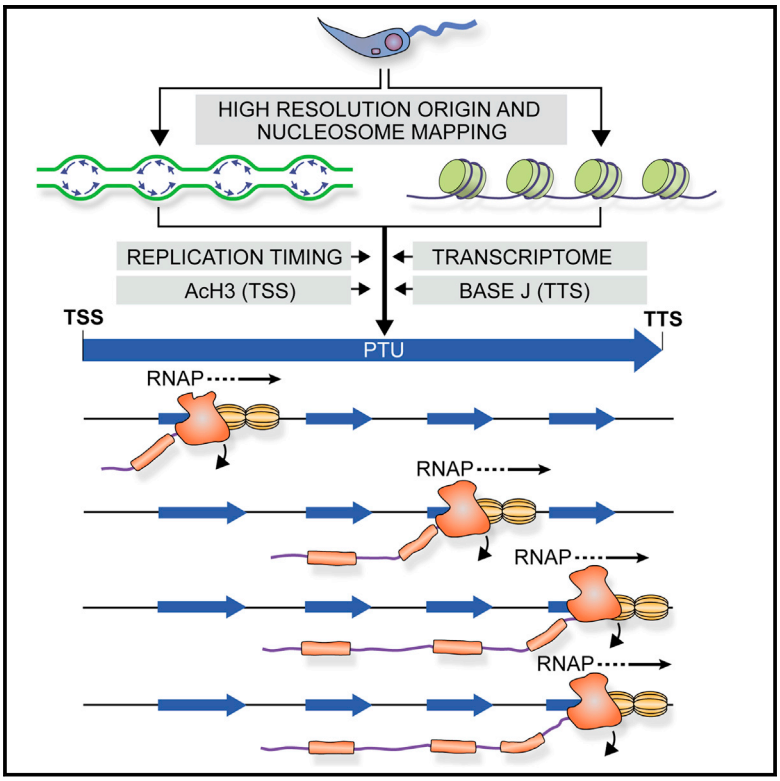
Resource

Cell Reports

# Transcriptionally Driven DNA Replication Program of the Human Parasite *Leishmania major*

## Graphical Abstract




## Authors

Rodrigo Lombraña, Alba Álvarez, José Miguel Fernández-Justel, ..., Arturo Calzada, José María Requena, María Gómez

## Correspondence

mgomez@cbm.csic.es

## In Brief

Though the mechanism of DNA replication origin activation is well understood and conserved among eukaryotes, the determinants of origin specification seem to vary across different model systems. Lombraña et al. found that active transcription is the primary determinant underlying the temporal and spatial replication of the *Leishmania major* genome.

## Highlights

- *L. major* DNA replication origins and nucleosome occupancies are mapped at high resolution
- Those datasets are integrated with mRNA profiling and epigenetic marks
- Nucleosome patterns recapitulate RNA polymerase kinetics
- DNA replication initiates at sites of RNA polymerase pausing and termination

## Accession Numbers

GSE81991

Lombraña et al., 2016, Cell Reports *16*, 1774–1786
August 9, 2016 
http://dx.doi.org/10.1016/j.celrep.2016.07.007

OPEN ACCESS
CellPress

Cell Reports
Resource

# Transcriptionally Driven DNA Replication Program of the Human Parasite *Leishmania major*

Rodrigo Lombraña,[1,3] Alba Álvarez,[1,3] José Miguel Fernández-Justel,[1] Ricardo Almeida,[1] César Poza-Carrión,[1] Fábia Gomes,[2] Arturo Calzada,[2] José María Requena,[1] and María Gómez[1,*]

[1]Functional Organization of the Genome Group, Centro de Biología Molecular Severo Ochoa (CSIC/UAM), Nicolás Cabrera 1, 28049 Madrid, Spain

[2]Centro Nacional de Biotecnología (CSIC), Darwin 3, 28049 Madrid, Spain

[3]Co-first author

*Correspondence: mgomez@cbm.csic.es

http://dx.doi.org/10.1016/j.celrep.2016.07.007

## SUMMARY

**Faithful inheritance of eukaryotic genomes requires the orchestrated activation of multiple DNA replication origins (ORIs). Although origin firing is mechanistically conserved, how origins are specified and selected for activation varies across different model systems. Here, we provide a complete analysis of the nucleosomal landscape and replication program of the human parasite *Leishmania major*, building on a better evolutionary understanding of replication organization in Eukarya. We found that active transcription is a driving force for the nucleosomal organization of the *L. major* genome and that both the spatial and the temporal program of DNA replication can be explained as associated to RNA polymerase kinetics. This simple scenario likely provides flexibility and robustness to deal with the environmental changes that impose alterations in the genetic programs during parasitic life cycle stages. Our findings also suggest that coupling replication initiation to transcription elongation could be an ancient solution used by eukaryotic cells for origin maintenance.**

## INTRODUCTION

Eukaryotic organisms replicate their large chromosomes from hundreds to thousands of DNA replication origins (ORIs). ORIs are specified by the binding of the origin recognition complex (ORC) that, in a cell-cycle-specific manner, recruits additional proteins that enable ORI licensing through the association of two Mcm2-7 complexes (generating the pre-replication complexes, or pre-RCs). Pre-RC activation occurs at the G1-S boundary, depending on the activity of Dbf4-dependent kinase (DDK) and cyclin-dependent kinase (CDK). Several phosphorylation events and the recruitment of additional proteins trigger Mcm2-7 helicase activation and the assembling of two divergent replication forks at each replication origin. The Mcm2-7 helicase then travels with the replication fork, unwinding the DNA ahead of it (Fragkos et al., 2015).

Licensing an adequate number of ORIs before entering the S phase is essential to avoid genome instability. The number of licensed ORIs is larger than the number of activated ones, what has been interpreted as a safeguard mechanism to the many challenges that can occur during S phase (Blow et al., 2011). The mechanism of ORI licensing and activation is relatively well understood and largely conserved among eukaryotes. However, the precise determinants of ORI specification and the selection of the licensed ORIs that are chosen for activation in each S phase are less well known and seem to differ across studied model systems, suggesting that they have varied across evolution. In the yeast *Saccharomyces cerevisiae*, for example, *cis*-acting elements at ORIs influence ORC recruitment to initiation sites, although chromatin environment and other determinants modulate origin choice and activation timing (Rhind and Gilbert, 2013). In *Drosophila* cells, ORC binds to genomic sites that contain no common sequence motifs but are characterized by a faster nucleosome turnover, such as the promoters of actively transcribed genes (Deal et al., 2010; MacAlpine et al., 2010). In mammalian systems, the scenario appears to be even more flexible, and several factors have been related to ORI function, such as the presence of CpG island promoters (Delgado et al., 1998; Sequeira-Mendes et al., 2009; Cayrou et al., 2011; Dellino et al., 2013), G-rich elements with G4-forming potential (Cayrou et al., 2012; Besnard et al., 2012; Picard et al., 2014), specific chromatin configurations comprising labile nucleosomes (Lombraña et al., 2013; Cayrou et al., 2015), or transcriptional activity (Cadoret et al., 2008; Sequeira-Mendes et al., 2009; Martin et al., 2011; Mesner et al., 2011; Lombraña et al., 2015). A plausible explanation for such flexibility in multicellular organisms might account for the accommodation of the replication program to the distinct transcriptional and chromatin patterns arising from a single genome to provide cell-type identity. A restrictive ORI specification might lead to spatial and temporal conflicts among replication, transcription, and chromatin structure that would result in decreased cell fitness. In contrast, coupling DNA replication initiation to cellular physiology would enhance the robustness of the replication program while allowing faster adaptation to environmental or developmental changes (Gilbert, 2010; Méchali, 2010; Sequeira-Mendes and Gómez, 2012).

To test the hypothesis that the genomic distribution of ORIs is integrated with other genomic functions, specifically with gene

transcription, here we choose a simple model system and generate a genome-wide high-resolution analysis of functional replication initiation sites and chromatin configurations from a eukaryotic human parasite: *Leishmania major*. Protozoan parasites of the genera *Leishmania* and *Trypanosoma* branched early in eukaryotic evolution (Baldauf, 2003). Many discoveries of general interest have been made in this group of organisms, emphasizing its value for understanding the evolution of core molecular processes and in particular the simplification and specialization of transcriptional components for parasitic life (Ivens et al., 2005). These microbes present an unusual eukaryotic genome arrangement in which genes are ordered in long polycistronic transcription units (PTUs) that are transcribed from a single promoter at the 5′ end of the cluster (Martínez-Calvillo et al., 2003). As a result of this genomic organization, most of the genome is being constitutively transcribed and RNA polymerase II (Pol II) promoters and termination sites constitute only 1%–2% of gene numbers, with gene expression being primarily controlled by post-transcriptional processes (Martínez-Calvillo et al., 2004). This peculiar genomic organization represents a unique scenario to directly address the interplay between replication and transcription on their shared chromatin template. Researchers started tackling this issue by measuring DNA sequencing depth ratios between early-S and G2 cells in *Trypanosoma brucei* and *L. major* (Tiengwe et al., 2012; Marques et al., 2015). These analyses have led to the conclusion that these microbes replicate their genomes from a remarkably low number of ORIs or even from a single region per chromosome, in the case of *Leishmania*. However, the replication profiles inferred from these approaches lack the resolution to resolve individual initiation sites, thus preventing detailed analysis of the relationships among replication origin distribution, chromatin structure, and gene expression.

We performed a whole-genome high-resolution analysis of active replication initiation sites and nucleosome organization in *L. major* and integrated these results with all published genomic datasets of epigenetic marks and transcript abundance. We found that active transcription is a driving force for the nucleosomal organization of the genome of this organism and that both the spatial and the temporal program of DNA replication arise from the dynamic genomic features associated with gene transcription. Specifically, we show that replication initiation sites occur at the genomic locations where Pol II stalls or terminates, strongly linking transcription kinetics to replication initiation. Thus, the relatively simple framework of the genome of this parasite is providing a well-suited model for evidencing the opportunistic nature of the replication process and its intrinsic robustness. Our findings also suggest that coupling replication initiation to transcription elongation is likely an ancient solution used by eukaryotic cells for ORI maintenance, as suggested in yeast cells defective for transcriptional termination (Gros et al., 2015) and generalized by the detailed genome-wide analysis presented here.

## RESULTS

### Integrative Analysis of *L. major* Whole-Genome Datasets

The *Leishmania* genome comprises 32.8 Mb of DNA distributed among 36 relatively small chromosomes (range 0.28–2.8 Mb) that lack clear subtelomeric regions or centromeres. Protein-coding genes are organized in long PTUs that are exclusively transcribed from one strand by Pol II, each containing tens to hundreds of genes (Figure S1) (Ivens et al., 2005). Regions between PTUs, where polycistronic sense changes, are called switch strand regions (SSRs) and can be convergent or divergent regarding the transcription sense (cSSR or dSSR, respectively). The third type of transcription regulatory region between PTUs, where transcription finishes and restarts at the same strand, is denoted as head-tail (HT). Polyadenylation of the upstream gene and *trans*-splicing of a capped spliced leader (SL) RNA to the downstream gene occur at almost all intergenic regions (IRs) of PTUs' pre-mRNAs to generate mature mRNAs (De Gaudenzi et al., 2011; Requena, 2011). Another peculiarity in this group of parasites is the presence of the non-canonical base J (β-D-glucosyl-hydroxymetiluracyl) in their genomes. Base J synthesis implies oxidation of thymine to 5-hydroxymethyluracil (HOMeU) and subsequent glycosylation of the latter modified nucleotide (Gommers-Ampt et al., 1993). It replaces about 1% of thymine in the genome of *L. major*, most of which (99%) are located at repetitive telomeric regions while the remaining 1% is located between PTUs and at ribosomal and SL loci (van Luenen et al., 2012). Loss of base J results in Pol II transcription readthrough at cSSR and HT regions and the generation of antisense RNAs, implying a role of this modified base in transcriptional termination (van Luenen et al., 2012; Reynolds et al., 2014). Readthrough is modest at cSSRs containing genes transcribed by Pol II, suggesting another mechanism to terminate Pol II transcription at those positions.

To provide a framework for the high-resolution genome-wide analysis of nucleosomal organization and replication initiation sites generated in this study, we integrated and curated the published data on AcH3 (Thomas et al., 2009), base J (van Luenen et al., 2012), and polyadenylated mRNA profiling (Rastrojo et al., 2013). We also identified miss-annotated genomic regions on the LmjF v6.0 reference genome by the presence of local abrupt increases in coverage of sequence reads derived from genomic DNA from promastigote cells grown to the stationary phase; these regions (169 regions comprising 5% of the genome, shade areas) (Table S1) were suppressed from the analysis. Most of these shade areas might represent loci containing genes in higher copy numbers than those currently annotated in the *L. major* genome database (GeneDB); similar findings have been noted in *L. donovani* regarding its reference genome database (Downing et al., 2011).

The annotated dSSRs were identified by the presence of one or two large AcH3 peaks, depending on the distance between the first transcripts of the divergent PTUs, confirming that AcH3 mark Pol II transcription start sites (TSSs) (Thomas et al., 2009). Representative examples are shown in Figure 1A (LmjF.05, position 385, and LmjF.35, position 735). Conversely, 36 of the 37 annotated cSSRs were marked by the presence of a base J peak (van Luenen et al., 2012) whose signal wideness correlates with the size of the untranscribed cSSR and the presence of Pol III-transcribed genes (Figure 1A, LmjF.05, positions 363 and 415, and LmjF.35, position 640, as well as Figure 2A, LmjF.03, position 259, and LmjF.23, position 227). To accurately annotate HT regions, where transcription terminates and restarts

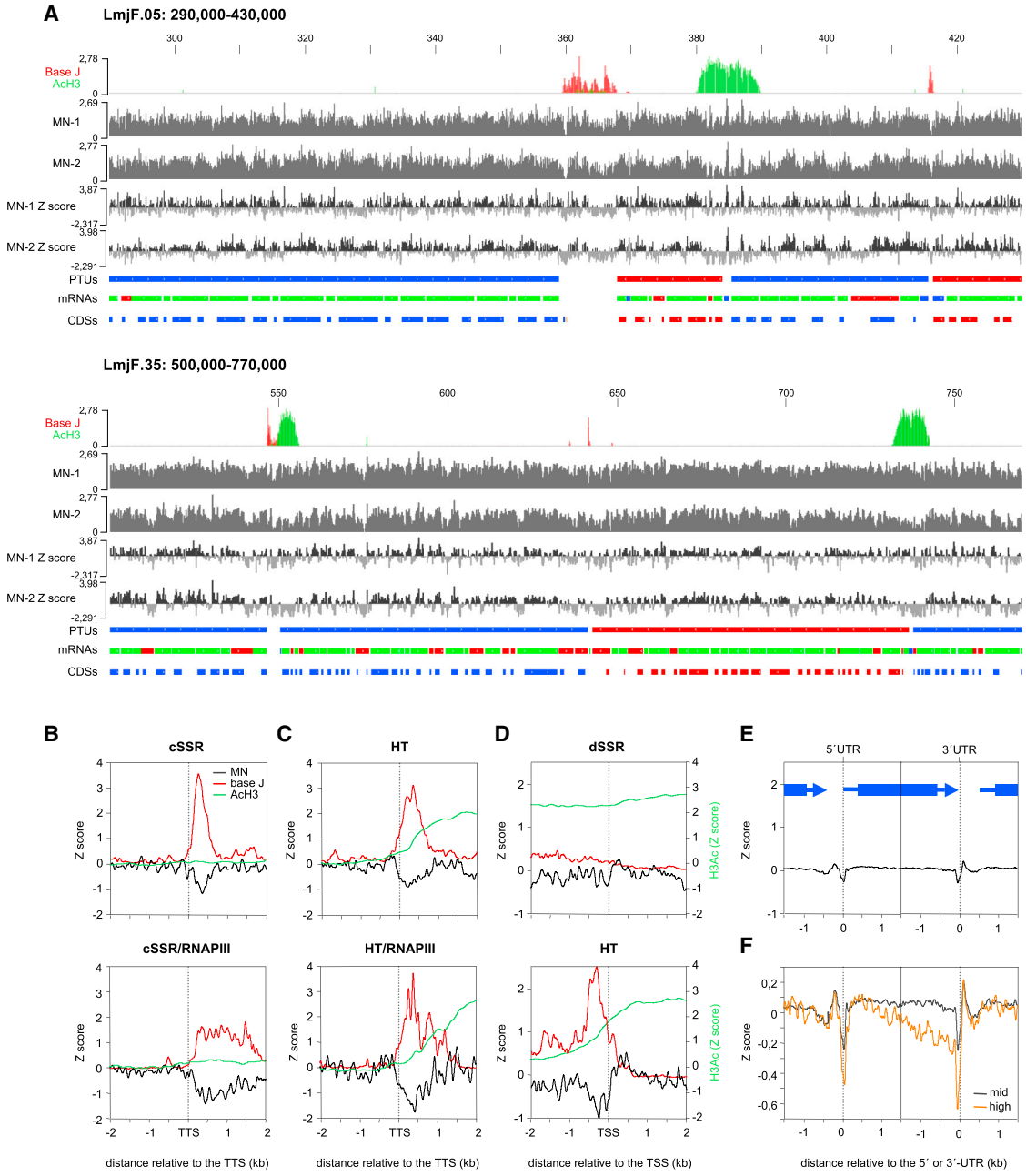



**Figure 1. Nucleosomal Organization of *L. major* Chromosomes**

(A) Representative integrative genomics viewer (IGV) snapshots showing the MN-seq coverage and MN *Z* scores from two biological replicates (micrococcal nuclease [MN]-1 and MN-2), alongside base J and AcH3 positioning (Thomas et al., 2009; van Luenen et al., 2012), PTUs, processed mRNAs (Rastrojo et al., 2013), and CDSs (TriTrypDB v.6.0). Numbers indicate chromosomal positions in kilobases. Blue and red PTUs represent Pol II transcription from the + strand or the − strand, respectively. The same code is used for CDSs. mRNAs are color-coded relative to reported abundance: red, green, and blue for high-, mid-, or low-abundant transcripts, respectively (Table S1) (Rastrojo et al., 2013).

(B) Aggregate plots of MN, base J, and AcH3 *Z* scores within ±2 kb relative to the TTS of cSSR without Pol III genes (n = 22, upper graph) or with Pol III genes (n = 14, lower graph).

(C) Same as in (B) for HT regions without Pol III genes (n = 38, upper graph) or with Pol III genes (n = 13, lower graph).

(D) Same as in (B) for dSSR (n = 52, upper graph) or HT (n = 38, lower graph) within ±2 kb relative to the TSS.

(E) MN profiling within ±1.5 kb relative to the 5′ or 3′ UTR of processed transcripts (n = 7,558). A diagram representing an average transcript is shown at the top, with the 5′ and 3′ UTRs shown as thin lines and CDS as a thick line.

(F) Same as in (E) for genes with mid-level mRNA abundance (88% of all transcripts, charcoal lines) or high mRNA abundance (10% of all transcripts, orange lines) (Table S1).

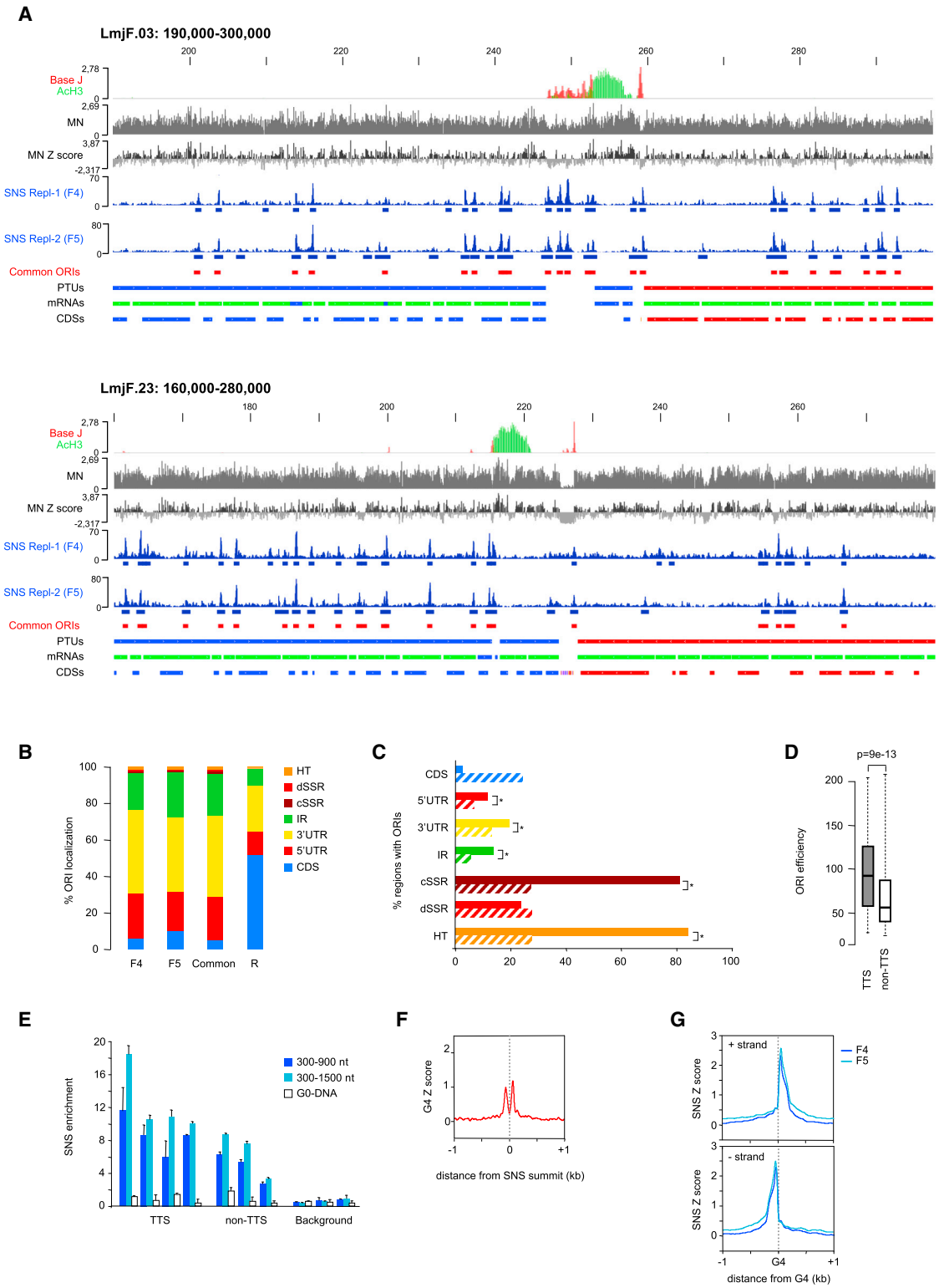



**Figure 2. High-Resolution Mapping of *L. major* DNA Replication Origins**

(A) Representative IGV snapshots showing the SNS-seq coverage and identified ORIs from two biological replicates derived from 300–900 nt SNS (F4) and 300–1,500 nt SNS (F5). See Figure S2A. Common ORIs between replicates (78%) are shown in red. MN-seq profiles and *Z* scores are also shown.

*(legend continued on next page)*

in the same strand, we cross-analyzed base J peaks of non-SSRs with the transcriptomic data and AcH3 locations. This analysis allowed us to identify 51 HT regions defined by the lack of annotated transcription and a presence of a base J peak followed by a strong AcH3 signal (Figure 1A, LmjF.35, position 550; Figure 2A, LmjF.03, position 250, and LmjF.23, position 215). In addition, we carefully curated the transcriptomic dataset to precisely mark the 5′ and 3′ UTR boundaries of most genes, as well as the Pol II TSS and transcription termination site (TTS) at dSSR, cSSR, and the HT region. Altogether, these analyses resulted in the compartmentalization of the *L. major* genome into two functionally distinct fractions: the 98.9% that is subjected to ongoing transcription by Pol II and the remaining ~1% (composed of 143 regions of an average length of 2.2 kb), flanked by the TSS and TTS of the PTUs and frequently containing clusters of Pol III-transcribed genes (Table S1).

## High-Resolution Nucleosomal Landscape of *L. major* Chromosomes

A crucial aspect of genome regulation strongly correlated with gene expression is chromatin structure. To address the chromatin configuration at *L. major* chromosomes, we generated genome-wide maps of nucleosome positioning by MNase digestion coupled to next-generation sequencing (MNase-seq) from exponentially growing cells. Briefly, native chromatin was exposed to MNase digestion, and two independent biological replicates of mononucleosomal DNA were sequenced at high depth. The large coverage obtained (>380-fold per nucleotide) allowed us to examine the nucleosome composition of the *L. major* genome with maximal resolution and accuracy. We found high nucleosome occupancy but little positioning along the long, constitutively transcribed PTUs comprising most of the genome of this parasite (Figure 1A). This is in sharp contrast with the lower nucleosomal occupancy detected at Pol II TTS regions, both at cSSRs and at HT regions (Figures 1B and 1C, upper panels), and with the strong deterring of nucleosomes found at Pol III-transcribed gene clusters (Figures 1B and 1C, lower panels; Figure 2A, LmjF.23, position 226). The extension of the region with decreased nucleosome occupancy mirrors the wideness of the base J signal, indicating a possible role of base J in excluding nucleosomes at TTS regions. The lack of well-phased nucleosomes along the body of the PTUs likely reflects the continuous passage of Pol II, as previously reported for highly transcribed genes in yeast (Rando and Winston, 2012). Similarly, the reduced nucleosomal occupancy at TTS regions is reminiscent of previous findings in yeast (Fan et al., 2010), thus suggesting that this nucleosome depletion might involve a transcription-based termination mechanism likely linked to base J positioning.

Well-positioned nucleosomes were almost exclusively found surrounding the TSS at Pol II promoter regions (both at dSSRs and at HT regions) (Figure 1D) and flanking Pol III gene clusters (Figures 1B and 1C, lower panels). The stereotypical pattern of nucleosome depletion at the TSS flanked by well-positioned −1 and +1 nucleosomes resembles the promoter architecture described in other studied systems (Jiang and Pugh, 2009), supporting the notion that chromatin structure also regulates transcription initiation at the PTUs of this parasite. A well-positioned nucleosome was also detected at the genomic positions corresponding to the spliced-out regions located between the 3′ and the 5′ UTR of contiguous genes (IRs) (Figure 1E). Moreover, the degree of positioning of this IR nucleosome, as charged by the depth of its flanking depleted region, is higher at the subset of genes displaying higher mRNA abundances (Figure 1F; Table S1). Because the positioning of nucleosomes at the intron-exon boundaries both in fission yeast and in mammalian cells has been proposed to be particularly relevant for splicing introns with weak splice sites (Patrick et al., 2013; Spies et al., 2009), our findings suggest a putative parallel role of chromatin configuration in regulating *trans*-splicing efficiency in this parasite. Thus, the nucleosomal landscape of *L. major* reflects the genic organization and expression constraints of this organism, illustrating the active role of both transcription elongation and termination in dynamically modulating chromatin structure.

## High-Resolution Mapping of DNA Replication Origins in *L. major*

To identify the genomic sites of DNA replication initiation in *L. major* promastigote cells, we adapted the most widely used high-throughput mapping procedure in metazoans: small leading nascent strand purification coupled to next-generation sequencing (SNS-seq) (Figure S2A). The accuracy and resolution of ORI detection by SNS-seq strongly relies in two steps: first, on the exclusion of Okazaki fragments by size selection and, second, on the efficient degradation of contaminant, unreplicated DNA by extensive digestion with λ-exonuclease, because the 5′ RNA primer of the leading strands specifically prevents their digestion with this enzyme. SNS-seq has been applied to ORI analysis in human, mouse, and *Drosophila* cell lines, revealing common features of metazoan replication origins

(B) Percentage of localization of ORIs identified at each SNS-seq library and common ORIs. The distribution of randomized ORI locations across the genome is shown (R).

(C) Percentage of genomic regions of each category containing common ORIs (solid bars) and the expected one after distribution of the same number of randomized ORI fragments (dashed bars), $p < 0.0001$. The number and genomic percentage represented within each category are shown in Table S1.

(D) Efficiency of TTS-ORIs (cSSR and HT) and non-TTS-ORIs (3′ UTR, IR, and 5′ UTR) measured as read counts normalized to the size of the peak expressed in kilobases and to library size. The p values are from Wilcoxon rank-sum test.

(E) qPCR validation of SNS enrichments for representative examples of TTS-ORIs and non-TTS-ORIs. The qPCR reactions were performed in duplicates on independent SNS preparations from two consecutive sucrose gradient fractions of the indicated sizes or on G0 DNA subjected to three rounds of λ-exonuclease digestion, as for the SNS samples (open histograms). Fold enrichments were calculated relative to those obtained at nearby genomic regions. Background enrichments and SD bars are indicated. See Figure S2B and Table S2 for primer sequences.

(F) Spatial distribution of G4 motifs (loop size 1–7 bp) within ±1 kb of common ORI summits. See also Figures S2C–S2E.

(G) Aggregate F4 and F5 SNS-seq signal within ±1 kb of ORI-associated (loop size 1–7 bp) G4 motifs occurring on the + DNA strand (upper plot) or − DNA strand (lower plot).

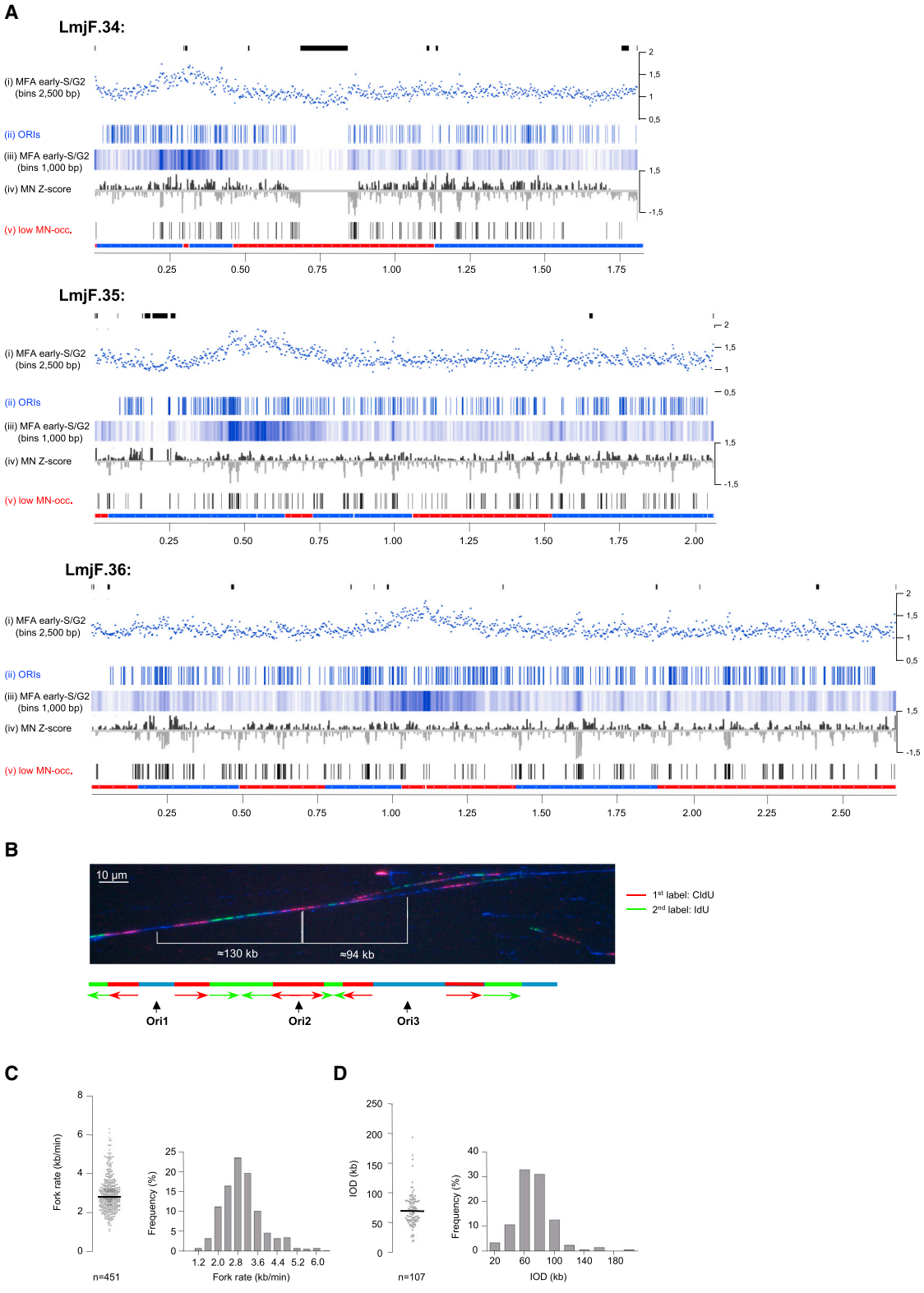



**Figure 3. High ORI Density at Early-Replicating Regions**

(A) IGV snapshots of the three largest *L. major* chromosomes showing, from top to bottom, (i) DNA enrichments in early S relative to G2 computed in 2,500 bp bins (MFA) (Marques et al., 2015), (ii) common ORIs summits, (iii) heatmap of MFA early S relative to G2 in sliding 1,000 bp bins, (iv) MN *Z* scores, and (v) regions of

*(legend continued on next page)*

including association to promoter regions, G-quadruplex motifs, DNA topology, and chromatin configurations (Besnard et al., 2012; Picard et al., 2014; Comoglio et al., 2015; Cayrou et al., 2015).

To map functional replication initiation sites, we purified replication intermediates of increasing sizes from consecutive sucrose gradient fractions, subjected them to three rounds of λ-exonuclease digestion, and performed two independent biological replicates of SNS-seq (Experimental Procedures) (Figure S2A). This enhanced purification protocol through three sequential rounds of λ-exonuclease digestion was found to be essential in eliminating contamination with unreplicated DNA and in minimizing digestion biases (Figure S2B). Deep sequencing of these SNS preparations showed excellent reproducibility, with 78% of ORI peaks independently detected in both replicates (n = 5,100) (Figure 2A). This vast number of ORIs likely reflects the sensitivity of the SNS technique to detect initiation events occurring in few cells of the population and suggests that replication initiates alternately within adjacent ORIs so that each cell in the population uses slightly different combinations of ORIs at a given cell cycle (described later). Peak-calling analysis with data from either individual SNS-seq experiments or common peaks found in both showed a strikingly similar nonrandom genomic distribution: 87%–90% of ORIs located at the boundaries of the *trans*-spliced regions (3′ UTR, IR, and 5′ UTR), 5%–10% mapped at coding DNA sequences (CDSs), and the remaining 3%–4% distributed to SSRs (Figure 2B). However, when ORI distribution across the annotated genomic regions was examined, a remarkable picture emerged: 81%–85% of all Pol II-TTS regions in the genome (cSSR and HT) contain ORIs, while only 10%–20% of all *trans*-spliced boundaries were associated with initiation sites (Figure 2C). Randomization of ORI locations across the genome indicated that these associations were significantly more frequent than would be expected by chance alone ($p < 0.0001$). TTS-associated ORIs showed significant higher firing efficiencies than did non-TTS-ORIs, as estimated by integrating SNS-seq read counts (Figure 2D). Altogether, these observations imply that although the TTS-ORI class accounts for less than 4% of the identified ORIs, this subset of ORIs is used with higher frequency within the cell population and, conversely, that genomic TTS regions are highly enriched in ORIs. Location and firing efficiency of a random subset of ORIs from both classes were confirmed by qPCR on independent preparations of SNS (Figure 2E; Table S2).

As reported for human, mouse, and *Drosophila* cell lines (Cayrou et al., 2012; Besnard et al., 2012; Picard et al., 2014; Comoglio et al., 2015), we also found a significant association between ORIs and predicted G-quadruplex motifs, with up to 74% of all identified ORIs containing at least one strict (loop size 1–7 bp) G-quadruplex motif (Figure S2C). Consistent with those previous findings, G4 motifs are located at both sides of the SNS apex (Figure 2F). To rule out that this association could be due to the presence of G4 motifs at the 5′ end of the sequencing reads, which could prevent λ-exonuclease digestion (Foulk et al., 2015), we eliminated those reads from the analysis and repeated the peak-calling procedure. The number of identified ORIs diminished only slightly in all categories, and their genomic distribution did not change significantly (Figures S2D and S2E). Moreover, we found a similar genomic distribution for ORIs using all sequenced reads and for only those that do not overlap with a G4 motif at their 5′ end, regardless of the peak-calling methodology used (data not shown). These results argue against biases in λ-exonuclease digestion toward GC-rich regions with the potential to form G-quadruplexes when properly controlled. They also support previous findings in various metazoan genomes, suggesting that G-quadruplexes could be among the factors contributing to ORI function (Besnard et al., 2012; Cayrou et al., 2012; Valton et al., 2014). We also examined whether G-quadruplexes could transiently stall replication forks in vivo, as reported in *Drosophila* cells (Comoglio et al., 2015). Similar to those findings, we observed that SNSs of increasing sizes were asymmetrically distributed relative to G4-motif location, thus suggesting that G4 structures could act as replication fork barriers in vivo in *L. major* (Figure 2G).

## Higher Origin Density at Early-Replicating Regions

Visual inspection of ORI location along the chromosomes showed an uneven distribution of replication initiation sites (Figure 3A, ORI track (ii)). The region of higher ORI density matches closely the single replicating region identified at each of the *L. major* chromosomes by analyzing read depth ratios between early-S and G2 cell-sorted cells (marker frequency analysis, or MFA) (Figure 3A, MFA early-S and G2 track (i); Figure S3) (Marques et al., 2015). Moreover, when early-S and G2 MFA data along each chromosome was computed at ORI resolution (sliding windows of 1,000 bp), secondary MFA peaks aligning to clustered ORIs could be clearly resolved (Figure 3A, MFA heatmap (iii); see as examples positions 0.25, 1.65, 1.75, 1.9, or 2.12 Mb in chromosome LmjF.36). Randomization of early-replicating regions across the genome indicates that the observed correlation with ORI density is highly significant ($p < 0.001$) (Figure S4B). This high degree of consistency is remarkable given the differences in the methodologies used between laboratories, constituting an excellent independent validation of our SNS ORI map. In addition, these results strongly suggest that early MFA peaks correspond to early-replicating regions containing a high density of ORIs.

significant low MN occupancies. Chromosome sizes are denoted in intervals of 0.25 Mb. See Figure S3 for data in all *L. major* chromosomes and Figure S4 for late-S and G2 MFA data. The black bars on the top represent shade areas excluded from the analysis (Table S1).

(B) Representative example of a DNA fiber labeled sequentially for 10 min with CldU (red) and IdU (green) used to estimate fork rates and IODs in *L. major* promastigote cells. Scale bar, 10 μm. The corresponding schematic diagram is shown underneath. Ori1 and Ori3 represent replication initiation events occurring before labeling was initiated, and Ori2 represents a replication event occurred during the initial CldU pulse.

(C and D) Fork rate values (C) and IODs (D) calculated from stretched fibers as shown in (B). Horizontal lines in the left graphs represent median values. Graphs on the right represent the percentage of replication signals that correspond to a particular range of replication rate (C) or IOD (D). Data are representative of three replicate experiments and pooled.

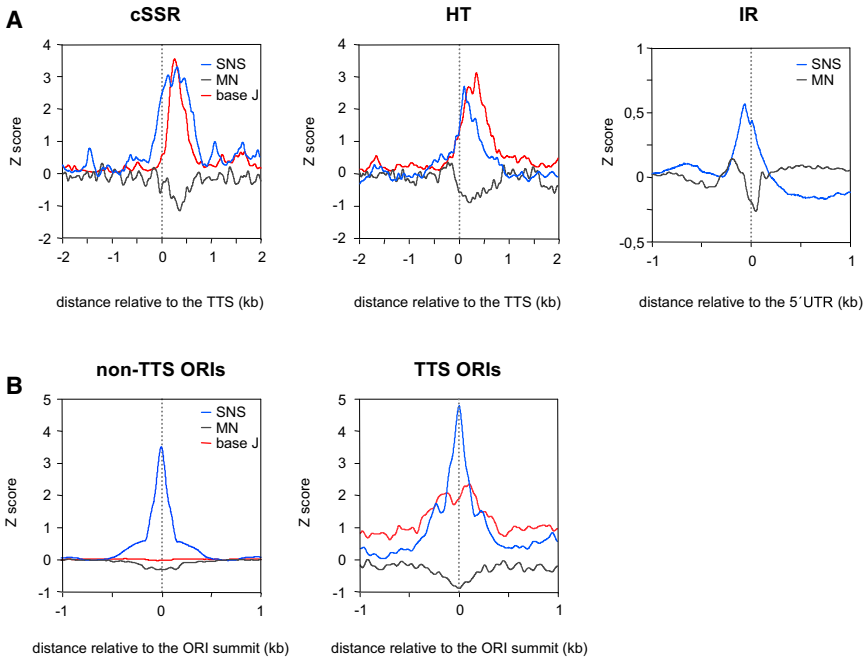



**Figure 4. Low Nucleosome Occupancies at *L. major* ORIs**

(A) Aggregate profiles of MN, base J, and SNS *Z* scores within ±2 kb relative to the TTS of cSSR and HT region (left and middle panels) or within ±1 kb relative to the 5′ UTR of IRs (right panel).

(B) Aggregate profiles of MN, base J, and SNS *Z* scores within ±1 kb relative to the ORI summit of both ORI classes.

The presence of multiple active ORIs per chromosome was confirmed by single-molecule replication analysis through sequential labeling of cells for 10 min with the thymidine analogs chloro-deoxyuridine (CldU) and iodo-deoxyuridine (IdU). Cells were then lysed on microscope coverslips, and DNA regions that contained CldU and IdU were fluorescently labeled via immunochemistry. DNA fibers were counterstained using an antibody directed against single-stranded DNA to ensure that origin tracks used to calculate interorigin distance (IOD) were on the same fiber (Figure 3B). Measurements of fork progression showed that *L. major* had a slower DNA synthesis rate (median of 2.87 kb/min) (Figure 3C) of that reported in *T. brucei* (median of 3.7 kb/min) (Calderano et al., 2015). Although the resolution of IODs by fiber analysis does not allow resolution of initiation sites separated by less than 20 kb, making it impossible to address whether the degree of ORI clustering detected by SNS-seq occurs at the single-cell level or only at the cell-population level, the IODs revealed by this analysis (median of 72 kb) (Figure 3D) demonstrate the firing of multiple ORIs per chromosome.

To test whether the location of higher-efficiency TTS-ORIs might account for the earlier replication of the MFA regions, we carefully examined the distribution of both ORI classes along the chromosomes and found no correlation between early-replicating regions and TTS-ORIs. This indicates that the replication timing profile of *L. major* likely arise from a high density of ORIs firing stochastically at a single early-replicating region per chromosome. This interpretation fits well with a model of replication kinetics in human cells, demonstrating that the timing program can be explained by the location of initiation sites alone, regardless of other factors such as exact initiation probabilities, and that timing profiles are optimally recapitulated by the distribution of DNAase hypersensitive sites (Gindin et al., 2014). To test whether DNA accessibility could predict early replication timing in *L. major*, we used the *Z* scores of our nucleosomal occupancy datasets to estimate the most exposed genomic regions (Figure 3A, tracks (iv) and (v)). Sequence feature analyses of these regions confirmed, as previously reported for other eukaryotes, the enrichment of AT-rich pentamers and the exclusion of GC-rich sequences in nucleosome-depleted DNA, supporting the prominent role of the underlying DNA sequence in nucleosome packaging (Kaplan et al., 2009; Tillo and Hughes, 2009). The density of accessible DNA sites along the chromosomes is highly correlated with the early-S MFA profile ($p < 0.001$) but not with the late-S MFA profile ($p = 0.84$) (Figure S4). These results imply that, as observed in human and *Drosophila* cells (Gindin et al., 2014; Eaton et al., 2011), DNA accessibility in *L. major* is increased at early-replicating regions of the genome.

## Low-Nucleosome Occupancies at Replication Origins

We next analyzed in detail the organization of replication initiation sites and nucleosomes at distinct genomic compartments and found good correlation between lower nucleosome occupancies and higher SNS enrichments (Figure 4A). This association was also observed when analyzing ORIs located at the *trans*-spliced regions, despite SNS signals being lower at those aggregate plots, because they represent *Z* scores and only about 20% of these regions in the genome associate with ORIs (Figure 4A, IR). A similar conclusion was reached when plotting nucleosome distribution across 2 kb surrounding the apex of SNS abundance of both non-TTS and TTS-ORI classes (Figure 4B). In both cases, ORI sites displayed lower nucleosome occupancies than their surrounding regions, although this depletion is more obvious at the TTS-ORI class, likely due to the lower nucleosome occupancy found at regions of transcription termination. Thus, similar to other systems, from yeasts to metazoans (Eaton et al., 2010; MacAlpine et al., 2010; Lombraña et al., 2013; Cayrou et al., 2015), these results indicate that DNA accessibility is one of the contributors to ORI activity in *L. major*.

## Transcriptionally Driven DNA Replication Initiation Landscape

The aggregate plots in Figure 4A show that at most initiation sites detected, SNS abundance peaks immediately upstream of the 5′ UTR of the genes (IR plot), just ahead of the region where Pol II is expected to pause to allow co-transcriptional

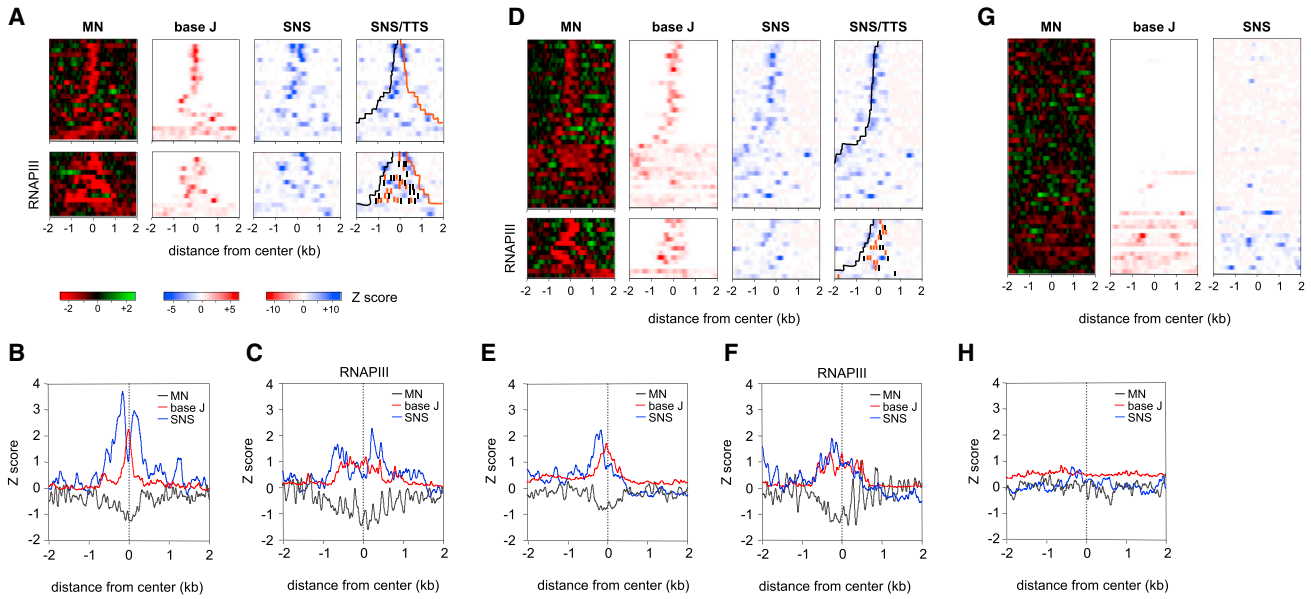



**Figure 5. Transcriptionally Driven ORI Activity at TTS Regions**

(A) Heatmaps of nucleosome occupancies (MN), base J positioning, and SNS enrichments across ±2 kb relative to the center of 36 cSSRs. Regions were ordered from top to bottom by the distance from upstream to downstream TTS and analyzed separately when containing or not containing clusters of Pol III-transcribed genes (Pol III, lower heatmaps). TTS positions are imposed on the SNS heatmaps at the right-most panels (in black or orange when PTUs or tRNA genes are transcribed from the + strand or the − strand, respectively).

(B) Spatial distribution of the MN, base J, and SNS signals shown in (A).

(C) Same as in (B) for cSSRs containing clusters of Pol III-transcribed genes.

(D) Same as in (A) for 51 HT regions ordered from top to bottom by the distance from upstream to downstream TSS.

(E) Spatial distribution of the MN, base J, and SNS signals shown in (D).

(F) Same as in (E) for HT regions containing clusters of Pol III-transcribed genes.

(G) Same as in (A) for 52 dSSRs ordered from top to bottom by the distance from upstream to downstream TSS. Only two dSSRs contain clusters of Pol III-transcribed genes (data not shown).

(H) Spatial distribution of the MN, base J, and SNS signals shown in (G).

*trans*-splicing. In addition, at Pol II termination regions containing a single TTS (85% of which were positive for replication initiation activity) (HT plot), SNS maxima occur immediately adjacent to the base J location and just downstream of the TTS. To dissect further the effect of PTU transcription directionality both on nucleosome occupancy and on SNS abundance, we stratified the cSSR and HT region according to the distance from the upstream neighbor TTS. The heatmaps in Figure 5 express nucleosomal occupancies, the base J location, and the SNS abundance for regions containing either two TTSs (cSSR) (Figure 5A) or a single TTS (HT) (Figure 5D), comprising or not comprising clustered Pol III-transcribed genes (lower and upper panels, respectively). These analyses confirm that nucleosome occupancies exhibit increased depletion at non-transcribed regions relative to the transcribed PTUs, with the transition from high occupancy to lower occupancy coinciding with the position of the TTS. In addition, clustered Pol III-transcribed genes further exclude nucleosomes and, together with base J, likely act as anti-nucleosomal barriers from which neighboring nucleosomes appear to be well phased (Figures 5C and 5F).

Two peaks of maximal SNS abundance were detected at cSSR, each located immediately adjacent to each TTS (Figure 5B). This physical arrangement is even more striking at HT regions, where a single SNS peak mapped adjacent to the single TTS (Figure 5E). A similar picture was observed at regions containing Pol III-transcribed genes, with either two or one SNS peaks detected at regions containing one or two TTSs (Figures 5C and 5F, respectively). In contrast, SNS organization at dSSRs did not show any particular positioning relative to the nearest TSS (Figures 5G and 5H). These results are consistent with Pol II elongation affecting the location of active replication initiation sites.

## DISCUSSION

Here we provide whole-genome mapping of DNA replication initiation sites, together with analysis of both the nucleosomal landscape and the replication timing of the protozoan parasite *L. major*. This study adds a model system that, from an evolutionary perspective, contributes to better understanding of replication organization in Eukarya. A common feature of *L. major* ORIs is their preferential location at genomic positions where Pol II is expected to slow or stall and eventually disassembles. Thus, we found that *L. major* ORIs are mainly located at regions related to the *trans*-splicing events (3′ UTR, IR, and 5′ UTR), and TTS regions. Although transcription kinetics has not been addressed in *L. major* in detail, Pol II pausing near gene promoters and polyadenylation sites is a general feature of eukaryotic cells,

thought to allow time for co-transcriptional 5′ and 3′ end processing (Moore and Proudfoot, 2009; Perales and Bentley, 2009). In addition, work demonstrates that Pol II pausing may assist co-transcriptional splicing in *S. cerevisiae* (Alexander et al., 2010; Carrillo Oesterreich et al., 2010), likely aided by nucleosome positioning at the intron-exon boundaries (Patrick et al., 2013; Spies et al., 2009).

The preferential initiation of DNA synthesis at Pol II pausing or termination sites suggests a high degree of coupling between Pol II elongation and ORI activity in *L. major*. The simplest scenario to accommodate these findings would be that DNA replication opportunistically initiates from genomic sites that have been dynamically rendered available by transcription. In agreement with this, ORI activity occurs at sites with lower nucleosome occupancies in all cases (Figure 4B). As such, the replication program would emerge as an adaptation to the chromatin constraints imposed by transcription. This interpretation fits well with results from our laboratory showing that in mammalian cells, local changes in nucleosome positioning due to increased transcriptional activity at highly efficient TSS-associated ORIs result in parallel changes at the replication initiation sites (Lombraña et al., 2013). Another more provocative scenario would posit that Pol II activity could drive ORI specification by displacing Mcm2-7 complexes along the PTUs. Evidence for this possibility comes from in vitro biochemical data showing that Mcm2-7 complexes remained attached to DNA when ORC complexes are eluted and retain their ability to initiate DNA replication after diffusion along DNA (Gros et al., 2014). Moreover, further work from the same laboratory elegantly demonstrated that in *S. cerevisiae* cells deficient for transcription termination, collisions with Pol II induce a redistribution of Mcm2-7 complexes along chromosomes, resulting in a shift in replication initiation sites (Gros et al., 2015). Along the same lines, work in *Drosophila* cells showed that Mcm2-7 complexes get redistributed from ORC binding sites to non-transcribed portions of the genome in cells blocked at the G1-S transition (Powell et al., 2015). Elongating Pol II machinery pushing Mcm2-7 complexes ahead of it would accommodate our findings: (1) more than 85% of the TTS regions in *L. major* chromosomes contain ORIs (Figure 2C), (2) this ORI class comprises the most efficient ORIs in the cell population (Figures 2D and 2E), and (3) replication initiation activity at these regions occurs precisely downstream of the TTS, at the site of base J-mediated Pol II stalling-induced release or displacement (Figure 5). Similarly, the encountering of Pol II elongation complexes with natural genomic barriers along the PTUs, such as positioned nucleosomes at IRs (Figure 1E), or the presence of G4 structures might impose the stalling and, in some cells, the eventual disassembly of Pol II that would leave Mcm2-7 complexes engaged on DNA at those positions. In addition, although most mapped ORIs are located between genes (Figure 2B), these replication initiation events occur at less than 20% of all IRs in the genome and are overall less efficiently used within the cell population (Figures 2D and 2E). Altogether, these observations suggest that ORIs, as well-defined genetic loci, might not exist in *L. major*. This interpretation is consistent with the long-known observation that *L. major* supports the maintenance of plasmids containing virtually any portion of the genome or even only exogenous DNA (Patnaik, 1997). We propose that at each cell, there is a certain probability of finding an ORI at a particular location, a probability strongly linked to the Pol II transcriptional kinetics of that cell, which at the population level, is maximal at the TTS. In agreement with this hypothesis, we found that the probability of finding an ORI at the 3′ UTR of a gene is higher for highly abundant transcripts: 36% versus 25% for high- versus mid-abundant mRNAs ($p < 0.00001$, chi-square test), as would be expected if Pol II preferentially stalls at those positions when encountering a well-positioned nucleosome (Figure 1F). This correlation does not exist when considering the 5′ UTR side of the genes ($p = 0.2$, chi-square test).

Transcriptionally driven Mcm2-7 complex distribution can also account for the observed replication timing of *L. major* chromosomes. By analyzing DNA sequencing depth ratios between early-S and G2 cells, it was suggested that each chromosome was replicated from a single ORI (Marques et al., 2015). We reanalyzed those datasets with increased resolution and show that early-replicating regions correspond to a higher density of low-efficiency ORIs (Figure 3). These findings provide an explanation for the small peak amplitude and the large constant width of the MFA sequencing profiles along *L. major* chromosomes as likely due to the heterogeneous, population-averaged firing of adjacent origins with similar firing times. However, a replication initiation landscape coupled to transcription kinetics can also help to explain the remarkable degree of synteny of the timing regions between related *Leishmania* species (Marques et al., 2015). In addition, the spatial resolution of our ORI maps that allows discernment of individual initiation events, as well as the average replication profile, might provide additional experimental support for the replication kinetics framework proposed by Bechhoefer and Rhind (2012). This working model posits that replication timing is a systemic phenomenon that emerges from the stochastic firing of ORIs at the single-cell level that, in the case of the genetically defined ORIs of *S. cerevisiae*, has been related to an increased number of Mcm2-7 complexes at early-replicating regions (Das et al., 2015).

While this study cannot establish a causal relationship between transcription elongation and ORI activity, the work described here is consistent with the view that both the replication initiation and the replication timing landscape of the *L. major* genome can be explained as arising from Pol II transcription dynamics, without the need to invoke specific regulatory mechanisms. If this hypothesis is correct, this simple solution would provide this organism with the required flexibility and robustness to deal with the environmental changes that impose alterations in the genetic programs that the parasite should tackle along its digenetic life cycle. Addressing whether this scenario is a curiosity of *Leishmania* or whether it represents an ancient solution used by eukaryotic cells for ORI maintenance constitutes an interesting open question.

## EXPERIMENTAL PROCEDURES

### Cell Culture

*L. major* strain Friedlin promastigote cells were grown at 26°C in M199 medium (Sigma-Aldrich) supplemented with 10% heat inactivated fetal bovine serum, 40 mM HEPES (pH 7.4), 0.1 mM adenine, 10 μg/ml hemin, 1 μg/ml biotin, 2 ng/ml biopterin, 100 U/ml penicillin G, and 0.1 mg/ml streptomycin sulfate.

### Mononucleosomal DNA Purification

For each nucleosome preparation, 100 ml cultures of *L. major* cells were harvested at an exponential phase of growth ($6 \times 10^6$ cells/ml) and lysed gently, and nuclei were purified through a sucrose cushion as previously described (Gong et al., 1996). Nuclei aliquots corresponding to $1.7 \times 10^7$ cells were digested with increasing concentrations of MNase over 6 min at 25°C, and DNA samples were electrophoretically separated. Mononucleosomal-sized fragments from samples containing a 80:20 mononucleosome-to-dinucleosome ratio were purified and subjected to standard library preparation (Illumina). Only paired reads with intermate distances of 150 ± 10 nt were retained for the analysis to ensure that reads derive from mononucleosome-protected DNA. For a complete description of the experimental procedures, see Supplemental Experimental Procedures.

### SNS Purification

For each SNS preparation, two independent 500 ml cultures of *L. major* cells were collected at an exponential phase of growth. Nuclei were obtained as described earlier and lysed gently, and nucleic acids were purified through careful phenol/chloroform extractions. SNS fractionation was performed by ultracentrifugation on SW40Ti rotors over 20 hr at 24,000 rpm and 20°C as previously described (Gómez and Antequera, 2008). Pooled fractions from individual gradients containing 300–900 and 300–1,500 nt SNS were subjected to three sequential rounds of T4 polynucleotide kinase (PNK) phosphorylation and λ-exonuclease digestion (custom-made, Thermo Scientific) to specifically enrich in replication intermediates. SNS were prepared for sequencing by random priming second-strand synthesis as previously described (Cadoret et al., 2008), followed by RNA-primer digestion with RNase A/T1 Mix (Thermo Scientific). DNA libraries were prepared with the NBNext kit (NE Biolabs) following the manufacturer's instructions, and library fragments of 400–1,500 bp were purified from polyacrylamide gels. For a complete description of the experimental procedures, see Supplemental Experimental Procedures.

### Genomic Annotations, Data Analysis, and Statistical Significances

*L. major* reference genome and genomic coordinates of CDSs were retrieved from TriTrypDB v.6.0 (http://www.tritrypdb.org). TSS and TTS coordinates were defined by intersecting Rastrojo et al. (2013) gene annotations with the coordinates of AcH3 (Thomas et al., 2009) and base J (van Luenen et al., 2012) datasets. MNase digestion coupled to next-generation sequencing (MN-seq) reads were aligned following the RUbioSeq pipeline (Rubio-Camarillo et al., 2013). Nucleosome occupancy maps were generated by applying NUCwave (Quintales et al., 2015). SNS-seq reads were aligned using Bowtie (Langmead et al., 2009), and peaks were obtained after applying the *callpeak* function from MACS2 (Zhang et al., 2008). G-quadruplex motif occurrences were predicted with QuadParser (Huppert and Balasubramanian, 2005). To account for the observed genomic distributions, number and sizes of ORI peaks for each chromosome were compared with the expected proportion calculated from genomic intervals randomly sampled from throughout the genome. The procedure was repeated 10,000 times, and statistical significances were determined by computing the empirical p value from the sampling distribution. For a complete description of the analysis procedures, see Supplemental Experimental Procedures.

### Fiber Stretching

Analysis of DNA replication by fiber stretching was adapted from Terret et al. (2009), except that exponentially growing cells were pulsed for 10 min with each nucleotide analog. For a complete description of the experimental procedures, see the Supplemental Experimental Procedures.

## ACCESSION NUMBERS

The accession number for the MN-seq and SNS-seq data reported in this paper is GEO: GSE81991.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.07.007.

## AUTHOR CONTRIBUTIONS

R.L., A.A., R.A., C.P.-C., and F.G. designed the experiments and carried out the experimental work. R.L., A.A., and J.M.F.-J. performed the computational analysis. A.C., J.M.R., and M.G. designed the experiments. M.G. supervised the general strategy of the work and wrote the article. All authors analyzed the data, discussed the results, and approved the final version of the manuscript.

## ACKNOWLEDGMENTS

We are grateful to Richard McCulloch for sharing data with us before publication and to Crisanto Gutierrez and Richard McCulloch for critical review of the manuscript. We also thank Ugo Bastolla for support with bioinformatics analysis, Sara Rodriguez-Acebes and Juan Méndez for advice on DNA fiber analysis, Diana Martín for technical support with *Leishmania* cell cultures, Pepe Belio for artwork, members of Crisanto Gutierrez and M.G. labs for discussions, and Fundación Parque Científico de Madrid (FPCM) for excellent library preparation. This work was supported by the Spanish Ministry of Economy and Competitiveness (BFU2013-45276-P to M.G. and SAF2013-47556-R to J.M.R.). R.A. is supported by a grant from the Portuguese Foundation for Science and Technology (SFRH/BD/81027/11).

Received: February 26, 2016
Revised: May 26, 2016
Accepted: July 1, 2016
Published: July 28, 2016

## REFERENCES

Alexander, R.D., Innocente, S.A., Barrass, J.D., and Beggs, J.D. (2010). Splicing-dependent RNA polymerase pausing in yeast. Mol. Cell *40*, 582–593.

Baldauf, S.L. (2003). The deep roots of eukaryotes. Science *300*, 1703–1706.

Bechhoefer, J., and Rhind, N. (2012). Replication timing and its emergence from stochastic processes. Trends Genet. *28*, 374–381.

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

Blow, J.J., Ge, X.Q., and Jackson, D.A. (2011). How dormant origins promote complete genome replication. Trends Biochem. Sci. *36*, 405–414.

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

Calderano, S.G., Drosopoulos, W.C., Quaresma, M.M., Marques, C.A., Kosiyatrakul, S., McCulloch, R., Schildkraut, C.L., and Elias, M.C. (2015). Single molecule analysis of *Trypanosoma brucei* DNA replication dynamics. Nucleic Acids Res. *43*, 2655–2665.

Carrillo Oesterreich, F., Preibisch, S., and Neugebauer, K.M. (2010). Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. Mol. Cell *40*, 571–581.

Cayrou, C., Coulombe, P., Vigneron, A., Stanojcic, S., Ganier, O., Peiffer, I., Rivals, E., Puy, A., Laurent-Chabalier, S., Desprat, R., and Méchali, M. (2011). Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. Genome Res. *21*, 1438–1449.

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

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

Comoglio, F., Schlumpf, T., Schmid, V., Rohs, R., Beisel, C., and Paro, R. (2015). High-resolution profiling of *Drosophila* replication start sites reveals a DNA shape and chromatin signature of metazoan origins. Cell Rep. *11*, 821–834.

Das, S.P., Borrman, T., Liu, V.W., Yang, S.C., Bechhoefer, J., and Rhind, N. (2015). Replication timing is regulated by the number of MCMs loaded at origins. Genome Res. *25*, 1886–1892.

De Gaudenzi, J.G., Noé, G., Campo, V.A., Frasch, A.C., and Cassola, A. (2011). Gene expression regulation in trypanosomatids. Essays Biochem. *51*, 31–46.

Deal, R.B., Henikoff, J.G., and Henikoff, S. (2010). Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. Science *328*, 1161–1164.

Delgado, S., Gómez, M., Bird, A., and Antequera, F. (1998). Initiation of DNA replication at CpG islands in mammalian chromosomes. EMBO J. *17*, 2426–2435.

Dellino, G.I., Cittaro, D., Piccioni, R., Luzi, L., Banfi, S., Segalla, S., Cesaroni, M., Mendoza-Maldonado, R., Giacca, M., and Pelicci, P.G. (2013). Genome-wide mapping of human DNA-replication origins: levels of transcription at ORC1 sites regulate origin selection and replication timing. Genome Res. *23*, 1–11.

Downing, T., Imamura, H., Decuypere, S., Clark, T.G., Coombs, G.H., Cotton, J.A., Hilley, J.D., de Doncker, S., Maes, I., Mottram, J.C., et al. (2011). Whole genome sequencing of multiple *Leishmania donovani* clinical isolates provides insights into population structure and mechanisms of drug resistance. Genome Res. *21*, 2143–2156.

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

Eaton, M.L., Prinz, J.A., MacAlpine, H.K., Tretyakov, G., Kharchenko, P.V., and MacAlpine, D.M. (2011). Chromatin signatures of the *Drosophila* replication program. Genome Res *21*, 164–174.

Fan, X., Moqtaderi, Z., Jin, Y., Zhang, Y., Liu, X.S., and Struhl, K. (2010). Nucleosome depletion at yeast terminators is not intrinsic and can occur by a transcriptional mechanism linked to 3′-end formation. Proc. Natl. Acad. Sci. USA *107*, 17945–17950.

Foulk, M.S., Urban, J.M., Casella, C., and Gerbi, S.A. (2015). Characterizing and controlling intrinsic biases of lambda exonuclease in nascent strand sequencing reveals phasing between nucleosomes and G-quadruplex motifs around a subset of human replication origins. Genome Res. *25*, 725–735.

Fragkos, M., Ganier, O., Coulombe, P., and Méchali, M. (2015). DNA replication origin activation in space and time. Nat. Rev. Mol. Cell Biol. *16*, 360–374.

Gilbert, D.M. (2010). Evaluating genome-scale approaches to eukaryotic DNA replication. Nat. Rev. Genet. *11*, 673–684.

Gindin, Y., Valenzuela, M.S., Aladjem, M.I., Meltzer, P.S., and Bilke, S. (2014). A chromatin structure-based model accurately predicts DNA replication timing in human cells. Mol. Syst. Biol. *10*, 722.

Gómez, M., and Antequera, F. (2008). Overreplication of short DNA regions during S phase in human cells. Genes Dev. *22*, 375–385.

Gommers-Ampt, J.H., Van Leeuwen, F., de Beer, A.L., Vliegenthart, J.F., Dizdaroglu, M., Kowalak, J.A., Crain, P.F., and Borst, P. (1993). beta-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan *T. brucei*. Cell *75*, 1129–1136.

Gong, Q.H., McDowell, J.C., and Dean, A. (1996). Essential role of NF-E2 in remodeling of chromatin structure and transcriptional activation of the epsilon-globin gene in vivo by 5′ hypersensitive site 2 of the beta-globin locus control region. Mol. Cell. Biol. *16*, 6055–6064.

Gros, J., Devbhandari, S., and Remus, D. (2014). Origin plasticity during budding yeast DNA replication in vitro. EMBO J. *33*, 621–636.

Gros, J., Kumar, C., Lynch, G., Yadav, T., Whitehouse, I., and Remus, D. (2015). Post-licensing specification of eukaryotic replication origins by facilitated Mcm2-7 sliding along DNA. Mol. Cell *60*, 797–807.

Huppert, J.L., and Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. Nucleic Acids Res. *33*, 2908–2916.

Ivens, A.C., Peacock, C.S., Worthey, E.A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M.A., Adlem, E., Aert, R., et al. (2005). The genome of the kinetoplastid parasite, *Leishmania major*. Science *309*, 436–442.

Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. Nat. Rev. Genet. *10*, 161–172.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J., and Segal, E. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362–366.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

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

Lombraña, R., Almeida, R., Álvarez, A., and Gómez, M. (2015). R-loops and initiation of DNA replication in human cells: a missing link? Front. Genet. *6*, 158.

MacAlpine, H.K., Gordân, R., Powell, S.K., Hartemink, A.J., and MacAlpine, D.M. (2010). *Drosophila* ORC localizes to open chromatin and marks sites of cohesin complex loading. Genome Res. *20*, 201–211.

Marques, C.A., Dickens, N.J., Paape, D., Campbell, S.J., and McCulloch, R. (2015). Genome-wide mapping reveals single-origin chromosome replication in *Leishmania*, a eukaryotic microbe. Genome Biol. *16*, 230.

Martin, M.M., Ryan, M., Kim, R., Zakas, A.L., Fu, H., Lin, C.M., Reinhold, W.C., Davis, S.R., Bilke, S., Liu, H., et al. (2011). Genome-wide depletion of replication initiation events in highly transcribed regions. Genome Res. *21*, 1822–1832.

Martínez-Calvillo, S., Yan, S., Nguyen, D., Fox, M., Stuart, K., and Myler, P.J. (2003). Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region. Mol. Cell *11*, 1291–1299.

Martínez-Calvillo, S., Nguyen, D., Stuart, K., and Myler, P.J. (2004). Transcription initiation and termination on *Leishmania major* chromosome 3. Eukaryot. Cell *3*, 506–517.

Méchali, M. (2010). Eukaryotic DNA replication origins: many choices for appropriate answers. Nat. Rev. Mol. Cell Biol. *11*, 728–738.

Mesner, L.D., Valsakumar, V., Karnani, N., Dutta, A., Hamlin, J.L., and Bekiranov, S. (2011). Bubble-chip analysis of human origin distributions demonstrates on a genomic scale significant clustering into zones and significant association with transcription. Genome Res. *21*, 377–389.

Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. Cell *136*, 688–700.

Patnaik, P.K. (1997). Studies with artificial extrachromosomal elements in trypanosomatids: could specificity in the initiation of DNA replication be linked to that in transcription? Parasitol. Today (Regul. Ed.) *13*, 468–471.

Patrick, E., Buckley, M., and Yang, Y.H. (2013). Estimation of data-specific constitutive exons with RNA-seq data. BMC Bioinformatics *14*, 31.

Perales, R., and Bentley, D. (2009). "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. Mol. Cell *36*, 178–191.

Picard, F., Cadoret, J.C., Audit, B., Arneodo, A., Alberti, A., Battail, C., Duret, L., and Prioleau, M.N. (2014). The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells. PLoS Genet. *10*, e1004282.

Powell, S.K., MacAlpine, H.K., Prinz, J.A., Li, Y., Belsky, J.A., and MacAlpine, D.M. (2015). Dynamic loading and redistribution of the Mcm2-7 helicase complex through the cell cycle. EMBO J. *34*, 531–543.

Quintales, L., Vázquez, E., and Antequera, F. (2015). Comparative analysis of methods for genome-wide nucleosome cartography. Brief. Bioinform. *16*, 576–587.

Rando, O.J., and Winston, F. (2012). Chromatin and transcription in yeast. Genetics *190*, 351–387.

Rastrojo, A., Carrasco-Ramiro, F., Martín, D., Crespillo, A., Reguera, R.M., Aguado, B., and Requena, J.M. (2013). The transcriptome of *Leishmania major* in the axenic promastigote stage: transcript annotation and relative expression levels by RNA-seq. BMC Genomics *14*, 223.

Requena, J.M. (2011). Lights and shadows on gene organization and regulation of gene expression in *Leishmania*. Front. Biosci. (Landmark Ed.) *16*, 2069–2085.

Reynolds, D., Cliffe, L., Förstner, K.U., Hon, C.C., Siegel, T.N., and Sabatini, R. (2014). Regulation of transcription termination by glucosylated hydroxymethyluracil, base J, in *Leishmania major* and *Trypanosoma brucei*. Nucleic Acids Res. *42*, 9717–9729.

Rhind, N., and Gilbert, D.M. (2013). DNA replication timing. Cold Spring Harb. Perspect. Biol. *5*, a010132.

Rubio-Camarillo, M., Gómez-López, G., Fernández, J.M., Valencia, A., and Pisano, D.G. (2013). RUbioSeq: a suite of parallelized pipelines to automate exome variation and bisulfite-seq analyses. Bioinformatics *29*, 1687–1689.

Sequeira-Mendes, J., and Gómez, M. (2012). On the opportunistic nature of transcription and replication initiation in the metazoan genome. BioEssays *34*, 119–125.

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

Spies, N., Nielsen, C.B., Padgett, R.A., and Burge, C.B. (2009). Biased chromatin signatures around polyadenylation sites and exons. Mol. Cell *36*, 245–254.

Terret, M.E., Sherwood, R., Rahman, S., Qin, J., and Jallepalli, P.V. (2009). Cohesin acetylation speeds the replication fork. Nature *462*, 231–234.

Thomas, S., Green, A., Sturm, N.R., Campbell, D.A., and Myler, P.J. (2009). Histone acetylations mark origins of polycistronic transcription in *Leishmania major*. BMC Genomics *10*, 152.

Tiengwe, C., Marcello, L., Farr, H., Dickens, N., Kelly, S., Swiderski, M., Vaughan, D., Gull, K., Barry, J.D., Bell, S.D., and McCulloch, R. (2012). Genome-wide analysis reveals extensive functional interaction between DNA replication initiation and transcription in the genome of *Trypanosoma brucei*. Cell Rep. *2*, 185–197.

Tillo, D., and Hughes, T.R. (2009). G+C content dominates intrinsic nucleosome occupancy. BMC Bioinformatics *10*, 442.

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

van Luenen, H.G., Farris, C., Jan, S., Genest, P.A., Tripathi, P., Velds, A., Kerkhoven, R.M., Nieuwland, M., Haydock, A., Ramasamy, G., et al. (2012). Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in *Leishmania*. Cell *150*, 909–921.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-seq (MACS). Genome Biol. *9*, R137.