ORChestrating the human DNA replication program

David M. MacAlpine[a,1]

In humans, more than 6 billion bp of DNA need to be copied accurately and completely every cell division. Given the size of human chromosomes, it would take a pair of bidirectional DNA replication forks, emanating from a single replication origin, more than 40 days to copy just chromosome 1. To duplicate the genome completely within the confines of S-phase, multiple start sites of DNA replication need to be distributed across each of the chromosomes. Understanding how start sites of DNA replication are selected and regulated across a diverse array of cell types has been a major question in the field since the discovery of the double helix. In PNAS, Miotto et al. (1) describe the genome-wide localization of a key replication initiation factor, the origin recognition complex (ORC), and provide insight into the genomic features that define the location and distribution of human replication origins.

ORC is a conserved and essential heterohexameric protein complex that is essential for the initiation of DNA replication (2). In G1 of the cell cycle, ORC, together with Cdt1 and Cdc6, directs the loading of the replicative helicase Mcm2–7 onto DNA to form the pre-replicative complex (pre-RC). The assembly of the pre-RC in G1 serves to "license" potential origins for activation in the subsequent S-phase (3). Cyclin-dependent kinase and Dbf4-dependent kinase activities during S-phase result in the recruitment of additional proteins to the pre-RC to form the preinitiation complex, unwinding of DNA, and initiation of DNA synthesis. The transactivating factors required for initiation of DNA replication are conserved from yeast to mammals, and the regulated initiation of DNA replication has recently been reconstituted from a minimal complement of ~42 proteins (4). Despite an in-depth understanding of the transacting factors required for the initiation of DNA replication, considerably less is known about *cis*-acting genomic features that define replication origins in the context of mammalian chromosomes.

In *Saccharomyces cerevisiae*, origins were first identified as short (~200 bp) autonomously replicating sequences (ARSs) required for the maintenance and propagation of episomal DNA (5). *S. cerevisiae* ORC binds to a *cis*-acting sequence element termed the ARS consensus sequence that is necessary, but not sufficient, for origin function. In contrast, ORC purified from higher eukaryotes exhibits little if any preference for binding specific sequences (6, 7). The lack of sequence specificity observed for ORC in higher eukaryotes suggests that additional chromatin features are likely involved in specifying replication origins.

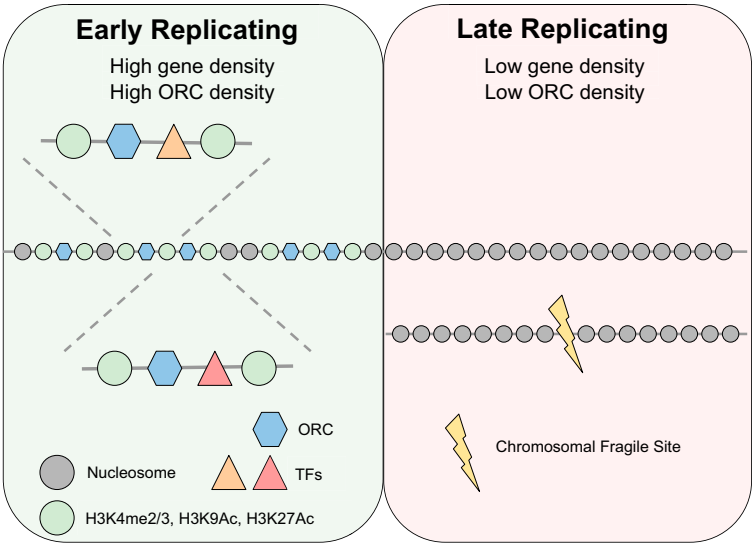

Fig. 1. ORC density is a determinant of replication timing and genome stability. Schematic of ORC localization in early- and late-replicating domains. ORC localizes to open and accessible chromatin marked by DNase I accessibility, transcription factor (TF) occupancy, and activating chromatin modifications. Late-replicating domains have decreased ORC density and frequently give rise to chromosomal fragile sites.

Chromatin immunoprecipitation (ChIP) coupled with genomic microarrays (ChIP-chip) or, more recently, next-generation sequencing allows for the comprehensive identification of genomic loci associated with specific DNA binding factors. In *S. cerevisiae*, ORC and the Mcm2–7 complex were among the first eukaryotic DNA binding factors to be mapped by ChIP-chip and provided the first genome-wide map of potential replication origins (8). The distribution of ORC was subsequently mapped by ChIP-chip in *Drosophila melanogaster* (9). However, despite these early successes

[a]Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, NC 27710
Author contributions: D.M.M. wrote the paper.
The author declares no conflict of interest.
See companion article 10.1073/pnas.1609060113.
[1]Email: david.macalpine@duke.edu.

in characterizing the genome-wide distribution of ORC, the identification of mammalian ORC binding sites remained a challenge. In the 15 years since the advent of genome-wide ChIP, numerous laboratories have attempted to map ORC binding sites precisely in mammalian cell lines with limited success. Anecdotally, the reasons for failure include low correlation among replicates, low signal-to-noise ratio, minimal overlap between different subunits, and little concordance with other genomic approaches for mapping replication initiation events. The failures were initially attributed to the lack of availability of ChIP-grade ORC antibodies; however, attempts to use tagged ORC subunits were also largely ineffective. Alternatively, due to the lack of observed sequence specificity for metazoan ORC, ORC may associate transiently or promiscuously with any accessible sequence in a stochastic manner. In 2013, Dellino et al. (10) overcame many of these challenges by "ChIPing" ORC from low-density chromatin associated with promoters and enhancers. In this partial map of human sequences associated with ORC, they identified ~8,000 binding sites throughout the human genome that were concordant between replicates. These 8,000 sites were far fewer than the estimated 50,000–80,000 replication origins required to duplicate the genome.

The Struhl group (1) now provides a more complete view of ORC localization throughout the human genome via the identification of nearly 50,000 sequences associated with ORC in K562 cells (Fig. 1). ORC localizes to open chromatin accessible by DNase I and is frequently found associated with promoters and enhancers. Not surprisingly, active chromatin marks, including H3K27Ac, H3K9Ac, and H3K4 methylation, are also enriched. No *cis*-acting sequence motifs were identified, but ORC binding sites were frequently co-occupied by transcription factors (including E2F, Myc, NF-κB, GATA, and AP-1). Despite the frequent colocalization with transcription factor binding sites, no predictive patterns of colocalization were identified. The chromatin states and joint occupancy with multiple transcription factors are similar to the observed occupancy of *Drosophila* ORC at highly occupied target regions of the fly genome (11) and are consistent with ORC localizing to accessible chromatin in a promiscuous manner independent of primary sequence. Given the colocalization with various DNA binding factors in these regions, the overall frequency of ORC occupancy across the population for a given locus remains to be determined.

The comprehensive identification of ORC binding sites provides a map of potential replication initiation sites throughout the human genome. Prior genome-wide experiments have identified replication origins based on the enrichment of specific DNA structures that arise during replication initiation. The identification of replication bubbles and short nascent strands and the distribution of Okazaki fragments have all been used to localize mammalian replication origins (12). A common theme to emerge is that replication origins are enriched at regulatory elements (promoters, enhancers, and CpG islands) in open and active chromatin. Despite gross similarities in the genomic features associated with origin activity, there has been only limited agreement and concordance between approaches at the nucleotide level. This lack of agreement is also true for the ORC mapping data presented here: Only about 12–15% of the origins mapped by short nascent strands are within 1–2 kb of an ORC binding site (1). Although these discrepancies may be due to technical limitations of either the ChIP (13) or short nascent strand (14) assay, they could also be due to initiation events being uncoupled from the site of ORC binding. For example, there is a significant excess of the Mcm2–7 helicase complex loaded onto DNA relative to ORC (15). Elegant in vitro and in vivo experiments from the Remus laboratory suggest that in *S. cerevisiae*, the Mcm2–7 complex can translocate along DNA and be displaced by transcription, resulting in de novo initiation sites (16). Similarly, in *Drosophila*, the genome-wide distribution of the Mcm2–7 complex as cells enter S-phase is distinct from ORC binding sites and shaped by active transcription (17). Thus, ORC may only pinpoint sites of helicase loading and not ultimately where initiation occurs.

The temporal order by which specific DNA sequences are copied during S-phase is referred to as the DNA replication timing program (18). Mammalian genomes are partitioned into discrete early- and late-replicating domains that span hundreds to thousands of kilobases. The time at which a sequence replicates is dependent on the density of replication origins and their time of activation during S-phase. The establishment of the replication timing program has been linked to chromatin environment (19), nuclear architecture (20) and the availability of limiting replication initiation factors (21, 22). In PNAS, Miotto et al. (1) provide evidence that ORC density is also a determinant of replication timing. Indeed, instead of simply reporting correlations and the density of ORC binding sites as a function of replication timing, they use mathematical modeling to predict replication timing accurately solely from the density of ORC binding sites and presumed initiation events. They also find a paucity of ORC binding in the proximity of chromosomal fragile sites and deletions frequently occurring in cancer cell lines. Together, these results suggest that the distribution and density of ORC are determinants of genome stability.

Miotto et al. describe the genome-wide localization of a key replication initiation factor, the origin recognition complex (ORC), and provide insight into the genomic features that define the location and distribution of human replication origins.

The comprehensive identification of ORC binding sites throughout the human genome provides insight into the mechanisms that specify potential replication origins. The sequence-independent localization of ORC to promoters and enhancers marked by active and accessible chromatin in both *Drosophila* and human cells suggests a conserved chromatin-based mechanism to specify ORC binding sites and potential replication origins in higher eukaryotes. Defining ORC binding sites by chromatin-based mechanisms provides a flexible system to specify the DNA replication program in different cell types, and the frequent colocalization at active promoters may decrease the likelihood of head-to-head collisions between RNA and DNA polymerases (23). The field is beginning to identify the genomic and chromatin features correlated with ORC binding and origin function; however, the precise chromatin-based rules that govern origin selection and activation at specific loci remain an enigma. Recent and rapid advances in CRISPR/Cas9 technologies to edit the genome precisely and modify chromatin structure (24) will hopefully allow for the rigorous testing of models predicted from the observed correlations between chromatin state and the DNA replication program.

Acknowledgments

This research is supported by the NIH (Grant R01-GM104097).

1 Miotto B, Ji Z, Stuhl K (2016) Selectivity of ORC binding sites and the relation to replication timing, fragile sites, and deletions in cancers. *Proc Natl Acad Sci USA*, 10.1073/pnas.1609060113.
2 Bell SP, Kaguni JM (2013) Helicase loading at chromosomal origins of replication. *Cold Spring Harb Perspect Biol* 5(6):a010124.
3 Siddiqui K, On KF, Diffley JFX (2013) Regulating DNA replication in eukarya. *Cold Spring Harb Perspect Biol* 5(9):a012930.
4 Yeeles JTP, Deegan TD, Janska A, Early A, Diffley JFX (2015) Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature* 519(7544):431–435.
5 Leonard AC, Méchali M (2013) DNA replication origins. *Cold Spring Harb Perspect Biol* 5(10):a010116.
6 Vashee S, et al. (2003) Sequence-independent DNA binding and replication initiation by the human origin recognition complex. *Genes Dev* 17(15):1894–1908.
7 Remus D, Beall EL, Botchan MR (2004) DNA topology, not DNA sequence, is a critical determinant for Drosophila ORC-DNA binding. *EMBO J* 23(4):897–907.
8 Wyrick JJ, et al. (2001) Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: High-resolution mapping of replication origins. *Science* 294(5550): 2357–2360.
9 MacAlpine HK, Gordân R, Powell SK, Hartemink AJ, MacAlpine DM (2010) Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading. *Genome Res* 20(2):201–211.
10 Dellino GI, 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):1–11.
11 modEncode Consortium, et al. (2010) Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science* 330(6012):1787–1797.
12 Hyrien O (2015) Peaks cloaked in the mist: The landscape of mammalian replication origins. *J Cell Biol* 208(2):147–160.
13 Teytelman L, Thurtle DM, Rine J, van Oudenaarden A (2013) Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc Natl Acad Sci USA* 110(46):18602–18607.
14 Foulk MS, Urban JM, Casella C, Gerbi SA (2015) Characterizing and controlling intrinsic biases of lambda exonuclease in nascent strand sequencing reveals phasing between nucleosomes and G-quadruplex motifs around a subset of human replication origins. *Genome Res* 25(5):725–735.
15 Edwards MC, et al. (2002) MCM2-7 complexes bind chromatin in a distributed pattern surrounding the origin recognition complex in Xenopus egg extracts. *J Biol Chem* 277(36):33049–33057.
16 Gros J, et al. (2015) Post-licensing specification of eukaryotic replication origins by facilitated Mcm2-7 sliding along DNA. *Mol Cell* 60(5):797–807.
17 Powell SK, et al. (2015) Dynamic loading and redistribution of the Mcm2-7 helicase complex through the cell cycle. *EMBO J* 34(4):531–543.
18 Rhind N, Gilbert DM (2013) DNA replication timing. *Cold Spring Harb Perspect Med* 3(7):1–26.
19 Birney E, et al.; ENCODE Project Consortium; NISC Comparative Sequencing Program; Baylor College of Medicine Human Genome Sequencing Center; Washington University Genome Sequencing Center; Broad Institute; Children's Hospital Oakland Research Institute (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447(7146):799–816.
20 Pope BD, et al. (2014) Topologically associating domains are stable units of replication-timing regulation. *Nature* 515(7527):402–405.
21 Mantiero D, Mackenzie A, Donaldson A, Zegerman P (2011) Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast. *EMBO J* 30(23):4805–4814.
22 Tanaka S, Nakato R, Katou Y, Shirahige K, Araki H (2011) Origin association of Sld3, Sld7, and Cdc45 proteins is a key step for determination of origin-firing timing. *Curr Biol* 21(24):2055–2063.
23 Petryk N, et al. (2016) Replication landscape of the human genome. *Nat Commun* 7:10208.
24 Thakore PI, Black JB, Hilton IB, Gersbach CA (2016) Editing the epigenome: Technologies for programmable transcription and epigenetic modulation. *Nat Methods* 13(2):127–137.