

UNIVERSITY OF BIRMINGHAM

University of Birmingham
Research at Birmingham

Metazoan DNA replication origins

Ganier, Olivier ; Prorok, Paulina; Akerman, Ildem; Méchali, Marcel

License:

Creative Commons: Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

Document Version

Peer reviewed version

Citation for published version (Harvard):

Ganier, O, Prorok, P, Akerman, I & Méchali, M 2019, 'Metazoan DNA replication origins', *Current Opinion in Cell Biology*, vol. 58, pp. 134-141.

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:

Checked for eligibility: 01/05/2019

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

METAZOAN DNA REPLICATION ORIGINS

Olivier Ganier, Paulina Prorok, Ildem Akerman, and Marcel Méchali*, Institute of Human Genetics, UMR 9004, CNRS-Université de Montpellier.

*Corresponding author

Abstract

DNA replication starts with the opening of DNA at sites called DNA replication origins. From the single sequence-specific DNA replication origin of the small *E. Coli* genome, up to the thousands of origins that are necessary to replicate the large human genome, strict sequence specificity has been lost. Nevertheless, genome-wide analyses performed in the recent years, using different mapping methods, demonstrated that there are precise locations along the metazoan genome from which replication initiates. These sites contain relaxed sequence consensus and epigenetic features. There is flexibility in the choice of origins to be used during a given cell cycle, probably imposed by evolution and developmental constraints. Here, we will briefly describe their main features.

Introduction: Choice and flexibility in the usage of DNA replication origins in eukaryotes

DNA replication starts with the opening of DNA at sites called DNA replication origins. A single replication origin is sufficient to completely duplicate the *E. coli* genome during the 30min-long cell cycle. The human genome is 700-fold larger, the replication fork speed is 30-fold lower. Therefore, with the same cell cycle length than in *E. Coli*, 21 000 origins would be necessary, 21 000 origins would be necessary to fully duplicate the human genome. Early experimental findings suggested that between 30 000 and 100 000 replication origins are activated in a mammalian cell cycle, and recent methods that allow characterizing origins at the genome-wide level gave similar numbers [1]. Two important features were also confirmed. First, from yeast to human cells, the number of potential origins is large, but only 20 to 30% of them are activated each cell cycle. Moreover, there is a large flexibility in the choice of the origins to be activated in each cell of a given population, and this choice is apparently stochastic. Second, selected origins are not activated all at the same time, but according to a highly regulated timing of individual activations during the entire S phase.

Identikit of a replication origin

From bacteria to higher eukaryotes, replication origins contain at least two elements: a replication initiation site (**IS**), where DNA synthesis is activated, and an upstream adjacent element where the pre-replication complex (**pre-RC**) is assembled (Figure 1). Pre-RC formation is initiated by ORC binding to the replication origin at the G1 phase of the cell cycle, followed by CDC6, CDT1, and finally the MCM DNA helicase, in a reaction called **replication origin licensing**. This complex remains inactive up to S-phase start, when Dbf4-dependent and cyclin-dependent kinases phosphorylate the MCM helicase that associates with the GINS tetramer,

MCM 10 and CDC45 to form the two active CMG helicases. At this stage, the DNA double helix is opened allowing the recruitment of the DNA polymerase machinery and the formation of the two replication forks [2]; [3], for reviews).

Due to the pre-RC steric occupancy on DNA and the neighboring nucleosome that cover at least 210 bp of DNA [3]; [4]), the IS cannot be at the pre-RC site, but at least 210 bp downstream of it. Other constraints also predict that the IS should be at a variable position downstream of the pre-RC site. First, the MCM helicase will have to unwind the DNA over a minimum length to allow the binding of the DNA polymerase machinery to the unwound DNA. Second, in some cases, MCM helicases can be displaced from their initial binding sites along DNA, leading to initiation of DNA replication further away from the pre-RC binding site [5]. Third, DNA polymerases do not require a specific sequence to start DNA synthesis. Indeed, the start of DNA synthesis might depend on the accessibility of the unwound DNA that may fluctuate according to the origin and its epigenetic features. These constraints are likely to explain the IS size variations from a sharp peak to an initiation zone.

Mapping DNA replication origins

In the last years, five main methods have been developed to map replication origins genome-wide. Table 1 summarizes these methods and the average size of IS zones and peaks.

The highest resolution is reached with the Small Nascent Strand-sequencing (SNS-seq) method that is based on the purification of nascent RNA-primed DNA synthesized at origins. Nascent DNA strands are the first to be produced at the leading strand of replication origins. They contain RNA primers (8 to 12 nucleotides) synthesized by DNA polymerase α -primase to start DNA replication. Usually, RNA-primed nascent DNA fragments (500-1500 bp in size) are first purified by incubation with lambda exonuclease to eliminate any contaminating DNA that was not RNA primed. This method is the most accurate to map replication origins.

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) against ORC also has been used for origin mapping. As ORC has additional functions in mitosis, ORC1 should be the most suitable target because this subunit is already present at pre-RCs in G1 and then is degraded during the S phase.

In the IS sequencing (Ini-Seq) method, BrdU immunoprecipitation is used to purify nascent DNA labeled with short BrdUTP pulses at the beginning of S phase (reagents are added to block DNA strand elongation). This approach is relatively accurate, but maps only early replication origins.

The replication bubble sequencing (bubble-seq) method relies on the structure of open replication origins to trap them on agarose gels. It requires the use of restriction enzymes to separate such structures that are then isolated and sequenced. This method allows defining relatively large initiation zones, but cannot distinguish between the IS and the replicator, due to the large size of the analyzed fragments.

The Okazaki fragment sequencing (OK-seq) method is based on the change of directionality in Okazaki fragments synthesized at replication origins. It only maps wide replication domains (30 Kbp on average) in which several replication origins that cannot be distinguished with this method might be included. Moreover, it assumes an equal level of background from both strands (a technical impossibility), and is insensitive to weaker initiation sites [6, 7]. A recent study described a Bayesian inference algorithm to analyze OK-seq data, and proposed that it can be used to detect origin obscuring suggesting that improvements of analysis methods may lead to more sensitive, accurate and higher resolution detection of DNA replication origins [8]. The replication origins identified with these five methods (i.e., the origin repertoire) reflect the sum of the activated origins in a cell population. Methods to visualize activated replication origins in single cells are still lacking. This information might be soon accessible thanks to a nanopore sequencing technology that can identify a base sequence from the electric readout produced when single-strand DNA passes through a protein pore [9] [10].

Motifs at DNA replication origins

In *S. cerevisiae*, DNA replication initiates from regions that contain an A/T-rich motif called the autonomously replicating sequence (ARS) [11]). The ARS is necessary, but not sufficient for origin specification, because only about 3% of ARS sequences in the genome are used as origins [12].

The identification of an ARS-like motif at metazoan DNA replication origins was more challenging. In mouse cells, genome-wide origin mapping using SNS-seq identified a G-rich region called Origin G-rich Repeat Element (OGRE) [13] [14]. This sequence motif is located ~280 bp from the IS, in a nucleosome-free region[15]) Other SNS-seq-based studies also linked replication origins to G-rich sequences, such as G-quadruplex (G4) elements and CpG islands [16], [17], [18], [19]. Similar G-rich sequences have also been identified around DNA replication IS using the Ini-seq method [20]. In fact, the majority of mammalian origins identified by SNS-seq and half of all ORC1 and ORC2 binding sites contain G4 structures [21] [22] and Akerman et al, unpublished data). Moreover, the requirement of a G4 element for replication initiation activity has been demonstrated for individual origins ([23] and Prorok et al, to be published), but has not been generalized to all origins yet. CpG islands also are highly associated with IS (about 70% of CpG islands host DNA replication origins). However, most DNA replication origins do not contain CpG islands, and the majority of putative G4 structures do not host IS. This suggests that a G-rich sequence, such as a G4 element or a CpG island, is necessary but not sufficient for DNA replication origin specification. Moreover, chromatin immunoprecipitation (ChIP) of pre-RC components suggests that ORC1 and ORC2 are localized in open chromatin regions, with a preference for G/C-rich sequences, but without a specific sequence motif [21, 24] [21, 24])

A study in *Caenorhabditis elegans* embryos found that in early embryogenesis, origins are associated with asymmetrical A/T-rich sequences, while in late embryogenesis they are associated with G/C-rich sequences, such as CpG islands [25]. This result confirms several previous observations showing that replication origins are more frequent in transcription promoter regions [26].

In *Drosophila melanogaster*, genome-wide mapping of ORC [27] did not identify any specific motif. Nevertheless, machine-learning algorithms could be used to discriminate between the DNA sequences of ORC-bound regions and random genomic sequences, through the detection of nucleosome-positioning signals that promote open chromatin [27]. A subsequent study that used SNS-seq to identify IS revealed a significant association between replication origins and putative G4 elements in the *D. melanogaster* genome) [28].

A recent work suggests that within the initiation zones identified using OK-seq method, regions characterized by poly-A or poly-T tracks are enriched in DNA replication origins [6]. These putative IS do not directly overlap with known ORC2-binding sites; however, about ~20% of them resides within 2Kb of known ORC2 binding sites (Akerman et al, unpublished data).

Transcription

Active transcription can be considered as a factor that positively affects replication activity. Genome-wide origin mapping performed in several organisms and using different mapping methods led to the conclusion that origins are strongly enriched in proximity of transcription start sites (TSS) in fly [28], plant [29], mouse [13, 15] and human cells [22], [7, 24, 30]. Interestingly, replication origins situated close to transcribed enhancers display higher replication activity compared with those close to non-transcribed enhancers, suggesting that transcription activity contributes to both replication origin position and strength [31].

Conversely, dormant replication origins (i.e., licensed but not activated) seem to be associated with GC-poor regions within transcriptionally silent regions. Interestingly, replication origins associated with cell transformation also are within transcriptionally inactive, late-replicating and GC-poor regions [32].

Epigenetic marks

In higher eukaryotes, the positioning of replication origins is not dictated by a single DNA sequence motif and appears to be more dependent on the chromatin environment. For instance, it has been suggested that ORC triggers a permissive chromatin environment through the direct recruitment of histone acetyltransferases, such as GCN5 [33]; [34]). Moreover, the OGRE/G4 element upstream of the pre-RC is nucleosome-free, whereas the IS contains a labile nucleosome that might be removed during the initiation step. Recent studies on the telomere-associated protein RIF1 have further highlighted the potential role of G4s in origin firing. Specifically, it has been suggested that RIF1 delays pre-RC activation until late S-phase at heterochromatin by organizing late replication-timing domains via its association with G4s that are enriched at telomeres [35], [36], [37]. Furthermore, a role of G4 in origin activation was also suggested by the presence of a functional G4-binding domain within MTBP [38], a protein cooperating with Treslin to trigger Cdc45 incorporation [39].

An exhaustive analysis of 43 epigenetic marks present at replication origins in mouse embryonic stem (ES) cells defined three main classes of replication origins that are

characterized by different chromatin environments [37]). Class 1 represents isolated origins that are poor in epigenetic marks and that are preferentially activated in late S phase. Class 2 origins are particularly rich in enhancer elements, and consequently correlate with the histone mark H3K4me1, but also with the specific DNA modifications 5fC and 5hmC. Class 3 origins are associated with open chromatin and polycomb protein-enriched regions, and with H3K4me2-3, H3K27ac, H3K9ac and H3K27me3. In mouse ES cells, H3K4me3 and H3K27me3 characterize polycomb-enriched bivalent domains that are strongly predictive of replication [15], [40]. Interestingly, H3K27me3 depletion in plants leads to re-replication of these genomic locations, arguing for an active role of the epigenetic environment in origin firing decision [41]. Similarly, conversion of H3K4me3 into H3K4me2 is catalyzed by the JARID1C demethylase that is crucial for origin activation in early S phase [42]. As JARID1C triggers the recruitment of CDC45 and PCNA that occurs simultaneously with H3K4me3 demethylation [43], it is tempting to hypothesize that the epigenetic environment might also contribute to origin firing coordination. Finally, H4K20 methylation also has been linked to origins. H4K20me1 seems to regulate late replicating regions in *D. melanogaster* [44] [45]. Conversely, in human cells, H4K20me1 is enriched at early replication domains [17, 46, 47]. In mammals, H4K20me3 regulates late-replication domains, likely through ORCA recruitment [48] and by controlling the nuclear chromatin organization. In adult cells, H4K20me3 depletion results in ORC and MCM overloading that perturb the replication program [49]. The decreased chromatin compaction observed in the absence of H4K20me3 is reminiscent of the open genome organization in early embryos that is not compatible with the H4K20me3 heterochromatin mark [50]. These observations suggest that lower chromatin compaction promotes higher replication origin density at very early developmental stages.

On the other hand, silencing of three H1 genes induces changes in genome organization and chromatin compaction that lead to more frequent abortive replication initiation, marked by a high number of stalled forks and ssDNA accumulation [51].

In conclusion, DNA replication initiation and pre-RC formation appear linked with genomic sequences that promote open chromatin, including G-rich motifs that potentially form G4 elements. These elements could represent a specific DNA structure, but cannot support a strict consensus motif (like the ARS element) at metazoan origins. They might be involved in pre-RC component recruitment and/or in origin specification by promoting nucleosome positioning and open chromatin. It is worth noting that while poly-A/T tracts play a role in nucleosome exclusion in yeast [52], the same function could be ensured by G-rich/G4 element in pluricellular organisms.

Licensing within the cell cycle

In eukaryotic cells, DNA replication and cell division are timely separated, ensuring that the whole genome is replicated once and only once per cell division. To prevent DNA re-replication within the same cell cycle and to ensure that DNA is fully replicated before cell division, the licensing reaction is regulated in time, integrated within the cell cycle, and occurs specifically in the G1 phase. To avoid re-replication, *de novo* pre-RC assembly on the newly synthesized

DNA is inhibited by different redundant mechanisms [53]; [54]; [55]; [56]. To date, three main mechanisms are known: i) pre-RC component inactivation by post-translational modification, ii) pre-RC component degradation, and iii) binding of the metazoan protein geminin (present from G1/S phase to metaphase) to the licensing factor CDT1 to inhibit its function.

Unexpectedly, it has been recently shown that the mechanisms preventing re-replication are not identical for all origins, and that two different degradation complexes act in parallel to prevent re-replication at distinct pools of origins. This suggests that replication origin activation may be differently regulated regarding their genomic localization [57].

Origin firing in time and space during S phase

Replication origins are not activated synchronously, but are fired in a coordinated manner. Indeed, monitoring of ongoing replicative forks in fixed or live nuclei revealed a robust spatio-temporal organization of replication during S phase [58], with origins gathered within replicating foci the size and location of which vary according to the replication origin timing. These replication foci have been characterized at the single replicon scale using 3D-SIM super resolution microscopy [59]. The results challenge the previous depiction of replication foci based on conventional microscopy studies as replication factories in which the replicative machinery is shared by several replicons. This vision of single replicons was completed by showing that replication domains typically contain four co-replicating regions that are gathered within a 150 nm region during their co-replication [59, 60].

It was originally postulated that the local gathering of replicative forks might result from a 3D genomic organization based on the formation of DNA loops that temporarily allow physical chromatin contacts [58]. Replication domains are megabase-sized regions the boundaries of which co-localize with CTCF binding sites [61, 62]. Remarkably, genome-wide analysis of cell populations and single cells associated the boundary of temporally regulated replication domains with the boundaries of Topological Associated Domains (TADs), suggesting an interconnection between the genome 3D organization and the timing of origin firing [61, 63]. Early replicating domains are enriched in origins, and are typically associated with highly transcribed regions, enriched in open chromatin marks (for review [64]). Conversely, late heterochromatin replicating domains are at the nuclear periphery [65]. Genome-editing approaches have been used to identify genetic elements that might regulate the timing of replication domains. Unexpectedly, they showed that replication timing is, at least in part, governed by genetic sequences involved in intra-domain contacts and not by the CTCF dependent-TAD boundaries. Importantly, these intra-domain contacts of DNA *in cis* confirmed the formation of DNA loops within replication domains and their importance in replication timing maintenance [62]. It has been recently proposed that RIF1, a known late-replication domain regulator that is involved in late replication timing maintenance [37]; [66]; [67], [68]), induces chromatin loop formation by simultaneous interactions with multiple G4 structures [69].

Moreover, direct evidence of interplays between replication timing, chromatin organisation and epigenome was provided by repositioning of a heterochromatin region to

the nuclear periphery. This relocation decreased levels of the constitutive heterochromatin marks H3K9me3, H3K9me2 and H4K20me3, marks and trigger a switch from late to early replicating timing [70].

Replication origins and human diseases

The importance of origin licensing regulation is highlighted by the deleterious phenotypes induced by its deregulation. A sub-optimal amount of licensing factors has been linked to Meier-Gorlin syndrome (MGS), a rare disease characterized by primordial dwarfism and specific developmental defects. Hypomorphic mutations of proteins involved in pre-RC formation or activation have been detected in nearly 70% of patients with MGS (reviewed by [71]; [72]; [73]; [74]; [75]. In addition, cilia formation impairment caused by pre-RC component deficiency [76] could contribute to MGS clinical features [77] and for review [78].

Origin licensing deregulation has also been involved in cancer. Indeed, pre-RC protein expression is often altered in cancer cells [79]; [80]. Moreover, deregulated licensing proteins can trigger cell transformation ([81], [56, 82], [83], [84], [85], [86], [87] [88] for review [79] and [56, 88]), Several studies have shown that licensing deregulation induces replication stress and subsequently genomic instability, two major features of cancer cells [56], [89], [87]. In normal conditions, origin licensing occurs preferentially at the TSS of transcribed genes, therefore ensuring the co-linearity of the replication and transcription machineries and preventing their deleterious collision [31, 90]. Importantly, when overexpressed, the cyclin E and Myc oncogene proteins trigger premature S-phase entry that results in DNA replication stress through improper activation of a set of origins that are not activated in normal conditions [91, 92]. Oncogene-induced initiation zones are within highly transcribed regions and generate conflicts between the replication and transcription machineries, leading to fork collapse, DNA damage, and genetic instability [92]).

All these observations indicate that although no DNA is synthesized in G1, this phase is crucial in normal somatic cells for ensuring correct origin licensing to maintain genomic stability. Interestingly, the G1 phase length is developmentally regulated. It is extremely short in ES cells and almost absent in early *Xenopus laevis* and *C. elegans* embryos [93, 94]. Importantly, in this context of rapid cell divisions, replication origin sites and density are different from those observed during late developmental stages [25, 95]. Rapid licensing during early development seems to be important for pluripotency maintenance in mouse ES cells, because experimentally slowing down licensing triggers their differentiation [96].

Perspectives

During the last years, genome-wide analyses allowed broadening our understanding of metazoan replication origins. Currently, the flexibility in the choice of the exact set of replication origins to be used and the plasticity of their sequence are the main reproducible features. Transcription boosts the activity of replication origins and their clustering, partly because of the open chromatin environment that is more favorable to replication factors binding. As opposed to what observed in bacteria and yeasts, strict consensus motifs have not

been found in metazoan. Nevertheless, some sequence features have been reproducibly observed, such as repeated G-rich motifs. It is not entirely clear whether their ability to form G4 allows defining a specific structure that can be used for initiation of DNA replication, or whether their upstream position relative to the IS is linked to a role in controlling pre-RC activation. Epigenetic features also clearly influence origin activity, but again, no universal strict universal feature that specifies all replication origins has been identified yet and this may reflect evolutionary constraints that imposed some flexibility specific to metazoan to adapt to different cell behaviors/identity and to different chromatin environments.

References

1. Prioleau, M.N. and D.M. MacAlpine, *DNA replication origins-where do we begin?* *Genes Dev*, 2016. **30**(15): p. 1683-97.
2. Bleichert, F., M.R. Botchan, and J.M. Berger, *Mechanisms for initiating cellular DNA replication*. *Science*, 2017. **355**(6327).
3. Remus, D. and J.F. Diffley, *Eukaryotic DNA replication control: lock and load, then fire*. *Curr Opin Cell Biol*, 2009. **21**(6): p. 771-7.
4. Belsky, J.A., et al., *Genome-wide chromatin footprinting reveals changes in replication origin architecture induced by pre-RC assembly*. *Genes Dev*, 2015. **29**(2): p. 212-24.
5. Edwards, M.C., et al., *MCM2-7 complexes bind chromatin in a distributed pattern surrounding the origin recognition complex in *Xenopus* egg extracts*. *J Biol Chem*, 2002. **277**(36): p. 33049-57.
6. Tubbs, A., et al., *Dual Roles of Poly(dA:dT) Tracts in Replication Initiation and Fork Collapse*. *Cell*, 2018. **174**(5): p. 1127-1142 e19.
7. Petryk, N., et al., *Replication landscape of the human genome*. *Nat Commun*, 2016. **7**: p. 10208.
8. Bazarova, A., et al., *Bayesian inference of origin firing time distributions, origin interference and licencing probabilities from Next Generation Sequencing data*. 2019.
9. Hennion, M., et al., *Mapping DNA replication with nanopore sequencing*. *BioRxiv*, 2018.
10. Mueller, C., et al., *Capturing the dynamics of genome replication on individual ultra-long nanopore sequence reads*. *BioRxiv*, 2018.
11. Stinchcomb, D.T., K. Struhl, and R.W. Davis, *Isolation and characterisation of a yeast chromosomal replicator*. *Nature*, 1979. **282**(5734): p. 39-43.
12. Nieduszynski, C.A., Y. Knox, and A.D. Donaldson, *Genome-wide identification of replication origins in yeast by comparative genomics*. *Genes Dev*, 2006. **20**(14): p. 1874-9.
13. Cayrou, C., et al., *Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features*. *Genome Res*, 2011. **21**(9): p. 1438-49.
14. Cayrou, C., et al., *New insights into replication origin characteristics in metazoans*. *Cell Cycle*, 2012. **11**(4): p. 658-67.
15. Cayrou, C., et al., *The chromatin environment shapes DNA replication origin organization and defines origin classes*. *Genome Res*, 2015. **25**(12): p. 1873-85.

**An exhaustive characterization of genetic and epigenetic marks at mouse DNA replication origins, which defines three main classes of origins.*

16. Cayrou, C., et al., *Genome-scale identification of active DNA replication origins*. *Methods*, 2012. **57**(2): p. 158-64.
17. Smith, O.K., et al., *Distinct epigenetic features of differentiation-regulated replication origins*. *Epigenetics Chromatin*, 2016. **9**: p. 18.
18. Bartholdy, B., et al., *Allele-specific analysis of DNA replication origins in mammalian cells*. *Nat Commun*, 2015. **6**: p. 7051.
19. Valenzuela, M.S., et al., *Preferential localization of human origins of DNA replication at the 5'-ends of expressed genes and at evolutionarily conserved DNA sequences*. *PLoS One*, 2011. **6**(5): p. e17308.
20. Langley, A.R., et al., *Genome-wide identification and characterisation of human DNA replication origins by initiation site sequencing (ini-seq)*. *Nucleic Acids Res*, 2016. **44**(21): p. 10230-10247.
21. Dellino, G.I., et al., *Genome-wide mapping of human DNA-replication origins: levels of transcription at ORC1 sites regulate origin selection and replication timing*. *Genome Res*, 2013. **23**(1): p. 1-11.
22. Picard, F., et al., *The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells*. *PLoS Genet*, 2014. **10**(5): p. e1004282.
23. Valton, A.L., et al., *G4 motifs affect origin positioning and efficiency in two vertebrate replicators*. *EMBO J*, 2014. **33**(7): p. 732-46.
24. Miotto, B., Z. Ji, and K. Struhl, *Selectivity of ORC binding sites and the relation to replication timing, fragile sites, and deletions in cancers*. *Proc Natl Acad Sci U S A*, 2016. **113**(33): p. E4810-9.
25. Rodriguez-Martinez, M., et al., *The gastrula transition reorganizes replication-origin selection in *Caenorhabditis elegans**. *Nat Struct Mol Biol*, 2017. **24**(3): p. 290-299.
26. Sequeira-Mendes, J., et al., *Transcription initiation activity sets replication origin efficiency in mammalian cells*. *PLoS Genet*, 2009. **5**(4): p. e1000446.
27. MacAlpine, H.K., et al., *Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading*. *Genome Res*, 2010. **20**(2): p. 201-11.
28. Comoglio, F., et al., *High-resolution profiling of Drosophila replication start sites reveals a DNA shape and chromatin signature of metazoan origins*. *Cell Rep*, 2015. **11**(5): p. 821-34.
29. Costas, C., et al., *Genome-wide mapping of Arabidopsis thaliana origins of DNA replication and their associated epigenetic marks*. *Nat Struct Mol Biol*, 2011. **18**(3): p. 395-400.
30. Sugimoto, N., et al., *Genome-wide analysis of the spatiotemporal regulation of firing and dormant replication origins in human cells*. *Nucleic Acids Res*, 2018. **46**(13): p. 6683-6696.

** Active replication origins are strongly enriched at TSS and within GC-rich and G4-forming regions. In contrast, dormant origins that are licenced but not activated under normal condition are situated in transcription silent and AT-rich regions.*

31. Chen, Y.H., et al., *Transcription shapes DNA replication initiation and termination in human cells*. *Nat Struct Mol Biol*, 2019. **26**(1): p. 67-77.

32. Wu, X., et al., *Developmental and cancer-associated plasticity of DNA replication preferentially targets GC-poor, lowly expressed and late-replicating regions*. *Nucleic Acids Res*, 2018. **46**(19): p. 10532.
 33. Espinosa, M.C., et al., *GCN5 is a positive regulator of origins of DNA replication in *Saccharomyces cerevisiae**. *PLoS One*, 2010. **5**(1): p. e8964.
 34. Giri, S., et al., *Orc5 induces large-scale chromatin decondensation in a GCN5-dependent manner*. *J Cell Sci*, 2016. **129**(2): p. 417-29.
- *The prereplication complex is involved in creation of a favourable environment for the replication initiation. ORC5 recruits GCN5 histone acetylase that increases chromatin accessibility and facilitate origin firing.*
35. Alver, R.C., et al., *Reversal of DDK-Mediated MCM Phosphorylation by Rif1-PP1 Regulates Replication Initiation and Replisome Stability Independently of ATR/Chk1*. *Cell Rep*, 2017. **18**(10): p. 2508-2520.
 36. Foti, R., et al., *Nuclear Architecture Organized by Rif1 Underpins the Replication-Timing Program*. *Mol Cell*, 2016. **61**(2): p. 260-73.
 37. Kanoh, Y., et al., *Rif1 binds to G quadruplexes and suppresses replication over long distances*. *Nat Struct Mol Biol*, 2015. **22**(11): p. 889-97.
- *Involvement of Rif1 in the activation of yeast dormant origins through binding to G quadruplexes regions.*
38. Kumagai, A. and W.G. Dunphy, *MTBP, the partner of Treslin, contains a novel DNA-binding domain that is essential for proper initiation of DNA replication*. *Mol Biol Cell*, 2017. **28**(22): p. 2998-3012.
 39. Boos, D., M. Yekezare, and J.F. Diffley, *Identification of a heteromeric complex that promotes DNA replication origin firing in human cells*. *Science*, 2013. **340**(6135): p. 981-4.
 40. Vergara, Z., et al., *Retrotransposons are specified as DNA replication origins in the gene-poor regions of *Arabidopsis* heterochromatin*. *Nucleic Acids Res*, 2017. **45**(14): p. 8358-8368.
 41. Jacob, Y., et al., *Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases*. *Nature*, 2010. **466**(7309): p. 987-91.
 42. Rondinelli, B., et al., *H3K4me3 demethylation by the histone demethylase KDM5C/JARID1C promotes DNA replication origin firing*. *Nucleic Acids Res*, 2015. **43**(5): p. 2560-74.
 43. Liang, Z., et al., *Proliferating cell nuclear antigen is required for loading of the SMCX/KMD5C histone demethylase onto chromatin*. *Epigenetics Chromatin*, 2011. **4**(1): p. 18.
 44. Li, Y., et al., *Methylation of histone H4 lysine 20 by PR-Set7 ensures the integrity of late replicating sequence domains in *Drosophila**. *Nucleic Acids Res*, 2016. **44**(15): p. 7204-18.
 45. Brustel, J., et al., *Histone H4K20 tri-methylation at late-firing origins ensures timely heterochromatin replication*. *Embo j*, 2017. **36**(18): p. 2726-2741.
 46. Beck, D.B., et al., *PR-Set7 and H4K20me1: at the crossroads of genome integrity, cell cycle, chromosome condensation, and transcription*. *Genes Dev*, 2012. **26**(4): p. 325-37.
 47. Kim, R., et al., *ColoWeb: a resource for analysis of colocalization of genomic features*. *BMC Genomics*, 2015. **16**: p. 142.

48. Wang, Y., et al., *Temporal association of ORCA/LRWD1 to late-firing origins during G1 dictates heterochromatin replication and organization*. Nucleic Acids Res, 2017. **45**(5): p. 2490-2502.

**ORCA establishes a repressive chromatin environment at a subset of origins, which primes them for late replication.*

49. Shoaib, M., et al., *Histone H4K20 methylation mediated chromatin compaction threshold ensures genome integrity by limiting DNA replication licensing*. Nat Commun, 2018. **9**(1): p. 3704.

**H4K20me3 mark is a regulator of nuclear organisation and contributes to high chromatin compaction within heterochromatin domains. Impaired heterochromatin domain formation upon H4K20me3 depletion leads to an increased loading of ORC and MCM on chromatin and severe defects in S-phase progression.*

50. Eid, A., et al., *SUV4-20 activity in the preimplantation mouse embryo controls timely replication*. Genes Dev, 2016. **30**(22): p. 2513-2526.

51. Almeida, R., et al., *Chromatin conformation regulates the coordination between DNA replication and transcription*. Nat Commun, 2018. **9**(1): p. 1590.

52. Sadeh, R. and C.D. Allis, *Genome-wide "re"-modeling of nucleosome positions*. Cell, 2011. **147**(2): p. 263-6.

53. Fragkos, M., et al., *DNA replication origin activation in space and time*. Nat Rev Mol Cell Biol, 2015. **16**(6): p. 360-74.

54. Hernandez-Carralero, E., et al., *Control of DNA Replication Initiation by Ubiquitin*. Cells, 2018. **7**(10).

55. Blow, J.J. and A. Dutta, *Preventing re-replication of chromosomal DNA*. Nat Rev Mol Cell Biol, 2005. **6**(6): p. 476-86.

56. Hills, S.A. and J.F. Diffley, *DNA replication and oncogene-induced replicative stress*. Curr Biol, 2014. **24**(10): p. R435-44.

57. Jang, S.M., et al., *The replication initiation determinant protein (RepID) modulates replication by recruiting CUL4 to chromatin*. Nat Commun, 2018. **9**(1): p. 2782.

58. Newport, J. and H. Yan, *Organization of DNA into foci during replication*. Curr Opin Cell Biol, 1996. **8**(3): p. 365-8.

59. Chagin, V.O., et al., *4D Visualization of replication foci in mammalian cells corresponding to individual replicons*. Nat Commun, 2016. **7**: p. 11231.

** see reference 60*

60. Xiang, W., et al., *Correlative live and super-resolution imaging reveals the dynamic structure of replication domains*. J Cell Biol, 2018. **217**(6): p. 1973-1984.

**These two articles describe super-resolution microscopy approaches to refine the 3D organization of replication foci and characterize the replicating domains as being formed by 4 coordinated active forks.*

61. Takahashi, S., et al., *Genome-wide stability of the DNA replication program in single mammalian cells*. Nat Genet, 2019. **51**(3): p. 529-540.

62. Sima, J., et al., *Identifying cis Elements for Spatiotemporal Control of Mammalian DNA Replication*. Cell, 2019. **176**(4): p. 816-830 e18.

** Gene editing experiments suggest that replication domains are defined by intra-domains cis elements interactions that imply loops, rather than by CTCF dependent borders.*

63. Pope, B.D., et al., *Topologically associating domains are stable units of replication-timing regulation*. *Nature*, 2014. **515**(7527): p. 402-5.
64. Rivera-Mulia, J.C. and D.M. Gilbert, *Replication timing and transcriptional control: beyond cause and effect-part III*. *Curr Opin Cell Biol*, 2016. **40**: p. 168-178.
65. Rhind, N. and D.M. Gilbert, *DNA replication timing*. *Cold Spring Harb Perspect Biol*, 2013. **5**(8): p. a010132.
66. Hiraga, S.I., et al., *Budding yeast Rif1 binds to replication origins and protects DNA at blocked replication forks*. *EMBO Rep*, 2018. **19**(9).
67. Hiraga, S.I., et al., *Human RIF1 and protein phosphatase 1 stimulate DNA replication origin licensing but suppress origin activation*. *EMBO Rep*, 2017. **18**(3): p. 403-419.
- * *A double control of initiation of DNA replication by Rif1: increase of licensing in G1 phase of the cell cycle, but coneracting replication origin activation*.
68. Hiraga, S., et al., *Rif1 controls DNA replication by directing Protein Phosphatase 1 to reverse Cdc7-mediated phosphorylation of the MCM complex*. *Genes Dev*, 2014. **28**(4): p. 372-83.
69. Moriyama, K., N. Yoshizawa-Sugata, and H. Masai, *Oligomer formation and G-quadruplex binding by purified murine Rif1 protein, a key organizer of higher-order chromatin architecture*. *J Biol Chem*, 2018. **293**(10): p. 3607-3624.
- * *Rif1 contributes to chromatin structure organization. Rif1 forms oligomeric complexes that bind simultaneously several G4s, which induces chromatin-loop formation*.
70. Heinz, K.S., et al., *Peripheral re-localization of constitutive heterochromatin advances its replication timing and impairs maintenance of silencing marks*. *Nucleic Acids Res*, 2018. **46**(12): p. 6112-6128.
71. de Munnik, S.A., et al., *Meier-Gorlin syndrome*. *Orphanet J Rare Dis*, 2015. **10**: p. 114.
72. Yao, L., et al., *Zebrafish cdc6 hypomorphic mutation causes Meier-Gorlin syndrome-like phenotype*. *Hum Mol Genet*, 2017. **26**(21): p. 4168-4180.
73. Fenwick, A.L., et al., *Mutations in CDC45, Encoding an Essential Component of the Pre-initiation Complex, Cause Meier-Gorlin Syndrome and Craniosynostosis*. *Am J Hum Genet*, 2016. **99**(1): p. 125-38.
74. Vetro, A., et al., *MCM5: a new actor in the link between DNA replication and Meier-Gorlin syndrome*. *Eur J Hum Genet*, 2017. **25**(5): p. 646-650.
75. Can, G., et al., *Helicase Subunit Cdc45 Targets the Checkpoint Kinase Rad53 to Both Replication Initiation and Elongation Complexes after Fork Stalling*. *Mol Cell*, 2019. **73**(3): p. 562-573 e3.
76. Stiff, T., et al., *Deficiency in origin licensing proteins impairs cilia formation: implications for the aetiology of Meier-Gorlin syndrome*. *PLoS Genet*, 2013. **9**(3): p. e1003360.
77. Maerz, L.D., et al., *Analysis of cilia dysfunction phenotypes in zebrafish embryos depleted of Origin recognition complex factors*. *Eur J Hum Genet*, 2019.
78. Klingseisen, A. and A.P. Jackson, *Mechanisms and pathways of growth failure in primordial dwarfism*. *Genes Dev*, 2011. **25**(19): p. 2011-24.
79. Blow, J.J. and P.J. Gillespie, *Replication licensing and cancer--a fatal entanglement?* *Nat Rev Cancer*, 2008. **8**(10): p. 799-806.
80. Williams, G.H. and K. Stoeber, *Cell cycle markers in clinical oncology*. *Curr Opin Cell Biol*, 2007. **19**(6): p. 672-9.
81. Lau, E., et al., *The role of pre-replicative complex (pre-RC) components in oncogenesis*. *FASEB J*, 2007. **21**(14): p. 3786-94.

82. Shima, N., et al., *A viable allele of Mcm4 causes chromosome instability and mammary adenocarcinomas in mice*. Nat Genet, 2007. **39**(1): p. 93-8.
83. Coulombe, P., et al., *A spontaneous Cdt1 mutation in 129 mouse strains reveals a regulatory domain restraining replication licensing*. Nat Commun, 2013. **4**: p. 2065.
84. Seo, J., et al., *Cdt1 transgenic mice develop lymphoblastic lymphoma in the absence of p53*. Oncogene, 2005. **24**(55): p. 8176-86.
85. Blanchard, Z., et al., *Geminin overexpression induces mammary tumors via suppressing cytokinesis*. Oncotarget, 2011. **2**(12): p. 1011-27.
86. Champeris Tsaniras, S., et al., *Geminin ablation in vivo enhances tumorigenesis through increased genomic instability*. J Pathol, 2018. **246**(2): p. 134-140.
87. Alvarez, S., et al., *Replication stress caused by low MCM expression limits fetal erythropoiesis and hematopoietic stem cell functionality*. Nat Commun, 2015. **6**: p. 8548.
88. Hernandez-Perez, S., et al., *DUB3 and USP7 de-ubiquitinating enzymes control replication inhibitor Geminin: molecular characterization and associations with breast cancer*. Oncogene, 2017. **36**(33): p. 4802-4809.
89. McIntosh, D. and J.J. Blow, *Dormant origins, the licensing checkpoint, and the response to replicative stresses*. Cold Spring Harb Perspect Biol, 2012. **4**(10).
90. Blin, M., et al., *Transcription-dependent regulation of replication dynamics modulates genome stability*. Nat Struct Mol Biol, 2019. **26**(1): p. 58-66.
- *Transcription activity positively influence replication and stimulation of transcription within common fragile site decreases genetic instability of these regions.*
91. Tanaka, S. and J.F. Diffley, *Deregulated G1-cyclin expression induces genomic instability by preventing efficient pre-RC formation*. Genes Dev, 2002. **16**(20): p. 2639-49.
92. Macheret, M. and T.D. Halazonetis, *Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress*. Nature, 2018. **555**(7694): p. 112-116.
- *Oncogenes can induce new initiation zones in transcribed regions, leading to transcription-replication conflicts and replication stress.*
93. Koreth, J. and S. van den Heuvel, *Cell-cycle control in Caenorhabditis elegans: how the worm moves from G1 to S*. Oncogene, 2005. **24**(17): p. 2756-64.
94. Newport, J.W. and M.W. Kirschner, *Regulation of the cell cycle during early Xenopus development*. Cell, 1984. **37**(3): p. 731-42.
95. Lemaitre, J.M., et al., *Mitotic remodeling of the replicon and chromosome structure*. Cell, 2005. **123**(5): p. 787-801.
96. Matson, J.P., et al., *Rapid DNA replication origin licensing protects stem cell pluripotency*. Elife, 2017. **6**.
- *An interesting link between replication origin licensing pluripotency.*

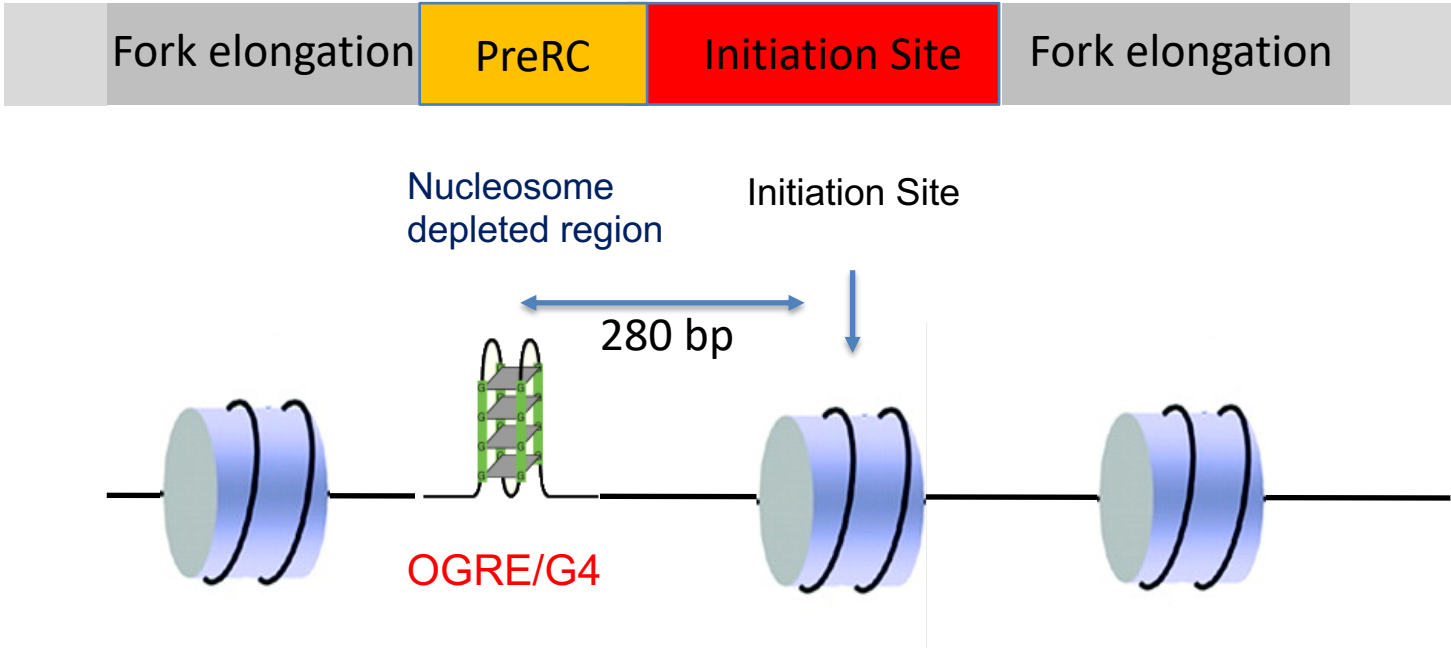


Figure 1: A scheme of a metazoan replication origin
 The preRC assembles to a sequence located upstream of the initiation site of DNA synthesis. Although its distance from the OGRE/G4 element could fit with preRC binding site, an experimental demonstration is lacking.

	Principle	Average peak size	Estimated resolution	Number of “origins” scored
SNS-Seq	RNA-primed Nascent strands	300 bp	< 2 Kbp	50 000- 200 000 IS
ChIP-seq (ORC2)	Chromatin immunoprecipitation	700 bp	< 2 Kbp	25 000 origins (Replicator - IS)
Ini-seq	Newly replicated DNA in mimosine synchronized nuclei	1200 bp	< 2 Kbp	25 000- 40 000 origins (Replicator - IS)
Bubble-Seq	Trapping of bubbles on agarose gels	6-20 kbp, unequal, depends on the restriction enzyme used	>4 Kbp	120 000 initiation zones
OK-seq	Transition in the strandedness of Okazaki fragments	34 Kbp, up to 150 Kbp	> 10 Kbp	5 000- 10 000 initiation domains

Main methods used in genome-wide identification of metazoan origins