R-loops and initiation of DNA replication in human cells: a missing link?

Rodrigo Lombraña, Ricardo Almeida, Alba Álvarez and María Gómez*

Functional Organization of the Genome Group, Centro de Biología Molecular "Severo Ochoa", Consejo Superior de Investigaciones Científicas/Universidad Autónoma de Madrid, Madrid, Spain

The unanticipated widespread occurrence of stable hybrid DNA/RNA structures (R-loops) in human cells and the increasing evidence of their involvement in several human malignancies have invigorated the research on R-loop biology in recent years. Here we propose that physiological R-loop formation at CpG island promoters can contribute to DNA replication origin specification at these regions, the most efficient replication initiation sites in mammalian cells. Quite likely, this occurs by the strand-displacement reaction activating the formation of G-quadruplex structures that target the origin recognition complex (ORC) in the single-stranded conformation. In agreement with this, we found that R-loops co-localize with the ORC within the same CpG island region in a significant fraction of these efficient replication origins, precisely at the position displaying the highest density of G4 motifs. This scenario builds on the connection between transcription and replication in human cells and suggests that R-loop dysregulation at CpG island promoter-origins might contribute to the phenotype of DNA replication abnormalities and loss of genome integrity detected in cancer cells.

Keywords: R-loops, CpG islands, ORC, G-quadruplex, DNA replication origins

R-loops are three-stranded nucleic acid structures formed upon the hybridization of an RNA strand to a complementary DNA strand. This RNA/DNA hybrid displaces the second DNA strand into a looped out state, giving this class of structures their name. *In vivo*, R-loops can be generated by RNA polymerase II transcribing a C-rich DNA template such that a G-rich transcript is produced. Although the mechanism through which R-loops are generated is still unclear, the prevalent model postulates that the newly synthesized RNA strand, upon leaving the RNA exit channel of the traveling RNA polymerase complex, competes with the non-template DNA strand for re-annealing to the template DNA strand. Once formed, R-loops are stable, as RNA/DNA interactions are thermodynamically far more stable than the corresponding DNA/DNA duplexes (Roberts and Crothers, 1992).

R-loops were first detected *in vivo* at prokaryotic ORIs (Masukata et al., 1987; Baker and Kornberg, 1988; Masukata and Tomizawa, 1990; Carles-Kinch and Kreuzer, 1997), the mitochondrial origin of replication (Xu and Clayton, 1996), the immunoglobulin class-switch region in activated B cells (Yu et al., 2003), and in yeast cells mutant for mRNA metabolism genes (Huertas and Aguilera, 2003). More recently, genome-wide approaches to measure R-loops showed that these nucleic acid structures are widespread in the human genome, being prevalently formed at promoter 5′- and terminator 3′-end regions of several genes (Ginno et al., 2012, 2013). R-loops are involved in multiple cellular processes including transcription repression, transcriptional termination, DNA methylation and histone modifications, as well as DNA replication and immunoglobulin class switch recombination. Importantly, unprogrammed R-loop formation or R-loop dysregulation can promote

DNA damage and genome instability that may lead to human diseases, placing this nucleic acid structure at the center of very active research in recent years (see the excellent reviews, by Aguilera and García-Muse, 2012; Groh and Gromak, 2014; Skourti-Stathaki and Proudfoot, 2014). Curiously, the aspect of R-loop biology that has been overlooked in recent investigations is their role in DNA replication initiation, which was actually the first biological function ascribed for R-loops more than 30 years ago. In this article, we revisit this issue and propose that, in human cells, persistent R-loop formation can play a role in replication origin specification likely through exposing and activating replication signals that are functional only in the single-stranded conformation by the strand-displacement reaction.

R-loops and Initiation of DNA Replication

The earliest evidence for a role for R-loops in initiation of DNA replication came from studies in the late 1980's on *Escherichia coli* plasmid ColEI. In this system, a transcript initiated from an upstream promoter forms a persistent hybrid with the template strand within specific origin elements (Dasgupta et al., 1987; Masukata and Tomizawa, 1990). The hybridized RNA is cleaved by RNAse H and then serves as a primer for DNA synthesis by DNA polymerase I (Itoh and Tomizawa, 1980). In the absence of RNAase H, the persistent DNA-RNA hybrid indirectly activates subsequent DNA synthesis instead of providing a primer as it occurs in the presence of the enzyme (Masukata et al., 1987). In both situations, the DNA-RNA hybrid activates DNA synthesis by displacing the non-transcribed DNA strand, thus exposing potential recognition sites of a helicase/primase that can simultaneously drive the replication fork forward and synthesize primers for the lagging strand (Marians, 1992). Importantly, the formation of the persistent hybrid between the RNA and the template DNA is necessary for ColEI replication. In particular, the interaction between the dC stretch in the DNA template strand and the rG stretch in the RNA is essential for the formation of the stable R-loop (Masukata and Tomizawa, 1990). Interestingly, the efficiency of persistent hybrid formation depends on the rate of elongation of the transcript, suggesting that its success requires the formation of a particular DNA-RNA structure at a particular time during transcription (Masukata and Tomizawa, 1990).

Another long-known example of R-loop-mediated DNA synthesis occurs early during bacteriophage T4 infection (Mosig, 1987). Several putative origins of replication have been identified in T4 and the best characterized (*oriF* and *oriG*) consist of two components: a middle-mode promoter and a downstream DNA unwinding element (DUE; Carles-Kinch and Kreuzer, 1997). In a first step, transcription initiates from the promoter; in a second step, a persistent DNA-RNA hybrid is formed within the DUE region, providing the primer for leading-strand synthesis at the 3′ end generated either by RNA polymerase or by RNAse cleavage. Alternatively, the R-loop structure allows T4 primase to synthesize RNA primers on the single-stranded non-template strand (Belanger and Kreuzer, 1998). It is possible that either mechanism can be used to prime leading-strand DNA synthesis depending on protein availability as reported for plasmid ColEI replication (Dasgupta et al., 1987; Masukata et al., 1987). The finding that non-origin plasmids are efficiently replicated *in vitro* by the T4 replisome, providing they carry a preformed R-loop within the DUE region, strongly implies that the R-loop itself supplies the signal for replisome assembly (Kreuzer and Brister, 2010).

Mitochondrial DNA replication at the leading-strand origin is also coupled to transcription through the formation of an R-loop (Chang and Clayton, 1985; Chang et al., 1985). The critical features of this origin are conserved from *Saccharomyces cerevisiae* to humans and include a promoter and a downstream short GC-rich cluster. *In vitro* transcription studies demonstrated that a short rG-dC sequence is the only necessary and sufficient *cis* element required for stable hybrid formation, although its efficiency depends on transcription by mtRNA polymerase and close proximity of the site of transcription initiation to the GC-rich cluster (Xu and Clayton, 1995). Once made, the RNA of the R-loop can serve as an effective primer for elongation by POLG, the mtDNA polymerase. These findings are reminiscent of those described for ColEI replication (Masukata et al., 1987), indicating that stable R-loop formation depends on C-rich clusters on template DNA. Additionally, the highly conserved nature of this essential template sequence element suggests that its role in stabilizing R-loop formation is ancient and likely pervasive in mitochondrial genomes (Xu and Clayton, 1995). More recently, an unorthodox mechanism of mtDNA replication involving long stretches of preformed RNA hybridized to the template-lagging strand has been described (Reyes et al., 2013). These long tracts of RNA are not products of on-going transcription and are therefore not directly related to the R-loops discussed here.

Replication of the *E coli* chromosome can also initiate by a mechanism involving R-loops in RNAse HI knock-out cells (Asai and Kogoma, 1994), and likely in wild-type cells under certain specific conditions such as entry into stationary phase or replication after DNA damage (Hong et al., 1996; Camps and Loeb, 2005; Wimberly et al., 2013). During this alternative mode of replication, named constitutive stable DNA replication (cSDR), RNAse HI-deficient cells initiate oriC-independent replication from multiple chromosomal sites termed *oriKs*. This results in global alterations of replication fork migration patterns, frequently in the opposite direction to normally initiated *oriC* synthesis and converging replication forks meeting in unusual places around the chromosome (Maduike et al., 2014). Notably, *E. coli* cells with a reduced capacity to remove R-loops display SOS constitutive phenotypes and an increase in hotspots for homologous recombination at the chromosomal terminus region flanked by the *Ter* sites (Nishitani et al., 1993; Horiuchi et al., 1994; Hong et al., 1995). It should be noted, however, that R-loop formation during cSDR occurs by transcript invasion, in contrast to the co-transcriptional R-loops generated in ColE1 replication (Kogoma, 1997).

Replication Origin Specification in Eukaryotes

In eukaryotic genomes, DNA synthesis initiates from multiple replication origins. Accurate duplication of the genetic material depends on a reliable mechanism that ensures that any given origin fires at most once per cell cycle by restricting "licensing" to late mitosis and "activation" to S phase. At the anaphase to

telophase transition the origin recognition complex (ORC) is recruited to replication origins; then, licensing rapidly occurs through the loading of the double hexameric minichromosome maintenance (MCM) complex together with other proteins such as Cdc6 and Cdt1 to form the pre-replication complex (pre-RC). Activation occurs when the pre-RC is converted to the pre-initiation complex (pre-IC) by the assembly of several replication factors facilitating the switch of the MCM complex to the active CMG (Cdc45-MCM-GINS) helicase during S-phase. The formation of the pre-IC depends on two protein kinases, cyclin-dependent kinase (CDK) and Dbf4-dependent kinase (DDK) that would ultimately trigger the unwinding of the origin DNA and the establishment of bidirectional replication forks (Siddiqui et al., 2013; Tanaka and Araki, 2013; and references therein). Replication initiation has been recently reconstituted *in vitro* with purified replication proteins, thus defining the initiation factors required for regulated eukaryotic DNA replication (Yeeles et al., 2015). Since there are far many more licensed ORIs than activated in each S phase, this is interpreted as a fail-safe mechanism used by cells to cope with replication stress (Ge et al., 2007; Ibarra et al., 2008). While the role of ORC in the pre-RC assembly at replication origins is conserved among various eukaryotes, the mechanism of origin recognition by ORC seems different across eukaryotic species (Bell et al., 2013). For example, ORC can be targeted to replication initiation sites by sequence-specific interaction as for *S. cerevisiae* (Marahrens and Stillman, 1992), or by non-specific binding to AT-rich sequences as for *Schizosaccharomyces pombe* or Drosophila (Kong and DePamphilis, 2001; Vashee et al., 2003). In addition, interactions can occur through sequence-specific binding proteins as for Drosophila and at certain loci in rat and human cells (Beall et al., 2002; Minami et al., 2006; Thomae et al., 2008), or through RNA-binding as in the case of 26T RNA during rDNA amplification in *Tetrahymena thermophila* (Mohammad et al., 2007) or during Epstein-Barr virus replication (Norseen et al., 2008).
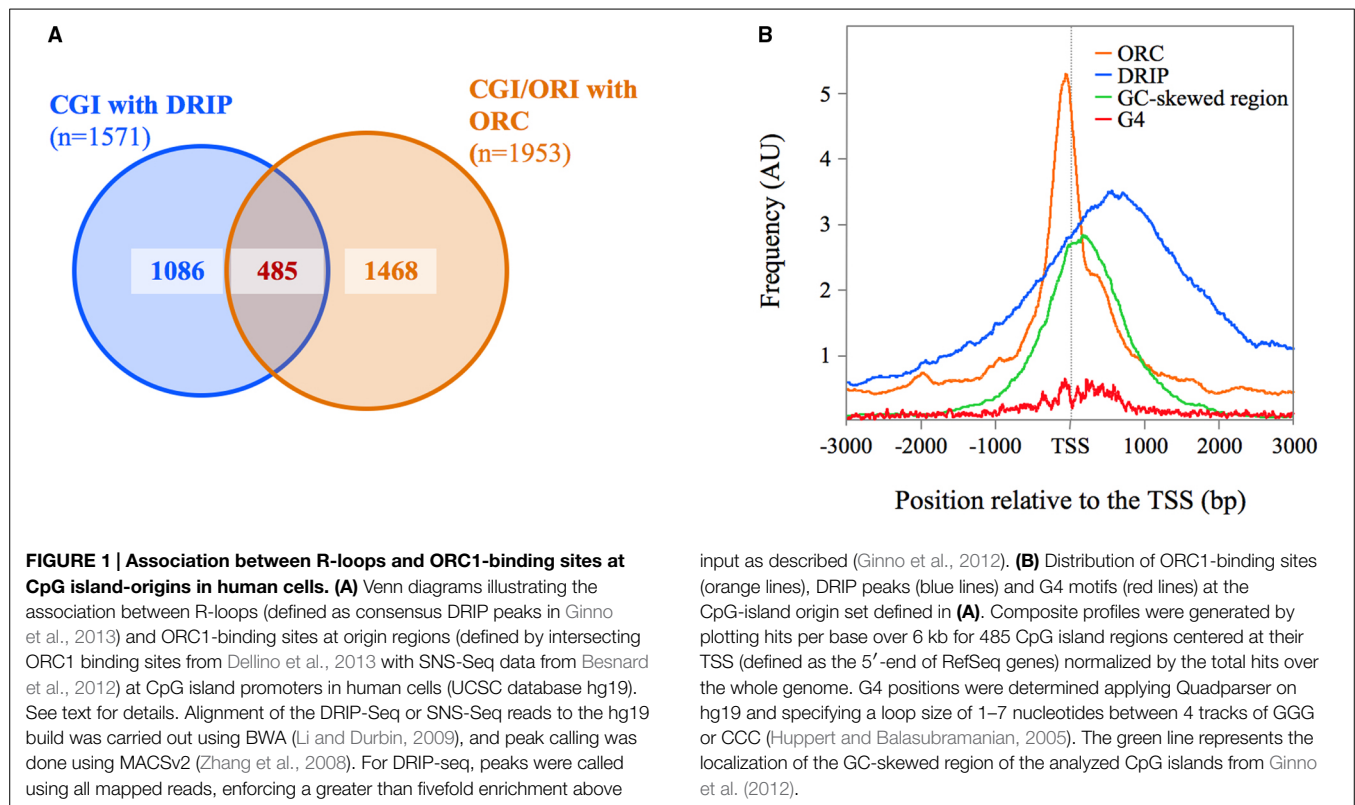
Although no DNA replication activity similar to *E. coli* cSDR has been found in eukaryotes, transcription activity is strongly associated with initiation of DNA replication in mammalian systems. Specifically, genome-wide maps of replication origins in mouse and human cells showed that the most efficiently activated and more conserved origins across all cell types examined are those associated with CpG island promoters (Sequeira-Mendes et al., 2009; Cayrou et al., 2011; Besnard et al., 2012; Picard et al., 2014). Intriguingly, these genomic analyses revealed that a G-rich repeat element with the potential to form G-quadruplex structures (G4) was present in most of the replication origins in mouse and human cells (Besnard et al., 2012; Cayrou et al., 2012). The role of G4 structures in origin specification is not clear, however, recent evidence demonstrated that some G4 motifs could stimulate replication initiation (Valton et al., 2014). An interesting possibility is that G4 structures could mediate ORC recruitment to initiation sites. This notion is supported by *in vitro* binding assays demonstrating that the human ORC has affinity for G4 motifs through a specific domain in the ORC1 protein. Notably, ORC1 affinity for G4 motifs is almost negligible on double-stranded DNA but it is highly increased when G4 motifs are present on RNA or single-stranded DNA (Hoshina et al., 2013). As many

CpG island promoters are prone to R-loop formation upon transcription (Ginno et al., 2012, 2013), this could provide a possible mechanism by which single-stranded G4 structures are formed within the origin region early during the cell cycle generating a potential substrate for ORC1 binding by the end of mitosis. To test this possibility, we analyzed the co-occurrence of ORC1 binding sites and R-loops at CpG island-origin regions in human cells.

Association of ORC1 Binding Sites and R-loops in Human Cells

As mentioned above, CpG island origins are the most conserved and highly-efficient origins in the mammalian genome, as determined by the increased levels of associated short nascent strands (SNS) detected either by array hybridization signals (Sequeira-Mendes et al., 2009; Cayrou et al., 2011) or by sequencing read depth relative to the length of the origin (Besnard et al., 2012; Picard et al., 2014). Consistent with their higher firing activity, 37% of the ORC1 binding sites identified genome-wide in HeLa cells were associated with CpG island promoters, and their corresponding ORC1-ChIP signal was significantly stronger (Dellino et al., 2013). Nevertheless, the large majority of the non-CpG island origins identified by SNS-Seq do not co-localize with ORC sites (Dellino et al., 2013; Picard et al., 2014). We therefore selected for our study the subset of CpG island promoters enriched in SNS (Besnard et al., 2012) that were positive for ORC1 binding (Dellino et al., 2013) in datasets derived from the same cell line. This set of CpG island-origins (1,661 from a total of 9,864 TSS-associated CpG island promoters identified at the UCSC database-hg19), although very restrictive, confidently comprises *bona-fide* highly efficient constitutive replication origins. Notably, this core of CpG island-origins display the highest density of G4 motifs (Picard et al., 2014), supporting the notion that G4 motifs can play a role in the control of origin selection in the human genome.

CpG island promoters, regardless of their enrichment in CpG dinucleotides, remain unmethylated in normal tissues (Illingworth and Bird, 2009). The majority of these unmethylated CpG island promoters show significant strand asymmetry in the distribution of guanines and cytosines, a property known as GC skew (Ginno et al., 2012, 2013). The Chédin lab found that nearly 75% of CpG island promoters (35% of al human promoters) displayed a positive GC skew, signifying that the non-template strand for transcription has an excess of G over C residues (Ginno et al., 2012). To test whether GC skew confers the ability to form R-loops upon transcription, the authors performed a genome-wide identification of R-loop by DRIP-Seq (DNA-RNA immunoprecipitation with the R-loop-specific S9.6 antibody coupled to deep-sequencing). A total of 4,181 DRIP-Seq peaks were identified in two complementary assays of DNA fragmentation, from which nearly 40% ($n = 1571$) mapped to GC-skewed CpG island promoters (Ginno et al., 2013). We chose this stringent DRIP peak dataset to study the association between R-loops and ORC1 binding sites. We found that 30% ($n = 485$) of the CpG islands displaying R-loops showed ORC1 binding just upstream of their TSS, compared with the 6% expected by chance (**Figure 1A**). ORC1 and DRIP signals overlap with each other precisely at the 1 kb GC-skewed

FIGURE 1 | Association between R-loops and ORC1-binding sites at CpG island-origins in human cells. (A) Venn diagrams illustrating the association between R-loops (defined as consensus DRIP peaks in Ginno et al., 2013) and ORC1-binding sites at origin regions (defined by intersecting ORC1 binding sites from Dellino et al., 2013 with SNS-Seq data from Besnard et al., 2012) at CpG island promoters in human cells (UCSC database hg19). See text for details. Alignment of the DRIP-Seq or SNS-Seq reads to the hg19 build was carried out using BWA (Li and Durbin, 2009), and peak calling was done using MACSv2 (Zhang et al., 2008). For DRIP-seq, peaks were called using all mapped reads, enforcing a greater than fivefold enrichment above input as described (Ginno et al., 2012). **(B)** Distribution of ORC1-binding sites (orange lines), DRIP peaks (blue lines) and G4 motifs (red lines) at the CpG-island origin set defined in **(A)**. Composite profiles were generated by plotting hits per base over 6 kb for 485 CpG island regions centered at their TSS (defined as the 5′-end of RefSeq genes) normalized by the total hits over the whole genome. G4 positions were determined applying Quadparser on hg19 and specifying a loop size of 1–7 nucleotides between 4 tracks of GGG or CCC (Huppert and Balasubramanian, 2005). The green line represents the localization of the GC-skewed region of the analyzed CpG islands from Ginno et al. (2012).

footprint defined for this set of CpG islands (Ginno et al., 2012; **Figure 1B**). It should be noted that the exact localization of the R-loops could not be extrapolated from the DRIP profiles as the resolution of the DRIP-Seq depends on the distribution of the restriction sites used to fragment the genome at each of the CpG islands analyzed.

These overlaps likely represent a great underestimate giving the fact that DRIP profiling experiments and origin mapping experiments were performed in different cell types (Ntera2 cells *versus* Hela cells), as well as the extremely stringent criteria used in selecting the significant regions from the various datasets for the analysis. Indeed, visual inspection of the sequencing reads on the UCSC genomic browser showed that many of the CpG island-origins defined by SNS enrichment displayed an ORC1 signal just below the threshold of significance. Similarly, several of the ORC1 positive CpG islands co-localize with R-loops defined in one set of DRIP data, but not in the other, due to the non-overlapping distribution of the cleavage sites of the restriction enzymes used to fractionate the genome in the two experiments analyzed. Nevertheless, these analyses show that R-loops co-localize with ORC1 binding sites within the same CpG island region at a fraction of the most efficient replication origins in human cells. More importantly, G4-motifs enrichment parallels the ORC1 signal at this CpG island-origin set (**Figure 1B**), consistent with the idea that G4 structures can mediate ORC recruitment to these specific sites.

By which mechanism could R-loops mediate ORC recruitment? The simplest possibility is that R-loop formation facilitates the generation of G4 structures on the displaced single-stranded non-template DNA strand that can target the ORC. Another possibility is that the ORC could be directly tethered to G4 structures formed on the RNA component of the hybrid. Indeed, EBNA1-ORC binding during EBV replication occurs through G-rich RNA and this binding is disrupted by G4-interacting drugs (Norseen et al., 2009). Finally, it is also possible that the ORC can bind hybrid G4 structures formed between the G-rich RNA and the G-rich displaced DNA strand generated by the R-loop. These hybrid G4 structures occur at the DNA replication leading-strand origin (OriH) in mammalian mitochondria and seem to regulate transcription termination at the replication-priming site (Wanrooij et al., 2012; Zheng et al., 2014). Any of these possible scenarios fulfill the affinity requirements described for ORC1 in *in vitro* assays (Hoshina et al., 2013) and would imply that G4 structures should persist through mitosis to mediate ORC recruitment. G4 structures formed on the displaced DNA strand could, in turn, stabilize the R-loops (Aguilera and García-Muse, 2012) and possibly inhibit nucleosome assembly at those sites (Wong and Huppert, 2009). Interestingly, we recently reported that ORC1 binding sites at efficient CpG island origins in human cells occupy the position marked by unstable nucleosome particles composed by H3.3/H2A.Z double-histone variants (Lombraña et al., 2013). Altogether, this suggests that stable R-loop formation at CpG island promoters could contribute to the generation of a permissive environment for ORC recruitment by, on one hand, exposing single-stranded DNA or RNA G4 structures and, on the other hand, by facilitating the assembly of labile nucleosome particles at those sites, thus preventing these regions from being occupied by adjacent stable nucleosomes or non-specific factors.

It is worth mentioning that the group of CpG island promoter-origins analyzed here mainly comprises strong promoters driving high transcriptional outputs mapping on gene-poor chromosomes (Ginno et al., 2013). As gene-poor genomic regions tend to be depleted from replication origins (Besnard et al., 2012; Picard et al., 2014), it is tempting to speculate that R-loop-mediated ORC recruitment, or G1-formed R-loops, could be one of the multiple factors influencing the choice of origins to be fired to initiate DNA synthesis during S-phase (Renard-Guillet et al., 2014). This mechanism could be especially relevant at this subset of CpG islands as a means to increase the probability of firing within genomic environments otherwise scarce in replication initiation sites. Indeed, replication origin paucity has been proposed as a major cause of the increased instability observed in common fragile sites in human cells (Letessier et al., 2011; Ozeri-Galai et al., 2011). Another interesting consideration is that CpG island firing activity during S phase generates short re-replicated DNA fragments (Gómez and Antequera, 2008) precisely derived from the position occupied by labile nucleosome particles where the ORC binds (Lombraña et al., 2013). Whether this phenomenon is related to R-loop formation or whether R-loop dysregulation can lead to extensive overreplication and genomic instability awaits elucidation.

The above scenario is consistent with the view that DNA replication origins in mammals are not unique entities with defined properties (Antequera, 2004; Gilbert, 2010; Méchali, 2010; Sequeira-Mendes and Gómez, 2012). On the contrary, increasing evidence suggests that the origin structure consists of redundant binding sites made accessible to the ORC by local chromatin conformation associated with gene transcription, including R-loop formation. This opportunistic coupling of DNA replication initiation to transcription presents the advantage that it links DNA synthesis to cellular physiology, enhancing the robustness of the replication program and, at the same time, allowing faster adaptation to environmental or developmental changes (Sequeira-Mendes and Gómez, 2012).

Conclusion

The proposal that R-loop formation at CpG islands can contribute to replication origin specification in human cells by exposing single-stranded DNA or RNA G-quadruplex structures strengthens the view that sequence-driven DNA structures may represent a new layer of regulatory information. Given the increasing evidence that abnormal R-loop accumulation can compromise genome stability, understanding how cells prevent the negative effects of R-loops yet allowing their positive effects is a challenge for the years to come. In the particular case discussed here, failure in the strict control of replication origin activation at CpG islands *via* R-loop dysregulation can lead to aberrant DNA replication, one of the hallmarks of cancer cells.

Acknowledgments

References

Aguilera, A., and García-Muse, T. (2012). R-loops: from transcription byproducts to threats to genome stability. *Mol. Cell* 46, 115–124. doi: 10.1016/j.molcel.2012.04.009

Antequera, F. (2004). Genomic specification and epigenetic regulation of eukaryotic DNA replication origins. *EMBO J.* 23, 4365–4370. doi: 10.1038/sj.emboj.7600450

Asai, T., and Kogoma, T. (1994). D-loops and R-loops: alternative mechanisms for the initiation of DNA replication in Eschherichia coli. *J. Bacteriol.* 176, 1807–1812.

Baker, T. A., and Kornberg, A. (1988). Transcriptional activation of initiation of replication from the *E. coli* chromosomal origin: an RNA-DNA hybrid near oriC. *Cell* 55, 113–123. doi: 10.1016/0092-8674(88)90014-1

Beall, E. L., Manak, J. R., Zhou, S., Bell, M., Lipsick, J. S., and Botcham, M. R. (2002). Role for a *Drosophila* Myb-containing protein complex in site-specific DNA replication. *Nature* 420, 833–837. doi: 10.1038/nature01228

Belanger, K. G., and Kreuzer, K. N. (1998). Bacteriophage T4 initiates bidirectional DNA replication through a two-step process. *Mol. Cell* 2, 693–701. doi: 10.1016/S1097-2765(00)80167-7

Bell, S. D., Méchali, M., and DePamphilis, M. L. (2013). *DNA Replication*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Besnard, E., Babled, A., Lapasset, L., Milhavet, O., Parrinello, H., Dantec, C., et al. (2012). Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat. Struct. Mol. Biol.* 19, 837–844. doi: 10.1038/nsmb.2339

Camps, M., and Loeb, L. A. (2005). Critical role of R-loops in processing replication blocks. *Front. Biosci.* 10:689–698. doi: 10.2741/1564

Carles-Kinch, K., and Kreuzer, K. N. (1997). RNA-DNA hybrid formation at a bacteriophage T4 replication origin. *J. Mol. Biol.* 266, 915–926. doi: 10.1006/jmbi.1996.0844

Cayrou, C., Coulombe, P., Puy, A., Rialle, S., Kaplan, N., Segal, E., et al. (2012). New insights into replication origin characteristics in metazoans. *Cell Cycle* 11, 658–667. doi: 10.4161/cc.11.4.19097

Cayrou, C., Coulombe, P., Vigneron, A., Stanojcic, S., Ganier, O., Peiffer, I., et al. (2011). Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res.* 21, 1438–1449. doi: 10.1101/gr.121830.111

Chang, D. D., and Clayton, D. A. (1985). Priming of human mitochondrial DNA replication occurs at the light-strand promoter. *Proc. Natl. Acad. Sci. U.S.A.* 82, 351–355. doi: 10.1073/pnas.82.2.351

Chang, D. D., Hauswirth, W. W., and Clayton, D. A. (1985). Replication priming and transcription initiate from precisely the same site in mouse mitochondrial DNA. *EMBO J.* 4, 1559–1567.

Dasgupta, S., Masukata, H., and Tomizawa, J. (1987). Multiple mechanisms for initiation of ColE1 DNA replication: DNA synthesis in the presence and absence of ribonuclease H. *Cell* 51, 1113–1122. doi: 10.1016/0092-8674(87)90597-6

Dellino, G. I., Cittaro, D., Piccioni, R., Luzi, L., Banfi, S., Segalla, S., et al. (2013). Genome-wide mapping of human DNA replication origins: levels of transcription at ORC1 sites regulate origin selection and replication timing. *Genome Res.* 23, 1–11. doi: 10.1101/gr.142331.112

Ge, X. Q., Jackson, D. A., and Blow, J. J. (2007). Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress. *Genes Dev.* 21, 3331–3341. doi: 10.1101/gad.457807

Gilbert, D. M. (2010). Evaluating genome-scale approaches to eukaryotic DNA replication. *Nat. Rev. Genet.* 11, 673–684. doi: 10.1038/nrg2830

Ginno, P. A., Lim, Y. W., Lott, P. L., Korf, I., and Chedin, F. (2013). GC skew at the 5′and 3′ends of human cells links R-loop formation to epigenetic regulation and transcription termination. *Genome Res.* 23, 1590–1600. doi: 10.1101/gr.158436.113

Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I., and Chedin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* 45, 814–825. doi: 10.1016/j.molcel.2012.01.017

Gómez, M., and Antequera, F. (2008). Overreplication of short DNA regions during S phase in human cells. *Genes Dev.* 22, 375–385. doi: 10.1101/gad.445608

Groh, M., and Gromak, N. (2014). Out of balance: R-loops in human disease. *PLoS Genet.* 10:e1004630. doi: 10.1371/journal.pgen.1004630

Hong, X., Cadwell, G. W., and Kogoma, T. (1995). *Escherichia coli* RecG and RecA proteins in R-loop formation. *EMBO J.* 14, 2385–2392.

Hong, X., Cadwell, G. W., and Kogoma, T. (1996). Activation of stable DNA replication in rapidly growing *Escherichia coli* at the time of entry to stationary phase. *Mol. Microbio.* 21, 953–961. doi: 10.1046/j.1365-2958.1996.591419.x

Horiuchi, T., Fujimura, Y., Nishitani, H., Kobayashi, T., and Hidaka, M. (1994). The DNA replication fork blocked at the Ter site may be an entrance for the RecBCD enzyme into duplex DNA. *J. Bacteriol.* 176, 4656–4663.

Hoshina, S., Yura, K., Teranishi, H., Kiyasu, N., Tominaga, A., Kadoma, H., et al. (2013). Human origin recognition complex binds preferentially to G-quadruplex-preferable RNA and single-stranded DNA. *J. Biol. Chem.* 288, 30161–30171. doi: 10.1074/jbc.M113.492504

Huertas, P., and Aguilera, A. (2003). Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol. Cell* 12, 711–721. doi: 10.1016/j.molcel.2003.08.010

Huppert, J. L., and Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. *Nucl. Acids Res.* 33, 2908–2916. doi: 10.1093/nar/gki609

Ibarra, A., Schwob, E., and Méndez, J. (2008). Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication. *Proc. Natl. Acad. Sci. U.S.A.* 105, 8956–8961. doi: 10.1073/pnas.0803978105

Illingworth, R. S., and Bird, A. P. (2009). CpG islands: a rough guide. *FEBS Lett.* 583, 1713–1720. doi: 10.1016/j.febslet.2009.04.012

Itoh, T., and Tomizawa, J. (1980). Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. *Proc. Natl. Acad. Sci. U.S.A.* 77, 2450–2454. doi: 10.1073/pnas.77.5.2450

Kogoma, T. (1997). Stable DNA replication: interplay between DNA replication, homologous recombination and transcription. *Microbiol. Mol. Biol. Rev.* 61, 212–238.

Kong, D., and DePamphilis, M. L. (2001). Site-specific DNA binding of the *Schizosaccharomyces pombe* origin recognition complex is determined by the Orc4 subunit. *Mol. Cell. Biol.* 21, 8095–8103. doi: 10.1128/MCB.21.23.8095-8103.2001

Kreuzer, K. N., and Brister, J. R. (2010). Initiation of bacteriophage T4 DNA replication and replication fork dynamics: a review in the Virology Journal series on bacteriophage T4 and its relatives. *Virol. J.* 7, 358–374. doi: 10.1186/1743-422X-7-358

Letessier, A., Millot, G. A., Koundrioukoff, S., Lachages, A. M., Vogt, N., Hansen, R. S., et al. (2011). Cell-type specific replication initiation programs set fragility of the FRA3B fragile site. *Nature* 470, 120–123. doi: 10.1038/nature09745

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Lombraña, R., Almeida, R., Revuelta, I., Madeira, S., Herranz, G., Saiz, N., et al. (2013). High-resolution analysis of DNA synthesis start sites and nucleosome architecture at efficient mammalian replication origins. *EMBO J.* 32, 2631–2644. doi: 10.1038/emboj.2013.195

Maduike, N. Z., Tehranchi, A. K., Wang, J. D., and Kreuzer, K. N. (2014). Replication of the escherichia coli chromosome in RNAse HI-deficient cells: multiple initiation regions and fork dynamics. *Mol. Microbiol.* 91, 39–56. doi: 10.1111/mmi.12440

Marahrens, Y., and Stillman, B. (1992). A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science* 255, 817–823. doi: 10.1126/science.1536007

Marians, K. J. (1992). Prokaryotic DNA replication. *Annu. Rev. Biochem.* 61, 673–719. doi: 10.1146/annurev.bi.61.070192.003325

Masukata, H., Dasgupta, S., and Tomizawa, J. (1987). Transcriptional activation of ColE1 DNA synthesis by displacement of the nontranscribed strand. *Cell* 51, 1123–1130. doi: 10.1016/0092-8674(87)90598-8

Masukata, H., and Tomizawa, J. (1990). A mechanism of formation of a persistent hybrid between elongating RNA and template DNA. *Cell* 62, 331–338. doi: 10.1016/0092-8674(90)90370-T

Méchali, M. (2010). Eukaryotic DNA replication origins: many choices for appropriate answers. *Nat. Rev. Mol. Cell Biol.* 11, 728–738. doi: 10.1038/nrm2976

Minami, H., Takahashi, J., Suto, A., Saitoh, Y., and Tsutsumi, K. (2006). Binding of AIF-C, an Orc1-binding transcriptional regulator, enhances replicator activity of the rat aldolase B origin. *Mol. Cell. Biol.* 16, 8770–8780. doi: 10.1128/MCB.00949-06

Mohammad, M. M., Donti, T. R., Yakisich, J. S., Smith, A. G., and Kapler, G. M. (2007). Tetrahymena ORC contains a ribosomal RNA fragment that participates in rDNA origin recognition. *EMBO J.* 26, 5048–5060. doi: 10.1038/sj.emboj.7601919

Mosig, G. (1987). The essential role of recombination in phage T4 growth. *Annu. Rev. Genet.* 21, 347–371. doi: 10.1146/annurev.ge.21.120187.002023

Nishitani, H., Hidaka, M., and Horiuchi, T. (1993). Specific chromosomal sites enhancing homologous recombination in *Escherichia coli* mutants defective in RNase H. *Mol. Gen. Genet.* 240, 307–314.

Norseen, J., Jonson, F. B., and Lieberman, P. M. (2009). Role for G-quadruplex RNA binding by Epstein-barr virus nuclear antigen 1 in DNA replication and metaphase chromosome attachment. *J. Virol.* 83, 10336–10346. doi: 10.1128/JVI.00747-09

Norseen, J., Thomae, A., Sridharan, V., Aiyar, A., Schepers, A., and Lieberman, P. M. (2008). RNA-dependent recruitment of the origin recognition complex. *EMBO J.* 27, 3024–3035. doi: 10.1038/emboj.2008.221

Ozeri-Galai, E., Lebofsky, R., Rahat, A., Bester, A. C., Bensimon, A., and Kerem, B. (2011). Failure of origin activation in response to fork stalling leads to chromosomal instability at fragile sites. *Mol. Cell* 43, 122–131. doi: 10.1016/j.molcel.2011.05.019

Picard, F., Cadoret, J. C., Audit, B., Arneodo, A., Alberti, A., Battail, C., et al. (2014). The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells. *PLoS Genet.* 10:e1004282. doi: 10.1371/journal.pgen.1004282

Renard-Guillet, C., Kanoh, Y., Shirahige, K., and Masai, H. (2014). Temporal and spatial regulation of eukaryotic DNA replication: from regulated initiation to genome-scale timing program. *Semin. Cell Dev. Biol.* 30, 110–120. doi: 10.1016/j.semcdb.2014.04.014

Reyes, A., Kazak, L., Wood, S. R., Yasukawa, T., Jacobs, H. T., and Holt, I. J. (2013). Mitochondrial DNA replication proceeds via a bootlace mechanism involving the incorporation of processed transcripts. *Nucl. Acids. Res.* 41, 5837–5850. doi: 10.1093/nar/gkt196

Roberts, R. W., and Crothers, D. M. (1992). Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. *Science* 258, 1463–1466. doi: 10.1126/science.1279808

Sequeira-Mendes, J., Diaz-Uriarte, R., Apedaile, A., Huntley, D., Brockdorff, N., and Gómez, M. (2009). Transcription initiation activity sets replication origin efficiency in mammalian cells. *PLoS Genet.* 5:e1000446. doi: 10.1371/journal.pgen.1000446

Sequeira-Mendes, J., and Gómez, M. (2012). On the opportunistic nature of transcription and replication initiation in the metazoan genome. *Bioessays* 34, 119–125. doi: 10.1002/bies.201100126

Siddiqui, K., On, K. F., and Diffley, J. F. (2013). Regulating DNA replication in eukarya. *Cold Spring Harb. Perspect. Biol.* 5, a012930. doi: 10.1101/cshperspect.a012930

Skourti-Stathaki, K., and Proudfoot, N. J. (2014). A double-edged sword: R-loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev.* 28, 1384–1396. doi: 10.1101/gad.242990.114

Tanaka, S., and Araki, H. (2013). Helicase activation and establishment of replication forks at chromosomal origins of replication. *Cold Spring Harb. Perspect. Biol.* 5, a010371. doi: 10.1101/cshperspect.a010371

Thomae, A. W., Pich, D., Brocher, J., Spindler, M. P., Berens, C., Hock, R., et al. (2008). Interaction between HMGA1a and the origin recognition complex creates site-specific replication origins. *Proc. Natl. Acad. Sci. U.S.A.* 105, 1692–1697. doi: 10.1073/pnas.0707260105

Valton, A. L., Hassan-Zadeh, V., Lema, I., Boggetto, N., Alberti, P., Saintomé, C., et al. (2014). G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *EMBO J.* 33, 732–746. doi: 10.1002/embj.201387506

Vashee, S., Cvetic, C., Lu, W., Simancek, P., Kelly, T. J., and Walter, J. C. (2003). Sequence-independent DNA binding and replication initiation by the human origin recognition complex. *Genes Dev.* 17, 1894–1908. doi: 10.1101/gad. 1084203

Wanrooij, P. H., Uhler, J. P., Shi, Y., Westerlund, F., Falkenberg, M., and Gustafsson, C. M. (2012). A hybrid G-quadruplex structure formed between RNA and DNA explains the extraordinary stability of the mitochondrial R-loop. *Nucl. Acids Res.* 40, 10334–10344. doi: 10.1093/nar/gks802

Wimberly, H., Shee, C., Thornton, P. C., Sivaramakrishnan, P., Rosenberg, S. M., and Hastings, P. J. (2013). R-loops and nicks initiate DNA breakage and genome instability in non-growing *Escherichia coli. Nat. Commun.* 4, 2115. doi: 10.1038/ncomms3115

Wong, H. M., and Huppert, J. L. (2009). Stable G-quadruplexes are found outside nucleosome-bound regions. *Mol. Biosyst.* 5, 1713–1719. doi: 10.1039/b905848f

Xu, B., and Clayton, D. A. (1995). A persistent RNA-DNA hybrid is formed during transcription at a phylogenetically conserved mitochondrial DNA sequence. *Mol. Cell. Biol.* 15, 580–589.

Xu, B., and Clayton, D. A. (1996). RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: an implication for RNA-DNA hybrids serving as primers. *EMBO J.* 15, 3135–3143.

Yeeles, J. T., Deegan, T. D., Janska, A., and Diffley, J. F. (2015). Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature* 519, 431–435. doi: 10.1038/nature14285

Yu, K., Chedin, F., Hsieh, C. L., Wilson, T. E., and Lieber, M. R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B-cells. *Nat. Immunol.* 4, 442–451. doi: 10.1038/ni919

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137. doi: 10.1186/gb-2008-9-9-r137

Zheng, K., Wu, R., He, Y., Xiao, S., Zhang, J., Liu, J., et al. (2014). A competitive formation of RNA:DNA hybrid G-quadruplex is responsible to the mitochondrial transcription termination at the DNA replication priming site. *Nucl. Acids Res.* 42, 10832–10844. doi: 10.1093/nar/gku764