Capturing the dynamics of genome replication on individual ultra-long nanopore sequence reads

Carolin A Müller[1+*], Michael A Boemo[1+], Paolo Spingardi[2], Benedikt M Kessler[3], Skirmantas Kriaucionis[2], Jared T Simpson[4,5], Conrad A Nieduszynski[1*]

[1]Sir William Dunn School of Pathology, University of Oxford, Oxford, UK; [2]Ludwig Institute for Cancer Research, Nuffield Department of Medicine, University of Oxford, Oxford, UK; [3]Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, UK; [4]Ontario Institute for Cancer Research, Toronto, Ontario, Canada; [5]Department of Computer Science, University of Toronto, Ontario, Canada

+ These authors contributed equally

* Correspondence should be addressed to C.A.M. (carolin.muller@path.ox.ac.uk) or C.A.N. (conrad.nieduszynski@path.ox.ac.uk)

Abstract

Replication of eukaryotic genomes is highly stochastic, making it difficult to determine the replication dynamics of individual molecules with existing methods. We report a sequencing method for the measurement of replication fork movement on single molecules by Detecting Nucleotide Analogue signal currents on extremely long nanopore traces (D-NAscent). Using this method, we detect BrdU incorporated by *Saccharomyces cerevisiae* to reveal, at a genomic scale and on single molecules, the DNA sequences replicated during a pulse labelling period. Under conditions of limiting BrdU concentration, D-NAscent detects the differences in BrdU incorporation frequency across individual molecules to reveal the location of active replication origins, fork direction, termination sites, and fork pausing/stalling events. We used sequencing reads of 20-160 kb, to generate the first whole genome single-molecule map of DNA replication dynamics and discover a new class of low frequency stochastic origins in budding yeast. The D-NAscent software is available at https://github.com/MBoemo/DNAscent.git.

Introduction

Genomic methods have provided insights into DNA replication and genome stability[1-3]. Within a population of cells, these methods mask heterogeneity in replication origin usage and replication fork dynamics; what happens in individual cells is difficult to ascertain[4]. Therefore, a high-throughput single-molecule approach is needed to reveal the heterogeneity in DNA replication dynamics.

Current single-molecule techniques to study replication have provided valuable insight, but have limitations. For example, DNA combing relies on antibody detection of nucleotide analogues incorporated on the nascent strand and can be used to determine the pattern of origin activation and fork progression in single molecules[5]. However, this approach is low-throughput and provides limited temporal and spatial resolution due to the difficulty of identifying genomic positions by probe hybridization. Alternative methods use nanochannels to stretch DNA molecules, which increases throughput and can help to reveal genomic location, but the temporal and spatial resolution are limited by analogue-pulse length and image-based detection, respectively[6, 7].

We present a nanopore-based sequencing method that can measure replication fork movement by Detecting Nucleotide Analogue signal currents on extremely long nanopore traces (D-NAscent) in nascent DNA. We demonstrated that nanopore sequencing platforms can reliably distinguish base analogues from natural bases. We developed software that detects BrdU on individual nanopore sequencing reads: When BrdU is incorporated by replication forks, D-NAscent can detect these regions of incorporation. In *S. cerevisiae*, a pulse-chase experiment revealed regions replicated during the pulse, providing information comparable to that from DNA combing, but at a genomic scale and with sequence-level information. We validated D-NAscent by comparison to mass-spectrometry and population-

level sequencing data. In experiments where BrdU was limiting, we showed that D-NAscent can identify the location of replication origins and detect the changes in BrdU incorporation frequency to reveal fork direction on individual molecules. Using this approach, we have created a whole-genome profile of replication fork dynamics and origin firing on each of over 100,000 molecules 20-160 kb in length.

Results

Nucleotide analogues produce a distinct signal in nanopore sequencing

Oxford Nanopore Technologies' (ONT) MinION instrument determines a base sequence from the electrical readout produced as DNA passes through a protein pore (**Fig. 1a**). During sequencing, the double-stranded DNA substrate is unwound and a single strand enters the pore, causing a disruption to the ionic current signal. For simplicity, a common assumption is that this current signal is characteristic of a short fixed-length sequence (*k*-mer) moving through the pore. The current signal for each *k*-mer fits a Gaussian distribution. For consistency with data released by ONT, we used a *k*-mer length of six. We and others demonstrated that signal-level events can distinguish methylated from unmodified bases[8, 9]. We hypothesised that nanopore sequencing might also distinguish nucleotide analogues from natural bases, and thus reveal genomic regions synthesised during analogue pulse-chase experiments (**Fig. 1b** and **1c**). To test this hypothesis, we sequenced substrates where thymidine (at various fixed positions) had been substituted by different synthetic analogues. We observed clear differences in the event distributions between thymidine and 5-bromodeoxyuridine (BrdU) using an earlier pore version (R9 pore; discontinued 2016) and the presently available chemistry (R9.5 pore; **Fig. 1d**). The shift in signal depended on the particular analogue, the sequence context, and the position of the analogue within the 6-mer (**Fig. 1d** and **Supplemental Fig. S1**). These observations indicated that MinION sequencing has the potential to detect nucleotide analogues in genomic DNA.

Identifying the characteristic nanopore signal of BrdU in genomic DNA

BrdU is a commonly used thymidine analogue with limited cytotoxicity compared to EdU or FdU[10-13]. Thus, we sought to determine the distribution of nanopore signal events for any

BrdU-containing 6-mer in genomic DNA. An *S. cerevisiae* strain, which is dependent upon exogenous thymidine[14] was grown in various proportions of thymidine and BrdU (see Online Methods). Genomic DNA was prepared and analysed by MinION sequencing. As controls, BrdU incorporation was quantified by mass spectrometry and immunoprecipitation followed by Illumina sequencing (BrdU-seq; **Supplemental Fig. S2**, **S3** and **Supplemental Table S1**)[15]. Mass spectrometry revealed that in our five genomic DNA samples the percentage of thymidines substituted by BrdU was 0%, 15%, 26%, 49% and 79%, respectively.

Nanopore signal events were aligned to the genomic reference using nanopolish[16]. These data revealed many thymidine-containing 6-mers where the distribution of signal events was bimodal; while one population matched the ONT model, there was a distinct second population (**Fig. 2a**). The relative proportions of the two populations reflected the concentration of incorporated BrdU. By contrast, 6-mers that did not contain thymidine were mono-modal and matched the ONT model (data not shown). We fitted a bimodal Gaussian mixture model to the signal events from the 49% BrdU sample that aligned to each thymidine-containing 6-mer (**Fig. 2b and Supplemental Fig. S4)** for results using other incorporation rates). We used the Kullback-Leibler (KL) divergence, which measures the average log-difference between two probability distributions, to quantify the difference between the ONT model and each of the two fit distributions. One distribution (fit 1) was close to the ONT model (only ~1% of 6-mers had a KL-divergence >0.5) while the second distribution (fit 2) was farther away from the ONT model (~62% 6-mers had a KL-divergence >0.5) and corresponded to the BrdU concentration-dependent population (**Fig. 2b** and **c**). We concluded that the second distribution represented the BrdU signal.

We note that even at high BrdU substitution levels 6-mers featuring multiple thymidines gave only a bimodal distribution of signal events (**Fig. 2a**). This is consistent with

the observation above that BrdU predominantly shifts the signal event at particular positions within the 6-mer (**Supplemental Fig. S1**). To assess this further, we examined the subset of 6-mers containing a single thymidine and observed the greatest shift in signal event when BrdU is the third or fourth base from the 5' end (**Fig. 2d**). These data indicate that it will be possible for MinION sequencing to distinguish BrdU from thymidine in genomic DNA.

Detection of BrdU incorporated in vivo

Detection relied on two thresholds: those 6-mers for inclusion in the model and a certainty above which a position is called as BrdU. Including only those 6-mers where BrdU caused a KL-divergence from the ONT model >2 (*N*=159; **Fig. 2c**) allowed assessment of BrdU incorporation on average every 21 nucleotides across the yeast genome (**Supplemental Fig. S5a**). Each time one of these 6-mers occurred in a sequencing read, we used a Hidden Markov model (HMM) to compute the log-likelihood that the 6-mer contained a BrdU (**Supplemental Fig. S6**). A position in a read was classified as BrdU if the log-likelihood exceeded a threshold (**Fig. 2e** and **Supplemental Fig S5b**); this threshold was determined by testing the HMM on unused training data and on an equivalent thymidine-only sample to determine true and false positive rates (**Fig. 2f** and **Supplemental Fig. S5b**). We set this threshold at log-likelihood >2.5, which gave a true positive rate of ~60% for a false positive rate of ~3%. We achieved a similar true positive rate using a DNA sample with an intermediary BrdU concentration (26% incorporation) that was unrelated to the training material.

To further test our detection strategy, we generated hemi-BrdU substituted yeast genomic DNA, by synchronizing a strain prototrophic for thymidine[17] and passing it through one S phase in media containing a high concentration (400 μg/ml) of BrdU. Material was validated by mass spectrometry (**Supplemental Fig. S2**) and BrdU-seq to reveal any

incorporation bias (**Supplemental Fig. S7**). Cell cycle synchrony was confirmed by flow cytometry of DNA content (**Supplemental Fig. S8**). MinION sequencing was performed and positions of BrdU incorporation were determined as described above. As anticipated, we observed reads with either low or predominantly high density of BrdU calls over the entire read, consistent with parental and nascent strands, respectively (**Fig. 3a**). To quantify the frequency of BrdU calls in each read, we fit the number of BrdU calls in non-overlapping 2 kb windows to a binomial distribution (see Online Methods). This allowed us to compare the number of BrdU calls in each window to what was expected if the window was BrdU-positive (determined from mass spectrometry data and the true positive rate). Computing the z-score of BrdU calls in each window against this binomial distribution revealed a bimodal density: One population was centered around the mean, indicating these windows had a BrdU frequency consistent with our expectation for a BrdU-positive window; the other population was centered ~5 s.d. below the mean, indicating these windows had significantly fewer BrdU calls than expected for BrdU-positive windows (**Fig. 3b**). We concluded that these two populations correspond to BrdU-containing and thymidine-only windows, respectively. Therefore, MinION sequencing can detect BrdU incorporated at replication forks; we call our method Detecting Nucleotide Analogue signal currents on extremely long nanopore traces (D-NAscent).

Single-molecule detection of replication origin activity in hydroxyurea

DNA synthesis can be slowed by addition of hydroxyurea (HU). To restrict BrdU incorporation to locations proximal to early activating replication origins, we released cells synchronously from G1 into S phase in the presence of HU and BrdU (40 μg/ml)[18-20]. After 60 minutes, cells were chased out of HU with excess thymidine (400 μg/ml), into nocodazole (**Fig. 3c**). After

158 completion of S phase, samples were collected for D-NAscent, mass spectrometry, and BrdU-
159 seq. Cell cycle synchrony was assessed by flow cytometry of DNA content (**Supplemental Fig.**
160 **S8**). BrdU detected in the MinION data was summed across all reads in non-overlapping 100
161 bp windows to allow comparison to the BrdU-seq data (**Fig. 3d**). This confirmed that an
162 ensemble of our single-molecule data was in good agreement with established short-read
163 methods (Pearson correlation coefficient, *R*=0.76). Visualising BrdU calls on single molecules
164 suggested that each sequencing read fell into one of two categories: there was either an
165 infrequent number of BrdU calls throughout the whole read, or the read contained short
166 patches of frequent BrdU calls (**Fig. 3e**). We concluded that these reads likely correspond to
167 parental and nascent strands, respectively. For each read, we assessed non-overlapping 2 kb
168 windows and quantified the frequency of BrdU calls by computing the z-score of positive calls
169 against a binomial distribution as before (**Fig. 3e**). For example, two early firing replication
170 origins on chromosome VI gave rise to BrdU-positive regions on a single >100 kb read (read 3
171 in **Fig. 3e**), indicating that both origins activated in a single cell.

172 We visualised the BrdU frequency z-scores for individual nascent-strand reads that
173 spanned known replication origins (reads covering >4 kb either side of the origin)[21]. In **Fig. 3f**,
174 each row represents an individual sequencing read centred upon an origin. The majority of
175 reads had the highest frequency of BrdU calls at the origin. Most of these nascent molecules
176 spanned efficient, early activating origins ('unchecked' by the intra-S phase checkpoint). On
177 average, early activating origins had incorporated more BrdU than later activating origins; in
178 both cases the signal was symmetric and centred on the origins (**Fig. 3g**). However, some
179 molecules showed asymmetric levels of BrdU incorporation relative to the origin, indicative
180 of different rates of sister fork progression (**Supplemental Fig. S9**). We note that the BrdU z-
181 score in individual molecules is highest at replication origins and falls away as a function of

distance from the origin (**Fig. 3e** and **f**), eventually dropping below our detection threshold and is called as a thymidine window. This implied a time-dependent drop in BrdU incorporation as forks progressed away from origins. This led us to hypothesize that we might be able infer fork-direction from the gradient of BrdU z-scores.

Single-molecule replication dynamics

To examine replication dynamics in the absence of stress, we synchronised thymidine-prototroph cells in G1 and released in the presence of BrdU (40 μg/ml) (**Fig. 4a** and **Supplemental Fig. S8**). Samples were collected for D-NAscent, mass spectrometry, BrdU-seq, and DNA copy number measurements[22]. We found good agreement between D-NAscent and population-level BrdU-seq data (Pearson correlation coefficient, R=0.75; **Fig. 4b**); both assessments confirmed that the population average level of BrdU incorporation was inversely correlated with average replication time (**Supplemental Fig. S10**). This is consistent with the intracellular ratio of BrdU to thymidine falling as cells progress through S phase; likely due to activation of endogenous thymidine synthesis pathways and the BrdU import rate being lower than incorporation[23, 24]. Within individual reads, we assessed non-overlapping 2 kb windows and computed the z-score of BrdU calls in each window to a binomial distribution (examples shown in **Fig. 4c**). We observed clear peaks of BrdU incorporation near known replication origins. As forks move away from initiation sites, the frequency of BrdU calls decreases as the concentration of cellular BrdU falls (**Fig. 4c**).

We determined the gradient of BrdU z-scores as a proxy for replication fork direction (**Fig. 4c**). Regions of diverging replication forks were used to call sites where replication initiated early in S phase. This provided a whole-genome map of replication origin activity on single molecules. Examining all identified replication initiation sites revealed two distinct

classes (**Supplemental Fig. S11**). The first class were found on multiple independent sequence reads, consistent with high-efficiency origins identified by population level analyses[21]. The second class were dispersed throughout the genome with each site identified in a minority of molecules. These sites did not correspond to known replication origins. We considered that these sites could represent BrdU incorporation from a DNA repair pathway prior to S phase or a sequencing/analysis artefact. However, DNA repair synthesis prior to S phase would be confined to parental strands and synthesis would be unidirectional; the novel initiation sites were present on nascent strands and showed bi-directional synthesis evidenced by bilateral BrdU-gradients (**Fig. 4c**). Furthermore, the BrdU call frequency and signal at novel initiation sites resembled that observed at previously reported origins, and applying more conservative criteria for origin identification (higher z-score and longer contiguous region of BrdU-positive windows) reduced the number of origin calls but did not diminish the proportion of initiation events at novel locations (data not shown). Therefore, D-NAscent allows the detection of replication initiation sites that are too infrequently used to be detected by population-level methods.

Identification of replication fork pausing on single molecules

We sought to determine whether D-NAscent allowed detection of replication fork pausing events. Yeast ribosomal DNA (rDNA) repeats each contain a replication origin and a unidirectional replication fork barrier (RFB) that pauses one of the sister forks (**Fig. 4d**)[25]. The repetitive nature of the rDNA limits analysis by short-read technologies, but our nanopore reads included thousands of molecules that each spanned multiple repeats. D-NAscent demonstrated an asymmetric peak in the BrdU average z-score signal across a single rDNA repeat (**Fig. 4d**). In this ensemble analysis, the average BrdU z-score was maximal at the

replication origin. The fall in z-score to the right of the origin was positioned over the RFB and is consistent with unidirectional fork pausing. As expected, the leftward moving fork showed no such delay. Visualisation of individual molecules (**Fig. 4e**) demonstrated origin firing in a subset of repeats and pausing of rightward moving forks at the RFB. Therefore, D-NAscent allowed single-molecule analysis of replication fork direction, initiation sites, termination sites and fork pausing/stalling.

Discussion

We have developed a genomic single-molecule method for detection of base analogues that we term D-NAscent. Base analogues are widely used for the study of chromosome biology[26], cell proliferation[27] and gene expression[28]. D-NAscent offers a powerful method for the advancement of each of these fields. Key features of nanopore sequencing[29] make D-NAscent possible: lack of an amplification step ensures that *in vivo* incorporated analogues are present in the sequenced strand; interrogation of single DNA strands permits direct detection of the analogue and provides single-molecule information; and extremely long sequence read lengths (>100 kb) allow detection of long-range *cis* interactions (**Supplemental Fig. S12**). The MinION nanopore sequencing platform allowed us to develop a model for detection of *in vivo* incorporated BrdU that we validated against mass spectrometry and population level BrdU-seq data (**Fig. 1** and **Fig. 2**). Our approach will allow detection of other analogues, such as EdU or IdU. The sensitivity of our BrdU-detection model allowed us to measure changes in BrdU incorporation frequency on nascent strands and thereby revealed the temporal order of DNA replication on single molecules. Given a replication fork velocity of ~2 kb/min[30, 31] and our 2 kb resolution, D-NAscent has the potential to provide a temporal resolution of close to 1 min.

DNA replication is a stochastic process and many aspects are masked in population-based approaches. By applying D-NAscent to the study of yeast chromosome biology, we have generated the first whole-genome map of replication dynamics at single-molecule level. Unexpectedly, we discovered a novel class of replication origin that could not have been discovered by established methods. While most initiation sites that we detected (~80%) were near known origins, approximately one fifth of replication initiation events occurred dispersed throughout the genome. Recent *in vitro* studies have demonstrated origin-dependent and independent DNA replication initiation, due to promiscuous binding of the origin recognition complex[32, 33]. Although *in vitro* origin-independent initiation was only observed in the absence of physiological levels of competitor DNA, it is consistent with our finding that many genomic locations can function at low frequency as replication origins. We propose that replication of the yeast genome is initiated from both well-defined, high-efficiency origins and a broadly distributed set of very low-efficiency origins, similar to the configuration observed in mammalian cells[2].

The D-NAscent single-molecule methodology will allow many unresolved questions in chromosome biology to be addressed. The power to explore *cis* regulatory mechanisms is enhanced by the extremely long sequencing reads; in this study we present reads <160 kb, but others have reported ultra-long reads of >1 Mb[34]. Single-cell approaches have relatively low spatial resolution, but they can provide *trans* information missing in single molecules[35, 36]. However, we and others have discovered that replication origin activity is generally regulated in *cis*[37-40] emphasising the importance of the single-molecule approach. Second, variants of the MinION library preparation allow capture of sequence information from both DNA strands ($1D^2$). Combining D-NAscent with $1D^2$ sequencing has the potential to reveal sites of conservative DNA replication, for example during recombination-dependent DNA

synthesis[41, 42]. Third, D-NAscent can examine replication patterns in complex genomic locations (e.g. non-unique or repetitive sequences) that are abundant in mammalian genomes and generally understudied. Fourth, the ability of D-NAscent to detect nascent DNA depends on incorporation of nucleotide analogues; achievable in many organisms and all commonly utilised model systems. This, and the gigabase throughput of nanopore sequencing platforms will allow application of D-NAscent to many organisms, including the study of large, stochastically replicated mammalian genomes. Applying D-NAscent to hemi-labelled human genomic DNA allowed us to discriminate nascent and parental strands based on BrdU detection frequency (personal communication, C Nieduszynski, R Wilson and J Carrington). In the future, this will allow the genome-wide identification of mammalian replication origins by D-NAscent.

Acknowledgments

We are grateful to John Diffley (Francis Crick Institute, UK) and Etienne Schwob (Institut de Génétique Moléculaire de Montpellier, France) for kindly providing strains; Amanda Williams and Rebecca Busby for Illumina NextSeq support; Michal Maj for assistance with flow cytometry; Joseph Caesar for IT infrastructure support; Microsoft and NVIDIA for computing resources. We thank Nieduszynski group members for helpful discussion and advice; Anthony Carr, Anne Donaldson, Shin-ichiro Hiraga and David Sherratt for critical reading of the manuscript.

This work was supported by Biotechnology and Biological Sciences Research Council grant BB/N016858/1 (CAM and CAN) and Wellcome Trust Investigator Award 110064/Z/15/Z (CAN). PS is funded by an MRC studentship; PS and SK are funded by Ludwig Cancer Research.

CAM is a Queen's College Extraordinary Junior Research Fellow in Physiology; MAB is a St. Cross College Emanoel Lee Junior Research Fellow.

Author contributions

CAN, CAM and MAB designed the study; CAM designed and performed the experiments; MAB designed and implemented the analogue-training and detection software; CAM and MAB undertook data analysis; PS undertook mass spectrometry, supervised by BMK and SK; CAN and JTS supervised the study; CAN, CAM and MAB wrote the paper.

Potential conflict of interest

JTS receives research funding from Oxford Nanopore Technologies and has received travel support to speak at meetings organized by Oxford Nanopore Technologies.

References

1. Gilbert, D.M. Evaluating genome-scale approaches to eukaryotic DNA replication. *Nat Rev Genet* **11**, 673-684 (2010).
2. Mechali, M. Eukaryotic DNA replication origins: many choices for appropriate answers. *Nat Rev Mol Cell Biol* **11**, 728-738 (2010).
3. Raghuraman, M.K. et al. Replication dynamics of the yeast genome. *Science* **294**, 115-121. (2001).
4. Hawkins, M. et al. High-resolution replication profiles define the stochastic nature of genome replication initiation and termination. *Cell Rep* **5**, 1132-1141 (2013).
5. Tuduri, S., Tourriere, H. & Pasero, P. Defining replication origin efficiency using DNA fiber assays. *Chromosome Res* **18**, 91-102 (2010).

6. Lacroix, J. et al. Analysis of DNA Replication by Optical Mapping in Nanochannels. *Small* **12**, 5963-5970 (2016).
7. De Carli, F. et al. High-Throughput Optical Mapping of Replicating DNA. *Small Methods* **2**, 1800146 (2018).
8. Simpson, J.T. et al. Detecting DNA cytosine methylation using nanopore sequencing. *Nat Methods* **14**, 407-410 (2017).
9. Rand, A.C. et al. Mapping DNA methylation with high-throughput nanopore sequencing. *Nat Methods* **14**, 411-413 (2017).
10. Hagenkort, A. et al. dUTPase inhibition augments replication defects of 5-Fluorouracil. *Oncotarget* **8**, 23713-23726 (2017).
11. Hua, H. & Kearsey, S.E. Monitoring DNA replication in fission yeast by incorporation of 5-ethynyl-2'-deoxyuridine. *Nucleic Acids Res* **39**, e60 (2011).
12. Diermeier-Daucher, S. et al. Cell type specific applicability of 5-ethynyl-2'-deoxyuridine (EdU) for dynamic proliferation assessment in flow cytometry. *Cytometry A* **75**, 535-546 (2009).
13. Talarek, N., Petit, J., Gueydon, E. & Schwob, E. EdU Incorporation for FACS and Microscopy Analysis of DNA Replication in Budding Yeast. *Methods Mol Biol* **1300**, 105-112 (2015).
14. Vernis, L., Piskur, J. & Diffley, J.F. Reconstitution of an efficient thymidine salvage pathway in Saccharomyces cerevisiae. *Nucleic Acids Res* **31**, e120 (2003).
15. Peace, J.M., Villwock, S.K., Zeytounian, J.L., Gan, Y. & Aparicio, O.M. Quantitative BrdU immunoprecipitation method demonstrates that Fkh1 and Fkh2 are rate-limiting activators of replication origins that reprogram replication timing in G1 phase. *Genome Res* **26**, 365-375 (2016).
16. Loman, N.J., Quick, J. & Simpson, J.T. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat Methods* **12**, 733-735 (2015).

17. Magiera, M.M., Gueydon, E. & Schwob, E. DNA replication and spindle checkpoints cooperate during S phase to delay mitosis and preserve genome integrity. *J Cell Biol* **204**, 165-175 (2014).

18. Alvino, G.M. et al. Replication in hydroxyurea: it's a matter of time. *Mol Cell Biol* **27**, 6396-6406 (2007).

19. Feng, W. et al. Genomic mapping of single-stranded DNA in hydroxyurea-challenged yeasts identifies origins of replication. *Nat Cell Biol* **8**, 148-155 (2006).

20. Poli, J. et al. dNTP pools determine fork progression and origin usage under replication stress. *EMBO J.* **31**, 883-894 (2012).

21. Siow, C.C., Nieduszynska, S.R., Muller, C.A. & Nieduszynski, C.A. OriDB, the DNA replication origin database updated and extended. *Nucleic Acids Res* **40**, D682-686 (2012).

22. Müller, C.A. et al. The dynamics of genome replication using deep sequencing. *Nucleic Acids Res* **42**, e3 (2014).

23. Vernis, L., Piskur, J. & Diffley, J.F. Reconstitution of an efficient thymidine salvage pathway in Saccharomyces cerevisiae. *Nucleic Acids Res* **31**, e120 (2003).

24. Koc, A., Wheeler, L.J., Mathews, C.K. & Merrill, G.F. Hydroxyurea arrests DNA replication by a mechanism that preserves basal dNTP pools. *J Biol Chem* **279**, 223-230 (2004).

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

26. Cavanagh, B.L., Walker, T., Norazit, A. & Meedeniya, A.C. Thymidine analogues for tracking DNA synthesis. *Molecules* **16**, 7980-7993 (2011).

27. Kee, N., Sivalingam, S., Boonstra, R. & Wojtowicz, J.M. The utility of Ki-67 and BrdU as proliferative markers of adult neurogenesis. *J Neurosci Methods* **115**, 97-105 (2002).

28. Dolken, L. et al. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* **14**, 1959-1972 (2008).
29. Jain, M., Olsen, H.E., Paten, B. & Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol* **17**, 239 (2016).
30. Hawkins, M. et al. High-resolution replication profiles define the stochastic nature of genome replication initiation and termination. *Cell reports* **5**, 1132-1141 (2013).
31. Sekedat, M.D. et al. GINS motion reveals replication fork progression is remarkably uniform throughout the yeast genome. *Mol Syst Biol* **6**, 353 (2010).
32. On, K.F. et al. Prereplicative complexes assembled in vitro support origin-dependent and independent DNA replication. *EMBO J.* **33**, 605-620 (2014).
33. Gros, J., Devbhandari, S. & Remus, D. Origin plasticity during budding yeast DNA replication in vitro. *EMBO J.* **33**, 621-636 (2014).
34. Payne, A., Holmes, N., Rakyan, V. & Loose, M. BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics* (2018).
35. Chen, C. et al. Single-cell whole-genome analyses by Linear Amplification via Transposon Insertion (LIANTI). *Science* **356**, 189-194 (2017).
36. Dileep, V. & Gilbert, D.M. Single-cell replication profiling to measure stochastic variation in mammalian replication timing. *Nat Commun* **9**, 427 (2018).
37. Natsume, T. et al. Kinetochores coordinate pericentromeric cohesion and early DNA replication by Cdc7-Dbf4 kinase recruitment. *Mol Cell* **50**, 661-674 (2013).
38. Fang, D. et al. Dbf4 recruitment by forkhead transcription factors defines an upstream rate-limiting step in determining origin firing timing. *Genes Dev* **31**, 2405-2415 (2017).
39. Hiraga, S.I. et al. Human RIF1 and protein phosphatase 1 stimulate DNA replication origin licensing but suppress origin activation. *EMBO Rep* **18**, 403-419 (2017).
40. Foti, R. et al. Nuclear Architecture Organized by Rif1 Underpins the Replication-Timing Program. *Mol Cell* **61**, 260-273 (2016).

41. Donnianni, R.A. & Symington, L.S. Break-induced replication occurs by conservative DNA synthesis. *Proc Natl Acad Sci U S A* **110**, 13475-13480 (2013).

42. Saini, N. et al. Migrating bubble during break-induced replication drives conservative DNA synthesis. *Nature* **502**, 389-392 (2013).

Figure legends

Figure 1: Nanopore sequencing can distinguish thymidine from analogues. (**a**) Graphic representation of DNA sequencing using a MinION. A processive enzyme (green) ratchets DNA into the pore (blue), causing a change in ionic current (ions shown as black dots) that is determined by the 6-mer in the central channel (purple box). The current is recorded over time (black trace, bottom right). (**b**) Schematic representation of pulse-labelling early replicating regions with thymidine analogues. (**c**) Outline of the experimental strategy for BrdU detection by nanopore sequencing. (**d**, **e**) For example 6-mer GCCTGA, each panel shows the distribution of signal events for thymidine (blue) and various analogues: BrdU (red); FdU (green); IdU (black); and EdU (yellow).

Figure 2: BrdU can be distinguished from thymidine in genomic DNA. (**a**) Signal event distributions for an example 6-mer from yeast genomic DNA containing various concentrations of BrdU compared to the ONT model (grey). (**b**) Bimodal Gaussian mixture model fit (purple and turquoise) for an example 6-mer from genomic DNA containing 49% BrdU. The ONT model is shown in grey. (**c**) Distribution of the KL-divergence between the ONT model and Gaussian fit 1 (upper) or fit 2 (lower) for all thymidine-containing 6-mers. (For detection (Fig. 2e), we make a BrdU call for all 6-mers that have a KL-divergence >2.0; dashed line.) (**d**) Distributions as in the lower plot from (c) but for the subset of 6-mers containing

just one thymidine; plotted by the position of the thymidine. (**e**) Signal event distributions from the ONT model (thymidine; grey) and from the bimodal Gaussian mixture model fit for BrdU (red). The KL-divergence of the two 6-mers is indicated. (**f**) Receiver operating characteristic (ROC) curve, using all 6-mers that have a KL-divergence >2.0, specifying the true positive and false positive rates for various log-likelihood thresholds of BrdU compared to thymidine (see Online Methods). Numbers near points specify the log-likelihood threshold above which a position in a read is classified as BrdU. The dashed lines demarcate the true and the false positive rates at a log-likelihood threshold >2.5.

Figure 3: Single-molecule detection of BrdU on nascent DNA. (**a**) Representative nanopore reads (>15 kb) showing BrdU calls in hemi-substituted yeast genomic DNA. Red ticks indicate BrdU calls and arrows give the read direction relative to the sacCer3 reference genome. (**b**) The distribution of BrdU call frequency measured as a z-score of a binomial distribution for non-overlapping 2 kb windows. (For later analysis we set a binomial z-score threshold >-2 for assigning a window as BrdU positive.) (**c**) Schematic of the experimental strategy for detection of replication origin activity in HU. (**d**) Comparison of BrdU-seq and an ensemble of D-NAscent data across chromosome II (from timepoint 4). Circles denote the location of 'unchecked' (yellow) and 'checked' (blue) replication origins[19]. (**e**) Four example nanopore sequencing reads that illustrate BrdU detection on parental (read 1) and nascent strands (reads 2 – 4) mapping to the right end of chromosome VI. Each read shows BrdU calls at individual 6-mers (upper track), BrdU-positive 2 kb windows (orange; middle track), and the z-score for each window where red bars are above the detection threshold (z-score $\geq$ -2) and are BrdU-positive (lower track). Confirmed replication origins from OriDB (yellow boxes) and genes (grey boxes) are shown. (**f**) Visualisation of D-NAscent data for 1,325 individual nanopore reads (rows) that span confirmed replication origins (OriDB). The colour gradient indicates the frequency of

BrdU calls within 2 kb windows, and reads are sorted vertically by the population average BrdU-seq data: reads that span the most active origins are near the top. Additional colour bars show population-level data for BrdU-seq, origin activation efficiency[4] and whether the origin is 'checked' by the intra-S phase checkpoint[19]. (**g**) Ensemble BrdU z-score from D-NAscent for all (N = 1,325) 'unchecked' (green) and 'checked' (black) origins (BrdU z-scores averaged for each column in (**f**); shaded areas show the standard error of the mean).

Figure 4: Single-molecule detection of replication dynamics. (**a**) Schematic of the experimental strategy for detection of replication dynamics by D-NAscent. (**b**) Comparison BrdU-seq data and an ensemble of D-NAscent data across chromosome II (from timepoint t_2). Origins annotated as confirmed in OriDB are shown (yellow circles). (**c**) An example 150 kb nanopore sequencing read showing BrdU calls at individual 6-mers (top track), the z-score for each 2 kb window where BrdU-positive window z-scores are shown in red and thymidine-only window z-scores are shown in blue (middle track), and called fork direction and replication initiation sites (lower tracks). Origin calls from all spanning nanopore reads (black bars: tall, close to known origins; short, >3.9 kb (**Supplemental Fig. S11**) from known origins) and origins annotated as confirmed or likely by OriDB (yellow boxes) are displayed. (**d**) (*top*) A schematic representation of a single rDNA repeat showing the origin, replication fork barrier (RFB), predominant replication fork direction (line arrows) and the major transcripts (open arrows). (*bottom*) An ensemble of D-NAscent z-scores averaged over all nanopore sequence reads that spanned an rDNA repeat and had at least one BrdU-positive 2 kb window. (**e**) The D-NAscent BrdU z-scores from selected molecules aligned to multiple rDNA repeats (origin, yellow; RFB, purple).

Methods

Defined substrates

Primers CA1218 and CA1219 (**Supplemental Table S2**) were annealed and extended with BIOTAQ DNA polymerase (Bioline) in the presence of dCTP, dGTP, dATP and either dTTP, BrdU-TP, FdU-TP, IdU-TP or EdU-TP (Jena Bioscience) each at 5 mM. Nanopore substrates must exceed a length of 250 bp. Thus, extended primers were digested with *Xma*I (NEB) and ligated to *Age*I-digested DNA sequences (>350 bp). Ligation products were gel purified prior to Nanopore sequencing.

Yeast DNA for model training

Thymidine-auxotrophic yeast strain YLV11 was grown overnight in YPG (Formedium) supplemented with 100 μM thymidine. Cells were then diluted to an OD_{600} of 0.06 into fresh YPG supplemented with 100 μM of BrdU and/or thymidine (0%, 40%, 60%, 80% or 100% BrdU). Cells were grown at 30°C for 24 hours before samples were taken for Nanopore sequencing, mass spectrometry analysis and BrdU-IP sequencing.

Yeast cell cycle experiments

Cell cycle experiments were performed with yeast strain E3087 (**Supplemental Table S3**)[17]. Cells were grown in YPD media and arrested in G1 phase using alpha-factor. BrdU was added to a final concentration of either 400 μg/ml (hemi-labelled genomic DNA relating to Figure 3a) or 40 μg/ml (HU experiment and limiting BrdU concentration experiment, Figures 3c-f and 4, respectively). BrdU was added 25 minutes prior to pronase-mediated release into S phase, followed by an arrest in G2/M by nocodazole treatment. For the HU experiment, 200 mM HU were added concomitantly with BrdU; cells were released into S phase for 45 minutes, then 400 μg/ml thymidine was added and 15 minutes later, cells transferred into fresh YPD with

thymidine (400 μg/ml). Flow cytometry samples were taken at regular time intervals to assess cell cycle progression of each time course. Samples were treated with RNaseA and Proteinase K prior to DNA staining with SYTOX Green (ThermoFisher S7020) and analysis on a BD LSRFortessa X-20 cell analyser. Samples for Nanopore sequencing, mass spectrometry analysis, DNA copy number measurements and BrdU-IP sequencing were taken at defined time points in every cell cycle experiment. Genomic DNA was purified using phenol-chloroform extraction, RNaseA and Proteinase K treatment followed by ethanol precipitation.

Mass spectrometry validation

The equivalent ratio of 1 μg of DNA in 100 μl of water was added to 200 μl of hydrolysis solution (100 mM NaCl, 20 mM MgCl2, 20 mM Tris pH 7.9, 1000 U/ml Benzonase, 600 mU/ml Phosphodiesterase I, 80 U/ml Alkaline phosphatase, 36 μg/ml EHNA hydrochloride, 2.7 mM deferoxamine). The mixture was incubated for two hours and then lyophylised by SpeedVac. The lyophylisate was resuspended in 100 μl of buffer A (10 mM ammonium acetate, pH 6) per 1 ug of DNA used and half was transferred into an LC-MS vial for analysis. For the analysis by HPLC– QQQ mass spectrometry, a 1290 Infinity UHPLC was fitted with a Zorbax Eclipse plus C18 column, (1.8 μm, 2.1 mm 150 mm; Agilent) and coupled to a 6495a Triple Quadrupole mass spectrometer (Agilent Technologies) equipped with a Jetstream ESI-AJS source. The data were acquired in dMRM mode using positive electrospray ionisation (ESI1). Mass spectrometry was used for rare nucleosides and abundant nucleosides were quantified by HPLC-UV. The AJS ESI settings were as follows: drying gas temperature 230 °C, the drying gas flow 14 lmin-1, nebulizer 20 psi, sheath gas temperature 400 °C, sheath gas flow 11 l/min, Vcap 2,000 V and nozzle voltage 0 V. The iFunnel parameters were as follows: high pressure RF 110 V, low pressure RF 80 V. The fragmentor of the QQQ mass spectrometer was set to

380 V and the delta EMV set to +200. The UV quantification wavelength was 254 nm. The gradient used to elute the nucleosides started by a 5-min isocratic gradient composed with 100% buffer A and 0% buffer B (composed of 100% methanol) with a flow rate of 0.4 ml/min and was followed by the subsequent steps: 5-8 min, 94.4% A; 8–9 min, 94.4% A; 9–16 min 86.3% A; 16–17 min 0% A; 17– 21 min 0% A; 21–24.3 min 100% A; 24.3–25 min 100% A. The gradient was followed by a 5 min post time to re-equilibrate the column. The raw mass spectrometry data was analysed using the MassHunter Quant Software package (Agilent Technologies, version B.07.01). For the identification of compounds, raw mass spectrometry data was processed using the dMRM extraction function in the MassHunter software.

Illumina sequencing

Yeast genomic DNA samples were assessed by BrdU-seq using the NextSeq 500 (Illumina). Genomic DNA was sheared to ~300 bp using a Bioruptor. Sheared DNA was end-repaired and A-tailed using the NEBNext Ultra II end-repair module (E7546). A-tailed genomic DNA was barcoded using Illumina-compatible primers and NEBNext Ultra II ligation mix. Equal quantities of barcoded DNA samples were pooled and 20 ng of pooled DNA was reserved as "Input". At least 1 microgram of pooled barcoded DNA was denatured and subjected to immunoprecipitation using an anti-BrdU antibody (BD 347580, see "Life Sciences Reporting Summary") and Protein-G dynabeads (ThermoFisher 10003D). Immuno-precipitated DNA was purified with AMPure XP bead. The immuno-precipitated and input DNA samples were PCR amplified separately, using Illumina-compatible indexing primers and the NEBNext Ultra II Q5 Master Mix. DNA samples were sequenced (80 bp single-end) on a NextSeq 500. Illumina sequencing reads were demultiplexed and the barcode sequences were trimmed using the FASTX toolkit. Sequencing reads were mapped to the sacCer3 genome assembly using

bowtie2 (v2.3.4., see “Life Sciences Reporting Summary”). Read tag counts were determined for the 5’ end of uniquely mapping reads without mismatches in 100 bp nonoverlapping regions. The ratio between IP and Input sample was calculated for each bin, excluding bins that had less than 20% of the expected number of reads in the input sample. Ratios were median-smoothed over 1 kb windows.

Nanopore sequencing

Samples were prepared for nanopore sequencing according to recommendations by Oxford Nanopore Technologies (ONT). The 2D library kit (SQK-LSK208) and $1D^2$ library kit (SQK-LSK308) were used for synthetic substrates (Fig. 1), the 1D ligation-based library kit (SQK-LSK109) was used for the HU (Fig. 3c-f) and BrdU-depletion experiments (Fig. 4), and the 1D Native barcoding genomic kit (EXP-NBD103 and SQK-LSK108) for yeast genomic training material (Fig. 2 and Fig. 3a,b). The yeast genomic training DNA was sheared to an average length of 8 kb using g-TUBE (Covaris, 520079). The input DNA for all other nanopore libraries was unsheared. Quantitities of input DNA were adjusted to average molecule lengths, ranging between 12 ng and 5 μg for short synthetic and unsheared high-molecular weight genomic DNA, respectively. Input DNA for all libraries was end-repaired using NEBNext End Repair Module (NEB, E6050). In addition, genomic input DNA was treated with NEBNext FFPE RepairMix (NEB, M6630) to repair nicks. End-repaired samples were purified using 1x (synthetic and genomic training material) or 0.4x (HU and BrdU-depletion experiment) AMPure XP beads (Beckman Coulter, A63880). Then, ONT barcodes and/or adaptors specific to each library kit were ligated onto the samples. For 1D libraries, adapters were ligated using NEBNext Quick T4 DNA Ligase (NEB, E6065), followed by library purification using 0.4x AMPure XP beads with ONT's wash buffer enriching for long molecules. For pooled 1D

libraries, end-repaired samples were first ligated to ONT barcodes using Blunt/TA Ligase Master Mix (NEB, M0367), cleaned up using 1x AMPure XP beads and pooled in equal amounts prior to adapter ligation and final purification as above. The $1D^2$ library preparation included Adapter ligation using Blunt/TA Ligase Master Mix, 0.4x AMPure XP bead purification, followed by sequencing adapter ligation using Blunt/TA Ligase Master Mix and AMPure XP bead clean up with proprietary ONT ABB wash buffer. 2D library preparation included ligation using Blunt/TA Ligase Master Mix and a proprietary mix of two adapters, one linear, the other a biotinylated hairpin. For purification after adaptor and tether ligation, My-One Streptavidin C1 Dynabeads (Thermo Fisher) were used to enrich for molecules with a hairpin. Nanopore libraries were sequenced on a MinION Mk1 sequencer using flow cell versions R9 (2D library), R9.4 (1D ligation libraries) and R9.5 ($1D^2$ library).

Model training

Nanopore reads sequenced from *S. cerevisiae* genomic material with 49% BrdU incorporation were basecalled using the Albacore basecalling software (v2.1.10, see "Life Sciences Reporting Summary") provided by ONT (see **Supplemental Fig. S4** for results using other incorporation rates). We aligned the reads to the *S. cerevisiae* sacCer3 genome assembly using minimap2 (v2.10, see "Life Sciences Reporting Summary") with the "-a map-ont" setting[43]. We excluded those reads that aligned to mitochondrial DNA or ribosomal DNA, and for each remaining read that mapped uniquely to the genome (mapping quality $\geq$ 20), we aligned the signal events to their respective positions on the reference using nanopolish eventalign. For each thymidine-containing 6-mer in our reads, we gathered all signal events that aligned to that 6-mer; hence, the events gathered for each 6-mer were taken from a range of genomic sequence contexts. We observed that the distribution for these events

sometimes had an elevated kurtosis; this broader tail can be the result of alignment artefacts or sequence context effects. Therefore, for each of these 6-mers that had greater than 200 aligned events, the signal events were filtered for outliers using a DBSCAN algorithm to eliminate any trace artefacts. The remaining events were used to fit a bimodal Gaussian mixture model. For each component of the fit mixture model, we computed the KL-divergence against the ONT 6-mer pore model. The distribution that had the higher KL-divergence against the ONT pore model was designated as the BrdU distribution.

BrdU detection

As in model training, Nanopore reads were basecalled using the Albacore basecalling software (v2.1.10, see "Life Sciences Reporting Summary") and the reads were aligned to the *S. cerevisiae* sacCer3 genome assembly using minimap2 (v2.10, see "Life Sciences Reporting Summary") on the "-a map-ont" setting. We found that incorporation of BrdU into nanopore reads disrupts the accuracy of Albacore basecalling (data not shown) so we designated the true sequence of the read to be the subsequence of the reference that the read aligned to.

Signal events were aligned to positions on the Albacore basecall using an adaptive banded alignment[44]. This allowed us to use our trained BrdU pore model in the alignment to account for the presence of BrdU in the sequence while also circumventing the high space and time complexities that can result from dynamic programming-based alignment approaches. With an alignment of events to the Albacore basecall, we then aligned the events to positions on the reference using the minimap2 alignment. We used this alignment to find the signal events that corresponded to each position on the subsequence of the genome that the read mapped to. We only attempted to make a BrdU call at 6-mers for which the KL-divergence between the BrdU distribution and the ONT thymidine-only distribution was

603 greater than two. For each of these 6-mers in the aligned reference subsequence, we
604 computed the log-likelihood that this 6-mer contained at least one BrdU by building a hidden
605 Markov model (HMM) for the subsequence consisting of the 6-mer of interest and the
606 surrounding 20 bases (**Supplementary Fig. S6**). Each match state for this surrounding
607 sequence was given the distribution from the ONT pore model corresponding to the 6-mer at
608 that position. We used the forward algorithm to calculate the probability of the events aligned
609 to this 41-mer when the match state at the central position was set to the ONT model
610 distribution (thymidine only) and again when the match state at the central position was set
611 to our trained BrdU distribution. Taking the log-ratio of these two probabilities specifies the
612 log-likelihood of BrdU at this position. We considered positions where the log likelihood of
613 BrdU exceeded 2.5 to be BrdU calls.

614 *Region calling and fork direction*

615 Using the detection output, for non-overlapping windows of approximately 2 kb in length, we
616 computed both the number of BrdU calls (*k*) and the total number of sites where a call (BrdU
617 or thymidine) was made (*n*). From the ROC curve analysis (see **Fig. 2f**) and the mass
618 spectrometry results, we can compute the probability of making a BrdU call for one of the 6-
619 mers in our trained BrdU pore model:
620
621 p = true positive probability × fraction of thymidine substituted for BrdU.
622
623 A binomial distribution with parameters *n* and *p* gives a model for the expected frequency of
624 BrdU calls if the window actually is BrdU-containing. We computed the z-score of making *k*
625 positive calls from this binomial distribution: positive z-scores indicate that there is a high

frequency of BrdU calls in the window while negative z-scores indicate that the frequency of BrdU calls is lower than expected if the window was BrdU-positive. It is possible that some regions of the genome may have fewer 6-mers from our BrdU model than expected (for example, GC-rich regions) so we required the window to include at least 65 attempted calls. If a window would have included fewer calls, the window was extended beyond 2 kb until it included 65 attempted calls, though this was seldom necessary (mean window length 2035 bp; s.d. 216 bp). We considered a window to be a region of BrdU incorporation if the z-score was greater than -2 (see **Fig. 3b**). Fork direction was determined by smoothing the z-scores across a read with a 10 kb moving average filter and computing the central derivative of the z-score for each window. Windows that had a z-score greater than -2 (called as a BrdU window) and had a negative z-score derivative were classified as rightward moving fork windows, and windows that were called as BrdU and had a positive derivative were classified as leftward moving fork windows. Positions that had at a leftward moving fork region of at least 4 kb to the left and a rightward moving fork region of at least 4 kb to the right were called as replication initiation sites.

We determined the number of reads with BrdU positive windows from a substrate prepared from cells grown in the absence of BrdU. Of the 1100 reads assessed only five had a BrdU positive window. In each of these five reads only a single window was called as BrdU positive.

Estimations for nanopore sequencing coverage required for large genomes

From the results in our manuscript, origins *ORI1622* and *ORI1623* (efficiencies of 67% and 90%) were detected 15 and 17 times, respectively (Fig 4c). Human replication origins tend to activate much more stochastically with approximately 30% of all origins initiating in any given

S phase[45-48]. Combined with the average interorigin distance of ~31 kb[45], the median replicon size in human cells (100 kb[45, 49-51]) is similar to that of yeast cells (75 kb [52]). In a human dataset of comparable coverage to Fig 4c, an origin will be detected 4 times on average. The difference in genome size between yeast and human cells (250 fold) will require ~3.5 Tb of data to be acquired for this coverage. Such an amount of data can be collected with a single run on the 'big brother' of the MinION - the PromethION (up to 15 Tb of data, according to Oxford Nanopore Technologies).

Alternatively, a particular genomic region can be enriched prior to nanopore sequencing. For example, enrichment of a 1.25 Mb region (e.g. the human Igh locus), followed by a typical MinION sequencing run with ~15 Gb data yield will provide 5,000x coverage of nascent strands, allowing detection of origins with firing probabilities below 1% (of note, data yields have multiplied by >3 times in the last 12 months and are likely to further increase in the near future). Methods for target enrichment include restriction enzyme digests coupled with gel extraction of anticipated fragments[53] or CATCH, a recently established Cas9-dependent approach that has been used in combination with nanopore sequencing[54, 55] (commercialised by Sage Science).

Code availability

The D-NAscent software is available at https://github.com/MBoemo/DNAscent.git.

Data availability

Raw and processed Illumina and MinION data are available from NCBI GEO under accession number GSE121941.

Additional references

43. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018).
44. Suzuki, H. & Kasahara, M. Acceleration Of Nucleotide Semi-Global Alignment With Adaptive Banded Dynamic Programming. *bioRxiv* (2017).
45. Moreno, A. et al. Unreplicated DNA remaining from unperturbed S phases passes through mitosis for resolution in daughter cells. *Proc Natl Acad Sci U S A* **113**, E5757-5764 (2016).
46. Ge, X.Q., Jackson, D.A. & Blow, J.J. Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress. *Genes Dev* **21**, 3331-3341 (2007).
47. Ibarra, A., Schwob, E. & Mendez, J. Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication. *Proc Natl Acad Sci U S A* **105**, 8956-8961 (2008).
48. Ge, X.Q. & Blow, J.J. Chk1 inhibits replication factory activation but allows dormant origin firing in existing factories. *J Cell Biol* **191**, 1285-1297 (2010).
49. Besnard, E. et al. Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nature structural & molecular biology* **19**, 837-844 (2012).
50. Conti, C. et al. Replication fork velocities at adjacent replication origins are coordinately modified during DNA replication in human cells. *Mol Biol Cell* **18**, 3059-3067 (2007).
51. Jackson, D.A. & Pombo, A. Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. *J Cell Biol* **140**, 1285-1295 (1998).
52. Müller, C.A. & Nieduszynski, C.A. DNA replication timing influences gene expression level. *J Cell Biol* **216**, 1907-1914 (2017).

53. Demczuk, A. et al. Regulation of DNA replication within the immunoglobulin heavy-chain locus during B cell commitment. *PLoS Biol* **10**, e1001360 (2012).
54. Gabrieli, T. et al. Selective nanopore sequencing of human BRCA1 by Cas9-assisted targeting of chromosome segments (CATCH). *Nucleic Acids Res* **46**, e87 (2018).
55. Jiang, W. et al. Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat Commun* **6**, 8101 (2015).

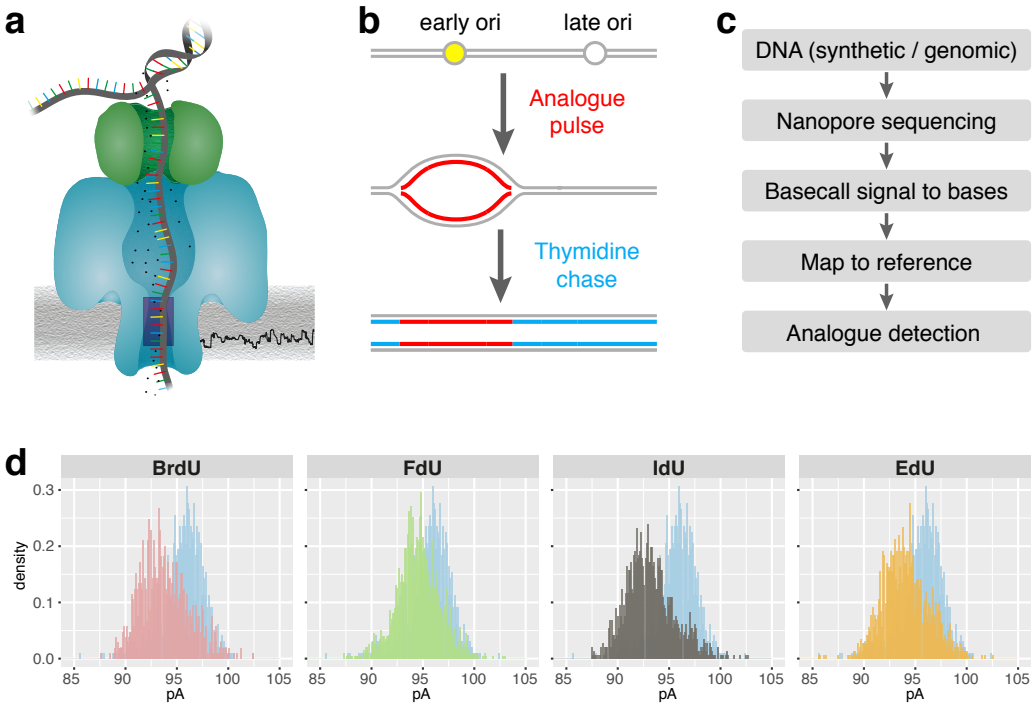

Figure 1: Nanopore sequencing can distinguish thymidine from analogues. (**a**) Graphic representation of DNA sequening using a MinION. A processive enzyme (green) ratchets DNA into the pore (blue), causing a change in ionic current (ions shown as black dots) that is determined by the 6-mer in the central chanel (purple box). The current is recorded over time (black squiggle, bottom right). (**b**) Schematic representation of pulse-labeling early replicating regions with thymidine analogues. (**c**) Outline of the experimental strategy for BrdU detection by nanopore sequencing. (**d**) For example 6-mer GCCTGA, each panel shows the distribution of signal events for thymidine (blue) and various analogues: BrdU (red); FdU (green); IdU (black); and EdU (yellow).

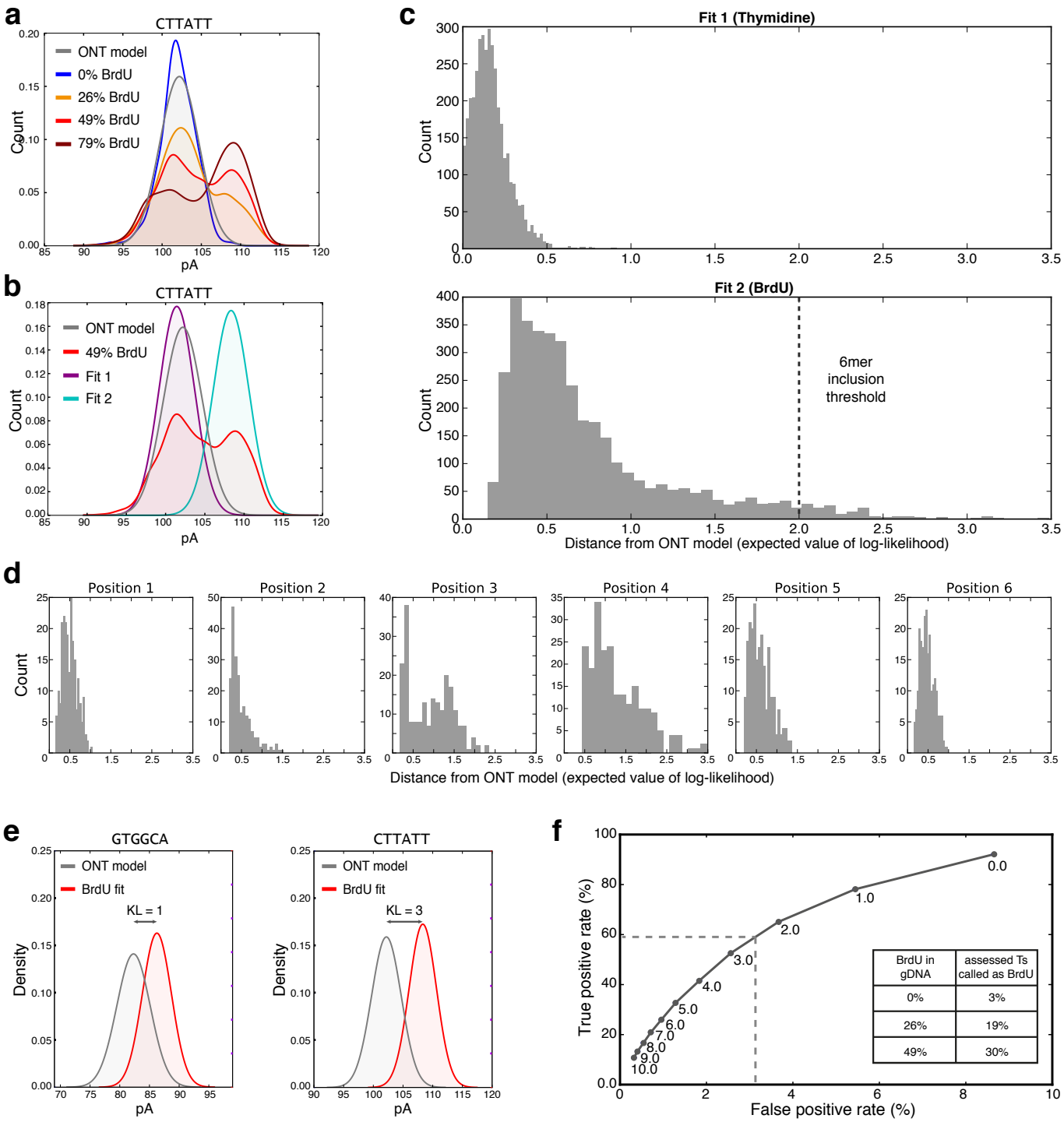

Figure 2: BrdU can be distinguished from thymidine in genomic DNA. (**a**) Signal event distributions for an example 6-mer from yeast genomic DNA containing various concentrations of BrdU (0% - blue; 26% - orange; 49% - red; 79% - crimson) compared to the ONT model (grey). (**b**) Bimodal Gaussian mixture model fit (purple and turquoise) for an example 6-mer from genomic DNA containing 49% BrdU (red). The ONT model is shown in grey. (**c**) Distribution of the KL-divergence between the ONT model and Gaussian fit 1 (upper) or fit 2 (lower) for all thymidine-containing 6-mers. (For detection (Fig. 2e), we make a BrdU call for all 6-mers that have a KL-divergence >2.0; dashed line, lower plot.) (**d**) Distributions as in the lower plot from (c) but for the subset of 6-mers containing just one thymidine; plotted by the position of the thymidine. (**e**) Signal event distributions from the ONT model (thymidine; grey) and from the bimodal Gaussian mixture model fit for BrdU (red). The KL-divergence of the two 6-mers is indicated. (**f**) Receiver operating characteristic (ROC) curve, using all 6-mers that have a KL-divergence >2.0, specifying the true positive and false positive rates for various log-likelihood thresholds of BrdU compared to thymidine (see Online Methods). Numbers near points specify the log-likelihood threshold above which a position in a read is classified as BrdU. The dashed lines demarcate the true and the false positive rates at a log-likelihood threshold >2.5.

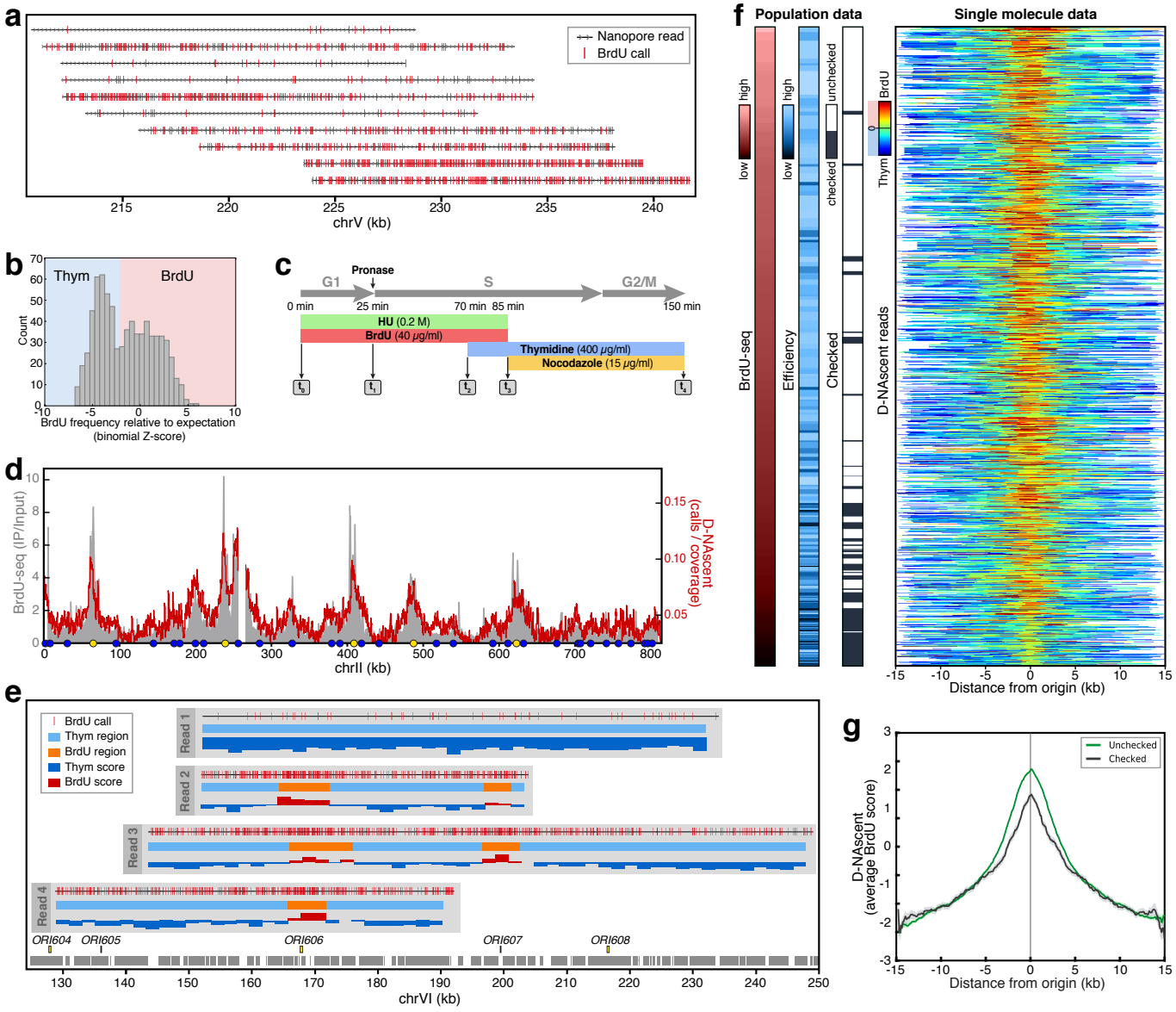

Figure 3: Single-molecule detection of BrdU on nascent DNA. (**a**) Representative nanopore reads (>15 kb) showing BrdU calls in hemi-substituted yeast genomic DNA. Red ticks indicate positive BrdU calls and arrows give the read direction relative to the sacCer3 reference genome. (**b**) The distribution of positive BrdU call frequency measured as a z-score of a binomial distribution for non-overlapping 2 kb windows. (For later analysis we set a binomial z-score threshold >-2 for assigning a window as BrdU positive.) (**c**) Schematic of the experimental strategy for detection of replication origin activity in HU. At each timepoint, samples were taken for mass spectrometry, DNA copy number measurement, BrdU-seq and D-NAscent. (**d**) Comparison of BrdU-seq and an ensemble of D-NAscent data across chromosome II (from timepoint 4). Circles denote the location of 'unchecked' (yellow) and 'checked' (blue) replication origins[17]. (**e**) Four example nanopore sequencing reads that illustrate BrdU detection on parental (read 1) and nascent strands (reads 2 - 4) mapping to the right end of chromosome VI. Each read shows BrdU calls at individual 6mers (upper track), BrdU-positive 2 kb windows (orange; middle track), and the z-score for each window where red bars are above the detection threshold (z-score ≥ -2) and are BrdU-positive (lower track). Confirmed replication origins from OriDB (yellow boxes) and genes (grey boxes) are shown. (**f**) Visualisation of D-NAscent data for 1,325 individual nanopore reads (rows) that span confirmed replication origins (OriDB), ordered by BrdU-seq data. Additional colour bars show population-level data for BrdU-seq, origin activation efficiency[4] and whether the origin is 'checked' by the intra-S phase checkpoint[17]. (**g**) Ensemble BrdU score from D-NAscent for all 'unchecked' (green) and 'checked' (black) origins (BrdU z-scores averaged for each column in (f); shaded areas show the standard error of the mean).

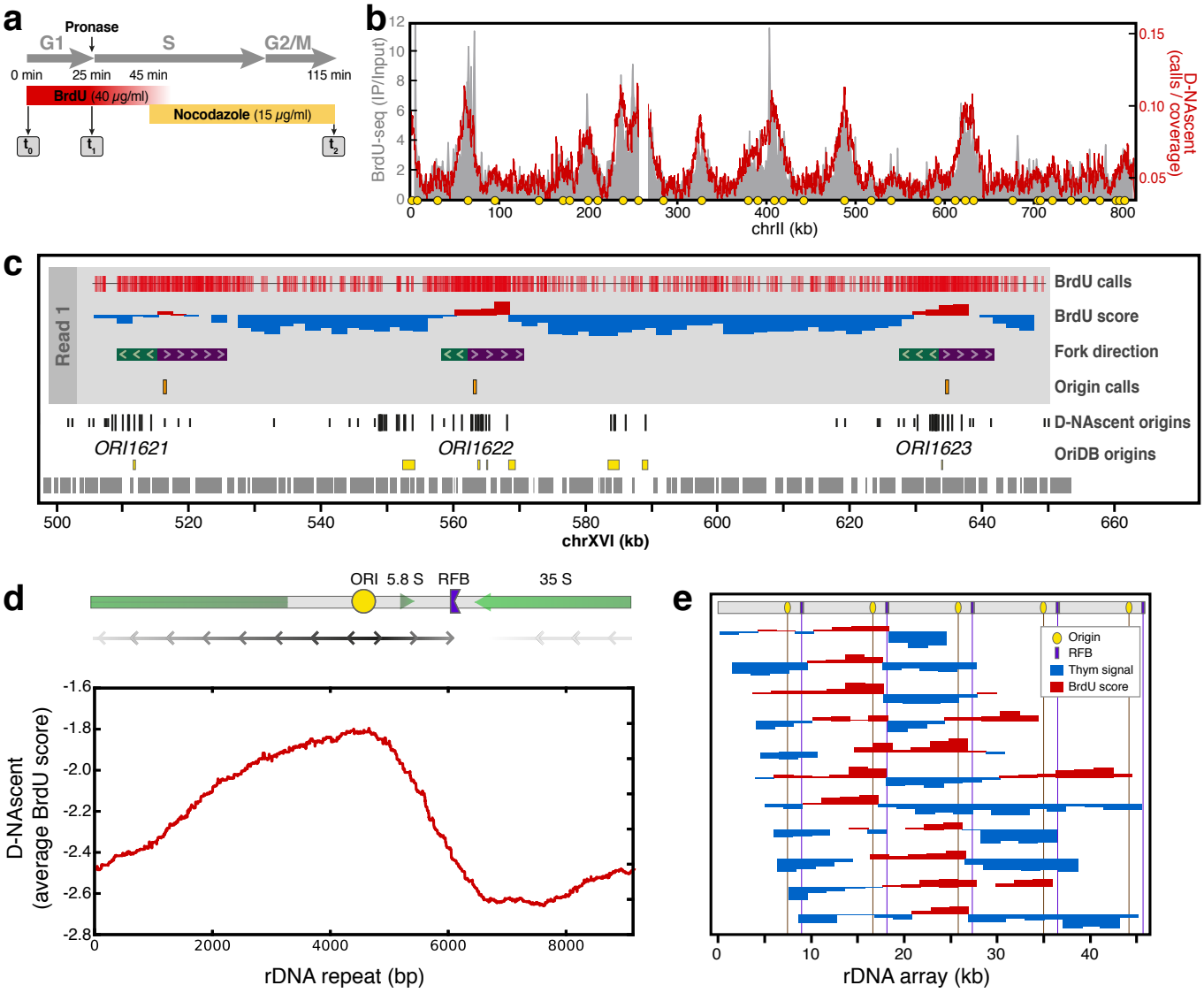

Figure 4: Single-molecule detection of replication dynamics. (**a**) Schematic of the experimental strategy for detection of replication dynamics by D-NAscent. At the indicated timepoints, samples were taken for mass spectrometry, DNA copy number measurements, BrdU-seq and D-NAscent. (**b**) Comparison BrdU-seq data and an ensemble of D-NAscent data across chromosome II (from timepoint t_2). Origins annotated as confirmed in OriDB are shown (yellow circles). (**c**) An example 150 kb nanopore sequencing read showing BrdU calls at individual 6-mers (top track), the z-score for each 2 kb window where BrdU-positive window scores are shown in red and thymidine-only window scores are shown in blue (middle track), and called fork direction and replication initiation sites (lower tracks). Origin calls from all spanning nanopore reads (black bars: tall, close to known origins; short, >3.9 kb (Supplemental Fig. S10) from known origins) and origins annotated as confirmed or likely by OriDB (yellow boxes) are displayed. (**d**) (top) A schematic representation of a single rDNA repeat showing the origin, replication fork barrier (RFB), predominant replication fork direction (line arrows) and the major transcripts (open arrows). (bottom) An ensemble of D-NAscent z-scores averaged over all nanopore sequence reads that spanned an rDNA repeat and had at least one BrdU-positive 2 kb window. (**e**) The D-NAscent BrdU scores from selected molecules aligned to multiple rDNA repeats (origin, yellow; RFB, purple).