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, South Parks Road, Oxford, OX1 3RE, UK; [2]Ludwig Institute for Cancer Research, Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7DQ, UK; [3]Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, OX3 7FZ, 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 now 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.

Introduction

Genomic methods have provided insights into DNA replication and genome stability[1-3]. Within a population of cells, these methods mask heterogeneity in both replication origin usage and replication fork dynamics; what happens in each individual cell is difficult to ascertain[4]. A high-throughput single-molecule approach is needed to reveal the heterogeneity in DNA replication dynamics. In addition, such an approach has the potential to identify origins used in organisms with very high levels of heterogeneity, in particular mammalian cells, for which population analysis is less useful.

Current single-molecule techniques to study DNA replication have provided valuable insight, but have limitations. 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: combed molecules are anonymous unless genomic positions are identified by probe hybridization, which is particularly challenging for large metazoan genomes. Alternative methods use nanochannels to stretch DNA molecules, which has led to increases in 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]. Recently, *in vitro* systems have been established that use single-molecule imaging to monitor replication protein kinetics on DNA[8, 9] . Visualizing individual, fluorescently-tagged proteins provided novel mechanistic insights into replication origin licensing and initiation. However, *in vitro* systems are presently limited to small DNA molecules (replicated from a single origin) and so cannot recapitulate *in vivo* replication dynamics.

Here, 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 presently available nanopore sequencing platforms can reliably distinguish base analogues from natural bases. We have 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. We demonstrated the power of D-NAscent in *S. cerevisiae* (the eukaryote in which genome replication is best characterized). A pulse-chase experiment revealed the 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 detect the changes in BrdU incorporation frequency to reveal the direction of replication forks and identify the location of replication origins on individual molecules. Using this approach, we have created a whole-genome profile that reveals the 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 characteristic disruption to the ionic current signal. Each short DNA sequence within the nanopore can be identified by the magnitude of the characteristic signal it produces. To simplify analysis, it is typically assumed that the observed signal only depends on a short fixed-length sequence, which is termed a *k*-mer. The current signal for each *k*-mer can be modelled by a Gaussian distribution. For consistency with the data released by ONT, we used a *k*-mer length of six. We and others have previously demonstrated that signal-level events can distinguish methylated from unmodified bases[10, 11]. 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 DNA 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 at a sequencing rate of 250bp/s; discontinued 2016) and the presently available chemistry (R9.5 pore at 450bp/s; **Fig. 1d** and **1e**). (Similar observations were made with the previously available generation of the pore, R7.3 – data not shown.) The shift in signal depended on the particular analogue, the sequence context, and the position of the analogue within the 6-mer; the greatest shift was observed when the analogue was substituted for thymidine at the fourth base from the 5' end of the 6-mer (**Fig.**

97 **1d-e** and **Supplemental Fig. S1**). These observations indicated that MinION sequencing has
98 the potential to detect nucleotide analogues in genomic DNA.

99 *Identifying the characteristic nanopore signal of BrdU in genomic DNA*

100 BrdU is a commonly used thymidine analogue with limited cytotoxicity compared to EdU or
101 FdU[12-15]. Thus, we sought to determine the distribution of nanopore signal events for any
102 BrdU-containing 6-mer in genomic DNA. A *Saccharomyces cerevisiae* strain, which is
103 dependent upon exogenous thymidine[16] was grown in various proportions of thymidine and
104 BrdU (media with 100 µM (BrdU + dT) consisting of 0, 40, 60, 80 or 100% BrdU; see Online
105 Methods). Genomic DNA was prepared and analysed by MinION sequencing. As a control,
106 BrdU incorporation was quantified by mass spectrometry and immunoprecipitation followed
107 by Illumina sequencing (BrdU-seq; **Supplemental Fig. S2**, **S3** and summarised in **Supplemental**
108 **Table S1**)[17]. The mass spectrometry data revealed that in our five genomic DNA samples the
109 percentage of thymidines substituted by BrdU was 0%, 15%, 26%, 49% and 79%, respectively.

110 Nanopore sequencing signal events were aligned to the genomic reference using
111 nanopolish[18]. These data revealed many thymidine-containing 6-mers where the distribution
112 of signal events was bimodal; while one population matched the ONT model, there was a
113 distinct second population (**Fig. 2a**). The relative proportions of the two populations reflected
114 the concentration of incorporated BrdU. By contrast, 6-mers that did not contain thymidine
115 were mono-modal and matched the ONT model (data not shown). We fit a bimodal Gaussian
116 mixture model to the signal events from the 49% BrdU sample that aligned to each thymidine-
117 containing 6-mer (**Fig. 2b;** see **Supplemental Fig. S11** for results using other incorporation
118 rates). We used the Kullback-Leibler (KL) divergence, which measures the average log-
119 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 our previous observation that BrdU predominantly shifts the signal event when present at the fourth base from the 5' end of the 6-mer (**Fig. 1b** and **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 indicated that it will be possible 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. S4a**). Each time one of these 6-mers occurred in our sequencing reads, we used a Hidden Markov model (HMM) to compute the log-likelihood that the 6-mer contained a BrdU (**Supplemental Fig. S5**). A position in a read was classified as BrdU if the log-likelihood exceeded a threshold (**Fig. 2e**); 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

143 rates (**Fig. 2f** and **Supplemental Fig. S4b**). We set this threshold at log-likelihood >2.5 which
144 gave a true positive rate of ~60% for a false positive rate of ~3%. We achieved a similar true
145 positive rate using a DNA sample with an intermediary BrdU concentration (26%
146 incorporation) that was unrelated to the training material.

147 To further test our detection strategy, we generated hemi-BrdU substituted yeast
148 genomic DNA, by synchronizing a strain prototrophic for thymidine[19] and passing it through
149 one S phase in media containing a high concentration (400 μg/ml) of BrdU. Material was
150 validated by mass spectrometry (**Supplemental Fig. S2**) and BrdU-seq to reveal any
151 incorporation bias (**Supplemental Fig. S6**). The cell cycle synchrony was confirmed by flow
152 cytometry of DNA content (**Supplemental Fig. S7**). MinION sequencing was performed and
153 positions of BrdU incorporation were determined as described above. As anticipated, we
154 observed reads with either low or predominantly high density of BrdU calls over the entire
155 read, consistent with parental and nascent strands, respectively (**Fig. 3a**). To quantify the
156 frequency of BrdU calls in each read, we fit the number of positive BrdU calls in non-
157 overlapping 2 kb windows to a binomial distribution (see Online Methods). This allowed us to
158 compare the number of positive BrdU calls in each window to what we would expect if the
159 window was BrdU-positive (determined from mass spectrometry data and the true-positive
160 rate). Computing the z-score of positive BrdU calls in each window against this binomial
161 distribution revealed a bimodal density: One population was centered around the mean,
162 indicating these windows had a BrdU frequency consistent with our expectation for a BrdU-
163 positive window; the other population was centered ~5 s.d. below the mean, indicating these
164 windows had significantly fewer positive BrdU calls than we expect if the window was BrdU-
165 positive (**Fig. 3b**). We conclude that these two populations correspond to BrdU-containing
166 windows and thymidine-only windows, respectively. These results indicated that our model

167 can distinguish parental DNA from nascent DNA containing BrdU. We call our method
168 Detecting Nucleotide Analogue signal currents on extremely long nanopore traces (D-
169 NAscent).

170 *Single-molecule detection of replication origin activity in hydroxyurea*

171 DNA synthesis can be slowed by the addition of hydroxyurea (HU), which inhibits
172 ribonucleotide reductase, thereby restricting BrdU incorporation to locations proximal to
173 early activating replication origins[20-22]. To investigate the pattern of origin usage on single-
174 molecules, we released cells synchronously from G1 into S phase in the presence of HU and
175 BrdU (40 μg/ml). After 60 minutes, cells were chased out of HU, with excess thymidine (400
176 μg/ml), into nocodazole (to prevent entry into a second cell cycle; **Fig. 3c**). After completion
177 of S phase, samples were collected for D-NAscent, mass spectrometry, and BrdU-seq. Cell
178 cycle synchrony was assessed by flow cytometry of DNA content (**Supplemental Fig. S7**). BrdU
179 detected in the MinION data was summed across all reads in non-overlapping 100 bp windows
180 to allow comparison to the BrdU-seq data (**Fig. 3d**). This confirmed that an ensemble of our
181 single-molecule data was in good agreement (Pearson correlation coefficient, R=0.76) with
182 established short-read methods. Visualising the individual positive BrdU calls on single-
183 molecules suggested that each sequencing read fell into one of two categories: there was
184 either an infrequent number of positive BrdU calls throughout the whole read, or the read
185 contained short patches of frequent positive BrdU calls (**Fig. 3e**). We concluded that these
186 reads likely correspond to parental and nascent strands, respectively. For each individual read,
187 we assessed non-overlapping 2 kb windows and quantified the frequency of positive BrdU
188 calls by computing the z-score of positive calls against a binomial distribution as before. This
189 allowed windows to be assigned as having either high or low BrdU z-scores (**Fig. 3e**). For

example, two early firing replication origins on chromosome VI both gave rise to BrdU positive regions on a single >100 kb read (read 3 in **Fig. 3e**), indicating that both origins activated in a single cell.

To explore this more widely, we visualised the BrdU frequency z-scores for individual nascent-strand reads that spanned known replication origins (each read considered covered >4kb either side of the origin location)[23]. In **Fig. 3f**, each row represents an individual nanopore sequencing read centred upon an origin. 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 will be near the top. In the majority of reads, we observed the highest frequency of positive BrdU calls at the origin. Most of these nascent molecules spanned efficient, early activating origins ('unchecked' by the intra-S phase checkpoint) indicative of origin firing during the BrdU pulse; there were only occasional examples of molecules where we detected BrdU incorporation at less efficient, late activating origins ('checked' by the intra-S phase checkpoint). Computing the average for each column, the BrdU z-score for unchecked (or checked) origins showed that on average early activating origins had incorporated more BrdU than later activating origins; in both cases the signal was symmetric and centred on the origins (**Fig. 3g**). However, some individual molecules showed asymmetric levels of BrdU incorporation relative to the origin, indicative of different rates of sister fork progression (**Supplemental Fig. S8**). We note that the BrdU z-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, the frequency of positive BrdU calls in a window drops below our detection threshold (z-score = -2, see **Fig. 3b**) and is called as a thymidine window. This implied a time-dependent drop in BrdU incorporation as the fork progressed

away from the origin. 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 replication stress, we synchronised thymidine prototroph cells in G1 and released in the presence of BrdU (40 μg/ml) (**Fig. 4a** and **Supplemental Fig. S7**). Samples were collected for D-NAscent, mass spectrometry, BrdU-seq, and DNA copy number measurements[24]. Using the D-NAscent results, we again summed all positive BrdU calls across all reads in non-overlapping 100 bp windows and found good agreement (Pearson correlation coefficient, R=0.75) with population-level BrdU-seq data (**Fig. 4b**); both assessments confirmed that the population average level of BrdU incorporation was inversely correlated with average replication time (**Supplemental Fig. S9**). This is consistent with the ratio of BrdU to thymidine falling (effective BrdU concentration) as cells progress through S phase; this is likely to be a consequence of the rate of BrdU import being lower than incorporation and the activation of endogenous thymidine synthesis pathways[25, 26]. Within individual reads, we assessed non-overlapping 2 kb windows and computed the z-score of positive BrdU calls in each window to a binomial distribution (examples shown in **Fig. 4c**). We observed clear peaks of BrdU incorporation at locations near known replication origins. The extremely long nanopore sequence reads allowed the identification of multiple active origins on single molecules. The BrdU z-score either side of each origin declined, indicative of the progression of bi-directional replication forks at a time when the effective concentration of cellular BrdU is falling. As forks move further away from initiation sites, the frequency of positive BrdU calls drops below our detection threshold (red to blue in **Fig. 4c**).

We determined the gradient of BrdU z-scores as a proxy for replication fork direction across all nascent strand sequence reads (examples shown in **Fig. 4c**). We used regions of diverging replication forks to call sites where replication initiated early in S phase prior to the BrdU concentration dropping below our detection threshold. This provided a whole genome map of DNA replication origin activity on single molecules. Examining all identified replication initiation sites revealed two distinct classes (**Supplemental Fig. S10**). The first class were found on multiple independent sequence reads, consistent with high efficiency origins used in many cells. These sites corresponded to replication origins identified in population level analyses[23]. 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, several lines of evidence argued in favour of these sites representing genuine bi-directional replication origins. First, DNA repair synthesis prior to S phase would be confined to parental strands and synthesis would be unidirectional, resulting in a unilateral BrdU gradient; the novel initiation sites we identified were present on nascent strands and showed bi-directional synthesis evidenced by bilateral BrdU-gradients (example shown in **Fig. 4c**). Second, the BrdU call frequency and signal detected at novel initiation sites closely resembled that observed at previously reported origins. Third, applying more conservative criteria for origin identification (a higher z-score and a longer contiguous region of BrdU-positive 2 kb windows) reduced the number of origin calls, but did not diminish the proportion of initiation events at novel locations (data not shown). Therefore, the single-molecule resolution of 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

Detection of BrdU incorporation differences across nanopore sequencing reads allowed us to infer replication initiation sites and fork direction. Next, we sought to determine whether this could allow the detection of replication fork pausing events. The yeast ribosomal DNA (rDNA) repeats each contain a replication origin and a programmed unidirectional replication fork barrier (RFB) that pauses one of the sister forks (**Fig. 4d**)[27]. The repetitive nature of the rDNA limits their analysis by short-read technologies, but we were able to sequence thousands of molecules that each spanned multiple repeats. An ensemble of D-NAscent data analysed across a single rDNA repeat clearly demonstrated an asymmetric peak in the BrdU average z-score signal (**Fig. 4d**). In this ensemble analysis, the population average BrdU z-score was maximal at the replication origin. The dramatic fall in z-score to the right of the origin indicated a substantial delay to the progress of the rightward moving fork. This delay was positioned over the RFB and is consistent with unidirectional fork pausing. By contrast, the leftward moving fork showed no such delay. Analysis of single molecules (**Fig. 4e**) demonstrated firing of the origin in a subset of repeats and pausing of rightward moving forks at the RFB. Therefore, time-dependent reductions in BrdU incorporation frequency allowed comprehensive analysis of replication dynamics on single molecules, revealing fork direction, initiation sites, termination sites and fork pausing/stalling.

Discussion

We have developed a genomic single-molecule method for the detection of base analogues that we term D-NAscent. Base analogues are widely used in modern biology for the study of chromosome biology[28], cell proliferation[29] and gene expression[30]. Therefore, D-NAscent offers a powerful method for the advancement of each of these fields. Key features of nanopore

sequencing[31] make D-NAscent possible: the lack of an obligatory amplification step ensures that *in vivo* incorporated analogues are present in the sequenced strand; the interrogation of single nucleic acid strands permits direct detection of the analogue and provides single molecule information; and the extremely long sequence read lengths (>100 kb) allow detection of long-range *cis* interactions. We demonstrated that the presently available ONT MinION nanopore sequencing platform gives robust detection of thymidine analogues across the full range of sequence contexts (**Fig. 1** and **Fig. 2**). This allowed us to develop a model for the detection of *in vivo* incorporated BrdU that we validated against mass spectrometry and population level BrdU-seq data. Using this approach should also allow for the detection of additional 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 that replication fork velocity is ~2 kb/ min[32, 33] and that we observed differences in BrdU incorporation at 2 kb resolution, this indicated that D-NAscent has the potential to provide a temporal resolution of close to 1 min.

DNA replication is a stochastic process and many aspects, including replication origin activity, are masked in population-based approaches. Historically, this has required the use of complex, low-throughput, and low-resolution methodologies to visualize DNA replication on single-molecules. By applying D-NAscent to the study of yeast chromosome biology, we have generated the first whole genome map of DNA replication dynamics at the 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 occured at sites dispersed throughout the genome. Yeast replication origins were first characterized by their

306 ability to support plasmid replication (as autonomously replicating sequences, called ARS
307 elements)[34] and it was subsequently shown that the same sequences can support replication
308 initiation at their endogenous chromosomal locations[35]. Neither the plasmid nor the
309 chromosomal assays have the sensitivity to detect very low efficiency origins. Recent *in vitro*
310 studies have demonstrated origin-dependent and independent DNA replication initiation, due
311 to promiscuity in the binding of the origin recognition complex[36, 37]. Although the *in vitro*
312 origin-independent replication initiation was only observed in the absence of physiological
313 levels of competitor DNA, it is consistent with our finding that many genomic locations can
314 function at low frequency as a replication origin. Thus, we propose that replication of the
315 yeast genome is initiated from both well-defined, high-efficiency origins and a broadly
316 distributed set of very low-efficiency origins, similar to the configuration observed in
317 mammalian cells[2].

318 The D-NAscent single-molecule methodology will allow many unresolved questions in
319 chromosome biology to be addressed. For example, the single-molecule nature will allow the
320 identification of *cis* regulatory mechanisms. The power to explore *cis* regulatory mechanisms
321 is enhanced by the extremely long sequencing reads; in this study we present reads <160 kb,
322 but others have reported ultra-long reads of >1 Mb[38]. As such, D-NAscent complements
323 recently developed single cell approaches for the study of DNA replication[39, 40]. Single cell
324 approaches have relatively low spatial resolution, but they can provide *trans* information
325 missing in single molecules. However, we and others have discovered that replication origin
326 activity is generally regulated in *cis*[41-44] emphasising the importance of the single molecule
327 approach. Second, variants of the MinION sequencing method allow capture of sequence
328 information from both DNA strands ($1D^2$). Combining D-NAscent with $1D^2$ sequencing has the
329 potential to reveal sites of conservative DNA replication, for example during recombination-

dependent DNA synthesis[45, 46]. Third, extremely long sequencing reads allow D-NAscent to examine patterns of DNA replication in complex genomic locations (e.g. non-unique or repetitive sequences; **Fig. 4d** and **4e**) that are abundant in mammalian genomes and generally understudied. Fourth, the ability of D-NAscent to detect nascent DNA depends on the 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 the application of D-NAscent to many organisms, including the study of large, stochastically replicated mammalian genomes. Existing single-molecule methods, such as DNA combing, have revealed extensive variability in replication initiation site usage and in fork progression rates. However, combed molecules are generally anonymous precluding the identification of chromatin features associated with variable fork velocity or replication initiation. Applying D-NAscent to hemi-labelled human genomic DNA allowed us to discriminate nascent and parental strands based on BrdU detection frequency (personal communication, R Wilson and J Carrington). In future work, we plan to generate D-NAscent data across the human genome using the PromethION platform (up to 15 Tb data output, with 3.5 Tb required for a coverage comparable to our yeast data, see Online Methods) or using MinION for very high coverage of specific, CATCH-enriched genomic regions[47, 48]. This will allow the genome-wide identification of mammalian replication origins by D-NAscent.

Acknowledgments

We are grateful to John Diffley and Etienne Schwob 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 and Wellcome Trust Investigator Award 110064/Z/15/Z. 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. Ticau, S., Friedman, L.J., Ivica, N.A., Gelles, J. & Bell, S.P. Single-molecule studies of origin licensing reveal mechanisms ensuring bidirectional helicase loading. *Cell* **161**, 513-525 (2015).
9. Duzdevich, D. et al. The dynamics of eukaryotic replication initiation: origin specificity, licensing, and firing at the single-molecule level. *Mol Cell* **58**, 483-494 (2015).
10. Simpson, J.T. et al. Detecting DNA cytosine methylation using nanopore sequencing. *Nat Methods* **14**, 407-410 (2017).
11. Rand, A.C. et al. Mapping DNA methylation with high-throughput nanopore sequencing. *Nat Methods* **14**, 411-413 (2017).
12. Hagenkort, A. et al. dUTPase inhibition augments replication defects of 5-Fluorouracil. *Oncotarget* **8**, 23713-23726 (2017).
13. Hua, H. & Kearsey, S.E. Monitoring DNA replication in fission yeast by incorporation of 5-ethynyl-2'-deoxyuridine. *Nucleic Acids Res* **39**, e60 (2011).
14. 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).

15. 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).
16. Vernis, L., Piskur, J. & Diffley, J.F. Reconstitution of an efficient thymidine salvage pathway in Saccharomyces cerevisiae. *Nucleic Acids Res* **31**, e120 (2003).
17. 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).
18. 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).
19. 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).
20. Alvino, G.M. et al. Replication in hydroxyurea: it's a matter of time. *Mol Cell Biol* **27**, 6396-6406 (2007).
21. 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).
22. Poli, J. et al. dNTP pools determine fork progression and origin usage under replication stress. *EMBO J.* **31**, 883-894 (2012).
23. 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).
24. Müller, C.A. et al. The dynamics of genome replication using deep sequencing. *Nucleic Acids Res* **42**, e3 (2014).

25. Vernis, L., Piskur, J. & Diffley, J.F. Reconstitution of an efficient thymidine salvage pathway in Saccharomyces cerevisiae. *Nucleic Acids Res* **31**, e120 (2003).
26. 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).
27. 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).
28. Cavanagh, B.L., Walker, T., Norazit, A. & Meedeniya, A.C. Thymidine analogues for tracking DNA synthesis. *Molecules* **16**, 7980-7993 (2011).
29. 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).
30. Dolken, L. et al. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* **14**, 1959-1972 (2008).
31. 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).
32. Hawkins, M. et al. High-resolution replication profiles define the stochastic nature of genome replication initiation and termination. *Cell reports* **5**, 1132-1141 (2013).
33. Sekedat, M.D. et al. GINS motion reveals replication fork progression is remarkably uniform throughout the yeast genome. *Mol Syst Biol* **6**, 353 (2010).
34. Stinchcomb, D.T., Struhl, K. & Davis, R.W. Isolation and characterisation of a yeast chromosomal replicator. *Nature* **282**, 39-43 (1979).
35. Huberman, J.A., Zhu, J.G., Davis, L.R. & Newlon, C.S. Close association of a DNA replication origin and an ARS element on chromosome III of the yeast, Saccharomyces cerevisiae. *Nucleic Acids Res* **16**, 6373-6384 (1988).

450 36. On, K.F. et al. Prereplicative complexes assembled in vitro support origin-dependent
451 and independent DNA replication. *EMBO J.* **33**, 605-620 (2014).
452 37. Gros, J., Devbhandari, S. & Remus, D. Origin plasticity during budding yeast DNA
453 replication in vitro. *EMBO J.* **33**, 621-636 (2014).
454 38. Payne, A., Holmes, N., Rakyan, V. & Loose, M. BulkVis: a graphical viewer for Oxford
455 nanopore bulk FAST5 files. *Bioinformatics* (2018).
456 39. Chen, C. et al. Single-cell whole-genome analyses by Linear Amplification via
457 Transposon Insertion (LIANTI). *Science* **356**, 189-194 (2017).
458 40. Dileep, V. & Gilbert, D.M. Single-cell replication profiling to measure stochastic
459 variation in mammalian replication timing. *Nat Commun* **9**, 427 (2018).
460 41. Natsume, T. et al. Kinetochores coordinate pericentromeric cohesion and early DNA
461 replication by Cdc7-Dbf4 kinase recruitment. *Mol Cell* **50**, 661-674 (2013).
462 42. Fang, D. et al. Dbf4 recruitment by forkhead transcription factors defines an upstream
463 rate-limiting step in determining origin firing timing. *Genes Dev* **31**, 2405-2415 (2017).
464 43. Hiraga, S.I. et al. Human RIF1 and protein phosphatase 1 stimulate DNA replication
465 origin licensing but suppress origin activation. *EMBO Rep* **18**, 403-419 (2017).
466 44. Foti, R. et al. Nuclear Architecture Organized by Rif1 Underpins the Replication-
467 Timing Program. *Mol Cell* **61**, 260-273 (2016).
468 45. Donnianni, R.A. & Symington, L.S. Break-induced replication occurs by conservative
469 DNA synthesis. *Proc Natl Acad Sci U S A* **110**, 13475-13480 (2013).
470 46. Saini, N. et al. Migrating bubble during break-induced replication drives conservative
471 DNA synthesis. *Nature* **502**, 389-392 (2013).
472 47. Gabrieli, T. et al. Selective nanopore sequencing of human BRCA1 by Cas9-assisted
473 targeting of chromosome segments (CATCH). *Nucleic Acids Res* **46**, e87 (2018).
474 48. Jiang, W. et al. Cas9-Assisted Targeting of CHromosome segments CATCH enables
475 one-step targeted cloning of large gene clusters. *Nat Commun* **6**, 8101 (2015).

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**)[19]. 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 population data

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) 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. 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) provided by ONT (see **Supplemental Fig. S11** for results using other incorporation rates). We aligned the reads to the *S. cerevisiae* sacCer3 genome assembly using minimap2 (v2.10) with the "-a map-ont" setting[49]. 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) and the reads were aligned to the *S. cerevisiae* sacCer3 genome assembly using minimap2 (v2.10) 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[50]. 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 greater than two. For each of these 6-mers in the aligned reference subsequence, we

computed the log-likelihood that this 6-mer contained at least one BrdU by building a hidden Markov model (HMM) for the subsequence consisting of the 6-mer of interest and the surrounding 20 bases (**Supplementary Fig. S5**). Each match state for this surrounding sequence was given the distribution from the ONT pore model corresponding to the 6-mer at that position. We used the forward algorithm to calculate the probability of the events aligned to this 41-mer when the match state at the central position was set to the ONT model distribution (thymidine only) and again when the match state at the central position was set to our trained BrdU distribution. Taking the log-ratio of these two probabilities specifies the log-likelihood of BrdU at this position. We considered positions where the log likelihood of BrdU exceeded 2.5 to be positive BrdU calls.

Region calling and fork direction

Using the detection output, for non-overlapping windows of approximately 2 kb in length, we computed both the number of positive BrdU calls (k) and the total number of sites where a call (BrdU or thymidine) was made (n). From the ROC curve analysis (see **Fig. 2f**) and the mass spectrometry results, we can compute the probability of making a positive BrdU call for one of the 6-mers in our trained BrdU pore model:

$$p = \text{true positive probability} \times \text{fraction of thymidine substituted for BrdU.}$$

A binomial distribution with parameters n and p gives a model for the expected frequency of positive BrdU calls if the window actually is BrdU-containing. We computed the z-score of making k positive calls from this binomial distribution: positive z-scores indicate that there is a high frequency of positive BrdU calls in the window while negative z-scores indicate that

636 the frequency of positive BrdU calls is lower than expected if the window was BrdU-positive.
637 It is possible that some regions of the genome may have fewer 6-mers from our BrdU model
638 than expected (for example, GC-rich regions) so we required the window to include at least
639 65 attempted calls. If a window would have included fewer calls, the window was extended
640 beyond 2 kb until it included 65 attempted calls, though this was seldom necessary (mean
641 window length 2035 bp; s.d. 216 bp). We considered a window to be a region of BrdU
642 incorporation if the z-score was greater than -2 (see **Fig. 3b**). Fork direction was determined
643 by smoothing the z-scores across a read with a 10 kb moving average filter and computing
644 the central derivative of the z-score for each window. Windows that had a z-score greater
645 than -2 (called as a BrdU window) and had a negative z-score derivative were classified as
646 rightward moving fork windows, and windows that were called as BrdU and had a positive
647 derivative were classified as leftward moving fork windows. Positions that had at a leftward
648 moving fork region of at least 4 kb to the left and a rightward moving fork region of at least 4
649 kb to the right were called as replication initiation sites.
650 We determined the number of reads with BrdU positive windows from a substrate
651 prepared from cells grown in the absence of BrdU. Of the 1100 reads assessed only five had
652 a BrdU positive window. In each of these five reads only a single window was called as BrdU
653 positive.

654 *Estimations for nanopore sequencing coverage required for large genomes*

655 From the results in our manuscript, origins *ORI1622* and *ORI1623* (efficiencies of 67% and
656 90%) were detected 15 and 17 times, respectively (Fig 4c). Human replication origins tend to
657 activate much more stochastically with approximately 30% of all origins initiating in any given
658 S phase[51-54]. Combined with the average interorigin distance of ~31 kb[51], the median replicon

size in human cells (100 kb[51, 55-57]) is similar to that of yeast cells (75 kb [58]). 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[59] or CATCH, a recently established Cas9-dependent approach that has been used in combination with nanopore sequencing[47, 48] (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

49. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018).

682 50. Suzuki, H. & Kasahara, M. Acceleration Of Nucleotide Semi-Global Alignment With
683 Adaptive Banded Dynamic Programming. *bioRxiv* (2017).
684 51. Moreno, A. et al. Unreplicated DNA remaining from unperturbed S phases passes
685 through mitosis for resolution in daughter cells. *Proc Natl Acad Sci U S A* **113**, E5757-
686 5764 (2016).
687 52. Ge, X.Q., Jackson, D.A. & Blow, J.J. Dormant origins licensed by excess Mcm2-7 are
688 required for human cells to survive replicative stress. *Genes Dev* **21**, 3331-3341 (2007).
689 53. Ibarra, A., Schwob, E. & Mendez, J. Excess MCM proteins protect human cells from
690 replicative stress by licensing backup origins of replication. *Proc Natl Acad Sci U S A*
691 **105**, 8956-8961 (2008).
692 54. Ge, X.Q. & Blow, J.J. Chk1 inhibits replication factory activation but allows dormant
693 origin firing in existing factories. *J Cell Biol* **191**, 1285-1297 (2010).
694 55. Besnard, E. et al. Unraveling cell type-specific and reprogrammable human replication
695 origin signatures associated with G-quadruplex consensus motifs. *Nature structural &*
696 *molecular biology* **19**, 837-844 (2012).
697 56. Conti, C. et al. Replication fork velocities at adjacent replication origins are
698 coordinately modified during DNA replication in human cells. *Mol Biol Cell* **18**, 3059-
699 3067 (2007).
700 57. Jackson, D.A. & Pombo, A. Replicon clusters are stable units of chromosome structure:
701 evidence that nuclear organization contributes to the efficient activation and
702 propagation of S phase in human cells. *J Cell Biol* **140**, 1285-1295 (1998).
703 58. Müller, C.A. & Nieduszynski, C.A. DNA replication timing influences gene expression
704 level. *J Cell Biol* **216**, 1907-1914 (2017).
705 59. Demczuk, A. et al. Regulation of DNA replication within the immunoglobulin heavy-
706 chain locus during B cell commitment. *PLoS Biol* **10**, e1001360 (2012).
707

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). The data were generated using the ONT MinION R9 and R9.5 pore with sequencing speeds of 250 bp/s (**d**) or 450 bp/s (**e**), respectively.

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-

731 divergence of the two 6-mers is indicated. (**f**) Receiver operating characteristic (ROC) curve,
732 using all 6-mers that have a KL-divergence >2.0, specifying the true positive and false positive
733 rates for various log-likelihood thresholds of BrdU compared to thymidine (see Online
734 Methods). Numbers near points specify the log-likelihood threshold above which a position
735 in a read is classified as BrdU. The dashed lines demarcate the true and the false positive rates
736 at a log-likelihood threshold >2.5.

737 **Figure 3: Single-molecule detection of BrdU on nascent DNA.** (**a**) Representative nanopore
738 reads (>15 kb) showing BrdU calls in hemi-substituted yeast genomic DNA. Red ticks indicate
739 positive BrdU calls and arrows give the read direction relative to the sacCer3 reference
740 genome. (**b**) The distribution of positive BrdU call frequency measured as a z-score of a
741 binomial distribution for non-overlapping 2 kb windows. (For later analysis we set a binomial
742 z-score threshold >-2 for assigning a window as BrdU positive.) (**c**) Schematic of the
743 experimental strategy for detection of replication origin activity in HU. At each timepoint,
744 samples were taken for mass spectrometry, DNA copy number measurement, BrdU-seq and
745 D-NAscent. (**d**) Comparison of BrdU-seq and an ensemble of D-NAscent data across
746 chromosome II (from timepoint 4). Circles denote the location of 'unchecked' (yellow) and
747 'checked' (blue) replication origins[21]. (**e**) Four example nanopore sequencing reads that
748 illustrate BrdU detection on parental (read 1) and nascent strands (reads 2 – 4) mapping to
749 the right end of chromosome VI. Each read shows BrdU calls at individual 6-mers (upper
750 track), BrdU-positive 2 kb windows (orange; middle track), and the z-score for each window
751 where red bars are above the detection threshold (z-score $\geq$ -2) and are BrdU-positive (lower
752 track). Confirmed replication origins from OriDB (yellow boxes) and genes (grey boxes) are
753 shown. (**f**) Visualisation of D-NAscent data for 1,325 individual nanopore reads (rows) that
754 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[21]. (**g**) Ensemble BrdU z-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 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. 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 z-scores from selected molecules aligned to multiple rDNA repeats (origin, yellow; RFB, purple).

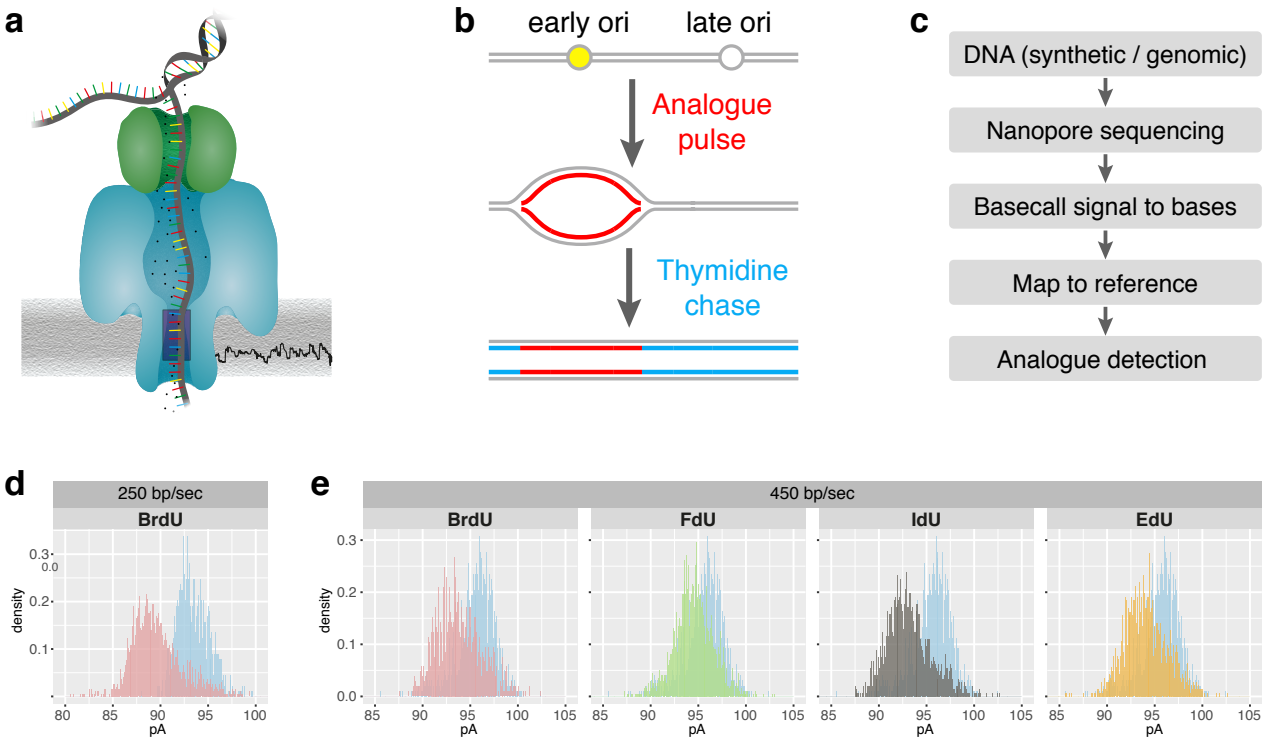

Figure 1: Nanopore sequencing can distinguish thymidine from analogues. (**a**) Graphic representation of DNA sequening using a MinION. A processive enzyme (green) ratchets DNA into the pore (blue), causing a change in ionic current (ions shown as black dots) that is determined by the 6-mer in the central 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**, **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). The data were generated using the ONT MinION R9 and R9.5 pore with sequencing speeds of 250 bp/s (**d**) or 450 bp/s (**e**), respectively.

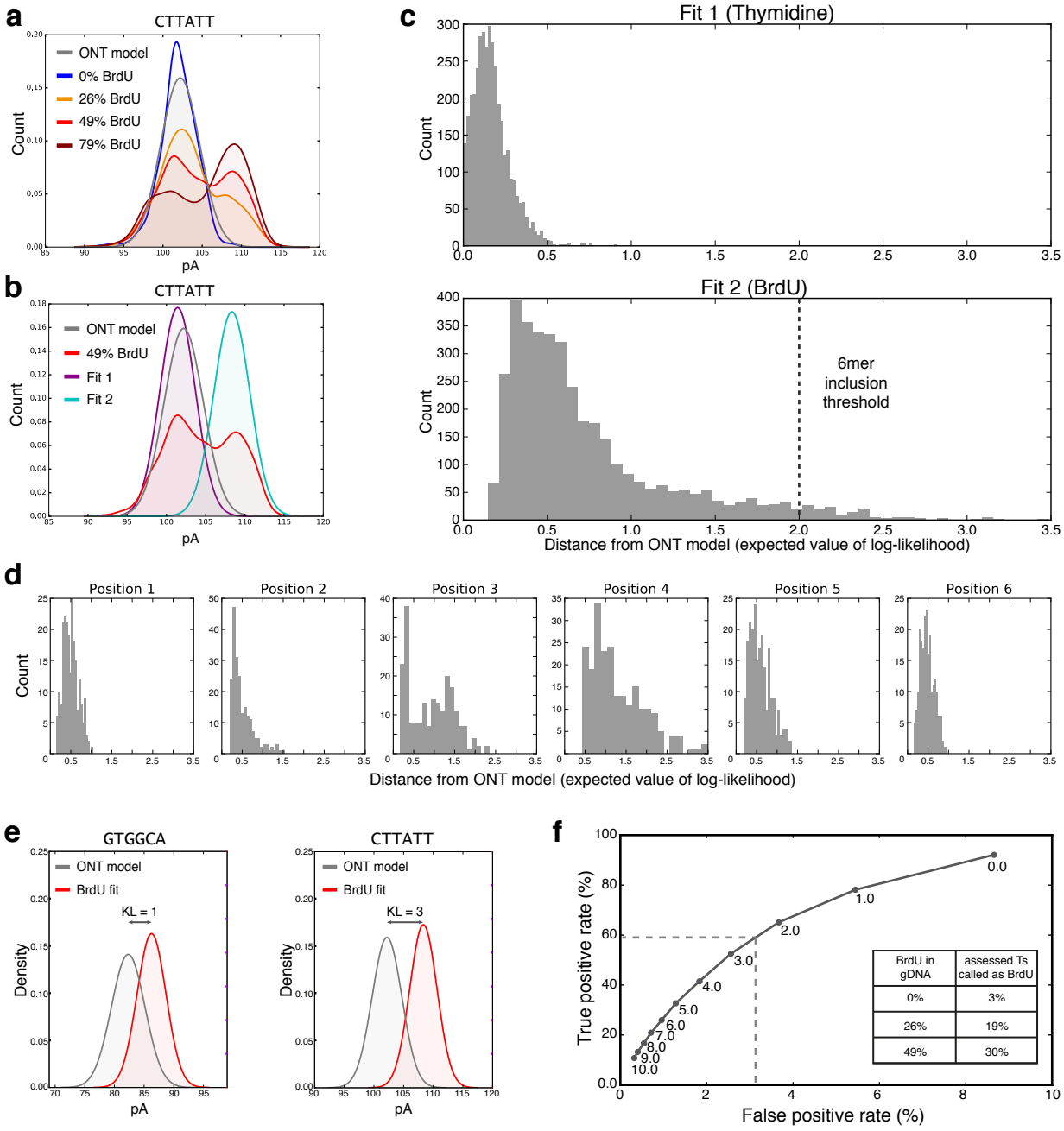

Figure 2: BrdU can be distinguished from thymidine in genomic DNA. (**a**) Signal event distributions for an example 6-mer from yeast genomic DNA containing various concentrations of BrdU (0% - blue; 26% - orange; 49% - red; 79% - crimson) compared to the ONT model (grey). (**b**) Bimodal Gaussian mixture 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.

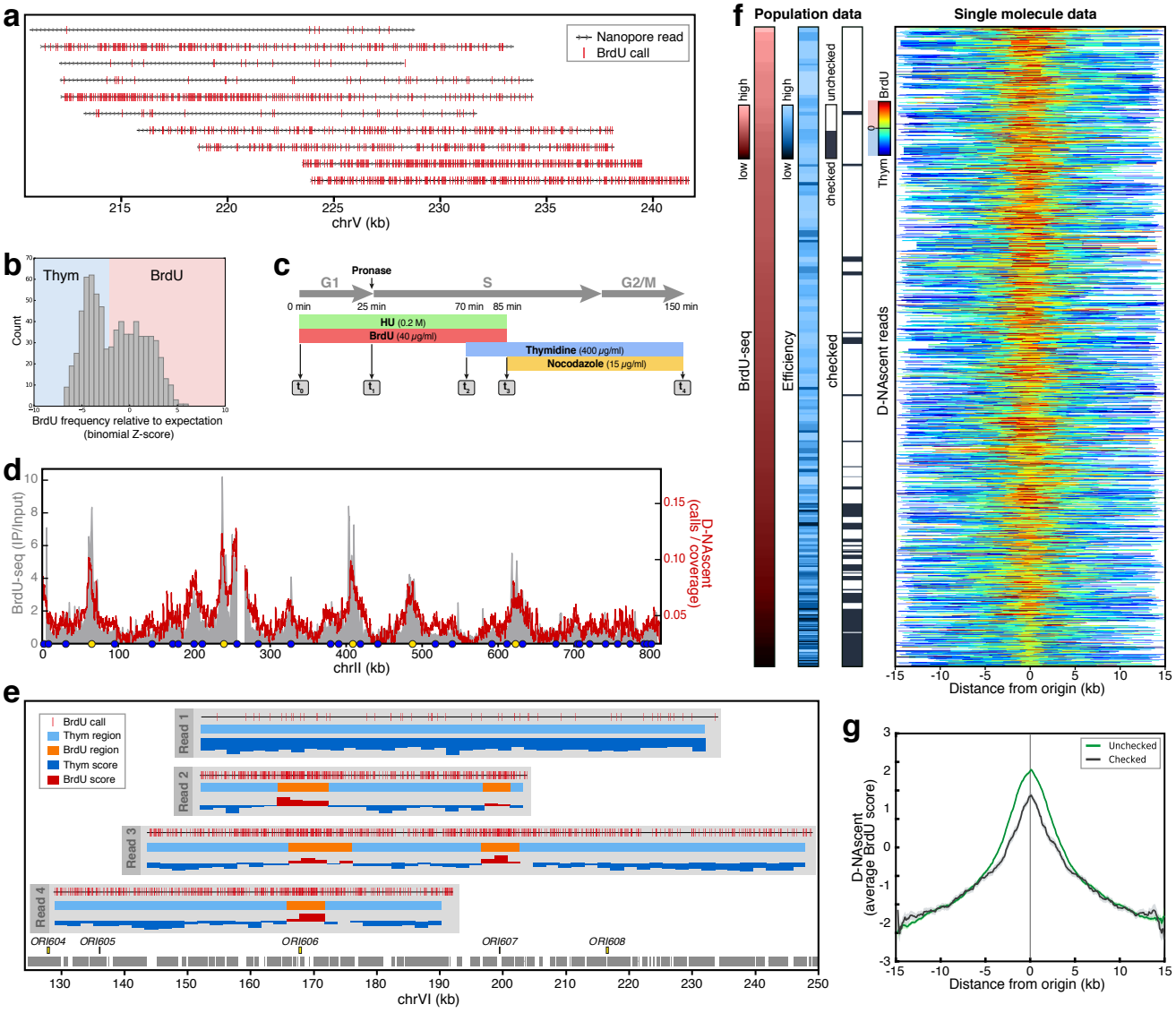

Figure 3: Single-molecule detection of BrdU on nascent DNA. (**a**) Representative nanopore reads (>15 kb) showing BrdU calls in hemi-substituted yeast genomic DNA. Red ticks indicate positive BrdU calls and arrows give the read direction relative to the sacCer3 reference genome. (**b**) The distribution 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).

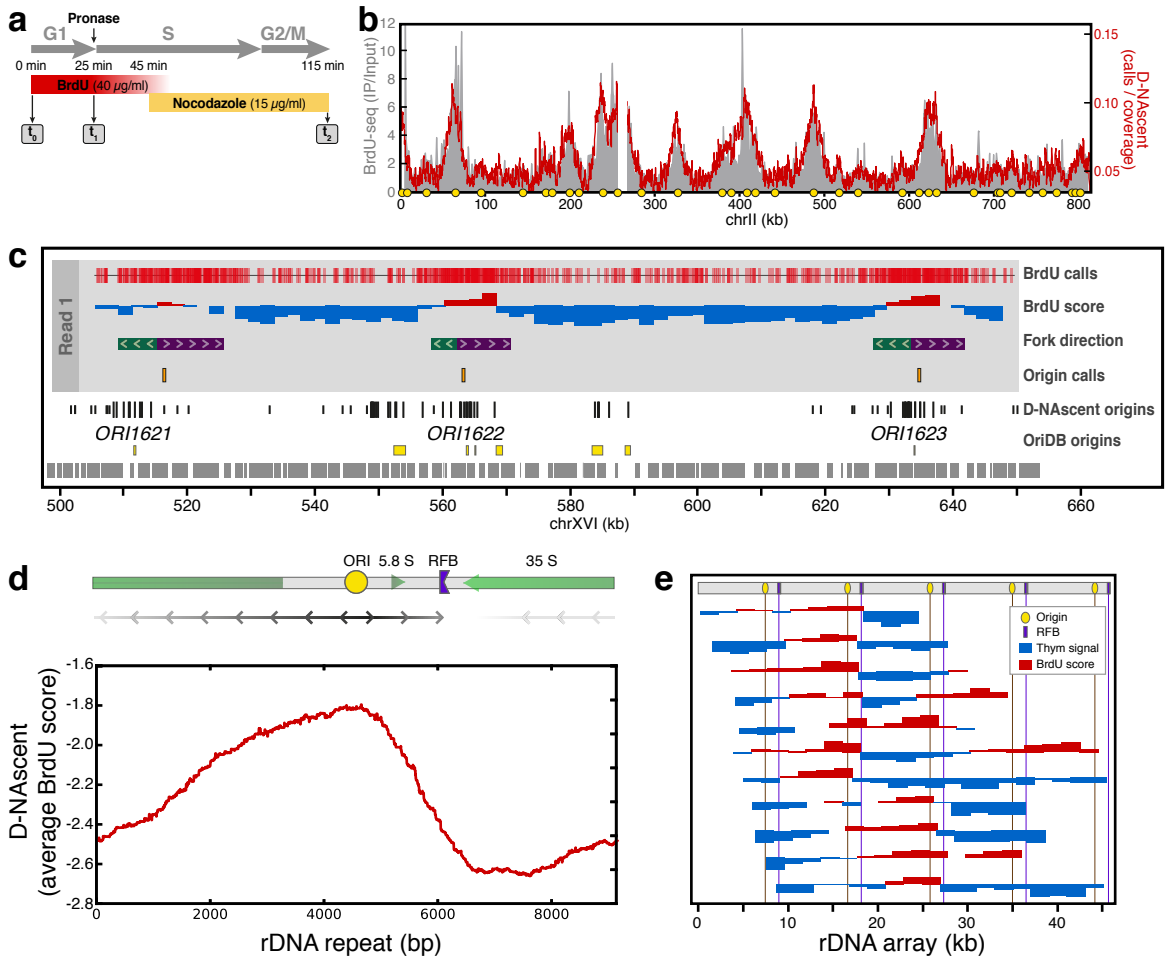

Figure 4: Single-molecule detection of replication dynamics. (**a**) Schematic of the experimental strategy for detection of replication dynamics by D-NAscent. At the indicated timepoints, samples were taken for mass spectrometry, DNA copy number measurements, BrdU-seq and D-NAscent. (**b**) Comparison 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).