Processive activity of the replicative DNA polymerases in the replisome of live eukaryotic cells

Nitin Kapadia, Ziad W. El-Hajj, Thomas R. Beattie, Angela Yu, Rodrigo Reyes-Lamothe*

Department of Biology, McGill University, 3649 Sir William Osler, Montreal, Qc, H3G 0B1 Canada

* Correspondence to: rodrigo.reyes@mcgill.ca

Short title

Dynamics at the eukaryotic replisome

One sentence summary

Stable binding of DNA polymerases at the active eukaryotic replisome explains processive DNA synthesis on both strands

Abstract

DNA replication is carried out by a multi-protein machine called the replisome. In *Saccharomyces cerevisiae*, the replisome is composed of over 30 different proteins arranged into multiple subassemblies, each performing distinct activities. Synchrony of these activities is required for efficient replication and preservation of genomic integrity, but how this is achieved remains an open question. Here we established novel single-molecule fluorescence microscopy protocols to

study the binding kinetics of individual replisome subunits in live *S. cerevisiae*. Our results show long residence times for most subunits at the active replisome, supporting a model where all subassemblies bind tightly and work in a coordinated manner for extended periods, hence defining the architecture of the active eukaryotic replisome.

A crucial step during cell proliferation is the duplication of the genome, composed of long DNA molecules. DNA replication is carried out by a specialized multiprotein machine called the replisome. In all organisms, the replisome accomplishes a few basic functions: progressive unwinding of the double-stranded DNA, the synthesis of an RNA molecule that serves as primer, and the polymerization of DNA. In eukaryotic cells, unwinding is catalysed by the CDC45/MCM2-7/GINS (CMG) helicase. Pol α carries out the RNA primer synthesis at the lagging strand – whose synthesis proceeds in an opposite direction to DNA unwinding – and extends primers by its DNA polymerase activity. Pol ε and Pol δ serve as the replicative DNA polymerases, synthesizing most of the DNA on the leading and lagging strands, respectively, and while Pol ε is thought to interact stably with CMG, Pol δ does not appear to have a stable interaction with the replisome [1,2]. In addition to these basic functions, the replisome contains multiple other proteins serving structural or other specialized roles, such as Ctf4, which is thought to act as a recruiting hub for other proteins, including Pol α (Fig. 2A) (reviewed in [3]).

It is thought that a stable architecture of the replisome helps ensure that the basic activities occur in an orderly and synchronized manner, thus limiting the accumulation of single-strand DNA (ssDNA) near the sites of replication, which can be a source of genome instability [4,5]. However, plasticity of the architecture over time – through dissociation of subunits – has been observed in bacteria [6-8], leading to questions on how this balance of synchronized work and dynamic behaviour of individual subunits is achieved. *In vitro* studies of the binding kinetics of individual subunits of the replisome have been limited in the past as they could not take into account the great number of components and the presence of post-translational modifications, both of which can influence the strength of interaction between proteins and replisome function. Here we characterized the

binding kinetics of eukaryotic replisome subunits in live cells, using single-molecule fluorescence microscopy, to gain insight into replisome function and architecture.

We first established a set of experimental and analysis protocols for the use of single-molecule fluorescence microscopy in budding yeast. We used a strain carrying a fusion of the histone H3 to HaloTag (H3-Halo) [9,10] as a control, due to the expected long-residence binding of H3 on chromatin [11]. Using H3-Halo we characterized the rate of bleaching of the fluorophore, the localization linking distance for tracking chromatin-bound proteins in our timelapses, and the traits of a single-fluorescent molecule in our system, which we exploited for robust classification using a machine-learning (ML) approach (Fig. 1B-E). For detection, we used a HaloTag ligand coupled to the cell permeable photoactivable dye PA-JF549 [12]. The imaging protocol used to track single-molecules is similar to that previously described for *Escherichia coli* [6], where a single exposure to a low dose of 405nm activating light was followed by multiple events of 561nm-excitation at fixed intervals. However, here we did multiple rounds of activation-imaging to maximize the sample size (Fig. 1A). Camera integration times of 500 milliseconds resulted in the motion-blurring of diffusive molecules, facilitating the tracking of chromatin-bound molecules represented as foci. All our imaging was done using HILO illumination to increase the signal-to-noise ratio [13]. In addition, to minimize light refraction caused by the cell wall [14], we did our experiments in presence of 30% of the refractive media-matching Optiprep [15](Fig. S4). Unbiased identification of chromatin-bound single-molecules was done using a random forest-based machine learning algorithm, which used multiple traits of molecular movement and intensity for classification (Fig. 1C-D) (Material and Methods).

We applied these newly developed strategies to evaluate the residence times of multiple replisome subunits. We expected long residence times for the CMG helicase and the leading-strand

Pol ε. In contrast, we expected much shorter binding for Pol δ, consistent with the use of a different copy of the polymerase for the synthesis of each Okazaki fragment. We constructed and characterized haploid strains carrying HaloTag fusions of replisome subunits as single gene copies at their original chromosomal location (Fig. S5-S7). Analysis was done in cells undergoing S-phase, as judged by the density and heterogeneity in the labelling of the nucleus by a Pol30-mNeonGreen fusion (Fig. 1F). Using 1-second intervals in a single-imaging plane, we observed lifetimes of fluorescent foci that were indistinguishable from H3-Halo (our bleaching control) in cells carrying Cdc45-Halo and Mcm4-Halo (Fig. 2B-C)(Table S5-S7), consistent with a reported tight binding of CMG at chromatin [16]. Similarly, Pol ε (Pol2-Halo and Dpb4-Halo) and the Ctf4 subunit – both of which have a direct interaction with the CMG helicase [1,17] – exhibited lifetimes similar to those of the CMG complex, suggesting they interact tightly with it. Unexpectedly, Pol3-Halo and Pol32-Halo, two subunits of Pol δ, also showed long track durations indicative of stable binding (Fig. 2C). This was surprising since there is no reported connection between Pol δ and the CMG helicase.

We corroborated and extended the results above by doing similar experiments using longer time intervals between acquisitions. To reduce the likelihood of chromatin-bound molecules moving out of the focal plane, we imaged three and four different planes separated by 500nm, for 8s and 20s intervals, respectively, and performed maximum intensity projections (MIP) on the resulting z-stacks. These experiments suggest that the CMG helicase, Ctf4, Pol δ, and Pol ε, remain bound to chromatin for an average time that exceeds 5 minutes (Fig. 2D)(Materials and methods). Note that the similarity to the bleaching control results in a high uncertainty in our estimates. Furthermore, given that our measurements originate from replisomes at varying stages between initiation and termination, the real residence times may be longer than our estimates. To put these

numbers into perspective, the average segment replicated by an individual replisome, defined as half the length of a replicon, is 18.5kb [18]. At an average replication rate of 1.6 kbp/min [19], the time for synthesizing this segment is ~11 minutes. Our estimates approach the time required for completion of synthesis, making it likely that some copies of the DNA polymerases remain associated to the replisome from initiation to termination.

Pol α was the only subcomplex that exhibited a dynamic behaviour in our experiments. Using 1-second intervals between consecutive pictures, both Pol12-Halo and Pri2-Halo had track durations shorter than the bleach time (Fig. 2E). This observation was corroborated by results using 8-second intervals for Pol12-Halo. A weighted average of the two subunits and intervals results in an estimated residence time of 63.5 seconds**,** indicating that Pol α can perform multiple cycles of priming before unbinding (Fig. 2F). To test if the binding kinetics of Pol α reflect a dynamic binding to the stably-bound Ctf4, we repeated the experiments of Pol12-Halo in a mutant strain of Pol1 lacking the Ctf4-interacting peptide (CIP), termed Pol12-Halo CIP- (Fig. 2E). Surprisingly, we observed no difference in the CIP- strain when compared to the wt strain, suggesting that the interaction with Ctf4 is not essential for retaining Pol α at the replisome. This is consistent with *in vitro* results that suggest that Ctf4 does not retain Pol α at the replisome [20].

Four different models can explain the observed stable-binding of Pol δ: first, stable binding at the end of Okazaki fragments; second, usage at both leading and lagging strands; third, efficient recycling and quick rebinding in replisome-dense nuclear regions; fourth, tight binding to the replication fork. Slow recycling after completion of the Okazaki fragments, possibly through binding to PCNA, is incompatible with the estimated copy numbers of the Pol δ subunits. Considering that at the peak of S-phase there are ~300 replisomes per nucleus [21], and that the estimated copy number of Pol31 is ~2200 (reportedly the least abundant subunit)[22], cells would

only be able to undergo up to 9 cycles at the lagging strand, spanning about 1 minute (165bp Okazaki fragments x 9 / 1.6kbp min^{-1} rate), before depleting the available Pol δ. The second model, where Pol δ works at both strands, would potentially result in two different binding regimes: long residence and short residence times at leading and lagging strands, respectively. However, we observe no evidence for the presence of two regimes in our data (Fig. 3A and Table S3). The third model considers that Pol δ efficiently rebinds chromatin after completion of an Okazaki fragment. This idea is especially compelling due to the high density of replisomes at mid S phase. To test this model, we artificially reduced the replisome density by using a strain carrying a deletion of clb5, which should result in the inactivation of about half of the origins of replication [23](Fig. S3). However, we did not observe a significant difference in the lifetimes of Pol32-Halo when comparing wt and clb5Δ cells, suggesting that fast rebinding cannot explain the long residence times observed (Fig. 3B). We also performed computer simulations of Pol δ binding in the context of high density of replisomes at mid S-phase and found no evidence of frequent rebinding under our experimental and biologically sensible parameters (Fig. 3C-E and Table S4)(Materials and methods). Furthermore, these simulations indicated that a model of transient Pol δ, where each copy dissociates every 4 seconds after completion of an Okazaki fragment, would lead to many replisomes being left without a Pol δ for a significant fraction of time (Fig. 3F-G). Hence, we propose that a single copy of Pol δ synthesizes multiple Okazaki fragments. This model would explain how the eukaryotic replisome achieves near identical rates of synthesis between the leading and lagging strand (discussed in [24]).

The structural, biophysical, and biochemical characterization of replisome architecture in bacteria shows a link between their replicative DNA polymerases and their helicase, and that there is some level of synchrony between helicase unwinding and polymerase activity [25,26]. This is

presumably important to limit the generation of ssDNA during elongation of the strands. Here we provide evidence that supports a similar architecture for the eukaryotic replisome. We hypothesize that a yet uncharacterized protein-protein interaction bridges Pol δ and the CMG helicase (Fig. 4). It is possible that interaction with Pol α may help to retain Pol δ at the replisome [27]; however, Pol α does have a shorter residency time than Pol δ, indicating that there are likely other important interactions retaining Pol δ at the replisome. In humans an interaction between Pol δ and AND1, the orthologue of Ctf4, has been described [28], which if present in budding yeast may contribute to the longer residence times. Future work will define how Pol δ is kept at the replication fork.

Long residence times of both the leading and lagging strand DNA polymerases in yeast contrast with the fast dynamics reported for the bacterial replisome [6-8], where DNA Pol III is exchanged every few seconds. However, the rate of replication, which is about 30 times slower in budding yeast compared to *E. coli*, would reduce the rate of accumulation of torsional stress near the fork – a possible factor triggering unbinding of the polymerases in bacteria [29] – and allow time for its dissipation by other means. Alternatively, strategies unique to eukaryotes may result in a similar transient binding of the DNA polymerases to DNA that do not require complete unbinding from the helicase. For example, despite the stable binding of Pol ε to CMG, its flexible linker potentially allows it to unbind from DNA while remaining bound to the rest of the replisome [30].

S. cerevisiae continues to be a powerful model organism for the understanding of basic eukaryotic cellular mechanisms due to its genetic tractability. However, the development of single-molecule imaging techniques, which have advanced the understanding and are widely applied in bacterial systems, have had limited use in this system due to higher background fluorescence and light scattering. We expect that the single-molecule protocols described here will facilitate its future use to study a wide range of questions in this organism.

References

1 Sengupta, S., van Deursen, F., de Piccoli, G. & Labib, K. Dpb2 integrates the leading-strand DNA polymerase into the eukaryotic replisome. *Curr Biol* **23**, 543-552, doi:10.1016/j.cub.2013.02.011 (2013).

2 Schauer, G. D. & O'Donnell, M. E. Quality control mechanisms exclude incorrect polymerases from the eukaryotic replication fork. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 675-680, doi:10.1073/pnas.1619748114 (2017).

3 Bell, S. P. & Labib, K. Chromosome Duplication in Saccharomyces cerevisiae. *Genetics* **203**, 1027-1067, doi:10.1534/genetics.115.186452 (2016).

4 Lopes, M., Foiani, M. & Sogo, J. M. Multiple mechanisms control chromosome integrity after replication fork uncoupling and restart at irreparable UV lesions. *Mol Cell* **21**, 15-27, doi:10.1016/j.molcel.2005.11.015 (2006).

5 Yang, Y., Sterling, J., Storici, F., Resnick, M. A. & Gordenin, D. A. Hypermutability of damaged single-strand DNA formed at double-strand breaks and uncapped telomeres in yeast Saccharomyces cerevisiae. *PLoS Genet* **4**, e1000264, doi:10.1371/journal.pgen.1000264 (2008).

6 Beattie, T. R. *et al.* Frequent exchange of the DNA polymerase during bacterial chromosome replication. *Elife* **6**, doi:10.7554/eLife.21763 (2017).

7 Lewis, J. S. *et al.* Single-molecule visualization of fast polymerase turnover in the bacterial replisome. *Elife* **6**, doi:10.7554/eLife.23932 (2017).

8 Liao, Y., Li, Y., Schroeder, J. W., Simmons, L. A. & Biteen, J. S. Single-Molecule DNA Polymerase Dynamics at a Bacterial Replisome in Live Cells. *Biophysical journal* **111**, 2562-2569, doi:10.1016/j.bpj.2016.11.006 (2016).

9 Los, G. V. *et al.* HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol* **3**, 373-382, doi:10.1021/cb800025k (2008).

10 Ball, D. A. *et al.* Single molecule tracking of Ace1p in Saccharomyces cerevisiae defines a characteristic residence time for non-specific interactions of transcription factors with chromatin. *Nucleic acids research* **44**, e160, doi:10.1093/nar/gkw744 (2016).

11 Kimura, H. & Cook, P. R. Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. *The Journal of cell biology* **153**, 1341-1353, doi:10.1083/jcb.153.7.1341 (2001).

12 Grimm, J. B. *et al.* Bright photoactivatable fluorophores for single-molecule imaging. *Nat Methods* **13**, 985-988, doi:10.1038/nmeth.4034 (2016).

13 Tokunaga, M., Imamoto, N. & Sakata-Sogawa, K. Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat Methods* **5**, 159-161, doi:10.1038/nmeth1171 (2008).

14 Smith, C. *et al.* In vivo single-particle imaging of nuclear mRNA export in budding yeast demonstrates an essential role for Mex67p. *The Journal of cell biology* **211**, 1121-1130, doi:10.1083/jcb.201503135 (2015).

15 Boothe, T. *et al.* A tunable refractive index matching medium for live imaging cells, tissues and model organisms. *Elife* **6**, doi:10.7554/eLife.27240 (2017).

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

17 Simon, A. C. *et al.* A Ctf4 trimer couples the CMG helicase to DNA polymerase alpha in the eukaryotic replisome. *Nature* **510**, 293-297, doi:10.1038/nature13234 (2014).

18 Liachko, I. *et al.* A comprehensive genome-wide map of autonomously replicating sequences in a naive genome. *PLoS Genet* **6**, e1000946, doi:10.1371/journal.pgen.1000946 (2010).

19 Sekedat, M. D. *et al.* GINS motion reveals replication fork progression is remarkably uniform throughout the yeast genome. *Mol Syst Biol* **6**, 353, doi:10.1038/msb.2010.8 (2010).

20 Kurat, C. F., Yeeles, J. T. P., Patel, H., Early, A. & Diffley, J. F. X. Chromatin Controls DNA Replication Origin Selection, Lagging-Strand Synthesis, and Replication Fork Rates. *Mol Cell* **65**, 117-130, doi:10.1016/j.molcel.2016.11.016 (2017).

21 Saner, N. *et al.* Stochastic association of neighboring replicons creates replication factories in budding yeast. *The Journal of cell biology* **202**, 1001-1012, doi:10.1083/jcb.201306143 (2013).

22 Ho, B., Baryshnikova, A. & Brown, G. W. Unification of Protein Abundance Datasets Yields a Quantitative Saccharomyces cerevisiae Proteome. *Cell Syst* **6**, 192-205 e193, doi:10.1016/j.cels.2017.12.004 (2018).

23 McCune, H. J. *et al.* The temporal program of chromosome replication: genomewide replication in clb5{Delta} Saccharomyces cerevisiae. *Genetics* **180**, 1833-1847, doi:10.1534/genetics.108.094359 (2008).

24 Kapadia, N. & Reyes-Lamothe, R. A quest for coordination among activities at the replisome. *Biochemical Society transactions*, doi:10.1042/BST20180402 (2019).

25 Kim, S., Dallmann, H. G., McHenry, C. S. & Marians, K. J. Coupling of a replicative polymerase and helicase: a tau-DnaB interaction mediates rapid replication fork movement. *Cell* **84**, 643-650 (1996).

26 Sanders, G. M., Dallmann, H. G. & McHenry, C. S. Reconstitution of the B. subtilis replisome with 13 proteins including two distinct replicases. *Mol Cell* **37**, 273-281, doi:S1097-2765(09)00956-3 [pii] 10.1016/j.molcel.2009.12.025 (2010).

27 Huang, M. E., Le Douarin, B., Henry, C. & Galibert, F. The Saccharomyces cerevisiae protein YJR043C (Pol32) interacts with the catalytic subunit of DNA polymerase alpha and is required for cell cycle progression in G2/M. *Mol Gen Genet* **260**, 541-550 (1999).

28 Bermudez, V. P., Farina, A., Tappin, I. & Hurwitz, J. Influence of the human cohesion establishment factor Ctf4/AND-1 on DNA replication. *The Journal of biological chemistry* **285**, 9493-9505, doi:10.1074/jbc.M109.093609 (2010).

29 Kurth, I., Georgescu, R. E. & O'Donnell, M. E. A solution to release twisted DNA during chromosome replication by coupled DNA polymerases. *Nature* **496**, 119-122, doi:10.1038/nature11988 (2013).

30 Zhou, J. C. *et al.* CMG-Pol epsilon dynamics suggests a mechanism for the establishment of leading-strand synthesis in the eukaryotic replisome. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 4141-4146, doi:10.1073/pnas.1700530114 (2017).

31 Simon, A. C. *et al.* A Ctf4 trimer couples the CMG helicase to DNA polymerase alpha in the eukaryotic replisome. *Nature* **510**, 293-+, doi:10.1038/nature13234 (2014).

32 Tinevez, J. Y. *et al.* TrackMate: An open and extensible platform for single-particle tracking. *Methods* **115**, 80-90, doi:10.1016/j.ymeth.2016.09.016 (2017).

33 Kitamura, E., Blow, J. J. & Tanaka, T. U. Live-cell imaging reveals replication of individual replicons in eukaryotic replication factories. *Cell* **125**, 1297-1308, doi:10.1016/j.cell.2006.04.041 (2006).

34 Shihavuddin, A. *et al.* Smooth 2D manifold extraction from 3D image stack. *Nat Commun* **8**, 15554, doi:10.1038/ncomms15554 (2017).

35 Breiman, L. Random Forests. *Machine Learning* **45**, 5-32, doi:10.1023/a:1010933404324 (2001).

36 Hajjoul, H. *et al.* High-throughput chromatin motion tracking in living yeast reveals the flexibility of the fiber throughout the genome. *Genome Res* **23**, 1829-1838, doi:10.1101/gr.157008.113 (2013).

37 Smith, D. J. & Whitehouse, I. Intrinsic coupling of lagging-strand synthesis to chromatin assembly. *Nature* **483**, 434-U480, doi:10.1038/nature10895 (2012).

38 Stodola, J. L. & Burgers, P. M. Resolving individual steps of Okazaki-fragment maturation at a millisecond timescale. *Nat Struct Mol Biol* **23**, 402-408, doi:10.1038/nsmb.3207 (2016).

Acknowledgments:

We thank Tatiana Karpova (NCI) for providing the yeast strains carrying H3-Halo and pdr5Δ;Luke Lavis (Janelia Farms) for providing the Halo substrate coupled to PA-JF549 dye; and Karim Labib for the strain carrying the CIP-binding mutant of Pol1. We thank Thomas Gligoris and Stephan Uphoff for initial help working with HaloTag fusions. **Funding:** The experiments were done using equipment from the Integrated Quantitative Bioscience Initiative (IQBI), funded by CFI 9. This work was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC# 435521-2013), the Canadian Institutes for Health Research (CIHR MOP# 142473), the Canada Foundation for Innovation (CFI# 228994), and the Canada Research Chairs program. **Author contributions:** NK, Conceptualization, Formal analysis, Investigation, Writing; ZWEH, Formal analysis, Investigation, Writing; AY, TRB, Investigation; RR-L, Conceptualization, Supervision, Funding acquisition, Writing. **Competing interests:** Authors declare no competing interests. **Data and materials availability:** All data, code, and materials used in this work are available by request.

Supplementary Materials

Supplementary Text

Figs. S1 to S7

Tables S1 to S7

Figure 1_

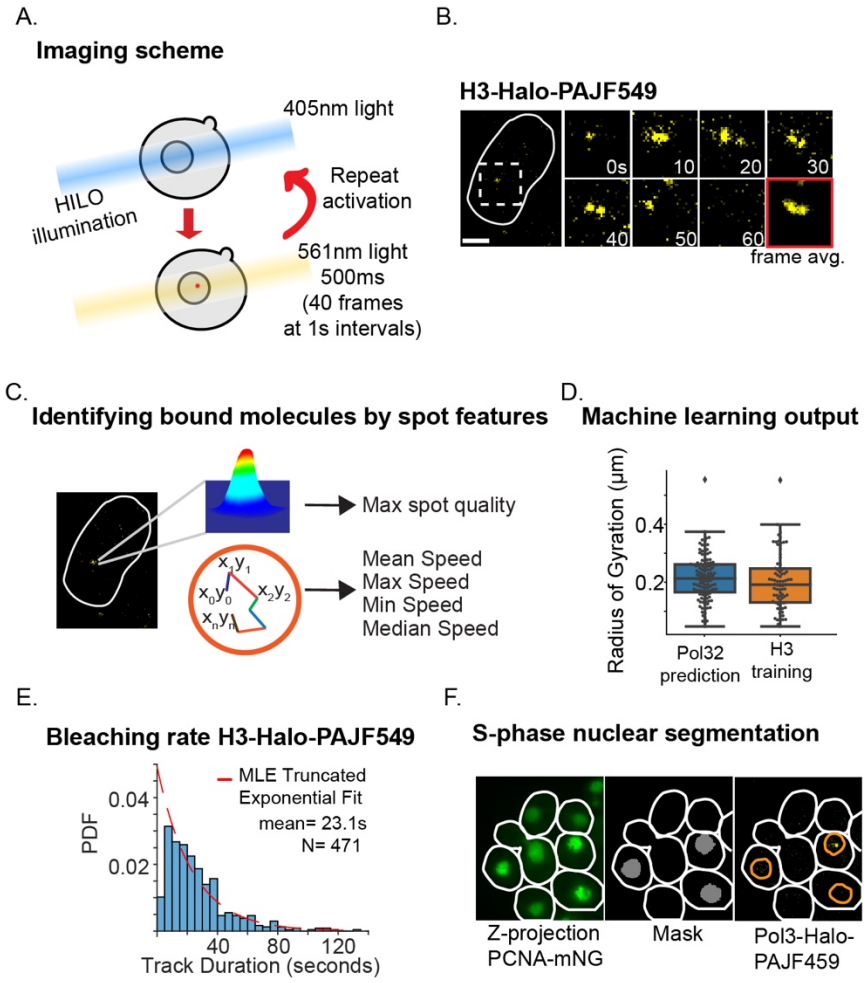


Figure 1. Single-molecule tracking in the yeast nucleus. (**A**) To track single-molecules, we exposed Halo-tagged yeast cells to a single exposure of low-intensity 405nm light, resulting in the stochastic activation of on average one molecule per nucleus. This was followed by time-lapse imaging using Highly Inclined and Laminated Optical sheet (HILO) illumination and 561nm light for fluorescence detection. Multiple rounds of activation and imaging were performed to increase sample size. (**B**) Representative images of single-molecules of H3-Halo-PAJF549 in live cells. The image on the left shows the entire cell. Images on the right show a magnified view of the square drawn on the left image, and show individual timepoints or a frame average of the

experiment over 60 seconds. Scale bar 2 μm. (**C**) Information on the intensity and speed of single molecules of H3-Halo-PAJF549 was extracted and used to identify chromatin-bound molecules of replisome subunits. (**D**) Example of the classification outcome of the machine-learning algorithm used to identify chromatin-bound molecules. The radius of gyration of individual molecules is shown for the training data of H3, and for molecules of Pol32 and H3 classified as bound. (**E**) Characteristic bleaching rate for the fluorophore PAJF549 using H3-Halo as control, whose molecules should bleach before unbinding. (**F**) Representative image of the procedure followed to segment nuclei of cells undergoing S-phase. Smooth Manifold Extraction (SME) was used to project the intensity of a Z-stack of PCNA-mNeonGreen covering the whole cell while preserving spot-like features. S-phase nuclei were selected based on high heterogeneity in the distribution of intensities.

Figure 2

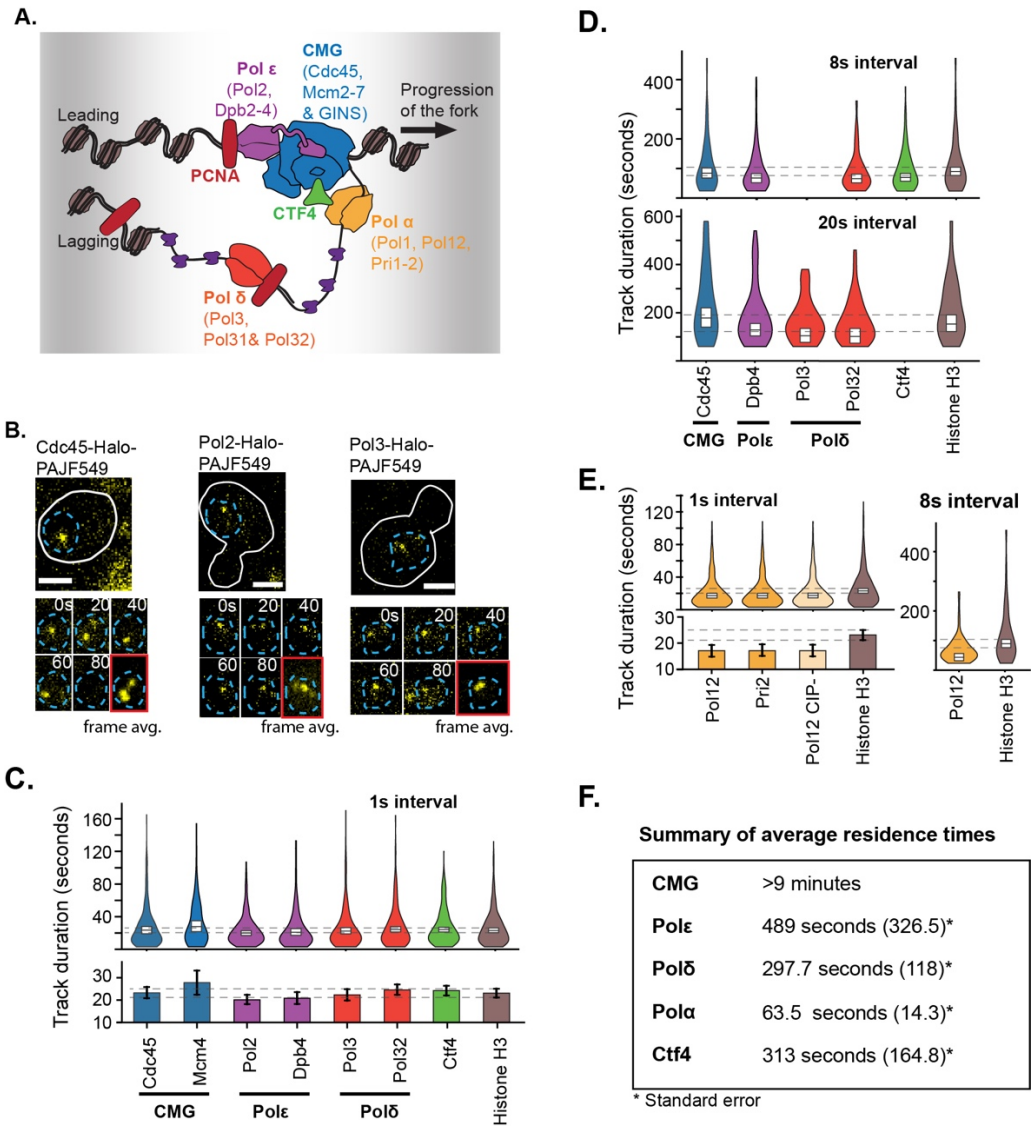
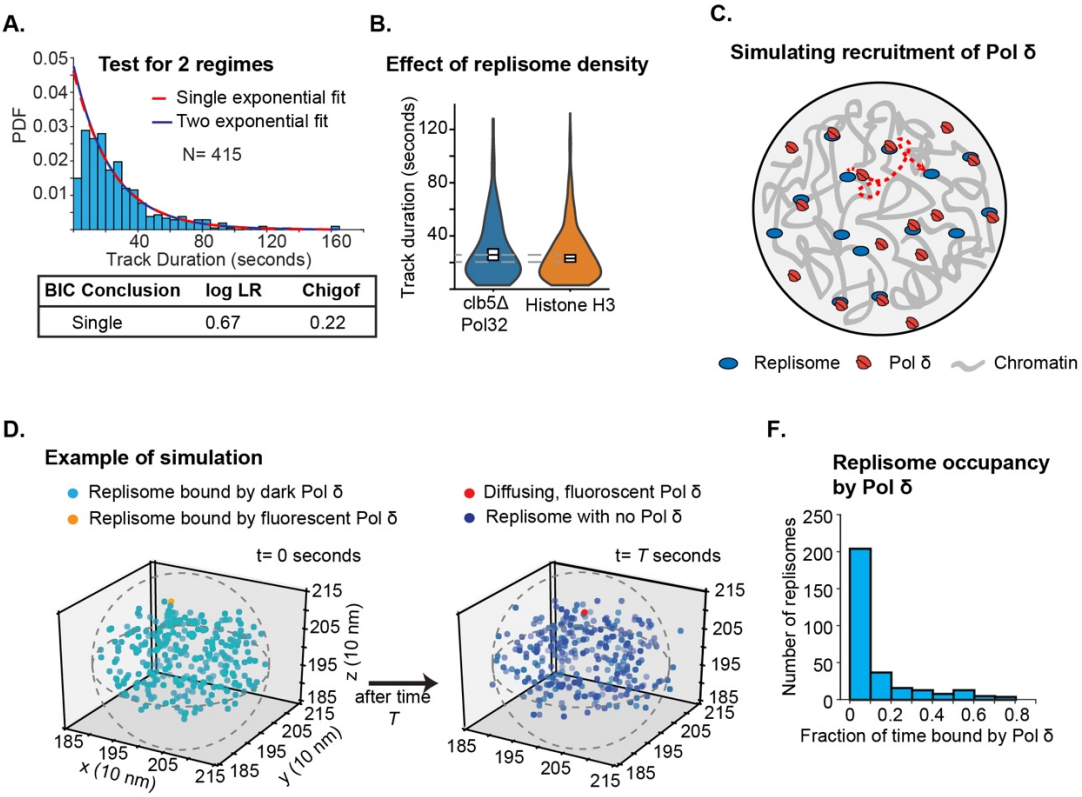


Figure 2. Binding kinetics of replisome subunits. (**A**) Diagram of the architecture of the replisome in *S. cerevisiae*. (**B**) Images for long-lived single-molecule fluorescent foci in strains carrying replisome fusions to Halo-PAJF549. Images on the top show the entire cells. Images on the bottom show a magnified view of the nuclei and show individual timepoints or a frame average of the experiment over 80 seconds. Scale bar 2 µm. (**C**) Distribution of track durations (spot lifetimes) for different replisome subunits imaged at intervals of 1 second (top). White boxes represent the

boundaries of confidence intervals of the mean (represented as the line in the box). Dashed lines, representing the confidence intervals for Histone H3, are drawn for comparison. Colours correspond to the diagram shown in A. Histogram showing only the average and confidence intervals to facilitate comparison (bottom). (**D**) Distribution of track durations for the same subunits as in C using intervals between pictures of 8 or 20 seconds. (**E**) Distribution of track durations for subunits for Pol α and comparison with histone H3. Plot on the top left shows the distribution obtained using 1-second intervals. Histogram is shown for these same distributions so that the difference with H3 can be appreciated (bottom left). The distribution of track durations using 8-second intervals is shown on the right. (**F**) Summary of the residence times obtained. With exception for CMG and Ctf4, the weighted averages were used from different time intervals and different subunits in the case of Pol δ and Pol α. For all subcomplexes except Pol α, the track durations were very close to the bleaching duration, resulting in large errors. For CMG, a lower-bound estimate was calculated from data obtained with the 8-second intervals, since the track duration was nearly identical to the bleaching duration.

Figure 3_



BIC Conclusion	log LR	Chigof
Single	0.67	0.22

E.

Expected spot lifetimes with a 4s residence time of Pol δ

$D_{Pol\delta}$ (μm²/s)	$D_{Replisome}$ (μm²/s)	T_{Track} [CI]
0.5	0.005	2.89 [2.00, 3.79]
0.5	0.05	3.17 [2.19, 4.16]
5	0.005	4.15 [2.86, 5.44]
5	0.05	2.94 [2.03, 3.85]

G.

Mean replisome occupancy by Pol δ

$D_{Pol\delta}$ (μm²/s)	$D_{Replisome}$ (μm²/s)	Mean Occupancy
0.5	0.005	0.1207
0.5	0.05	0.1275
5	0.005	0.1219
5	0.05	0.1269

Figure 3. Testing for recycling of Pol δ. (**A**) Statistical testing of track durations obtained from Pol32 with 500ms interval, to check if a more complex model fits better than a single-exponential, indicating that recycling may be frequent or if there are multiple binding behaviours at the replisome. The tests used were the loglikelihood ratio (LLR) test and the Bayesian information criterion (BIC) test. (**B**) Comparison of track durations obtained of Pol32 in a clb5Δ background with a 1s interval, and from the histone H3 data. (**C**) Diagram for recycling of Pol δ among multiple replisomes. (**D**) Diagram of general simulation setup zoomed into the sphere where all the

replisomes were initially placed. We used a 4-second residence time for Pol δ to account for a model where a different copy is needed for synthesis of each Okazaki fragment (Materials and methods). Initially, 300 replisomes were placed inside a sphere that was 300nm in diameter to represent a diffraction-limited region, represented by the dotted line. This unnaturally high density was used to maximize the probability of recycling. The replisomes are subsequently allowed to move outside the initial sphere and throughout a nucleus of diameter 1.5μm. Unbound, non-fluorescent copies of Pol δ are not shown for clarity. **E)** Track durations obtained from simulations, over a range of diffusion coefficients. The initial fraction of replisomes bound by Pol δ was 0.75. The estimated track duration for a single unbinding event was estimated to be 3.41 seconds, based on a 4s residence time and a 23s bleach time. **F)** Representative histogram of the fraction of time replisomes had a bound Pol δ, over the course of 200s. **G)** Results for mean occupancy of replisomes (i.e. the average fraction of time replisomes had a bound Pol δ), over a range of diffusion coefficients.

Figure 4_

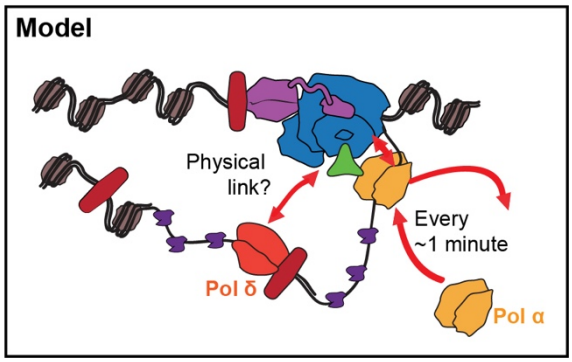


Figure 4. Model for the architecture of the eukaryotic replisome. The CMG helicase, Pol ε and CTF4 bind tightly to chromatin and act processively. In addition, our data suggest the existence of a direct or indirect interaction between Pol δ and the stably bound CMG helicase. Pol δ acts processively, with a single copy of this polymerase potentially synthesizing all the Okazaki fragments of an individual replisome from initiation to termination. Pol α interacts with the CMG independently of Ctf4 and turnovers at rates that permit it to synthesize multiple primers at every binding event. The function of the interaction between Pol α and Ctf4 at the replisome is unclear.

Materials and Methods

Strains and constructions

Strains used in this study are all from a BY4741 background (with the exception of the CIP-mutant, see below) and are shown in Table S1. Plasmids used in this study are derivatives of pUC18 (ColE1 origin, Ampicillin resistance); pTB16 carries mNeonGreen and a downstream NatMX marker, while pSJW01 carries the HaloTag gene with a HygB marker. Both mNeonGreen and HaloTag genes encode an 8 amino acid linker at the 5' end (sequence: GGTGACGGTGCTGGTTTAATTAAC). Plasmids were maintained in *E. coli* DH5α and were extracted by growing in LB with 100 µg/ml ampicillin then using the Presto Mini Plasmid Kit (Geneaid).

All PCRs were made using either Phusion or Q5 (NEB). PCR reaction were in a volume of 50 µl and included water, 3% DMSO, the reaction buffer, 2.5 mM of each dNTP, 0.2 µM of each primer, either 1ng of plasmid DNA (for insertions) or 1 µl of genomic DNA (for screening insertions), and 0.5 µl of polymerase.

Fluorescent fusions were made by PCR amplification from pTB16 or pSJW01 using the primers listed in Table S2. PCR products were transformed into wild-type diploid BY4743. A single colony was grown at 30°C in 5 ml yeast peptone dextrose (YPD) overnight, then diluted to a final OD_{600} of 0.1 in 10ml of YPD. Cells were taken at OD_{600}=0.5-0.6 and centrifuged at 4000 rpm for 5 min. The pellet was washed twice with 25 ml of sterile deionized water, then once with 1ml of 100mM lithium acetate. Cells were then resuspended in 240 µl of 50% PEG, then 50 µl of salmon sperm DNA (thawed at 95°C for 5min, then incubated on ice for at least 10min), 50 µl of the PCR product and 36 µl of 1M lithium acetate were added in this order. The mixture was thoroughly mixed by pipeting and incubated on a rotator at 25°C for 45min, followed by heat

shock at 42°C for 30min. Cells were pelleted in a microcentrifuge, washed in 500 µl of sterile water, then resuspended in 200 µl of YPD and plated on YPD agar. After growing at 30°C overnight, the cell lawn was replica-plated onto selective YPD agar, either with 100 µg/ml cloNAT (Werner) for mNeonGreen, or with 200 µg/ml Hygromycin B (Life Technologies) for HaloTag. Transformants were screened for the presence of an insert by PCR using the indicated primers in Table S2.

Confirmed clones were then sporulated by taking 750 µl of a YPD overnight cultures, washing 4 times with 1 ml sterile deionized water, then washing once with 1 ml of potassium acetate sporulation medium (Kac), and finally resuspending in 2 ml of KAc and incubating at 25°C wish shaking. After 5 days the sporulating cultures were checked by microscopy for the appearance of numerous tetrads, then 750 µl was taken and washed 3 times in sterile water before final resuspension in 1ml water and storage at 4°C. For dissection, 45 µl of spores were treated with 5 µl of zymolase for 10 min, then tetrads were dissected on YPD plates to isolate haploids with the tagged fusion. Genomic DNA was isolated from the haploid by vortexing the cells in the presence of zirconia/silica beads, followed by phenol extraction and ethanol precipitation. The insertion site was once again amplified using the same screening primers as above, and the PCR product was sequenced to confirm that the tag and linker were both mutation-free.

The HaloTag haploids were combined with PCNA-mNeonGreen (from YTB31) and the pdr5Δ::KanMX deletion (from a haploid sporulated from YTK1414) by mating. To do this, 10 µl of water was spotted on a YPD plate, and colonies from the Mat a and Mat α haploids to be combined were mixed together into the water drop and incubated at 30°C overnight. Cells were then restreaked on an auxotrophic -Met-Lys plate on which only the mated diploid could grow.

Diploids resulting from a mating were dissected as above, and eventually haploids with all three markers were isolated and used for imaging. These haploids are all listed in Table S1.

The CIP- strain YCE449 is from a W303 background and carries a Pol1-4A allele, where four amino acids are substituted with alanine at D141, D142, L144 and F147 [31]. Pol12-Halo and CIP- were combined by mating YJL10 with YCE449, then sporulating and dissecting the resulting diploid to get YAJ05.

Western blot

To prepare crude cell lysates, YPD cultures were grown to exponential phase (OD600=0.5) and fixed in 10% TCA. The pellet was washed in cold acetone then in Beating Buffer (8 M urea, 50 mM ammonium bicarbonate, 5 mM EDTA), then resuspended in Beating Buffer with glass beads and vortexed at 4°C for 5min. The bottom of the tube was pierced with a needle, placed inside another tube and the cell debris and buffer was centrifuged into the clean tube. This was centrifuged again and the supernatant was collected. Protein concentration was determined using a Bradford assay.

SDS-PAGE was performed using 4-20% Mini-Protean TGX precast gels (Biorad). Lysates were prepared in Laemmli sample buffer with β-mercaptoethanol. Gels were run in Tris-Glycine-SDS at 100V for 2-3h. Proteins were transferred onto a nitrocellulose membrane using Biorad's Trans-Blot Turbo system and transfer packs running the MixedMW preset program. The membrane was incubated in Blocking Buffer (5% milk and 3% BSA in TBS-Tween) for 1h. The membrane was probed with α-Halotag mouse monoclonal antibody (Promega, diluted 1:1000 in Blocking Buffer), washed 3 times in TBS-Tween, then probed with goat α-mouse HRP-conjugated secondary antibody (Promega, diluted 1:10000 in Blocking Buffer) and finally washed 3 times in

TBS-Tween. The membrane was treated with Clarity Western ECL substrate (Biorad) and exposed to autoradiography film (Diamed).

Flow cytometry

Cells were prepared for flow cytometry by growing cultures in YPD to exponential phase (OD600=0.5) and fixing them in 70% ethanol at 4°C. Cells were washed in Tris-Cl and incubated with RNAse A at 42°C for 3h, then with Proteinase K at 50°C for 30min. Cells were centrifuged and the pellet was resuspended in Tris-Cl and stained with propidium iodide (8 µg/ml). Samples were run on a FACSCalibur (Becton Dickinson) using the following filters and detector settings: FSC E01; SSC 396V and 4.61 gain; FL2 730V and 4.10 gain. Cytometer was calibrated using parental haploid (BY4741) and diploid (BY4743) asynchronous exponential cultures.

Microscopy

A single colony from a YPD plate was placed in 5mL synthetic complete (SC) medium and grown with shaking at 30 degrees Celsius for ~5-6 hours. This culture was diluted by transferring ~50uL into 5mL of fresh S.C and grown overnight at 30 degrees. The overnight culture was diluted to 0.15 the next day and grown until the optical density (OD) reached 0.30. 1mL of this culture was spun down for 1 min @ 13000RPM, and the pellet was resuspended in 500uL of fresh S.C. Janelia Farms photoactivatable 549 (JF-PA549) was added to the 500uL culture for a final dye concentration of 50nM, except for YTK1434-Halo (Histone H3), where a concentration of 10nM was used to compensate for the higher copy number. This culture was placed in a thermomixer at 30 degrees and 500RPM for 40 minutes. After incubation, 3 wash cycles using fresh S.C were done to wash away unbound dye. After the final wash step, the pellet was resuspended in 50uL of

S.C, and 3uL of the culture was placed on an agarose pad consisting of SC and Optiprep (Sigma), within a Gene Frame (Thermo Scientific). The pad was made by taking a 2% agarose Optiprep mixture (0.02g in 1mL Optiprep) - that was heated to 90 degrees - and mixing 500uL with 500uL 2xSC, resulting in a 1% agarose 30% Optiprep SC mixture. Approximately 140uL of this mixture was placed within the Gene Frame, with excess being removed with a KimWipe. Prior to imaging, we waited ~15 minutes to let any unbound dye be released.

Coverslips were cleaned with the following steps: 1) Place in 2% VersaClean detergent solution overnight. 2) Wash with MilliQ water 3x. 3) Sonicate in acetone for 30 minutes. 4) Wash with MilliQ water 3x. 5) Place in methanol and flame coverslips using Bunsen burner. 6) Place in Plasma Etch plasma oven for 10 minutes.

Microscopy was done at 23 degrees, on a Leica DMi8 inverted microscope with a Roper Scientific iLasV2 (capable of ring total internal reflection fluorescence (TIRF)), and an Andor iXON Ultra EMCCD camera. An Andor ILE combiner was used, and the maximum power from the optical fiber was 100mW for the 405nm wavelength, and 150mW for the 488nm and 561nm wavelengths. The iLasV2 was configured to do ring highly inclined and laminated optical sheet (HILO), for selective illumination and single-molecule sensitivity. Metamorph was used to control acquisition. A Leica HCX PL APO 100x/1.47 oil immersion objective was used, with 100nm pixel size. Any z-stacks were doing using a PInano piezo Z controller.

Single-particle photoactivated localization microscopy (sptPALM) experiments were performed by activating molecules with low power (0.5-2% in software) 405nm light to photoactivate ~1 molecule/cell, followed by stroboscopic, long-exposure (500ms) illumination with 561nm light (5% in software) to image primarily bound molecules. The time intervals used were 1s, 8s, and 20s, with cycles of activation every 40 time points for the 0.5s and 1s interval

data, and every 15 and 10 timepoints, for the 8s and 20s interval data, respectively. For the 8s and 20s time intervals, a mini 561nm z-stack of 1um and 1.5um was done, respectively, with 0.5um step sizes, in order retain molecules in focus. Maximum intensity projections (MIP) of these stacks were used for subsequent tracking analysis. A brightfield image and a z-stack of 6um (0.3um step size) in the 488nm channel, was taken before and after each timelapse, to ensure normal cell health and to find cells undergoing S-phase through the presence of POL30-mNG foci.

Tracking Analysis

Aside from tracking, all analysis was done using Matlab.

Tracking was done using Trackmate [32]. First, molecules were localized in each frame using a Laplacian of Gaussian (LoG) method, with an estimated diameter of 2.5 pixels. An intensity threshold was chosen that was slightly low, to still detect molecules that were moving out of the focal plane. Subsequent classification helped discard potential false-positive tracks from analysis that will be discussed later. After localization, tracks were formed using the Linear Assignment Problem algorithm by linking molecules in consecutive frames. The linking distance was chosen based on calculating the cumulative distribution function of the step sizes from the Histone H3 data, thus providing information on the step size of chromatin-bound proteins. Essentially, we varied the linking distance when analyzing the Histone H3 data until a noticeable plateau in the cumulative distribution function (CDF) was observed. We determined the linking distance as being the step size giving the 0.95 probability value in the CDF and did this for multiple time intervals: 3 pixels for 0.5s and 1s interval, 5 pixels, and 7 pixels for 8s and 20s time intervals,

respectively. A gap frame – to allow for missed localization – of 1 was used the gap-linking distance was set to 2 pixels more than the linking distance for that time interval (e.g. 5 pixel gap-linking distance when using a 3 pixel linking distance). Linking also had cost of 0.3 for the "Quality" parameter to ensure that correct molecules were linked. We also allowed for track merging and splitting with a 5 pixel distance for all time intervals, in some cases where multiple molecules were activated within the region and they were near one another. Tracks with less than four localizations were discarded as they were unreliable. Furthermore, since we are using track merging and splitting, we can exploit this property to discard tracks that are found in noisy regions or too many molecules active. In these cases, one would expect more spots localized in a track than predicted given a certain track duration (e.g. for a track duration of 20 frames, one would expect 21 spots, but may get 40). Therefore, we calculated a ratio of number of spots in track vs. number of spots estimated based on track duration and used a cut-off of 1.5 to remove tracks with too many spots localized.

To isolate tracks found in the nuclei of cells undergoing S-phase, we used the Pol30-mNG z-stacks, as PCNA is active during S-phase, thus resulting in fluorescent foci[33]. First, we performed a Smooth Manifold Extraction (SME) [34] on the z-stack, as an alternative to MIP, to distinguish more clearly Pol30-mNG foci. We then generated a binary mask of this image, resulting in all the nuclei regions have intensity values of 1, and zero elsewhere. To isolate S-phase nuclei, we used a threshold on the standard deviation of the intensities within the regions, as an indicator of heterogenous intensity caused by fluorescent foci. From the resulting binary mask, we isolated tracks found only in S-phase nuclei by calculating the mean positions of tracks and determining if those positions land in the binary mask.

For robust classification of tracks representing bound molecules, we employed a machine learning (ML) approach, using the output from Trackmate. First, we generated a training data set of ~750 tracks from the Histone H3 data, taken with 0.5s interval (Table S5, date: 20180511). Given that we used long-exposure to blur out diffusing molecules, it is quite easy to detect bound molecules from the raw timelapses. Therefore, we manually classified tracks by assigning a value of 1 to tracks representing genuine bound molecules vs 0 for tracks representing diffusing molecules or noise. Once we had the classifications, we performed the learning procedure using the algorithm, Random Forest ([35]), to develop two classification models, referred to as Model 1 and Model 2, with out-of-bag errors (equivalent to estimated cross-validation error) of 0.10 and 0.05, respectively. Model 1 had the predictor variables from Trackmate: mean speed, max speed, min speed, and median speed. Model 2 had all those variables as well, but in addition, included max quality. Both models were built with 6000 trees during the learning procedure, with following hyperparameters: InBagFraction (fraction of training data set given to each tree) = 0.50, MinLeafSize (minimum leaf size) = 50, NumPredictorstoSample (number of predictors to sample at each node) = 2. We then could output the predictor variables from tracks derived from timelapses of replisomal proteins, and input them into the models, and they would classify tracks as being bound.

The motive for building two classification models was to scale the predictor variables accordingly based on different time intervals and quality of data (e.g. slightly different laser intensities), compared to the training data set. First, we calculated the mean of the mean speed and the mean of the max quality from the Histone H3 training data set. Next, we performed a Gaussian Mixture Model (GMM) fit, with two components, on the log-transformed mean speed values of a given data set. We then took the mean of the first Gaussian component, likely

representing bound molecules, and with the mean of mean speed values obtained from the training data set, we scaled all 4 speed values accordingly. For example, if the mean of mean speed from the training data set was 1.71 and the mean of the first Gaussian component was 2.5 (assuming a longer time interval), then the scaling factor would be 1.71/2.5 = 0.68. We then performed classification with Model 1.

From the tracks extracted with Model classification, we again performed a GMM fit, but this time on the log-transformed max quality values (Figure S1).

We then took the mean of the second Gaussian component (likely representing genuine molecules), and used the mean of the max quality value from the training data, to scale the quality values in a similar manner as described before. This would help ensure that despite slight differences in the quality of data compared to the training data set, tracks would still be robustly classified. After scaling, we input the tracks into Model 2, as the final classification step. To confirm that ML approach can extract bound molecules we calculated the radius of gyration (R_g) of tracks from the training data set as well as tracks from Pol32 data. The R_g was calculated as follows:

1)

$$R_g = \sqrt{\left(\frac{1}{n}\right)\sum_{i}^{n}((x_i - \bar{X})^2 + (y_i - \bar{Y})^2)}$$

where n is the number of localizations of the track, $\overline{X}$ is the mean x-position, $\overline{Y}$ is the mean y-postion, and x_i and y_i, are the x and y coordinates, respectively of the ith localization.

We performed a GMM fit, with the Bayesian Information Criterion (BIC) test incorporated to prevent overfitting, on the mean intensities of tracks followed by clustering, to isolate tracks representing single-molecules and not multiple ones (Figure S2). Also, given that we use very low dye concentrations, thus <<100% dye labelling, and the very low laser power for activation, it is unlikely that we are activating multiple molecules at once.

We then fit the track durations of the remaining tracks with a truncated exponential model, to compensate for discarding short duration tracks, using Maximum Likelihood Estimation (MLE) through Matlab's "mle" function, to calculate the mean track duration.

2)

$$PDF = \left(\frac{1}{\tau}\right) e^{\frac{-(x-L)}{\tau}}$$

where τ is the mean track duration, and L is the truncation point.

The 95% confidence intervals were calculated by bootstrapping 1000 samples. Bound times were calculated using the following equation:

3)

$$Tbound = Ttrack * Tbleach/(Tbleach - Ttrack)$$

To calculate the errors on the estimate, we performed bootstrap sampling on the track durations from the combined data set (when multiple data sets of the same time interval were taken) to the following equation:

4)

$$(\frac{1}{T_{track}})e^{-t/T_{track}} = \left((\frac{1}{T_{bound}}) + (\frac{1}{T_{bleach}})\right) e^{-((\frac{1}{T_{bound}})+(\frac{1}{T_{bleach}}))t}$$

With 10% variation allowed for the T_{bleach} estimate, in order to obtain biologically sensible results.

For Cdc45, since its track durations overlapped strongly with bleaching duration, we could not calculate a reliable estimate, so we estimated a minimum bound time based on the bleach duration with 8s time interval (Figure S3). Using equation 3, we calculated track durations by varying the bound time with a fixed bleaching time. The minimum bound time we chose was the one that gave us a track duration time equal to 1 second lower than the lower bound of the CI from the estimated bleaching time.

To check for two-exponential mixtures, the track durations were fit with the following two-exponential model:

5)

$$p\left(\frac{1}{\tau_1}\right)e^{\frac{-(x-L)}{\tau_1}} + (1-p)\left(\frac{1}{\tau_2}\right)e^{\frac{-(x-L)}{\tau_2}}$$

where τ_1 = (*Tbleach* + *Tboundalpha*)/(*Tbleach* **Tboundalpha*), τ_2= (*Tbleach* + *Tboundbeta*)/(*Tbleach* * *Tboundbeta*), p is the mixture proportion, and L is the truncation point.

The lower and upper bounds on the two binding timescales were 0.1s and 900s, respectively, while allowing for a 20% variation in the bleaching estimate.

To check for overfitting and to identify whether the two-exponential model significantly fit the data better, we used the BIC test and the Loglikelihood ratio (LLR) test, as described in [6]. We looked for cases when the two-exponential model estimates did not simply return the lower and/or upper bounds as this would indicate that no physically sensible solution was found. We also performed a chi-square goodness of fit test, under the null hypothesis that the data comes from a single-exponential distribution.

Simulations of Pol δ

Simulations were performed using Python 3.7.

A 3D array of 400 x 400 x 400, was generated, with each element corresponding to a 10nm x 10nm x 10nm region. The nucleus was modelled as a sphere of diameter of 1.5um. Inside the nucleus was another sphere – termed the replisome sphere – that was 300nm in diameter, to resemble a diffraction-limited region. Both of these spheres had their centers placed at position (200, 200, 200) of the 3D array. Replisomes were placed randomly within the replisome sphere but are subsequently allowed to move throughout the nucleus. A fraction of Pol δ were bound to

replisomes while the unbound fraction (referred to as excess copy numbers) were placed throughout the nucleus. One of the bound Pol δ molecules was assigned to be fluorescent, in order to test for any rebinding behavior of a fluorescent molecule. To simulate diffusive behavior of the unbound fraction, we calculated the step size in each direction, by sampling from a normal distribution of mean 0, and a standard deviation of $\sqrt{2 * D * dt^{\alpha}}$, where dt is the time lag (0.05 seconds), α is the diffusion exponent (equal to 1 for diffusing Pol δ, and 0.4 for replisomes, representing chromosomal loci movement), and D is the diffusion coefficient. If the step size caused the molecule to go out of the nucleus, the step size was resampled until it remained in the nucleus. Movement of replisomes based on genomic loci movement, was assumed to be subdiffusive as explained by the Rouse model [36]. To represent the 200 seconds of acquisition time for the 1 second time interval data, we performed 200s/(0.05s/time step) = 4000 time steps. The residence time of Pol δ – assuming one copy per Okazaki fragment – was estimated to be 4s, based on the size of Okazaki fragments measured *in vivo* (~160bp) [37], and rate of synthesis of Pol δ under physiological conditions (~50bp/s) [38], while also allowing for delays due to strand-displacement activity and engagement with PCNA.

Two sets of transition probabilities were setup; one for the fluorescence state kinetics (e.g. photobleaching), with two states (fluorescent vs. bleaching) and the other for the molecule kinetics (e.g. unbinding), with two states (bound vs. unbound) since these were assumed to be independent from one another. Binding of Pol δ to a replisome was assumed to happen when the molecule entered a region occupied by a replisome and bound molecules would move along with their respective replisomes until dissociation. The next state of the molecule after a time step, was calculated by a sampling from a multinomial distribution with proportions given by the appropriate transition probabilities.

To determine rebinding of a fluorescent Pol δ, we calculated the overall time it spent in the fluorescent, bound state, regardless of the specific replisome it was bound to. This gives us a track duration time for one molecule within a nucleus. A fraction of 0.75 of the replisomes had a Pol δ bound at the start of the simulation. We also performed with the simulation with a fraction 0.25 with no significant effect (data not shown). We repeated this simulation 40 times (representing 40 nuclei sampled) under the same parameter set, in order to obtain a distribution of track duration times, which we can fit to using MLE, to extract the average track duration. Based on the bound time we input (4 seconds), and the measured photobleaching time with 1s interval (23 seconds), we can calculate the predicted track duration using equation 3, representing a model where there is no rebinding, and compare it to the estimated average track duration from the simulations.

For the simulations regarding the occupancy of replisomes by Pol δ, they were identical to the ones described above, except we only performed one simulation per parameter set. We then calculated for each replisome, the fraction of time it was bound by a Pol δ.