, P. Coulombe, I. Peiffer, L. Lacroix, A. Guédin, J.-L. Mergny, J. Damaschke, A. Schepers, C. Cayrou, M.-P. Teulade-Fichou, B. Ballester, M. Méchali, Involvement of G-quadruplex regions in mammalian replication origin activity. *Nat Commun*. **10**, 3274 (2019).

10. E. Besnard, A. Babled, L. Lapasset, O. Milhavet, H. Parrinello, C. Dantec, J.-M. Marin, J.-M. Lemaitre, Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat. Struct. Mol. Biol.* **19**, 837–844 (2012).

11. F. Picard, J.-C. Cadoret, B. Audit, A. Arneodo, A. Alberti, C. Battail, L. Duret, M.-N. Prioleau, The Spatiotemporal Program of DNA Replication Is Associated with Specific Combinations of Chromatin Marks in Human Cells. *PLoS Genet.* **10**, e1004282 (2014).

12. H. Long, L. Zhang, M. Lv, Z. Wen, W. Zhang, X. Chen, P. Zhang, T. Li, L. Chang, C. Jin, G. Wu, X. Wang, F. Yang, J. Pei, P. Chen, R. Margueron, H. Deng, M. Zhu, G. Li, H2A.Z facilitates licensing and activation of early replication origins. *Nature*. **577**, 576–581 (2020).

13. C. S. K. Lee, M. F. Cheung, J. Li, Y. Zhao, W. H. Lam, V. Ho, R. Rohs, Y. Zhai, D. Leung, B.-K. Tye, Humanizing the yeast origin recognition complex. *Nat Commun*. **12**, 33 (2021).

14. C. Cayrou, B. Ballester, I. Peiffer, R. Fenouil, P. Coulombe, J.-C. Andrau, J. van Helden, M. Méchali, The chromatin environment shapes DNA replication origin organization and defines origin classes. *Genome Res.* **25**, 1873–1885 (2015).

15. G. Guilbaud, P. Murat, H. S. Wilkes, L. K. Lerner, J. E. Sale, T. Krude, Determination of human DNA replication origin position and efficiency reveals principles of initiation zone organisation. *Nucleic Acids Res*. **50**, 7436–7450 (2022).

16. A. Kumagai, W. G. Dunphy, Binding of the Treslin-MTBP Complex to Specific Regions of the Human Genome Promotes the Initiation of DNA Replication. *Cell Rep*. **32**, 108178 (2020).

17. C. Brossas, A.-L. Valton, S. V. Venev, S. Chilaka, A. Counillon, M. Laurent, C. Goncalves, B. Duriez, F. Picard, J. Dekker, M.-N. Prioleau, Clustering of strong replicators associated with active promoters is sufficient to establish an early-replicating domain. *EMBO J*. **39**, e99520 (2020).

18. D. J. Massey, A. Koren, High-throughput analysis of single human cells reveals the complex nature of DNA replication timing control. *Nat Commun*. **13**, 2402 (2022).

19. V. Hassan-Zadeh, S. Chilaka, J.-C. Cadoret, M. K.-W. Ma, N. Boggetto, A. G. West, M.-N. Prioleau, USF binding sequences from the HS4 insulator element impose early replication timing on a vertebrate replicator. *PLoS Biol.* **10**, e1001277 (2012).

20. R. Lombraña, R. Almeida, I. Revuelta, S. Madeira, G. Herranz, N. Saiz, U. Bastolla, M. Gómez, High-resolution analysis of DNA synthesis start sites and nucleosome architecture at efficient mammalian replication origins. *EMBO J.* **32**, 2631–2644 (2013).

21. K.-W. Zheng, J.-Y. Zhang, Y. He, J.-Y. Gong, C.-J. Wen, J.-N. Chen, Y.-H. Hao, Y. Zhao, Z. Tan, Detection of genomic G-quadruplexes in living cells using a small artificial protein. *Nucleic Acids Res* (2020), doi:10.1093/nar/gkaa841.

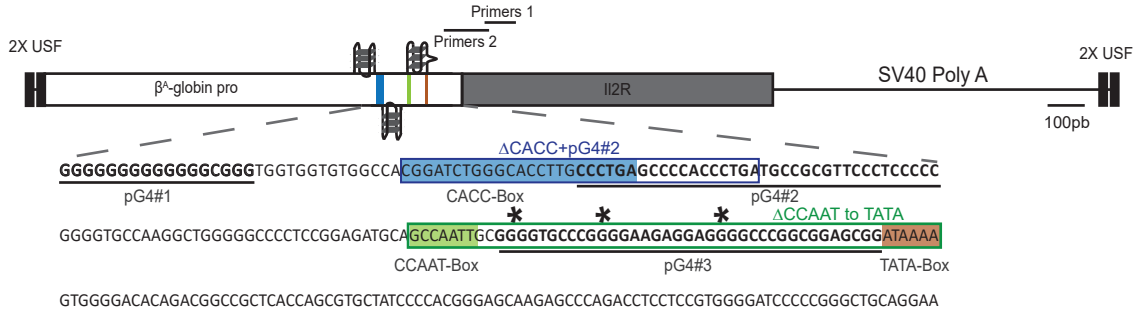
22. S. Hoshina, K. Yura, H. Teranishi, N. Kiyasu, A. Tominaga, H. Kadoma, A. Nakatsuka, T. Kunichika, C. Obuse, S. Waga, Human Origin Recognition Complex Binds Preferentially to G-Quadruplex-Preferable RNA and Single-Stranded DNA. *J. Biol. Chem.* (2013), doi:10.1074/jbc.M113.492504.

23. A. Kumagai, W. G. Dunphy, MTBP, the partner of Treslin, contains a novel DNA-binding domain that is essential for proper initiation of DNA replication. *Mol. Biol. Cell*. **28**, 2998–3012 (2017).

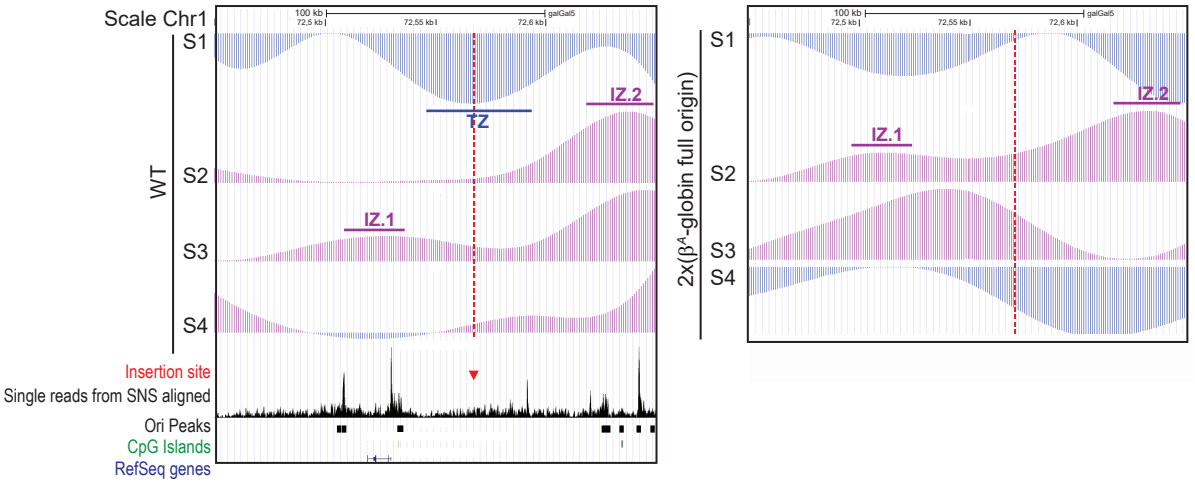
24. F. Massip, M. Laurent, C. Brossas, J. M. Fernández-Justel, M. Gómez, M.-N. Prioleau, L. Duret, F. Picard, Evolution of replication origins in vertebrate genomes: rapid turnover despite selective constraints. *Nucleic Acids Res.* **47**, 5114–5125 (2019).

Figure 1

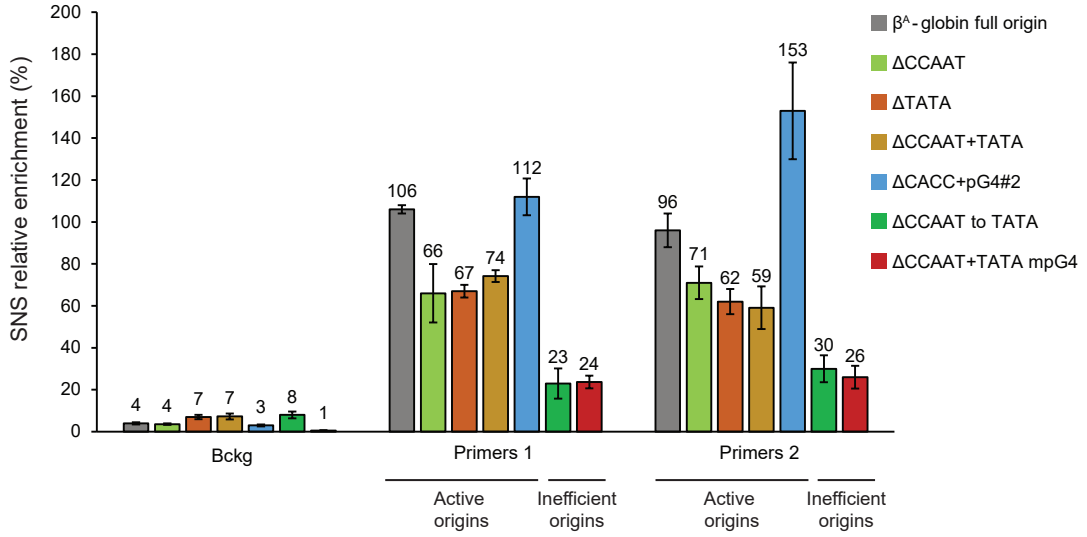
A



B



C



D

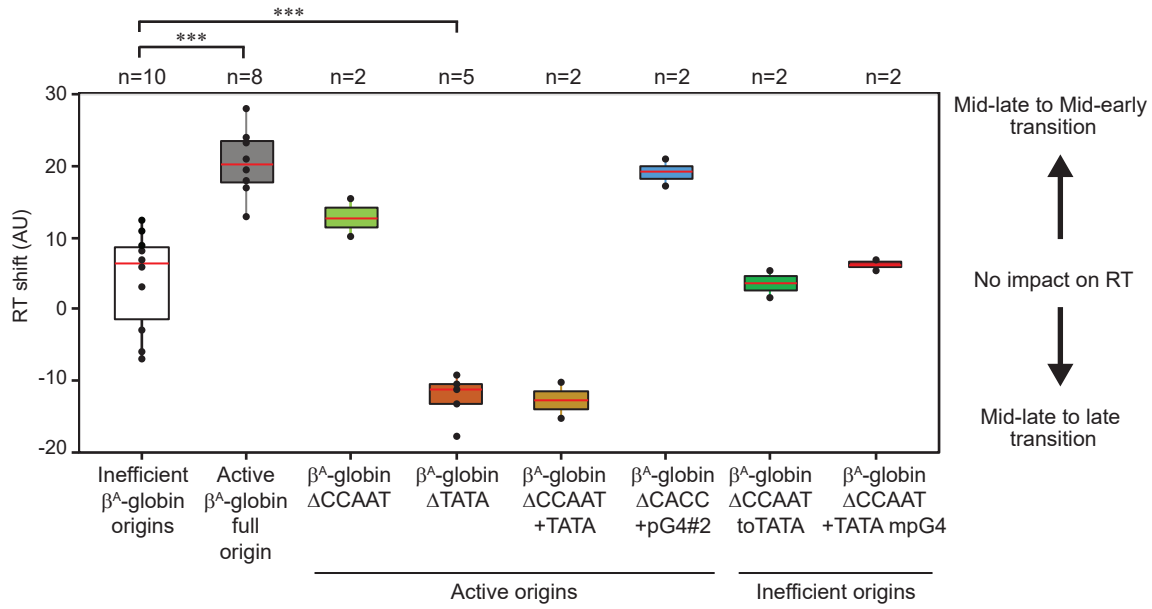
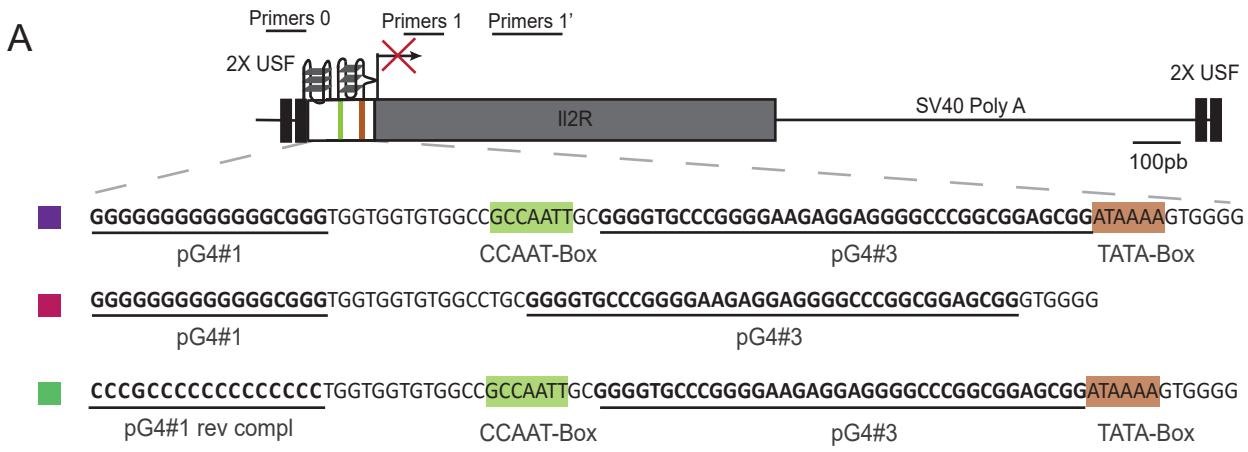
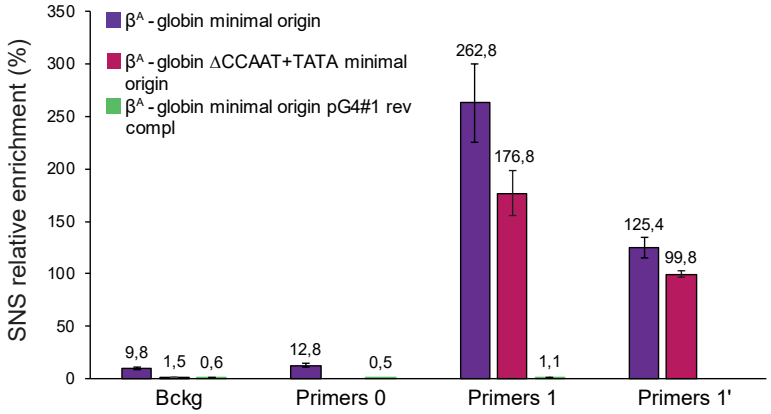


Figure 2

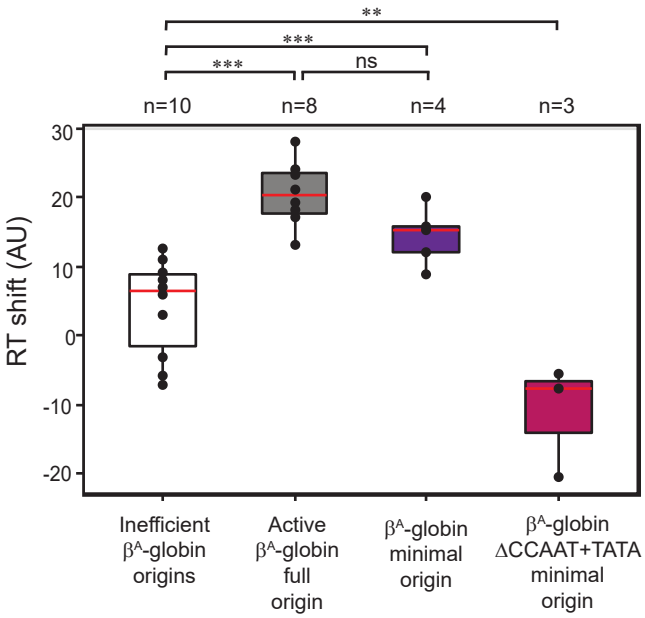
A



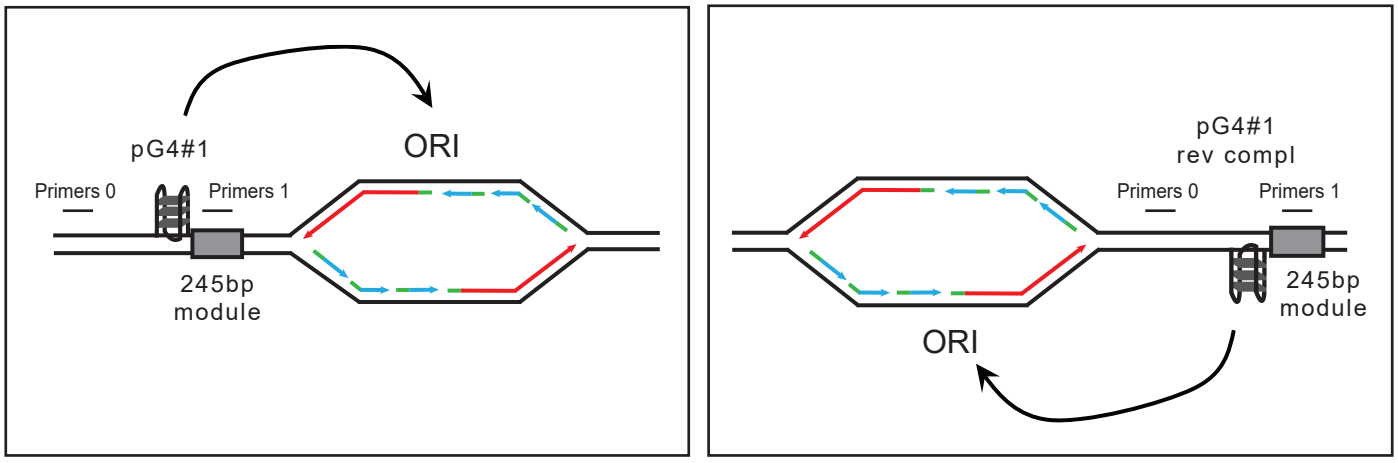
B



C



D



pG4#1+pG4#3

pG4#1 rev compl+pG4#2

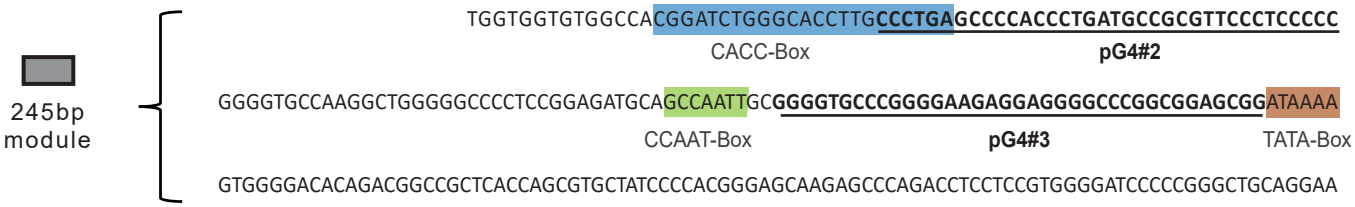
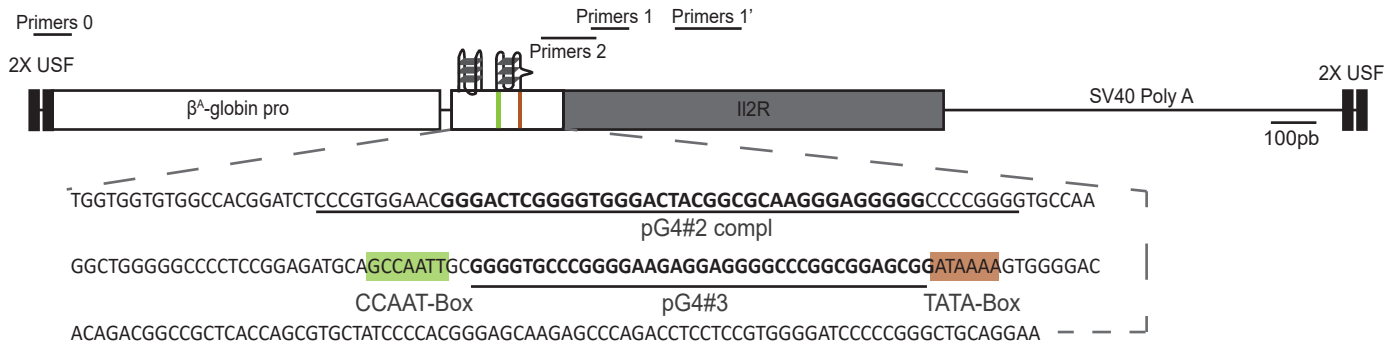
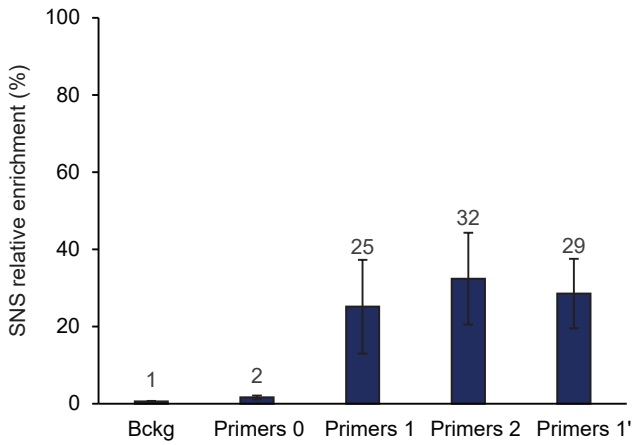


Figure 3

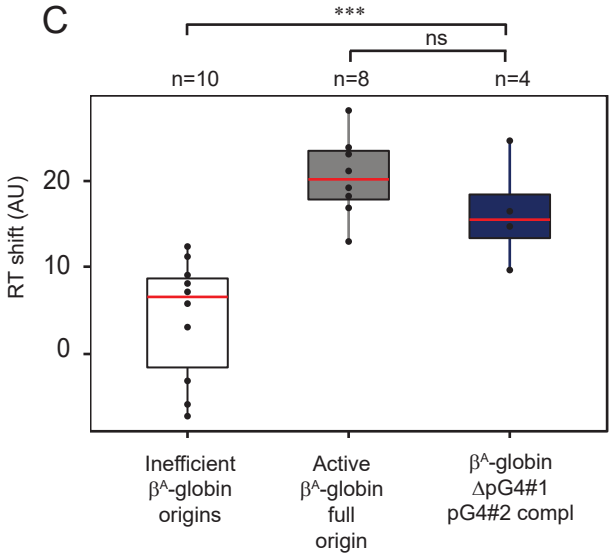
A



B



C



D

Functional constructs	Distance between the two pG4s (bp)	pG4 combination
β^A- globin full origin (Valton et al. 2014)	111	1+3
ΔCCAAT	104	
ΔCACC	88	
β^A-globin minimal origin	22	
pG4#1 rev compl (Valton et al. 2014)	31	1+2
ΔpG4#1-pG4#2 compl	40	2+3

Figure 4

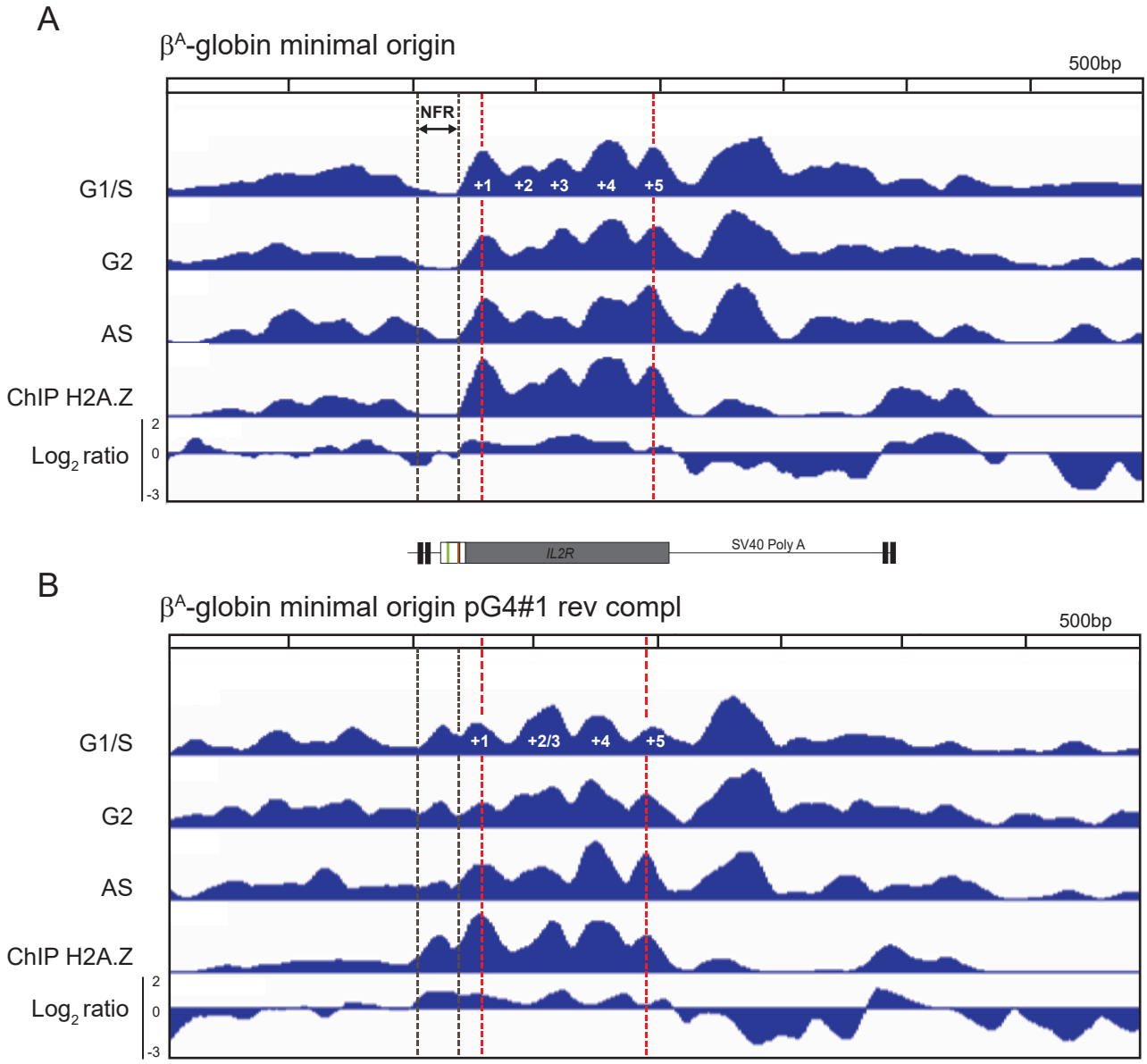

A
βA-globin minimal origin
500bp
NFR
+1
+2
+3
+4
+5
G1/S
G2
AS
ChIP H2A.Z
Log2 ratio
2
0
-3
IL2R
SV40 Poly A
B
βA-globin minimal origin pG4#1 rev compl
500bp
+1
+2/3
+4
+5
G1/S
G2
AS
ChIP H2A.Z
Log2 ratio
2
0
-3

Figure 5

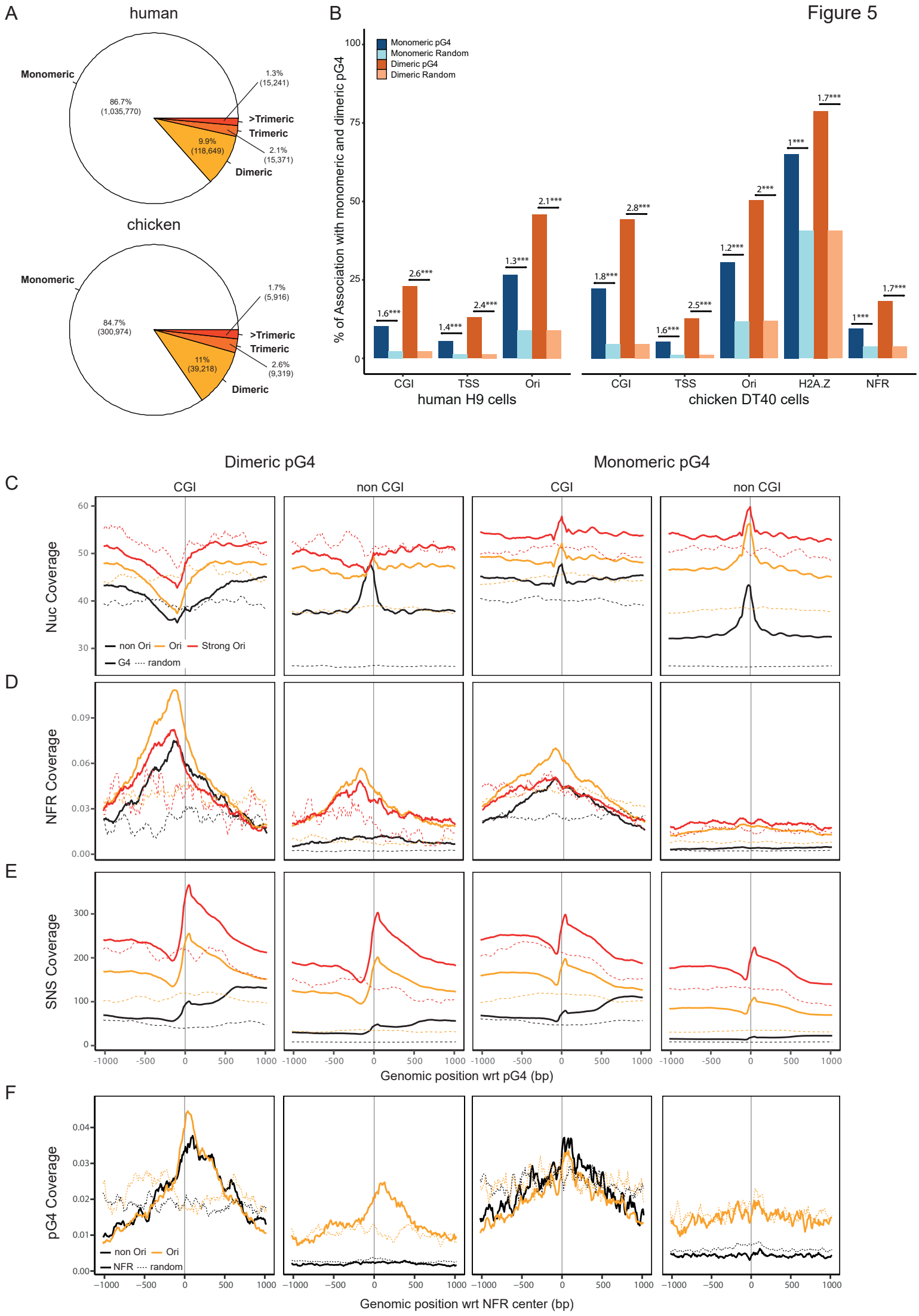
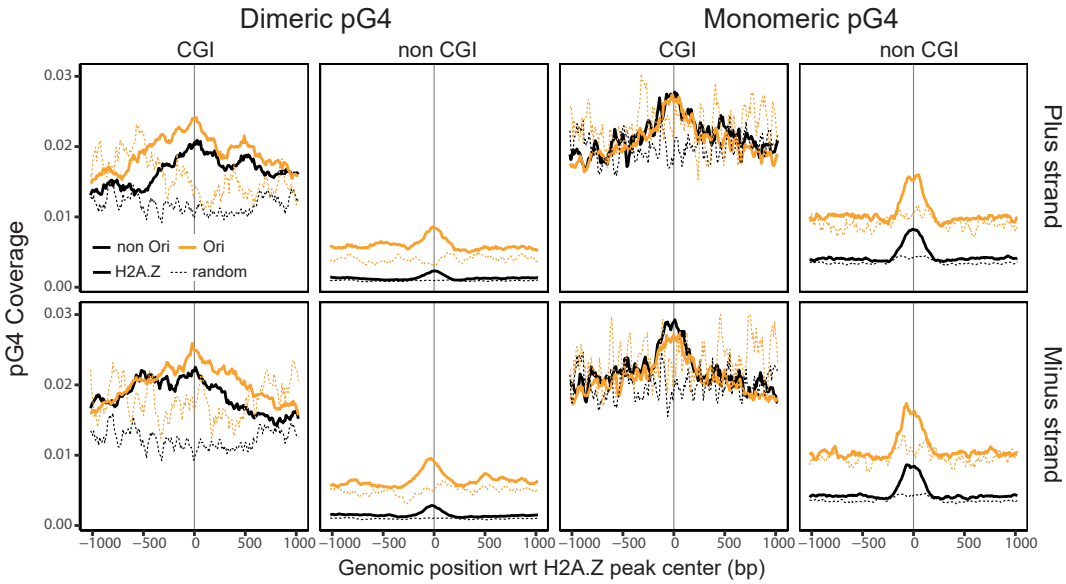
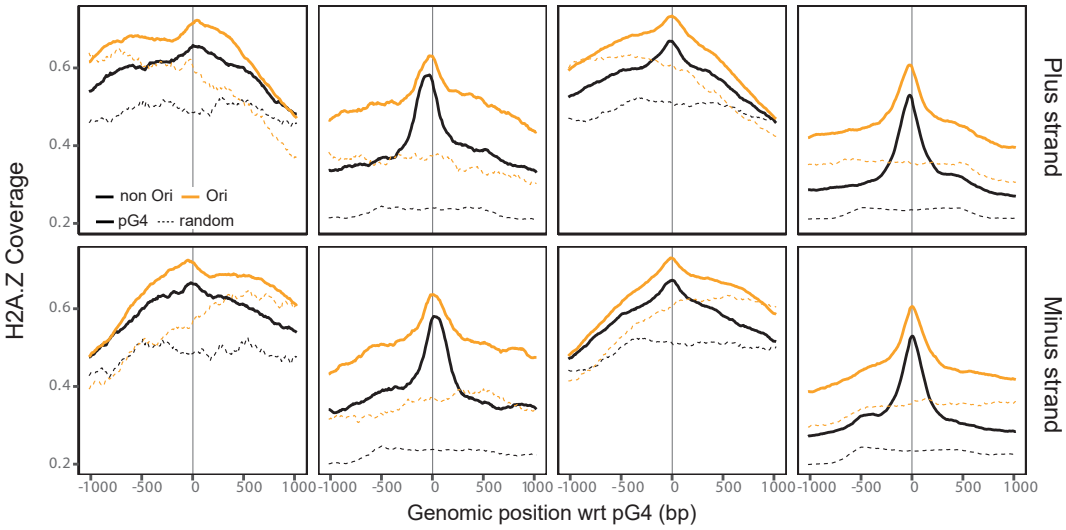

A
human
Monomeric
86.7%
(1,035,770)
1.3%
(15,241)
>Trimeric
Trimeric
9.9%
(118,649)
2.1%
(15,371)
Dimeric
chicken
Monomeric
84.7%
(300,974)
1.7%
(5,916)
>Trimeric
Trimeric
11%
(39,218)
2.6%
(9,319)
Dimeric
B
Monomeric pG4
Monomeric Random
Dimeric pG4
Dimeric Random
% of Association with monomeric and dimeric pG4
CGI
TSS
Ori
human H9 cells
CGI
TSS
Ori
H2A.Z
NFR
chicken DT40 cells
Dimeric pG4
Monomeric pG4
CGI
non CGI
CGI
non CGI
C
Nuc Coverage
non Ori
Ori
Strong Ori
G4
random
D
NFR Coverage
E
SNS Coverage
Genomic position wrt pG4 (bp)
F
pG4 Coverage
non Ori
Ori
NFR
random
Genomic position wrt NFR center (bp)

Figure 6

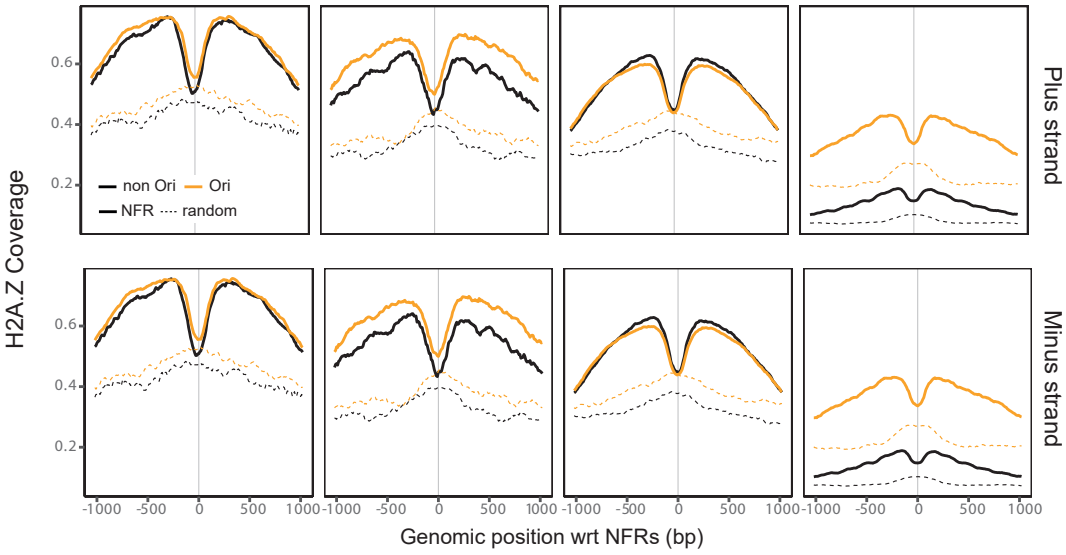
A



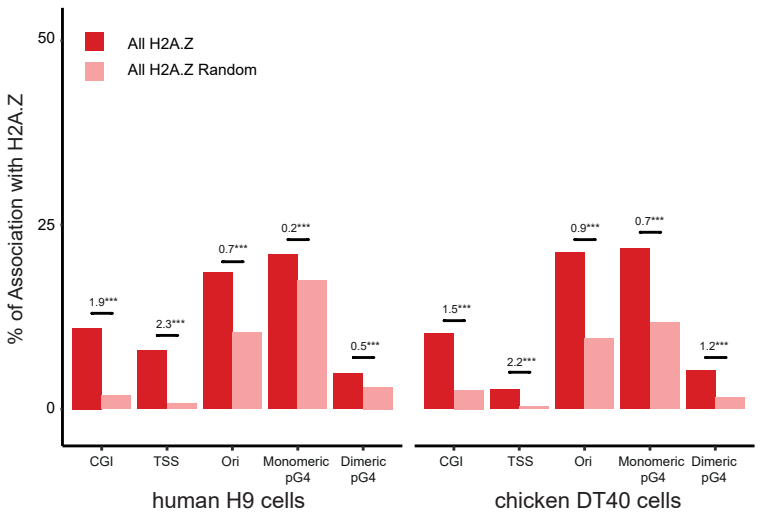
B



C



D



Materials and methods

Plasmid construction

The targeting vectors used for homologous recombination in DT40 cells were constructed with the multisite Gateway Pro kit (Thermo Fischer Scientific #12537100) as previously described (*1*). Vectors containing the 5' and 3' target arms used for specific recombination at the mid-late genomic insertion site (chr1:72,565,520 bp, galGal5) were described previously (*1*). We used four entry vectors to generate the new β^A-globin +β-actin-BsR construct inserted at the mid-late genomic site: two entry vectors containing the 5′ and 3′ target arms for specific insertion, one entry vector (β-actin-BsR) containing the β-actin promoter linked with the *blasticidin* resistance gene (BsR), flanked by *loxP* sites and one entry vector pDONR221 P5-P4 containing the β^A-globin fused to the *IL2R* gene and SV40 PolyA sequence, flanked with two USF binding sites. The corresponding final vector was generated by recombining compatible *att* sites between the entry vectors, with LR clonase (Thermo Fisher Scientific #12538120). For electroporation, the final vector was linearized with *ScaI* (NEB #R3122S). β^A-globin origin mutagenesis was made using overlapping primers replicating the entire entry vector, using the Herculase II fusion DNA polymerase (Agilent #600679) according to manufacturer recommendations. Primers were designed with around 15 bp of overlapping sequence to ensure proper circularization with the In-Fusion HD cloning plus kit (Takara #638909). Mach1 competent cells (Thermo Fisher Scientific #C862003) were used for plasmid cloning.

Cell culture condition and transfection

DT40 cells were grown in RPMI 1640 medium supplemented with Glutamax (Thermo Fisher Scientific #61870010), containing 10% FBS, 1% chicken serum, 0.1mM β-mercaptoethanol, 200U/mL penicillin, 200µg/mL streptomycin and 1.75 µg/mL of amphotericin B at 37°C, under an atmosphere containing 5% CO_2. Cells were electroporated as previously described (*1*). Cell clones were selected on media containing a final concentration of 20 µg/ml blasticidin. Genomic DNA was extracted from cells in lysis buffer (10 mM Tris pH8.0; 25 mM NaCl; 1 mM EDTA and 200 µg/mL proteinase K). Clones into which the plasmid DNA was integrated were screened by PCR with primer pairs designed to bind on one side of the insertion site such that one primer bound within the construct and the other primer bound just upstream or downstream from the arm used for recombination, as previously described (*1*) (fig. S13 and table 7).

The *BsR* resistance gene was excised from positive clones using the Cre-LoxP system. DT40 cells constitutively express a tightly regulated Cre recombinase fused to a mutated oestrogen receptor (Mer) (*2*). This inactive Mer-Cre-Mer fusion protein can be transiently activated in the presence of 4-hydroxytamoxifen, resulting in the efficient excision of genomic regions flanked by two recombination signals (*loxP* sites) inserted in the same direction. For the excision of the genomic DNA flanked by loxP

sites, we treated 3 x 10^5 cells with 5 µM 4-hydroxytamoxifen (Sigma Aldrich #T176) for 24 h. Subclones were obtained by plating dilutions of the treated cell suspension at a density of 50, 150, and 1500 viable cells per 10 ml in 96-well flat-bottomed microtiter plates. Genomic DNA was extracted from single subclones and analyzed by PCR with specific primer pairs (Table 7). We assessed the excision of the β-actin-BsR selection cassette with a primer pair amplifying the polyA sequence of the both *BsR* and *Il2R* genes and the downstream part of the 3' arm of the insertion site (figure S13 C and F, primer pair #3/#4, table 7). All clones were cultured for 72 h in selective media containing the appropriate antibiotic, to confirm PCR results. For each clonal line, the copy number of the construct was quantified by qPCR as previously described ((*1*) and table 8). Homozygous insertions were made in a two-step process. Cells were first heterozygously modified as described earlier with excision of the BsR gene. Once cell lines were established, a second homologous recombination was made to target the other chromosome. Cells were then double screened, one to confirm the maintenance of the first inserted construct (fig. S13E, primer pair #5a/6) and the second to validate proper insertion of the second construct (fig. S13E, primer pair #5b/6).

SNS purification

SNS purification was performed as previously described (*3*) with some slight modifications. Fresh cultured cells were used for total genomic DNA extraction and the T4 polynucleotide kinase (Biolabs #M0201S) concentration was adjusted to 100U and incubated for 30 min at 37°C. Proteinase K (Thermo Fischer Scientific #EO0491) digestion was realized at a final concentration of 625µg/ml for 30min at 50°C.

Replication Timing analysis

The RT experiments were performed as previously described (*3*). Briefly, about 10^7 exponentially growing cells were pulse-labelled with 5-Bromo-2'-deoxyuridine (BrdU, Sigma-Aldrich #B9285) for 1 h and sorted into four S-phase fractions, from early to late S phase. The collected cells were treated with lysis buffer (50 mM Tris pH 8.0; 10 mM EDTA pH 8.0; 300 mM NaCl; 0.5% SDS, 0.2 mg/ml of freshly added proteinase and 0.5 mg/ml of freshly added RNase A), incubated at 56°C for 2 h and stored at -20°C, in the dark. Genomic DNA was isolated from each sample by phenol-chloroform extraction and alcohol precipitation and sonicated four times for 30s each, at 30s intervals, in the high mode at 4°C in a Bioruptor water bath sonicator (Diagenode), to obtain fragments of 500 to 1000 bp in size. The sonicated DNA was denatured by incubation at 95°C for 5 minutes. We added monoclonal anti-BrdU antibody (BD Biosciences #347580) at a final concentration of 3.6 µg/ml in 1x IP buffer (10 mM Tris pH 8.0, 1 mM EDTA pH 8.0, 150 mM NaCl, 0.5% Triton X-100, and 7 mM NaOH). We used 50 µl of protein-G-coated magnetic beads (from Thermo Fisher Scientific #10004D) per sample to pull down the anti-BrdU antibody. Beads and BrdU-labelled nascent DNA were incubated for 2-3 hours at 4°C, on a rotating wheel. The beads were then washed once with 1x IP buffer, twice with wash buffer (20 mM

Tris pH 8.0, 2 mM EDTA pH 8.0, 250 mM NaCl, 0.25% Triton X-100) and then twice with 1x TE buffer pH 8.0. The DNA was eluted by incubating the beads at 37°C for 2hrs in 250 µl 1x TE buffer pH 8.0, to which we added 1% SDS and 0.5 mg/ml proteinase K. DNA was purified by phenol-chloroform extraction and alcohol precipitation and resuspended in 50 µl TE. For Repli-seq analyses, immunoprecipitated NS from the four S-phase fractions collected by flow cytometry or from an asynchronous cell population were amplified by whole-genome amplification (GenomePlex Complete Whole Genome Amplification kit #WGA2; Sigma) according to the manufacturer's recommendations to obtain sufficient DNA amount. After amplification, libraries were constructed as described in the sequencing library preparation section.

Repli-Seq data processing

Paired-end sequencing data were mapped on the galGal5 chicken genome using bowtie2 version 2.3.4.1. For each timing fraction, replication timing was computed using 50 kb sliding windows at 10 kb intervals, normalized by the global and local genomic coverage of the asynchronous cell population, to normalize for total and local coverage variations. Then the centered and standardized timing profiles were smoothed using cubic splines (smooth.spline function of R). In order to provide a single RT profile combining all fractions, we used the method proposed by (*4*), by computing the weighted average WA = (0.750*S1) + (0.583*S2) + (0.417*S3) + (0.250*S4). An increase in WA indicates an earlier timing.

Flow cytometry analysis

After BrdU incorporation, DT40 cells were washed twice with PBS, fixed in 75% ethanol and stored at -20°C. On the day of sorting, fixed cells were resuspended at a final concentration of 2.5 x 10^6 cells/mL in 0.1% IGEPAL in PBS (Sigma, #CA-630), 50 µg/ml propidium iodide and 0.5 mg/ml RNase A, and incubated for 30 minutes at room temperature. Cells were sorted with an INFLUX 500 cell sorter (Cytopeia, BD Biosciences) or a FACSAria Fusion (BD Biosciences). Four fractions of S-phase cells (S1-S4), each containing 5x10^4 cells, were collected and further treated for locus-specific RT analyses.

MNase Digestion

We cross-linked 30 x 10^6 exponential growing cells by incubation for 5 minutes with 1% freshly prepared formaldehyde (Thermo Fisher Scientific, #28908) at room temperature. Fixation was stopped by adding 0.125 M glycine-PBS for five minutes at room temperature. After three ice cold PBS washing, nuclei were extracted in lysis buffer (10mM Tris-HCl, pH7.5, 10mM NaCl, 3mM $MgCl_2$, 0.2% triton X-100, 0.5mM EGTA, 1mM DTT, 1x protease inhibitor cocktail (Sigma, #P8340)) for 5min on ice, centrifuged and resuspended in digestion buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 1 mM $CaCl_2$, 1x protease inhibitor cocktail (Sigma, #P8340)). Micrococcal Nuclease (MNase; Thermo

Fisher Scientific, #EN0181) digestions were performed for 15 minutes at 37°C, using a series of four increasing concentrations of MNase (2.5, 10, 40 and 160 U/mL). The lowest MNAse concentration generated a mixture of oligo-, di- and mono-nucleosomes, whereas the highest concentration produced mostly mono-nucleosomes (fig. S14). The final concentration of 160 U/mL was used for ChIp and nucleosome coverage analyses. The reaction was stopped by adding 0.1 volume of stop buffer (200 mM EDTA pH 8.0, 40 mM EGTA pH 8.0). Chromatin was then sonicated during 30 sec at 7°C in a buffer adjusted to a final SDS concentration of 0.1% with a Covaris sonicator using 75W intensity, 200 cycles per burst, 5% of duty cycle. For nucleosome coverage analyses, DNA molecules were recovered after RNAse A and proteinase K digestion and phenol-chloroform extraction and ethanol precipitation. For ChIp analyses, chromatin was adjusted in 1x IP buffer (20 mM Tris-HCl pH 8.0, 2 mM EDTA pH 8.0, 150 mM NaCl, 1% Triton-X100 and 0.1% SDS).

Chromatin immunoprecipitation (ChIp)

Immunoprecipitation was performed overnight at 4°C, in a final volume of 300 µl of 1X IP buffer (20 mM Tris-HCl pH 8.0, 2 mM EDTA pH 8.0, 150 mM NaCl, 1% Triton-X100 and 0.1% SDS) on an amount of MNAse digested chromatin corresponding to 10 µg of DNA, with anti-H2A.Z antibody (Active Motif, #39013), according to manufacturer recommendations. Immuno-complexes were pulled down with 50 µl of protein-G coated magnetic beads (Thermo Fisher Scientific, Dynabeads protein G, #10004D) per sample. Beads and immuno-complexes were incubated for 2 h at 4°C, on a rotating wheel. The beads were then washed once with 1× IP buffer, twice with wash B buffer (20 mM Tris pH 8.0, 2 mM EDTA pH 8.0, 500 mM NaCl, 0.25% Triton X-100) and then twice with 1× TE buffer pH 8.0. The DNA was eluted by incubating the beads for 2 h at 37°C with 250 µl 1× TE buffer pH 8.0, to which we added 1% SDS and 0.5 mg/ml proteinase K. Cross-linking was reversed by overnight incubation at 65°C, and samples were further treated with 10 µg of RNase A for 15 min at 37°C, and with 20 µg of proteinase K for 1 h at 56°C. DNA was purified by phenol–chloroform extraction, precipitated in alcohol, resuspended in 100 µl TE and quantified by QuBit (dsDNA HS assay kit, Thermo Fisher Scientific, #Q32851). Input samples were obtained from MNAse digested chromatin after RNAse A and proteinase K digestion, phenol-chloroform extraction and ethanol precipitation. These samples were used as a reference for H2AZ peak detection, performed using bowtie2 version 2.3.4.1 and MAC2s version 2.1.2.

RNA extraction and reverse transcription

Total RNA were extracted from $5x10^6$ cells with the Nucleospin RNA kit (Macherey Nagel, #740955). 20 µg of total RNA was then treated with 4 units of DNAse I (NEB, #M0303S) for 1 h at 37°C. The enzyme was inactivated by adding 5 mM EDTA and incubating the reaction mixture for 10 min at 75°C. The RNA was then purified by phenol-chloroform extraction and ethanol precipitation. Reverse transcription reactions (RT +) were then performed with 5 µg of RNA and random hexamers (NEB,

#S1330S), using the Superscript III Reverse Transcriptase (Thermo Fisher Scientific, #18080093) according to the manufacturer's instructions. Negative controls (RT -) were performed with the same procedure, but without the addition of reverse transcriptase. The comparison of RT + and RT - samples was used to validate DNAse I treatment and the complete digestion of the genomic DNA in the RNA samples.

Real-time PCR quantification of DNA

Real-time qPCRs were executed according to the MIQE guideline. The Techne Prime Pro48 apparatus and the QPCR-SYBR Green mix (Thermo Fisher Scientific, #AB1285B) were used for the real-time PCR quantification of BrdU-labeled nascent strands (NS), genomic DNA extracted from 4-hydroxytamoxifen-treated clonal cell lines, short nascent strands or cDNA. Each sample was quantified at least in duplicates. For all reactions on the Techne Prime Pro48, real-time PCR was performed under the following cycling conditions: initial denaturation at 95°C for 15 minutes, followed by 50 cycles of 95°C for 15 s, 61°C for 30 s, 72°C for 20 s, and fluorescence measurement. Following PCR, a thermal melting profile was used for specific amplicon validation.

Cells sorting by centrifugal elutriation

Elutriation was performed as previously described in (*5*) with slight modifications. We used a Beckman Coulter Avanti® J-26 XP with a JE-5.0 rotor. For MNAse experiments, around 1.5 x 10^9 cells were resuspended into 200ml of elutriation buffer (PBS1X, 1% FBS and 1mM EDTA; filtered on 0.22µm). Cells are injected into a large elutriation chamber of 40 ml subjected to a rotation of 2700 RPM, with a pump flow of 80 ml/min. Once injected, cell population was incubated for 30 min into the chamber for recovery and equilibration. Cells are sorted by size with decreasing centrifugal speed. We recover five fractions of cells (from F1 to F5) with the following parameters of centrifugal force / recovered volume: 2530 RPM/400ml; 2380 RPM/800ml; 2260 RPM/800ml; 2130 RPM/800ml; 1980 RPM/800ml. The Fraction 2 corresponds to G1 cells and the fraction 5 to G2/M cells. Cells were then cultured with classical medium culture at a concentration of 1 x 10^6 cells/ml. They were either directly treated with L-mimosine for 3h, or directly cross-linked and extracted for MNase digestion. Cells synchronisation in G1/S was performed using 160 µg/ml of L-mimosine for 3h (Sigma Aldrich #M0253).

For ATAC experiments, around 3 x 10^8 cells were resuspended into 100 ml of elutriation buffer and injected into a standard elutriation chamber of 4 ml subjected to a rotation of 3800 RPM, with a pump flow of 40 ml/min. We recover six fractions of cells (from F1 to F6) with the following parameters of centrifugal force / recovered volume: 3600 RPM/150ml; 3400 RPM/150ml; 3200 RPM/150ml; 3000 RPM/150ml; 2800 RPM/150ml; 2600 RPM/150ml. The fraction 2 corresponds to G1 cells and the fraction F6 to G2/M cells. Cells were used directly for ATAC experiments.

Assay for transposase accessible chromatin (ATAC) experiment

ATAC experiments were performed as described in (*6*). Briefly, 1 x 10^5 asynchronous, G1 and G2/M cells collected by elutriation were washed once with cold PBS. The pellet was resuspended in 50µl of Lysis buffer (0.3M Sucrose, 10mM Tris pH8, 10mM NaCl, 3mM MgCl2, 0.2% Triton X100) and the transposition reaction was incubated for 30 min at 37°C in a thermomixer shaking at 700rpm. Samples were purified with the MinElute PCR purification kit (Qiagen #28004) following manufacturer instructions and eluted in 11µl. The tagmentation and library preparation was performed with the Nextera DNA Library Preparation Kit with 8 cycles of PCR amplification (Illumina #FC-121-1030).

Sequencing library preparation

Sequencing libraries were prepared using NEBNext Ultra II DNA library prep Kit for Illumina (NEB #E7645S) according to the manufacturer's instructions for MNase, ChIp and input libraries. Samples were not subjected to size selection, but only cleaned-up for adaptor-ligated DNA using SPRISelect Reagent kit (Beckman coulter #B23317), to ensure a MNase digested small fragment overall recovery. Libraries were prepared with a starting DNA input of 100 ng for MNAse libraries and 10 ng for ChIp libraries. For ChIp libraries prep, we used a 10 fold diluted adaptor for the adaptor ligation. The library amplification was performed using NEB-Next Multiplex Oligos for Illumina (NEB #E7710S) with different NEB-Next index primers and the NEB-Next Universal PCR Primer (NEB #E6861A) with three PCR cycles for MNAse samples or eight PCR cycles for ChIp samples. Sequencing libraries for ATAC samples were prepared using the Nextera DNA library preparation kit (Illumina, #FC-121-1030) according to the manufacturer's instructions. The tagmented DNA was amplified with eight cycles of PCR amplification and the libraries were cleaned up with SPRselect beads (Beckman Coulter, #B23317). The mean size of the library molecules and the quality of the libraries were determined on an Agilent Bio-analyser High Sensitivity DNA chip (Agilent technologies, #5067–4626).

Sequencing

Sequencing was made at the GENOM'IC Cochin institute facility (Paris) on a NextSeq 500 Illumina sequencer. Samples were sequenced with a High Output 150 cycles Flow Cell in paired-end (paired-end reads of 75 bp) according to standard procedures. For MNAse accessibility and H2A.Z ChIp analyses, 200 M of reads were generated. For ATAC analysis, 200M reads were generated.

Databases

CGI annotations were downloaded from the UCSC database (*7*) and TSSs from the Eukaryotic Promoter Database (*8*) for human, and from the Ensembl website (*9*) for chicken. pG4 detections were downloaded from (*10*) and correspond to automatic detections combining pattern matching and manual curation. Clustered pG4s were detected using mergeBed of the BedTools toolset (*11*), with a maximum

distance between pG4s allowed for pG4s to be merged of 100bp (-d option). Replication origins were downloaded from (*12*), and human H2AZ peaks were downloaded from (*13*).

Statistical Analysis

Overlaps between genomic features were performed using intersectBed of the BedTools (*11*), with genomic segments (pG4s, origins, H2AZ, NFR), 1kb extended from their peak coordinate. Regarding the association of origins with pG4s, we inspected separately the association with pG4s ([+] strand) and pC4s ([-] strand) upstream and downstream origins peak respectively. Random segments were sampled using the random function of BedTools (*11*) and match the size and the GC-content distribution of the corresponding feature. The significance of enrichments was assessed using logistic regression (*14*). For instance, the analyses of replication origins associations rely on their relative enrichment in genomic features as compared with random segments. Logistic regression produces log-odds-ratios (logOR) of association: logOR=0 means that the association of this feature is the same in origins and random segments, if logOR>0 the association is stronger in origins, if logOR<0 the association is weaker in origins. Similar analyses were performed to assess the associations of pG4s in their mono or dimeric form, H2A.Z and NFR with genomic features. In the case of origins, H2A.Z and NFR, we also computed their associations with other genomic features when they carry a pG4, as compared with random segments that also carry a pG4 (Supp Tables 3-6). These analyses show that when associated with pG4s, the relative enrichment of origins (or H2A.Z, or NFR) concerns dimeric pG4s and not monomeric pG4s. Details on the analysis methodology are provided in the Supplementary Material. Coverage plots represent the average number of genomic features in sliding windows of 20 bp, overlapping by 10 bp.

ATAC-seq

ATAC-seq reads were trimmed and then mapped to the chicken gal5 genome using version 2.3.4.1 of bowtie2. We removed discordant pairs (inserts >2000bp) as well as reads with quality score below 40 and reads mapping to mitochondrial DNA. Finally, duplicated reads were removed using Picard Toolkit (http://broadinstitute.github.io/picard/). Reads from the three different replicates were pooled together to increase the statistical power. To do so, we first called open regions using macs2 (with option --broad), and peaks close to each other (<30bp) were merged together. We then called NFRs in open regions using the nucleoatac pipeline (*15*).

1. V. Hassan-Zadeh, S. Chilaka, J.-C. Cadoret, M. K.-W. Ma, N. Boggetto, A. G. West, M.-N. Prioleau, USF binding sequences from the HS4 insulator element impose early replication timing on a vertebrate replicator. *PLoS Biol.* **10**, e1001277 (2012).

2. H. Arakawa, D. Lodygin, J. M. Buerstedde, Mutant loxP vectors for selectable marker recycle and conditional knock-outs. *BMC Biotechnol.* **1**, 7 (2001).

3. A.-L. Valton, V. Hassan-Zadeh, I. Lema, N. Boggetto, P. Alberti, C. Saintomé, J.-F. Riou, M.-N. Prioleau, G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *EMBO J.* **33**, 732–746 (2014).

4. Q. Du, S. A. Bert, N. J. Armstrong, C. E. Caldon, J. Z. Song, S. S. Nair, C. M. Gould, P.-L. Luu, T. Peters, A. Khoury, W. Qu, E. Zotenko, C. Stirzaker, S. J. Clark, Replication timing and epigenome remodelling are associated with the nature of chromosomal rearrangements in cancer. *Nat Commun*. **10**, 1–15 (2019).

5. B. Duriez, S. Chilaka, J.-F. Bercher, E. Hercul, M.-N. Prioleau, Replication dynamics of individual loci in single living cells reveal changes in the degree of replication stochasticity through S phase. *Nucleic Acids Res.* **47**, 5155–5169 (2019).

6. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods*. **10**, 1213–1218 (2013).

7. W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, D. Haussler, The human genome browser at UCSC. *Genome Res*. **12**, 996–1006 (2002).

8. R. Dreos, G. Ambrosini, R. Groux, R. Cavin Périer, P. Bucher, The eukaryotic promoter database in its 30th year: focus on non-vertebrate organisms. *Nucleic Acids Res*. **45**, D51–D55 (2017).

9. F. Cunningham, J. E. Allen, J. Allen, J. Alvarez-Jarreta, M. R. Amode, I. M. Armean, O. Austine-Orimoloye, A. G. Azov, I. Barnes, R. Bennett, A. Berry, J. Bhai, A. Bignell, K. Billis, S. Boddu, L. Brooks, M. Charkhchi, C. Cummins, L. Da Rin Fioretto, C. Davidson, K. Dodiya, S. Donaldson, B. El Houdaigui, T. El Naboulsi, R. Fatima, C. G. Giron, T. Genez, J. G. Martinez, C. Guijarro-Clarke, A. Gymer, M. Hardy, Z. Hollis, T. Hourlier, T. Hunt, T. Juettemann, V. Kaikala, M. Kay, I. Lavidas, T. Le, D. Lemos, J. C. Marugán, S. Mohanan, A. Mushtaq, M. Naven, D. N. Ogeh, A. Parker, A. Parton, M. Perry, I. Piližota, I. Prosovetskaia, M. P. Sakthivel, A. I. A. Salam, B. M. Schmitt, H. Schuilenburg, D. Sheppard, J. G. Pérez-Silva, W. Stark, E. Steed, K. Sutinen, R. Sukumaran, D. Sumathipala, M.-M. Suner, M. Szpak, A. Thormann, F. F. Tricomi, D. Urbina-Gómez, A. Veidenberg, T. A. Walsh, B. Walts, N. Willhoft, A. Winterbottom, E. Wass, M. Chakiachvili, B. Flint, A. Frankish, S. Giorgetti, L. Haggerty, S. E. Hunt, G. R. IIsley, J. E. Loveland, F. J. Martin, B. Moore, J. M. Mudge, M. Muffato, E. Perry, M. Ruffier, J. Tate, D. Thybert, S. J. Trevanion, S. Dyer, P. W. Harrison, K. L. Howe, A. D. Yates, D. R. Zerbino, P. Flicek, Ensembl 2022. *Nucleic Acids Res.* **50**, D988–D995 (2022).

10. K.-W. Zheng, J.-Y. Zhang, Y. He, J.-Y. Gong, C.-J. Wen, J.-N. Chen, Y.-H. Hao, Y. Zhao, Z. Tan, Detection of genomic G-quadruplexes in living cells using a small artificial protein. *Nucleic Acids Res*. **48**, 11706–11720 (2020).

11. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. **26**, 841–842 (2010).

12. F. Massip, M. Laurent, C. Brossas, J. M. Fernández-Justel, M. Gómez, M.-N. Prioleau, L. Duret, F. Picard, Evolution of replication origins in vertebrate genomes: rapid turnover despite selective constraints. *Nucleic Acids Res.* **47**, 5114–5125 (2019).

13. H. Long, L. Zhang, M. Lv, Z. Wen, W. Zhang, X. Chen, P. Zhang, T. Li, L. Chang, C. Jin, G. Wu, X. Wang, F. Yang, J. Pei, P. Chen, R. Margueron, H. Deng, M. Zhu, G. Li, H2A.Z facilitates licensing and activation of early replication origins. *Nature*. **577**, 576–581 (2020).

14. J. J. Faraway, *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models, Second Edition* (Chapman and Hall/CRC, New York, ed. 2, 2016).

15. A. N. Schep, J. D. Buenrostro, S. K. Denny, K. Schwartz, G. Sherlock, W. J. Greenleaf, Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. *Genome Res*. **25**, 1757–1770 (2015).