Regulation of replication timing in *Saccharomyces cerevisiae*

Rosie Berners-Lee [1], Eamonn Gilmore [2], Francisco Berkemeier [2,3,✉], and Michael A. Boemo [2,3,✉]

[1]University of St Andrews, KY16 9AJ, St Andrews, United Kingdom
[2]Department of Pathology, University of Cambridge, CB2 1QP, Cambridge, United Kingdom
[3]Department of Genetics, University of Cambridge, CB2 3EH, Cambridge, United Kingdom

✉Corresponding authors: fp409@cam.ac.uk, mb915@cam.ac.uk

In order to maintain genomic integrity, DNA replication must be highly coordinated. Disruptions in this process can cause replication stress which is aberrant in many pathologies including cancer. Despite this, little is known about the mechanisms governing the temporal regulation of DNA replication initiation, thought to be related to the limited copy number of firing factors. Here, we present a high (1-kilobase) resolution stochastic model of *Saccharomyces cerevisiae* whole-genome replication in which origins compete to associate with limited firing factors. After developing an algorithm to fit this model to replication timing data, we validated the model by reproducing experimental inter-origin distances, origin efficiencies, and replication fork directionality. This suggests the model accurately simulates the aspects of DNA replication most important for determining its dynamics. We also use the model to predict measures of DNA replication dynamics which are yet to be determined experimentally and investigate the potential impacts of variations in firing factor concentrations on DNA replication.

DNA Replication | Replication Stress | Origin Efficiency | Firing Factors

Introduction

To maintain genomic integrity, DNA replication must be tightly controlled to ensure that the entire genome is replicated precisely once per cell cycle (Blow and Dutta, 2005). In eukaryotes, replication initiates from multiple sites across the genome known as origins of replication. While the assembly and major enzymatic activity of the replisome is conserved across eukaryotes, there is considerable variability between species as to what genomic features cause loci to act as origins of replication. In *Saccharomyces cerevisiae*, origins occur at defined sequences known as autonomously replicating sequences (ARS) (Stinchcomb et al., 1979). This, along with their short cell cycle and simple, well-understood genome (Goffeau et al., 1996), makes *S. cerevisiae* a good model organism for studying DNA replication. Each ARS shares a common ARS consensus sequence (ACS), which is essential but not sufficient for replication initiation. Local chromatin architecture and nucleosome positioning are also thought to influence the locations of origins (Eaton et al., 2010; Li et al., 2022). The locations of 829 *S. cerevisiae* origins have been mapped and are available in OriDB (Siow et al., 2012).

As depicted in Figure 1a, DNA replication initiation at origins is divided into two temporally distinct stages: licensing and firing (reviewed in Costa and Diffley, 2022). Licensing involves loading of the core of the replicative helicase, the MCM2-7 complex, onto replication origins and occurs at low levels of cyclin-dependent kinase (CDK) and Dbf4-dependent kinase (DDK) activity during late-M and G1 phases of the cell cycle. The process begins with the origin recognition complex (ORC) binding to origins in an ATP-dependent manner. Subsequent recruitment of Cdc6 and MCM2-7/Cdt1 proteins facilitates the loading of a head-to-head double-hexamer of the MCM2-7 replicative helicase around the duplex origin DNA. Importantly, in this form, the MCM2-7 complex is inactive. The resulting pre-replication complex (pre-RC) licences the origin for replication (Bell and Dutta, 2002). The transition from G1 to S phase is accompanied by increased CDK and DDK activity that facilitates origin firing, which involves conversion of the inactive MCM2-7 double hexamer to two active CMG helicases (Reusswig and Pfander, 2019). Increased CDK activity also inhibits further licensing, thereby ensuring only one round of replication per cell cycle. The process of origin firing is driven by the coordinated action of multiple proteins, collectively referred to as “firing factors”. DDK-mediated phosphorylation of the pre-RC drives recruitment of the Sld7-Sld3 complex and Cdc45 (Ramer et al., 2013). Away from the DNA, Sld2 binds to GINS and Pol-ϵ. Phosphorylation of Sld2 and Sld3 by CDK allows both to bind to Dpb11, bringing together their associated proteins at the origin to form the pre-initiation complex (pre-IC) (Zegerman and Diffley, 2007). Sld2, Sld3, and Dpb11 then depart, leaving the remaining CMG helicase (Cdc45, MCM2-7, GINS) and pol-ϵ. Subsequent binding of MCM10, along with other replication factors, results in origin firing by allowing rearrangements of CMG and origin DNA to form the active replisome (Boos and Ferreira, 2019). DNA synthesis then progresses bidirectionally from each fired origin until replication forks are terminated by either converging or reaching the ends of chromosomes (Dewar and Walter, 2017).

Three important and related quantities in genome replication are DNA replication timing, origin efficiency, and

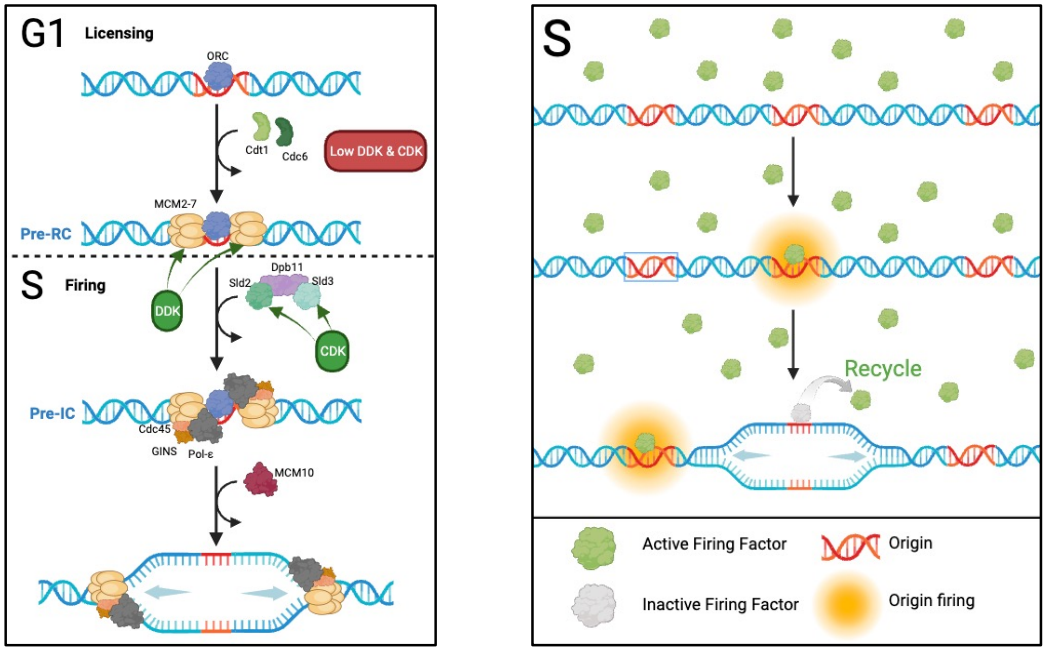


Figure 1. Initiation of DNA replication.
(a) Mechanism of DNA replication initiation at origins. Licensing occurs during G1 at low levels of DDK and CDK activity. Firing occurs during S phase at high levels of DDK and CDK activity. **(b)** DNA replication initiation as represented in our model. Our model simplifies DNA replication initiation to the S phase, representing all necessary proteins for origin firing as a single "firing factor". This factor binds to origins with different affinities and is subsequently recycled for future use. Created in BioRender.com.

origin firing probability (reviewed in (Rhind, 2022a)). DNA replication timing refers to the time in S phase when a genomic locus replicates, often measured over a population of cells. However, only a fraction of licensed origins fire in each cell and the subset of putative replication origins that fires is stochastic; it varies cell-to-cell and over successive cell cycles (McCune et al., 2008; Czajkowsky et al., 2008; Wang et al., 2021b; Bechhoefer and Rhind, 2012). Hence, the replication timing within a particular cell may deviate from the population average (Dileep and Gilbert, 2018). This stochastic nature leads to the concepts of origin efficiency and origin firing probability. An origin's efficiency is defined as the fraction of cells in which that origin fires (Newlon and Theis, 1993) while an origin's firing probability is the probability that an origin fires at a particular time in S phase. Firing probability is often modelled as an exponential distribution such that the rate of origin firing is the sole parameter that determines the shape of the distribution. Origins that do not fire are passively replicated by forks from neighbouring origins. However, these dormant origins still serve an important purpose: They increase the robustness of whole-genome replication by serving as a reserve pool that can be activated if primary origins fail or if replication is challenged, thereby ensuring complete and efficient DNA duplication (Das et al., 2014).

The essential firing factors Sld2, Sld3, Cdc45, Dpb11, and the DDK subunit Dbf4 have all been found to be in low abundance in *S. cerevisiae*. Overexpression of these leads to premature firing of late origins (Mantiero et al., 2011; Tanaka et al., 2011). An origin's affinity for firing factors is influenced by multiple factors including chromatin accessibility and nuclear localisation (Knott et al., 2012; Kitamura et al., 2006; Boos and Ferreira, 2019). Stochastic firing means that, whilst the replication timing profile is reproducible at a population level, the set of origins that fire and their precise firing times differ between cells and cell cycles (Czajkowsky et al., 2008; Wang et al., 2021b; Retkute et al., 2011). Therefore, single-molecule methods such as DNA fibre analysis are needed to study details of DNA replication which are smoothed out in population averaging obtained from bulk experiments (Bianco et al., 2012; Quinet et al., 2017). However, these techniques tend to have low throughput, and while recent long-read sequencing (Müller et al., 2019; Hennion et al., 2020; Theulot et al., 2022) and optical mapping (Wang et al., 2021a) techniques have expanded the toolkit to study DNA replication, mathematical models are a valuable complement to these techniques to study DNA replication dynamics.

Previous work on mathematical modelling of DNA replication in *S. cerevisiae* has employed analytical models based on the Kolmogorov-Johnson-Mehl-Avrami equation to derive local, time-dependent firing rates, capturing genome-wide stochastic initiation patterns (Baker et al., 2012; Yang et al., 2010). Deterministic models have also used origin positions and fork migration rates to predict replication timing profiles (Spiesser et al., 2009), while other studies have highlighted how stochastic origin usage enhances replication robustness (Hyrien and Goldar, 2010). Additionally, other models characterised origin firing on chromo-

some 6 by defining each origin's competence, reflecting firing probability through pre-RC assembly, and a Gaussian-distributed activation time (de Moura et al., 2010; Retkute et al., 2012). Whilst fast to compute, none of these modelling approaches includes firing factors. Rate-limiting firing factors were incorporated into a stochastic model by Arbona et al. (2018) in which firing factors were loaded into the system gradually and traveled with the replication forks, being released upon fork termination. While this is thought to be the case for Cdc45 (Tanaka et al., 2011), it does not account for other firing factors. Stochastically varying fork speeds were incorporated into the model by Yousefi and Rowicka (2019). However, this contrasts with data from (Theulot et al., 2022) which showed a consistent fork speed across the budding yeast genome. A Bayesian algorithm was used by Bazarova et al. (2019) to infer the origin firing time distributions of *S. cerevisiae* chromosome 10 from Okazaki fragment analysis. However, this model only considered three consecutive origins at a time and used "replicated base pairs" rather than conventional time units. A more complex model developed by Brümmer et al. (2010) attempted to describe all of the core processes governing DNA replication initiation in *S. cerevisiae*. However, the challenge of incorporating every relevant factor restricted their model to capturing the kinetics of early, but not late, firing origin.

Whilst a broad range of different measures of DNA replication dynamics have been investigated across the myriad of existing models, differences in assumptions and generalisations between these models makes it difficult to compare their findings. Furthermore, many rely on complex, rigid mathematical equations, making them difficult to interpret and modify. To gain a better understanding of the core mechanisms governing DNA replication dynamics, there is therefore still a need for a DNA replication model which is as simple as possible whilst still capturing the main features and from which many aspects of DNA replication dynamics can be extracted.

Here, we present a stochastic model for *S. cerevisiae* whole-genome DNA replication in which origins compete to associate with limited firing factors, needed for initiation, which then recycle to be used again (Figure 1b) (Mantiero et al., 2011).

Materials and methods

The simulation

Origin positions. We obtained the positions of origins from OriDB (Siow et al., 2012), which catalogues 829 origins, classifying them as 'Confirmed' (410), 'Likely' (216), or 'Dubious' (203). In our model, we only include 'Confirmed' origins (verified by ARS assays and/or 2D gel analysis) and 'Likely' origins (identified in two or more microarray studies). We excluded 'Dubious' origins, as these were identified by only a single microarray study.

Model formulation. We created a stochastic model for *S. cerevisiae* whole-genome replication in which origins compete to associate with limited firing factors, needed for activation, which then recycle to be used again (Figure 1b). To construct our model, we used Beacon Calculus, a process algebra designed for modelling biological systems (Boemo et al., 2020) (Figure S1). In Beacon Calculus, each component of the system is represented as a process capable of performing one or more actions. The Beacon Calculus model is simulated according to a modified Gillespie algorithm, with each action being associated with a rate that represents the parameter of an exponential distribution. This makes models written in the Beacon Calculus inherently stochastic and therefore well-suited to modelling DNA replication. The simplicity of Beacon Calculus also makes it intuitive, straightforward to interpret, and easily modifiable.

In our model, each origin is characterised by four parameters: its chromosome, `[ch]`; position on the chromosome `[i]`; chromosome length `[length]`; and the rate at which it associates with firing factors, `[fire]`. The model was formulated such that chromosomal positions had a spatial resolution of one kilobase (kb). Once an origin associates with a firing factor, it initiates two replication forks which move bidirectionally, replicating DNA at a constant rate of 1.4 kb/min (Friedman et al., 1997; Yousefi and Rowicka, 2019). Forks terminate when they reach either the end of the chromosome or an oncoming replication fork (Dewar and Walter, 2017). In our model, an origin is passively replicated and therefore removed from the system when a fork moves through its position before it can fire. After associating with an origin, firing factors are temporarily inactive while they are recycled for reuse. We assume that firing factors are available immediately following the transition from G1 to S phase (Reusswig and Pfander, 2019). Therefore, all firing factors are present from the beginning of the simulation.

Each simulation of the model can be interpreted as a single S-phase within an individual cell. Given the stochastic nature of the model, each simulation yields variable results, similar to the variability observed in individual cells (Czajkowsky et al., 2008). To capture a representative overview of DNA replication dynamics at the population level, unless otherwise stated, we based our analysis on results averaged from 500 simulations of each model.

Fitting the origin firing rates. The model was fitted to experimental replication timing data iteratively through cycles of model prediction, comparing the predicted and experimental replication timings, and adjusting the firing rate to better fit the replication timing data. For iteration n and origin i, we denote the origin's firing rate as $f_{i,n}$. For the initial iteration $n = 0$, the firing rate at each origin was calculated using the experimentally determined replication timing at that origin (Müller et al., 2014), T_i, and the number of firing factors in the model, F, accord-

ing to

$$f_{i,0} = \frac{1}{F}\frac{1}{T_i}. \quad (1)$$

For subsequent iterations, $n+1$, new firing rates at each origin, $f_{i,n+1}$, were calculated based on the previous firing rate, $f_{i,n}$, and a power (α) of the ratio of its simulated replication timing, $\tilde{T}_{i,n}$, to its experimental replication timing:

$$f_{i,n+1} = f_{i,n}\left(\frac{\tilde{T}_{i,n}}{T_i}\right)^{\alpha}. \quad (2)$$

The parameter α was set to 1.2 to achieve efficient fitting by balancing faster fitting times against the risk of taking large steps that cause instability. While there are alternative fitting methods for unreplicated DNA fractions as discussed in Baker et al. (2012); Baker and Bechhoefer (2014); Gispan et al. (2017), this method is sufficient to capture the prevailing firing rate distribution for given replication timing profiles in *S. cerevisiae*.

Total error. The total error in replication timing was quantified using the Mean Absolute Error (MAE), which was calculated as the mean absolute difference between our simulated replication timing and the experimental data from (Müller et al., 2014) for each genomic locus with 1-kb resolution. The model was fit for 15 iterations, chosen because the MAE's rate of change approached zero beyond this point indicating that further iterations would not substantially improve the fit. The model with the lowest MAE over the fitting iterations was selected as the final model used for the results in this paper.

Model configuration. The essential firing factor identified to have the lowest copy number is Dpb11, estimated at 200 (Tanaka et al., 2011). For simplicity, we assume that Dpb11 is rate-limiting and therefore set F =200. While the recycling rate is more difficult to estimate, this parameter was set conservatively to 0.05 such that the expected recycling time would be approximately one-third of S phase. While the model's robustness meant that it was also able to be fitted using a range of different combinations of firing factor copy numbers and recycling rates, all analysis in this study was based on the fitted model configured with F =200 and a recycling rate of 0.05.

Software and processing

Models were simulated using Beacon Calculus version 1.1.0 (the latest available at the time of writing) https://github.com/MBoemo/bcs, (Boemo et al., 2020). Models, fitting algorithms, and analysis scripts are available at https://github.com/rb2065/S_cerevisiae_DNA_rep_model.

Results

Model convergence. The replication timing predicted by the model converged with the replication timing data from (Müller et al., 2014) over successive iterations to produce a close match (Figure 2). Figure 2a shows how the model's prediction of replication timing compares to data when using the selected model with the lowest MAE against the data over 15 iterations. Figure 2b demonstrates how our initial predictions of origin firing rates were improved upon by successive fitting iterations. As the iterations progressed, we observed the MAE converging to a minimum value of 1.32 minutes by the 15^{th} iteration (as shown in Figure 2c). The R^2 value obtained from comparing the experimentally determined and simulated replication timings was 0.94.

Length of S phase. In our model, S phase duration is defined by the time required, from the beginning of the simulation, to replicate the entire genome. The mean simulated length of S phase was 93.7 $\pm$ 10.8 minutes (Figure 3b). This aligns with experimental findings that place the completion of S phase at approximately 90-100 minutes (Müller et al., 2014; Rivin and Fangman, 1980). The low variation in simulated S phase duration shows that, despite its inherent stochasticity, our model consistently achieves DNA replication completion within a biologically realistic timeframe. Moreover, consistent with experimental results, our model indicates that the majority of DNA replication occurs within the first 60 minutes of S phase (Figure 3a).

Model validation

To validate the model, features of DNA replication dynamics were calculated from the output of 500 model simulations and compared to experimental data that was not used to fit the model.

Inter-origin distances. Inter-origin distance (IOD) is defined as the length of DNA between successive origins which fire during the same cell cycle. The mean IOD determined from the model simulation was 57.4 $\pm$ 30.9 kb (Figure 3c) and was relatively consistent across chromosomes. This aligns closely with the mean IOD of 55.6 $\pm$ 30.3 kb which has been previously established by DNA combing experiments (Devault et al., 2002).

Origin efficiency. Origin efficiencies were calculated as the percentage of simulations in which each origin fired. As expected, more efficient origins tended to have earlier replication timings. The efficiencies of 433 of the 626 origins included our model have been previously determined experimentally from Okazaki fragment analysis (Smith and Whitehouse, 2012). The mean absolute difference between the experimental and simulated origin efficiencies was 18.7% $\pm$ 14%. The mean signed difference of 6.2% $\pm$ 22% is close to zero and this was relatively consistent across all chromosomes (Figure 3d). This indicates that our model's estimation of origin efficiencies is not markedly skewed towards either overestimation or underestimation.

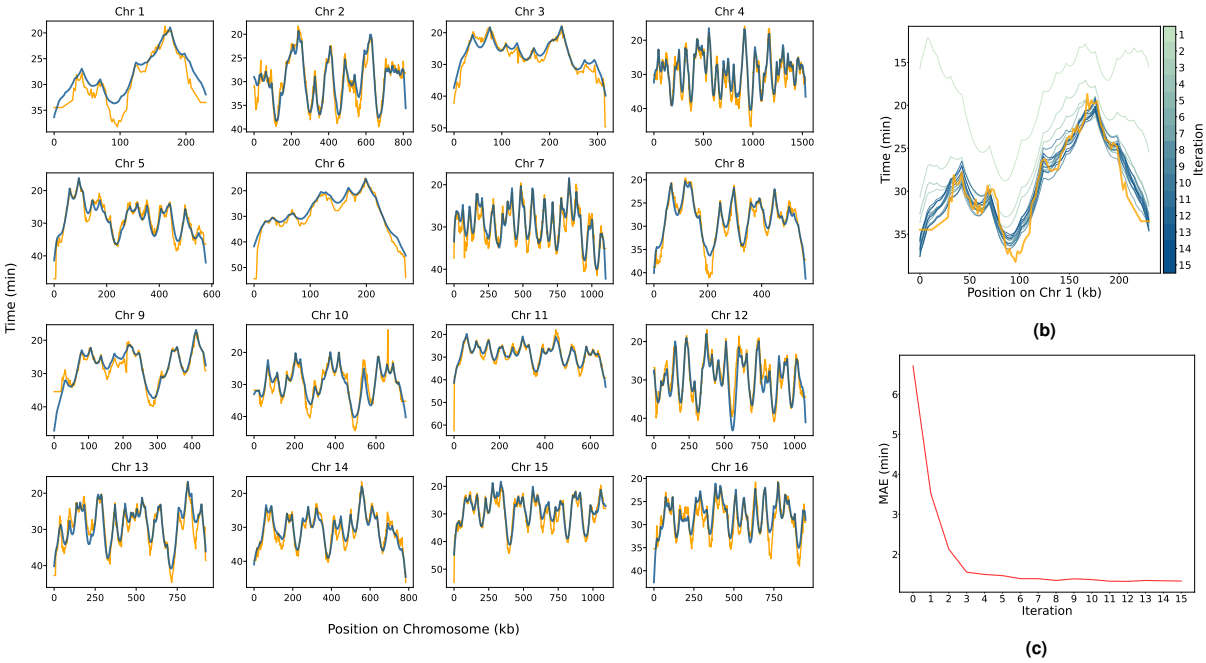


Figure 2. Fitting the model to replication timing data.
(a) Replication timing profiles from simulations using the fitted model (*blue*) and experimental data from (Müller et al., 2014) (*orange*). **(b)** Replication timing profile of chromosome 1 over the fitting, highlighting the incremental convergence towards the real timing profile. Model fitting iterations are represented as a colour gradient from *light green* to *blue*. Experimental data (*orange*) is also shown for comparison. **(c)** The change in Mean Absolute Error (MAE) in replication timing over the fitting iterations.

Replication fork directionality. Replication fork directionality (RFD) is a measure of the proportion of cell cycles in which a particular position is replicated by either the leftward or rightward moving fork. RFD ranges from -1 to 1 where a RFD of -1 corresponds to positions which are always replicated by leftward moving replication forks whilst positions exclusively replicated by rightward moving replication forks have a RFD of 1. Wu et al. (2023) have used Okazaki fragment sequencing (OK-seq) (Petryk et al., 2016) to determine the RFD for the entire *S. cerevisiae* genome. As demonstrated for chromosome 2 in Figure 3e, the simulated RFD from our model was in relatively close agreement with this experimentally determined RFD.
The mean percentage absolute difference between these experimentally determined RFDs and those determined by our model was 19.9% $\pm$ 16.0%. This difference can be partly attributed to the increased noise in the experimental RFD.

Model predictions

Our model is also capable of predicting various aspects of DNA replication that have not yet been fully explored experimentally. While this means these predictions cannot currently be verified by experiments, they offer valuable insights into the dynamics of DNA replication and suggest directions for future research.

Active replication forks. Our model predicts the mean number of replication forks which are active at each time point over the simulated S phase (Figure 4a). This reached a maximum of 200 active replication forks at 22 minutes into S phase.

Distribution of origin firing times. From our model, the distribution of firing times for each origin can be extracted. Using a subset of chromosome 2 origins as an example, Figure 4b illustrates how firing time distributions correspond to origin efficiencies. This is shown for all chromosome 2 origins in Figure S2. More efficient origins do not necessarily fire earlier or have less varied firing time distributions. Given the limited components in our minimal model, this suggests positioning and context among other origins are more important than efficiency in determining firing time which is consistent with the discussion in (Rhind, 2022a). Additionally, many origins exhibited weakly bimodal firing time distributions, which also did not correlate with their efficiencies.

Replicons. Our model's capacity to track individual replication forks also enables the investigation of replicons, defined as the segment of DNA replicated from a single origin. Replicon length is an important metric for investigating DNA replication dynamics because, as well as the efficiency of the origin in question, it is also influenced by its proximity to neighboring origins and their respective efficiencies. From our model, we calculated the average replicon length for each origin. Figure 4c exemplifies this for chromosome 2. As expected, there was a strong positive correlation between an origin's efficiency and its average replicon length. Although Claussin et al. (2022) have used the single molecule technique Replicon-seq to study replicons in *S. cere-*

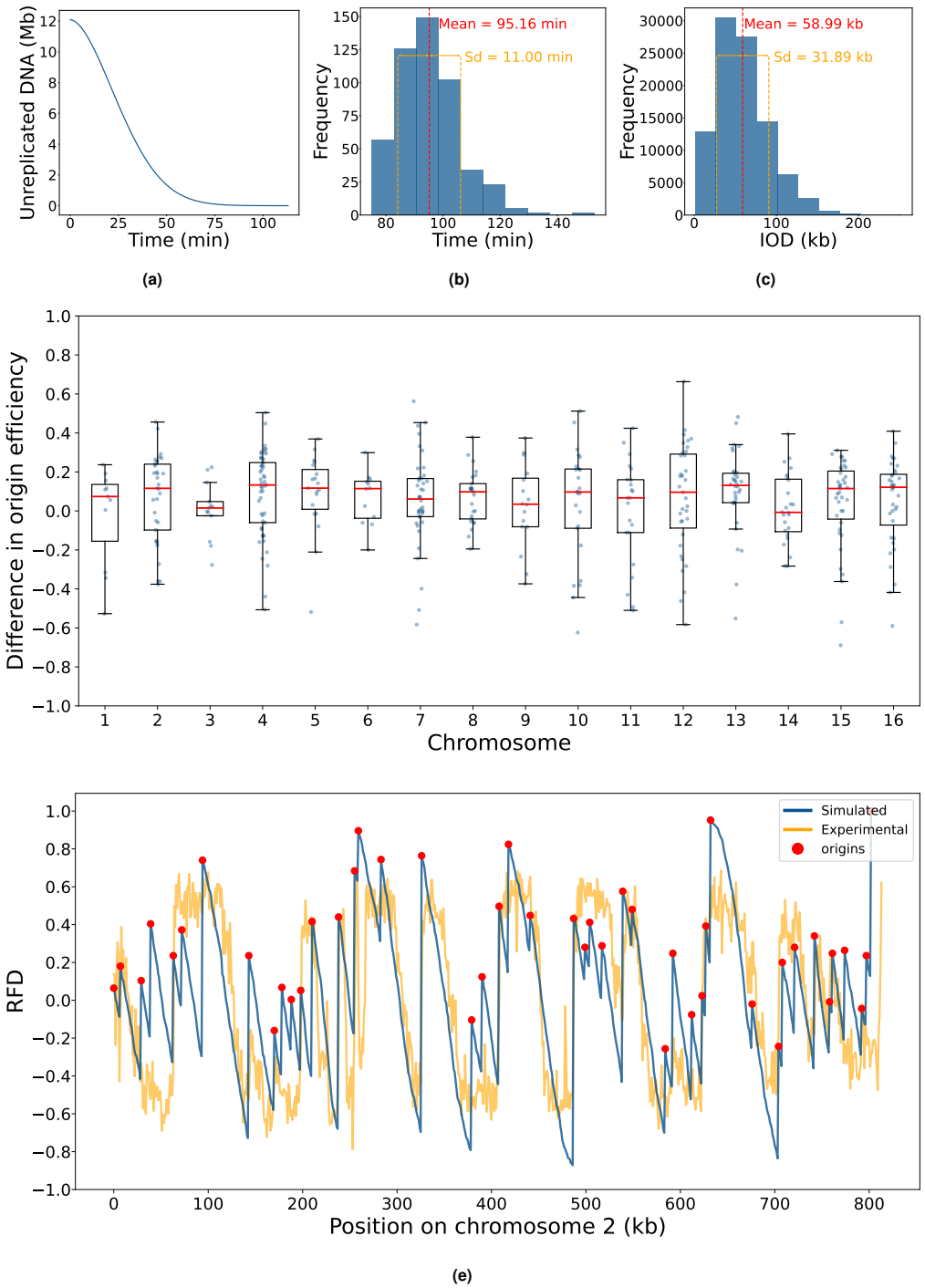


Figure 3. Comparing simulated and experimental replication dynamics.
(a) Length of unreplicated DNA over time. **(b)** Distribution of the total time taken to complete DNA replication (lengths of S phase). **(c)** Distribution of inter-origin distances (IODs). The mean (*red*) and standard deviation (*orange*) are shown. **(d)** Box plots of the signed difference between simulated and experimentally determined origin efficiency across each Chromosome. Efficiency differences for individual origins are shown as scatter points (*blue*). **(e)** Simulated (*blue*) and experimentally determined (*orange*) replication fork drectionality (RFD) for chromosome 2. Positions of origins are shown as *red points*. RFD represents the proportion leftward and rightward moving forks which replicate each position, ranging from -1 (exclusively leftward moving forks) to 1 (exclusively rightward forks). All plots were derived from 500 simulation a version of the model in which F =200 and the recycling rate was 0.05.

visiae experimentally, their focus on early S phase and the potential bias of Replicon-seq towards detecting shorter replicons limits its use for validating our model. Furthermore, by determining replicon lengths beyond just the early stages of replication, our model provides a broader overview which captures the dynamics of both early and late firing origins.

Variation in replication timing. Our model enables us to predict variability in replication timing, as shown for chromosome 2 in Figure 4d. Notably, we observed local maxima in variability at the boundaries of origin clusters. The average standard deviation for replication timings at the origins stood at 16.82 $\pm$ 2.36 minutes whereas the average standard deviation for replication timings across the entire genome was slightly lower, at 15.50 $\pm$ 2.14 minutes.

Influence of firing factors on DNA replication

Impact of firing factor availability. In our model, the rate at which origins fire is determined by the interplay between three factors: the affinity of individual origins for firing factors, the abundance of "free" firing factors able to associate with origins, and the number of available origins yet to fire or be passively replicated. This dynamic, averaged over 500 simulations, is illustrated in Figure 5a. The simulation begins with maximal availability of firing factors and origins, causing high initial origin firing. This initial swell of activity reduces the pool of both free firing factors and available origins, subsequently leading to a decline in the rate of origin firing. As the simulation progresses, firing factors are recycled back into the available pool. This increases the probability of some of the few remaining origins fire, thereby facilitating the timely completion of genome replication within the duration of the S phase. Of the 626 origins included our model, a mean of only 206.43 $\pm$ 6.72 origins fire per simulation.

Altering the number of firing factors. We explored the influence of firing factor concentration variations on DNA replication dynamics by manipulating the copy number of firing factors introduced into the model. To isolate the effects of firing factor concentration changes, we maintained the constant recycling rate of 0.05 and kept the origin firing rates which were derived from fitting the baseline model configuration in which F =200. We quantitatively assessed the impact of altering the firing factor copy numbers by computing the MAE in replication timing across a range of values from F =25 to F =800. This data was subsequently fitted a fifth-degree polynomial (Figure 5b). As expected, increasing the number of firing factors led to earlier replication timing and a higher number of origins firing, whereas decreasing the number of firing factors produced the opposite effect. However, the shape of the replication timing curve remained remarkably robust despite this severe perturbation. This is illustrated for chromosome 2 in Figure 5c.

As shown in Figure 5b, successive increases in firing factor copy numbers causes diminishing increases in error as they become less limiting and therefore their availability has less influence on origin firing dynamics. Conversely, reducing the copy number of firing factors incrementally increases the error. This is because, in our model, firing factors are essential for initiating origin firing; therefore, a minimum number is required to ensure sufficient origins can activate to complete DNA replication within the expected timeframe.

Discussion

Despite being central to maintaining genetic integrity, the mechanisms underlying the temporal regulation of DNA replication remain poorly understood. Since DNA replication is a stochastic process and therefore differs between cells, valuable information is lost in bulk experiments which use population averages (Czajkowsky et al., 2008; Wang et al., 2021b). Whilst the development of high-throughput, single-molecule experimental techniques have advanced our ability to investigate DNA replication, the complex cellular environment means it is still difficult to identify the features most important for determining its dynamics (Bianco et al., 2012; Quinet et al., 2017; Georgieva et al., 2020; Claussin et al., 2022; Rhind, 2022b; Theulot et al., 2022).

Here, we created a stochastic model for *S. cerevisiae* whole-genome replication from first principles. By making the model as simple as possible whilst still able to reproduce experimentally determined measures of DNA replication dynamics, we highlight the features which are most important for determining these dynamics. The ability to rapidly conduct thousands of simulations of our model offers a scale of repetition far surpassing what experimental methods can feasibly achieve. Moreover, conventional experimental techniques typically focus on only a few aspect of DNA replication dynamics. In contrast, our model allows numerous features to be calculated from a single output, providing a more comprehensive insight.

In simplifying the model, we used firing factors, recycling rates, and origin firing rates to combine many complex features affecting DNA replication into a few basic concepts. In our model, firing factors represent the completed pre-IC. The number of available firing factors represents any protein or protein complex essential for forming the pre-IC that may be in limiting abundance at any given time during S phase. For instance, this may represent the availability of the complex formed during pre-IC assembly between Sld2, Sld3, and Dpb11 which are all present in limited abundance (Tanaka et al., 2011). This complex is released before origin firing and therefore does not move with the replication forks (Boos and Ferreira, 2019; Zegerman and Diffley, 2007). Based on this, in our model, firing factor recycling is independent of replication fork termination and, instead,

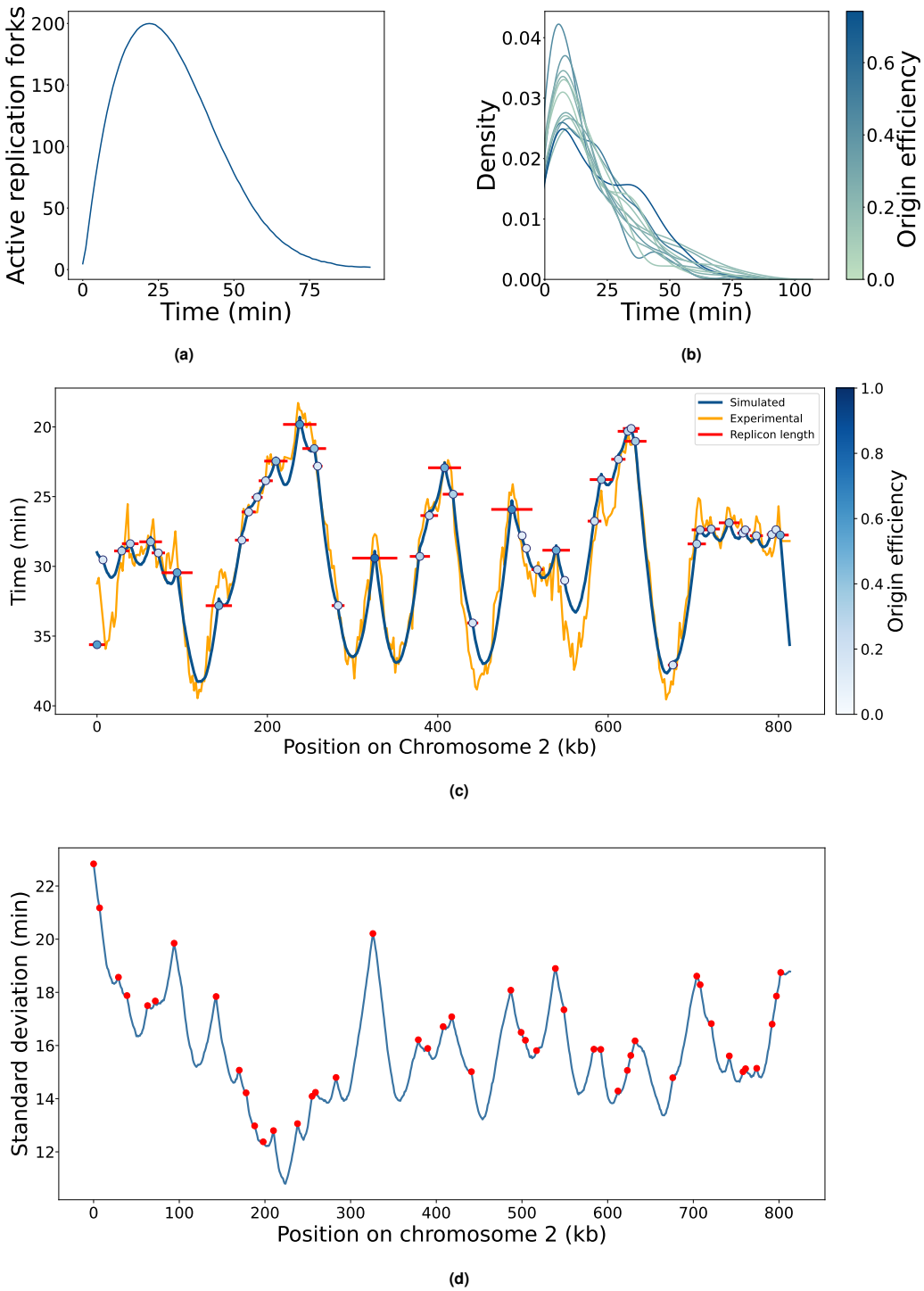
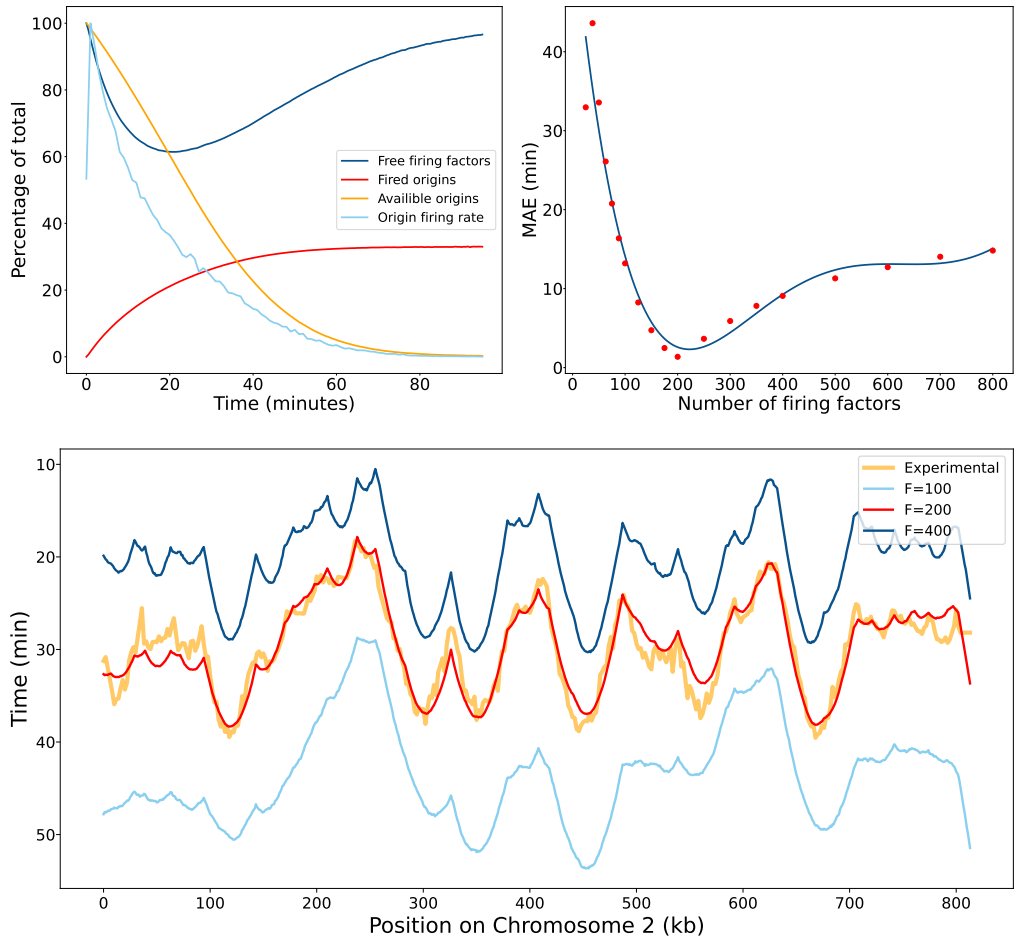


Figure 4. Predictions of DNA replication dynamics.
(a) The number of active replication forks over time. For this plot, the time has been cropped at the mean simulation length. These plots were derived from a version of the model which used $F = 200$ and a recycling rate of 0.05. **(b)** Kernel density plot showing the distribution of firing times for origin on chromosome 2. Each line represents a separate origin, with the shade colour gradient from *light green* to *dark blue* reflecting its efficiency. For improved visualisation, only 1 in 4 origins, ordered by efficiency, are shown. **(c)** Chromosome 2 replication dynamics. Simulated and experimental replication timing profiles are shown in *blue* and *orange* respectively. Origins are shown as *circles* with the shade of *blue* reflecting their efficiencies. The average replicon length of each origin is shown in *red*. **(d)** The standard deviation of simulated replication timing for each kb on chromosome 2. The positions of origins are shown in *red*. All plots were derived from 500 simulation a version of the model in which $F = 200$ and the recycling rate was 0.05.

recycling begins at a set rate immediately following firing. Of the essential firing factors thought to be in limited abundance, only Cdc45 is known to travel with replication forks. Over-expression of Cdc45 alone is insufficient to cause premature firing of late origins, suggesting it is not solely responsible for limiting origin firing (Mantiero et al., 2011). Despite this, the DNA replication model developed by Arbona et al. (2018) employed a fork-dependent mechanism for recycling firing factors. This assumes that the firing factors travel with the replication forks, only becoming available again once the forks terminate. However, the ability of our model to



(c)

Figure 5. Influence of firing factor availability on origin firing dynamics.
(a) The percentage of active firing factors (*dark blue*), fired origins (*orange*), origins available to fire (*red*) and rate of origin firing over time (*light blue*). Here, available origins refers to origins which have not already either fired or been passively replicated. For this plot, the time has been cropped at the mean simulation length. **(b)** Mean Absolute Error (MAE) in replication timing with different numbers of firing factors. Data points from simulations (*red*) have been fit to a fifth-degree polynomial (*blue*). **(c)** Simulated replication timing profiles for chromosome 2 from models using F =100 (*light blue*), F =200 (*red*), and F =400 (*dark blue*). In all models, origin firing fates were fitted using F =200. The experimentally determined replication timing profile is also shown for comparison (*orange*).

be fitted and validated without fork-dependent recycling suggests that it is not strictly necessary to model DNA replication dynamics. In our model, recycling rates represent a generic time-lag which encapsulates any processes which may delay the ability of firing factors to activate new origins following their release from a fired origin. This could include the time required for diffusion and the assembly of protein complexes required to construct the replisome.

Based on the assumption of a rapid transition from G1 to S phase, in our model, all firing factors are available from the beginning of the simulation. However, although the transition from G1 to S-phase is accompanied by a switch-like increase in CDK and DDK activity, it is possible that the subsequent activation of firing factors required for initiating DNA replication is more gradual (Reusswig and Pfander, 2019). This would be expected to increase the competition for firing factors at the beginning of S phase, thereby decreasing the initial rate of origin firing.

Whilst the number of firing factors and their recycling rates affect the firing probability of all origins uniformly, individual differences in firing probability arise from each origin's specific assigned firing rate. An origin's firing rate encompasses anything which may impact its ability to associate with firing factors, including sequence composition and local chromatin structure. There is evidence suggesting that the probabilities of origins firing are dynamically regulated throughout S phase by mechanisms such as cooperative firing, the formation and dissolution of replication factories, and changes in chromatin environments (Fangman et al., 2008; Kitamura et al., 2006; Knott et al., 2012; Meister et al., 2007; Hao et al., 2023; Kurat et al., 2017). However, our model simplifies these dynamics by assuming constant origin firing rates for the entirety of S phase.

Our model also assumes a constant rate of fork movement in order to be consistent with findings from Theulot et al. (2022) which demonstrated a remarkable uniformity in replication fork speeds across the *S. cerevisiae* genome. Theulot et al. (2022) also observed a slight acceleration of fork speeds during the S phase

and identified specific genomic regions, such as centromeres, telomeres, and tRNA genes, where replication forks progress more slowly, likely due to increased pausing and stalling at replication barriers (Mirkin and Mirkin, 2007; Yeung and Smith, 2020). Neither increasing fork speeds nor fork stalling are included in our model. Despite these simplifications, our model's ability to successfully reproduce experimentally determined measures of *S. cerevisiae* DNA replication suggests that such features may not be central to determining the overall dynamics of DNA replication in *S. cerevisiae*. Whilst changing fork speeds and fork stalling could also easily be incorporated into our model, the constant rate of fork movement also limits the maximum gradient of the DNA replication timing profile which helps to avoid over fitting.

Our model's capacity to predict aspects of DNA replication dynamics which have not yet been determined experimentally offers novel insights and directs future research. For instance, by calculating the number of active replication forks throughout S phase, our model could highlight periods when cells may be particularly susceptible to replication stress or damage. Moreover, our model predicted that the distribution of an origin's firing time is not associated with its efficiency, indicated a non-trivial relationship between the average replication time and the standard deviation of replication time, and predicted that replication timing curves do not particularly change shape under overexpression and underexpression of firing factors. The simplicity of our model not only identifies the minimal components necessary to reconstruct replication timing profiles, but it also shows how the shape of the replication timing curve remains remarkably robust to insults such as changes in the copy number of firing factors.

ACKNOWLEDGEMENTS

We thank all members of the Boemo Lab (University of Cambridge Departments of Pathology and Genetics) and Gideon Coster (Institute of Cancer Research) for helpful discussions and comments that greatly improved the manuscript. We also thank the University of St Andrews MBiochem program for their support and guidance. This work was made possible by the Leverhulme Trust Research Project Grant RPG-2022-028. This work was performed using resources provided by the Cambridge Service for Data Driven Discovery (CSD3) operated by the University of Cambridge Research Computing Service (www.csd3.cam.ac.uk), provided by Dell EMC and Intel using Tier-2 funding from the Engineering and Physical Sciences Research Council (capital grant EP/T022159/1), and DiRAC funding from the Science and Technology Facilities Council (www.dirac.ac.uk).

COMPETING INTERESTS

The authors declare no competing interests that could have influenced the work in this paper.

References

Jean-Michel Arbona, Arach Goldar, Olivier Hyrien, Alain Arneodo, and Benjamin Audit. The eukaryotic bell-shaped temporal rate of DNA replication origin firing emanates from a balance between origin activation and passivation. *Elife*, 7:e35192, 2018.

A Baker and J Bechhoefer. Inferring the spatiotemporal DNA replication program from noisy data. *Physical Review E*, 89(3):032703, 2014.

A Baker, B Audit, SC-H Yang, J Bechhoefer, A Arneodo, et al. Inferring where and when replication initiates from genome-wide replication timing data. *Physical review letters*, 108(26):268101, 2012.

Alina Bazarova, Conrad A Nieduszynski, Ildem Akerman, and Nigel J Burroughs. Bayesian inference of origin firing time distributions, origin interference and licencing probabilities from Next Generation Sequencing data. *Nucleic acids research*, 47(5):2229–2243, 2019.

John Bechhoefer and Nicholas Rhind. Replication timing and its emergence from stochastic processes. *Trends in Genetics*, 28(8):374–381, 2012.

Stephen P Bell and Anindya Dutta. DNA replication in eukaryotic cells. *Annual review of biochemistry*, 71(1):333–374, 2002.

Julien N Bianco, Jérôme Poli, Julie Saksouk, Julien Bacal, Maria Joao Silva, Kazumasa Yoshida, Yea-Lih Lin, Hélène Tourrière, Armelle Lengronne, and Philippe Pasero. Analysis of DNA replication profiles in budding yeast and mammalian cells using DNA combing. *Methods*, 57(2):149–157, 2012.

J Julian Blow and Anindya Dutta. Preventing re-replication of chromosomal DNA. *Nature reviews Molecular cell biology*, 6(6):476–486, 2005.

Michael A Boemo, Luca Cardelli, and Conrad A Nieduszynski. The Beacon Calculus: A formal method for the flexible and concise modelling of biological systems. *PLoS computational biology*, 16(3):e1007651, 2020.

Dominik Boos and Pedro Ferreira. Origin firing regulations to control genome replication timing. *Genes*, 10(3):199, 2019.

Anneke Brümmer, Carlos Salazar, Vittoria Zinzalla, Lilia Alberghina, and Thomas Höfer. Mathematical modelling of DNA replication reveals a trade-off between coherence of origin activation and robustness against rereplication. *PLoS computational biology*, 6(5): e1000783, 2010.

Clémence Claussin, Jacob Vazquez, and Iestyn Whitehouse. Single-molecule mapping of replisome progression. *Molecular Cell*, 82(7):1372–1382, 2022.

Alessandro Costa and John FX Diffley. The initiation of eukaryotic DNA replication. *Annual Review of Biochemistry*, 91:107–131, 2022.

Daniel M Czajkowsky, Jie Liu, Joyce L Hamlin, and Zhifeng Shao. DNA combing reveals intrinsic temporal disorder in the replication of yeast chromosome VI. *Journal of molecular biology*, 375(1):12–19, 2008.

Mitali Das, Sunita Singh, Satyajit Pradhan, and Gopeshwar Narayan. MCM paradox: abundance of eukaryotic replicative helicases and genomic integrity. *Molecular biology international*, 2014, 2014.

Alessandro PS de Moura, Renata Retkute, Michelle Hawkins, and Conrad A Nieduszynski. Mathematical modelling of whole chromosome replication. *Nucleic acids research*, 38 (17):5623–5633, 2010.

Alain Devault, Elizabeth A Vallen, Tina Yuan, Stephen Green, Aaron Bensimon, and Etienne Schwob. Identification of Tah11/Sid2 as the ortholog of the replication licensing factor Cdt1 in Saccharomyces cerevisiae. *Current Biology*, 12(8):689–694, 2002.

James M Dewar and Johannes C Walter. Mechanisms of DNA replication termination. *Nature Reviews Molecular Cell Biology*, 18(8):507–516, 2017.

V. Dileep and D.M. Gilbert. Single-cell replication profiling to measure stochastic variation in mammalian replication timing. *Nature Communications*, 9, 2018.

Matthew L Eaton, Kyriaki Galani, Sukhyun Kang, Stephen P Bell, and David M MacAlpine. Conserved nucleosome positioning defines replication origins. *Genes & development*, 2010.

WL Fangman, BJ Brewer, and MK Raghuraman. The temporal program of chromosome replication: genome-wide replication in clb5D Saccharomyces cerevisiae. *Genetics. doi: http://dx. doi. org/10.1534/genetics*, 108, 2008.

Katherine L Friedman, Bonita J Brewer, and Walton L Fangman. Replication profile of Saccharomyces cerevisiae chromosome VI. *Genes to Cells*, 2(11):667–678, 1997.

Daniela Georgieva, Qian Liu, Kai Wang, and Dieter Egli. Detection of base analogs incorporated during dna replication by nanopore sequencing. *Nucleic Acids Research*, 48(15): e88–e88, 2020.

Ariel Gispan, Miri Carmi, and Naama Barkai. Model-based analysis of DNA replication profiles: predicting replication fork velocity and initiation rate by profiling free-cycling cells. *Genome Research*, 27(2):310–319, 2017.

André Goffeau, Bart G Barrell, Howard Bussey, Ronald W Davis, Bernard Dujon, Heinz Feldmann, Francis Galibert, Jörg D Hoheisel, Claude Jacq, Michael Johnston, et al. Life with 6000 genes. *Science*, 274(5287):546–567, 1996.

Lei Hao, Ruixin Fang, and Haizhen Long. Chromatin-based DNA replication initiation regulation in eukaryotes. *Genome Instability & Disease*, 4(5):275–288, 2023.

Magali Hennion, Jean-Michel Arbona, Laurent Lacroix, Corinne Cruaud, Bertrand Theulot, Benoît Le Tallec, Florence Proux, Xia Wu, Elizaveta Novikova, Stefan Engelen, Arnaud Lemainque, Benjamin Audit, and Olivier Hyrien. Fork-seq: replication landscape of the saccharomyces cerevisiae genome by nanopore sequencing. *Genome Biology*, 21(1): 125, May 2020.

Olivier Hyrien and Arach Goldar. Mathematical modelling of eukaryotic DNA replication. *Chromosome Research*, 18(1):147–161, 2010.

Etsushi Kitamura, J Julian Blow, and Tomoyuki U Tanaka. Live-cell imaging reveals replication of individual replicons in eukaryotic replication factories. *Cell*, 125(7):1297–1308, 2006.

Simon RV Knott, Jared M Peace, A Zachary Ostrow, Yan Gan, Alexandra E Rex, Christopher J Viggiani, Simon Tavaré, and Oscar M Aparicio. Forkhead transcription factors establish origin timing and long-range clustering in S. cerevisiae. *Cell*, 148(1):99–111, 2012.

Christoph F Kurat, Joseph TP Yeeles, Harshil Patel, Anne Early, and John FX Diffley. Chromatin controls DNA replication origin selection, lagging-strand synthesis, and replication fork rates. *Molecular cell*, 65(1):117–130, 2017.

Sai Li, Michael R Wasserman, Olga Yurieva, Lu Bai, Michael E O'Donnell, and Shixin Liu. Nucleosome-directed replication origin licensing independent of a consensus DNA sequence. *Nature communications*, 13(1):4947, 2022.

Davide Mantiero, Amanda Mackenzie, Anne Donaldson, and Philip Zegerman. Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast. *The EMBO journal*, 30(23):4805–4814, 2011.

Heather J McCune, Laura S Danielson, Gina M Alvino, David Collingwood, Jeffrey J Delrow, Walton L Fangman, Bonita J Brewer, and MK Raghuraman. The temporal program of chromosome replication: genomewide replication in clb5 δ Saccharomyces cerevisiae. *Genetics*, 180(4):1833–1847, 2008.

Peter Meister, Angela Taddei, Aaron Ponti, Giuseppe Baldacci, and Susan M Gasser. Replication foci dynamics: replication patterns are modulated by S-phase checkpoint kinases in fission yeast. *The EMBO journal*, 26(5):1315–1326, 2007.

Ekaterina V Mirkin and Sergei M Mirkin. Replication fork stalling at natural impediments. *Microbiology and molecular biology reviews*, 71(1):13–35, 2007.

Carolin A Müller, Michelle Hawkins, Renata Retkute, Sunir Malla, Ray Wilson, Martin J Blythe, Ryuichiro Nakato, Makiko Komata, Katsuhiko Shirahige, Alessandro PS de Moura, et al. The dynamics of genome replication using deep sequencing. *Nucleic acids research*, 42 (1):e3–e3, 2014.

Carolin A. Müller, Michael A. Boemo, Paolo Spingardi, Benedikt M. Kessler, Skirmantas Kriaucionis, Jared T. Simpson, and Conrad A. Nieduszynski. Capturing the dynamics of genome replication on individual ultra-long nanopore sequence reads. *Nature Methods*, 16(5):429–436, May 2019.

Carol S. Newlon and James F. Theis. The structure and function of yeast ARS elements. *Current Opinion in Genetics & Development*, 3:752–758, 1993.

Nataliya Petryk, Malik Kahli, Yves d'Aubenton Carafa, Yan Jaszczyszyn, Yimin Shen, Maud Silvain, Claude Thermes, Chun-Long Chen, and Olivier Hyrien. Replication landscape of the human genome. *Nature communications*, 7(1):10208, 2016.

Annabel Quinet, Denisse Carvajal-Maldonado, Delphine Lemacon, and Alessandro Vindigni. DNA fiber analysis: mind the gap! *Methods in enzymology*, 591:55–82, 2017.

Matthew D Ramer, Evelyin S Suman, Hagen Richter, Karen Stanger, Martina Spranger, Nicole Bieberstein, and Bernard P Duncker. Dbf4 and Cdc7 proteins promote DNA replication through interactions with distinct Mcm2–7 protein subunits. *Journal of Biological Chemistry*, 288(21):14926–14935, 2013.

Renata Retkute, Conrad A Nieduszynski, and Alessandro De Moura. Dynamics of DNA replication in yeast. *Physical review letters*, 107(6):068103, 2011.

Renata Retkute, Conrad A Nieduszynski, and Alessandro De Moura. Mathematical modeling of genome replication. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 86(3):031916, 2012.

Karl-Uwe Reusswig and Boris Pfander. Control of eukaryotic DNA replication initiation—mechanisms to ensure smooth transitions. *Genes*, 10(2):99, 2019.

Nicholas Rhind. DNA replication timing: Biochemical mechanisms and biological significance. *BioEssays*, 44(11):2200097, 2022a.

Nicholas Rhind. Mapping replication forks, one replicon at a time. *Molecular cell*, 82(7): 1246–1248, 2022b.

Carol J Rivin and Walton L Fangman. Cell cycle phase expansion in nitrogen-limited cultures of saccharomyces cerevisiae. *The Journal of cell biology*, 85(1):96–107, 1980.

Cheuk C Siow, Sian R Nieduszynska, Carolin A Müller, and Conrad A Nieduszynski. OriDB, the DNA replication origin database updated and extended. *Nucleic acids research*, 40 (D1):D682–D686, 2012.

Duncan J Smith and Iestyn Whitehouse. Intrinsic coupling of lagging-strand synthesis to chromatin assembly. *Nature*, 483(7390):434–438, 2012.

Thomas Wolfgang Spiesser, Edda Klipp, and Matteo Barberis. A model for the spatiotemporal organization of dna replication in saccharomyces cerevisiae. *Molecular Genetics and Genomics*, 282:25–35, 2009.

DT Stinchcomb, K Struhl, and RW Davis. Isolation and characterisation of a yeast chromosomal replicator. *Nature*, 282(5734):39–43, 1979.

Seiji Tanaka, Ryuichiro Nakato, Yuki Katou, Katsuhiko Shirahige, and Hiroyuki Araki. Origin association of Sld3, Sld7, and Cdc45 proteins is a key step for determination of origin-firing timing. *Current Biology*, 21(24):2055–2063, 2011.

Bertrand Theulot, Laurent Lacroix, Jean-Michel Arbona, Gael A Millot, Etienne Jean, Corinne Cruaud, Jade Pellet, Florence Proux, Magali Hennion, Stefan Engelen, et al. Genome-wide mapping of individual replication fork velocities using nanopore sequencing. *Nature Communications*, 13(1):3295, 2022.

Weitao Wang, Kyle N. Klein, Karel Proesmans, Hongbo Yang, Claire Marchal, Xiaopeng Zhu, Tyler Borrman, Alex Hastie, Zhiping Weng, John Bechhoefer, Chun-Long Chen, David M. Gilbert, and Nicholas Rhind. Genome-wide mapping of human dna replication by optical replication mapping supports a stochastic model of eukaryotic replication. *Molecular Cell*, 81(14):2975–2988.e6, 2021a.

Weitao Wang, Kyle N Klein, Karel Proesmans, Hongbo Yang, Claire Marchal, Xiaopeng Zhu, Tyler Borrman, Alex Hastie, Zhiping Weng, John Bechhoefer, et al. Genome-wide mapping of human DNA replication by optical replication mapping supports a stochastic model of eukaryotic replication. *Molecular cell*, 81(14):2975–2988, 2021b.

Xia Wu, Yaqun Liu, Yves d'Aubenton Carafa, Claude Thermes, Olivier Hyrien, Chun-Long Chen, and Nataliya Petryk. Genome-wide measurement of DNA replication fork directionality and quantification of DNA replication initiation and termination with Okazaki fragment sequencing. *Nature Protocols*, 18(4):1260–1295, 2023.

Scott Cheng-Hsin Yang, Nicholas Rhind, and John Bechhoefer. Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing. *Molecular systems biology*, 6(1):404, 2010.

Rani Yeung and Duncan J Smith. Determinants of replication-fork pausing at tRNA genes in Saccharomyces cerevisiae. *Genetics*, 214(4):825–838, 2020.

Razie Yousefi and Maga Rowicka. Stochasticity of replication forks' speeds plays a key role in the dynamics of DNA replication. *PLoS Computational Biology*, 15(12):e1007519, 2019.

Philip Zegerman and John FX Diffley. Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature*, 445(7125):281–285, 2007.

Supplementary Information

The Beacon Calculus model

```
fast = 100000; //fast rate
v = 1.4; //fork velocity in kilobases per minute

//process definitions

FF[] = {@factor![0],1}.{dwell,0.05}.FF[];

ORI[i,ch,length,fire] = {@factor?[0],fire}.(FL[i,ch,length]||FR[i,ch,length])
    + {ch?[i],fast};

FR[i,ch,length] = {ch![i],fast}.[i < length] -> {~ch?[i+1],v}.FR[i+1,ch,length];

FL[i,ch,length] = {ch![i],fast}.[i > 0] -> {~ch?[i-1],v}.FL[i-1,ch,length];
```

Figure S1. The Beacon Calculus model.
Beacon Calculus code used for the model. Processes and comments are highlighted in *pink* and *green* respectively. The code includes process definitions for firing factors, `FF`, origins, `ORI`, and replication forks, `FR` and `FL`. Comments within the code are indicated by "\\". Actions are enclosed within "{ }" and are defined as ordered pairs, specifying the action followed by the rate at which it occurs. Handshake communications are denoted by `@factor!` for sending and `@factor?` for receiving on the `factor` channel. Beacon actions are represented by `ch!` for sending, `~ch?` for checking, and `ch?` for receiving, all on the `ch` channel. The values within "`[]`" following handshake or beacon are transmitted. The code syntax includes "`.`" for sequential statements, "`|`" for parallel statements, and "`+`" for making exclusive choices. Condition gates are represented by "`->`". All origin and firing factor processes are initiated from the beginning of the simulation. However, this has been omitted from this representation for conciseness.

Origin firing rates

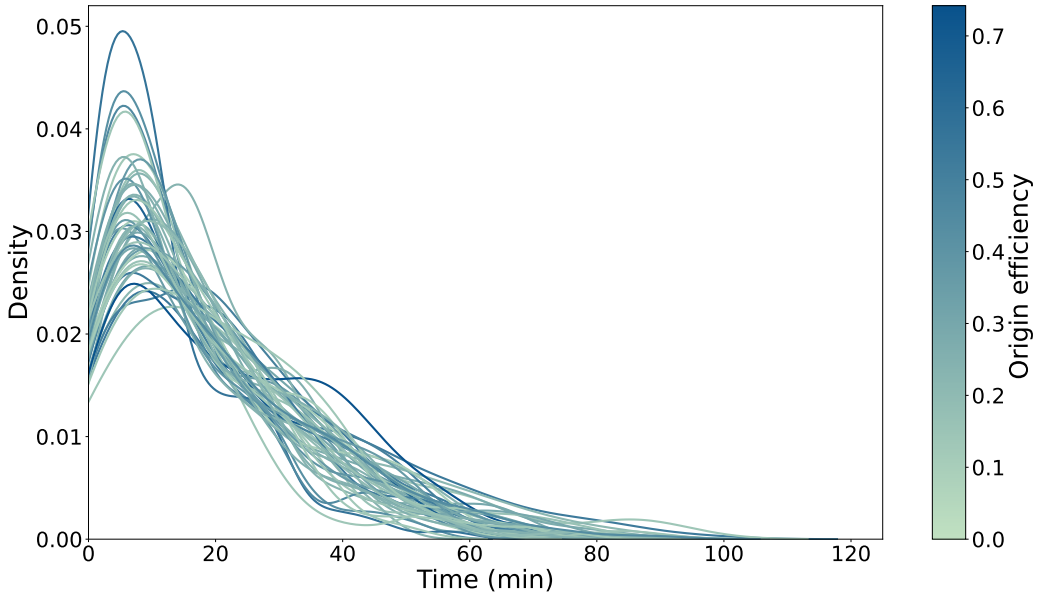


Figure S2. Chromosome 2 origin firing time distributions.
Kernel density plot showing the distributions of origin firing times for all origins on chromosome 2. Each line represents a separate origin, with the colour gradient from *light green* to *dark blue* reflecting its efficiency.