ON THE OPPORTUNISTIC NATURE OF TRANSCRIPTION AND REPLICATION INITIATION IN THE METAZOAN GENOME

Joana Sequeira-Mendes and María Gómez*

Centro de Biología Molecular. Consejo Superior de Investigaciones Científicas (CSIC/UAM). Nicolás Cabrera
1. 28049-Madrid, Spain

Instituto de Biología Funcional y Genómica (CSIC/USAL). Edificio Departamental. Campus Miguel de
Unamuno. 37007-Salamanca, Spain

* Corresponding author (mgomez@cbm.uam.es)

Abstract

Cellular identity and its response to external or internal signalling variations are encoded in a cells genome as regulatory information. The genomic regions that specify this type of information are highly variable and degenerated in its sequence determinants, as is becoming increasingly evident through the application of genome-scale methods to study gene expression. Here, we speculate that the same scenario applies to the regulatory regions controlling where DNA replication starts in the metazoan genome. We propose that replication origins cannot be defined as unique genomic features, but rather that DNA synthesis initiates opportunistically from accessible DNA sites, making cells highly robust and adaptable to environmental or developmental changes.

Introduction

The availability of the complete genomic sequence of several model organisms, and the development of high-density genomic microarrays and deep-sequencing techniques has provided a full landscape of the structural information contained in DNA. The regulatory information contained in the genomes, however, is more difficult to decode as its sequence determinants are highly degenerated and fine-tuned by epigenetic factors. A paradigm of this complexity is our current vision of gene regulation. The recent development of high throughput technologies for studying gene expression has led to the idea that genomes are complex systems: pervasive transcription, regulatory non-coding RNAs that modulate the action of transcription factors; engaged RNA polymerase II (RNAPolII) associated with non-expressed genes to produce short non-coding transcripts at the promoter; differentially marked genes poised for activation upon development [1]. To further complicate the scenario, recent in-depth analysis of chromatin types has unveiled a new landscape with uncharacterised transcriptionally active and inactive chromatin domains which are organised non-randomly in the nucleus and may influence the expression of the residing genes [2]. Even the assumption that chromatin modifications are heritable epigenetic marks has been challenged by recent data from S. Henikoff's group. By directly measuring on genome-wide scale the kinetics of nucleosome exchange, the authors proposed that the active or silent gene state may be perpetuated by regulated nucleosome turnover, rather than by the histone modifications themselves [3].

In this essay we primarily focus on animal genomes to explore how this genomic complexity, whose role in the regulation of gene expression we are just beginning to comprehend, applies to the regulation of replication initiation. We propose that metazoan DNA replication origins do not exist as discrete genetic entities. Instead, we hypothesise that replication initiation starts opportunistically from differently exposed DNA sites, rendering cells highly robust and rapidly adaptable to environmental or developmental changes.

Transcription regulation is highly degenerated

Transcriptome analyses have shown that much of the eukaryotic genome is transcribed, and that transcripts from both strands of specific genomic loci are detectable in a diverse range of organisms [4-9]. Some of these transcripts map to known transcription units, but many derive

from unannotated regions of the chromosomes, what is consistent with the presence of RNAPolII bound to those non-genic regions. Pervasive transcription not only originates from canonical promoter regions; in both yeast and mammalian cells a number of non-coding RNAs (ncRNA) seem to start at the 3' ends of genes [7,10]. Exonic cryptic transcripts have also been identified in *Saccharomyces cerevisiae*, in addition to telomeric transcripts and ncRNA transcripts derived from repetitive regions of many organisms [11-14]. The functionality of this pervasive transcription, particularly the ncRNAs derived from outside of protein-coding genes, is subject of considerable debate [15-17]. Despite the relative proportion between biologically significant ncRNAs and transcriptional noise, several ncRNA transcripts are known to interact with proteins, are differentially expressed, and conserved across species. Some well-characterised examples include mouse *Xist*, *Air* or human *HOTAIR* ncRNAs [18-20]. These findings suggest that at least a subset of ncRNAs can be active players in the regulation of genomic plasticity.

Promoters themselves are diffuse genetic entities. Transcription initiation can start at either localised sites on TATA-box regulated genes or broadly dispersed transcription start sites (TSS) spanning across CpG island (CGI) promoters in mammalian cells. These diverse initiation patterns are strongly correlated with differences in nucleosome organisation and chromatin structure [21]. For most of the promoters with dispersed TSS, transcription initiates in both directions, consistent with the finding of two peaks of the activated form of the RNAPolII. This phenomenon has also been reported to occur in the yeast genome [6,10,22-24]. However, the role of widespread divergent transcription at eukaryotic promoters remains unclear. It has been hypothesised that divergent transcription could function by shifting away nucleosomes at promoter regions, thus facilitating the access of transcription factors [6,10,22-24]. Other authors suggest that it would provide a quick available pool of active RNAPolII molecules for protein-coding mRNA expression [25], or just represent opportunistic RNAPolII-mediated transcription within TSS [26].

On top of the diversity in promoter architecture that governs differences in the initiation process, transcription can also be regulated at several levels in eukaryotes. One of them occurs during the multiple steps of the RNAPolII transcription cycle [27]. This includes: (i) the formation of the pre-initiation complex at the promoter, mediated by the recruitment of general transcription factors; (ii) the initiation of transcription; (iii) the clearance of the RNAPolII from the promoter-bound factors; (iv) the pausing of the RNAPolII after transcribing 20-50 nucleotides and; (v) the escape from the paused state into productive

elongation. Recent findings indicate that progression of a promoter-proximal, paused RNAPolII, into productive elongation is a rate limiting step in the transcription of nearly 40% of genes in mouse embryonic stem (ES) cells and mouse embryonic fibroblasts [28]. Interestingly, key pluripotency regulatory genes such as *Oct4*, *Sox2* and *Nanog* exhibit a regulated rate of escape from pausing, suggesting that RNAPolII pausing may provide a responsive transcriptional regulation control during cell differentiation. At each promoter, a particular combination of transcription factors, elongation factors, nucleosomes, and underlying DNA sequence act together to determine the kinetics of the RNAPolII capture and release, orchestrating the regulation of DNA transcription by this enzyme. Most likely, the degenerate and combinatorial nature of this regulatory information enhances the robustness of biological systems. Degeneracy, at the same time, might allow for faster adaptation to signalling variations, an essential feature to achieve fitness and evolutionary success.

Chromatin accessibility is a hallmark of replication origins

DNA replication is another basic genomic process and, as such, thought to be highly regulated in time and space to ensure a complete and accurate duplication of the genetic information prior to cell division. In the recent years, high throughput technologies have also been applied genome-wide to address the temporal domains of DNA replication and the location of the replication initiation sites (ORIs) in a variety of organisms. The lack of identifiable sequences combined with the preference of the Origin Recognition Complex (ORC) for nucleosome-free chromatin suggests that DNA accessibility is a major determinant of the replication program.

Sequence requirements for origin specification decrease from yeast to metazoans

The first eukaryotic model organism to be studied by these approaches was *S. cerevisiae*, aided by the well-established genetic and biochemical knowledge of the sequence requirements of the replication initiation sites in its genome [29,30]. The analysis of the replication initiation landscape in *S. cerevisiae* demonstrated that, while the primary sequence of the DNA is a strong determinant of origin selection, it is not the unique driving force. Indeed, of the thousands of consensus sites that exist in its genome, only around 400 are bound specifically by ORC and function as *bona fide* origins of DNA replication [31,32].

Although ORC binding to DNA in *S. cerevisiae* requires a consensus 11-bp A/T-rich

sequence, each ORI contains a variable number of auxiliary elements that individually contribute to replication activity, constituting a first degree of variability between ORIs within the same genome [33]. This diversity between replicator elements is more extreme in the other yeast model, *Schizosaccharomyces pombe*. In this organism, replication initiation sites consist of asymmetric stretches of A/T-rich DNA, which vary in length and composition and do not share any apparent common consensus [34-37]. Most likely, these degenerated motifs are decoded by the *SpOrc4* subunit that shows a high preference for A/T-rich DNA [38]. ORC binding in both budding and fission yeasts is biased towards A/T-rich sequences and these elements are known to destabilise nucleosome assembly. Low occupancy of nucleosomes around an ORI may be important for efficient replication initiation, since positioning of nucleosomes over the consensus region greatly diminishes replication activity [39]. Indeed, genome-wide ORC localisation in *S. cerevisiae* is highly coincident with nucleosome-depleted regions [40,41] and, in *S. pombe*, the efficiency of an ORI is negatively correlated with nucleosome occupancy [42]. Similarly, recent genome-wide analysis of ORC binding in *Drosophila* identified around 5000 sites that contain no common sequence motifs [43]. ORC sites tend to localise to open chromatin regions that are enriched in the histone variant H3.3 and depleted of bulk nucleosomes, such as the promoters of actively transcribed genes and sites of cohesin loading. The association of replication ORIs with active transcription and regions enriched in the histone variant H2A.Z and other activating marks have also been described genome-wide in *Arabidopsis* [44]. Interestingly, in *Drosophila*, the density of bound ORC at a given ORI correlates well with its time of replication [43]. As early replicating domains tend to be enriched in euchromatin and actively transcribed genes, this observation probably reflects the enhanced accessibility of these regions. Supporting this idea, the majority of ORC binding sites correspond to regions with a high nucleosome turnover rate [3], many of which are associated with a subset of RNAPII binding sites [45]. Thus, the intrinsic positions of bulk nucleosomes encoded in the underlying DNA sequence seem to be the principle determinant of ORI sites in these systems, probably by restricting ORC association with the DNA. This is consistent with previous reports demonstrating that metazoan ORC shows no specificity or higher affinity for ORI sequences *in vitro*. However, negative supercoiling, such as that introduced *in vivo* by the removal of a nucleosome, can increase ORC binding affinity for DNA by several-fold [46-48].

DNA accessibility is a major determinant of the replication program in metazoans

In multicellular eukaryotes, high-throughput technologies were applied to cartograph the temporal domains of DNA replication long before the first large-scale maps of ORI location

were produced. These studies showed a clear correlation between transcriptionally active domains and early replication timing in the *Drosophila* genome [45,49,50], a scenario that also occurs in both the mammalian [51-53] and the *Arabidopsis* genome [54]. These results indicate that permissive chromatin environments promote both transcription and replication. Interestingly, it was recently reported that replication timing domains are strongly correlated with regions of long-range chromatin interactions [55]. These observations suggest that chromatin in close spatial proximity replicates at a similar time, reinforcing the idea that both the temporal and the spatial regulatory information of DNA replication depend on chromatin accessibility to replication factors.

In recent years, several reports have addressed the large-scale localisation of ORIs by microarray hybridisation or deep-sequencing of purified short DNA nascent strands in various cell types of human and mouse [56-60]. In spite of the diversity of the genomic programs represented by these studies, they collectively found a clear association of ORIs with annotated transcriptional units, specifically gene promoters enriched in activating epigenetic marks. A similar conclusion was reached on a recent report employing a very different technology based on the selective entrapment of bubble-containing fragments in gelling agarose and their subsequent hybridisation on genomic microarrays [61]. However, these studies also revealed a lower-than-expected concordance between the various ORI datasets. These observations likely reflect the different levels of saturation achieved in the preparation of ORI libraries, and thus, the intrinsic diversity of the cohort of replication initiation sites activated in each cell. A well-studied example that illustrates the cell-specific plasticity in the replication program is the mouse *Igh* locus. Differential activation of DNA replication origins occurs within the locus during B-cell development, parallel with the developmentally regulated changes in transcriptional activity and chromatin structure [62]. Similarly, normally silent or dormant origins can be activated in response to genotoxic stresses, as demonstrated in the case of hidroxyurea treatment of Chinese hamster cells [63,64].

Therefore, mammalian ORIs seem to represent an extreme case of degeneracy in their coding information as the only common feature shared by the identified replication initiation sites is that a large majority of them lie on an accessible chromatin environment. In fact, the ORIs that show a higher probability of being activated in a cell population (*i.e.*, the most efficient ones) are those located close to the TSS of CGI promoters, landmarks of permissive chromatin state [59]. Intriguingly, the nucleosome-free region surrounding the TSS of active promoters undergoes several rounds of abortive replication cycles in a single S phase [65,66].

This phenomenon resembles the production of short bidirectional transcripts occurring at CGIs and may represent the intrinsic open structure of the chromatin at these regulatory sites. An interpretation of all these findings could be that ORIs do not exist as unique, well-defined regulatory units. Instead, DNA replication may start from a continuous gradient of initiation sites throughout the genome with preferential locations reflecting a higher probability of replication factor binding (Figure 1). Consistent with this hypothesis, a recent report indicates that ORC localises at nucleosome-depleted sites within the replication initiation zone at the *DHFR* locus in Chinese hamster cells [67]. The above scenario can also accommodate the earlier observations that mammalian replicators are able to initiate DNA synthesis at ectopic sites depending on the chromosomal context (reviewed in [68]), and raises the question of whether ORIs are truly genomic features or simply represent opportunistic ORC recruitment within accessible regions of the genome.

Opportunistic initiation of DNA replication?

DNA sequence is a potent driver of nucleosome organisation at certain regions, but the majority of the metazoan genome seems to exhibit very little nucleosome positioning [69]. Nucleosome positioning preferences at regulatory sites might, *per se*, facilitate transcription initiation or help direct transcription factors towards the functional subset of their target sequences [70]. Interestingly, the classically-referred nucleosome-free gaps are in fact regulatory regions occupied by inherently unstable nucleosomes containing the H2A.Z and H3.3 histone variants [71] and probably have high nucleosome turnover rates [3]. The instability of the double-variant nucleosomes most likely maintains regions such as promoters and enhancers temporarily free of nucleosomes and accessible to regulatory elements. It is possible that ORC could take advantage of this permissiveness and opportunistically colonise these sites. In support of this, the modENCODE consortium in *Drosophila* showed that, of the dozens of chromatin features studied, ORC occupancy mainly correlates with sites that are characterised by nucleosome depletion, a high turnover of H3.3, and an enrichment of chromatin remodelers of the NURF complex [40].

Based on the findings of the diverse chromatin architectures found at functional ORIs in *S. cerevisiae*, it has been proposed that, upon binding by ORC and recruitment of chromatin remodelling and modification activities, the flanking nucleosomes are precisely positioned [40,72]. These positioned nucleosomes then become important for the assembly of the pre-

replication complex (pre-RC) prior to ORI firing. Interestingly, the *S. cerevisiae* Orc1 bromo-adjacent homology (BAH) domain is particularly necessary to stabilise ORC binding at a subset of ORIs that are characterised by smaller-than-average nucleosome free region [73]. As human ORC binding to chromosomes is facilitated by the BAH domain [74], one possible mechanism for ORC stabilisation in the metazoan genome could be mediated through a direct interplay between nucleosomes and ORC. In addition, several transcription factors have been implicated in interacting with ORC and components of the pre-RC. Whether transcription factors act redundantly to recruit ORC to DNA or, as part of larger complexes with activating or repressive properties, may alter the local chromatin environment and regulate downstream replication initiation events is not yet clear. Recent reports on mammalian systems have shown that the local chromatin environment at ORIs can affect pre-RC assembly [75,76]. This suggests that the recruitment of ORC might not itself be sufficient for pre-RC formation and, instead, that the local chromatin environment may be critical for replication initiation activity. Strengthening this idea, a recent report in human cells indicates that the chromatin remodeler SNF2H is loaded at DNA replication origins in a Cdt1-dependent manner [77]. Binding of SNF2H to origin DNA specifically during G1 phase promotes pre-RC loading, suggesting a direct role of chromatin remodelers in the regulation of DNA replication activity and in the maintenance of genome integrity.

ORIs as specific, fixed genomic sites are not strictly required, as has been noted previously [78-80]. On one hand, there are biological systems in which replication can initiate from random locations, such as *Xenopus* early embryos [81]. On the other, the temporal separation between pre-RC assembly and ORI firing into discrete phases of the cell cycle provides an exquisite solution that ensures the lack of re-replication that could be deleterious, without the need of specific start sites. Decreasing the sequence requirements of replication initiation and placing the selective pressure on the final readout of the process (the complete and accurate duplication of the genome before cell division) provides metazoan genomes the necessary flexibility and adaptation to variation in signalling information needed to respond to either environmental changes or determinants of cell identity. This opportunistic scenario relieves the genome from the constraints of maintaining selective pressure on individual ORIs. It also fits in well with the myriad of regulatory mechanisms that redundantly control DNA replication to precisely ensure the firing of ORIs only once per cell cycle (reviewed in [82]). In addition, the coupling between nucleosome dynamics and replication initiation guarantees an excess of ORIs for completion of genome duplication during S phase, while simultaneously allowing cells to exploit this flexibility to achieve a tailored, fine-tuned

regulation of both replication and transcription. The opportunistic nature of ORIs, however, might jeopardise the integrity of the genomes at certain specific regions that are naturally refractory to ORC or pre-RC recruitment. This has been exemplified in two recent studies that addressed common fragile site formation in human lymphocytes [83,84]. Both papers suggest that the sensitivity of fragile sites to replication stress is the combination of late replication and paucity of initiation events required for completion of DNA replication under stress conditions.

Conclusions

The genome can be understood as an ecosystem where the selective pressure on any basic process (only transcription and replication were discussed here) mainly acts on its final outcome. This biological environment may lead to the opportunism or even symbiosis of components of the molecular machineries that compete between themselves, and with the nucleosomes, for access to the DNA. This view can help explain the flexibility and variability of the regulatory information encoded in the metazoan genome. The intrinsic degeneracy shared between these systems could, in turn, generate variation upon which natural selection acts, contributing to rapid adaptation to changing conditions and enhancing robustness. Although disconcerting, the genomic complexity found through the application of large-scale approaches reflects the power of these techniques to unveil the enormous variety of ways that cells use to decode their genomes and opens up exciting avenues of future research.

Acknowledgments

We thank Francisco Antequera for several years of stimulating conversations that have generated many of the ideas contained in this essay. We are also grateful to Francisco Antequera, Anwyn Chapman, Mauro Lodolo and Crisanto Gutierrez for critical review of the manuscript, José Belio for artwork, and members of MG laboratory for discussions. Research at MG lab is funded by the Spanish Ministry of Sciences and Innovation (BFU2010-18992). JSM is a recipient of a postdoctoral Juan de La Cierva (Spanish Ministry of Sciences and Innovation) contract.

Keywords

DNA replication origins; transcription; opportunism; CpG islands; nucleosomes

Abbreviations

RNApolII, RNA polymerase II

TSS, transcription start sites

CGI, CpG island

ORI, origin of DNA replication initiation

ORC, origin recognition complex

Pre-RC, pre-replication complex

FIGURE LEGEND

Figure 1. Continuous gradient of DNA replication initiation sites in the mammalian genome. The same portion of the genome containing two annotated transcriptional units and several ORIs is represented in six different cells. The blue and red boxes indicate the exons of the genes codified in the upper or lower DNA strand, respectively, and the same color code is used to illustrate RNApolII transcribing either DNA strand. The long and short tips protruding from the RNApolII represent elongating mRNAs and short or non-coding RNAs, respectively. In this example, regions of constitutively low nucleosome occupancy, such as CpG islands, are landing pads for transcription factors and RNApolII and can recruit pre-RC complexes in most cells, resulting in higher measurements of ORI activity in the cell population (*i.e.*, higher relative efficiency). In addition to these high efficient ORIs, DNA synthesis can initiate from different genomic locations, which firing efficiencies will reflect the distinct accessibility of the replication factors to the DNA in each cell. In some cases, like the one represented between both transcriptional units, pre-RC recruitment could be enhanced due to the lower nucleosome occupancy at the 3' end of the genes associated with RNApolII antisense transcription. Genes, CpG islands, and protein complexes are not drawn to scale.

References

1. **Mellor J.** 2010. Transcription: from regulatory ncRNA to incongruent redundancy. *Genes Dev* **24**:1449-55.
2. **Filion GJ, van Bemmelen JG, Braunschweig U, Talhout W, Kind J, et al.** 2010. Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* **143**:212-24.
3. **Deal RB, Henikoff JG, Henikoff S.** 2010. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* **328**:1161-4.
4. **Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, et al.** 2005. The transcriptional landscape of the mammalian genome. *Science* **309**:1559-63.
5. **David L, Huber W, Granovskaia M, Toedling J, Palm CJ, et al.** 2006. A high-resolution map of transcription in the yeast genome. *Proc Natl Acad Sci U S A* **103**:5320-5.
6. **He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW.** 2008. The antisense transcriptomes of human cells. *Science* **322**:1855-7.
7. **Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, et al.** 2007. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**:1484-8.
8. **Li L, Wang X, Stolc V, Li X, Zhang D, et al.** 2006. Genome-wide transcription analyses in rice using tiling microarrays. *Nat Genet* **38**:124-9.
9. **Stolc V, Gauhar Z, Mason C, Halasz G, van Batenburg MF, et al.** 2004. A gene expression map for the euchromatic genome of *Drosophila melanogaster*. *Science* **306**:655-60.
10. **Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Munster S, et al.** 2009. Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**:1033-7.
11. **Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ.** 2007. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316**:744-7.
12. **Azzalin CM, Reichenbach P, Khoriatuli L, Giulotto E, Lingner J.** 2007. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* **318**:798-801.
13. **Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, et al.** 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128**:1089-103.
14. **Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, et al.** 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**:581-92.
15. **Berretta J, Morillon A.** 2009. Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO Rep* **10**:973-82.
16. **Ponting CP, Belgard TG.** 2010. Transcribed dark matter: meaning or myth? *Hum Mol Genet* **19**:R162-8.
17. **Struhl K.** 2007. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol* **14**:103-5.
18. **Chow J, Heard E.** 2009. X inactivation and the complexities of silencing a sex chromosome. *Curr Opin Cell Biol* **21**:359-66.
19. **Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, et al.** 2008. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**:1717-20.

20. **Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, et al.** 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**:1311-23.
21. **Rach EA, Winter DR, Benjamin AM, Corcoran DL, Ni T, et al.** 2011. Transcription initiation patterns indicate divergent strategies for gene regulation at the chromatin level. *PLoS Genet* **7**:e1001274.
22. **Core LJ, Waterfall JJ, Lis JT.** 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**:1845-8.
23. **Neil H, Malabat C, d'Aubenton-Carafa Y, Xu Z, Steinmetz LM, et al.** 2009. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* **457**:1038-42.
24. **Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, et al.** 2008. Divergent transcription from active promoters. *Science* **322**:1849-51.
25. **Preker R, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, et al.** 2008. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**:1851-4.
26. **Layer JH, Weil PA.** 2009. Ubiquitous antisense transcription in eukaryotes: novel regulatory mechanism or byproduct of opportunistic RNA polymerase? *F1000 Biol Rep* **1**:33.
27. **Fuda NJ, Ardehali MB, Lis JT.** 2009. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* **461**:186-92.
28. **Min IM, Waterfall JJ, Core LJ, Munroe RJ, Schimenti J, et al.** 2011. Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes Dev* **25**:742-54.
29. **Bell SP, Stillman B.** 1992. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* **357**:128-34.
30. **Newlon CS, Collins I, Dershowitz A, Deshpande AM, Greenfeder SA, et al.** 1993. Analysis of replication origin function on chromosome III of *Saccharomyces cerevisiae*. *Cold Spring Harb Symp Quant Biol* **58**:415-23.
31. **Feng W, Collingwood D, Boeck ME, Fox LA, Alvino GM, et al.** 2006. Genomic mapping of single-stranded DNA in hydroxyurea-challenged yeasts identifies origins of replication. *Nat Cell Biol* **8**:148-55.
32. **Xu W, Aparicio JG, Aparicio OM, Tavaré S.** 2006. Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in *S. cerevisiae*. *BMC Genomics* **7**:276.
33. **Rao H, Marahrens Y, Stillman B.** 1994. Functional conservation of multiple elements in yeast chromosomal replicators. *Mol Cell Biol* **14**:7643-51.
34. **Cotobal C, Segurado M, Antequera F.** 2010. Structural diversity and dynamics of genomic replication origins in *Schizosaccharomyces pombe*. *Embo J* **29**:934-42.
35. **Hayashi M, Katou Y, Itoh T, Tazumi A, Yamada Y, et al.** 2007. Genome-wide localization of pre-RC sites and identification of replication origins in fission yeast. *Embo J* **26**:1327-39.
36. **Heichinger C, Penkett CJ, Bahler J, Nurse P.** 2006. Genome-wide characterization of fission yeast DNA replication origins. *Embo J* **25**:5171-9.
37. **Segurado M, de Luis A, Antequera F.** 2003. Genome-wide distribution of DNA replication origins at A+T-rich islands in *Schizosaccharomyces pombe*. *EMBO Rep* **4**:1048-53.
38. **Chuang RY, Kelly TJ.** 1999. The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks. *Proc Natl Acad Sci U S A* **96**:2656-61.
39. **Lipford JR, Bell SP.** 2001. Nucleosomes positioned by ORC facilitate the initiation of DNA replication. *Mol Cell* **7**:21-30.

40. **Eaton ML, Galani K, Kang S, Bell SP, MacAlpine DM.** 2010. Conserved nucleosome positioning defines replication origins. *Genes Dev* **24**:748-53.
41. **Field Y, Kaplan N, Fondufe-Mittendorf Y, Moore IK, Sharon E, et al.** 2008. Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput Biol* **4**:e1000216.
42. **Lantermann AB, Straub T, Stralfors A, Yuan GC, Ekwall K, et al.** 2010. Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. *Nat Struct Mol Biol* **17**:251-7.
43. **MacAlpine HK, Gordan R, Powell SK, Hartemink AJ, MacAlpine DM.** 2010. Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading. *Genome Res* **20**:201-11.
44. **Costas C, de la Paz Sanchez M, Stroud H, Yu Y, Oliveros JC, et al.** 2010. Genome-wide mapping of Arabidopsis thaliana origins of DNA replication and their associated epigenetic marks. *Nat Struct Mol Biol* **18**:395-400.
45. **MacAlpine DM, Rodriguez HK, Bell SP.** 2004. Coordination of replication and transcription along a Drosophila chromosome. *Genes Dev* **18**:3094-105.
46. **Remus D, Beall EL, Botchan MR.** 2004. DNA topology, not DNA sequence, is a critical determinant for Drosophila ORC-DNA binding. *Embo J* **23**:897-907.
47. **Schaarschmidt D, Baltin J, Stehle IM, Lipps HJ, Knippers R.** 2004. An episomal mammalian replicon: sequence-independent binding of the origin recognition complex. *Embo J* **23**:191-201.
48. **Vashee S, Cvetic C, Lu W, Simancek P, Kelly TJ, et al.** 2003. Sequence-independent DNA binding and replication initiation by the human origin recognition complex. *Genes Dev* **17**:1894-908.
49. **Schubeler D, Scalzo D, Kooperberg C, van Steensel B, Delrow J, et al.** 2002. Genome-wide DNA replication profile for Drosophila melanogaster: a link between transcription and replication timing. *Nat Genet* **32**:438-42.
50. **Schwaiger M, Stadler MB, Bell O, Kohler H, Oakeley EJ, et al.** 2009. Chromatin state marks cell-type- and gender-specific replication of the Drosophila genome. *Genes Dev* **23**:589-601.
51. **Farkash-Amar S, Lipson D, Polten A, Goren A, Helmstetter C, et al.** 2008. Global organization of replication time zones of the mouse genome. *Genome Res* **18**:1562-70.
52. **Hiratani I, Ryba T, Itoh M, Rathjen J, Kulik M, et al.** 2010. Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. *Genome Res* **20**:155-69.
53. **Hiratani I, Ryba T, Itoh M, Yokochi T, Schwaiger M, et al.** 2008. Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol* **6**:e245.
54. **Lee TJ, Pascuzzi PE, Settlege SB, Shultz RW, Tanurdzic M, et al.** 2010. Arabidopsis thaliana chromosome 4 replicates in two phases that correlate with chromatin state. *PLoS Genet* **6**:e1000982.
55. **Ryba T, Hiratani I, Lu J, Itoh M, Kulik M, et al.** 2010. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res* **20**:761-70.
56. **Cadore JC, Meisch F, Hassan-Zadeh V, Luyten I, Guillet C, et al.** 2008. Genome-wide studies highlight indirect links between human replication origins and gene regulation. *Proc Natl Acad Sci U S A* **105**:15837-42.
57. **Cayrou C, Coulombe P, Vigneron A, Stanojcic S, Ganier O, et al.** 2011. Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res.*

58. **Karnani N, Taylor CM, Malhotra A, Dutta A.** 2010. Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. *Mol Biol Cell* **21**:393-404.
59. **Sequeira-Mendes J, Diaz-Uriarte R, Apedaile A, Huntley D, Brockdorff N, et al.** 2009. Transcription initiation activity sets replication origin efficiency in mammalian cells. *PLoS Genet* **5**:e1000446.
60. **Valenzuela MS, Chen Y, Davis S, Yang F, Walker RL, et al.** 2011. Preferential localization of human origins of DNA replication at the 5'-ends of expressed genes and at evolutionarily conserved DNA sequences. *PLoS One* **6**:e17308.
61. **Mesner LD, Valsakumar V, Karnani N, Dutta A, Hamlin JL, et al.** 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**:377-89.
62. **Norio P, Kosiyatrakul S, Yang Q, Guan Z, Brown NM, et al.** 2005. Progressive activation of DNA replication initiation in large domains of the immunoglobulin heavy chain locus during B cell development. *Mol Cell* **20**:575-87.
63. **Anglana M, Apiou F, Bensimon A, Debatisse M.** 2003. Dynamics of DNA replication in mammalian somatic cells: nucleotide pool modulates origin choice and interorigin spacing. *Cell* **114**:385-94.
64. **Courbet S, Gay S, Arnoult N, Wronka G, Anglana M, et al.** 2008. Replication fork movement sets chromatin loop size and origin choice in mammalian cells. *Nature* **455**:557-60.
65. **Gomez M.** 2008. Controlled rereplication at DNA replication origins. *Cell Cycle* **7**:1313-4.
66. **Gomez M, Antequera F.** 2008. Overreplication of short DNA regions during S phase in human cells. *Genes Dev* **22**:375-85.
67. **Lubelsky Y, Sasaki T, Kuipers MA, Lucas I, Le Beau MM, et al.** 2011. Pre-replication complex proteins assemble at regions of low nucleosome occupancy within the Chinese hamster dihydrofolate reductase initiation zone. *Nucleic Acids Res* **39**:3141-55.
68. **Aladjem MI, Fanning E.** 2004. The replicon revisited: an old model learns new tricks in metazoan chromosomes. *EMBO Rep* **5**:686-91.
69. **Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, et al.** 2011. Determinants of nucleosome organization in primary human cells. *Nature* **474**:516-20.
70. **Segal E, Fondufe-Mittendorf Y, Chen L, Thastrom A, Field Y, et al.** 2006. A genomic code for nucleosome positioning. *Nature* **442**:772-8.
71. **Jin C, Zang C, Wei G, Cui K, Peng W, et al.** 2009. H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat Genet* **41**:941-5.
72. **Berbenetz NM, Nislow C, Brown GW.** 2010. Diversity of eukaryotic DNA replication origins revealed by genome-wide analysis of chromatin structure. *PLoS Genet* **6**.
73. **Muller P, Park S, Shor E, Huebert DJ, Warren CL, et al.** 2010. The conserved bromo-adjacent homology domain of yeast Orc1 functions in the selection of DNA replication origins within chromatin. *Genes Dev* **24**:1418-33.
74. **Noguchi K, Vassilev A, Ghosh S, Yates JL, DePamphilis ML.** 2006. The BAH domain facilitates the ability of human Orc1 protein to activate replication origins in vivo. *Embo J* **25**:5372-82.
75. **Miotto B, Struhl K.** 2010. HBO1 histone acetylase activity is essential for DNA replication licensing and inhibited by Geminin. *Mol Cell* **37**:57-66.

76. **Tardat M, Brustel J, Kirsh O, Lefevbre C, Callanan M, et al.** 2010. The histone H4 Lys 20 methyltransferase PR-Set7 regulates replication origins in mammalian cells. *Nat Cell Biol* **12**:1086-93.
77. **Sugimoto N, Yugawa T, Iizuka M, Kiyono T, Fujita M.** 2011. The chromatin remodeler sucrose nonfermenting 2 homolog (SNF2H) is recruited onto DNA replication origins through interaction with CDC10-dependent transcript 1 (CDT1) and promotes the pre-replication complex formation. *J Biol Chem.*
78. **Gilbert DM.** 2010. Evaluating genome-scale approaches to eukaryotic DNA replication. *Nat Rev Genet* **11**:673-84.
79. **Mechali M.** 2010. Eukaryotic DNA replication origins: many choices for appropriate answers. *Nat Rev Mol Cell Biol* **11**:728-38.
80. **Antequera F.** 2004. Genomic specification and epigenetic regulation of eukaryotic DNA replication origins. *Embo J* **23**:4365-70.
81. **Hyrien O, Maric C, Mechali M.** 1995. Transition in specification of embryonic metazoan DNA replication origins. *Science* **270**:994-7.
82. **Diffley JF.** 2010. The many faces of redundancy in DNA replication control. *Cold Spring Harb Symp Quant Biol* **75**:135-42.
83. **Letessier A, Millot GA, Koundrioukoff S, Lachages AM, Vogt N, et al.** 2011. Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature* **470**:120-3.
84. **Ozeri-Galai E, Lebofsky R, Rahat A, Bester AC, Bensimon A, et al.** 2011. Failure of origin activation in response to fork stalling leads to chromosomal instability at fragile sites. *Mol Cell* **43**:122-31.