Chromosome Res (2020) 28:49–67
https://doi.org/10.1007/s10577-019-09624-y

REVIEW

Genomic methods for measuring DNA replication dynamics

Michelle L. Hulke · Dashiell J. Massey · Amnon Koren

Received: 1 October 2019 / Revised: 30 November 2019 / Accepted: 3 December 2019 / Published online: 17 December 2019

Abstract Genomic DNA replicates according to a defined temporal program in which early-replicating loci are associated with open chromatin, higher gene density, and increased gene expression levels, while late-replicating loci tend to be heterochromatic and show higher rates of genomic instability. The ability to measure DNA replication dynamics at genome scale has proven crucial for understanding the mechanisms and cellular consequences of DNA replication timing. Several methods, such as quantification of nucleotide analog incorporation and DNA copy number analyses, can accurately reconstruct the genomic replication timing profiles of various species and cell types. More recent developments have expanded the DNA replication genomic toolkit to assays that directly measure the activity of replication origins, while single-cell replication timing assays are beginning to reveal a new level of replication timing regulation. The combination of these methods, applied on a genomic scale and in multiple biological systems, promises to resolve many open questions and lead to a holistic understanding of how eukaryotic cells replicate their genomes accurately and efficiently.

Keywords DNA replication · replication timing · replication origin · genomics · single cell

Abbreviations

MFA	Marker frequency analysis
ORC	Origin recognition complex
MCM	Mini-chromosome maintenance
ChIP	Chromatin immunoprecipitation
ARS	Autonomous replicating sequence
SNS-seq	Short nascent strand sequencing
NSCR	Nascent strand capture and release
HU	Hydroxyurea
ini-seq	Initiation site sequencing
TSS	Transcription start sites
G4	G-quadruplex
LCL	Lymphoblastoid cell line
ESC	Embryonic stem cell
FISH	Fluorescence in situ hybridization
SMARD	Single-molecule analysis of replicating DNA
WGA	Whole-genome amplification
DOP-PCR	Degenerate-oligonucleotide-primed PCR
MALBAC	Multiple annealing and looping-based amplification cycles
LIANTI	Linear amplification via transposon insertion

Michelle L. Hulke and Dashiell J. Massey contributed equally to this work.

Responsible Editor: Beth A Sullivan

M. L. Hulke · D. J. Massey · A. Koren (✉)
Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA
e-mail: koren@cornell.edu

Introduction

The replication of genomic DNA is mediated by the sequential activation of DNA replication origins along chromosomes. Different origins are activated at different times during the S phase of the cell cycle to mediate bidirectional DNA synthesis that continues until the ensuing replication forks merge with opposing forks that originated from nearby origins. Ultimately, this process leads to the duplication of the entire genomic material, with different regions of the genome replicating at different times (Fig. 1). DNA replication origins are located at specific sites along chromosomes and are activated at characteristic and reproducible times, prompting questions regarding both the mechanisms and function of this spatiotemporal regulation. The "program" of replication origin activation and resulting replication timing has been linked to chromatin structure, chromosome conformation, gene activity, and genome stability and development. On average, early-replicating genomic regions tend to reside within open chromatin regions in the "A" compartment and contain a high density of active genes, while late-replicating regions tend to be gene-poor and are more susceptible to accumulating mutations during evolution or somatic cell divisions. The replication program has been shown to be extensively remodeled during embryonic development and cellular differentiation, while disruption of the normal replication program can lead to cell lethality, developmental defects, and cancer (reviewed in (Aladjem and Redon 2017; Gaboriaud and Wu 2019; Rhind and Gilbert 2013)).

Not with standing the central role of DNA replication and its temporal regulation in cellular and organismal biology, several fundamental questions remain incompletely resolved despite many years of intensive research. For instance, the locations of replication origins in higher eukaryotic genomes remain to be unequivocally established, as does the actual nature of replication initiation sites as single origins, clusters of nearby origins, or a combination thereof. Similarly, although the replication timing of different genomic regions appears to be highly

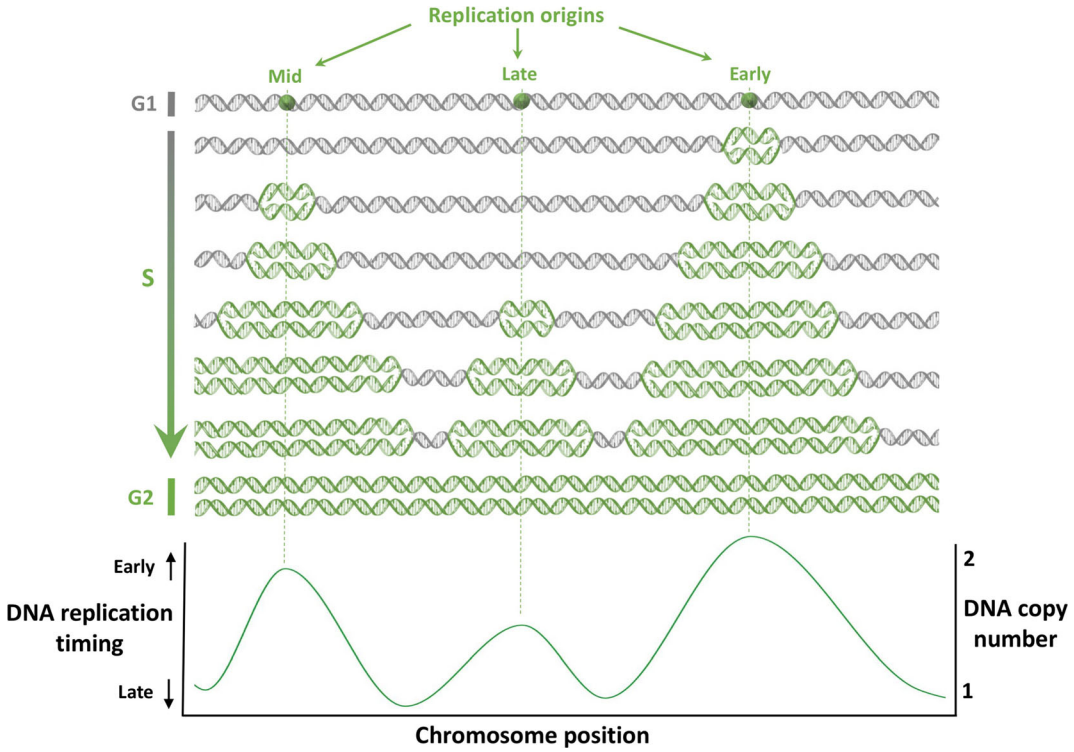

Fig. 1 DNA replication timing. DNA initiates at replication origins, each of which is activated at a characteristic time during S phase. Replication continues bidirectionally until all replication forks converge and the entire genome is replicated. The DNA replication timing program can be represented in the form of a replication profile (bottom), in which the amplitude of the profile represents the replication timing along the chromosome and peaks correspond to the locations of replication origins. These replication profiles also correspond to the average DNA copy number in a population of replicating cells, in which earlier-replicating regions will have a greater DNA content compared to later-replicating genomic regions

reproducible, there is compelling evidence for substantial heterogeneity in origin activation timing among individual cells. The molecular mechanisms determining the activation times of different replication origins are another fundamental yet unresolved aspect of DNA replication. Last, while replication timing has been linked to cellular and organismal development, it remains unclear to what extent that different functional aspects have driven the evolution of the replication timing program. For example, is a specific replication program required in order to support particular patterns of gene expression (Muller and Nieduszynski 2012) or for maintaining the stability of specific parts of the genome (Gaboriaud and Wu 2019)? Or perhaps are there other "functions" for the replication timing program?

For several decades, a plethora of genetic, molecular, biochemical, and cellular assays have been applied to study DNA replication dynamics. These studies have revealed the existence of a replication timing program, the locations and nature of replication origins in bacteria and in budding yeast, and a great deal of detail regarding the mechanisms of replication initiation and DNA synthesis. Classical studies have also started to point out the intricacies of DNA replication timing, including the complexity of metazoan replication origins and their regulation. These studies set the stage for the current era of DNA replication research, where replication dynamics are studied predominantly on a genomic scale. A genomic approach has proven particularly important for studying DNA replication: not only does it provide the advantage of scale, but it also enables a much deeper interrogation of the concerted process of genome replication.

DNA replication encompasses the entire genome in a single "mega-reaction" in which replication of one genomic region affects that of other genomic regions. This is due to several reasons. First, since replication is continuous along chromosomes, replication initiation from one locus affects the replication timing of adjacent loci. These dynamics also lead to a form of replication origin interference, in which incoming replication suppresses the activity of replication origins that have yet to fire. Thus, replication timing is best interpreted in the context of the activity of all surrounding origins and, by extrapolation, of all origins on a chromosome. Second, replication origins compete for limited amounts of initiation factors (Lynch et al. 2019; Mantiero et al. 2011; Tanaka et al. 2011; Wong et al. 2011). Thus, replication activity in one part of the genome can in principle influence the replication of other parts (Yoshida et al. 2014). Last, replication timing is relative: in some instances, an entire chromosome can be generally replicated before or after the replication of another chromosome (Koren and McCarroll 2014). By extension, a region that replicates relatively early compared to the rest of the genome may actually be late-replicating in the context of its chromosomal region. Thus, a full account of replication timing necessitates a genome-wide perspective.

Here, we review genomic approaches to studying DNA replication dynamics. We separate our discussion into three parts. First, we review genomic methods for measuring DNA replication timing itself and how these have evolved over the past few years. We then describe developing methods that aim to map the actual locations of DNA replication origins in higher eukaryotes. We end by discussing recent approaches for genome-wide replication profiling in single cells. Throughout, we focus on recent developments and challenges for the future, while we point the reader to several excellent reviews that provide more detail about various methods and approaches (Hyrien 2015; Prioleau and MacAlpine 2016; Raghuraman and Brewer 2010; Urban et al. 2015).

Measuring DNA replication timing in eukaryotic genomes

The first measurement of replication timing of a eukaryotic genome used a clever genomic implementation of the classic Meselson-Stahl experiment. *Saccharomyces cerevisiae* (budding yeast) cells were cultured in isotopically dense media, after which cells were transferred to light media and synchronously released into S phase. Replicated (heavy-light) DNA was separated from unreplicated (heavy-heavy) DNA at several time points, and genomic tiling microarrays were used to reveal the locations of replicated DNA, generating a genome-wide profile of replication timing (Raghuraman et al. 2001). The genomic scale of this first replication timing measurement revealed the locations of replication origins in the form of peaks that replicated before their adjacent regions, as well as the activation times of replication origins as indicated by the height of each peak. Similarly, troughs in the profile represented the locations and timing of replication termination events, while the slopes between peaks and valleys were interpreted to correspond to the rate of replication fork progression (see Fig. 1). This study revealed several important principles that were largely inaccessible with pre-genomic analyses. For instance, origin activation times appeared

to lie on a continuum rather than being bimodally distributed to early and late. In addition, replication was earlier near the centromeres and later near telomeres, a property that had been alluded to with locus-specific studies yet is best evaluated with chromosome-level data. Notwithstanding the transformative impact of this first genomic analysis of replication timing, several limitations of such an approach also became clear. For instance, not all replication origins can be detected, in particular locally weak or inefficient origins that may be undetected or missed in data analysis (e.g., oversmoothed). Similarly, the inferred locations of replication origins are not precise, and single origins cannot be discriminated from tight clusters of several nearby origins. Replication fork progression inferred from replication timing profiles can be confounded by integrating information from multiple, often superimposed replication forks (de Moura et al. 2010). To a large extent, these limitations are due to the ensemble cell-population-level measurement of replication timing in bulk cultures. Heterogeneity between cells has emerged as an important layer of complexity and a missing source of information in such genomic studies (see further below).

Following the first characterization of the replication dynamics of a eukaryotic genome, a branching-out of techniques ensued. Notable among those are the probing of replicating DNA via detection of single-stranded DNA (Feng et al. 2006), immunoprecipitation of nucleotide analogs (specifically, the thymidine analog BrdU) incorporated into replicating DNA (Knott et al. 2012; Peace et al. 2016), and detection of Okazaki fragments by DNA ligase inactivation and genomic analysis of short single-stranded DNA fragments (McGuffee et al. 2013; Smith and Whitehouse 2012). Another approach, which was used in parallel to the first characterization of the yeast replication program, is based on DNA copy number *per se*: the doubling of DNA can be followed along chromosomes as cells traverse through S phase (Yabuki et al. 2002).

All of the above techniques proved consistent and highly informative with regard to the replication of the budding yeast genome. However, interrogating the genomes of metazoan species introduced further challenges, most notably the greater difficulty of obtaining synchronized cell cultures and of genetic manipulation of the cells. While specialized techniques to synchronize mammalian cells have been implicated in genomic replication timing measurements (Farkash-Amar et al. 2008; Jeon et al. 2005), two complementary approaches ultimately became the current mainstay of mammalian replication timing profiling. Both of these are genomic versions mirroring (or following) yeast techniques or locus-specific molecular techniques (Gartler et al. 1999; Hansen et al. 1995; Hansen et al. 1993; Koren et al. 2010a; Yabuki et al. 2002). The first approach is immunoprecipitation of BrdU-labeled DNA from cells flow-sorted into different fractions of S phase. Cells are incubated with BrdU for a certain amount of time (typically 30 min to 2 h) to allow incorporation into actively replicating DNA. This is followed by sorting of S phase cells, immunoprecipitation of BrdU-labeled DNA, and genomic analysis of the DNA by either microarrays or, more recently, next-generation sequencing. This approach obviated the need for synchronization and enabled genome-wide replication timing measurements in higher eukaryotes such as flies (Schubeler et al. 2002), mice (Hiratani et al. 2008), and humans (Chen et al. 2010; Hansen et al. 2010; Rivera-Mulia et al. 2015). Retrospective synchronization achieved by flow sorting enables a more native experiment without chemical perturbation of live cells. On the other hand, the reliance on analog-labeled DNA prevents the ability to use non-replicating G1 cells as a direct control. Accordingly, some experiments compare cells in late S phase to cells in early S phase (Hiratani et al. 2008; Rivera-Mulia et al. 2015; Ryba et al. 2010). Alternatively, several studies have used up to six fractions of cells within S phase (Chen et al. 2010; Du et al. 2019; Hansen et al. 2010) to measure replication timing. The use of multiple S phase fractions in particular reveals loci that replicate at different times between the two chromosomal copies (asynchronously replicating regions) (Hansen et al. 2010). This family of approaches is typically referred to as "Repli-seq."

Repli-seq provides a highly reproducible assay and has led to the emergence of the "replication domain" model, in which long regions of 400 Kb to 1.2 Mb with consistent replication timing manifest as relatively flat plateaus on replication profiles (Hiratani et al. 2008; Pope and Gilbert 2013). Two limitations of Repli-seq are, however, notable: the use of BrdU requires continuous incubation with the analog which therefore labels long stretches of DNA as single units. These stretches are similar in length to replication domains. A second potential shortcoming is the sorting of sub-S phase fractions. This is expected to limit resolution, particularly at the beginning and end of S phase, if not all S phase cells are included in the sorting gates. In addition, the

use of one S phase fraction in comparison to another, instead of relative to a non-replicating G1 control, could also compromise the accuracy of the replication timing measurements. It remains an open question whether these restrictions account for some of the global properties of replication timing profiles measured using Repli-seq, such as broad domains.

A second general approach for measuring replication timing is based on assaying DNA copy number (see Fig. 1), similar to the original yeast experiment (Yabuki et al. 2002), yet by using sorted S phase cells instead of synchronized cells (Koren et al. 2010a; Woodfine et al. 2004). The quantitative nature of copy number measurements and the ability to assay G1 cells in parallel enable sorting a single, entire S phase fraction instead of several sub-S fractions. This offers a more complete and uniform representation of the replication profile, while G1 cells serve to correct for any technical deviation from euploid copy number. The power of this technique lies in its independence from in vivo manipulation of cells and in its experimental simplicity. It has accordingly been applied to various systems, from yeast where it enabled scaled-up experiments in many strains (Gispan et al. 2014, 2017; Koren et al. 2010a) and species (Agier et al. 2018; Koren et al. 2010b; Liachko et al. 2014; Muller and Nieduszynski 2012), to model organisms such as zebrafish (Siefert et al. 2017) and mice (Gdula et al. 2019; Gitlin et al. 2015; Yaffe et al. 2010; Yehuda et al. 2018), and to humans (Brody et al. 2018; Desprat et al. 2009; Koren et al. 2012; Massey et al. 2019; Mukhopadhyay et al. 2014; Yaffe et al. 2010). The more uniform representation of cells in S phase provides higher-resolution replication profiles than Repli-seq techniques, especially with deeper sequencing (Koren et al. 2014; Koren et al. 2012; Mukhopadhyay et al. 2014). However, the reliance on cell sorting means that resolution may still be lost in some very early or very late-replicating regions, since the corresponding cells may reside within the G1 or G2/M cell cycle fractions.

A more recent development of the copy number approach has eliminated the need for cell sorting. While direct measurement of DNA copy number in proliferating cells has been used for detection of replication origins in bacteria (referred to as MFA, marker frequency analysis; (Rudolph et al. 2013; Sueoka and Yoshikawa 1965; Yoshikawa and Sueoka 1963)) and also suggested in yeast (although stationary phase cells were still used as a reference; (Muller et al. 2014)), this approach has more recently been used to probe the replication profile of entire human genomes (Koren et al. 2014). Proliferating cell cultures include a fraction of cells in S phase that enables detection of subtle fluctuations in DNA copy number along chromosomes by whole-genome sequencing. As little as 5–10% of cells in S phase are sufficient in order to generate high-resolution replication profiles. This approach is the least manipulative of all and for the same reason also the easiest to implicate. It does, however, require high-quality (and deeper coverage compared to other techniques) whole-genome sequence data and is more computationally challenging. The extra computational processing steps account for the lack of an empirical control sample (such as a G1 cell sample). Instead, the sequence data in combination with the reference genome is used to in silico simulate copy number profiles in non-replicating cells (Koren et al. 2014). Ultimately, this sequencing-only, copy number-based method is proving to yield the most accurate replication timing profiles, with sharp peaks and valleys even in mammalian genomes (Fig. 2).

Overall, genome-wide replication timing has by now been measured in various yeast species, plants, flies, zebrafish, mice, ape species, and normal and cancer human cells (Concia et al. 2018; Farkash-Amar and Simon 2010; Rhind and Gilbert 2013; Yang et al. 2018). It has been applied to cultured cells as well as in vivo animal specimens and low-input samples (Rivera-Mulia et al. 2019; Siefert et al. 2017; Yehuda et al. 2018). It has been performed in an allele-specific manner by relying on polymorphisms in offspring of genotyped family trios and quartets (Koren and McCarroll 2014; Mukhopadhyay et al. 2014) and of mouse strain crosses (Rivera-Mulia et al. 2018). Most recently, it has been advanced to the level of single-cell measurements (see below).

Mapping DNA replication origins genome-wide

While replication timing profiles provide a comprehensive account of a genome's replication dynamics, they only indirectly inform of the locations and activation times of replication origins. Accordingly, peaks in replication timing profiles of yeasts correspond to replication origins but can nevertheless be located several kilobases away from the actual initiation sites (due to resolution limitations). In higher eukaryotes, Repli-seq

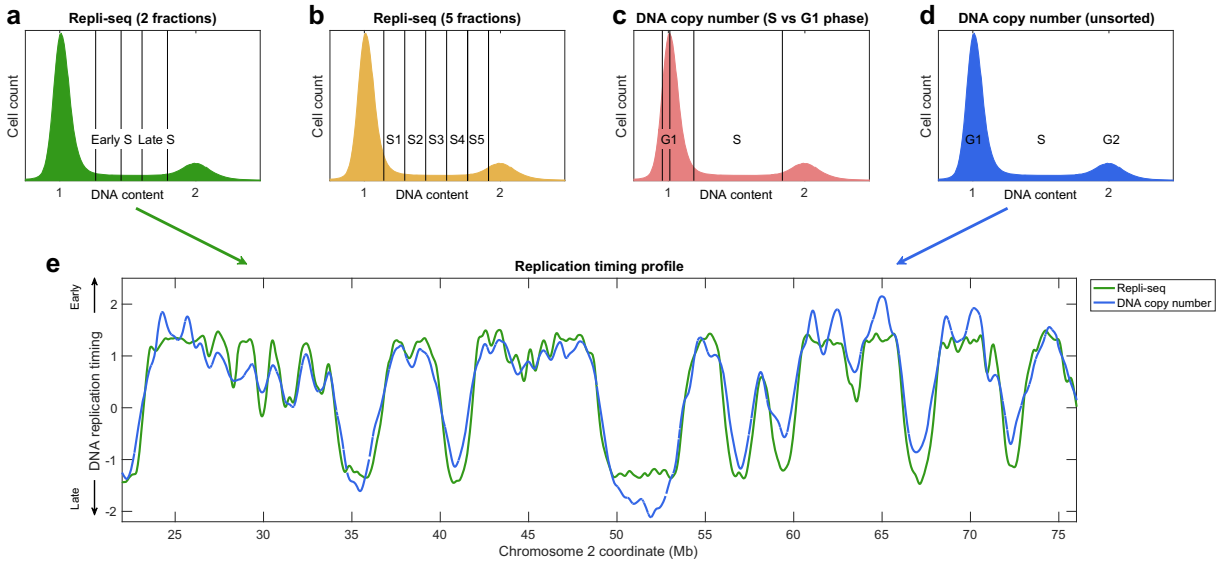

Fig. 2. DNA replication timing. (**A**–**D**) Different cell isolation schemes for measuring DNA replication timing. In Repli-seq (**A**–**B**), cells are incubated with BrdU, two to six fractions of S phase cells are sorted (five fractions are shown in **B**), and BrdU-containing DNA is sequenced. The choice of gate locations varies from one experiment to another, with some experiments using wider gates that together span most of S phase. Alternatively, the entire S phase fraction as well as G1 cells can be sorted and DNA sequenced without BrdU incorporation (**C**); DNA copy number is analyzed directly to infer replication dynamics. The control G1 cells are sorted from the left side of the G1 peak, in order to avoid inclusion of contaminating S phase cells (which would distort the replication pattern in early S phase). G2 cells could also be used as controls instead of G1 cells. (**D**) Last, DNA can be sequenced from proliferating cells without any sorting. This reduces technical influences and avoids arbitrary choices of gate locations. (**E**) Representative DNA replication timing profiles for the lymphoblastoid cell line GM12878 generated by either two-fraction Repli-seq (Pope et al. 2014) or DNA copy number without sorting (Koren et al. 2014). Replication profiles from multi-fraction Repli-seq and from S-/G1-sorted cells (not shown) typically fall in between the other two approaches. Greater dynamic range and sharper peaks may be obtained by avoiding FACS sorting of S phase cells

approaches are typically thought to reveal replication domains rather than individual initiating sites, while copy number approaches can reveal distinct peaks yet cannot resolve single origins from origin clusters nor predict the precise locations of initiation sites. In parallel to studies of DNA replication timing, methods that directly identify replication initiation sites are therefore instrumental in understanding eukaryotic DNA replication.

Eukaryotic replication origins have originally been identified in budding yeast using several elegant techniques, some of which were later adapted to the genomic level. Originally, DNA sequences that confer autonomous replication ability to plasmids were isolated, which was later followed by functional 2-D gel assays in which locus-specific replication bubble structures were identified in their native chromosomal location (reviewed in (Fangman and Brewer 1991)). Such experiments opened the way to comprehensive genetic and biochemical characterization of yeast replication origins and ultimately led to the uncovering of the origin recognition complex (ORC), the mini-chromosome maintenance (MCM) complex that forms the core of the replicative helicase, and other factors essential for replication initiation ((Bell and Stillman 1992; Diffley and Cocker 1992; Maine et al. 1984); reviewed in (Raghuraman and Brewer 2010; Tye 1999)). Subsequently, ORC and MCM binding were used to probe for origin sequences by chromatin immunoprecipitation (ChIP), which was later used on a genomic scale to reveal active as well as inactive origins genome-wide (Wyrick et al. 2001). The autonomous replicating sequence (ARS) assay has also been applied on a genome-wide level in budding yeast ("ARS-seq"), which revealed new ARSs, enabled identification of essential ARS sub-sequences, and tested the influence of sequence alterations on ARS function (Liachko et al. 2013). ARS-seq also enabled the identification of replication origins in other yeast species, such as *Pichia pastoris*, in which two different ARS elements were

identified (Liachko et al. 2014). Last, the identification of replication origin sequences in *S. cerevisiae* enabled their further refinement and dissection by studying their conservation among closely related *Saccharomyces* species (Nieduszynski et al. 2006). Together, the combination of focused and genome-scale assays has led to a detailed understanding of the locations, nature, and mechanisms of yeast replication origin activation.

While the biochemistry of replication initiation turned out to be highly conserved throughout eukaryotic evolution, the specification and regulation of replication origin sites are not. For instance, DNA sequences that are essential for mammalian replication origin activity have not been found, and the ARS assay has not been successful in higher eukaryotes (Gilbert 2004; Hyrien 2015; Prioleau and MacAlpine 2016). Instead, metazoan ORC binds DNA promiscuously throughout the genome (MacAlpine et al. 2010; Vashee et al. 2003), suggesting that a conserved sequence is not a primary determinant of replication origins in higher eukaryotes. Currently, searches for mammalian replication origins rely mostly on the molecular and biochemical rather than genetic properties of origins. Many of these methods are extensions of earlier yeast techniques for origin mapping. For instance, ChIP-seq has been applied to map human replication origins (Dellino et al. 2013; Miotto et al. 2016; Sugimoto et al. 2018). An alternative, ChEC-seq, fuses a protein of interest – for example, MCM subunits – to micrococcal nuclease, and uses DNA sequencing to examine the patterns of DNA cleavage reflecting the native binding region of the protein (Foss et al. 2019; Zentner et al. 2015). Bubble-seq extracts DNA bubble structures trapped in electrophoretic gels followed by next-generation sequencing to map their locations (Mesner et al. 2013). Another technique, short nascent strand sequencing (SNS-seq), size-selects and sequences short single-stranded DNA fragments that represent the early products of DNA replication (Besnard et al. 2012; Cadoret et al. 2008; Cayrou et al. 2011). In order to prevent false signals arising from random chromosomal breaks that occur during DNA isolation, SNS-seq makes use of another biochemical property specific to nascent DNA strands: the presence of short RNA primers upstream of newly replicated DNA. In yet another translation of a classic yeast technique to the genomic era, lambda exonuclease, an enzyme that digests DNA but is blocked by RNA (Bielinsky and Gerbi 1998), has been implicated to enhance the specificity of SNS-seq to replicating DNA. Alternatively, RNA primers can be used to positively select for nascent DNA strands by binding short DNA molecules to a column and then using RNase I to specifically release RNA-primed DNA fragments ("nascent strand capture and release," NSCR; (Kunnev et al. 2015)).

Instead of relying on RNA primers, nascent DNA can be isolated by pulse-labeling cells with BrdU or EdU, followed by isolation and sequencing analog-containing short double-stranded DNA fragments (Li et al. 2014; Mukhopadhyay et al. 2014; Smith et al. 2016). This approach has also been applied to cells arrested in early S phase using hydroxyurea (HU) in a technique called EdUseq-HU (Macheret and Halazonetis 2018; Tubbs et al. 2018). Alternatively, ini-seq (initiation site sequencing) uses isolated G1 nuclei that initiate replication in a cell-free system with soluble human cell extracts in the presence of digoxigenin-dUTP (Langley et al. 2016). The cell-free system used in ini-seq supports much slower replication than normal, facilitating the labeling of short DNA strands. Of note, the incubation periods with nucleotide analogs differ among these various techniques, which could potentially affect the resolution of origin calling.

Okazaki fragment sequencing (Smith and Whitehouse 2012) has also been applied to mammalian genomes. In OK-seq, cells are briefly pulse-labeled with EdU, after which short DNA fragments are isolated and EdU-containing DNA is enriched and sequenced in a strand-specific manner. Shifts in the ratios of DNA content between the two strands identify the locations of replication origins or, reciprocally, termini (Chen et al. 2019; Petryk et al. 2018; Petryk et al. 2016; Tubbs et al. 2018; Wu et al. 2018). Another approach based on the asymmetry of the replication fork uses bioinformatic sequence analysis to predict the locations of replication origins, adopting classical work in bacteria. The differences in DNA polymerase usage and DNA synthesis kinetics between the leading and lagging strands lead to a mutational signature – nucleotide composition skews – that reverses sign exactly at replication initiation sites (Huvet et al. 2007; Touchon et al. 2005). This principle has also been applied to particularly stable centromeric origins in certain yeast species (Koren et al. 2010b) and has been used to link somatic cancer mutations to replication direction shifts and, by inference, initiation sites (Haradhvala et al. 2016; Shinbrot et al. 2014; Tomkova et al. 2018). Interestingly, mutational patterns reminiscent of bidirectional replication around origins have been observed at CpG islands, which are associated with transcription start sites (TSS;

see further below (Polak and Arndt 2009)). The asymmetrical usage of DNA polymerases within the eukaryotic replication fork can also be used to experimentally map replication initiation sites and overall replication dynamics. This has been achieved by modifying different DNA polymerases so that they incorporate an excess of ribonucleotides into the sites and strands that they replicate, which can then be assayed by DNA sequencing (Clausen et al. 2015; Daigaku et al. 2015; Reijns et al. 2015).

Naturally, these diverse techniques each have strengths and limitations. For instance, bubble-seq is generalizable and unobstructive to replication since it does not require any in vivo modification or manipulation of DNA. In addition, the low background signal in bubble-seq could enable the identification of rare origins that might be lost in methods that rely on signal averages (reviewed in (Prioleau and MacAlpine 2016)). However, bubble-seq relies on restriction enzymes that generate the target lengths of DNA fragments to be assayed by 2-D gels; this limits the resolution of origin identification and also which origins can be identified. The use of two restriction enzymes in parallel has been proposed to at least partially overcome these limitations (Mesner et al. 2013). While bubble-seq has generally shown low reproducibility, possibly requiring far more input to reach full saturation of origin calling, SNS-seq has proven reproducible among different cell types and labs and has been applied in various species and systems, under replication stress, as well as for allele-specific origin mapping (Almeida et al. 2018; Bartholdy et al. 2015; Besnard et al. 2012; Cayrou et al. 2015; Cayrou et al. 2011; Comoglio et al. 2015; Jodkowska et al. 2019; Lombrana et al. 2016; Martin et al. 2011; Massip et al. 2019; Mukhopadhyay et al. 2014; Picard et al. 2014; Sequeira-Mendes et al. 2019; Smith et al. 2016; Yudkin et al. 2014). SNS-seq boasts a finer resolution compared to bubble-seq as the reads are not limited by restriction cut sites. However, because replication origins are identified by calling peaks of sequence read coverage, SNS-seq will primarily detect frequently used origins within a population of cells. Another concern that has been raised with regards to SNS-seq is that lambda exonuclease may have lower efficiency for digesting not only RNA-primed DNA but also other species such as single-stranded DNA, GC-rich DNA, and G-quadruplexes (G4s; (Foulk et al. 2015; Perkins et al. 2003)). This has been suggested to explain the enrichment of these genomic features in SNS-seq libraries, although further controls may argue against that ((Prioleau 2017) and see further below). Similarly, ribonucleotides, which are abundantly embedded in genomic DNA (reviewed in (Jinks-Robertson and Klein 2015; Williams et al. 2016)), may also create false-positive origin calls by directly blocking DNA degradation or by distorting the DNA helix (Klein 2017). A comparison with nucleotide skew patterns even suggested that some sites called by SNS-seq and bubble-seq may correspond to sites of replication termination, not just initiation (Langley et al. 2016).

High resolution of origin calling has also been suggested to be obtained by methods such as ini-seq and EdUseq-HU, although these methods by design will only identify the earliest replication origins in the genome. In fact, part of the signal in nucleotide analog-based short nascent strand sequencing may come from enriching early-replicating DNA per se, independently of replication origins. For example, normalizing by a control of newly replicated (input) DNA without immunoprecipitation of labeled DNA actually decreased the reproducibility of ini-seq (Langley et al. 2016). Similarly, BrdU-seq sites correspond significantly to sites enriched in input DNA (Li et al. 2014). ChIP-seq techniques identify active as well as inactive, dormant origins, which is a strength but also a limitation without the ability to functionally discriminate the two. A common potential source of bias in virtually all origin mapping techniques to date is the need to amplify the small amounts of DNA that are obtained. Single-molecule approaches (see below) may circumvent this need for DNA amplification (Fig. 3).

More fundamental, however, is the lack of consistency among different origin mapping techniques. Specifically, different human replication origin maps reveal initiation sites that show only partial overlap and sometimes not much more overlap than those obtained with random sets of loci (Dellino et al. 2013; Hyrien 2015; Langley et al. 2016; Miotto et al. 2016; Petryk et al. 2016; Picard et al. 2014; Urban et al. 2015). Some techniques, such as EdUseq-HU (Macheret and Halazonetis 2018; Tubbs et al. 2018) or OK-seq (Petryk et al. 2016), map several thousand sites, while SNS-seq and BrdU-seq often map several hundreds of thousands of sites (Bartholdy et al. 2015; Besnard et al. 2012; Massip et al. 2019; Mukhopadhyay et al. 2014). Different techniques reach different and sometimes opposing conclusions with regard to whether origins are discrete sites (e.g., SNS-seq, ini-seq), broad initiation

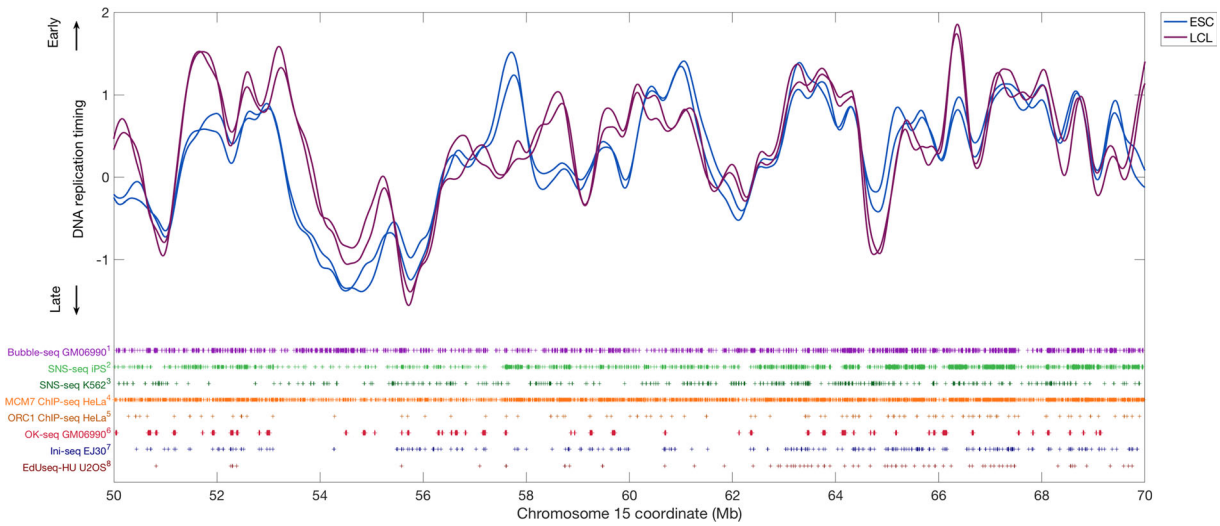

Fig. 3 DNA replication origin mapping. Various techniques have been used to map the locations of DNA replication origins in the human genome. Shown are the mapped origin locations in several of these techniques, together with replication timing profiles in human lymphoblastoid cell lines (LCL) and embryonic stem cells (ESCs, two cell lines each) measured by DNA copy number sequencing. Replication origin or initiation zone locations are based on the mapping used by each respective study and are represented as plus signs with widths that correspond to the size of the initiation sites or zones. At some locations, replication origins are consistent between techniques and also correspond to peaks in the replication timing profiles. However, there are also fundamental differences among origin mapping methods in the number of initiation sites along chromosomes, their widths (precise sites vs broader initiation zones), and the actual locations of predicted origins. [1] (Mesner et al. 2013); [2] (Smith et al. 2016); [3] (Picard et al. 2014); [4] (Sugimoto et al. 2018); [5] (Dellino et al. 2013); [6] (Petryk et al. 2016); [7] (Langley et al. 2016); [8] (Macheret and Halazonetis 2018). GM06990 is an LCL

zones (bubble-seq, OK-seq), or peaks that may themselves organize into larger zones (EdUseq-HU). It is almost inevitable to conclude that at least some of the current origin maps contain a substantial fraction of false negatives and/or false positives. One approach to circumvent this has been to look at the intersection of sites mapped by more than one technique (Karnani et al. 2010; Mukhopadhyay et al. 2014). While intersecting origin maps would reduce false positives, it also entails the risk of increasing the burden of false negatives (i.e., true origins not called by one method) and cannot substitute for a more complete understanding of the potential biases of the various techniques.

Notwithstanding inconsistencies between techniques, several genetic and epigenetic features are consistently emerging as associated with replication origins. For instance, various techniques have provided evidence for the association of origins with transcribed genes or transcription start sites ((Besnard et al. 2012; Cadoret et al. 2008; Cayrou et al. 2015; Chen et al. 2019; Dellino et al. 2013; Langley et al. 2016; Picard et al. 2014; Sequeira-Mendes et al. 2009; Sugimoto et al. 2018); but see (Macheret and Halazonetis 2018; Martin et al. 2011)). One of the most intriguing findings that has emerged from origin mapping techniques is an association with G4s. Although it remains possible that this association is a side effect, for example, of lambda exonuclease being blocked by G4s and not just RNA primers ((Foulk et al. 2015); see further below), evidence is accumulating for G4s being directly associated with mammalian replication origins. This association is observed in SNS-seq even when including control G1 cells ((Smith et al. 2016); but see (Foulk et al. 2015)) or an RNAse-treated control that accounts for background signals of lambda exonuclease and replication-independent short nascent strands (Cayrou et al. 2015). In addition, BrdU nascent strands and ini-seq have also linked G4 sequences to replication initiation independently of lambda exonuclease (Langley et al. 2016; Mukhopadhyay et al. 2014). Experiments in chicken DT40 cells showed directly that G4 sequences are required (although they are not sufficient) for replication origin activity (Valton et al. 2014). A recent comprehensive study of the effect of G4s on DNA replication provided compelling evidence for the importance of G4s for origin activity, their ability to drive origin

activity de novo, and their ability to drive the autonomous replication of episomal vectors in frog egg extracts (Prorok et al. 2019). The involvement of G4s in DNA replication is also suggested by their binding to ORC (Hoshina et al. 2013) and the replication factors RecQL4 (Keller et al. 2014), MTBP (Kumagai and Dunphy 2017), and Rif1 (Masai et al. 2018).

Still, different studies report varying levels of association of G4 sequences with replication origins, and several have suggested that G4s are not necessary for origin activity (e.g., (Comoglio et al. 2015; Dellino et al. 2013; Foulk et al. 2015; Mesner et al. 2013; Miotto et al. 2016; Sugimoto et al. 2018)). G4s are highly abundant across the genome and tend to occur in promoters and untranslated regions (Prioleau 2017); origins that are enriched for G4s are often found at TSSs; thus, it remains possible that G4s are not an independent signal of replication origins (Langley et al. 2016). In addition, origins mapped by EdUseq-HU have not been reported to associate with G4s and instead appear to be linked to long asymmetrical poly(dA:dT) tracts (Tubbs et al. 2018). Similarly, OK-seq initiation zones were enriched for active genes and CpG islands at their borders but were not enriched for G-quadruplex structures nor CpG islands within the initiation zones. There is also evidence that G4s cause replication fork stalling, as opposed to directly driving initiation ((Comoglio et al. 2015; Tubbs et al. 2018); reviewed in (Prioleau 2017)). Replication restart at G4 sites or other sites of replication stalling could in principle produce strong SNS signals that do not represent canonical replication origin sites. Allele-specific SNS-seq (Bartholdy et al. 2015) has suggested that G4 sequences do not contribute to origin firing efficiency. Specifically, SNP alleles that disrupt G4 sequences increased origin activity more often than they decreased origin activity. In addition, the enrichment of G4 sequences at origin sites was suggested to be secondary to the high GC content of these sites rather than being specific to G4s. Instead, origins correlated with asymmetric G/C and A/T sequences which were suggested to be the actual drivers of origin activity by means of facilitating origin unwinding (Bartholdy et al. 2015). Taken together, it remains to be settled whether G4s or other sequence elements and DNA structures contribute to replication origin activity or are side effects of technical manipulations or simply correlative findings driven by other genomic features.

Going forward, the locations, nature, and genetic and epigenetic associations of metazoan replication origins remain to be fully established. A useful benchmark for understanding the outputs of different methods would be to map origins in model organisms in which their locations are already well known and finely mapped. For instance, SNS-seq has never (to our knowledge) been applied to budding yeast or even bacteria, although the expectation would be that it will uncover the known initiation sites in these species or alternatively reveal non-origin elements that are enriched by SNS-seq (of note, one study that utilized lambda exonuclease to enrich for Okazaki fragments in yeast was unable to link them to known replication origins (Yanga and Lib 2013)). Zebrafish would also be a useful model species, since replication timing and origins are well mapped and, in contrast to many other species, are not correlated with GC content (Siefert et al. 2017). In addition, it is important to systematically include controls such as non-replicating DNA and DNA specifically subject to (or excluded from) the same enzymatic reactions used in any given technique. Another important aspect of any genomic method is rigorous statistical analysis. For instance, not all current studies control for confounding correlated factors when looking for genetic and epigenetic enrichments at replication origins. These would include sequence composition, gene density, open chromatin, replication timing, nucleosome positioning, and other factors that tend to correlate with each other along chromosomes. An association with one of these factors may not always point to a biologically relevant mechanism.

Last, it remains possible that unexpected aspects of DNA synthesis may be at the basis of some of the apparent discrepancies between studies. For instance, some initiation events may not lead to productive DNA synthesis but instead result in aborted replication forks, extrachromosomal circular DNA (also known as "microDNA"), or other outcomes (Artemov et al. 2019; Moller et al. 2018; Shibata et al. 2012; Urban et al. 2015). Other replication-independent DNA synthesis events such as repair intermediates may also be detected by some techniques that attempt to map replication origins. For instance, transcription-associated DNA damage at highly transcribed genes can lead to DNA repair synthesis that results in high rates of ribonucleotide incorporation into DNA (Owiti et al. 2018). Similarly, several studies have shown that RNA can prime DNA replication outside of replication origins, typically

as a result of collisions between RNA and DNA polymerases or the formation of RNA-DNA hybrids (R-loops) (Pomerantz and O'Donnell 2008; Ravoityte and Wellinger 2017; Stuckey et al. 2015; Xu and Clayton 1996). Such processes, to the extent that they influence origin mapping techniques, might be more readily detectable by some methods compared to others. They may also themselves be biased to certain genomic features, such as G4 structures (Hoshina et al. 2013). In that respect, assays such as replication timing, nucleotide composition skews, OK-seq, and polymerase usage are useful readouts of DNA replication that do not directly depend on measuring replication origin activity but nonetheless represent the outcomes of productive initiation events at canonical origin sites. With higher-resolution replication timing profiles, it may prove beneficial to consider the locations of replication timing peak sites as bona fide markers of initiation events.

Toward single-cell genomics of DNA replication

While genome-wide replication timing assays have proven highly reproducible and informative, a clear limitation of these assays is that they rely on ensemble population analyses and do not provide information about replication progression in individual cells. Similarly, replication origin mapping techniques rely on many cells, largely masking heterogeneity that might be present among cells. This could potentially also underscore some of the discrepancies between techniques, for example, if some techniques have a greater propensity to reveal replication origins that are only active in a subset of cells.

It has long been argued that DNA replication is stochastic, in the sense that different cells (or the descendants of the same cell in different cell cycles) activate different subsets of replication origins. From the perspective of studying replication origins, the manifestation of this stochasticity is that a given origin may or may not fire in a given cell cycle. Thus, origins are characterized by a genomic location and a preferred time of activation but also by the probability of being activated. Importantly, many DNA replication assays measure a combination of these factors, and it may be difficult to disentangle what is actually being measured. For instance, an origin that fires frequently in late S phase may appear similar in an ensemble analysis to an origin that fires early but only in a subset of cells. In addition, if origins are clustered along chromosomes, the firing of different origins within the same region in different cells will give the impression of an extended initiation zone, masking the activity of the individual origins.

The inefficiency of replication origins has been observed early on with techniques such as 2-D gels. Another line of evidence in support of stochastic activation of replication origins came from DNA combing experiments, which assay nascent DNA replication in single molecules (Bensimon et al. 1994; Michalet et al. 1997; Norio and Schildkraut 2001) and are extensions of earlier DNA fiber autoradiography analyses (Huberman and Riggs 1968). In these experiments, replication is allowed to proceed in the presence of a succession of labeled nucleotide analogs, followed by stretching of individual DNA fibers on glass slides. The location and orientation of replication forks are inferred from immunofluorescent detection of the incorporated nucleotides. Combing-based analyses of DNA replication have been performed in frog cell extracts (Blow et al. 2001; Herrick et al. 2000; Labit et al. 2008), yeasts (Czajkowsky et al. 2008; Patel et al. 2006), flies (Cayrou et al. 2011), mice (Cayrou et al. 2011; Norio et al. 2005; Piunti et al. 2014), and human cells (Bester et al. 2011; Frum et al. 2009; Guilbaud et al. 2011; Lamm et al. 2015). These studies have shown that firing of an origin in one cell is not correlated to its firing in other cells (Czajkowsky et al. 2008) or in a subsequent cell cycle (Labit et al. 2008; Patel et al. 2006). However, at scales larger than an individual origin, single-molecule replication is conserved: averaging data across all single molecules recapitulates the population-level replication profile (Czajkowsky et al. 2008), and broad regions appear to have consistent replication timing across cell cycles (Cayrou et al. 2011; Labit et al. 2008). Thus, stochastic firing of individual replication origins, parameterized by differential firing potential (Rhind et al. 2010; Yang et al. 2010), still predicts overall replication timing profiles that are consistent with observed ensemble measurements.

A main limitation of DNA combing, however, is throughput: a typical experiment involves two nucleotide analog pulses, each of which requires a distinct antibody detection step. In addition, combing alone cannot determine where the observed replication initiation events are located within the genome. Approaches that combine DNA fiber analysis with fluorescence in situ hybridization (FISH) of sequence-specific probes

(Lebofsky et al. 2006; Pasero et al. 2002), most notably a technique called SMARD (single-molecule analysis of replicating DNA; (Norio et al. 2005; Norio and Schildkraut 2001)), overcome this last restriction yet introduce further experimental challenges.

Recently, two strategies have been put forward for single-molecule replication timing analysis on a genomic scale. The first, optical mapping with microfluidic nanochannels, attempts to alleviate the bottlenecks that reduce the throughput of traditional DNA combing experiments. The use of fluorescently tagged nucleotides instead of nucleotide analogs dramatically speeds up detection of nucleotide incorporation (i.e., nascent DNA replication), obviating the need for antibody detection steps prior to imaging (De Carli et al. 2016; Lacroix et al. 2016). In addition, fluorescent tag locations can be mapped to the genome by treating stretched DNA molecules with a nicking endonuclease and identifying restriction patterns (De Carli et al. 2016; Xiao et al. 2007). Together, these innovations make it practical to automate DNA combing for higher throughput: DNA fibers can be imaged as they are flowed (and stretched) through a microfluidic nanochannel and retrospectively mapped back to the reference genome by their endonuclease fingerprints (De Carli et al. 2018; Lacroix et al. 2016). This strategy has been employed to develop a high-coverage single-molecule map of origins in *Xenopus* egg extracts (De Carli et al. 2018) and to study early-firing origins in synchronized, aphidicolin-treated HeLa S3 cells (Klein et al. 2017). The vast majority of origins detected in single molecules by optical mapping overlap with origins called by OK-seq and are enriched for ORC1 binding. Additionally, origins can be identified that are used by as few as 1% of the cells in a population (Klein et al. 2017). However, this approach requires specialized equipment, and, more fundamentally, resolution is inherently limited by the distribution of endonuclease cleavage sites in the genome.

The second strategy for single-molecule origin detection uses DNA sequencing, which enables direct genomic mapping. The Oxford Nanopore Technologies sequencer produces reads that can reach upward of 100 Kb in length from single DNA molecules without the need for DNA amplification. Within the sequencer, a single strand of DNA is translocated through a protein nanopore by electrophoresis. As the DNA traverses the nanopore, characteristic changes in ionic current can be detected and interpreted as sequence readout (Clarke et al. 2009). It has been shown that nanopore sequencing can distinguish BrdU from thymidine and, thus, replicated from unreplicated DNA. This strategy has been used to map BrdU tracks to early-replicating origins in yeast (Hennion et al. 2018; Muller et al. 2019). Intriguingly, ~20% of origins observed at the single-molecule level were *not* detected in population-scale sequencing (Muller et al. 2019). The long-read approach generates large amounts of contextual information, making it easier to map reads to the genome and providing information on how neighboring origins may impact one another. However, it remains technically challenging to distinguish nucleotide analogs from canonical nucleotides. For instance, nanopore base calling relies on the shifts in ionic current characteristic of *k*-mers, rather than single nucleotides – and BrdU can only be distinguished from thymidine in certain 6-mer contexts (Muller et al. 2019). Another concern is that this strategy compares BrdU-labeled DNA to non-labeled DNA; biases in base calling and mappability between BrdU and the native thymidine must be accounted for (Muller et al. 2019). The recent demonstration that 11 different thymidine analogs can be detected using Oxford Nanopore Technologies sequencing provides a promising avenue for using analog combinations to identify the location and direction of DNA synthesis events (Georgieva et al. 2019).

Finally, recent advances have extended DNA copy number-based replication timing assays (that use short-read sequencing depth) to single cells. Measuring genome-wide replication timing in single cells is appealing: direct measurements can be made without cell synchronization or other perturbations, and the set of possible copy number values is theoretically limited to integers. Single-cell replication timing requires isolation of individual cells, which has so far been done using flow cytometry. A greater challenge is that, unlike with nanopore sequencing, short-read DNA sequencing requires whole-genome amplification (WGA), and there is rightful concern that even small biases in amplification will be exponentially magnified given the paucity of starting material. To date, several WGA strategies have been developed (reviewed in (Blainey 2013; Gawad et al. 2016)). Some, like degenerate-oligonucleotide-primed PCR (DOP-PCR; (Arneson et al. 2008; Telenius et al. 1992)), amplify the genome exponentially. Other methods like multiple annealing and looping-based amplification cycles (MALBAC; (Zong et al. 2012)) combine reduced amplification bias from the low-temperature isothermal MDA polymerase

with the higher amplification efficiency of traditional PCR once the amount of input template has been increased sufficiently. More recently, an approach called linear amplification via transposon insertion (LIANTI) was developed, which uses in vitro RNA transcription to linearly amplify the genome (Chen et al. 2017).

The possibility that single cells contain sufficient copy number variation between early- and late-replicating regions to be detected on a genomic level was first proposed by analyzing microarray data of single lymphoblastoid cells in S phase (Van der Aa et al. 2013). Much higher resolution, however, has been obtained using LIANTI and DNA sequencing of 11 single human BJ fibroblast cells in early S phase (Chen et al. 2017). Although replication profiles were correlated between cells and the average replication timing correlated well with a Repli-seq ensemble profile, there was also strong evidence of discrepant local copy number between cells, suggesting stochasticity of replication timing. In contrast, more recent studies that used DOP-PCR to amplify DNA from hundreds of mouse embryonic stem cells found limited heterogeneity between cells (Dileep and Gilbert 2018; Miura et al. 2019; Takahashi et al. 2019). This heterogeneity may be non-uniformly distributed in the cell cycle: Dileep and Gilbert reported that variability in replication timing was comparable between early- and late-firing regions, while Takahashi et al. claim that heterogeneity peaks in mid-S phase with lower levels observed in early and late S phase. These studies also showed that single-cell replication timing profiling can distinguish between different cell types and even between the two copies of each chromosome. It cannot, however, identify individual replication initiation events. Future improvements in data resolution, which will likely require minimally biased amplification regimes such as LIANTI, may enable high-resolution single-cell replication timing analysis that could explicitly examine the activity of replication origins.

At present, the primary drawback of short-read sequencing for single-cell replication analysis is scalability: applying standard library preparation methods to single cells is laborious. This limits the size of available datasets, constraining the conclusions that can be drawn from them. We anticipate that further improvements in barcoding strategies (e.g., Vitak et al. 2017; Yin et al. 2019; Zahn et al. 2017) will allow larger numbers of cells to be pooled during time-consuming steps of library preparation and that increased accessibility of microfluidic devices will enable further automation of this process (Davis Bell et al. 2019; Laks et al. 2019; Zahn et al. 2017).

Conclusion

Recent years have brought about many exciting developments in the genomic interrogation of DNA replication. Replication timing profiling is becoming more accurate and comprehensive and is being applied evermore broadly. Innovative techniques for mapping replication origin sites are constantly being developed and improved. And single-molecule and single-cell approaches are adding a new dimension of DNA replication exploration. Together, these approaches promise to resolve decades-long questions regarding the regulation of DNA replication progression in eukaryotes. In parallel, similar developments are being made in measuring DNA damage and repair on a genomic scale (e.g., (Biernacka et al. 2018; Bryan et al. 2014; Canela et al. 2016; Hu et al. 2015; Mao et al. 2017)) as well as in ensemble and single-cell epigenomic methods. Combined analyses of these vast emerging data with DNA replication dynamics would enable to answer additional key questions regarding the molecular causes and consequences of the spatiotemporal replication timing program.

Author contributions MLH, DJM, and AK wrote the paper.

Funding information Work in the Koren lab is supported by grants DP2GM123495 from the National Institutes of Health and MCB-1921341 from the National Science Foundation. MLH is supported by the National Science Foundation Graduate Research Fellowship DGE-1650441.

References

Agier N, Delmas S, Zhang Q, Fleiss A, Jaszczyszyn Y, van Dijk E, Thermes C, Weigt M, Cosentino-Lagomarsino M, Fischer G (2018) The evolution of the temporal program of genome replication. Nat Commun *9*:2199

Aladjem MI, Redon CE (2017) Order from clutter: selective interactions at mammalian replication origins. Nat Rev Genet *18*:101–116

Almeida R, Fernandez-Justel JM, Santa-Maria C, Cadoret JC, Cano-Aroca L, Lombrana R, Herranz G, Agresti A, Gomez M (2018) Chromatin conformation regulates the coordination between DNA replication and transcription. Nat Commun *9*: 1590

Arneson, N., Hughes, S., Houlston, R., and Done, S. (2008). Whole-genome amplification by degenerate oligonucleotide primed PCR (DOP-PCR). CSH Protoc *2008*, pdb prot4919.

Artemov AV, Andrianova, Bazykin, Seplyarskiy (2019) POLD replicates both strands of small kilobase-long replication bubbles initiated at a majority of human replication origins. BioRxiv. https://doi.org/10.1101/174730

Bartholdy B, Mukhopadhyay R, Lajugie J, Aladjem MI, Bouhassira EE (2015) Allele-specific analysis of DNA replication origins in mammalian cells. Nat Commun 6:7051

Bell SP, Stillman B (1992) ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature *357*:128–134

Bensimon A, Simon A, Chiffaudel A, Croquette V, Heslot F, Bensimon D (1994) Alignment and sensitive detection of DNA by a moving interface. Science 265:2096–2098

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

Bester AC, Roniger M, Oren YS, Im MM, Sarni D, Chaoat M, Bensimon A, Zamir G, Shewach DS, Kerem B (2011) Nucleotide deficiency promotes genomic instability in early stages of cancer development. Cell *145*:435–446

Bielinsky AK, Gerbi SA (1998) Discrete start sites for DNA synthesis in the yeast ARS1 origin. Science 279:95–98

Biernacka A, Zhu Y, Skrzypczak M, Forey R, Pardo B, Grzelak M, Nde J, Mitra A, Kudlicki A, Crosetto N et al (2018) i-BLESS is an ultra-sensitive method for detection of DNA double-strand breaks. Commun Biol 1:181

Blainey PC (2013) The future is now: single-cell genomics of bacteria and archaea. FEMS Microbiol Rev 37:407–427

Blow JJ, Gillespie PJ, Francis D, Jackson DA (2001) Replication origins in Xenopus egg extract are 5-15 kilobases apart and are activated in clusters that fire at different times. J Cell Biol 152:15–25

Brody Y, Kimmerling RJ, Maruvka YE, Benjamin D, Elacqua JJ, Haradhvala NJ, Kim J, Mouw KW, Frangaj K, Koren A et al (2018) Quantification of somatic mutation flow across individual cell division events by lineage sequencing. Genome Res *28*:1901–1918

Bryan DS, Ransom M, Adane B, York K, Hesselberth JR (2014) High resolution mapping of modified DNA nucleobases using excision repair enzymes. Genome Res *24*:1534–1542

Cadoret JC, Meisch F, Hassan-Zadeh V, Luyten I, Guillet C, Duret L, Quesneville H, Prioleau MN (2008) Genome-wide studies highlight indirect links between human replication origins and gene regulation. Proc Natl Acad Sci U S A 105:15837–15842

Canela A, Sridharan S, Sciascia N, Tubbs A, Meltzer P, Sleckman BP, Nussenzweig A (2016) DNA breaks and end resection measured genome-wide by end sequencing. Mol Cell *63*:898–911

Cayrou C, Ballester B, Peiffer I, Fenouil R, Coulombe P, Andrau JC, van Helden J, Mechali M (2015) The chromatin environment shapes DNA replication origin organization and defines origin classes. Genome Res *25*:1873–1885

Cayrou C, Coulombe P, Vigneron A, Stanojcic S, Ganier O, Peiffer I, Rivals E, Puy A, Laurent-Chabalier S, Desprat R 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

Chen C, Xing D, Tan L, Li H, Zhou G, Huang L, Xie XS (2017) Single-cell whole-genome analyses by linear amplification via transposon insertion (LIANTI). Science 356:189–194

Chen CL, Rappailles A, Duquenne L, Huvet M, Guilbaud G, Farinelli L, Audit B, d'Aubenton-Carafa, Y., Arneodo, A., Hyrien, O., et al. (2010) Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. Genome Res 20:447–457

Chen YH, Keegan S, Kahli M, Tonzi P, Fenyo D, Huang TT, Smith DJ (2019) Transcription shapes DNA replication initiation and termination in human cells. Nat Struct Mol Biol 26:67–77

Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H (2009) Continuous base identification for single-molecule nanopore DNA sequencing. Nat Nanotechnol 4:265–270

Clausen AR, Lujan SA, Burkholder AB, Orebaugh CD, Williams JS, Clausen MF, Malc EP, Mieczkowski PA, Fargo DC, Smith DJ et al (2015) Tracking replication enzymology in vivo by genome-wide mapping of ribonucleotide incorporation. Nat Struct Mol Biol 22:185–191

Comoglio F, Schlumpf T, Schmid V, Rohs R, Beisel C, Paro R (2015) High-resolution profiling of Drosophila replication start sites reveals a DNA shape and chromatin signature of metazoan origins. Cell Rep 11:821–834

Concia L, Brooks AM, Wheeler E, Zynda GJ, Wear EE, LeBlanc C, Song J, Lee TJ, Pascuzzi PE, Martienssen RA et al (2018) Genome-wide analysis of the Arabidopsis replication timing program. Plant Physiol *176*:2166–2185

Czajkowsky DM, Liu J, Hamlin JL, Shao Z (2008) DNA combing reveals intrinsic temporal disorder in the replication of yeast chromosome VI. J Mol Biol 375:12–19

Daigaku Y, Keszthelyi A, Muller CA, Miyabe I, Brooks T, Retkute R, Hubank M, Nieduszynski CA, Carr AM (2015) A global profile of replicative polymerase usage. Nat Struct Mol Biol *22*:192–198

Davis Bell, A., Mello, Nemesh, Brumbaugh, Wysoker, and McCarroll (2019). Insights about variation in meiosis from 31,228 human sperm genomes. BioRxiv https://doi.org/10.1101/625202.

De Carli F, Gaggioli V, Millot GA, Hyrien O (2016) Single-molecule, antibody-free fluorescent visualisation of replication tracts along barcoded DNA molecules. Int J Dev Biol 60:297–304

De Carli F, Menezes B, Barbe G, Hyrien O (2018) High-throughput optical mapping of replicating DNA. Small Methods. https://doi.org/10.1002/smtd.201800146

de Moura AP, Retkute R, Hawkins M, Nieduszynski CA (2010) Mathematical modelling of whole chromosome replication. Nucleic Acids Res 38:5623–5633

Dellino GI, Cittaro D, Piccioni R, Luzi L, Banfi S, Segalla S, Cesaroni M, Mendoza-Maldonado R, Giacca M, Pelicci PG (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

Desprat R, Thierry-Mieg D, Lailler N, Lajugie J, Schildkraut C, Thierry-Mieg J, Bouhassira EE (2009) Predictable dynamic program of timing of DNA replication in human cells. Genome Res 19:2288–2299

Diffley JF, Cocker JH (1992) Protein-DNA interactions at a yeast replication origin. Nature *357*:169–172

Dileep V, Gilbert DM (2018) Single-cell replication profiling to measure stochastic variation in mammalian replication timing. Nat Commun 9:427

Du Q, Bert SA, Armstrong NJ, Caldon CE, Song JZ, Nair SS, Gould CM, Luu PL, Peters T, Khoury A et al (2019) Replication timing and epigenome remodelling are associated with the nature of chromosomal rearrangements in cancer. Nat Commun 10:416

Fangman WL, Brewer BJ (1991) Activation of replication origins within yeast chromosomes. Annu Rev Cell Biol *7*:375–402

Farkash-Amar S, Lipson D, Polten A, Goren A, Helmstetter C, Yakhini Z, Simon I (2008) Global organization of replication time zones of the mouse genome. Genome Res *18*:1562–1570

Farkash-Amar S, Simon I (2010) Genome-wide analysis of the replication program in mammals. Chromosom Res 18:115–125

Feng W, Collingwood D, Boeck ME, Fox LA, Alvino GM, Fangman WL, Raghuraman MK, Brewer BJ (2006) Genomic mapping of single-stranded DNA in hydroxyurea-challenged yeasts identifies origins of replication. Nat Cell Biol 8:148–155

Foss EJ, Sripathy G-S, Kwak T, Lao, Bedalov (2019) Chromosomal Mcm2-7 distribution is a primary driver of genome replication timing in budding yeast, fission yeast and mouse. BioRxiv. https://doi.org/10.1101/737742

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*:725–735

Frum RA, Khondker ZS, Kaufman DG (2009) Temporal differences in DNA replication during the S phase using single fiber analysis of normal human fibroblasts and glioblastoma T98G cells. Cell Cycle 8:3133–3148

Gaboriaud J, Wu PJ (2019) Insights into the link between the organization of DNA replication and the mutational landscape. Genes (Basel) 10

Gartler SM, Goldstein L, Tyler-Freer SE, Hansen RS (1999) The timing of XIST replication: dominance of the domain. Hum Mol Genet *8*:1085–1089

Gawad C, Koh W, Quake SR (2016) Single-cell genome sequencing: current state of the science. Nat Rev Genet 17:175–188

Gdula MR, Nesterova TB, Pintacuda G, Godwin J, Zhan Y, Ozadam H, McClellan M, Moralli D, Krueger F, Green CM et al (2019) The non-canonical SMC protein SmcHD1 antagonises TAD formation and compartmentalisation on the inactive X chromosome. Nat Commun 10:30

Georgieva D, Liu Q, Wang K, Egli D (2019) Detection of base analogs incorporated during DNA replication by nanopore sequencing. BioRxiv. https://doi.org/10.1101/549220

Gilbert DM (2004) In search of the holy replicator. Nat Rev Mol Cell Biol *5*:848–855

Gispan A, Carmi M, Barkai N (2014) Checkpoint-independent scaling of the Saccharomyces cerevisiae DNA replication program. BMC Biol 12:79

Gispan A, Carmi M, Barkai N (2017) Model-based analysis of DNA replication profiles: predicting replication fork velocity and initiation rate by profiling free-cycling cells. Genome Res 27:310–319

Gitlin AD, Mayer CT, Oliveira TY, Shulman Z, Jones MJ, Koren A, Nussenzweig MC (2015) HUMORAL IMMUNITY. T cell help controls the speed of the cell cycle in germinal center B cells. Science 349:643–646

Guilbaud G, Rappailles A, Baker A, Chen CL, Arneodo A, Goldar A, d'Aubenton-Carafa Y, Thermes C, Audit B, Hyrien O (2011) Evidence for sequential and increasing activation of replication origins along replication timing gradients in the human genome. PLoS Comput Biol *7*:e1002322

Hansen RS, Canfield TK, Gartler SM (1995) Reverse replication timing for the XIST gene in human fibroblasts. Hum Mol Genet *4*:813–820

Hansen RS, Canfield TK, Lamb MM, Gartler SM, Laird CD (1993) Association of fragile X syndrome with delayed replication of the FMR1 gene. Cell *73*:1403–1409

Hansen RS, Thomas S, Sandstrom R, Canfield TK, Thurman RE, Weaver M, Dorschner MO, Gartler SM, Stamatoyannopoulos JA (2010) Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proc Natl Acad Sci U S A 107:139–144

Haradhvala NJ, Polak P, Stojanov P, Covington KR, Shinbrot E, Hess JM, Rheinbay E, Kim J, Maruvka YE, Braunstein LZ et al (2016) Mutational strand asymmetries in cancer genomes reveal mechanisms of DNA damage and repair. Cell 164:538–549

Hennion M, Arbona, Cruaud, Proux, Tallec L, Novikova, Engelen, Lemainque, Audit, Hyrien (2018) Mapping DNA replication with nanopore sequencing. BioRxiv. https://doi.org/10.1101/426858

Herrick J, Stanislawski P, Hyrien O, Bensimon A (2000) Replication fork density increases during DNA synthesis in X. laevis egg extracts. J Mol Biol 300:1133–1142

Hiratani I, Ryba T, Itoh M, Yokochi T, Schwaiger M, Chang CW, Lyou Y, Townes TM, Schubeler D, Gilbert DM (2008) Global reorganization of replication domains during embryonic stem cell differentiation. PLoS Biol 6:e245

Hoshina S, Yura K, Teranishi H, Kiyasu N, Tominaga A, Kadoma H, Nakatsuka A, Kunichika T, Obuse C, Waga S (2013) Human origin recognition complex binds preferentially to G-quadruplex-preferable RNA and single-stranded DNA. J Biol Chem 288:30161–30171

Hu J, Adar S, Selby CP, Lieb JD, Sancar A (2015) Genome-wide analysis of human global and transcription-coupled excision repair of UV damage at single-nucleotide resolution. Genes Dev *29*:948–960

Huberman JA, Riggs AD (1968) On the mechanism of DNA replication in mammalian chromosomes. J Mol Biol 32:327–341

Huvet M, Nicolay S, Touchon M, Audit B, d'Aubenton-Carafa, Y., Arneodo, A., and Thermes, C. (2007) Human gene organization driven by the coordination of replication and transcription. Genome Res 17:1278–1285

Hyrien O (2015) Peaks cloaked in the mist: the landscape of mammalian replication origins. J Cell Biol 208:147–160

Jeon Y, Bekiranov S, Karnani N, Kapranov P, Ghosh S, MacAlpine D, Lee C, Hwang DS, Gingeras TR, Dutta A (2005) Temporal profile of replication of human chromosomes. Proc Natl Acad Sci U S A *102*:6419–6424

Jinks-Robertson S, Klein HL (2015) Ribonucleotides in DNA: hidden in plain sight. Nat Struct Mol Biol 22:176–178

Jodkowska K, Pancaldi A, Rigau G-C, Fernández-Justel R-A, Rubio-Camarillo P, C.-d.S., Pisano et al (2019) Three-dimensional connectivity and chromatin environment mediate the activation efficiency of mammalian DNA replication origins. BioRxiv. https://doi.org/10.1101/644971

Karnani N, Taylor CM, Malhotra A, Dutta A (2010) Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. Mol Biol Cell 21:393–404

Keller H, Kiosze K, Sachsenweger J, Haumann S, Ohlenschlager O, Nuutinen T, Syvaoja JE, Gorlach M, Grosse F, Pospiech H (2014) The intrinsically disordered amino-terminal region of human RecQL4: multiple DNA-binding domains confer annealing, strand exchange and G4 DNA binding. Nucleic Acids Res 42:12614–12627

Klein HL (2017) Genome instabilities arising from ribonucleotides in DNA. DNA Repair (Amst) *56*:26–32

Klein K, Wang B, Chan Z, Weng H, Chen G, Rhind (2017) Genome-wide identification of early-firing human replication origins by optical replication mapping. BioRxiv. https://doi.org/10.1101/214841

Knott SR, Peace JM, Ostrow AZ, Gan Y, Rex AE, Viggiani CJ, Tavare S, Aparicio OM (2012) Forkhead transcription factors establish origin timing and long-range clustering in S. cerevisiae. Cell 148:99–111

Koren A, Handsaker RE, Kamitaki N, Karlic R, Ghosh S, Polak P, Eggan K, McCarroll SA (2014) Genetic variation in human DNA replication timing. Cell 159:1015–1026

Koren A, McCarroll SA (2014) Random replication of the inactive X chromosome. Genome Res *24*:64–69

Koren A, Polak P, Nemesh J, Michaelson JJ, Sebat J, Sunyaev SR, McCarroll SA (2012) Differential relationship of DNA replication timing to different forms of human mutation and variation. Am J Hum Genet 91:1033–1040

Koren A, Soifer I, Barkai N (2010a) MRC1-dependent scaling of the budding yeast DNA replication timing program. Genome Res *20*:781–790

Koren A, Tsai HJ, Tirosh I, Burrack LS, Barkai N, Berman J (2010b) Epigenetically-inherited centromere and neocentromere DNA replicates earliest in S-phase. PLoS Genet *6*:e1001068

Kumagai A, Dunphy WG (2017) MTBP, the partner of Treslin, contains a novel DNA-binding domain that is essential for proper initiation of DNA replication. Mol Biol Cell 28:2998–3012

Kunnev D, Freeland A, Qin M, Leach RW, Wang J, Shenoy RM, Pruitt SC (2015) Effect of minichromosome maintenance protein 2 deficiency on the locations of DNA replication origins. Genome Res *25*:558–569

Labit H, Perewoska I, Germe T, Hyrien O, Marheineke K (2008) DNA replication timing is deterministic at the level of chromosomal domains but stochastic at the level of replicons in Xenopus egg extracts. Nucleic Acids Res 36:5623–5634

Lacroix J, Pelofy S, Blatche C, Pillaire MJ, Huet S, Chapuis C, Hoffmann JS, Bancaud A (2016) Analysis of DNA replication by optical mapping in nanochannels. Small *12*:5963–5970

Laks E, McPherson A, Zahn H, Lai D, Steif A, Brimhall J, Biele J, Wang B, Masud T, Ting J et al (2019) Clonal decomposition and DNA replication states defined by scaled single-cell genome sequencing. Cell 179:1207–1221 e1222

Lamm N, Maoz K, Bester AC, Im MM, Shewach DS, Karni R, Kerem B (2015) Folate levels modulate oncogene-induced replication stress and tumorigenicity. EMBO Mol Med 7:1138–1152

Langley AR, Graf S, Smith JC, Krude T (2016) Genome-wide identification and characterisation of human DNA replication origins by initiation site sequencing (ini-seq). Nucleic Acids Res 44:10230–10247

Lebofsky R, Heilig R, Sonnleitner M, Weissenbach J, Bensimon A (2006) DNA replication origin interference increases the spacing between initiation events in human cells. Mol Biol Cell *17*:5337–5345

Li B, Su T, Ferrari R, Li JY, Kurdistani SK (2014) A unique epigenetic signature is associated with active DNA replication loci in human embryonic stem cells. Epigenetics 9:257–267

Liachko I, Youngblood RA, Keich U, Dunham MJ (2013) High-resolution mapping, characterization, and optimization of autonomously replicating sequences in yeast. Genome Res 23:698–704

Liachko I, Youngblood RA, Tsui K, Bubb KL, Queitsch C, Raghuraman MK, Nislow C, Brewer BJ, Dunham MJ (2014) GC-rich DNA elements enable replication origin activity in the methylotrophic yeast Pichia pastoris. PLoS Genet 10:e1004169

Lombrana R, Alvarez A, Fernandez-Justel JM, Almeida R, Poza-Carrion C, Gomes F, Calzada A, Requena JM, Gomez M (2016) Transcriptionally driven DNA replication program of the human parasite Leishmania major. Cell Rep 16:1774–1786

Lynch KL, Alvino GM, Kwan EX, Brewer BJ, Raghuraman MK (2019) The effects of manipulating levels of replication initiation factors on origin firing efficiency in yeast. PLoS Genet *15*:e1008430

MacAlpine HK, Gordan R, Powell SK, Hartemink AJ, MacAlpine DM (2010) Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading. Genome Res *20*:201–211

Macheret M, Halazonetis TD (2018) Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress. Nature *555*:112–116

Maine GT, Sinha P, Tye BK (1984) Mutants of S. cerevisiae defective in the maintenance of minichromosomes. Genetics *106*:365–385

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 30:4805–4814

Mao P, Brown AJ, Malc EP, Mieczkowski PA, Smerdon MJ, Roberts SA, Wyrick JJ (2017) Genome-wide maps of alkylation damage, repair, and mutagenesis in yeast reveal mechanisms of mutational heterogeneity. Genome Res *27*:1674–1684

Martin MM, Ryan M, Kim R, Zakas AL, Fu H, Lin CM, Reinhold WC, Davis SR, Bilke S, Liu H et al (2011) Genome-wide depletion of replication initiation events in highly transcribed regions. Genome Res 21:1822–1832

Masai H, Kakusho N, Fukatsu R, Ma Y, Iida K, Kanoh Y, Nagasawa K (2018) Molecular architecture of G-

quadruplex structures generated on duplex Rif1-binding sequences. J Biol Chem 293:17033–17049

Massey DJ, Kim D, Brooks KE, Smolka MB, Koren A (2019) Next-generation sequencing enables spatiotemporal resolution of human centromere replication timing. Genes (Basel) 10

Massip F, Laurent M, Brossas C, Fernandez-Justel JM, Gomez M, Prioleau MN, Duret L, Picard F (2019) Evolution of replication origins in vertebrate genomes: rapid turnover despite selective constraints. Nucleic Acids Res 47:5114–5125

McGuffee SR, Smith DJ, Whitehouse I (2013) Quantitative, genome-wide analysis of eukaryotic replication initiation and termination. Mol Cell *50*:123–135

Mesner LD, Valsakumar V, Cieslik M, Pickin R, Hamlin JL, Bekiranov S (2013) Bubble-seq analysis of the human genome reveals distinct chromatin-mediated mechanisms for regulating early- and late-firing origins. Genome Res 23: 1774–1788

Michalet X, Ekong R, Fougerousse F, Rousseaux S, Schurra C, Hornigold N, van Slegtenhorst M, Wolfe J, Povey S, Beckmann JS et al (1997) Dynamic molecular combing: stretching the whole human genome for high-resolution studies. Science 277:1518–1523

Miotto B, Ji Z, Struhl K (2016) Selectivity of ORC binding sites and the relation to replication timing, fragile sites, and deletions in cancers. Proc Natl Acad Sci U S A 113:E4810–E4819

Miura H, Takahashi S, Poonperm R, Tanigawa A, Takebayashi SI, Hiratani I (2019) Single-cell DNA replication profiling identifies spatiotemporal developmental dynamics of chromosome organization. Nat Genet 51:1356–1368

Moller HD, Mohiyuddin M, Prada-Luengo I, Sailani MR, Halling JF, Plomgaard P, Maretty L, Hansen AJ, Snyder MP, Pilegaard H et al (2018) Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. Nat Commun *9*:1069

Mukhopadhyay R, Lajugie J, Fourel N, Selzer A, Schizas M, Bartholdy B, Mar J, Lin CM, Martin MM, Ryan M et al (2014) Allele-specific genome-wide profiling in human primary erythroblasts reveal replication program organization. PLoS Genet 10:e1004319

Muller CA, Boemo MA, Spingardi P, Kessler BM, Kriaucionis S, Simpson JT, Nieduszynski CA (2019) Capturing the dynamics of genome replication on individual ultra-long nanopore sequence reads. Nat Methods 16:429–436

Muller CA, Hawkins M, Retkute R, Malla S, Wilson R, Blythe MJ, Nakato R, Komata M, Shirahige K, de Moura AP et al (2014) The dynamics of genome replication using deep sequencing. Nucleic Acids Res 42:e3

Muller CA, Nieduszynski CA (2012) Conservation of replication timing reveals global and local regulation of replication origin activity. Genome Res *22*:1953–1962

Nieduszynski CA, Knox Y, Donaldson AD (2006) Genome-wide identification of replication origins in yeast by comparative genomics. Genes Dev *20*:1874–1879

Norio P, Kosiyatrakul S, Yang Q, Guan Z, Brown NM, Thomas S, Riblet R, Schildkraut CL (2005) Progressive activation of DNA replication initiation in large domains of the immunoglobulin heavy chain locus during B cell development. Mol Cell *20*:575–587

Norio P, Schildkraut CL (2001) Visualization of DNA replication on individual Epstein-Barr virus episomes. Science 294: 2361–2364

Owiti N, Wei S, Bhagwat AS, Kim N (2018) Unscheduled DNA synthesis leads to elevated uracil residues at highly transcribed genomic loci in Saccharomyces cerevisiae. PLoS Genet 14:e1007516

Pasero P, Bensimon A, Schwob E (2002) Single-molecule analysis reveals clustering and epigenetic regulation of replication origins at the yeast rDNA locus. Genes Dev *16*:2479–2484

Patel PK, Arcangioli B, Baker SP, Bensimon A, Rhind N (2006) DNA replication origins fire stochastically in fission yeast. Mol Biol Cell *17*:308–316

Peace JM, Villwock SK, Zeytounian JL, Gan Y, Aparicio OM (2016) Quantitative BrdU immunoprecipitation method demonstrates that Fkh1 and Fkh2 are rate-limiting activators of replication origins that reprogram replication timing in G1 phase. Genome Res 26:365–375

Perkins TT, Dalal RV, Mitsis PG, Block SM (2003) Sequence-dependent pausing of single lambda exonuclease molecules. Science 301:1914–1918

Petryk N, Dalby M, Wenger A, Stromme CB, Strandsby A, Andersson R, Groth A (2018) MCM2 promotes symmetric inheritance of modified histones during DNA replication. Science *361*:1389–1392

Petryk N, Kahli M, d'Aubenton-Carafa Y, Jaszczyszyn Y, Shen Y, Silvain M, Thermes C, Chen CL, Hyrien O (2016) Replication landscape of the human genome. Nat Commun *7*:10208

Picard F, Cadoret JC, Audit B, Arneodo A, Alberti A, Battail C, Duret L, Prioleau MN (2014) The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells. PLoS Genet 10:e1004282

Piunti A, Rossi A, Cerutti A, Albert M, Jammula S, Scelfo A, Cedrone L, Fragola G, Olsson L, Koseki H et al (2014) Polycomb proteins control proliferation and transformation independently of cell cycle checkpoints by regulating DNA replication. Nat Commun 5:3649

Polak P, Arndt PF (2009) Long-range bidirectional strand asymmetries originate at CpG islands in the human genome. Genome Biol Evol 1:189–197

Pomerantz RT, O'Donnell M (2008) The replisome uses mRNA as a primer after colliding with RNA polymerase. Nature 456: 762–766

Pope BD, Gilbert DM (2013) The replication domain model: regulating replicon firing in the context of large-scale chromosome architecture. J Mol Biol 425:4690–4695

Pope BD, Ryba T, Dileep V, Yue F, Wu W, Denas O, Vera DL, Wang Y, Hansen RS, Canfield TK et al (2014) Topologically associating domains are stable units of replication-timing regulation. Nature 515:402–405

Prioleau MN (2017) G-quadruplexes and DNA replication origins. Adv Exp Med Biol *1042*:273–286

Prioleau MN, MacAlpine DM (2016) DNA replication origins-where do we begin? Genes Dev 30:1683–1697

Prorok P, Artufel M, Aze A, Coulombe P, Peiffer I, Lacroix L, Guedin A, Mergny JL, Damaschke J, Schepers A et al (2019) Involvement of G-quadruplex regions in mammalian replication origin activity. Nat Commun 10:3274

Raghuraman MK, Brewer BJ (2010) Molecular analysis of the replication program in unicellular model organisms. Chromosom Res 18:19–34

Raghuraman MK, Winzeler EA, Collingwood D, Hunt S, Wodicka L, Conway A, Lockhart DJ, Davis RW, Brewer BJ, Fangman WL (2001) Replication dynamics of the yeast genome. Science 294:115–121

Ravoityte B, Wellinger RE (2017) Non-canonical replication initiation: you're fired! Genes (Basel) *8*

Reijns MAM, Kemp H, Ding J, de Proce SM, Jackson AP, Taylor MS (2015) Lagging-strand replication shapes the mutational landscape of the genome. Nature *518*:502–506

Rhind N, Gilbert DM (2013) DNA replication timing. Cold Spring Harb Perspect Biol *5*:a010132

Rhind N, Yang SC, Bechhoefer J (2010) Reconciling stochastic origin firing with defined replication timing. Chromosom Res *18*:35–43

Rivera-Mulia JC, Buckley Q, Sasaki T, Zimmerman J, Didier RA, Nazor K, Loring JF, Lian Z, Weissman S, Robins AJ et al (2015) Dynamic changes in replication timing and gene expression during lineage specification of human pluripotent stem cells. Genome Res *25*:1091–1103

Rivera-Mulia JC, Dimond A, Vera D, Trevilla-Garcia C, Sasaki T, Zimmerman J, Dupont C, Gribnau J, Fraser P, Gilbert DM (2018) Allele-specific control of replication timing and genome organization during development. Genome Res 28: 800–811

Rivera-Mulia JC, Sasaki T, Trevilla-Garcia C, Nakamichi N, Knapp D, Hammond CA, Chang BH, Tyner JW, Devidas M, Zimmerman J et al (2019) Replication timing alterations in leukemia affect clinically relevant chromosome domains. Blood Adv *3*:3201–3213

Rudolph CJ, Upton AL, Stockum A, Nieduszynski CA, Lloyd RG (2013) Avoiding chromosome pathology when replication forks collide. Nature 500:608–611

Ryba T, Hiratani I, Lu J, Itoh M, Kulik M, Zhang J, Schulz TC, Robins AJ, Dalton S, Gilbert DM (2010) Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. Genome Res *20*:761–770

Schubeler D, Scalzo D, Kooperberg C, van Steensel B, Delrow J, Groudine M (2002) Genome-wide DNA replication profile for Drosophila melanogaster: a link between transcription and replication timing. Nat Genet 32:438–442

Sequeira-Mendes J, Diaz-Uriarte R, Apedaile A, Huntley D, Brockdorff N, Gomez M (2009) Transcription initiation activity sets replication origin efficiency in mammalian cells. PLoS Genet 5:e1000446

Sequeira-Mendes J, Vergara Z, Peiro R, Morata J, Araguez I, Costas C, Mendez-Giraldez R, Casacuberta JM, Bastolla U, Gutierrez C (2019) Differences in firing efficiency, chromatin, and transcription underlie the developmental plasticity of the Arabidopsis DNA replication origins. Genome Res 29: 784–797

Shibata Y, Kumar P, Layer R, Willcox S, Gagan JR, Griffith JD, Dutta A (2012) Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. Science 336:82–86

Shinbrot E, Henninger EE, Weinhold N, Covington KR, Goksenin AY, Schultz N, Chao H, Doddapaneni H, Muzny DM, Gibbs RA et al (2014) Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. Genome Res 24:1740–1750

Siefert JC, Georgescu C, Wren JD, Koren A, Sansam CL (2017) DNA replication timing during development anticipates transcriptional programs and parallels enhancer activation. Genome Res 27:1406–1416

Smith DJ, Whitehouse I (2012) Intrinsic coupling of lagging-strand synthesis to chromatin assembly. Nature 483:434–438

Smith OK, Kim R, Fu H, Martin MM, Lin CM, Utani K, Zhang Y, Marks AB, Lalande M, Chamberlain S et al (2016) Distinct epigenetic features of differentiation-regulated replication origins. Epigenetics Chromatin *9*:18

Stuckey R, Garcia-Rodriguez N, Aguilera A, Wellinger RE (2015) Role for RNA:DNA hybrids in origin-independent replication priming in a eukaryotic system. Proc Natl Acad Sci U S A 112:5779–5784

Sueoka N, Yoshikawa H (1965) The chromosome of Bacillus subtilis. I. Theory of marker frequency analysis. Genetics *52*:747–757

Sugimoto N, Maehara K, Yoshida K, Ohkawa Y, Fujita M (2018) Genome-wide analysis of the spatiotemporal regulation of firing and dormant replication origins in human cells. Nucleic Acids Res 46:6683–6696

Takahashi S, Miura H, Shibata T, Nagao K, Okumura K, Ogata M, Obuse C, Takebayashi SI, Hiratani I (2019) Genome-wide stability of the DNA replication program in single mammalian cells. Nat Genet 51:529–540

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*:2055–2063

Telenius H, Carter NP, Bebb CE, Nordenskjold M, Ponder BA, Tunnacliffe A (1992) Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. Genomics 13:718–725

Tomkova M, Tomek J, Kriaucionis S, Schuster-Bockler B (2018) Mutational signature distribution varies with DNA replication timing and strand asymmetry. Genome Biol 19:129

Touchon M, Nicolay S, Audit B, Brodie of Brodie, E.B, d'Aubenton-Carafa Y, Arneodo A, Thermes C (2005) Replication-associated strand asymmetries in mammalian genomes: toward detection of replication origins. Proc Natl Acad Sci U S A 102:9836–9841

Tubbs A, Sridharan S, van Wietmarschen N, Maman Y, Callen E, Stanlie A, Wu W, Wu X, Day A, Wong N et al (2018) Dual roles of poly(dA:dT) tracts in replication initiation and fork collapse. Cell 174:1127–1142 e1119

Tye BK (1999) MCM proteins in DNA replication. Annu Rev Biochem *68*:649–686

Urban JM, Foulk MS, Casella C, Gerbi SA (2015) The hunt for origins of DNA replication in multicellular eukaryotes. F1000Prime Rep 7, 30

Valton AL, Hassan-Zadeh V, Lema I, Boggetto N, Alberti P, Saintome C, Riou JF, Prioleau MN (2014) G4 motifs affect origin positioning and efficiency in two vertebrate replicators. EMBO J 33:732–746

Van der Aa N, Cheng J, Mateiu L, Zamani Esteki M, Kumar P, Dimitriadou E, Vanneste E, Moreau Y, Vermeesch JR, Voet T (2013) Genome-wide copy number profiling of single cells in S-phase reveals DNA-replication domains. Nucleic Acids Res 41:e66

Vashee S, Cvetic C, Lu W, Simancek P, Kelly TJ, Walter JC (2003) Sequence-independent DNA binding and replication initiation by the human origin recognition complex. Genes Dev 17:1894–1908

Vitak SA, Torkenczy KA, Rosenkrantz JL, Fields AJ, Christiansen L, Wong MH, Carbone L, Steemers FJ, Adey A (2017) Sequencing thousands of single-cell genomes with combinatorial indexing. Nat Methods *14*:302–308

Williams JS, Lujan SA, Kunkel TA (2016) Processing ribonucleotides incorporated during eukaryotic DNA replication. Nat Rev Mol Cell Biol *17*:350–363

Wong PG, Winter SL, Zaika E, Cao TV, Oguz U, Koomen JM, Hamlin JL, Alexandrow MG (2011) Cdc45 limits replicon usage from a low density of preRCs in mammalian cells. PLoS One 6:e17533

Woodfine K, Fiegler H, Beare DM, Collins JE, McCann OT, Young BD, Debernardi S, Mott R, Dunham I, Carter NP (2004) Replication timing of the human genome. Hum Mol Genet 13:191–202

Wu X, Kabalane H, Kahli M, Petryk N, Laperrousaz B, Jaszczyszyn Y, Drillon G, Nicolini FE, Perot G, Robert A et al (2018) Developmental and cancer-associated plasticity of DNA replication preferentially targets GC-poor, lowly expressed and late-replicating regions. Nucleic Acids Res 46:10157–10172

Wyrick JJ, Aparicio JG, Chen T, Barnett JD, Jennings EG, Young RA, Bell SP, Aparicio OM (2001) Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins. Science 294:2357–2360

Xiao M, Phong A, Ha C, Chan TF, Cai D, Leung L, Wan E, Kistler AL, DeRisi JL, Selvin PR et al (2007) Rapid DNA mapping by fluorescent single molecule detection. Nucleic Acids Res 35:e16

Xu B, Clayton DA (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

Yabuki N, Terashima H, Kitada K (2002) Mapping of early firing origins on a replication profile of budding yeast. Genes Cells 7:781–789

Yaffe E, Farkash-Amar S, Polten A, Yakhini Z, Tanay A, Simon I (2010) Comparative analysis of DNA replication timing reveals conserved large-scale chromosomal architecture. PLoS Genet 6:e1001011

Yang SC, Rhind N, Bechhoefer J (2010) Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing. Mol Syst Biol *6*:404

Yang Y, Gu Q, Zhang Y, Sasaki T, Crivello J, O'Neill RJ, Gilbert DM, Ma J (2018) Continuous-trait probabilistic model for comparing multi-species functional genomic data. Cell Syst 7(208-218):e211

Yanga W, Lib X (2013) Next-generation sequencing of Okazaki fragments extracted from Saccharomyces cerevisiae. FEBS Lett 587:2441–2447

Yehuda Y, Blumenfeld B, Mayorek N, Makedonski K, Vardi O, Cohen-Daniel L, Mansour Y, Baror-Sebban S, Masika H, Farago M et al (2018) Germline DNA replication timing shapes mammalian genome composition. Nucleic Acids Res 46:8299–8310

Yin Y, Jiang Y, Lam KG, Berletch JB, Disteche CM, Noble WS, Steemers FJ, Camerini-Otero RD, Adey AC, Shendure J (2019) High-throughput single-cell sequencing with linear amplification. Mol Cell

Yoshida K, Bacal J, Desmarais D, Padioleau I, Tsaponina O, Chabes A, Pantesco V, Dubois E, Parrinello H, Skrzypczak M et al (2014) The histone deacetylases sir2 and rpd3 act on ribosomal DNA to control the replication program in budding yeast. Mol Cell 54:691–697

Yoshikawa H, Sueoka N (1963) Sequential replication of Bacillus subtilis chromosome. I. Comparison of marker frequencies in exponential and stationary growth phases. Proc Natl Acad Sci U S A 49:559–566

Yudkin D, Hayward BE, Aladjem MI, Kumari D, Usdin K (2014) Chromosome fragility and the abnormal replication of the FMR1 locus in fragile X syndrome. Hum Mol Genet *23*: 2940–2952

Zahn H, Steif A, Laks E, Eirew P, VanInsberghe M, Shah SP, Aparicio S, Hansen CL (2017) Scalable whole-genome single-cell library preparation without preamplification. Nat Methods 14:167–173

Zentner GE, Kasinathan S, Xin B, Rohs R, Henikoff S (2015) ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo. Nat Commun 6: 8733

Zong C, Lu S, Chapman AR, Xie XS (2012) Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science 338:1622–1626