Published in final edited form as:
Nat Rev Genet. 2010 October ; 11(10): 673–684. doi:10.1038/nrg2830.

Evaluating genome-scale approaches to eukaryotic DNA replication

David M. Gilbert
Department of Biological Science, Florida State University, 319 Stadium Drive, Tallahassee, Florida 32306, USA.

David M. Gilbert: gilbert@bio.fsu.edu

Abstract

Mechanisms regulating where and when eukaryotic DNA replication initiates remain a mystery. Recently, genome-scale methods have been brought to bear on this problem. The identification of replication origins and their associated proteins in yeasts is a well-integrated investigative tool, but corresponding data sets from multicellular organisms are scarce. By contrast, standardized protocols for evaluating replication timing have generated informative data sets for most eukaryotic systems. Here, I summarize the genome-scale methods that are most frequently used to analyse replication in eukaryotes, the kinds of questions each method can address and the technical hurdles that must be overcome to gain a complete understanding of the nature of eukaryotic replication origins.

The replication of DNA exactly once per cell cycle is fundamental to all biological systems, and the basic copying mechanism to ensure complete and accurate duplication of each DNA strand is highly conserved across all kingdoms[1]. Ironically, however, mechanisms regulating the initiation of replication are quite variable among living systems and remain largely obscure in most eukaryotic organisms[1]. Eukaryotes have long linear chromosomes that must initiate at many sites to ensure complete replication within the duration of S phase. As a fail-safe, pre-replication complexes (pre-RCs) are assembled at many more potential origins of replication than are used in any given cell cycle. A subset of these pre-RCs is chosen for initiation by as yet poorly understood mechanisms, and the rest serve as dormant origins: 'backups' used if the cell experiences problems completing replication[2–5]. These multiple origins of replication also fire in a defined temporal order during the course of S phase[6]. Hence, there are three layers of regulation to consider in eukaryotic DNA replication: the locations of potential origins (where), the selection of initiation sites from a pool of potential origins (which) and the time of firing during S phase (when).

The major impediment to understanding origins in most eukaryotic systems has been finding them in the first place. In yeast systems, specific DNA segments containing origin activity can be identified by their ability to confer autonomous replication on small circular plasmid

Competing interests statement: The author declares no competing financial interests.

Further Information
A complete protocol for genome-scale analysis of replication timing in mammalian cells: http://dx.doi.org/10.1038/nprot.2010.151
DNAReplication (a knowledge base for the eukaryotic
DNA replication community that includes a forum and a list of new papers): http://www.dnareplication.net
The ENCODE Project: http://www.genome.gov/10005107
OriDB (a catalogue of confirmed and predicted DNA replication origin sites): http://www.oridb.org
ReplicationDomain (an interactive database of replication timing and related chromosomal properties, supplemented with protocols): http://www.replicationdomain.org
All Links are Active in the Online Pdf

molecules. Unfortunately, when this assay is used in multicellular organisms, virtually any DNA fragment of sufficient size has origin activity[7]. For this reason, investigators have searched for the locations of replication origins in their natural chromosomal context using every imaginable characteristic of a replication origin (FIG. 1) and have thereby identified a few dozen animal replication origins[1,8,9]. In some cases, different approaches have led to conflicting data, so the list of universally accepted animal origins is even smaller. For these reasons, genome-scale methods offer new hope for identifying hundreds or thousands of origins that could provide much needed information regarding the recognition features of potential versus active origins. Indeed, genomic studies in budding yeast have revealed novel chromatin-based mechanisms of origin recognition[10] that are likely to be relevant to more complex organisms[11,12]. In mammalian cells, genome complexity and cell-to-cell heterogeneity in origin choice continue to impede efforts to identify origins. Origin site flexibility, however, does not seem to substantially affect the temporal order in which segments of the genome replicate, as this 'replication timing programme' has been successfully profiled from yeast to humans and seems to be more highly conserved among species than the positions of replication origins.

Here, I have divided genome-scale methods into those that identify potential origins versus sites of active initiation and those that query replication timing. I first describe the most common methods that have been used to date. I only briefly summarize the findings from these studies, and instead refer the reader to the many recent reviews on this topic, but I elaborate on the advantages and disadvantages of each method, the technical challenges for the future and the potential solutions to these challenges.

Mapping potential replication origins

What constitutes a replication origin in the eukaryotic world? Based on findings in bacteria and viruses, it was naturally thought to be a specific DNA sequence recognized by an initiator protein. The eukaryotic initiator (the heterohexameric origin recognition complex (ORC)) has been identified but it seems that, of our major model systems, only the budding yeast ORC shows DNA-sequence-specific binding, and even in budding yeast there is a lot of sequence flexibility[13]. In this section, I discuss progress using chromatin immunoprecipitation (ChIP) followed by microarray (ChIP–chip) or sequencing (ChIP–seq) to decipher the initial steps of origin recognition.

Insights from yeasts

Theoretically, finding replication origins seems simple enough. According to Jacob–Brenner dogma, there should be an initiator protein that binds the origin sequence to initiate replication nearby. Therefore, in principle, standard ChIP–chip or ChIP–seq methods for finding ORC should identify replication origins (FIG. 1a). What is more, ORC is a heterohexamer and seeds the assembly of the additional pre-RC proteins cell division cycle protein 6 (Cdc6), DNA replication factor Cdt1 and the heterohexameric mini-chromosome maintenance complex (MCM complex). Therefore, there are at least 14 different subunits that should be enriched at the same site for confirmation of an origin. Indeed, in budding yeast, studies using ChIP–chip[14,15] or ChIP–seq[10] identified approximately 300 ORC-bound sites, with general concordance between data sets. ORC prefers to bind between genes and, for the most part, wherever there is ORC there is almost always MCM[16]. In fission yeast, ChIP–chip showed colocalized Orc1 and Mcm6 at 460 sites in intergenic AT-rich stretches, and 80% of ORC sites colocalized with MCM[17]. In both yeasts, not all sites of ORC and MCM binding are active origins, which is consistent with the presence of dormant origins that are not used under normal growth conditions.

In budding yeast, a consensus DNA sequence for ORC binding has been identified and ChIP–seq for ORC has revealed that many consensus sequences are not bound by ORC[10,18]. A recent study comparing genome-wide nucleosome positioning data led to the finding that ORC-bound consensus sites are nucleosome-free regions (NFRs) flanked by phased nucleosomes (positioned at precise intervals from the NFR), whereas non-bound consensus sites are only moderately associated with NFRs and lack nucleosome phasing[10]. Moreover, it was clear from this study that not just any NFR can promote ORC binding. Instead, only a subset of NFRs with specific flanking sequence features — which allow the ORC to position nucleosomes with sufficient space for Mcm protein loading — can promote binding of the ORC. Origin consensus sequences have not been identified in any other eukaryotic organisms. Fission yeast can initiate replication within any sufficiently extensive stretch of AT-rich DNA[19,20], which is recognized by ORC via a species-specific AT-hook in the ORC4 subunit[21]. The rules governing nucleosome positioning may also differ in fission yeast, but origins still seem to align with NFRs[22], although improvements in the resolution of this alignment will be necessary to determine how well they align and whether they are flanked by phased nucleosomes.

Are nucleosome-free regions a universal feature of replication origins?

Metazoan origins are determined by a complex, poorly understood set of structural and topological features of DNA and chromatin in which DNA sequence motifs do not have an obvious role[23]. Nonetheless, in most metazoa, replication does initiate preferentially at specific sites, which can be either highly localized or part of a broad de-localized ‘initiation zone’[1,8,9]. In some cases, specific origin DNA sequences have been shown to be both necessary and sufficient to direct initiation of replication when inserted into some, but not all, ectopic locations[24–27]. However, sequences that do not normally function as origins, even bacterial sequences, can direct local initiation in certain ectopic locations[28]. These observations suggest some role for DNA sequence composition in positioning origins.

Box 1 | The elusive nature of replication origins

The hunt for replication origins in higher eukaryotes has produced few sites that investigators agree are genuine sites of initiation[23]. We are beginning to understand some of the reasons why they have been so difficult to find.

Origin recognition complex binding sites: a question of enrichment

- Origin recognition complex (ORC) binding sites have been mapped in yeasts and *Drosophila melanogaster* but not in mammals.
- ORC binding sites do not exhibit sequence specificity, except in budding yeast.
- ORC may be bound to different sites in different cells.
- ORC binding sites do not reveal origin efficiency or timing.
- Some ORC binding sites may not function as origins at all.
- Initiation may occur remotely from the ORC binding site.

Sites of initiation of replication: a question of heterogeneity

- There are many more potential initiation sites than are used.
- Most origins fire in a small fraction of cell cycles.
- Each individual cell fires from a different set of origins.
- Many initiations occur in ‘zones’ of clustered, inefficient sites.

- Synchronization methods can activate ‘dormant’ origins.

Metazoan ORC has no intrinsic affinity for any DNA sequence motifs, but negative supercoiling can increase its binding affinity by 30-fold[29,30]. *In vivo*, anything that removes a nucleosome to create an NFR can introduce negative supercoils. NFRs are encoded by physical properties of the primary DNA sequence that cannot be described by a sequence motif[31]. NFRs can also be influenced by local chromatin composition, a dynamic property of cellular state. Hence, genome-wide studies of ORC binding in budding yeast have suggested a potential origin determinant that is consistent with the elusive characteristics of origins in higher eukaryotes. Indeed, a genome-scale mapping of *Drosophila melanogaster* ORC localization found ORC localized to previously mapped NFRs[11], and sites of rapid nucleosome turnover were found to align with ORC[12]. Of the ~5,000 sites identified, ~66% localized near transcription start sites (TSSs) of active genes, but ORC localized to NFRs even when not associated with a TSS. Consistent with the *in vitro* binding data[29,30], no ORC-binding sequence motifs could be identified, but an *in silico* learning approach revealed a complex code of short sequences that could predict ORC binding, and these same sequences also predicted NFRs. These data make a compelling case that ORC targets to NFRs in both yeast and *D. melanogaster*, and suggest that properties of the primary DNA sequence that influence the location of origins may eventually be decipherable.

Challenges remain in mammalian cells

To date there are no published examples of ChIP–chip or ChIP–seq for any pre-RC proteins in mammalian cells owing to a lack of significant enrichment over background (BOX 1 and reviewed by [REF. 32]). ORC and MCM enrichment by ChIP has been detected at specific sites by directed PCR, but to do so the PCR primer sites — particularly the negative control unbound sites — must be carefully selected[32]. It is unlikely that epitopes for ChIP are buried deep in the chromatin because similar difficulties are experienced for any pre-RC protein. In a particularly revealing study, ChIP–chip in Epstein Barr virus (EBV)-positive human B cells readily detected ORC and MCM enrichment at the EBV origin oriP but did not detect comparable enrichment in the same ChIP preparations anywhere in the host genome[32]. As pre-RC proteins are assembled during G1 phase and evicted during S phase, cell synchronization before ChIP may improve their detection. However, it should be acknowledged that even in yeast and *D. melanogaster*, ORC enrichment by ChIP–chip is poor (often not more than twofold over background) relative to transcription-factor enrichment[10,11], so it is possible that pre-RCs are simply not highly enriched at particular sites, and discriminating their relative occupancy may be a function of genome complexity. In short, successful genome-scale ChIP methods for identifying pre-RC protein binding sites in mammalian cells would be a welcome addition, but the obstacles to this goal remain obscure.

Mapping replication initiation sites

Even with successful identification of all pre-RC binding sites, ChIP alone can only identify origin potential and cannot tell us where, when or how often an ORC-bound site will initiate replication. Catching an origin in the act of initiation is neither simple nor obvious, considering that intermediates are short lived, each site fires considerably less than one time per cell cycle and each site has a different firing efficiency. Even in budding yeast, most origins are used in less than 50% of cell cycles[33,34], and many fission yeast[20,35–37] and most metazoan[27,28,38–41] origins seem to fire in less than 10% of cell cycles (a systematic quantification of origin efficiencies in metazoans has yet to be performed). As a result, each cell in a population uses a different cohort of origins. However, there are at least some efficient origins, even in mammalian cells[42], and with so few validated origins in higher eukaryotes, confidence is

currently more important than numbers. A method for identifying only the most efficient origins would be a valuable contribution.

The hallmarks of an initiation event that lend themselves to detection are shown in FIG. 1b. All of these properties have been exploited to map origins at genome-scale. Three general approaches and the findings that have emerged from them are described below. It should be noted that no method can catch every origin, and all of them provide an ensemble view of the average activity in a population of cells.

Trapping the earliest replicated DNA by replication fork arrest

Cells can be synchronized before S phase and allowed to initiate replication in the presence of a replication fork inhibitor, such as hydroxyurea, to accumulate nascent strands within a few kilobases of the origins that fire at the onset of S phase. Replicated sequences are then detected by their twofold copy number increase on duplication[35,43] or by labelling the nascent strands with tagged nucleotide precursors either before fork arrest[11,17,36,44,45] or after primer extension of the arrested forks[46,47]. One or more of these methods has been successfully applied at genome scale in both budding[44–46] and fission[17,35,36] yeasts, and recently in *D. melanogaster*[11].

In principle, these methods are limited to mapping origins that fire very early in S phase because arrested replication forks trigger a checkpoint response that inhibits origins that would normally fire later in S phase[1]. However, a substantial number of origins, more than can be accounted for by just the earliest firing origins, seem to initiate before the replication forks arrest. Indeed, approximately one-third of the total number of ORC binding sites was detected using these methods in budding[43,46] and fission[17] yeast and nearly that many in *D. melanogaster*[11]. Consistent with checkpoint inhibition of late-firing and dormant origins, additional origins were detected with yeast strains harbouring mutations in proteins essential for the S-phase checkpoint[17,46]. Furthermore, all origins eventually fired in their normal temporal order when hydroxyurea was administered to wild-type cells for an extended period of time[48], suggesting that it may be possible to map most or all origins using this method under the appropriate conditions.

One concern with this method is that it is hard to evaluate the extent to which replication fork inhibitors obscure the relative efficiencies or normal firing time of origins[6,49–51]. Nonetheless, this has become a routine assay in yeasts that has been used to assess the effects of specific mutations on the firing of origins throughout the genome[52]. However, it has not been exploited at genome scale in mammalian cells, partly owing to the complexity of the genome. A variation of this method has been used to map origins on an array covering 120 kb of the Chinese hamster dihydrofolate reductase locus in a cell line harbouring 1,000 amplified copies of the locus[47], but this is clearly an underexplored methodology for identifying mammalian origins.

Small nascent leading strands

Arguably, the most precise way to identify origins is to map small nascent leading strands (SNSs). By definition, there are two leading strands emanating bidirectionally from origins, and their 5′ ends define the site of initiation (FIG. 1b). Denaturing genomic DNA from proliferating cells releases single-stranded nascent DNA, which can be fractionated by size to identify strands closer to the origin (FIG. 2A). Moreover, nascent strands from asynchronous cells derive from all origins firing throughout S phase, and their relative abundance should be a direct reflection of their efficiency of firing within a population of cells. Thus — in principle — hybridizing SNSs to a microarray or sequencing them should provide a satisfyingly high-resolution origin map.

In practice, however, SNSs are truly needles in a haystack, and obtaining sufficiently enriched preparations of SNSs is a work in progress. First, such strands should be as small as possible but must be large enough to be cleanly separated from the 50–350 bp Okazaki fragments that are assembled throughout the genome. However, the most difficult aspect is that size-fractionated SNSs must be distinguished and separated from DNA that has been inadvertently sheared or digested by the inevitable contaminating endonucleases during purification[53,54]. Nascent DNA enrichment is accomplished either by labelling newly synthesized DNA with 5-bromodeoxy-uridine (BrdU) and purifying the BrdU-substituted DNA with anti-BrdU antibodies (BrdU immunoprecipitation (BrdU-IP)) or by digesting DNA preparations with λ exonuclease (Lexo), which enriches for DNA strands that are protected from digestion by the presence of RNA primers at their 5′ ends[55,56]. Both methods have their drawbacks. BrdU-substituted DNA is more prone to breakage than unsubstituted DNA; breakage can create small strands that are nascent but not close to origins. Lexo is cleaner in principle, but its activity must be carefully controlled. Successful mapping of some human origins has been reported without either enrichment protocol[57], but the yield of nascent strands obtained is far greater than expected, making contamination a lingering concern.

Three studies using the Lexo method to map mammalian origins have been reported: two in HeLa cells mapped across the 1% of the human genome designated by The ENCODE Project[54,58] and one in mouse embryonic stem (ES) cells mapped across 10 Mb (0.4%) of the murine genome[59]. The two ENCODE studies used the Lexo method and each identified approximately 300 origins. One of these studies[58] also used BrdU-IP and identified over 800 origins that overlapped with half of the Lexo origins. In single-site ChIP analyses, 12 of these overlapping origins out of 15 tested were also enriched for ORC. All three studies found that origins were significantly enriched near TSSs and transcriptional regulatory elements; half aligned with CpG islands and almost half were in the bodies of genes. The extensive data available for ENCODE regions enabled alignment of these origins to chromatin marks, but no specific feature of chromatin (histone marks or DNase I hypersensitive sites) aligned to more than half of the origins[53,58,60].

What has been most disconcerting is that the two ENCODE studies that used the same method (albeit with different amplification protocols and array platforms) in the same cell line (albeit using independent subclones of the HeLa cancer cell line) show less than 14% overlap[54,58]. Hence, these methods may each be detecting a different subset of origins in the cell. Given that mammalian cells are estimated to replicate from 100,000 origins per cell cycle and most origins are used in less than 10% of cell cycles[27,28,38–41], a conservative estimate is that these data sets are detecting only a few per cent of expected origins, which are probably the most efficient origins (FIG. 2). So far, this method has not been applied in model systems with smaller genomes, which would assist the further development of the methodology. In short, the SNS method holds great promise, but improved reproducibility will be necessary before it can be expanded as a standardized assay, and improved sensitivity will be necessary to detect less efficient origins.

Bubble trap

The bubble-trap method relies on the fact that newly initiated sites of bidirectional replication or replication bubbles are essentially circular DNA molecules that will become selectively trapped as materials, such as agarose, polymerize in their presence (FIG. 2B). Hence, if genomic DNA from proliferating cells is digested with a restriction enzyme and embedded in an agarose plug, extensive electrophoresis will eventually remove linear double-stranded molecules and any molecules that were replicating from forks initiated outside the restriction fragment, but fragments that contain recently initiated origins will be unable to escape. Trapped bubbles can then be cloned into a plasmid library by virtue of their sticky ends and then either

hybridized to an array or sequenced, although the resolution gained by sequencing is limited by the relatively large sizes of the fragments analysed. This method is best at detecting origins that are positioned at the centre of a restriction fragment because bubbles generated from origins near the restriction site, even if they fire efficiently, will be severed by the restriction enzyme (FIG. 2C). However, unlike the SNS method, this method can also detect clusters of inefficient origins that reside throughout the fragment because there is a higher overall probability that initiation will occur nearer the centre of the fragment.

Plasmid libraries of bubble-containing fragments from one Chinese hamster and two human cell lines (HeLa and GM06990) have been made and hybridized to ENCODE microarrays[41,61] (J. Hamlin, personal communication). These results have identified several hundred origins, many of which encompass more than one contiguous restriction fragment, which is suggestive of large 'initiation zones'. In HeLa cells, bubble-trap origins align weakly to either of the sets of origins found using the SNS method in the same cell line[54,58] (~10% of bubble-trap origins overlap with either set of SNS origins and ~33% of each SNS set overlaps with the bubble-trap set). This is not surprising, as the bubble-trap and SNS methods might enrich for different types of origins (FIG. 2C). Despite the discordance between them, however, bubble-trap origins, like SNS origins, were strongly enriched in TSSs and the bodies of genes.

Future directions in mapping origins

Sequencing can improve both sensitivity and resolution

Sequencing provides a direct read-out of the nucleic acid sequences derived from any genomics method. It avoids nonlinear hybridization signal artefacts and broadening of signal peaks owing to the bridging of partially annealed target sequences to additional complementary target sequences, which is a problem with array-based methods. It is also a truly genome-wide method, as it is independent of the limitations of probe design (such as having to avoid repetitive sequences) and automatically queries the entire genome. As discussed, ChIP–seq for ORC in budding yeast allowed for a more precise definition of ORC binding sites, and a similar approach in *D. melanogaster* would probably do the same and would enable further investigation of the generality of NFRs as a universal requirement for ORC binding. Sequencing should also help ORC detection in mammalian cells. However, if the cause of poor enrichment in ChIP-based studies is highly dispersed ORC binding, sequencing will have to be extremely deep and the coverage broad to show relative binding frequencies. However, if ORC and/or MCM has a nonspecific chromatin-scanning state and a more site-specific functional-binding state, it might be possible to identify the conformational properties of the protein complexes that enrich for the functional state[62] or use mutations that are capable of capturing functional binding states, as has been done for chromatin remodellers[63]. Such enrichment would greatly increase the benefits of sequencing.

In the case of initiation site mapping, it is not clear how much insight will be gained from sequencing the products of replication fork arrest because arrest occurs hundreds of base pairs from the origin. Genome-scale adaptation of nucleotide resolution methods, however, should be explored. For example, replication initiation point mapping uses primer extension to map the 5′ ends of SNSs and has been previously used to map individual origins at nucleotide resolution in yeasts and human cells[55,56,64]. It should not be insurmountable to adapt these methods in yeasts towards a comprehensive understanding of the precise *in vivo* structural relationship of initiation start sites to the positions of pre-RC proteins. In higher eukaryotes, sensitivity and reliability is a higher priority than resolution, and in this sense sequencing the products of any of the origin-mapping methods would help to reduce the noise that results from hybridization, should enable the detection of larger numbers of initiation sites and should improve the reproducibility of origin detection.

Single-molecule methods can evaluate heterogeneity

All genomics methods examine populations of cells, but it is clear that each chromosome uses a different cohort of replication origins. We know this because origins are used less than once per cell cycle and studies examining the positions of initiation sites on individual DNA fibres directly demonstrate the heterogeneity of origin usage[65]. If obtaining reliable data on origins in cell populations is daunting, determining what is happening on each individual fibre is truly futuristic. However, methods for improving and automating DNA fibre analysis are worthwhile long-term goals (BOX 2). Moreover, methods for examining single-molecule protein-binding sites are slowly but steadily improving[66–68]. What is clear is that the field needs to advance beyond ensemble methods if we are ever to gain a complete understanding of origin function.

Do origin locations matter?

Does it matter where replication initiates, or are origins simply sites where ORC can opportunistically bind and assemble a pre-RC? We know that replication seems to be entirely random in some systems, such as in rapid-early-cleavage-stage embryos and nematodes[69]. Moreover, our increasingly sophisticated understanding of how replication is limited to exactly one round per cell cycle still does not invoke any selective pressure for initiation at specific sites[69]. Two mutually exclusive periods of the cell cycle, one (G1 phase) in which pre-RCs can assemble but not initiate, followed by another (S phase) during which assembled pre-RCs can initiate but not re-assemble, ensure exactly one round of replication regardless of the positions of the initiation sites. Hence, any selective pressure to maintain the position of an origin must transcend the basic need to duplicate the genome. Preventing collisions between replication and transcription is a popular hypothesis for such a selective pressure[60,70–72], but evidence in many systems suggests that such collisions are a common occurrence[73–76] and origin positions do not seem to be highly conserved nor excluded from transcription units[53,60,77–79]. It is disconcerting to be an origin hunter when the origin may simply be any and all places where an origin can be assembled. However, this question is so extraordinarily fundamental that we have no choice but to pursue the answer, and genome-scale mapping of origins is essential for this goal.

Box 2 | Can DNA fibre analysis become a genome-scale approach?

The most direct means of evaluating origin efficiency is by examining the location and frequency of initiation sites on individual DNA fibres, where the relative number of times that an origin is used on any given DNA fibre can be directly counted[65]. Cells are metabolically labelled with tagged nucleotide precursors, then purified DNA is stretched on microscope slides and the sites of DNA synthesis are directly observed along the length of the DNA template by visualizing the nucleotide tag. The very first methods labelled sites of DNA synthesis with the low-energy radioactive precursor tritiated thymidine, which required months of autoradiographic exposure to detect[118]. Modern methods use fluorescently tagged nucleotides[119] or tags that can be indirectly visualized with fluorescent probes, such as halogens (chloro-, iodo- or bromo-probes), biotin or digoxigenin[65]. The polarity of replication can be determined either by observing a trailing off of a single label as the precursor pool is depleted during DNA synthesis or by using two consecutive labels. Origins can be identified as the centres of regions where forks are emanating bidirectionally. The method is quite robust for viewing overall origin distributions and fork elongation rates without map position information. The primary limitation at this level is fibre length; it is difficult to prepare fibres longer than 1 Mb, so inter-origin distance calculations (which require at least two origins on a single fibre) are biased towards those within a few hundred kilobases. Despite this bias, fibre data have directly demonstrated that origins are spaced farther apart than predicted by ensemble molecular methods, and hence only a fraction of origins used in a population are utilized in each cell[65].

A much bigger challenge is identifying the map positions of origins and forks, which requires hybridization, typically with fluorescent probes. Fibres tend to wash off during hybridization, so specialized surfaces have been developed to retain them. Perhaps more importantly, in complex genomes only one in several thousand sub-megabase-sized fibres contains the sequence of interest, so the mapping of origins requires either exhaustive searching or prospective enrichment methods that further reduce the size of fibres. Even with substantial enrichment, obtaining a statistically significant sampling of fibres is laborious, although valuable information can be gained from such studies of defined loci.

Could this process be automated to achieve genome-scale analyses? Doing so will require collaborative advances in molecular and nanoscience technologies to develop novel methods for capturing and/or identifying specific DNA molecules while preserving longer DNA molecules to reduce search times. In addition, methods will be needed to align these long linear molecules in a systematic grid-like fashion that can be interpreted by machine-driven detection methods[120]. At one extreme, for example, if it were possible to stretch DNA in such a way as to retain entire chromosomes as intact fibres, the entire constellation of simultaneously active origins and forks across a chromosome could be visualized and only 46 molecules would be necessary per diploid human genome. Alternatively, if megabase-sized molecules in solution could be stably hybridized to unique DNA sequences printed in a pattern onto a surface template[121], the positions of stretched molecules on a well-aligned, ordered array could serve as a unique identifier, which would allow origin and fork data to be collected by automated analysis of a few thousand molecules per human genome.

Mapping replication timing

In contrast to the slow progress in mapping origins, there is a lot to celebrate in terms of our ability to collect reliable data on the relative order in which segments of chromosomes replicate in cell populations. In this section I discuss genome-scale methods for mapping replication timing. These studies have confirmed links between replication timing and other structural and functional properties of chromatin and are being routinely applied in hypothesis-driven mechanistic studies.

Synchronization

When one is interested in the timing of a cell-cycle-regulated event, synchronization of cells is a must. Synchronization can either be done prospectively, before cell collection, or retroactively, after the cells have been collected. In yeasts, prospective synchrony methods are well established, which permits the same methods to be compared across different strains. Budding yeast can be synchronized in G1 phase with the mating pheromone α-factor, and both budding and fission yeast can be synchronized at the G1/S boundary using hydroxyurea and/or with mutants that arrest cells at specific cell cycle stages[6,17]. Meselson and Stahl-style density transfer, copy number increases or metabolic labelling can be used to detect replication at different times after release from the synchrony block[6]. These methods have been applied to evaluate timing of origin firing under conditions of replication stress and in cells harbouring mutations in relevant pathways such as replication, S-phase checkpoint and histone modification[6].

In cell lines derived from multicellular organisms, most synchronization schemes are cumbersome and only applicable to a limited number of cell lines[47,80,81]. Furthermore, methods that require the use of metabolic inhibitors can interfere with normal cell cycle regulation[42,50]. Hence, in cases in which multiple cell lines are to be compared, retroactive synchronization using a fluorescence-activated cell sorter (FACS) to select cells based upon

the increase in DNA content during S phase is the method of choice because it can be applied to any proliferating cell population that can be dissociated into a single cell suspension (FIG. 3). In the original version of the method[82,83], cells were labelled with BrdU for a fraction of S phase and then sorted into several different time points during S phase. BrdU-substituted DNA can then be isolated based on its increased density[82,83] or by using BrdU-IP[84], and specific loci can be examined by hybridization[82,83] or PCR[84]. Using microarray analysis[85], replication of the entire genome can be queried in a single array hybridization; two differentially labelled samples are analysed together and each probe is assigned an internally normalized relative replication timing value. This method has been applied in *D. melanogaster*[85–87], many mouse and human cell types[88–96] and *Arabidopsis thaliana*[97], and was recently used to compare replication timing across 14 yeast mutants[5].

FIG. 3 summarizes the two most popular permutations of the FACS method. In one method, cells are pulsed with BrdU and sorted into early- and late-S-phase populations; BrdU-IP DNA that was synthesized either early or late in S phase is hybridized to a microarray. Because BrdU-IP substantially enriches for DNA synthesized in each half of S phase, this method produces a high signal-to-noise ratio. However, BrdU-IP must be carefully monitored for immunoprecipitation efficiency. In the second method, unlabelled cells are sorted into total S phase versus G1 phase populations and DNA from these stages is differentially labelled and used as the target. This obviates the BrdU-IP step but the signal-to-noise ratio is low due to reliance on DNA copy number increase, which is at most twofold for the earliest replicating sequences. Both methods have been used and produce comparable results; a direct comparison in the same cell line has been done in one study[93]. A link to detailed protocols for these methods can be found in the 'Further Information' section.

Modifications for studying replication asynchrony

One limitation of these rapid profiling methods is that they only assign a single replication timing value for any given map position and so cannot distinguish cases in which regions of the two homologous chromosomes replicate at different times (asynchronous replication). This limitation has been overcome by sorting additional fractions of S phase[82,83,88,97,98]. Using these methods genome-wide, estimates of the fraction of the genome that replicates asynchronously range from 0.32%[88] to 10%[94,99]; the variability is probably due to the fact that most asynchronies are quite small and escape detection. Data collection, normalization and computation for multiple S phase fractions per experiment is considerably more demanding, but for those interested in the mechanisms of monoallelic expression and imprinting, retroactive FACS synchronization is still the method of choice.

Insights into replication timing from genome-scale studies

Genome-wide replication timing profiling is now a routine assay that can be used to assess the effects of mutations on replication timing[52,87,92,100] or to localize replication timing changes during development, including cells differentiated to finely separated differentiation states[89] or changes in cancer cells (D. Battaglia, T. Ryba and D.M.G., unpublished observations). The results of these studies, which have been discussed in detail in numerous recent reviews[79,81,101–104], have substantially advanced our understanding of replication timing programmes.

Many findings are in common from yeast to humans. We can now say that replication timing is regulated at the level of multi-replicon domains (replication domains) in all eukaryotic species, and the sizes of these domains are generally proportional to the genome size[17,43,86,103,105,106]. Replication domains are punctuated by large origin-less regions in which single forks move unidirectionally for long distances[90,91,93,94]. Replication is well conserved between distantly related species and considerably more conserved than the positions of origins themselves[77–79,90,94,107] (N. Rhind, personal communication; K. Lindstrom and B. Brewer,

personal communication). The global replication timing programme is generally resilient to many mutations in replication and chromatin proteins[5,48,92,108], but can exhibit profound localized responses to such mutations[36,52,87,109–111].

One major difference that has emerged between multicellular and single-celled organisms is that a strong statistical correlation between early replication and transcription does not exist in either budding or fission yeasts, whereas in all multicellular organisms, including plants, there is strong correlation, particularly when transcriptional output or its associated chromatin modifications are integrated over replication domain-sized regions[93,97,106]. Another difference is that yeasts do not show developmental control of replication timing, whereas at least half of the mammalian genome is subject to developmentally regulated changes in replication timing that occur in units of 400–800 kb and are coordinated with changes in transcriptional regulation from a certain class of promoters[90,93]. Similar changes occur in *D. melanogaster*, but with smaller domains[86,87].

Finally, genomics has allowed investigators to look for other properties that correlate with replication timing. In mammalian cells, the strongest of these correlations is a close alignment with the density of chromatin interactions within the nucleus[90], which provides strong support for the hypothesis that replication domains are self-interacting structural and functional units of chromosomes. It remains to be seen whether this correlation will also be found in yeasts, but chromatin interaction maps are now available for comparing chromatin interactions and replication timing in yeast[112].

What is next for replication timing?

What can sequencing do for replication timing?

In budding yeast, in which many origins fire in more than 50% of cell cycles, it is possible that resolution may be improved by sequencing nascent DNA from temporally staged cells. However, in mammalian cells the resolution of replication timing experiments is limited by the fact that replication forks move at 1–2 kb per minute within a 500–600 min S phase. Given that even very short BrdU pulses label hundreds of kilobases, sequencing is something akin to weighing a truck on a precision balance. A more important area of improvement in replication timing technology is to reduce labelling times and optimize the purification of replication intermediates, which would improve resolution and increase the value of sequencing. However, because each mammalian cell initiates from a different cohort of origins, even with improved resolution the replication timing profiles are at best a smoothed average of origin choices in a cell population. Sequencing is more expensive and, so far, has required more cells than microarray hybridization (5–20 million cells[88,91] are required for sequencing compared with as few as 0.5 million cells[89] for microarrays). However, cost and sensitivity will continue to improve, and sequencing is not disadvantageous; direct comparisons reveal complete concordance of sequencing and microarray data[90]. Moreover, there is an area in which sequencing might eventually provide a clear advantage: one can envision that in the future, SNPs and their linkage relationships will be sufficiently well mapped to allow the separate profiles of individual maternal and paternal chromosomes to be distinguished by extremely deep and broad sequencing.

Single-cell methods will be challenging

One important question that has not been tackled in any system is the extent to which there may be cell-to-cell variation in replication timing. Such information would aid tremendously in understanding the mechanism of replication timing changes during cell-fate transitions in development[89]. Replication timing can be measured in single cells using fluorescence *in situ* hybridization (FISH); the detection of two adjacent FISH signals indicates the presence of

specific duplicated regions in individual cells[113]. Currently, one can envision comparisons of relative replication timing between two specific sites within individual cells, or possibly more (with improvements in colour discrimination). However, the frequency of false-positive FISH doublets currently limits sensitivity, and FISH data are generally pooled to give an average doublet frequency. DNA fibres isolated from synchronized cells could provide information on the percentage of specific molecules that replicate at an inappropriate time, but tracing individual DNA fibres back to their cell of origin is hard to envision. Clearly, addressing the issue of replication-timing heterogeneity will require substantial conceptual breakthroughs.

What is the significance of replication timing?

Genome-scale methods have confirmed that replication timing is conserved across distantly related species[90,94,107], suggesting that it either represents or reflects something biologically important. However, replication timing could be a passive reflection of some other chromosomal property, such as chromosome architecture[90], and it is unlikely that the genome has to duplicate in any particular temporal sequence to complete replication. Because it is not just DNA that replicates but also chromatin, and because different types of chromatin are assembled at different times during S phase[114], it is reasonable to suggest that the replication fork provides an important opportunity for cells to either maintain or rapidly change the status of chromatin across large multi-replicon chromosome domains, but direct evidence for this hypothesis is lacking[101]. Recent genome-wide analyses have also suggested that mutation rates are significantly higher in late-replicating regions[115,116], although the mechanism through which this occurs is unknown and it is difficult to determine whether replication timing is a driving force in evolution or reflects an evolutionary cost of having genes in regions that happen to be late-replicating for some other reason.

Regardless of whether replication is the driving force behind chromosome domain organization or whether chromosome architecture is driving replicon structure, replication timing provides a clear and measurable property with which to evaluate higher-order chromosome structure and function. Genome-wide replication timing profiles can identify megabase-sized regions that behave as organizational units and can rapidly identify which of these higher-order units is remodelled during development or in specific diseases.

Conclusions

The field of DNA replication needs a breakthrough, and the advent of genomics gives hope that detailed maps can shed light on fresh directions for research. It is now clear that a temporal programme for replication is measurable, developmentally regulated and evolutionarily conserved. We know the sizes of timing units, we have confirmed long-standing correlations with transcription and we know that timing closely reflects the spatial proximity of chromatin in the nucleus. Replication-timing profiling provides a technology that can identify domain-level aberrations or alterations in replication timing associated with disease and can identify important epigenetic distinctions between closely related cell types[90]. Genome-scale studies of DNA replication timing have also sparked the development of a number of mathematical models that simulate various aspects of replication control and permit the development of novel hypotheses and the iterative testing of predictions about regulation[117]. Although these studies have focused our hypotheses, the issues of significance — cause and effect — remain a mystery.

Lagging far behind studies of replication timing is the methodology for identifying metazoan origins; developments in this area have been frustrated by the degree of site flexibility and cell-to-cell variation in origin choice. Indeed, is not clear whether the positions of origins matter to the cell. Despite this frustration, genomic studies in yeast and *D. melanogaster* have lent substantial credence to the hypothesis developed in budding yeast that ORC binding sites consist of NFRs that, together with ORC, position nucleosomes in such a way as to provide

space for loading the remainder of the pre-RC. As both sequence and chromatin composition can influence nucleosome positions, this kind of feature is consistent with the variability and plasticity of replication origins in multicellular organisms and provides a glimmer of hope that some universal rules will eventually be decipherable.

Acknowledgments

I thank J. Huberman, D. Macalpine, O. Aparicio, A. Beilinsky, N. Rhind, M.-N. Prioleau, A. Dutta, H. Masukata, A. Schepers, J. Hamlin, K. Plath, B. Papp, L. Mesner, M. Mechali, K. Ekwall and members of my laboratory for helpful discussions during the preparation of this article.

References

1. Masai H, Matsumoto S, You Z, Yoshizawa-Sugata N, Oda M. Eukaryotic chromosome DNA replication: where, when, and how? Annu Rev Biochem 2010;79:89–130. [PubMed: 20373915]
2. Woodward AM, et al. Excess Mcm2–7 license dormant origins of replication that can be used under conditions of replicative stress. J Cell Biol 2006;173:673–683. [PubMed: 16754955]
3. Ibarra A, Schwob E, Mendez J. Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication. Proc Natl Acad Sci USA 2008;105:8956–8961. [PubMed: 18579778]
4. Doksani Y, Bermejo R, Fiorani S, Haber JE, Foiani M. Replicon dynamics, dormant origin firing, and terminal fork integrity after double-strand break formation. Cell 2009;137:247–258. [PubMed: 19361851]
5. Koren A, Soifer I, Barkai N. MRC1-dependent scaling of the budding yeast DNA replication timing program. Genome Res 2010;20:781–790. [PubMed: 20219942]
6. Raghuraman MK, Brewer BJ. Molecular analysis of the replication program in unicellular model organisms. Chromosome Res 2010;18:19–34. [PubMed: 20012185]
7. Krysan PJ, Smith JG, Calos MP. Autonomous replication in human cells of multimers of specific human and bacterial DNA sequences. Mol Cell Biol 1993;13:2688–2696. [PubMed: 8386315]
8. Aladjem MI. Replication in context: dynamic regulation of DNA replication patterns in metazoans. Nature Rev Genet 2007;8:588–600. [PubMed: 17621316]
9. Hamlin JL, et al. A revisionist replicon model for higher eukaryotic genomes. J Cell Biochem 2008;105:321–329. [PubMed: 18680119]
10. Eaton ML, Galani K, Kang S, Bell SP, MacAlpine DM. Conserved nucleosome positioning defines replication origins. Genes Dev 2010;24:748–753. [PubMed: 20351051] This paper combined genome-wide ChIP–seq analysis of ORC and nucleosome positions with in vitro nucleosome reconstitution to demonstrate that ORC binds to sequence-defined NFRs and directs local nucleosome positioning at budding yeast origins.
11. MacAlpine HK, Gordan R, Powell SK, Hartemink AJ, MacAlpine DM. *Drosophila* ORC localizes to open chromatin and marks sites of cohesin complex loading. Genome Res 2010;20:201–211. [PubMed: 19996087] This paper used ChIP–chip to show that ORC localizes to NFRs in D. melanogaster cells. In combination with reference 10, it suggests that nucleosome organization may be a defining feature of all eukaryotic origins.
12. Deal RB, Henikoff JG, Henikoff S. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. Science 2010;328:1161–1164. [PubMed: 20508129]
13. Breier AM, Chatterji S, Cozzarelli NR. Prediction of *Saccharomyces cerevisiae* replication origins. Genome Biol 2004;5:R22. [PubMed: 15059255]
14. Wyrick JJ, et al. Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins. Science 2001;294:2357–2360. [PubMed: 11743203]
15. Xu W, Aparicio JG, Aparicio OM, Tavare S. Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in *S cerevisiae*. BMC Genomics 2006;7:276. [PubMed: 17067396]
16. MacAlpine DM, Bell SP. A genomic view of eukaryotic DNA replication. Chromosome Res 2005;13:309–326. [PubMed: 15868424]

17. Hayashi M, et al. Genome-wide localization of pre-RC sites and identification of replication origins in fission yeast. EMBO J 2007;26:1327–1339. [PubMed: 17304213]
18. Nieduszynski CA, Knox Y, Donaldson AD. Genome-wide identification of replication origins in yeast by comparative genomics. Genes Dev 2006;20:1874–1879. [PubMed: 16847347]
19. Dai J, Chuang RY, Kelly TJ. DNA replication origins in the *Schizosaccharomyces pombe* genome. Proc Natl Acad Sci USA 2005;102:337–342. [PubMed: 15623550]
20. Cotobal C, Segurado M, Antequera F. Structural diversity and dynamics of genomic replication origins in *Schizosaccharomyces pombe*. EMBO J 2010;29:934–942. [PubMed: 20094030]
21. Chuang RY, Kelly TJ. The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks. Proc Natl Acad Sci USA 1999;96:2656–2661. [PubMed: 10077566]
22. Lantermann AB, et al. *Schizosaccharomyces pombe* genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of *Saccharomyces cerevisiae*. Nature Struct Mol Biol 2010;17:251–257. [PubMed: 20118936]
23. Gilbert DM. In search of the holy replicator. Nature Rev Mol Cell Biol 2004;5:848–855. [PubMed: 15459665]
24. Paixao S, et al. Modular structure of the human lamin B2 replicator. Mol Cell Biol 2004;24:2958–2967. [PubMed: 15024083]
25. Altman AL, Fanning E. Defined sequence modules and an architectural element cooperate to promote initiation at an ectopic mammalian chromosomal replication origin. Mol Cell Biol 2004;24:4138–4150. [PubMed: 15121836]
26. Liu G, Malott M, Leffak M. Multiple functional elements comprise a mammalian chromosomal replicator. Mol Cell Biol 2003;23:1832–1842. [PubMed: 12589000]
27. Guan Z, et al. Decreased replication origin activity in temporal transition regions. J Cell Biol 2009;187:623–635. [PubMed: 19951913]
28. Lin HB, Dijkwel PA, Hamlin JL. Promiscuous initiation on mammalian chromosomal DNA templates and its possible suppression by transcription. Exp Cell Res 2005;308:53–64. [PubMed: 15904920]
29. Vashee S, et al. Sequence-independent DNA binding and replication initiation by the human origin recognition complex. Genes Dev 2003;17:1894–1908. [PubMed: 12897055]
30. Remus D, Beall EL, Botchan MR. DNA topology, not DNA sequence, is a critical determinant for *Drosophila* ORC-DNA binding. EMBO J 2004;23:897–907. [PubMed: 14765124]
31. Gupta S, et al. Predicting human nucleosome occupancy from primary sequence. PLoS Comput Biol 2008;4:e1000134. [PubMed: 18725940]
32. Schepers A, Papior P. Why are we where we are? Understanding replication origins and initiation sites in eukaryotes using ChIP-approaches. Chromosome Res 2010;18:63–77. [PubMed: 19904620]
33. Friedman KL, Brewer BJ, Fangman WL. Replication profile of *Saccharomyces cerevisiae* chromosome VI. Genes Cells 1997;2:667–678. [PubMed: 9491801]
34. Poloumienko A, Dershowitz A, De J, Newlon CS. Completion of replication map of *Saccharomyces cerevisiae* chromosome III. Mol Biol Cell 2001;12:3317–3327. [PubMed: 11694569]
35. Heichinger C, Penkett CJ, Bahler J, Nurse P. Genome-wide characterization of fission yeast DNA replication origins. EMBO J 2006;25:5171–5179. [PubMed: 17053780]
36. Wu PY, Nurse P. Establishing the program of origin firing during S phase in fission yeast. Cell 2009;136:852–864. [PubMed: 19269364] This report used genome-wide replication profiling of the earliest BrdU-labelled DNA synthesized in the presence of hydroxyurea to show that holding cells in mitosis increases the binding of ORC to certain origins in fission yeast and causes these origins to become earlier replicating.
37. Patel PK, Arcangioli B, Baker SP, Bensimon A, Rhind N. DNA replication origins fire stochastically in fission yeast. Mol Biol Cell 2006;17:308–316. [PubMed: 16251353]
38. Norio P, et al. Progressive activation of DNA replication initiation in large domains of the immunoglobulin heavy chain locus during B cell development. Mol Cell 2005;20:575–587. [PubMed: 16307921]
39. Norio P, Schildkraut CL. Plasticity of DNA replication initiation in epstein-barr virus episomes. PLoS Biol 2004;2:e152. [PubMed: 15208711]

40. Lebofsky R, Heilig R, Sonnleitner M, Weissenbach J, Bensimon A. DNA replication origin interference increases the spacing between initiation events in human cells. Mol Biol Cell 2006;17:5337–5345. [PubMed: 17005913]
41. Mesner LD, Crawford EL, Hamlin JL. Isolating apparently pure libraries of replication origins from complex genomes. Mol Cell 2006;21:719–726. [PubMed: 16507369]
42. Anglana M, Apiou F, Bensimon A, Debatisse M. Dynamics of DNA replication in mammalian somatic cells: nucleotide pool modulates origin choice and interorigin spacing. Cell 2003;114:385–394. [PubMed: 12914702]
43. Yabuki N, Terashima H, Kitada K. Mapping of early firing origins on a replication profile of budding yeast. Genes Cells 2002;7:781–789. [PubMed: 12167157]
44. Katou Y, et al. S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex. Nature 2003;424:1078–1083. [PubMed: 12944972]
45. Viggiani CJ, Knott SR, Aparicio OM. Genome-wide analysis of DNA synthesis by BrdU immunoprecipitation on tiling microarrays (BrdU-IP-chip). Saccharomyces cerevisiae Cold Spring Harb Protoc 2010;2010 pdb.prot5385.
46. Feng W, et al. Genomic mapping of single-stranded DNA in hydroxyurea-challenged yeasts identifies origins of replication. Nature Cell Biol 2006;8:148–155. [PubMed: 16429127]
47. Sasaki T, et al. The Chinese hamster dihydrofolate reductase replication origin decision point follows activation of transcription and suppresses initiation of replication within transcription units. Mol Cell Biol 2006;26:1051–1062. [PubMed: 16428457]
48. Alvino GM, et al. Replication in hydroxyurea: it's a matter of time. Mol Cell Biol 2007;27:6396–6406. [PubMed: 17636020]
49. Mickle KL, et al. Checkpoint independence of most DNA replication origins in fission yeast. BMC Mol Biol 2007;8:112. [PubMed: 18093330]
50. Gilbert DM. Replication origin plasticity, Taylor-made: inhibition vs recruitment of origins under conditions of replication stress. Chromosoma 2007;116:341–347. [PubMed: 17404750]
51. Courbet S, et al. Replication fork movement sets chromatin loop size and origin choice in mammalian cells. Nature 2008;455:557–560. [PubMed: 18716622]
52. Knott SR, Viggiani CJ, Tavare S, Aparicio OM. Genome-wide replication profiles indicate an expansive role for Rpd3L in regulating replication initiation timing or efficiency, and reveal genomic loci of Rpd3 function in Saccharomyces cerevisiae. Genes Dev 2009;23:1077–1090. [PubMed: 19417103] In this report, the authors examined the role of the budding yeast histone deacetylase Rpd3 in the timing of origin firing using genome-wide replication profiling of the earliest BrdU-labelled DNA synthesized in the presence of hydroxyurea.
53. Cadoret JC, Prioleau MN. Genome-wide approaches to determining origin distribution. Chromosome Res 2010;18:79–89. [PubMed: 19921448]
54. Cadoret JC, et al. Genome-wide studies highlight indirect links between human replication origins and gene regulation. Proc Natl Acad Sci USA 2008;105:15837–15842. [PubMed: 18838675] A genome-scale analysis of sites of enrichment of small nascent single-stranded DNA within the human ENCODE regions using the Lexo method and hybridization to microarrays.
55. Gerbi SA, Bielinsky AK. Replication initiation point mapping. Methods 1997;13:271–280. [PubMed: 9441853]
56. Das-Bradoo S, Bielinsky AK. Replication initiation point mapping: approach and implications. Methods Mol Biol 2009;521:105–120. [PubMed: 19563103]
57. Lucas I, et al. High-throughput mapping of origins of replication in human cells. EMBO Rep 2007;8:770–777. [PubMed: 17668008]
58. Karnani N, Taylor CM, Malhotra A, Dutta A. Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. Mol Biol Cell 2010;21:393–404. [PubMed: 19955211] A second genome-scale analysis of sites of enrichment of small nascent single-stranded DNA within the human ENCODE regions using both the Lexo and BrdU-IP methods and hybridization to microarrays. Reasons for the low concordance with results in reference 54 are discussed although many features of the sites were in common.

59. Sequeira-Mendes J, et al. Transcription initiation activity sets replication origin efficiency in mammalian cells. PLoS Genet 2009;5:e1000446. [PubMed: 19360092] A genome-scale analysis of sites of enrichment of small nascent single-stranded DNA across a segment of the mouse genome in ES cells using hybridization to microarrays. Similarities in the features found for these origins and those in reference 54 are discussed.
60. Maric C, Prioleau MN. Interplay between DNA replication and gene expression: a harmonious coexistence. Curr Opin Cell Biol 2010;22:277–283. [PubMed: 20363609]
61. Mesner LD, Hamlin JL. Isolation of restriction fragments containing origins of replication from complex genomes. Methods Mol Biol 2009;521:315–328. [PubMed: 19563114]
62. Francis LI, Randell JC, Takara TJ, Uchima L, Bell SP. Incorporation into the prereplicative complex activates the Mcm2–7 helicase for Cdc7-Dbf4 phosphorylation. Genes Dev 2009;23:643–654. [PubMed: 19270162]
63. Gelbart ME, Bachman N, Delrow J, Boeke JD, Tsukiyama T. Genome-wide identification of Isw2 chromatin-remodeling targets by localization of a catalytically inactive mutant. Genes Dev 2005;19:942–954. [PubMed: 15833917]
64. Romero J, Lee H. Asymmetric bidirectional replication at the human DBF4 origin. Nature Struct Mol Biol 2008;15:722–729. [PubMed: 18536724]
65. Tuduri S, Tourriere H, Pasero P. Defining replication origin efficiency using DNA fiber assays. Chromosome Res 2010;18:91–102. [PubMed: 20039120]
66. Cohen SM, et al. BRG1 co-localizes with DNA replication factors and is required for efficient replication fork progression. Nucleic Acids Res. Jun 22;2010 10.1093/nar/gkq559
67. Sullivan BA, Karpen GH. Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. Nature Struct Mol Biol 2004;11:1076–1083. [PubMed: 15475964]
68. Cipriany BR, et al. Single molecule epigenetic analysis in a nanofluidic channel. Anal Chem 2010;82:2480–2487. [PubMed: 20184350]
69. Gilbert DM. Making sense of eukaryotic DNA replication origins. Science 2001;294:96–100. [PubMed: 11588251]
70. Knott SR, Viggiani CJ, Aparicio OM. To promote and protect: coordinating DNA replication and transcription for genome stability. Epigenetics 2009;4:362–365. [PubMed: 19736523]
71. Huvet M, et al. Human gene organization driven by the coordination of replication and transcription. Genome Res 2007;17:1278–1285. [PubMed: 17675363]
72. Jorgensen FG, Schierup MH. Increased rate of human mutations where DNA and RNA polymerases collide. Trends Genet 2009;25:523–527. [PubMed: 19853958]
73. Bermejo R, et al. Genome-organizing factors Top2 and Hmo1 prevent chromosome fragility at sites of S phase transcription. Cell 2009;138:870–884. [PubMed: 19737516]
74. Tuduri S, et al. Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. Nature Cell Biol 2009;11:1315–1324. [PubMed: 19838172]
75. Necsulea A, Guillet C, Cadoret JC, Prioleau MN, Duret L. The relationship between DNA replication and human genome organization. Mol Biol Evol 2009;26:729–741. [PubMed: 19126867]
76. Azvolinsky A, Giresi PG, Lieb JD, Zakian VA. Highly transcribed RNA polymerase II genes are impediments to replication fork progression in *Saccharomyces cerevisiae*. Mol Cell 2009;34:722–734. [PubMed: 19560424]
77. Liachko I, et al. A comprehensive genome-wide map of autonomously replicating sequences in a naive genome. PLoS Genet 2010;6:e1000946. [PubMed: 20485513]
78. Aladjem MI, et al. Replication initiation patterns in the β-globin loci of totipotent and differentiated murine cells: evidence for multiple initiation regions. Mol Cell Biol 2002;22:442–452. [PubMed: 11756541]
79. Hiratani, I.; Gilbert, DM. The Cell Biology of Stem Cells. Meshorer, E.; Plath, K., editors. Landes and Springer; Austin, Texas: 2010. p. 41-58.
80. Karnani N, Taylor CM, Dutta A. Microarray analysis of DNA replication timing. Methods Mol Biol 2009;556:191–203. [PubMed: 19488880]

81. Farkash-Amar S, Simon I. Genome-wide analysis of the replication program in mammals. Chromosome Res 2009;18:115–125. [PubMed: 20205353]

82. Gilbert DM, Cohen SN. Bovine papilloma virus plasmids replicate randomly in mouse fibroblasts throughout S-phase of the cell cycle. Cell 1987;50:59–68. [PubMed: 3036365]

83. Gilbert DM. Temporal order of replication of *Xenopus laevis* 5S ribosomal RNA genes in somatic cells. Proc Natl Acad Sci USA 1986;83:2924–2928. [PubMed: 3458252]

84. Hansen R, Canfield T, Lamb M, Gartler S, Laird C. Association of fragile X syndrome with delayed replication of the *FMR1* gene. Cell 1993;73:1403–1409. [PubMed: 8324827]

85. Schubeler D, et al. Genome-wide DNA replication profile for *Drosophila melanogaster*: a link between transcription and replication timing. Nature Genet 2002;32:438–442. [PubMed: 12355067]

86. Schwaiger M, et al. Chromatin state marks cell-type- and gender-specific replication of the *Drosophila* genome. Genes Dev 2009;23:589–601. [PubMed: 19270159] The first comprehensive genome-wide mapping of replication timing in two D. melanogaster tissue culture cell lines derived from different tissues. The study revealed developmental and gender-specific differences.

87. Schwaiger M, Kohler H, Oakeley EJ, Stadler MB, Schubeler D. Heterochromatin protein 1 (HP1) modulates replication timing of the *Drosophila* genome. Genome Res 2010;20:771–780. [PubMed: 20435908] A genome-wide analysis of replication timing in D. melanogaster cultured cells lacking heterochromatin protein 1 (HP1). Surprisingly, the study revealed roles for HP1 both in very late replication of centromeric DNA and in early replication of certain euchromatic regions with high levels of repeats.

88. Hansen RS, et al. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proc Natl Acad Sci USA 2010;107:139–144. [PubMed: 19966280] A genome-wide replication timing analysis in four human cell lines using the BrdU-IP method followed by sequencing. This study revealed many differences in and links to the accessibility of chromatin by DNase I.

89. Hiratani I, et al. Genome-wide dynamics of replication timing revealed by *in vitro* models of mouse embryogenesis. Genome Res 2010;20:155–169. [PubMed: 19952138] A genome-wide analysis of replication timing in 22 cell lines representing 10 stages of early mouse development. It revealed widespread changes in replication timing, including a set of early to late changes during epiblast maturation that harboured genes which became difficult to reactivate when late replicating.

90. Ryba T, et al. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. Genome Res 2010;20:761–770. [PubMed: 20430782] A genome-wide replication timing analysis of eight human cell types, including four human ES cell lines, and a comparison to corresponding mouse profiles. The study revealed overall cell-type-specific evolutionary conservation of replication timing and a much closer alignment of human ES cells to mouse epiblast-derived stem cells than to mouse ES cells, as well as a surprisingly close alliance of replication timing to chromatin spatial proximity. This report also showed the equivalence of microarray to sequencing methods in the same cell line.

91. Desprat R, et al. Predictable dynamic program of timing of DNA replication in human cells. Genome Res 2009;19:2288–2299. [PubMed: 19767418] A replication timing analysis of 3% of the genome in three human cell lines using the S/G1 microarray hybridization method, followed by genome-wide analysis by S/G1 sequencing. Equivalent results were obtained by both methods, and this group also identified a close linkage between early replication and proximity to expressed genes.

92. Yokochi T, et al. G9a selectively represses a class of late-replicating genes at the nuclear periphery. Proc Natl Acad Sci USA 2009;106:19363–19368. [PubMed: 19889976] Genome-wide replication timing analysis is used as a tool to show that loss of the G9a histone methyltransferase in mouse ES cells results in the de-repression of a set of genes that are almost exclusively late replicating. However, loss of this enzyme has no detectable effect on replication timing of the genome, except for peri-centromeric regions.

93. Hiratani I, et al. Global reorganization of replication domains during embryonic stem cell differentiation. PLoS Biol 2008;6:e245. [PubMed: 18842067] The first genome-wide analysis of replication timing during the course of directed differentiation in three mouse ES cell lines. It revealed widespread consolidation of replication domains, which resulted in the production of fewer segments of discordant replication during neural differentiation. Furthermore, the study highlighted links to

transcriptional contol, promoter classes and chromatin marks, and also compared the S/G1 and BrdU-IP profiling methods.

94. Farkash-Amar S, et al. Global organization of replication time zones of the mouse genome. Genome Res 2008;18:1562–1570. [PubMed: 18669478] The authors used an elegant synchronization scheme to create a high-resolution genome-wide map of regions that replicate within fine cell cycle windows.
95. Woodfine K, et al. Replication timing of human chromosome 6. Cell Cycle 2005;4:172–176. [PubMed: 15611667]
96. Woodfine K, et al. Replication timing of the human genome. Hum Mol Genet 2004;13:191–202. [PubMed: 14645202]
97. Lee TJ, et al. Arabidopsis thaliana chromosome 4 replicates in two phases that correlate with chromatin state. PLoS Genet 2010;6:e1000982. [PubMed: 20548960] The first genome-scale analysis of replication timing across a segment of the A. thaliana genome using the BrdU-IP method. Many parallels to animal cells were found.
98. Gilbert DM, Cohen SN. Position effects on the timing of replication of chromosomally integrated simian virus 40 molecules in Chinese hamster cells. Mol Cell Biol 1990;10:4345–4355. [PubMed: 2164638]
99. Karnani N, Taylor C, Malhotra A, Dutta A. Pan-S replication patterns and chromosomal domains defined by genome-tiling arrays of ENCODE genomic areas. Genome Res 2007;17:865–876. [PubMed: 17568004]
100. Lu J, Li F, Murphy CS, Davidson MW, Gilbert DM. G2 phase chromatin lacks determinants of replication timing. J Cell Biol 2010;189:967–980. [PubMed: 20530209]
101. Hiratani I, Takebayashi S, Lu J, Gilbert DM. Replication timing and transcriptional control: beyond cause and effect — part II. Curr Opin Genet Dev 2009;19:142–149. [PubMed: 19345088]
102. Hiratani I, Gilbert DM. Replication timing as an epigenetic mark. Epigenetics 2009;4:93–97. [PubMed: 19242104]
103. Pope BD, Hiratani I, Gilbert DM. Domain-wide regulation of DNA replication timing during mammalian development. Chromosome Res 2010;18:127–136. [PubMed: 20013151]
104. Mendez J. Temporal regulation of DNA replication in mammalian cells. Crit Rev Biochem Mol Biol 2009;44:343–351. [PubMed: 19780641]
105. McCune HJ, et al. The temporal program of chromosome replication: genomewide replication in clb5 δ Saccharomyces cerevisiae. Genetics 2008;180:1833–1847. [PubMed: 18832352] This report used density transfer to isolate DNA replicated at specific times during S phase, followed by array hybridization. The study showed that, in Clb5-deleted budding yeast that can only initiate at early origins, large blocks of chromosomes became considerably delayed in their replication, providing evidence for clustering of temporally related origins in budding yeast.
106. MacAlpine DM, Rodriguez HK, Bell SP. Coordination of replication and transcription along a *Drosophila* chromosome. Genes Dev 2004;18:3094–3105. [PubMed: 15601823]
107. Yaffe E, et al. Comparative analysis of DNA replication timing reveals conserved large-scale chromosomal architecture. PLoS Genet 2010;6:e1001011. [PubMed: 20617169] Genome-wide replication timing profiles in mouse and human fibroblast and lymphoblast cell lines using the S/G1 method revealed high cell-type-specific conservation of replication timing, consistent with reference 90. These authors further showed that evolutionary break points are usually located between regions of similar replication time that are in close proximity.
108. Jorgensen HF, et al. The impact of chromatin modifiers on the timing of locus replication in mouse embryonic stem cells. Genome Biol 2007;8:R169. [PubMed: 17705870]
109. Li J, Santoro R, Koberna K, Grummt I. The chromatin remodeling complex NoRC controls replication timing of rRNA genes. EMBO J 2004;24:120–127. [PubMed: 15577942]
110. Wu R, Singh PB, Gilbert DM. Uncoupling global and fine-tuning replication timing determinants for mouse pericentric heterochromatin. J Cell Biol 2006;174:185–194. [PubMed: 16831888]
111. Hayashi MT, Takahashi TS, Nakagawa T, Nakayama JI, Masukata H. The heterochromatin protein Swi6/HP1 activates replication origins at the pericentromeric region and silent mating-type locus. Nature Cell Biol 2009;11:357–362. [PubMed: 19182789]
112. Duan Z, et al. A three-dimensional model of the yeast genome. Nature 2010;465:363–367. [PubMed: 20436457]

113. Azuara V, et al. Heritable gene silencing in lymphocytes delays chromatid resolution without affecting the timing of DNA replication. Nature Cell Biol 2003;5:668–674. [PubMed: 12833066]
114. Lande-Diner L, Zhang J, Cedar H. Shifts in replication timing actively affect histone acetylation during nucleosome reassembly. Mol Cell 2009;34:767–774. [PubMed: 19560427]
115. Pink CJ, Hurst LD. Timing of replication is a determinant of neutral substitution rates but does not explain slow Y chromosome evolution in rodents. Mol Biol Evol 2010;27:1077–1086. [PubMed: 20026481]
116. Chen CL, et al. Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. Genome Res 2010;20:447–457. [PubMed: 20103589]
117. Hyrien O, Goldar A. Mathematical modelling of eukaryotic DNA replication. Chromosome Res 2010;18:147–161. [PubMed: 20205354]
118. Huberman JA, Riggs AD. Autoradiography of chromosomal DNA fibers from Chinese hamster cells. Proc Natl Acad Sci USA 1966;55:599–606. [PubMed: 5221245]
119. Takebayashi SI, Manders EM, Kimura H, Taguchi H, Okumura K. Mapping sites where replication initiates in mammalian cells using DNA fibers. Exp Cell Res 2001;271:263–268. [PubMed: 11716538]
120. Guan J, Lee LJ. Generating highly ordered DNA nanostrand arrays. Proc Natl Acad Sci USA 2005;102:18321–18325. [PubMed: 16352724]
121. Sekula S, et al. Multiplexed lipid dip-pen nanolithography on subcellular scales for the templating of functional proteins and cell culture. Small 2008;4:1785–1793. [PubMed: 18814174]

Glossary

Pre-replication complex (pre-RC) — A complex of proteins that forms at the origin of replication during the initiation step of DNA replication. All pre-RC proteins are essential for DNA replication. The pre-RC is typically thought to consist of origin recognition complex (ORC), DNA replication factor Cdt1, cell division cycle protein 6 (Cdc6) and mini-chromosome maintenance (MCM) complex

Dormant origin — Cells have a large excess of replication origins over what they need to complete DNA replication. Origins that are in the vicinity of a recently initiated origin normally will be replicated passively when the replication fork passes through them. However, if DNA damage or low-nucleotide pools slow the replication forks, these origins can fire to complete duplication of nearby DNA in a timely fashion

Replication origin — A site where replication is initiated during S phase. It is bound by the origin recognition complex

Replication timing programme — All eukaryotic cells replicate segments of their genomes in a defined temporal sequence. This process is referred to as replication timing. The temporal order in which segments of DNA are replicated is specific to specific cell types, and that temporal order is its replication timing programme

Origin recognition complex — A complex of six subunits that binds to the origins of DNA replication in an ATP-dependent manner before initiation to recruit additional protein members of the pre-replication complex

Chromatin immunoprecipitation — A technique that is used to identify the location of DNA-binding proteins and epigenetic marks in the genome. Genomic sequences containing the mark of interest are enriched by

	binding soluble DNA chromatin extracts (complexes of DNA and protein) to an antibody that recognizes the mark
Mini-chromosome maintenance complex	An oligomeric complex that is suggested to be the helicase involved in replication
Phased nucleosomes	Nucleosomes that are evenly spaced. This usually occurs when a nucleosome is positioned by a DNA sequence or chromatin protein, which restricts the possible locations of its nearest neighbours
Efficiency	The percentage of replication cycles in which any given origin is used as an initiation site
Replication fork	The branch-point structure that forms at the site of active DNA synthesis, where helicases break the hydrogen bonds tethering the two DNA strands and unwind the DNA
Primer extension	Any configuration in which a partially single-stranded nucleic acid is annealed with a 5′ overhang to a smaller complementary strand. The 3′ hydroxyl of the annealed complementary strand can serve as a primer that can be extended by DNA polymerase along the remaining single-stranded portion of the larger template molecule
DNase I hypersensitive site	A region of the genome that is readily degraded by the enzyme DNase I owing to decreased nucleosome occupancy (an ‘open’ chromatin structure)
Replication bubble	The structure formed where two replication forks, derived from the same replication origins, are moving bidirectionally away from the site of initiation. The intervening DNA consists of two newly synthesized strands

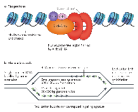

Figure 1. How to find an origin

a | Pre-replication complexes consist of at least 14 different proteins conserved in all eukaryotes: cell division cycle protein 6 (Cdc6), DNA replication factor Cdt1, the heterohexameric origin recognition complex (composed of Orc1 to Orc6), and the heterohexameric mini-chromosome maintenance complex (composed of Mcm2 to Mcm7)[1]. The ORC competes with nucleosomes to bind to DNA and, once bound, is able to position nucleosomes in such a way as to leave sufficient space for MCM complex loading[10]. **b |** Summary of the unique nucleic acid features found near origins of replication. When cells that have been synchronized before the onset of S phase initiate replication in the presence of replication fork inhibitors, replication forks are arrested close to sites of initiation so that any DNA synthesized must be close to origins. The sites where forks are arrested consist of primed templates that can be labelled at the sites of arrest by extension. The leading strands of DNA synthesis quickly become larger than Okazaki fragments and can be isolated as small single-stranded molecules that can be verified to be nascent either by metabolic labelling or by virtue of the fact that nascent strands have small stretches of RNA at their 5′ ends that render them resistant to λ exonuclease[55,56]. Finally, the physical structure of replication origins shortly after initiation is that of a bubble structure, which can be trapped in gelling agarose (FIG. 2).

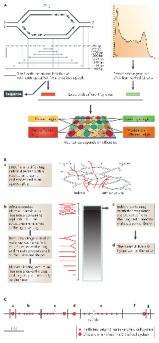

Figure 2. Different methods may enrich for different origins

A | Mapping the positions of small nascent leading strands (SNSs). Genomic DNA from proliferating cells is denatured and single-stranded DNA (ssDNA) is size-fractionated, usually by sucrose density gradient centrifugation. Strands that are larger than Okazaki fragments (50–350 bp) but small enough to be representative of initiation sites (usually 500–1,500 bp) are isolated. Genomic DNA is isolated from cells synchronized in G1 phase by sorting cells with an unreplicated DNA content using a fluorescence-activated cell sorter (FACS). G1 DNA is used as a reference for microarray analysis because all sequences have an identical copy number. SNSs and genomic DNA are then differentially labelled and hybridized to a microarray or directly subjected to deep sequencing. In the microarray schematic, the colours represent the efficiencies of the origins. In principle, the degree of enrichment of probe sequences in the SNS preparation should be proportional to the fraction of cells in which initiation takes place close to any given probe, but contamination from broken, unreplicated DNA must be carefully controlled. **B |** Principle of the bubble-trap method[61]. **Ba |** Initiation of DNA creates replication bubble structures that essentially behave as circular molecules. Hence, when genomic DNA is digested with a restriction enzyme and then mixed with molten agarose and allowed to cool to form an agarose plug, recently initiated origin-containing fragments become trapped in the polymerized agarose. **Bb |** The plug is then subjected to exhaustive electrophoresis, which removes all unreplicated fragments as well as replicating fragments containing branched structures resulting from the entry of replication forks initiated outside the DNA fragment. **Bc |** Origin-containing DNA within the plug is then isolated and cloned. **Bd |** This cloned DNA can be either hybridized to a microarray or sequenced. **C |** Types of origins enriched by each method. Vertical lines depict restriction enzyme cutting sites and red circles depict the positions of replication origins, simplified as two types termed efficient or inefficient. The following describes which method would most easily detect the activity of origins positioned as shown. Note that replicate experiments using different restriction enzymes can overcome some of the limitations of the bubble-trap method. **Ca |** SNSs but not bubble trap. The origin is efficient and localized but the fragment is small and any bubbles formed will be quickly converted to branched or linear structures. **Cb |** Bubble trap but not SNSs. The individual origins fire too infrequently for their localized sites to be detected by SNSs, but most of them will make bubble structures that will trap the fragment regardless of where initiation occurs. **Cc |** Bubble trap and SNSs. The right-most origin will be detected by SNSs and the collection of origins will make detectable bubbles. **Cd |** Bubble trap and SNSs. Because it is an efficient, localized origin, it will be detected by SNS, and because it is positioned at the centre of a sufficiently large restriction fragment, it will generate bubble structures. **Ce |** Bubble trap, but may be difficult for either method. The origins are too inefficient for SNS detection. For bubble trap, large restriction fragments can be retained false-positives. Large fragments are also more difficult to clone into a plasmid library. **Cf |** SNSs but not bubble trap. These efficient origins can be detected by SNSs but are too close to the edge of the fragment to be detected by the bubble-trap method. When initiation occurs off-centre, the fragment rapidly converts to a branched structure when the replication fork crosses the restriction site. **Cg |** Neither method. The origin

is too inefficient for detection by SNSs and the fragment is too small for detection by bubble traps. Part **B** is modified, with permission, from REF. 61 © (2009) Springer.

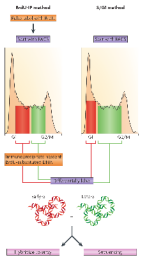

Figure 3. Replication timing analysis by retroactive fluorescence-activated cell sorter synchronization

This method can be applied to any proliferating cell population that can be dissociated into a single-cell suspension. Cells are stained with any dye that fluoresces proportionally to DNA content to produce a histogram of the number of cells with unreplicated (G1 phase), fully replicated (G2/M phase) or increasing (S phase) DNA content. In the 5-bromodeoxyuridine immunoprecipitation (BrdU-IP) method, cells are first pulse-labelled with BrdU to tag replicating DNA in a short (10–20%) interval of S phase and then separated into early- and late-S-phase populations by DNA content using the fluorescence-activated cell sorter (FACS). BrdU-substituted nascent DNA from these populations is immunoprecipitated, differentially labelled and co-hybridized to a high-density whole-genome oligonucleotide microarray. Alternatively, the BrdU-IP DNA can be sequenced. In the S/G1 method, unlabelled cells are sorted into either G1 phase populations (in which the copy numbers of all genomic sequences are equivalent) or total S phase populations (in which the copy numbers are proportional to how early the sequence replicates). DNA is isolated and replication timing is determined as the copy number ratio of S/G1 across the genome either on an array or by sequencing. A link to protocols for these methods, and more information, is provided in the ‘Further Information’ section.