The Protecting Role of Dormant Origins in Response to Replicative Stress

Lilas Courtot, Jean-Sébastien Hoffmann and Valérie Bergoglio*

CRCT, Université de Toulouse, Inserm, CNRS, UPS; Equipe labellisée Ligue Contre le Cancer, Laboratoire d’excellence Toulouse Cancer, 2 Avenue Hubert Curien, 31037 Toulouse, France

*Correspondence: valerie.bergoglio@inserm.fr; Tel.: +33 (0)5-82-74-16-57

Abstract: Maintenance of the human chromosomes stability requires a tight regulation of DNA replication to duplicate once and only once the entire genome of a single cell. In mammalian cells, origin activation is controlled in space and time by a cell specific and robust program called replication timing. About 100 000 of potential origins are loaded onto the chromatin at the G1 phase but only 20-30% are selected and active during the replication of a given cell. When the replication fork is slowed down by exogenous or endogenous sources, the cell need to activate more origins to complete the replication on time. Thus, the large choice of origins that can be activated may be a key player in the protection of the genome. The aim of this review is to discuss about the role of these dormant origins as housekeepers of the human genome in response to replicative stress.

Keywords: Dormant origins; replicative stress; replication timing; DNA damage; genome instability; cancer

1. Introduction: Eukaryotic origins and replication program

Because of their large genome, mammalian cells need thousands of replication forks, which are initiated from replication origins, to fully ensure the duplication of their DNA on a definite time before dividing. In human cells, this process is performed within about 10 hours and involves roughly the activation of 30 000 replication origins. Replication origins spread on about 100kb of DNA form a replicon cluster in which only one selected origin is going to be activated in normal condition. Clustering of selected origins from several replicons can be visualized as DNA replication foci [1]. The sequential activation of potential origins within groups is thought to play a direct role in defining the S-phase programme or replication program. At any given time of the S-phase, about 10% of replicons are activated and replicate simultaneously [2]. In addition, the temporal activation of origins in a specific region of the genome correlates with distinct pattern of replication sites as cells progress from early to late S phase. This multi-layered system has been adopted by metazoans cells to finely control the challenging goal to replicate the DNA in a limited time and to counteract obstacles that replication forks can encounter.

1.1. Origin licensing

Complete and robust DNA duplication requires the loading of Minichromosome maintenance (MCM2-7) helicase complex at many specific loci in the DNA that is named replication origins (ROs). The step of origin licensing is restricted to G1 phase of the cell cycle.

A key initial step in origin licensing is the binding of Pre-Recognition complex (Pre-RC) starting with the loading of origin recognition complex (ORC1-6) onto the chromatin. This ORC complex marks all potential origins providing spatial control of origin position. In higher eukaryotes, ORC binding sites has not been proven to be related to DNA sequence, in contrast to other organisms such
as yeast and bacteria [3,4]. It is currently assumed that multiple factors can characterize an origin such as CpG islands, G-quadruplexes, epigenetic marks, chromatin accessibility, sites of active transcription, or secondary DNA structures [5–10]. This is the reason why it has been so difficult to identify metazoan replication origins. ORCs are required for the chromatin loading of downstream replication factors. From late mitosis to G1 phase, ORCs are recognized by loading factor Cdc6 thanks to the interaction between Cdt1 and MCM subunits, allowing the formation of the pre-RC. The last step of licensing requires the loading of CDC45 and GINS on MCM2-7 to finally form the Pre-IC. This complex need DDK and CDK activities for its activation at the G1/S transition, then the polymerases and other replication factors are recruited to allow the origin firing (Figure 1).

Total MCMs level does not change throughout cell cycle but the amount of loaded MCMs is increasing from telophase in mitosis to the end of the G1/S transition. The repression of MCM loading during S phase and G2 ensures that re-replication of DNA does not occur. The inability to license new origins after the onset of S phase is a challenge for the cell because it needs to fully replicate the genome using its finite supply of licensed origins. Thus, the control of origin licensing is crucial to avoid re-replication, which can lead to aneuploidy, double-strand break, gene amplification and general genome instability [11–13]. On the other hand, un-replicated DNA due to double fork stalling and/or lack of origins can also lead to genome rearrangement and instability if the checkpoint is inactive or deficient [14–16].

**Figure 1**: Scheme describing origin licensing and firing. In late mitosis, the first step before licensing is the binding of ORC to the origin that will determine where the replication fork can initiate. ORC complex binding to DNA is required for the recruitment of the Cdc6 and Cdt1 in G1. Both Cdc6 and Cdt1 are necessary for the subsequent association of the MCM2–7 helicases onto chromatin. The presence of two Cdt1 binding sites on ORC is consistent with cooperative loading of two MCMs hexamers delivered by two Cdt1 molecules. MCM2–7 double hexamers encircle double stranded DNA and is able to slide along it but remains catalytically inactive until the G1/S transition when it is phosphorylated by both Cdk and Cdc7 activities. Once the
principal origin is fired, adjacent origins from the same replicon (flexible or dormant) are repressed (red dotted lines) by unclear mechanism involving several protein kinases (ATR, ATM, Chk1 and Chk2). Excess of MCMs that are not initiated will be removed by replication fork passing [17].

1.2. Spatial and temporal organization of replication origins

Eukaryotes origin usage is mainly dependant on two important and interdependent factors: space and time. Replication origins fire at a defined timing that remains the same among cell generations and is closely related to their spatial organization. Early replicating origins are mainly observed in transcriptionally active gene-rich domains with active epigenetic marks [18–23]. These chromosomal regions are enriched in potential origins with ORCs and MCMs, possibly explaining why they replicate early in the S phase. Late replication is observed in origin-poor regions that have low gene density and are enriched in heterochromatin hallmarks [23–26].

Several studies, that have compared replication timing (RT) and genome topology, have suggested that replicons are clustered into large (~1 Mb) chromatin units, close to the estimated size of replication foci, called replication domains, which are located at distinct areas of the nucleus during G1 and replicate concomitantly [27–29]. Replication factories are organized in the 3D nuclear space, with early replicating domains preferentially positioned inside the nucleus while late replicating domains are located at the nuclear periphery (Figure 2B). This spatial organization of early and late replicating domains can be observed by chromatin conformation mapping methods (Hi-C) [30,31].

There are several lines of evidence showing that nuclear matrix attachment step is necessary for initiation events [32–36]. The internal nuclear matrix maintains chromosomes within their respective territories and has been implicated in replication foci formation [37]. The organization of a replicon in clusters might thus reflect chromatin looping to bring origin of different replicon to a unique domain and exclude the flexible and/or dormant origins from this replication factory (Figure 2C). Cohesin complex may be a key player in chromatin looping because it has been found to physically interact with MCM2-7 complex and to be enriched at origin sites [38]. Replication domains are created by topological reorganization of the chromatin in nuclear space. In metazoans, the association of particular replication domains with sub-nuclear compartments will determine their replication timing. The setting-up of this compartmentalization occurs at a specific time of the G1 phase and is called the timing decision point (TDP) [39,40] (Figure 2).
Figure 2: Spatial organization of origins and replication timing (inspired from Fragkos et al., 2015 [17]). In light brown are represented nucleus structures: lamina and nuclear envelop. (A) Right after mitosis (early G1), pre-RC proteins (in black) are loaded onto the chromatin and mark potential origins. Only after the timing decision point (TDP), chromatin organization will define the replication timing of each replication domains. (B) At late G1, early replicating domains (in green) are close to nucleus center whereas late replication domains (in red) are located within lamina-associated domains, close to the nuclear periphery. (C) Active origins (in yellow) cluster in replication factories and are associated to the nuclear matrix (NM) leaving inactive (dormant or flexible) origins in DNA loops (in grey).

1.3. The different techniques allowing origin detection and identification

DNA fiber autoradiography provided the first quantitative assessment of origin densities in metazoan genomes [41]. Nowadays, this time-consuming assay has been replaced with fiber fluorography approaches (DNA combing or spreading), in which newly replicated DNA is substituted with halogenated nucleotide derivatives, such as bromo-, chloro-, or iododeoxyuridine (BrdU, CldU and IdU) and is visualized by indirect immunofluorescence using specific antibodies [42].

The use of next-generation DNA sequencing has led to the discovery of tens of thousands of potential replication origins in the human genome. Many independent approaches have been used and exploit the direct identification of DNA replication initiation intermediates. The first approach is based on the purification and quantification of short nascent strands (SNS) DNA [20]. In this method, 1.5–2.5 kb nascent strands specific to replication origins are purified thanks to their resistance to λ-exonuclease digestion due to the incorporation, by the primase, of small RNA primers at their 5’ ends [43]. This step is important to have a complete digestion of the large excess of broken genomic DNA
that would generate a background signal if not correctly removed. These genome-wide SNS-seq studies have extended and consolidated earlier microarray hybridization data, consistently showing that active origin sites often correlate with transcription start sites (TSS) and are located in GC-rich regions, near CpG islands and G-quadruplex [18,19,44]. A second approach is based on the sequencing of DNA replication bubbles [23]. Replication bubbles are formed as early intermediates after establishment of the two divergent replication forks. To isolate such bubbles, replicating DNA is fragmented by a restriction endonuclease and then embedded into agarose gel. Circular replication bubbles are trapped topologically by the polymerising agarose fibres whereas linear DNA fragments and Y-shaped replication forks are going out of the set gel [23]. Next-generation DNA sequencing of these trapped bubbles (bubble-seq) has identified more than 100,000 origin sites in the human genome. A third approach is using the sequencing of purified Okazaki fragments (OK-seq) for a genome-wide determination of replication fork polarity enabling the mapping of initiation and termination sites [45]. This analysis identified between 5000 and 10,000 broad initiation zones of up to 150kb that are mostly non-transcribed, often flanked by active genes, and typically contain a single but randomly located initiation event. Finally, a fourth method to identify metazoans replication origins has been recently described in a paper from Langley et al., (2016) [46]. Initiation site sequencing (ini-seq) consist in the direct labelling and subsequent immunoprecipitation of newly replicated DNA, synthesised a few minutes after highly synchronous initiation in a cell-free system. The biochemically controlled cell-free approach of ini-seq offers the important advantage to allow functional genome-wide studies of origin activation. Overall, these methods are giving a large set of new information on origins characteristics while being more and more accurate and complementary between one another.

1.4. Origin competence, efficiency and dormancy

The replication initiation program of metazoan cells exhibits a remarkably large flexibility with many origins that fire at disparate frequencies depending on cell lineage. MCMs and all the components of the Pre-RC are loaded in excess onto the chromatin in G1 to give the flexibility of choice. In addition to inter-lineage differences, origin flexibility is also observed within a cell population [36,47]. Very few origins activate almost 100% of the time, they are called constitutive origins [48]. Origins that do not initiate replication in all cell cycles are called flexible origins and represent the majority of origins. Both constitutive and flexible origins are detectable by whole genome analysis. By contrast, dormant origins are not detectable in whole-genome analyses and might be activated only if replication from adjacent origins is compromised. Origin choice may explain the observation that inter-origin distances measured by whole-genome sequencing are shorter than those measured by single-fibre analyses that gives information at single-cell level [49]. Whole-genome and single-fibre analyses have proven that, in many metazoan loci, replication initiates randomly within clusters of adjacent origins such that each cell within a population can uses different combinations of replication origins. It was suggested that this flexibility might help to coordinate DNA replication with transcription [50,51] and other nuclear processes, in a cell type-specific manner, and also to facilitate recovery when replication is challenged. Both the flexibility in establishing replication initiation sites and the lack of DNA consensus sequences raise the question of how nonspecific chromatin interactions can lead to accurate initiation at consistent origins [52].
The reason why some origins are activated preferentially to other is still unclear. There are currently two theories to explain how origins are selected. The first theory relies on the idea that origin choice occurs in G1, after the timing decision point, called origin decision point (ODP) that will determine which origin will be activated during the replication [53]. Although chromosomal loops and loop anchors are still poorly defined biochemically, for now we know that such chromosome architecture plays a predominant role in the regulation of DNA replication origin localization and activation [54]. The second theory is origin efficiency. It is an alternative model based on the stochastic firing of origins that may also explain replication timing. This model assumes varying origin efficiency instead of a strict origin-timing programme [55]. This difference of efficiency can first depend on replication origin location in the nucleus, chromatin structure and its epigenetic marks but also can be due to the amount of loaded MCM [4,56,57] and other pre-RC proteins.

2. Dormant origin activation in response to replicative stress

2.1 The notion of DNA replication stress

During DNA replication, the presence of endogenous or exogenous sources of stress causes individual replication forks slowing or stalling. Exogenous sources are mainly induced by genotoxic chemical agents and UV or ionic radiations. They can be many endogenous sources of stress, which are considered as replication barriers such as repetitive sequences, secondary structures (i.e. G-quadruplexes), telomeres, DNA–RNA hybrids, wrong incorporation of ribonucleotides, collisions between replication and transcription complexes, hypo-acetylation and compaction of chromatin, deregulation of origin activity or else reduction of the dNTP pool. Some region of the genome such as early-replicating fragile sites (ERFSs) and common fragile sites (CFSs) are more prone to replicative stress. Finally, overexpression or constitutive activation of oncogenes such as HRAS, c-Myc and cyclin E is an emerging source of replication stress [58]. All three oncogenes promote increased replication initiation or origin firing, leading to an elevated risk of nucleotides pools depletion and/or increased collisions with transcription complexes [59,60]. This may explain why supplementing cancer cells with exogenous nucleotides helps to decrease genomic instability [61].

The first consequence of replication stress is fork collapse, creating DNA single-strand (SSB) and/or double-strand breaks (DSB). These lesions need to be resolved before cell division by repair mechanisms such as homologous recombination (HR), Non-homologous end-joining (NHEJ) or else Micro-homology mediated end-joining (MMEJ). In a non-pathological context, checkpoint pathways (ATM and ATR signalling cascades) do not let the cell divide with an impaired genome. When some proteins of the checkpoint are mutated, such as p53, cell can divide while harbouring DNA lesions (breaks or un-replicated DNA) which leads to DNA breaks, chromosomal rearrangements and genomic instability [62–65].

2.2 Dormant origins discovery and their link with replicative stress

In 1977, J. Herbert Taylor [66] described for the first time that cells license more origins than the actual number of origin activated during the DNA replication process in CHO cells. Moreover, several studies in a range of eukaryotes, including Saccharomyces cerevisiae, humans, and Xenopus laevis, have demonstrated that Mcm2–7 complexes are loaded onto DNA in a large excess compared to DNA-bound ORC molecules and over the number of active replication origins [67–72]. Later, it
was observed in xenopus laevis [73] and in human cells that excess of Mcm2–7 provide a reservoir of dormant origins that is unused under normal replication conditions, but get activated when replication forks are challenged by replicative stress agents such as Aphidicolin or HU [74,75]. In these papers, they prove that the reduction of MCM2 loading by small interfering RNA (siRNA) leads to hypersensitivity to replication inhibitors due to the lack of dormant origins [74,75]. Moreover, Ge et al. [74] demonstrate that Chk1 activation is required for the firing of dormant origins within active replication clusters as well as for the repression of other replicons that are not yet active. This observation suggest the link between DNA damage response and dormant origin activation. Indeed, in vertebrates, inactivation or depletion of different proteins involved in genome maintenance, such as ATR [76,77], Chk1 [78–80] [81], Wee1 [82,83], BLM [84], Claspin [85,86], BRCA2 or Rad51 [87], elicits decrease in replication fork speed and, when studied, an increase in the rate of initiation events. This underline a link between fork speed and the amount of active origins.

2.3 The density of active origins depends on replication fork speed

Under normal condition, dormant origins do not fire and are passively replicated by the fork coming from adjacent activated origins. Thus, it makes sense to assume that replication fork speed can be a regulator of active origin density. In the papers of Anglana et al., and Courbet et al., [54,88] they used the AMPD2 locus in Chinese hamster ovary (CHO) cells to prove that indeed replication fork speed has a direct impact in the number of active origins. When the fork is slowed down by HU treatment, the density of active origins increases, not only the principal origin is active but also the adjacent ones, that normally are dormant. In contrast, under condition that accelerate fork speed (addition of adenine and uridine in the culture medium), less origins are active. They further showed that the cell starts to compensate the fork speed decrease within half an hour of treatment by setting in motion dormant origins, which are then able to change their statute within S phase. They observed that regulation of initiation events density occurs at the level of individual clusters, which is consistent with the fact that origins are functionally organized in replicon clusters [89]. Finally, using chemical inhibition of origin activity (CDC7 kinase inhibitor) and of DNA synthesis (APH), a more recent paper found that primary effects of replicative stress on fork rate can be distinguished from primary effects of replicative stress on origin firing [90]. All things considered, these results prove that the pattern of initiation depends on fork speed and thus is impacted by endogenous or exogenous replicative stress.

2.4 CFSs fragility due to the lack of dormant origins

Common fragile sites (CFSs) are chromosomic regions that play a major role in cancer initiation because of their specific instability under replication stress conditions. CFSs were first described in 1984 by Glover et al., [91] as gaps and constrictions in metaphase chromosomes of human lymphocytes grown under mild replication stress conditions (low dose of APH). These observations have been then confirmed in other organisms and are very likely to be the consequence of under-replication and/or DNA breaks caused by replication stress [92,93].

CFSs have been described for a long time now, but the cause of their fragility is still controversial [94,95]. It was first thought that CFSs fragility was linked to the multiple DNA sequences within CFS able to adopt secondary structures such as AT-rich sequences which constitute barriers to replication
forks [96–99]. However, the deletion of these sequences in some cancer cell lines does not avoid breaks at these loci [100–102] and the appearance of fragile sites depends on the cell type, which argues against a model where the DNA sequence would be the sole mechanism responsible for their instability. Genome-wide analysis of replication and DNA combing experiments showed that CFSs were localized in replication origin-poor regions of the genome [103,104]. This underlines the fact that replication of these regions is based on the fork capacity to replicate DNA with multiple non-B sequences over long distances, and their fragility is correlated with the absence of replication origin firing even though replication is slowed down. Most CFSs correspond to long genes (>300 kb), which might increase the risk of collision between transcription and replication machineries [105]. Although it has been demonstrated lately that the transcription of large genes does not systematically dictate CFS fragility [106], other studies indicate that active large transcription units drive extreme locus- and cell-type-specific genomic instability under replication stress, resulting in CFSs as different manifestations of perturbed replication dynamics [107,108]. Currently, it is believed that CFSs result from mitotic entry prior to the completion of replication in late-replicating regions [109,110], which are demonstrated as origin-poor or dormant origins deficient regions [111]. Overall, replication defects at fragile sites may be due to a low density of licensed origins or may reflect inefficient or delayed activation of replication forks under replication stress.

3. Dormant origins regulation: passive or active mechanism?

3.1 Activation of dormant origins by a “passive” mechanism

It is currently not clear what drives the firing of dormant origins when forks are slowed down or inhibited. One hypothesis could be that it does not involve an active mechanism but occurs as a consequence of the stochastic nature of origin firing [11,74]. Dormant origins have a precise laps of time to fire before being passively replicated, then inactivated, by forks from adjacent origins. When fork progression is impeded, the replication at dormant origins is delayed and therefore they have an increased probability to fire. The work described in Blow et al (2009) [111] uses a computer model to show that such passive mechanism can lead to similar levels of dormant origin activation to those seen in vivo, protecting thereby against the effects of fork stalling. This is managed essentially ‘for free’, just relying on origin firing stochasticity and without any need for additional regulatory pathways.

This simple mechanism can be sufficient to not require additional active pathways to activate dormant origins when the cell undergoes replication stress. However, it is possible that dormant origins are also, at least in part, regulated by active mechanisms.

3.2 Regulation of dormant origins by “active” mechanisms

3.2.1 ATR-Chk1 kinases as modulator of origin activation

The inhibition of replication forks activates DNA damage checkpoint kinases ATR-Chk1 and ATM-Chk2, which play many different functions such as stabilizing the forks, to delay or block the progression through the cell cycle, and promote lesion repair [112–114]. It is quite surprising that in response to replication stresses, the cell can both activate dormant origins and suppress overall origin
initiation. However, when replication forks are stalled, it only makes sense for dormant origins to be activated in the vicinity of the stalled forks and not elsewhere in the genome.

In normal S phase, Chk1 affects replication fork speeds by inhibiting excess origin firing [16,80,115]. It has been proven that, in response to low levels of replication fork inhibition induced by HU or APH, ATR and Chk1 preferentially inhibit the activation of new replication factories while allowing dormant origins to fire within the existing factories experiencing replicative stress [116]. This redirects origin activation within active factories and away from un-replicated regions of the genome, thereby avoiding the deleterious impact of replication fork stalling. The mechanism by which this happens is unclear, but one possibility is that ATR and Chk1 modestly reduce S phase Cdk levels, which has been shown to affect the level of active replication factories [117]. Alternatively, Chk1 could directly regulate negatively the initiation process through an interaction with Treslin, required for stabilizing Cdc45, GINS, MCM complex together with TOPBP1 [118–122]. Moreover, it has been proven lately that ATR inhibitor not only targets origins firing but also reveals another mechanism of origin regulation through a Cdc7-dependent association between GINS and And-1 [123]. Finally, a very recent paper showed that ATR-activation domain of TopBP1 was required to suppress origin firing during the S Phase [124], supporting further the important role of ATR-Chk1 pathway in the regulation of origins activation.

3.2.2 MRC1/Claspin is a central regulator of origin firing under normal and stressed replication

During the normal DNA replication process of eukaryotes, Claspin/MRC1 is required for efficient fork progression [85,86,125,126]. Claspin interacts with various replication factors including ATR, Chk1, Cdc7 kinase, Cdc45, Tim, MCM4, MCM10, PCNA, DNA polymerases α, δ, ε, and And-1 [127–130], suggesting its role at the replication forks to link the helicase components to the replicative polymerases. A new role for Claspin has been described more lately in the initiation of DNA replication during normal S phase through the recruitment of CDC7 kinase that facilitates phosphorylation of MCM proteins [131]. Besides its checkpoint function, it has been recently discovered that the loss of MRC1 checkpoint activity leads to aberrant activation of late or dormant origins in the presence of replication inhibitors HU [132]. MRC1, with these two crucial functions, can be placed at the heart of origin firing regulation: one regulating late/dormant origins through its well-established checkpoint function and the other regulating early-firing origins through checkpoint-independent mechanism [132].

3.2.3 Fanconi Anemia proteins in regulation of dormant origins

The role of the Fanconi Anemia (FA) pathway in DNA repair has been highly studied and a clear model has emerged describing how FA proteins coordinate the convergence of multiple DNA repair pathways, including homologous recombination (HR) and translesion synthesis (TLS), for the repair of Interstrand Cross Link (ICL) [169–171]. However, treatment of cells with a low dose APH, robustly activates the FA pathway, indicating a role of the FA proteins during DNA replication [172].

FANCI has been recently shown to be involved in the regulation of dormant origin firing upon low replication stress [173]. In this study, the authors provide evidence that this occurs through an FA pathway-independent mechanism involving ATR phosphorylation of FANCI which is a negative
regulator of dormant origin firing. Interestingly, the loss of FANCD2 also resulted in reduced stalled forks in the presence of low-dose of HU, which is likely due to increased origin firing to help alleviate replication stress. The depletion of FANCD2 inhibits FANCI mono-ubiquitination and leads to an increase in the number of active origins suggesting that the FA core complex (ubiquitin ligase complex for FANCI and FANCD2 mono-ubiquitination) is probably not required for FANCI-mediated origin firing during replication stress. Interestingly, only the loss of FANCD2, and not members of the FA core complex, enhanced origin firing in a FANCI-dependent manner, suggesting that FANCD2 binding to FANCI may be inhibitory for origin firing. This supports the notion that an intermediary modification of FANCI preceding mono-ubiquitination could be the trigger for the modulation of its role in dormant origins activation.

More recently, FANCD2 has been shown to facilitate replication through repeat-rich genomic regions such as CFSs by ameliorating DNA:RNA hybrid accumulation and by influencing dormant origin firing, even during unperturbed replication [174]. In absence of FANCD2, dormant origins are activated at CFSs due to an increase of replication fork pausing. These results underline a role for FANCD2 in efficient replication origin firing. Because changes in origin usage can be attributed to changes in chromatin looping [55,175], this role of FANCD2 is possibly associated with changes of chromatin looping and/or with the histone chaperone activity of FANCD2 [176].

3.2.4 RIF1 orchestrate origins and replication timing

RIF1 (Rap1-interacting factor 1) was first identified as a telomeric chromatin interacting protein required for telomere length regulation in budding yeast via its interaction with Rap1 [133,134]. It was later demonstrated that S. cerevisiae RIF1 impede activation of the DNA damage checkpoint near telomeres [135,136] and affects telomere replication timing [137]. The RIF1 protein is evolutionarily conserved, but in higher eukaryotes, it has been shown to play non-telomeric roles, such as directing the DNA double-strand break repair pathway and DNA recombination [138–142].

More recently, studies have implicated the fission yeast and mammalian RIF1 in the regulation of DNA replication genome-wide. Hayano and colleagues (2012) [143] showed that fission yeast RIF1 selectively bind not only to telomeres, but also specific regions of the genome, and may regulate the choice and timing of origin firing throughout late replicating regions of chromosomes. Importantly, in addition to activation of dormant/late origins, some active, early-firing origins are suppressed in RIF1 deficient cells, indicating that RIF1 is not simply a repressor of origin activation, but rather a critical determinant of the genome-wide origin activation program in fission yeast.

In the paper of Yamazaki et al., (2012) [144] they observed that the depletion of human RIF1 results in specific loss of mid-S replication foci profiles, stimulation of initiation events in early-S-phase and changes in long-range replication timing domain structures. Analyses of replication timing showed that sequences normally replicating early are delayed, whereas that normally replicating late are advanced, suggesting that replication timing regulation is abrogated without RIF1. Another important finding is that RIF1 tightly binds to nuclear-insoluble structures at late mitosis to early G1 and regulates chromatin-loop sizes. Overall, their results indicate that RIF1 plays crucial roles in determining the replication timing domain structures in human cells through regulating higher-order chromatin architecture.

Very interestingly, in a paper from Kanoh et al., (2015) [145], they have identified a RIF1 consensus sequences in fission yeast that are G-quadruplex-like. These latest tend to be near dormant
origins and the binding of RIF1 on these sites would allow their repression. These results suggest that RIF1 recognizes and binds G-quadruplex-like structures generating local chromatin structures that may exert long-range suppressive effects on origin firing.

In a more recent paper from Hariga et al., (2017) [146], they observe that RIF1 depletion leads not only to an increase in origin spacing in normal S phase, but also to a reduction in the availability of dormant origins following replication stress. One consequence of limiting the number of available dormant origins is increased sensitivity to replication-inhibiting drugs, such as HU or APH. Overall, increasing set of data suggest a role for RIF1 in the regulation of dormant origin availability in response to replicative stress.

4. Dormant origins deficiency, genome stability and pathologies

4.1 MCM mutants and dormant origins in mice

Mcm2-7 genes are essential for DNA replication and homozygosity for a null allele of the respective Mcm genes causes embryonic lethality [147–149]. Only hypomorphic alleles such as Mcm4\textsuperscript{Chao3} and Mcm2\textsuperscript{IRES-CreERT2} can result in viable homozygous mice surviving until adulthood. Mcm4\textsuperscript{Chao3} encodes a Phe345Ile mutation, which reduces the efficiency of MCM2-7 assembly but does not lead to helicase activity defect in vitro [150]. Mcm2\textsuperscript{IRES-CreERT2} allele was engineered to express tamoxifen-inducible form of Cre recombinase (CreERT2) that is inserted into the 3’-UTR of the endogenous Mcm2 locus. This modification might be responsible for 65% reduction of Mcm2 expression compared to wild type cells [151].

Surprisingly, MEFs from Mcm4\textsuperscript{Chao3} mice have also a reduced MCM7 protein level in addition to Mcm4 [152]. Moreover, SV40-immortalized homozygous Mcm4\textsuperscript{Chao3} display less stable association of MCM2-7 at replication forks compared to wild type cells [153]. Finally, Mcm4\textsuperscript{Chao3}\textsuperscript{Chao3} MEFs exhibit about half reduction in chromatin bound MCM2-7 that causes a fewer ability to activate dormant origins in response to treatment with low dose of Aphidicolin (APH) [148,150].

Subsequently, it was reported that mice containing one-third of the normal MCM2 level succumbed to lymphomas at a very young age, and had diverse stem cell proliferation defects. These mice also had 27% reduced levels of MCM7 protein, and, even in the presence of hydroxyurea (HU), cells exhibited decreased replication origin usage due to less dormant origins availability; proved by DNA combing experiments [151,154].

Altogether, these two mouse models are close phenotypically: showing dormant origins deficiency due to reduced level of loaded MCM onto the chromatin. This loss of dormant origins results in an accumulation of stalled replication forks in unchallenged S phase. Furthermore, despite the activation of multiple DNA repair pathways, a significant fraction of stalled forks persists into M phase and interfere with chromosome segregation. Both phenotypes lead to improper chromosome segregation and premature tumorigenesis, with several differences in the latency of the disease development.

4.2 MCM mutants and dormant origins in stem/progenitor cells

Notwithstanding the fact that most Mcm2\textsuperscript{IRES-CreERT2} mice develop tumours and that this is generally the cause of their death, these mice showed a spectrum of additional hallmarks of ageing-related dysfunction. One potential explanation for these additional phenotypes is that reduced Mcm2
expression has a general effect on proliferating cells within multiple tissues. A recent study from Pruitt and colleagues [151], attempted to determine the effect of Mcm2 deficiency on somatic stem cells and proliferative progenitors. Even though they did not observe any effect on the rate at which proliferative progenitors cycle, they showed an approximately three-fold reduction in the level of neurogenesis within the Sub Ventricular Zone (SVZ) of Mcm2RES-CreERT2 mice brain. They also observed a reduced stem cell number in intestinal crypt and skeletal muscle with a modest increase in DNA damage.

Similarly, neural stem cells progenitors derived from Mcm4Chaos3/Chaos3 embryos show an increased number of γH2AX and 53BP1 foci with accumulation in G2/M, leading to a reduced ability to form neurospheres in vitro [155]. In Mcm4Chaos3/Chaos3 mice, the renewal of stem cells in the brain appears to be normal but the ability to differentiate in intermediate progenitors is highly reduced due to an increase of apoptotic cells in the sub-ventricular and intermediate zones.

These studies suggest that a full expression of MCM2-7 proteins is essential for stem/progenitor cells function by reducing the risk of replication associated genome instability. Several recent studies go in line with this idea. One first paper demonstrating that human embryonic stem cells, that have a remarkably short G1 phase, display a very fast MCM loading rate compared to differentiated cells, to reach similar total amount of loaded MCM at the G1/S transition [156]. A second paper showing that, in mouse strain with hypomorphic expression of the origin licensing factor MCM3, limiting origin licensing in vivo affects the functionality of hematopoietic stem cells and the differentiation of rapidly-dividing erythrocyte precursors. These results indicate that hematopoietic progenitors are particularly sensitive to replication stress, and full origin licensing ensures their correct differentiation and functionality. This is the first demonstration that the rate of MCM loading is crucial during organism development [149].

Intriguingly, aging hematopoietic stem cells suffer from replication stress even in wild type mice. This must be because old stem cells have reduced expression of MCM2-7 resulting in reduce amount of dormant origins and as a consequence more chromosome instability and cell cycle defects [157].

### 4.3 The consequence of limited licensing and firing in humans

A set of human patient with growth delay, natural killer cell deficiency, adrenal insufficiency and genome instability were shown to carry a mutation in Mcm4 gene resulting in a truncated form of this protein with disruption in its N-terminal serine-threonine-rich domain [158–160]. This truncated form of MCM4 does not impact MCM2-7 loading but patient’s SV40 fibroblasts exhibit a high level of chromosome breakage, defect in cell cycle progression and cells are sensitized to low dose of APH [159]. These findings indicate that the first 50 and 74 amino acids are not required for MCM complex formation and the loading of MCM onto chromatin, at least in dermal fibroblast cell lines. However, the higher rate of DNA breakage in patients’ leukocytes and dermal fibroblasts suggests that the N-terminal domain of MCM4 is involved in DNA replication and, specifically, in the maintenance of genome integrity during DNA replication. Further study need to be done to elucidate the mechanism by which normal MCM4 ensure genome maintenance but one possibility is the role of MCM4 phosphorylation in the checkpoint response knowing that although the eukaryotic N-terminal domain is non-essential, it is involved in protein kinase regulation of cell cycle progression [161].
Meier-Gorlin syndrome (MGS) is an autosomal recessive primordial dwarfism syndrome characterized by pre- and post-natal impaired growth. Although microcephaly is often evident, intellect is usually normal in this syndrome. Several studies have reported marked locus heterogeneity, and identified mutations in five separate genes from the pre-RC: Orc1, Orc4, Orc6, Cdt1 and Cdc6 [162,163]. Molecular and cellular phenotypes observed were impaired licensing, altered S phase progression and proliferation defects that partially overlap with MCM mutations except for chromosomal instability or an increased predisposition to cancer. Nonetheless, MGS mutations (in Orc1 and Orc6) can cause quite significant reduction in MCM loading and replication origin licensing [162,164,165]. It can also happen that some MGS individuals have an increased risk of cancer, but this has not become apparent in the clinical record.

The mice and human phenotypes caused by mutations in the licensing system underline the limited understanding of what happens to cells when the DNA replication programme is compromised. The threshold value for the limiting number of licensed origins that will activate the licensing checkpoint is still not known neither whether this varies between cell types or not.

5. Dormant origins activation, chromatin loops and cellular memory

5.1 Changes to chromatin loops correlate with dormant origin activation

In addition to their discovery on replication fork speed and dormant origin activation, Courbet et al., (2008) [54] also observed a strict correlation between replication fork speed during a given S phase and chromatin loop size in the next G1 phase. For that, they used the fluorescent DNA halo technique to estimate the average length of DNA loops in the G1/S transition nuclei combined with FISH. When cells are permeabilized with detergent and depleted of soluble proteins by extraction with high-salt buffers, supercoiled DNA loops unwind and form a halo around an insoluble scaffold that can be visualized by fluorescence staining [166]. This technique has been first essential to establish the link between chromatin loops and replicon size [167] and then to describe replicon remodelling events in Xenopus [168].

Another experiment with DNA halo technique indicates that cohesin determines the size of interphase chromatin loops and the absence of cohesin leads to an increase in chromatin loops due to a limited origin usage observed measuring interfork distances by DNA spreading [38]. These results imply that the presence of cohesin at origins modulates their activity, providing a novel link between the DNA replication and cohesion machineries, which is independent from the reported effect of cohesin acetylation on fork progression [169].

Chromatin loop sizes increase in RIF1-depleted cells, indicating that RIF1 is required for correct chromatin loop formation [144]. The strong association of RIF1 with nuclear-insoluble structures suggests a possibility that RIF1 may be a crucial factor for generating higher-order chromatin architecture including special organization of chromatin loops.

5.2 Long term adaptation of origin usage

Besides the direct correlation between origin activation and chromatin loops, Courbet et al., (2008) also proved that origins located near the sites of anchorage of chromatin loops are preferentially activated in the S phase of the following cell generation. This is a key observation proving that, in addition to the rapid response of origin activation, cells also respond to changes in
fork dynamics by adapting origin usage in the next cell cycles. It appears that cells can adapt to grow under condition of fork slowing by increasing the efficiency of some origins that are usually dormant in normal growth condition.

To explain mechanistically this phenomenon, we can assume a role of epigenetic marks that are known to be modified in response to developmental and environmental stimuli and to play an important role in cellular memory [170–172]. Moreover, many models in plants and other eukaryotes predict that strong positive feedback chromatin loops are necessary and sufficient for spreading of histone modifications and histone-based epigenetic long-term memory but this mechanism remains to be proven in human cells [173–176].

6. Conclusion

Dormant origins are now recognized as a major safeguard against under-replication allowing genome maintenance. Activation of dormant origins plays a central role in the rescue of stalled forks in the context of replicative stress, contributing to complete DNA replication. The functional interplay between dormant origins and other mechanisms (such as translesion synthesis and homology-mediated fork restart) is largely unknown even though some link with DNA damage checkpoint or Fanconi Anemia pathways is becoming more and more obvious. However, pathway choice between dormant origins and these mechanisms in response to fork-stalling still remains to be investigated.

The molecular details of how replication factories and replicon clusters are activated remain obscure, but knowing that factory activation is regulated by both CDKs and CHK1 might help to tackle this problem. RIF1 protein might be the most interesting actor in this molecular process knowing that it is present both at the replication fork and replication origins where it plays a role in DNA damage response as well as replication timing regulation. Perhaps most exciting is the prospect that the regulation of dormant origins could be different in cancer cells. Indeed, proteins of the MCM complex are often misregulated at the early stage of cancer [11,177,178] and a recent paper from Zimmerman et al. [141] show that tumour cells are more sensitive to replicative stress when they have a reduced origin licensing capacity. On the other hand, MCM hypomorphic mice show the potential importance of dormant origins, but it remains to be determined whether spontaneous cancers show similar defects and whether this information can be used to direct anti-cancer treatment more precisely.
Figure 3: Summary diagram showing the importance of dormant origin activation in response to replicative stress. During normal replication, only the principal origin is activated. If there is no replicative stress, this principal origin will also be activated in the next S phase. Under low replicative stress, adjacent or dormant origins fire to compensate for fork slowing and to allow the complete replication on time. Many proteins (ATR/Chk1, Mrc1/Claspin, FANCI/FANCD2, RIF1) are thought to be involved in the activation of dormant origins under replicative stress. RIF1 and Cohesin are two good candidates to explain this mechanism. Finally, when cells have a low origins reservoir or dormant origin deficiency, the addition of replicative stress leads inevitably to fork stalling, DNA breaks and genomic instability that give an open window for tumorigenesis.
Author Contributions: L.C. wrote the manuscript, JS.H. proofread the manuscript and V.B designed and proofread the review.

Funding: Work in JSH’s laboratory is supported by funding from the Institut National du Cancer (PLBIO 2016), the Agence Nationale de la Recherche (ANR PRC 2016), Labex TOUCAN and La Ligue contre le Cancer (Equipe labellisée 2017)

Acknowledgments:

Conflicts of Interest: No conflict

Abbreviations:

53BP1: p53 binding protein 1
APH: aphidicolin
ATM: ataxia telangiectasia mutated
ATR: ataxia telangiectasia and Rad3-related protein
BLM: bloom syndrome RecQ like helicase
BRCA2: breast cancer 2
BrdU: bromodeoxyuridine
Cdc45: Cell division cycle protein 45
Cdc6/7: cell division cycle 6/7
CDK: cyclin-dependent kinase
Cdt1: Chromatin licensing and DNA replication factor 1
CFS: common fragile site
Chk1: checkpoint kinase 1
Chk2: checkpoint kinase 2
CHO: Chinese hamster ovary
CldU: chlorodeoxyuridine
c-Myc: Myelocytomatosis
DDK: Dbf4-dependent kinase
DNA: deoxyribonucleic acid
dNTP: deoxyribonucleotides
ERFS: early-replicating fragile site
FANCI/D2: Fanconi Anemia Complementation Group I/D2
FISH: fluorescence in situ hybridization
GINS: go-ichi-ni-san
HR: homologous recombination
HRAS: Harvey Rat Sarcoma
HU: hydroxyurea
ICL: interstrand cross link
IdU: iododeoxyuridine
MCM: minichromosome maintenance
MGS: Meier-Glorin syndrome
MMEJ: micro-homology mediated end-joining
MRC1: Mannose Receptor C-Type 1
NHEJ: non-homologous end-joining
ODP: origin decision point
ORC: origin recognition complex
PCNA: proliferating cell nuclear antigen
Pre-IC: pre-initiation complex
Pre-RC: pre-recognition complex
RIFI: Rap1-interacting factor 1
RNA: ribonucleic acid
RO: replication origin
SNS: short nascent strand
SVZ: sub ventricular zone
TDP: timing decision point
TOPBP1: topoisomerase 2-binding protein 1
TSS: transcription start sites
UV: ultraviolet

References


