MOLECULAR BIOLOGY

Nanopore-based sequencing of active DNA replication reveals key principles of metazoan replication dynamics

Dongsheng Han[1]†, Cole Shepherd[1]†, Mary Lauren Benton[2], Jared T. Nordman[1]*

Balancing replication fork progression and origin usage is essential to maintain genome stability, but measuring replication fork progression rates and origin usage throughout the genome has been challenging. Here, we use nanopore sequencing combined with DNAscent to measure replication fork progression together with origin and termination site usage with single-molecule precision throughout the Drosophila genome with nearly full genome coverage. We find that replication fork progression rates are not uniform throughout the genome. Rather, fork progression is slowest in euchromatin, and this is not correlated with active transcription. Replication origins are also influenced by chromatin, but the exact position of initiation is highly variable and are often several kilobases away from ORC (origin recognition complex) binding sites. Termination sites lack any chromatin or sequence motifs and appear nearly random throughout the genome. By measuring DNA replication dynamics at near full genome coverage, our work reveals key principles of metazoan replication dynamics.

INTRODUCTION

Complete and faithful replication of the genome is essential to maintain genome stability. DNA replication begins when the origin recognition complex (ORC) binds to thousands of sites throughout the genome and loads a double hexamer of the minichromosome maintenance (MCM)2-7 replicative helicase complex in a head-to-head orientation (*1*, *2*). Cell cycle–dependent kinase activation of the MCM2-7 helicase results in a pair of bidirectional replication forks. Activation of the MCM2-7 helicase and the first synthesis of DNA defines a replication origin. Understanding the factors that control where replication origins are specified and how origin usage is regulated throughout the genome remains an exciting challenge in eukaryotic DNA replication (*3*).

Although ORC binding in metazoans is largely sequence independent, it is highly influenced by chromatin (*4*–*6*). ORC preferentially binds to nucleosome free regions and directly contributes to the generation of nucleosome free regions (*5*–*8*). ORC binding is not uniform throughout the genome. Rather, there is a higher density of ORC binding sites in early-replicating regions of the genome relative to late-replicating regions (*4*, *5*). Several chromatin factors are also enriched in ORC binding sites (*4*, *9*–*13*). Although chromatin clearly influences ORC binding, the relationship between metazoan ORC binding sites and origins of replication (start sites of DNA synthesis) is not completely clear. For example, a recent integrative analysis of ORC binding sites and replication origins in human cultured cells revealed replication origins are often many kilobases from ORC binding sites (*14*).

Multiple population-based approaches have been used to define replication origins on a genome-wide scale (*15*–*21*). Although powerful, many of the current techniques for mapping replication initiation sites vary in resolution and concordance. Not only is there variability in origin mapping between techniques, but origin mapping by a single technique, short nascent strand sequencing (SNS-seq), also reveals extensive variation in datasets between laboratories (*14*). To circumvent problems with population-level mapping of replication origins and the biases that may be introduced during isolation of nascent DNA, nanopore-based sequencing approaches have been developed to map replication origins in budding yeast and human cells (*22*–*24*). Nanopore-based sequencing of origins has two major advantages to previous origin mapping studies. First, origins can be directly identified from patterns of nucleotide analog incorporation, alleviating the need for experimental manipulation of samples before sequencing. Second, nanopore-based sequencing of origins provides a single-molecule view of replication initiation. A recent nanopore-based origin mapping study in human cells has revealed that most origins are dispersed throughout the genome with no defining chromatin or sequence features (*23*). Although powerful, nanopore-based identification of origins with full coverage of the human genome is cost prohibitive.

Nanopore-based sequencing approaches also have the power to integrate single-molecule measurements of replication fork progression with DNA sequence information. Now, DNA combing is widely used to measure replication fork progression rates and inter-origin distances (*25*, *26*). Although powerful, DNA combing has notable limitations. First, combing does not provide genomic coordinates for replication tracks, making it impossible to relate fork dynamics to genome-wide chromatin or sequence features (*26*, *27*). Second, DNA combing is a low-throughput and labor-intensive approach and is prone to technical variability due to inconsistent fiber stretching and analog detection (*26*, *28*). Third, the efficiency of 5-chloro-2′-deoxyuridine (CldU) and 5′-iododeoxyuridine (IdU) incorporation can vary between species and cell types, and detection by immunofluorescence can limit both sensitivity and resolution (*26*, *28*). For instance, in Drosophila cultured cells, only CldU is reliably incorporated under standard conditions, limiting the ability to monitor bidirectional fork movement (*29*).

Nanopore-based single-molecule sequencing strategies have emerged as all-in-one assays to measure replication fork progression together with initiation and termination sites in a single assay (*22*, *24*, *30*–*33*). Most current studies using nanopore-based sequencing of replication dynamics have been conducted in human or yeast systems (*22*, *24*, *30*, *31*, *33*). Human studies are limited by low genome coverage (typically <10%) (*33*), whereas yeast systems, despite their tractability, lack the complex chromatin environment characteristic of metazoans (*24*, *31*, *34*). Here, we use *Drosophila*

[1]Department of Biological Sciences, Vanderbilt University, Nashville, TN 37212, USA. [2]Department of Computer Science, Baylor University, Waco, TX 76798, USA.
*Corresponding author. Email: jared.nordman@vanderbilt.edu
†These authors contributed equally to this work.

melanogaster as a model to study metazoan fork progression, origin, and termination sites throughout a metazoan genome with near full genome coverage. We find that fork progression rates are slower in euchromatic and early-replicating regions compared to heterochromatic and late-replicating regions. This difference in fork rate does not correlate with active transcription. We identified 1513 replication origins and 1984 termination sites. Replication origins are influenced by chromatin, but the exact position where initiation occurs is variable, whereas termination sites have no defining chromatin or sequence-related elements.

RESULTS

Nanopore can measure replication initiation, termination, and fork progression in Drosophila cultured cells

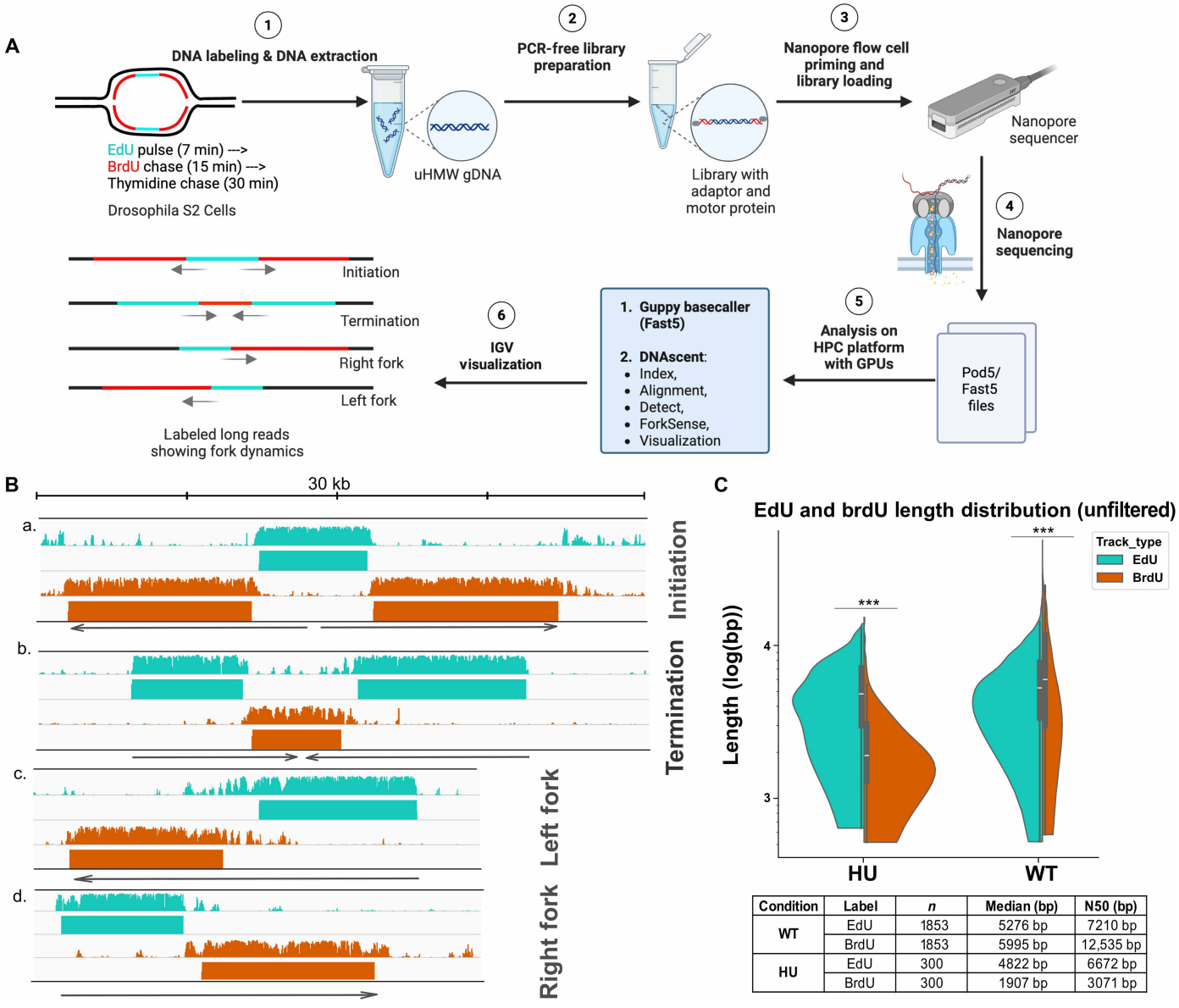
To measure DNA replication dynamics throughout the entire Drosophila genome with nearly full genome coverage and DNA sequence resolution, we adapted and optimized an approach that integrates Oxford Nanopore long-read sequencing with DNA thymidine analog labeling of nascent DNA (*22*, *24*, *30*, *31*, *33*). Specific for Drosophila S2 cells, we refined the timing and concentration of nucleotide analog pulses and optimized DNA extraction protocols to maximize yield and DNA integrity. Cells were sequentially labeled with 5-ethynyl-2′-deoxyuridine (EdU) for 7 min, followed by a 15-min chase with 5-bromo-2′-deoxyuridine (BrdU) (Fig. 1A). Both analogs can be distinguished from thymidine during nanopore sequencing based on their unique current signal profiles (*22*, *24*, *30*, *31*, *33*). After extracting genomic DNA (gDNA), we prepared polymerase chain reaction (PCR)–free libraries and performed sequencing using the Oxford Nanopore platform (Fig. 1A). We then used the machine learning–based software DNAscent (*35*) to accurately identify EdU, BrdU, and thymidine incorporation patterns along individual DNA molecules. In wild-type (WT) cells, both EdU and BrdU were efficiently incorporated into nascent DNA (Fig. 1B). We were able to capture various aspects of replication dynamics—including fork directionality, initiation, and termination events—at single-molecule resolution (Fig. 1B).

To validate this method, we exposed cells to the DNA replication inhibitor hydroxyurea (HU) at the start of the BrdU chase (*36*). This experimental design allowed EdU tracks to reflect fork progression before replication stress, whereas BrdU tracks captured fork behavior under stress. In untreated cells, BrdU-labeled tracks were significantly longer than EdU-labeled tracks (Wilcoxon signed-rank test, $P < 0.001$), with median lengths of 5995 base pairs (bp) (N50: 12,535 bp) and 5276 bp (N50: 7210 bp), respectively. N50 is a weighted midpoint of the read length distribution and is related to both the median and mean of the distribution. Upon HU treatment, BrdU track lengths were markedly reduced (median: 1907 bp; N50: 3071 bp), whereas EdU-labeled tracks remained relatively unchanged (median: 4822 bp; N50: 6672 bp) (Fig. 1C). These results demonstrate that nanopore-based approach is highly sensitive to replication stress and enables high-resolution, single-molecule analysis of replication fork dynamics in Drosophila cultured cells.

Nanopore-based sequencing can measure replication dynamics with near full genome coverage in Drosophila

Having established nanopore sequencing to measure replication dynamics in Drosophila cultured cells, we wanted to extend this approach to measure replication dynamics with higher genome coverage. To achieve this, we applied the same labeling and extraction approach but sequenced the resulting library across two PromethION flow cells. This approach yielded a total of 25,057 unfiltered replication forks, 1513 replication origins, and 1984 replication termination sites. To ensure data quality, we excluded replication fork reads when the BrdU tracks extended to the terminal ends of the molecule (*33*, *35*). After filtering, we retained 11,449 high-quality labeled molecules, which we consider high-confidence replication forks. EdU-labeled segments of these high-confidence forks had a mean length of 6223 bp, a median of 5512 bp, and an N50 of 7668 bp (Fig. 2A). In contrast, BrdU-labeled segments were much longer (Wilcoxon signed-rank test, $P < 0.001$), with a mean of 10,525 bp, a median of 8815 bp, and an N50 of 14,543 bp. To calculate forks progression rate, we choose high-confidence reads with consecutive EdU and BrdU labeling tracks and only quantified the length of the BrdU track. This ensures labeled tracks initiated during the EdU pulse and continued through the BrdU pulse, thus eliminating BrdU tracks that may have initiated during the BrdU pulse. On the basis of the N50 length over a 15-min labeling period, this translates to a fork progression rate of 969.5 bp/min in S2 cells (fork rate = BrdU track length/BrdU pulse duration; see Materials and Methods). The ratio of BrdU to EdU segment lengths had a median of 1.50 and a mean of 2.13 (Fig. 2B), consistent with the relative durations of our EdU and BrdU pulses. The median gap size between the EdU- and BrdU-labeled tracks was 1436 bp, consistent with the 3-min wash time between nucleotide pulses (fig. S1A).

Next, we sought to determine whether the number of replication forks, origin, and termination sites we obtained was sufficient to achieve full genome coverage and, if so, whether that coverage was uniformly distributed across the genome. Fork coverage was calculated as the total nonoverlapping base pairs of called forks divided by each chromosome length. We found nearly complete coverage of the mappable genome (Fig. 2C). For example, before filtering, *chr2L* exhibited 94.3% coverage (23.5 Mb), whereas *chrX* showed slightly lower coverage at 78.4% (23.6 Mb) (Fig. 2D). After filtering, fork coverage across chromosomes ranged from 81.3% (*chr2L*) to 53.8% (*chrX*) (Fig. 2D). When visualized in 100–kb bins, fork distribution was nonuniform across the genome (Fig. 2C), with regions of high and low fork density. Drosophila S2 cells carry two copies of the X chromosome and four copies of the major autosomes (near-triploid karyotype), with some segmental variation (*37*). The density of replication forks was correlated with DNA copy number in Drosophila S2 cells (fig. S1C). We next identified replication origins and termination sites based on the spatial arrangement of EdU and BrdU labels within individual molecules (Fig. 1A). To ensure uniformity, we took the midpoint of each origin and termination site and added 100 bp to each side. We identified a total of 1513 origin sites and 1984 termination events throughout the genome. Replication origins were unevenly distributed across the genome (Fig. 2E), with a median inter-origin distance of 49.69 kb and a mean of 82.78 kb (Fig. 2F), consistent with previously reported inter-origin distances in *Drosophila* cultured cells using DNA fibers (*38*). Termination sites also exhibited variable genomic density (Fig. 2G), with a median interterminination distance of 37.89 kb and a mean of 63.03 kb (Fig. 2H). Similar to replication forks, replication origins and termination sites were not uniformly distributed across the genome. Instead, both features exhibited regions of local enrichment and depletion (Fig. 2, E and G), reflecting the heterogeneous nature of replication initiation and completion across different chromosomal contexts and chromosomal copy number changes in Drosophila S2 cells (*37*, *39*). Together, data



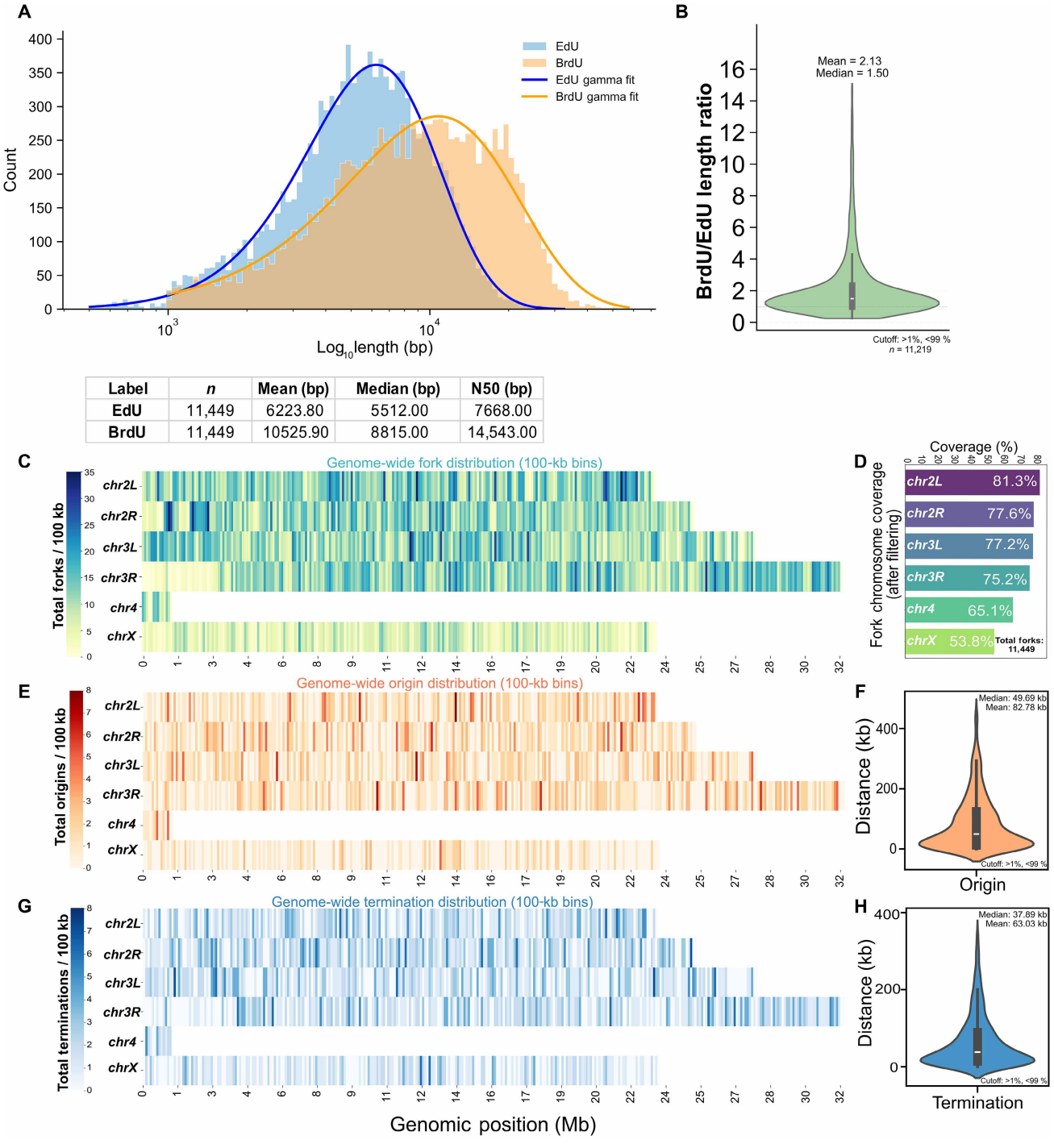
Condition	Label	*n*	Median (bp)	N50 (bp)
WT	EdU	1853	5276 bp	7210 bp
	BrdU	1853	5995 bp	12,535 bp
HU	EdU	300	4822 bp	6672 bp
	BrdU	300	1907 bp	3071 bp

Fig. 1. Nanopore-based replication fork capture and analysis. (**A**) Schematic overview of the workflow for nanopore-based sequencing of replication dynamics in Drosophila S2 cells. S2 cells are subjected to sequential pulses of EdU (cyan) and BrdU (red) to label nascent DNA. After DNA extraction, uHMW gDNA is prepared for nanopore sequencing without PCR amplification. The prepared library is loaded onto the nanopore flow cell and sequenced. The raw data are then analyzed on a high-performance computing (HPC) platform Guppy basecaller and the DNAscent software suite for signal processing and visualization. The final output consists of labeled long reads that show the dynamics of replication forks. (**B**) Examples of labeled long reads showing different replication fork dynamics. Cyan tracks correspond to EdU-labeled DNA, and orange tracks correspond to BrdU-labeled DNA. The gray line with arrows indicates the direction of the replication forks. The scale bar represents 30 kb. (a) Example of a replication initiation event. (b) Example of a replication termination event. (c) Example of a leftward moving fork. (d) Example of a rightward moving fork. (**C**) Unfiltered violin plots showing the length distribution of EdU- and BrdU-labeled segments in untreated and HU-treated cells. The horizontal lines within the violins indicate the median. The table below the plot shows the number reads (*n*), the median length, and the N50 length for each condition. Statistical significance was calculated using a two-sided Mann-Whitney *U* test. ***$P < 0.001$.

from two PromethION flow cells can provide sufficient labeled molecules to achieve up to 80% coverage of each Drosophila chromosome, enabling comprehensive genome-wide profiling of replication initiation, progression, and termination events.

Replication fork progression is influenced by chromatin

Like replication tracks measured by DNA combing, DNAscent revealed substantial variation in BrdU track lengths, potentially reflecting differences in replication fork speeds across the genome. To assess whether variation in track length is associate with chromatin state, we mapped BrdU tracks on the established nine-state Drosophila chromatin model (*39*), including updated regions that are predominantly repetitive and heterochromatic (State 0) (fig. S2A). During the conversion of the nine-state chromatin model from the dm3 to the dm6 genome assembly, regions present in dm6 but absent in dm3 were assigned to the heterochromatin state, based on these regions



Label	n	Mean (bp)	Median (bp)	N50 (bp)
EdU	11,449	6223.80	5512.00	7668.00
BrdU	11,449	10525.90	8815.00	14,543.00

Fig. 2. Nanopore based sequencing of replication dynamics achieves near genome-wide coverage. (**A**) Log-transformed length distributions of filtered EdU and BrdU tracks. Histogram shows the count of tracks at various lengths. Gamma distribution curves (orange and blue lines) are fitted to the BrdU and EdU distributions, respectively. The table provides statistics for each labeled population. Wilcoxon signed-rank test, pairing EdU and BrdU lengths, both one-sided and two-sided $P < 0.001$. (**B**) Violin plot showing the distribution of the BrdU-to-EdU length ratio for 11,219 replication tracks. The mean and median values for the ratio are shown above the plot. The plot excludes values outside the 1st and 99th percentiles. (**C**) Heatmap showing the density of replication forks across different chromosomes (*chr2L*, *chr2R*, *chr3L*, *chr3R*, *chr4*, and *chrX*) in 100-kb bins. The percentage of chromosomal coverage is shown in the table to the right. 0 represents the leftmost coordinate of each chromosome arm and may correspond to either a telomere or centromere [same for (E) and (G)]. (**D**) Table summarizing the chromosomal coverage (%) for the genome-wide fork distribution shown in (C). (**E**) Heatmap showing the density of replication origins across the chromosomes in 100-kb bins. The color intensity represents the origin density. (**F**) Violin plot showing the inter-origin distances. The mean and median values are indicated above the plot. The plot excludes values outside the 1st and 99th percentiles. (**G**) Heatmap showing the density of replication termination events across the chromosomes in 100-kb bins. (**H**) Violin plot showing the intertermination distances. The mean and median values are indicated above the plot. The plot excludes values outside the 1st and 99th percentiles.

being predominantly composed of mappable repetitive sequences and are typically annotated as heterochromatin in Drosophila (*40*). Analysis of BrdU track lengths across chromatin states revealed differences (Fig. 3, A and B). Unexpectedly, chromatin states associated with heterochromatin or repressive features—including State 0 (Dm3-not-covered), State 6 (Polycomb-repressed), and State 7 (Heterochromatin)—exhibited longer BrdU tracks than euchromatin-associated states (States 1 to 5) (*39*). State 9, which lacks defined histone modifications (*39*), also showed longer BrdU tracks relative to most euchromatic states (Fig. 3B). To more easily compare fork progression between broad chromatin compartments, we grouped the chromatin states into euchromatin (States 1 to 5) and heterochromatin (States 0, 6, 7, and 8) (fig. S2A). Consistent with individual state analysis, BrdU track length was significant longer (two-sided Mann-Whitney *U* test, $P < 0.001$) in heterochromatin (median = 7758 bp) than in euchromatin (median = 4547 bp) (Fig. 3B). To account for potential bias driven by the DNA molecule size during filtering (fig. S2, B and C), we normalized the BrdU track length to total read length for each track. Normalized BrdU track lengths revealed the same trend with longer BrdU tracks in heterochromatic relative to euchromatic states (fig. S2D).

To independently validate this observation, we grouped BrdU tracks into quintiles by length (Fig. 3C) and analyzed the chromatin state composition in each quintile. The longest track quintiles (Q4 and Q5) were enriched for heterochromatin-associated states, such as State 0 and State 7. In contrast, euchromatic states such as State 2 and State 5 were overrepresented in the shortest quintiles (Q1 and Q2). When normalized to the genomic representation of each state, this trend persisted and heterochromatin states were consistently overrepresented among the longest BrdU track quintiles, whereas euchromatin states were enriched among the shortest (fig. S2E). State 9 was also overrepresented in longer track quintiles (Fig. 3C and fig. S3E), suggesting that defined heterochromatin chromatin marks are not necessarily the driver of increased fork progression rates.

Given that replication timing (RT) is tightly linked to chromatin state (Fig. 3, B to E), we next asked whether fork progression differs between early- and late-replicating regions (*6*, *41*–*44*). We first compared BrdU lengths across four discrete RT groups using existing RT datasets generated in S2 cells (Fig. 3F) (*6*). The BrdU track lengths in "Mid-Late S" and "Latest S" regions were significantly longer (two-sided Mann-Whitney *U* test, $P < 0.05$) than those in the earliest-replicating regions (Fig. 3F). To assess this relationship more rigorously, we assigned an RT score to each BrdU-labeled track and measured its correlation with track length. This analysis revealed a weak but statistically significant negative correlation (Pearson's $R = -0.04$; $P = 1.8 \times 10^{-6}$), indicating that replication fork progression is slower in early-replicating regions (fig. S2F). We confirmed that, as expected, early-replicating regions were predominantly euchromatic, whereas late-replicating regions were enriched for heterochromatin (fig. S2G). Together, these data indicate that replication fork progression is faster in later-replicating, heterochromatin-rich regions when compared to early replicating euchromatin. This trend has recently been observed in human cells, suggesting a more universal principle in control of replication fork progression rates (*33*, *45*).

Differences in fork progression across chromatin states are largely independent of transcription

Given the significant differences in replication fork speed observed across distinct chromatin states, we investigated whether these differences could be attributed to transcriptional activity. Because replication-transcription conflicts shape fork speed and stability (*46*, *47*), we hypothesized that slower fork speeds observed in euchromatic regions might be related to higher levels of gene transcription. To test this, we examined the relationship between BrdU track length (as a proxy for fork speed) and active transcription using published precision nuclear run-on sequencing (PRO-seq) data in Drosophila S2 cells (*48*). First, we compared activated transcription levels across BrdU track length quintiles. As shown in Fig. 4A, mean reads per kilobase million (RPKM) values did not differ significantly between the shortest (Q1) and longest (Q5) BrdU track groups (Mann-Whitney *U* tests, $P > 0.05$). A cumulative distribution analysis of transcription levels between all quintiles revealed substantial overlap, consistent with minimal differences between BrdU track length and active transcription (Fig. 4B). We next performed a linear correlation analysis between BrdU track length and transcription levels, which revealed no correlation between BrdU track length and transcription levels (Pearson's $R = -0.01$, $P = 2.91 \times 10^{-1}$) (Fig. 4C). To determine whether chromatin context might influence this relationship, we performed correlation analyses within specific chromatin subtypes. We restricted our analysis to BrdU tracks located entirely within either euchromatin (States 1 to 5) or heterochromatin (States 0, 6, 7, and 8) (*39*). A slightly significant weak correlation between transcription levels and BrdU track length was observed in euchromatin (Pearson's $R = -0.07$, $P = 3.28 \times 10^{-2}$, $n = 817$), but no significant correlation was determined in heterochromatin (Pearson's $R = -0.02$, $P = 6.11 \times 10^{-1}$, $n = 762$) (Fig. 4C).

We asked whether the orientation of transcription relative to replication fork direction or S phase–specific gene expression might influence fork progression. For regions where transcription and replication were oriented counterdirectional (head-on), we observed no significant correlation between BrdU track length and head-on transcription (Pearson's $R = -0.006$; $P = 7.58 \times 10^{-1}$; $n = 2425$). Similarly, no significant correlation was observed in regions with codirectional transcription and replication (Pearson's $R = -0.032$; $P = 8.7 \times 10^{-2}$; $n = 2799$) (fig. S3, A and B). We then asked whether chromatin context was obscuring a connection between gene transcription and fork length. We found no significant correlation between BrdU track length and head-on or codirectional transcription, however, for forks and genes located solely in euchromatin (fig. S3, C and D). Last, we asked whether E2F-dependent transcripts that represent genes induced in S phase could influence fork progression (*49*). No change in fork progression was seen for E2F-dependent transcripts whether they were codirectional or counterdirectional (fig. S3, E and G). Together, these results demonstrate that transcription is unlikely to explain the observed differences in replication fork progression rates across chromatin states. Thus, additional features associated with these chromatin environments are likely to drive differential fork progression rates.

Nanopore sequencing revealed nonuniform replication fork directionality genome-wide

Nanopore-based sequencing provides a tool to measure replication fork directionality (RFD). To test for biases in fork directionality throughout the Drosophila genome, we measured the RFD throughout the genome in 100-kb window sizes (Fig. 5A). We found the vast majority of windows were replicated with both leftward and rightward forks (Fig. 5, A and B). Although we did observe a small percentage of windows that did show bias in fork directionality (fig. S4A), there was

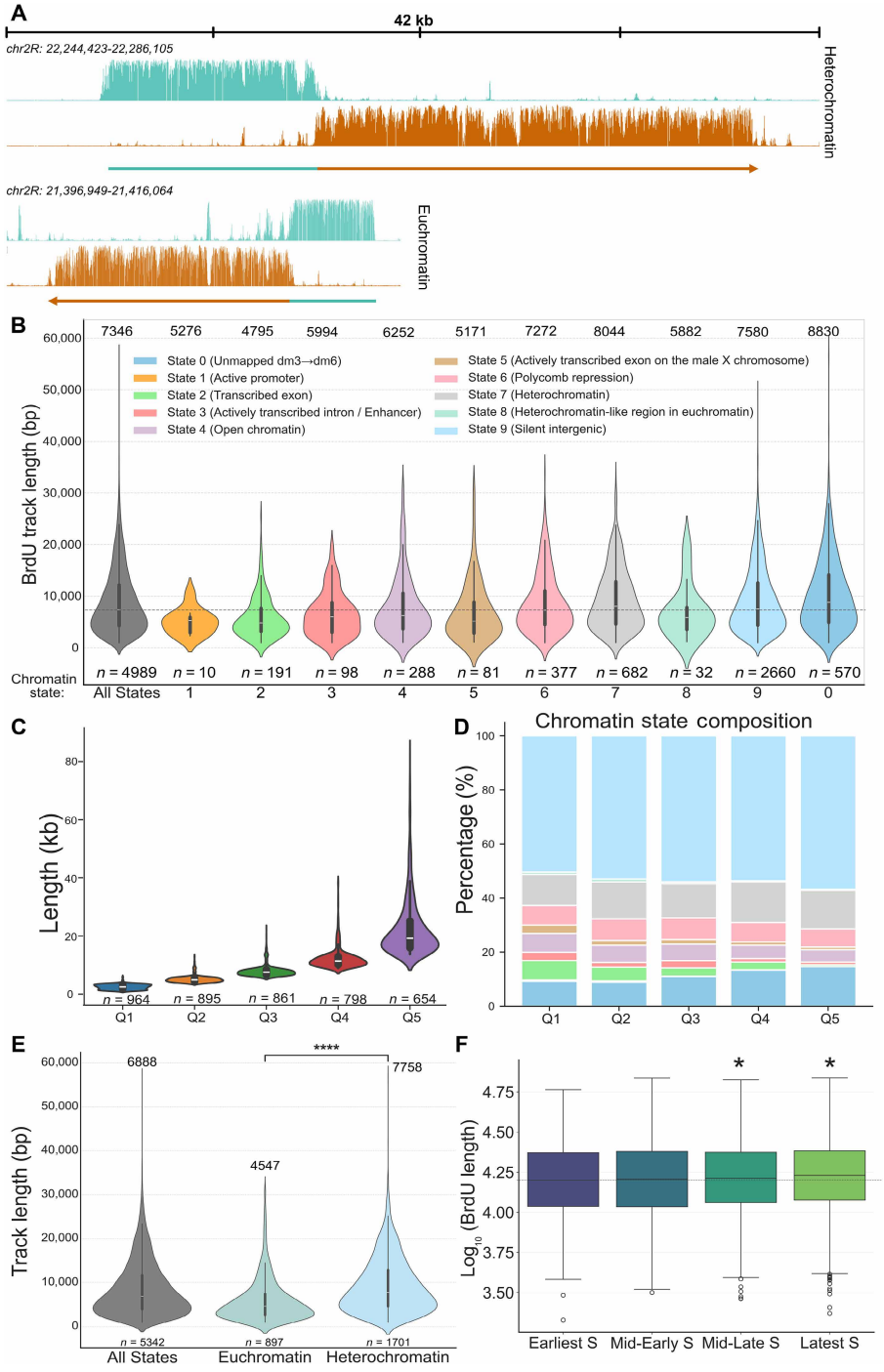


Fig. 3. Distribution of BrdU track lengths across chromatin states and S phase stages. (**A**) Representative image of EdU and BrdU track lengths in heterochromatin and euchromatin. Arrows indicate direction of fork movement. (**B**) Violin plot showing the distribution of BrdU track lengths [in base pairs (bp)] across 10 different chromatin states. The first violin represents all BrdU tracks ("All States"), whereas the remaining violins correspond to individual states defined in fig. S2A. The number of tracks (*n*) for each state is shown below the plot. The median BrdU track length (bp) for each is above the plot. (**C**) Violin plot showing the distribution of BrdU track lengths [kilobases (kb)] within each quintiles. The number of tracks (*n*) per quintiles is indicated below the plot. (**D**) Stacked bar chart depicting the percentage composition of chromatin states within BrdU track length quintiles (Q1 to Q5). The *y* axis represents the proportion of each quintiles comprised of different chromatin states, color coded as in (A). (**E**) Violin plot comparing the distribution of BrdU track lengths (bp) between broad chromatin domains. BrdU tracks are grouped into euchromatin (States 1 to 5), heterochromatin (States 0, 6, 7, and 8), and an all-reads "combined" group. The number of tracks (*n*) for each group is shown below the plot. Statistical significance was determined using a two-sided Mann-Whitney *U* test. ****$P < 0.001$. Dashed line represents the median value for all states. (**F**) Box plot showing the distribution of $\log_{10}$-transformed BrdU track lengths across distinct S-phase stages (Earliest S, Mid-Early S, Mid-Late S, and Latest S). Boxes represent the interquartile range (IQR), with the median indicated by the central line. Outliers are plotted as individual points. Statistical comparisons between each stage and the earliest S phase were performed using a two-sided Mann-Whitney *U* test. *$P < 0.05$.

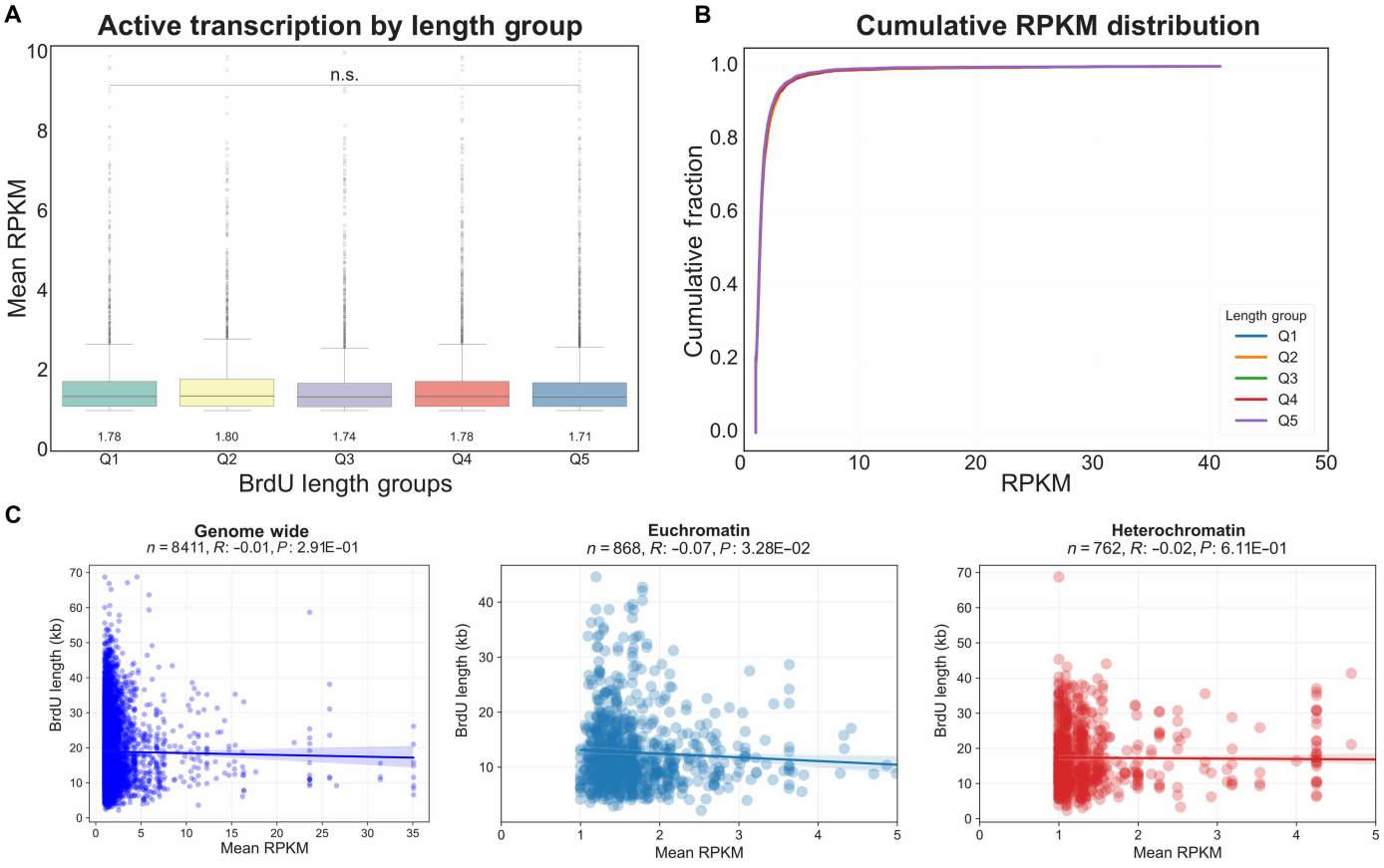


Fig. 4. Differences in fork progression rates across chromatin states do not correlate with transcript levels. (**A**) Box plot showing the distribution of mean transcription activity (in RPKM) across five BrdU track length quintiles (Q1 to Q5). Mean RPKM values (listed under the box plots) did not differ significantly between the shortest (Q1) and longest (Q5) track groups (n.s., not significant). Statistical differences between groups were tested using Mann-Whitney *U* tests. (**B**) Cumulative distribution function (CDF) plot showing the cumulative fraction of BrdU tracks as a function of transcription activity for each length quintile (Q1 to Q5). The substantial overlap of the curves for all five length groups indicates that the overall active transcriptional profiles are similar regardless of BrdU track length. (**C**) Scatterplots showing the correlation between BrdU track lengths and active transcription. The analyses are shown for all nonzero genes in three different genomic contexts.

no large-scale unidirectional replication of the genome in S2 cells (Fig. 5A).

PCR-free nanopore-based sequencing strategies can be used to capture strand-specific replication dynamics of the mitochondrial genome (*50*). The mitochondrial genome is replicated through a strand-displacement mechanism, and leftward and rightward forks are expected to map to a specific template strand (Fig. 5C). Therefore, we mapped replication tracks to the mitochondrial genome in a strand-specific manner. We detected strand-specific patterns consistent with the strand-displacement model of mitochondrial DNA (mtDNA) replication (*51*) (Fig. 5D). The small size of the mitochondrial genome and the different rate for fork progression in the mitochondrial genome will require separate optimization for ideal measurements of fork progression rates. Regardless, nanopore-based sequencing can be used to measure both nuclear and mtDNA replication dynamics in Drosophila.

Origins are influenced by chromatin and are often distant from ORC sites

DNAscent identified 1513 replication origins genome-wide with a size range from 1 to 167,590 bp and a median of 7140 bp (Fig. 6A and fig. S5A). This represents a population-level median inter-origin distance of 49.69 kb (Fig. 2F). This population-level approximation is close to a reported inter-origin distance of ~40 kb measured in Drosophila cultured cell by DNA fiber analysis (*38*). Therefore, we were able to analyze origin usage throughout the entire Drosophila genome. To assess whether origin density is influenced by RT, we compared origin density between early- and late-replicating regions. Replication origins are initially licensed through ORC-dependent MCM loading (*52*–*54*), and the density of ORC2 binding sites is higher in early-replicating regions relative to late-replicating regions (Fig. 6B) (*4*, *6*). Similar to ORC binding sites, origin density was also higher in early domains than in late domains (1.45 versus 0.73 origins per 100 kb), although the difference was smaller than that observed for ORC2 binding (Fig. 6B). Next, we examined whether chromatin state influences origin usage. Like the densities in early- and late-replicating regions, the density of origins was higher in euchromatic regions of the genome, which was also true for ORC binding sites (Fig. 6C). Together, these results indicate that origin density is influenced by RT and/or chromatin structure.

The differential origin density between euchromatin and heterochromatin suggests that chromatin structure could modulate origin usage. Because ORC binding is influenced by chromatin features such as nucleosome positioning and chromatin-associated factors (*55*–*58*), we asked whether specific chromatin factors are positively

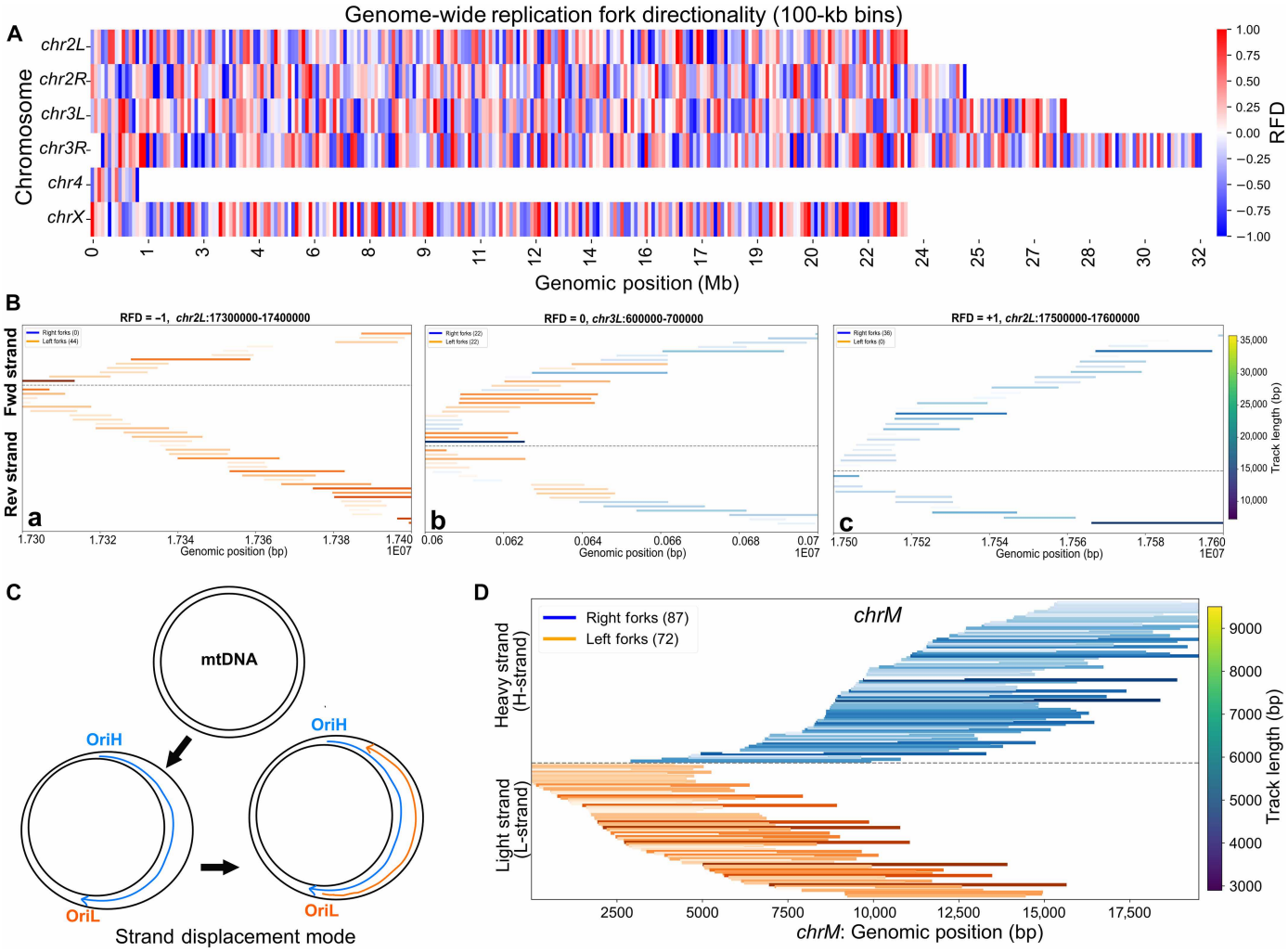


Fig. 5. Genome-wide and mtDNA RFD. (**A**) Heatmap showing the genome-wide distribution of RFD in 100-kb bins across Drosophila chromosomes. The color scale (red to blue) represents the RFD value, with positive values indicating a bias toward one direction and negative values indicating a bias toward the other. The lack of large red or blue blocks indicates no substantial unidirectional replication across the genome. (**B**) Screenshots of individual 100-kb window showing (a) leftward fork bias, (b) leftward and rightward fork replication, and (c) rightward fork bias. (**C**) Schematic of the mtDNA replication through the strand displacement mechanism. (**D**) Visualization of strand-specific replication patterns on mtDNA. Reads are plotted on either the forward (fwd) or reverse (rev) strand, with each bar representing a single read. The length of each bar corresponds to its genomic position on the chromosome (*chrM*), and its color indicates the BrdU length [in base pairs (bp)] as shown by the color bar.

or negatively associated with replication origins. To this end, we measured the frequency of overlap between origins and chromatin-associated factors with chromatin immunoprecipitation (ChIP) data available through modENCODE (*39*, *59*). To determine whether enrichments were more or less than expected by chance, we shuffled regions throughout the genome and calculated the amount of overlap for each permutation (*60*, *61*). Many chromatin factors showed significant enrichment at origins (Fig. 6D). Most of these chromatin factors are also significantly enriched at ORC2 binding sites (*9*) (Fig. 6E). More ORC2 binding sites overlapped with origins than would be expected by chance (Fig. 6C).

Although ORC2 binding sites and origins could share common chromatin features, there were also differences. The median distance between an origin and the nearest ORC2 biding site was 6.309 kb, which was closer than the distance expected by chance (median: 13.099 kb) (Fig. 6F). This is consistent with a recent meta-analysis of initiation sites and ORC binding sites in mammalian cells (*14*) and suggests a functional relationship between ORC2 binding and replication initiation. ORC binding is enriched in nucleosome-depleted regions (*6*, *8*) (Fig. 6G). In contrast, replication origins mapped by single-molecule assays show only a weak correlation with nucleosome-free regions identified by population-based methods (Fig. 6G). This indicates that, unlike ORC sites, initiation sites do not require well-defined nucleosome-free regions.

We next examined sequence features associated with replication origins. We found no specific sequence motifs enriched at origins (fig. S5B) nor any consistent bias in AT or GC content (fig. S5C). Instead, origins preferentially localize to promoters and transcription termination sites while being depleted within gene bodies, including exons and introns (fig. S5D). Together, this suggests that ORC-loaded MCMs are likely influenced by transcription, resulting in origins many kilobases away from MCM (*6*, *8*, *23*, *62*, *63*).

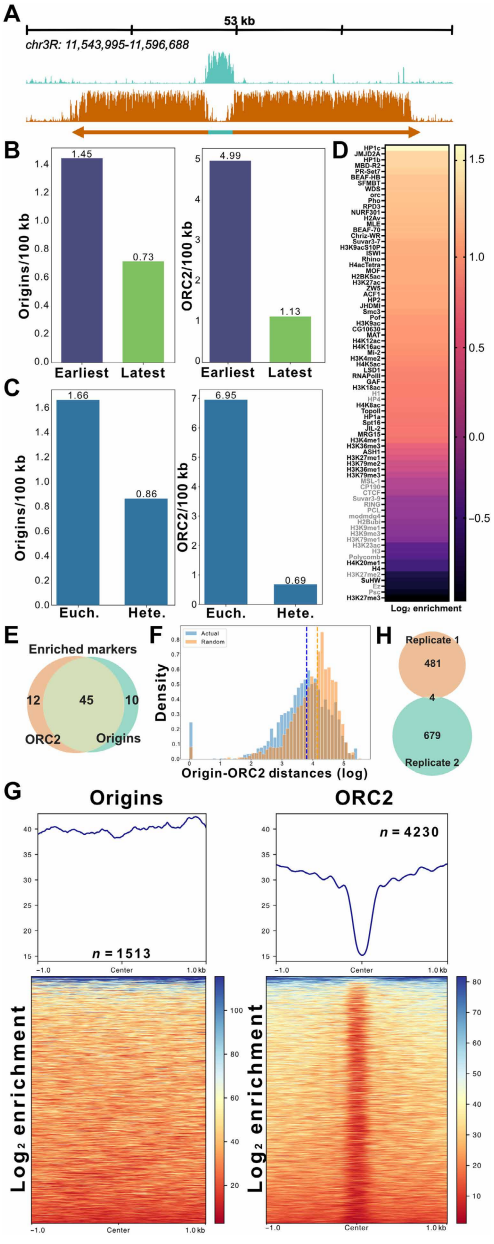


Fig. 6. Replication origin distribution, chromatin context, and usage at the single-molecule level. (**A**) Representative EdU and BrdU tracks for a single replication origin. Arrows indicate direction of fork movement. (**B**) Bar chart showing the replication origin (left) and ORC2 binding site (right) density across early- and late-replicating domains. (**C**) Bar chart comparing the replication origin (left) and ORC2 binding site (right) density between euchromatin and heterochromatin. (**D**) Heatmap showing the $\log_2$ enrichment values of chromatin factors and histone modifications at replication origins. $\log_2$ enrichment values were calculated on the basis of the observed overlap relative to the expected overlap for each chromatin factor. Black indicates statistically significant enriched or depleted factors (FDR < 0.05%). Gray indicates nonsignificant enriched or depleted factors. (**E**) Venn diagram illustrating the overlap between chromatin factors significantly enriched at replication origins and those significantly enriched at ORC2 binding sites. (**F**) Histogram comparing the distribution of distances between replication origins and the nearest ORC2 peak ("Actual") or a randomized control ("Random"). (**G**) Heatmap visualizing the chromatin accessibility by MNase-seq at replication origins and ORC2 sites. (**H**) Venn diagram showing the overlap of identified replication origins (size = 200 bp) between two biological replicates.

In two independent experiments, we identified 485 and 683 origins. Unexpected, only four origins were shared between these replicates (Fig. 6G). Even when the overlap window was expanded tenfold, only 36 origins overlapped between the two biological replicates (fig. S5E). We then pooled all origins from multiple individual replicates ($n = 1513$) and found that only 29 origins that appeared more than once. Assuming an inter-origin distance of 40 kb (*38*), we expect 3593 origins in the mappable genome (see Materials and Methods). We conducted simulated sampling to determine the number of overlapping origins we would expect to identify based on random chance (see Materials and Methods). Random sampling of 3593 sites would result in 277 shared sites based on the size of our dataset. Furthermore, only ~20% of replication origins were found in clusters (defined as three or more origins within a 40-kb window; fig. S5F). Together with our results showing that origins lack any sequence bias or motifs, we conclude that origin usage is influenced by chromatin, but the precise sites of initiation are dispersed throughout the Drosophila genome. Last, we compared our single-molecule mapping of replication origins with a published population-based SNS-seq origin mapping dataset performed in the same Drosophila cell line (*64*). We found only 99 overlaps between the two datasets, suggesting that most origins we identified by a single-molecule assay have been missed by this population-based approach (fig. S5G).

Although we identified enough origins to cover the Drosophila genome with a median inter-origin distance of ~50 kb at the population level, it is possible that undersampling is responsible for the lack of overlap between origins. To address this directly, we performed a gradient BrdU pulse approach to map additional origins in S2 cells (*23*). Briefly, increasing amounts of BrdU are applied to cells in 2.5-min intervals over the course of 1 hour (0 to 12 mM) (fig. S6A). The increased labeling period, together with the increasing concentration of BrdU, are ideal to maximize origin identification with nanopore sequencing (*23*). Using this approach, we identified an additional 1529 origins with a similar size distribution to our EdU-BrdU called origins (fig. S6, B and C). Analysis of these origins, in combination with our EdU-BrdU–identified origins, revealed similar chromatin signatures (fig. S6, D to I). Critically, there was very little overlap between origins identified between all three datasets whether origins were called at 200-bp or 2-kb resolution (fig. S6, J and K). Thus, we conclude that origins are likely dispersed throughout the genome with some chromatin constraints.

Termination sites do not correlate with chromatin features or sequence motifs

Population-based studies have not been able to identify precise sites of replication termination (*15*, *16*). This is because termination happens in zones that cannot be precisely defined by lagging-strand mapping methods (*15*, *16*). Nanopore has been used to map termination events in budding yeast and human cells; however, genome-wide analysis of termination sites has only been performed in budding yeast (*22*). Here, it was found that termination sites are more dispersed than previously thought based on population-based studies. DNAscent identified 1984 termination sites that were distributed throughout the entire Drosophila genome with a median population-level intertermination distance of 37.89 kb, similar to the median population-level inter-origin distance of 49.69 kb (Figs. 2, G and H, and 7A).

In contrast to origins, there was an increased number of termination sites in late-replicating regions of the genome (Fig. 7B). To ask

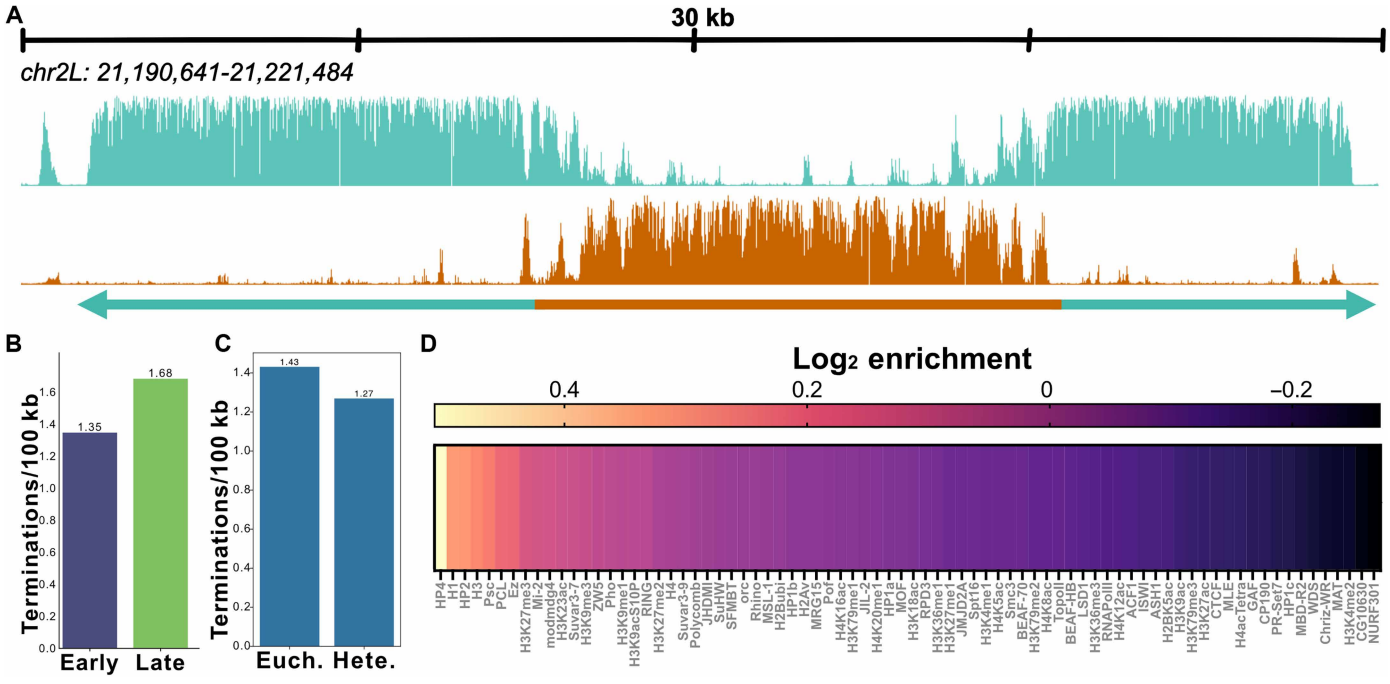


Fig. 7. Termination events appear to be stochastic. (**A**) Representative EdU and BrdU tracks for a single termination event. Arrows indicate direction of fork movement. (**B**) Bar chart comparing the density of termination events (peaks per 100 kb) between early- and late-replicating regions. (**C**) Bar chart comparing the replication origin density between euchromatin and heterochromatin. (**D**) Heatmap showing the $\log_2$ enrichment values of chromatin factors and histone modifications at replication termination sites. Log_2 enrichment values were calculated on the basis of the observed overlap relative to the expected overlap for each chromatin factor. Black indicates statistically significant enriched or depleted factors (FDR < 0.05%). Gray indicates nonsignificant enriched or depleted factors.

whether the density of termination events was influenced by chromatin, we quantified the density of termination sites in euchromatin and heterochromatin and found only a small difference (Fig. 7C). We used a random shuffling method to determine whether specific chromatin factors are enriched at termination sites. Although there were some enrichment and depletions, none of these reached the level of significance [false discovery rate (FDR) < 0.05; Fig. 7D]. Furthermore, no AT/GC sequence bias or sequence motif was found at termination sites (fig. S7, A and B). Termination sites were found to be slightly enriched in regions of noncoding RNA, but no significant enrichment or depletion was found in any other regions (fig. S7C). Termination sites between biological replicates showed very little overlap (fig. S7, D and E), and only 16 of 1984 termination sites were mapped more than once from the combined datasets. Therefore, we conclude that termination sites are potentially random throughout the genome with no defining chromatin or sequence features.

DISCUSSION

Nanopore-based sequencing strategies have the potential to answer long-standing questions about metazoan DNA replication that have been difficult to address due to technical limitations (*22*–*24*, *31*). Here, we have used nanopore-based sequencing of active DNA replication using DNAscent to understand key principles of metazoan replication fork dynamics, origin, and termination site selection with single-molecule precision (*33*, *35*). Given the genome size of *D. melanogaster*, we were able to define single-molecule replication events that cover nearly the entire genome with two PromethION flow cells. Thus, we have provided a single-molecule analysis of replication dynamics covering nearly an entire metazoan genome.

DNA combing has been the gold standard method to measure rates of replication fork progression throughout the genome (*22*, *24*, *30*–*33*). Although powerful, this method is unable to provide sequence-level information. Therefore, it is impossible to know whether track length variability in DNA combining experiments is caused by random fluctuations in nucleotide incorporation rates or reflect differences in fork progression rates throughout the genome driven by biological constraints to fork progression. We were able to address this issue by combining BrdU and EdU pulses with DNAscent, generating single-molecule measurements of fork progression throughout the Drosophila genome. To our surprise, we found that replication fork progression is slower in euchromatic and early-replicating regions of the genome relative to heterochromatic and late replication portions of the genome. Nanopore-based measurements of replication fork progression in budding yeast found no difference in fork progression rates throughout most of the genome (*31*). In contrast, our work in Drosophila and recent nanopore-based measurements of fork progression in human cells have revealed that fork progression is slower in early-replicating regions of the genome relative to late-replicating regions (*33*, *45*). Furthermore, quantification of replication fork speed in single cells have also found the same trend (*45*). It is likely that the higher-order chromatin structure found in metazoans, and lacking in budding yeast, could contribute to differences in fork rates throughout the genome. For example, budding yeast lacks large blocks of pericentric heterochromatin and the linker histone H1 (*65*, *66*).

The driving force causing reduced fork rate speeds in euchromatin relative to heterochromatin remains unresolved. One potential driver of fork slowing in early and euchromatic regions are replication-transcription conflicts (*46*, *67*, *68*). Single-cell measurements of fork

speeds in human cells revealed that transcription is at least partially responsible for the slowing of fork progression in early-replicating regions of the genome (*45*). Our work in Drosophila, however, suggests that transcription is unlikely to contribute to fork slowing in euchromatic regions of the genome. We were unable to find a correlation between transcript abundance and fork progression that would suggest that transcription shapes the rate of fork progression throughout the genome. We also did not see a difference in fork progression rates when forks were traversing transcription units in head-on versus codirectional orientation, as would be predicted if slowing was driven by replication-transcription conflicts (*46*, *67*, *68*). The difference between our results and single-cell sequencing in human cells could be driven by single-molecule versus population-level approaches or differences between human and Drosophila DNA replication. An independent study of fork progression rate in human cells, however, identified the same slowing of fork progression in early-replicating regions that appears independent of transcription (*33*). If not transcription, what is driving slower replication in early and euchromatic regions of the genome? One possibility is that higher origin density in these early-replicating regions limits the availability of replication factors, such as deoxynucleoside triphosphates, thereby slowing fork progression as more forks compete for shared nucleotide pools (*69*, *70*). Alternatively, it is possible that topological stress due to more frequent origin firing in early replicating and euchromatic regions of the genome could slow fork progression. We do observe an increased density of origins in these regions.

DNAscent allows for the identification of replication forks together with replication initiation and termination sites (*22*–*24*, *31*). Thus, we were able to define replication origins from single-molecule replication events throughout the entire genome. Several strategies have been used to define replication origins from using population-based approaches. Nanopore-based sequencing, however, provides a single-molecule view of replication initiation, allowing a much more precise view of where initiation actually starts in the genome. By performing this analysis in Drosophila, we were able to provide coverage of nearly the entire genome, thus revealing key features of metazoan replication initiation. Unexpectedly, of the >3000 replication initiation sites we identified, very few were identified more than once. Given the genome coverage, number of events identified, and expected inter-origin distance, we would have expected to identify individual origins more than once. This suggests that, at the single-molecule level, where initiation starts is relatively stochastic. Nanopore-based sequencing of origins in human cells and a meta-analysis of initiation sites in human cells came to a similar conclusion (*14*, *23*).

Although DNA sequence does not influence where initiation starts, chromatin context does. For example, we found a higher density of origins in early and euchromatic regions of the genome relative to late and heterochromatic regions. Furthermore, we were able to identify chromatin features that were statistically enriched at origins. Many of these features were similar to features enriched at ORC binding sites (*9*). Although there was some overlap between origins and ORC binding sites, the median distance between an origin and ORC binding site was closer than expected by chance. A recent meta-analysis of ORC binding sites and origins in human cell lines found the vast majority of origins were >1 kb away from an ORC binding site (*14*). Thus, physical separation of ORC binding sites and replication origins seems to be a consistent trend in metazoan replication. Although there is a formal possibility that specification of initiation sites is independent of ORC, MCM distribution is shaped by transcription and Cyclin E activity. Therefore, we proposed that ORC-loaded MCMs moved by transcription and other chromatin forces are stochastically activated by Dbf4-dependent kinase. Thus, the MCMs are a moving target rather than a static complex waiting to be activated. This would explain the distance between ORC and initiation sites and why actual sites of initiation, defined by a single-molecule approach, are dispersed throughout the genome.

We also identified 1984 termination sites. To our knowledge, there has been no formal analysis of metazoan termination sites at the single-molecule level. Population-based approaches such as OK-seq or SNS-seq cannot define precise sites of replication termination because termination occurs in zones between initiation sites (*71*–*73*). Nanopore-based strategies have been used to define termination sites in budding yeast and revealed that they were often dispersed between replication origins (*24*, *31*). We failed to identify any sequence bias or chromatin signature associated with termination sites in metazoan. Thus, replication termination events occur stochastically throughout the genome without sequence specificity or a chromatin signature.

Nanopore sequencing technology has the potential to revolutionize our understanding of DNA replication dynamics. This single-molecule, long-read sequence technology can reveal properties of fork progression, origin, and termination site usage throughout the genome in a single assay. Although powerful, nanopore-based detection of replication dynamics with EdU and BrdU has limitations. For example, defining the precise boundary of EdU- and BrdU-labeled regions could generate artifacts in track length. In addition, this method still requires mapping nanopore-generated reads to the genome. Therefore, unmapped sequences and repetitive DNA elements are areas that will require advancements for complete genome analysis. In addition, changes in dNTP concentrations, whether in the context of cell type or duration of S phase, could change the incorporation EdU and BrdU incorporation frequency. Although none of these limitations are insurmountable, there is still work needed to optimize this approach. Our work reveals that nanopore-sequencing based approaches can measure replication dynamics with almost full coverage of a metazoan genome. Given the only requirement is the incorporation of BrdU and EdU, this technology can be adapted to many organisms, cell types, and tissues. Furthermore, as a telomere-to-telomere genome assembly becomes available for the Drosophila genome, our datasets can be reanalyzed to define replication dynamics in regions of the genome that have been difficult to map, including satellite sequences.

MATERIALS AND METHODS

Cell culture

S2 cell stocks were maintained in T225 flasks with room temperature Schneider S2 media supplemented with 10% fetal bovine serum (FBS), penicillin (100 U/ml), and streptomycin (100 μg/ml). Cells were passaged 1:5 dilution (1-part old media:5-part final volume) when flasks reached 90 to 100% confluency (fully confluent). New stocks were thawed and maintained after 30 passages.

DNA labeling and validation for nanopore sequencing

Twenty-four hours before labeling, a fully confluent flask was passaged 1:2 or 1:3 to achieve 60 to 80% confluency at the time of labeling to ensure that cells were actively replicating. The two flasks of cells resulting from this split were labeled at the same time and combined

during a later step. The day of labeling, old media were removed and immediately replaced with 25 ml of new media supplemented with 10 μM EdU. Cells were returned to the 25°C incubator for 3 or 7 min for labeling. After the appropriate time had passed, media were removed, and the cells were gently washed three times with 10 ml of 1x phosphate-buffered saline (1x PBS). Cells were then incubated with 25 ml of new media supplemented with 10 μM BrdU. For HU treatment, 1 mM was added to the BrdU media. Cells were returned to the 25°C incubator for 6 or 15 min for labeling. After the appropriate time had passed, media were removed, and the cells were gently washed three times with 10 ml of 1x PBS. The cells were then incubated with 25 ml of new media supplemented with 50 μM thymidine. Cells were returned to the 25°C incubator for 20 min before harvesting. Cells from pairs of flasks labeled at the same time were combined into a single 50-ml conical flask, and 200 μl of cells were set aside on ice for later counting and labeling validation. The remaining cells were pelleted at 600*g* for 5 min, and the old media were removed. The cells were then washed via resuspending in 10 ml of 1x PBS and subsequent pelleting at 600*g* for 5 min. This wash step was repeated once. After washing, the cells were resuspended into 1 ml of 1x PBS, transferred to a 1.5-ml Eppendorf tube, and pelleted at 600*g* for 5 min. The supernatant was discarded, and the cell pellets were stored at −80°C for up to 2 weeks until DNA extraction and sequencing.

Gradient labeling for nanopore sequencing and origin identification

Twenty-four hours before labeling, a fully confluent flask was passaged 1:3 to achieve 60 to 80% confluency at the time of labeling to ensure that cells were actively replicating. The two flasks of cells resulting from this split were labeled at the same time and combined during a later step. The day of labeling, old media were removed and immediately replaced with new media. The flasks were treated with a gradient of BrdU in 0.5 μM steps to a maximum of 12 μM over a period of an hour, with each step lasting for 2.5 min. Between each step, the cells were returned to the 25°C incubator. After the gradient, the flasks remained in the incubator for a further half hour at 12 μM before being scraped and cells from pairs of flasks labeled at the same time were combined into a single 50-ml conical flask and 200 μl of cells were set aside on ice for later counting and labeling validation. The remaining cells were pelleted at 600*g* for 5 min, and the old media were removed. The cells were then washed via resuspending in 10 ml of 1x PBS and subsequent pelleting at 600*g* for 5 min. This wash step was repeated once. After washing, the cells were resuspended into 1 ml of 1x PBS, transferred to a 1.5-ml Eppendorf tube, and pelleted at 600*g* for 5 min. The supernatant was discarded, and the cell pellets were stored at −80°C for up to 2 weeks until DNA extraction and sequencing.

Extraction of uHMW gDNA for nanopore sequencing

Ultrahigh-molecular-weight (uHMW) gDNA extraction and library preparation were conducted as a variation on New England Biolab's adaption for their T3050L Kit for Oxford Nanopore Sequencing. EB+ was prepared by supplementing 1 ml of Monarch Elution Buffer 2 with 2 μl of 10% Triton X-100. One hundred million previously labeled and frozen cells were resuspended to a concentration of 50 million cells per 100 μl of 1x PBS and divided into two 1.5-ml Eppendorf tubes for parallel extraction. The cells were lysed using 100 μl of Nuclei Prep Buffer and 50 μl of RNase A (10 mg/ml) and mixed at room temperature for 2 min. Nuclei Lysis Solution was prepared using 200 μl of Nuclei Lysis Buffer and 80 μl of Proteinase K (1 mg/ml). The nuclei were then lysed using 150 μl of Nuclei Lysis Solution. The sample was inverted to mix, incubated at 56°C and 300 rpm for 10 min, and then incubated at room temperature for 20 min. The sample was precipitated using 150 μl of Precipitation Enhancer, two DNA Capture Beads, and 700 μl of isopropanol and incubated on a vertical rotating mixer at 6 rpm for 8 min. The samples were then washed twice with 500 μl each of Wash Buffer before incubating in 200 μl of Proteinase K (1 mg/ml) for 16 hours at room temperature. The following morning, the precipitation step was repeated with 75 μl of Precipitation Enhancer and 350 μl of isopropanol and incubated on a vertical rotating mixer at 6 rpm for 8 min. The samples were washed twice more with 500 μl each of Wash Buffer, and the beads with naked DNA were transferred to a spin column and quickly tapped on an absorbent pad to remove excess buffer and then transferred to 385 μl of EB+ buffer. The DNA was allowed to solubilize at room temperature for 16 hours. The sample was poured into a bead retainer seated in a 1.5-ml LoBind Eppendorf tube and spun at 16,000*g* for 1 min. The sample was examined, and if DNA strands were visible between the flow through and the beads, the sample was spun again. The sample was mixed gently with a wide-bore pipette until maximally solubilized, pipetted roughly 10 to 20 times, and then stored in the fridge.

A 10-μl sample was aspirated using a wide-bore pipette tip, with effort being made to aspirate from a homogeneous region of the sample if separation is present. The 10-μl sample was added to a 2-ml Eppendorf tube with a DNA Precipitation Bead and vortexed for 2 min to shear the DNA. The sample concentration and purity were recorded via NanoDrop.

Library preparation for nanopore sequencing

Sample libraries were prepared using Oxford Nanopore's Ultra-Long Sequencing Kit (catalog no. SQK-ULK114; batch no. ULK114.20.0003). A dilute fragmentation mix was prepared using 6 μl of Fragmentation Mix (FRA) and 244 μl of Fragmentation dilution buffer (FDB) before the dilute fragmentation mix was mixed with the sample via gently pipetting with a wide-bore pipette for 2 min at room temperature. The sample was then incubated for an additional 8 min at room temperature followed by 10 min at 75°C, 5 min on ice, and lastly 5 min at room temperature. Five microliters of Rapid Adaptor (RA) was added and mixed thoroughly via wide-bore pipetting, and the sample was incubated at room temperature for 30 min. Precipitation and sample cleanup were carried out by adding a precipitation star sample followed by 500 μl of Precipitation Buffer (PTB). The sample was precipitated on a vertical rotating mixer for 20 min at 6 rpm. The sample was then spun briefly, and the supernatant was removed. The sample was eluted in 300 μl of Elution Buffer (EB) overnight at room temperature before being stored at 4°C.

PromethION loading, reloading, and sequencing

A 90-μl sample was mixed with 100 μl of Sequencing Buffer (SBU) and 10 μl of Loading Solution Buffer (LSU) to prepare for loading onto the flow cell. This mixture was mixed thoroughly via wide-bore pipetting, incubated at room temperature for 1 hour, and, if any insoluble DNA was visible in the sample at this point, spun down at 2000*g* for 2 min immediately before loading. The flow cell primer was prepared with 30 μL Priming Tether (FTU) and 1170 μL Flush

Buffer (FB). The flow cell was prepared for loading by setting a P1000 pipette to 200 µl, inserting the tip into the sample loading port, and turning the wheel to draw up liquid until ~10 µl of the liquid was visible in the pipette tip. A 500-µl flow cell primer was loaded onto the flow cell. After 5 min, flow cell priming was completed by loading another 500 µl of the flow cell primer onto the flow cell. The sample was then immediately loaded dropwise onto the flow cell using a wide-bore pipette. After 10 min, sequencing was initiated. The flow cell was washed and reloaded following ONT's wash protocol twice, as published on 25 April 2024, and reloaded following the initial loading protocol, every 24 hours.

MinION loading, reloading, and sequencing

A 33.8-µl sample was mixed with 37.5 µl of Sequencing Buffer (SBU) and 3.7 µl of Loading Solution Buffer (LSU) to prepare for loading onto the flow cell. This mixture was mixed thoroughly via wide-bore pipetting, incubated at room temperature for 1 hour, and, if any insoluble DNA was visible in the sample at this point, spun down at 2000*g* for 2 min immediately before loading. The flow cell primer was prepared with 30 µl of Priming Tether (FTU) and 1170 µl of Flush Buffer (FB). The flow cell was prepared for loading by setting a P1000 pipette to 200 µl, inserting the tip into the priming port, and turning the wheel to draw up liquid until ~10 µl of the liquid was visible in the pipette tip. A 500-µl flow cell primer was loaded onto the flow cell. After 5 min, flow cell priming was completed by loading another 500 µl of the flow cell primer onto the flow cell via the priming port. The sample was then immediately loaded dropwise onto the flow cell using a wide-bore pipette via the SpotON sample port. After 10 min, sequencing was initiated. The flow cell was washed and reloaded following ONT's wash protocol twice, as published on 25 April 2024, and reloaded following the initial loading protocol, every 24 hours.

Bioinformatic analysis

Nanopore sequencing data processing and DNAscent replication analysis

Raw Oxford Nanopore sequencing data in POD5 format were processed using a custom DNAscent analysis pipeline. First, POD5 files were converted to FAST5 using the Oxford Nanopore pod5 utility. Basecalling was performed with Guppy v6.4.2 (dna_r10.4.1_e8.2_400bps_5khz_hac configuration, GPU-enabled). Reads were aligned to the Drosophila melanogaster dm6 reference genome using Minimap2 v2.24 with the map-ont preset, followed by conversion and indexing using Samtools v1.10. DNAscent indexing was used to generate searchable indices from the FAST5 files and sequencing summary data. Replication detection was performed with DNAscent (GPU-accelerated) using a minimum read length threshold of 15 kb. ForkSense analysis was conducted with all detection modes enabled, including forks, origins, terminations, analogs, and stress signatures, using EdU and BrdU as replication markers. Bedgraph files were generated using DNAscent utility scripts to visualize replication features across the genome.

Track filtering and quality control for fork progression analysis

Labeled replication tracks were filtered to retain high-confidence replication fork calls. The DNAscent forkSense output BED files (leftForks_DNAscent_forkSense_stressSignatures.bed and rightForks_DNAscent_forkSense_stressSignatures.bed) were merged. Fork entries with negative DNAscent-assigned stall scores were excluded as these indicate unreliable or ambiguous events: −1 (termination site), −2 (segmentation error), −3 (read end), and −4 (proximity to indels > 100 bp). Because of the sequential EdU-then-BrdU labeling, initiation events were retained. In these cases, BrdU incorporation following EdU reliably marks fork progression from replication origins. By contrast, termination events lacking BrdU labeling likely reflect forks that completed before BrdU was incorporated and were excluded due to insufficient labeling continuity. This filtering strategy ensured that only high-confidence replication forks were included in downstream analyses.

Gradient track and origin identification

Raw Oxford Nanopore sequencing data in POD5 format were processed using a custom pipeline using DNAscent v4.0.3 and Dorado v0.9.1 basecaller. POD5 file were basecalled using Dorado (na_r10.4.1_e8.2_400bps_5khz_hac configuration, GPU-enabled). DNAscent index and detect steps were used with identical options as the other analysis (including production of a human readable DETECT file rather than the updated BAM file format). The DETECT file was split by read, and the second column (first analog probability value column) was removed in preparation for ori-ter_promethion_calls_server_TVDiff.R from Carrington *et al.* (*23*). Individual read files less than 300 lines in length were removed before gradient detection due to the ori-ter_promethion_calls_server_TVDiff.R script seeming to encounter errors in processing reads approaching the window size. The script was then run with a window size and step size of 250 bases, a probability threshold of 0.5, and a gradient threshold of 1. Origin calls were visually validated using the limited_plot_read.sh script from the fork_arrest repository (*74*).

BrdU and EdU track length determination, ratio calculation, and statistical analysis

Replication fork progression data were obtained from DNAscent forkSense output BED files after quality filtering. EdU and BrdU track lengths, representing base pairs synthesized during each labeling period, were extracted from these files. Entries with negative stall scores or zero EdU track length were excluded. The BrdU-to-EdU ratio was calculated as BrdU track length divided by EdU track length for each fork, with ratios above 100 removed as outliers. Ratio distributions were summarized by descriptive statistics and compared across groups using Wilcoxon signed-rank tests. Ratios were analyzed alongside track lengths, stall scores, and fork direction to characterize replication dynamics.

RFD calculation

RFD was calculated using fork direction data from unfiltered DNAscent forkSense BED files, which contain genomic coordinates and strand annotations ("left" or "right"). The genome was divided into nonoverlapping 100-kb windows. For each window, the number of rightward- and leftward-moving forks was counted, and RFD was computed using the formula: RFD = (rightward − leftward)/(rightward + leftward). This yielded values ranging from −1 (fully leftward) to +1 (fully rightward), with 0 indicating balanced fork directionality.

Genome-wide RFD heatmap generation

Genome-wide RFD heatmaps were generated by binning RFD values into 100-kb windows across all chromosomes. A matrix was constructed with chromosomes as rows and genomic bins as columns. Each cell represented the RFD value in a given genomic region. Heatmaps were visualized using a custom blue-white-red color scale centered at zero to highlight replication polarity patterns across the genome. Axes were labeled with chromosome names (*y* axis) and genomic coordinates in megabases (*x* axis). A color bar indicated the RFD scale.

Linear genomic visualization of mitochondrial replication fork tracks

DNAscent ForkSense output BED files containing BrdU-labeled mitochondrial replication tracks were filtered to include only those overlapping the mitochondrial genome. Tracks were classified by strand orientation (forward or reverse) and fork direction (left or right) and then sorted by genomic start position within each category. Visualization consisted of linear plots with forward strand tracks displayed above and reverse strand tracks below a central axis, where each track is depicted as a horizontal line proportional to its genomic length. Fork direction was indicated by a color scale—blue for rightmoving forks and orange for left-moving forks—with intensity reflecting track length.

Fork coverage calculation and heatmap visualization

Fork coverage was calculated by summing leftward and rightward fork counts within each nonoverlapping 100-kb window, based on DNAscent forkSense BED files. A fork coverage matrix was created with chromosomes as rows and genomic bins as columns, where each cell contained the total fork count for that region. Missing values were assigned to bins without any detected forks. Heatmaps were generated using the viridis color scale to represent replication fork density, with darker colors indicating higher coverage. Matrix structure and axes were standardized as described above to enable direct comparison with RFD heatmaps.

Origin and termination coverage heatmap generation and analysis

Replication origin or termination sites identified from DNAscent forkSense BED files were separately assigned to nonoverlapping 100-kb genomic windows based on their genomic start positions. For each window, coverage was defined as the number of origins or terminations within that bin. Genome-wide coverage heatmaps were generated with chromosomes as rows and 100-kb bins as columns. Origin and termination densities were visualized using continuous color scales optimized to capture dynamic range and density variation.

Inter-origin and intertermination distance calculation and distribution analysis

Replication origin or termination coordinates were obtained from BED files generated by DNAscent ForkSense outputs, containing chromosome names and start and end positions. Midpoints were calculated as the average of start and end coordinates. Coordinates were grouped by chromosome and sorted by midpoint in ascending order. Inter-event distances were calculated as the difference between consecutive midpoints within the same chromosome and converted to kilobases. When plotting, outliers outside the 1st to 99th percentile range were excluded.

BrdU track mapping to chromatin states and comparative analysis

Previously filtered high-confidence BrdU-labeled replication tracks were used to assess replication dynamics across distinct chromatin environments. Chromatin annotations were based on a nine-state ChromHMM model, originally trained on the Drosophila dm3 genome and converted to dm6 coordinates. Regions that failed to map during the conversion process were designated as State 0. To ensure accurate chromatin state assignment, only BrdU tracks that were fully contained within a single chromatin state were included in the analysis. This was achieved using a strict containment criterion via bedtools intersect. For each retained BrdU track, we quantified (i) the length of the BrdU-labeled region, (ii) the total length of the parent DNA molecule, and (iii) the ratio of BrdU track length to read length as a normalized replication metric.

To simplify comparisons across different chromatin environments, the nine ChromHMM states were grouped into broader chromatin categories based on their functional characteristics. States 1 to 5, corresponding to transcriptionally active or accessible regions, were classified as euchromatin. States 0 and 6 to 8, typically associated with repressive or compact chromatin features, were categorized as heterochromatin. State 9, lacking hallmark chromatin modifications and generally considered transcriptionally inert, was labeled as a silent region. Differences between the euchromatin and heterochromatin groups were assessed using nonparametric Mann-Whitney *U* tests.

Chromatin state composition analysis of BrdU tracks

Previously chromatin state–labeled BrdU tracks were ranked by length and stratified into length-based percentile groups (Q1 to Q5) to enable balanced comparisons across replication track lengths. Within each length group, overlapping BrdU tracks were merged to generate nonredundant genomic intervals, thereby avoiding redundancy bias. For each merged interval, the proportion of sequence annotated as each chromatin state was calculated. Additional normalization strategies were applied, including abundance normalization based on raw percent coverage, to facilitate relative comparisons of chromatin state biases across length groups.

BrdU-labeled replication tracks were analyzed in relation to genome-wide RT annotations to investigate the relationship between track length and RT. DNAscent ForkSense output BED files containing BrdU track coordinates were used alongside RT annotation files, which assign genomic intervals to four timing groups (0 to 3) with associated timing scores.

BrdU track length analysis relative to RT

BrdU-labeled replication tracks were analyzed in relation to genome-wide RT annotations to investigate the relationship between track length and RT. Previous filtered high-confidence ForkSense BED files containing BrdU track coordinates were used alongside RT annotation files, which assign genomic intervals to four timing groups (0 to 3) with associated timing scores. Because of the small size of RT segments (~100 bp) relative to larger BrdU tracks, fixed nonoverlapping 30-kb timing windows were generated across each chromosome to provide a standardized spatial framework. Each 30-kb window was assigned a dominant RT group based on the largest cumulative overlap with the smaller timing segments. Specifically, for each window, the total overlap length of all small timing regions belonging to each timing group was calculated, and the group with the greatest total overlap was assigned as the dominant group. This approach ensures that the timing group covering most of the window's genomic space is selected. In addition, mean timing scores and region counts—representing the average timing score and number of small timing segments overlapping the window—were computed to characterize the timing signal strength and density per window. BrdU tracks were then mapped to these timing windows by identifying overlaps. For tracks overlapping multiple windows, the window with the greatest overlap was assigned to the track, ensuring accurate spatial mapping. Track length distributions were first compared across all four RT groups using the Kruskal-Wallis *H* test to assess overall differences. Subsequently, pairwise Mann-Whitney *U* tests were performed for specific group comparisons. Significance thresholds were set at $P \leq 0.05$. In addition, to explore the relationship between

RT and replication track size, scatterplots were generated with the *x* axis representing $\log_{10}$-transformed mean RT scores and the *y* axis representing $\log_{10}$-transformed BrdU track lengths. This visualization highlights genome-wide trends between RT and track size on a continuous scale. The analysis pipeline was implemented in Python 3 using pandas for data handling, matplotlib and seaborn for visualization, and scipy for statistical testing.

Copy number correlation

Copy number variance (CNV) data from Drosophila S2 cells (S2-DRSC-BO) (*75*) were converted to BED file format and converted from dm3 to dm6 genome builds using CrossMap BED (*76*) on usegalaxy.org (*77*) with dm3toDm6.over.chain.gz from the UCSC. The genome was then binned into 100-kb windows in Python, and for each window, the mean nonzero copy number was determined, and the number of replication forks overlapping with the region was counted.

Transcription levels for the S2 genome

PRO-seq data from Drosophila S2 cells (GSM1032758) (*48*) were processed to quantify gene-level transcription activity. Data containing RPKM values were converted from dm3 to dm6 genome builds using CrossMap BED (*76*) on usegalaxy.org (*77*) with dm3ToDm6.over.chain.gz from the UCSC. The files were converted from bedgraph to bigWig using bedGraphToBigWig (*78*). A BED file of all genes in the dm6 genome was generated from the dm6.gff. The converted data were overlapped with the BED file using UCSC's bigWigAverageOverBed tool. This was done in a strand-wise manner, and genes with zero RPKM values were filtered out, resulting in a BED file containing location, strand, and RPKM values for each gene. In the case of analysis involving E2F-regulated genes, only genes identified by Dimova *et al.* [see table 1 (A to C) in the study] were used to generate the BED file used for later overlap (*49*).

Transcriptional activity within BrdU tracks

BrdU-labeled replication tracks from filtered high-confidence forkSense BED files were analyzed alongside transcription activity (RPKM) data from Drosophila S2 cells. Tracks were divided into five length-based quantile groups (Q1 to Q5). For each individual track, the mean RPKM of overlapping genes was calculated and assigned in a strand-specific manner, with each track represented as a single data point. Genes with an RPKM of 0 were excluded from this analysis. This analysis was completed with all expressed genes and genes that had previously been defined as E2F1 or E2F2 dependent (*49*). The *x* axis corresponds to track length groups, and the *y* axis shows mean RPKM per track, enabling comparison of transcriptional activity across replication track lengths. Statistical differences between groups were tested using Mann-Whitney *U* tests. Transcription distributions were visualized with box plots and cumulative distribution functions on logarithmic scales. Analyses were conducted in Python 3 using pandas, intervaltree, scipy, matplotlib, and seaborn.

Analysis of replication fork length and transcriptional activity by chromatin state

Genome-wide analysis used all BrdU tracks that were not part of the chromatin state fitting and overlapped at least one gene with nonzero RPKM. For chromatin state analysis, BrdU track coordinates with chromatin state annotations were obtained from previously generated single-state fits. Only tracks overlapping at least one gene with nonzero RPKM were retained. For each track, the mean RPKM of all overlapping genes was computed. Tracks were grouped by chromatin state (0 to 9) and by broader functional categories: euchromatin (States 1 to 5), heterochromatin (States 0 and 6 to 8), and silent intergenic (State 9). Scatterplots were generated for each group, plotting mean RPKM versus BrdU length. Linear regression lines and Pearson correlation coefficients were calculated to assess the relationship between active transcription and replication fork length.

Analysis of replication fork length and transcriptional activity by orientation

Replication-transcription orientation analysis was performed on BrdU-labeled replication tracks categorized as head-on or codirectional based on the direction of fork movement and overlapping gene transcription. Fork direction was obtained from DNAscent forkSense output, and gene annotations with strand and RPKM values were used to define orientation: rightward-moving forks overlapping leftward-transcribed genes (−) and leftward-moving forks overlapping rightward-transcribed genes (+) were classified as head-on, with the inverse defined as codirectional. For each track, mean RPKM of overlapping genes was calculated, and active transcription was assessed by plotting replication fork length against the mean RPKM. Scatterplots were generated separately for head-on and codirectional forks, with RPKM on the *x* axis and fork length on the *y* axis. Pearson correlation coefficients were computed for each group. This analysis was performed using both genome-wide forks and on just euchromatic forks as defined by the Analysis of replication fork length and transcriptional activity by chromatin state analysis.

Defining origin and termination sites

Origin and termination BED files produced by DNAscent forkSense for each sequencing run were combined per run and then concatenated across runs to create comprehensive origin and termination datasets. Origin and termination events were defined by taking the midpoint of each called origin or termination event and adding 100 bp to each side, resulting in a uniform set of origin and termination sites that were all 200 bp. These final center-coordinate BED files were generated for downstream analyses using Python (v3.12) and pandas (v2.2.2) (*79*).

RT data preparation

RT data were retrieved from a published study [GSE17280 (*6*)]. The chromosome, start, stop, and average of the three M-values for each entry were extracted and exported as a bed file. This bed file was converted from dm3 to dm6 using CrossMap (*76*) BED on usegalaxy.org (*77*). For conversion, the liftover file from the UCSC was used (dm3ToDm6.over.chain.gz).

RT distance analysis

RT regions were grouped into quartiles, merging adjacent intervals within the same quartile to ensure coverage. The centers of adjacent inter-origin or intertermination events are used to calculate the inter-event distance. The midpoint of each distance is then used to classify that distance into an RT quartile by determining the overlap of the midpoint with the RT bed file. Distances with midpoints that do not overlap with any RT region are discarded for this analysis. The distances are then grouped by quartile and the earliest and latest quartile groups are plotted using seaborn (v0.13.2) violin plots (*80*). The mean and median for each quartile are calculated, and the means are compared with Student's *t* test.

RT density analysis

Origin and termination events are classified into RT quartiles by dividing the RT bed file into quartiles, and adjacent reads in the same quartile are merged (this is necessary to ensure sufficient genome coverage when assigning replication quartiles based on single base overlaps). For each event, we assign the event an RT quartile based on the overlap of the RT bed file with the center of the event. Events where the center does not overlap with any region in the RT bed file

were discarded. Once the RT quartile is assigned, the total number of events in each quartile is summed and divided by the total genomic length mapped by that RT quartile.

Chromatin state distance analysis

The centers of adjacent events are used to calculate the inter-event distance. The midpoint of each distance is then used to classify that distance into a chromatin state by determining the overlap of the midpoint base with the chromatin state bed file. The distances in States 1 to 5 are grouped as euchromatin, and the distances in States 6 to 8 and 0 are grouped as heterochromatin and plotted as violin plots. The mean and median for each group are calculated, and the means are compared with Student's *t* test.

Chromatin state density analysis

We assign a chromatin state to each event based on the overlap of the center of that event with regions in the chromatin state bed file. If an event does not overlap with any region in the chromatin state bed file, it is discarded for this analysis. Once the chromatin state is assigned, events in States 1 to 5 are grouped as euchromatin and events in States 6 to 8 and 0 are grouped as heterochromatin events. The total number of events in each group is summed and divided by the total genomic length mapped by the chromatin state bed file for each grouping.

ORC2 data acquisition, preparation, and analysis

ORC2 ChIP sequencing (ChIP-seq) data were downloaded from the GEO (GSE20887) as GFF3 files, converted to BED format, and converted from the dm3 to dm6 assembly using CrossMap with UCSC chain files. The processed ORC2 peak BED file was used for downstream analyses, including interpeak distance calculations and density comparisons substituting for origin event files. Inter-ORC2 distance is calculated as the distance between the centers of adjacent ORC2 peaks and plotted as a violin plot using seaborn.

Random permutation origin-ORC2 distances

For each origin, the distance to the nearest ORC2 peak is calculated and plotted as a histogram on a $\log_{10}$ scale (one is added to each distance to prevent logarithmic compute errors). A randomized imitation dataset is then generated containing the same number of events per chromosome as the true dataset but is distributed randomly along the genome. The distance from each imitated event to the nearest ORC2 peak is calculated and stored. This process of dataset imitation and distance calculation is repeated 1000 times, and the normalized results are plotted as a histogram on a log scale. The mean and median for the original and imitation dataset are calculated, and a Student's *t* test was performed.

MNase heatmaps

Micrococcal nuclease sequencing (MNase-seq) data were obtained from the GEO (GSE128689), converted from TDF to bedgraph format using igvtools tdftobedgraph (v2.17) (*81*), and analyzed with deepTools (*82*) computeMatrix and plotHeatmap on usegalaxy.org to visualize nucleosome occupancy centered on origins and ORC2 peaks within 2-kb windows.

Overlaps: Expected versus observed

We compute the number of overlaps found within the origin dataset and compare that to the number of overlaps expected from a set of 3593 achieved by simulated, random, independent sampling equal to the amount of the original origin dataset (1513). We calculated 3593 as the expected number of mappable origins in the S2 genome given a mappable genome size of 143,726,002 bp divided by an average inter-origin distance of 40 kb (*36*). This was achieved in Python using a series ranging from 0 to 3593. Values from the series were independently sampled from that 1513 times. The number of repeated sampling events was recorded as the number of expected overlaps for that simulation. The simulation and sampling process was repeated 1000 times, and the mean and SD of the expected number of overlaps was computed.

Random permutation analysis

The random permutation analysis was performed as previously described (*9*, *60*). Briefly, ChIP-chip and ChIP-seq datasets were downloaded from modENCODE, and for each factor, base pair overlap was calculated between the origin peaks or the termination peaks. For each ChIP-chip or ChIP-seq marker, we calculated the number of overlapping base pairs between the marker and the origin/termination centers. We applied a permutation-based approach to determine whether the observed amount of overlap was more than what was expected by chance. Briefly, we calculated an empirical *P* value for the observed amount of overlap by comparing the number of overlapping base pairs to a simulated null distribution. We simulated the null distribution by randomly shuffling regions throughout the genome and calculating the amount of overlap for each permutation. The *P* values were adjusted for multiple testing using the Bonferroni method.

When permuting, we matched the number and length distribution of the shuffled chromatin markers to the original set of markers and excluded all blacklisted regions from consideration. For peaks obtained from ChIP-chip data, we required that the shuffled peaks maintained both the overall length distribution and the probe density of the original peak. We reshuffled any peaks that were more than 0.1 units away from the original probe density until at least 99% of the original peaks were appropriately matched. We performed 1000 permutations for each marker and set of origin or termination centers.

UpSet diagram of origins, ORC2, and SNS-seq

SNS-seq data were retrieved from a published study (GSE65692) and converted to dm6 using CrossMap BED on usegalaxy.org. For conversion, the liftover file from the UCSC was used (dm3ToDm6.over.chain.gz). This bed file, the origins bed file, and the previously converted ORC2 ChIP-seq bed file were inputted into the UpSet (*83*) diagram on usegalaxy.org.

Nucleotide frequency and motif enrichment

HOMER (*84*) (v5.1) was installed locally to evaluate nucleotide bias or motif enrichment about origins. findMotifsGenome.pl was used to identify potential sequence motifs in the origin bed files. The tool was run with default inputs, with the origin bed file as the position file and dm6 as the reference genome.

annotatePeaks.pl with the -annStats tag was used to determine origin enrichment in the various genomic regions. The –hist and –nuc tags were used to determine nucleotide frequency about the origins. The AC genome frequency determined using the UCSC faCount (v482) tool and was plotted as dashed lines.

Supplementary Materials

This PDF file includes:

Figs. S1 to S7

REFERENCES

1. M. W. Parker, M. R. Botchan, J. M. Berger, Mechanisms and regulation of DNA replication initiation in eukaryotes. *Crit. Rev. Biochem. Mol. Biol.* **52**, 107–144 (2017).
2. B. Stillman, J. F. X. Diffley, J. H. Iwasa, Mechanisms for licensing origins of DNA replication in eukaryotic cells. *Nat. Struct. Mol. Biol.* **32**, 1143–1153 (2025).
3. B. L. Thakur, A. Ray, C. E. Redon, M. I. Aladjem, Preventing excess replication origin activation to ensure genome stability. *Trends Genet.* **38**, 169–181 (2022).

4. B. Miotto, Z. Ji, K. Struhl, Selectivity of ORC binding sites and the relation to replication timing, fragile sites, and deletions in cancers. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E4810–E4819 (2016).
5. H. K. MacAlpine, R. Gordân, S. K. Powell, A. J. Hartemink, D. M. MacAlpine, Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading. *Genome Res.* **20**, 201–211 (2010).
6. M. L. Eaton, J. A. Prinz, H. K. MacAlpine, G. Tretyakov, P. V. Kharchenko, D. M. MacAlpine, Chromatin signatures of the Drosophila replication program. *Genome Res.* **21**, 164–174 (2011).
7. J. R. Lipford, S. P. Bell, Nucleosomes positioned by ORC facilitate the initiation of DNA replication. *Mol. Cell* **7**, 21–30 (2001).
8. M. L. Eaton, K. Galani, S. Kang, S. P. Bell, D. M. MacAlpine, Conserved nucleosome positioning defines replication origins. *Gene Dev* **24**, 748–753 (2010).
9. L. Richards, C. L. Lord, M. L. Benton, J. A. Capra, J. T. Nordman, Nucleoporins facilitate ORC loading onto chromatin. *Cell Rep.* **41**, 111590 (2022).
10. Y. Zhang, L. Huang, H. Fu, O. K. Smith, C. M. Lin, K. Utani, M. Rao, W. C. Reinhold, C. E. Redon, M. Ryan, R. Kim, Y. You, H. Hanna, Y. Boisclair, Q. Long, M. I. Aladjem, A replicator-specific binding protein essential for site-specific initiation of DNA replication in mammalian cells. *Nat. Commun.* **7**, 11748 (2016).
11. Z. Shen, K. M. Sathyan, Y. Geng, R. Zheng, A. Chakraborty, B. Freeman, F. Wang, K. V. Prasanth, S. G. Prasanth, A WD-repeat protein stabilizes ORC binding to chromatin. *Mol. Cell* **40**, 99–111 (2010).
12. Y. Wang, A. Khan, A. B. Marks, O. K. Smith, S. Giri, Y.-C. Lin, R. Creager, D. M. MacAlpine, K. V. Prasanth, M. I. Aladjem, S. G. Prasanth, Temporal association of ORCA/LRWD1 to late-firing origins during G1 dictates heterochromatin replication and organization. *Nucleic Acids Res.* **45**, gkw1211 (2016).
13. S. Giri, V. Aggarwal, J. Pontis, Z. Shen, A. Chakraborty, A. Khan, C. Mizzen, K. V. Prasanth, S. Ait-Si-Ali, T. Ha, S. G. Prasanth, The preRC protein ORCA organizes heterochromatin by assembling histone H3 lysine 9 methyltransferases on chromatin. *Elife* **4**, e06496 (2015).
14. M. Tian, Z. Wang, Z. Su, E. Shibata, Y. Shibata, A. Dutta, C. Zang, Integrative analysis of DNA replication origins and ORC-/MCM-binding sites in human cells reveals a lack of overlap. *Elife* **12**, RP89548 (2024).
15. D. J. Smith, I. Whitehouse, Intrinsic coupling of lagging-strand synthesis to chromatin assembly. *Nature* **483**, 434–438 (2012).
16. N. Petryk, M. Kahli, Y. d'Aubenton-Carafa, Y. Jaszczyszyn, Y. Shen, M. Silvain, C. Thermes, C.-L. Chen, O. Hyrien, Replication landscape of the human genome. *Nat. Commun.* **7**, 10208 (2016).
17. A. R. Langley, S. Gräf, J. C. Smith, T. Krude, Genome-wide identification and characterisation of human DNA replication origins by initiation site sequencing (ini-seq). *Nucleic Acids Res.* **44**, 10230–10247 (2016).
18. P. A. Zhao, T. Sasaki, D. M. Gilbert, High-resolution Repli-Seq defines the temporal choreography of initiation, elongation and termination of replication in mammalian cells. *Genome Biol.* **21**, 76 (2020).
19. W. Wang, K. N. Klein, K. Proesmans, H. Yang, C. Marchal, X. Zhu, T. Borrman, A. Hastie, Z. Weng, J. Bechhoefer, C.-L. Chen, D. M. Gilbert, N. Rhind, Genome-wide mapping of human DNA replication by optical replication mapping supports a stochastic model of eukaryotic replication. *Mol. Cell* **81**, 2975–2988.e6 (2021).
20. M. S. Foulk, J. M. Urban, C. Casella, S. A. Gerbi, Characterizing and controlling intrinsic biases of lambda exonuclease in nascent strand sequencing reveals phasing between nucleosomes and G-quadruplex motifs around a subset of human replication origins. *Genome Res.* **25**, 725–735 (2015).
21. L. D. Mesner, V. Valsakumar, M. Cieślik, R. Pickin, J. L. Hamlin, S. Bekiranov, Bubble-seq analysis of the human genome reveals distinct chromatin-mediated mechanisms for regulating early- and late-firing origins. *Genome Res.* **23**, 1774–1788 (2013).
22. M. Hennion, J.-M. Arbona, L. Lacroix, C. Cruaud, B. Theulot, B. L. Tallec, F. Proux, X. Wu, E. Novikova, S. Engelen, A. Lemainque, B. Audit, O. Hyrien, FORK-seq: Replication landscape of the *Saccharomyces cerevisiae* genome by nanopore sequencing. *Genome Biol.* **21**, 125 (2020).
23. J. T. Carrington, R. H. C. Wilson, E. de L. Vega, S. Thiyagarajan, T. Barker, L. Catchpole, A. Durrant, V. Knitlhoffer, C. Watkins, K. Gharbi, C. A. Nieduszynski, Most human DNA replication initiation is dispersed throughout the genome with only a minority within previously identified initiation zones. *Genome Biol.* **26**, 122 (2025).
24. C. A. Müller, M. A. Boemo, P. Spingardi, B. M. Kessler, S. Kriaucionis, J. T. Simpson, C. A. Nieduszynski, Capturing the dynamics of genome replication on individual ultra-long nanopore sequence reads. *Nat. Methods* **16**, 429–436 (2019).
25. X. Michalet, R. Ekong, F. Fougerousse, S. Rousseaux, C. Schurra, N. Hornigold, M. van Slegtenhorst, J. Wolfe, S. Povey, J. S. Beckmann, A. Bensimon, Dynamic molecular combing: Stretching the whole human genome for high-resolution studies. *Science* **277**, 1518–1523 (1997).
26. J. N. Bianco, J. Poli, J. Saksouk, J. Bacal, M. J. Silva, K. Yoshida, Y.-L. Lin, H. Tourrière, A. Lengronne, P. Pasero, Analysis of DNA replication profiles in budding yeast and mammalian cells using DNA combing. *Methods* **57**, 149–157 (2012).
27. R. Lebofsky, A. Bensimon, Single DNA molecule analysis: Applications of molecular combing. *Brief. Funct. Genom.* **1**, 385–396 (2003).
28. G. Moore, J. J. Sainz, R. B. Jensen, DNA fiber combing protocol using in-house reagents and coverslips to analyze replication fork dynamics in mammalian cells. *STAR Protoc.* **3**, 101371 (2022).
29. A. Munden, M. T. Wright, D. Han, R. Tirgar, L. Plate, J. T. Nordman, Identification of replication fork-associated proteins in Drosophila embryos and cultured cells using iPOND coupled to quantitative mass spectrometry. *Sci. Rep.* **12**, 6903 (2022).
30. D. Georgieva, Q. Liu, K. Wang, D. Egli, Detection of base analogs incorporated during DNA replication by nanopore sequencing. *Nucleic Acids Res.* **48**, e88–e88 (2020).
31. B. Theulot, L. Lacroix, J.-M. Arbona, G. A. Millot, E. Jean, C. Cruaud, J. Pellet, F. Proux, M. Hennion, S. Engelen, A. Lemainque, B. Audit, O. Hyrien, B. Le Tallec, Genome-wide mapping of individual replication fork velocities using nanopore sequencing. *Nat. Commun.* **13**, 3295 (2022).
32. F. I. G. Totañes, J. Gockel, S. E. Chapman, R. Bártfai, M. A. Boemo, C. J. Merrick, A genome-wide map of DNA replication at single-molecule resolution in the malaria parasite Plasmodium falciparum. *Nucleic Acids Res.* **51**, 2709–2724 (2023).
33. M. J. K. Jones, S. K. Rai, P. L. Pfuderer, A. Bonfim-Melo, J. K. Pagan, P. R. Clarke, F. I. G. Totañes, C. J. Merrick, S. E. McClelland, M. A. Boemo, A high-resolution, nanopore-based artificial intelligence assay for DNA replication stress in human cancer cells. *Nat. Commun.* **16**, 7732 (2025).
34. O. Hyrien, G. Guilbaud, T. Krude, The double life of mammalian DNA replication origins. *Genes Dev.* **39**, 304–324 (2025).
35. M. A. Boemo, DNAscent v2: Detecting replication forks in nanopore sequencing data with deep learning. *BMC Genomics* **22**, 430 (2021).
36. G. M. Alvino, D. Collingwood, J. M. Murphy, J. Delrow, B. J. Brewer, M. K. Raghuraman, Replication in hydroxyurea: It's a matter of time. *Mol. Cell. Biol.* **27**, 6396–6406 (2007).
37. Y. Zhang, J. H. Malone, S. K. Powell, V. Periwal, E. Spana, D. M. MacAlpine, B. Oliver, Expression in aneuploid *Drosophila* S2 cells. *PLOS Biol.* **8**, e1000320 (2010).
38. A. B. Blumenthal, H. J. Kriegstein, D. S. Hogness, The units of DNA replication in *Drosophila melanogaster* chromosomes. *Cold Spring Harb. Symp. Quant. Biol.* **38**, 205–223 (1974).
39. P. V. Kharchenko, A. A. Alekseyenko, Y. B. Schwartz, A. Minoda, N. C. Riddle, J. Ernst, P. J. Sabo, E. Larschan, A. A. Gorchakov, T. Gu, D. Linder-Basso, A. Plachetka, G. Shanower, M. Y. Tolstorukov, L. J. Luquette, R. Xi, Y. L. Jung, R. W. Park, E. P. Bishop, T. K. Canfield, R. Sandstrom, R. E. Thurman, D. M. MacAlpine, J. A. Stamatoyannopoulos, M. Kellis, S. C. R. Elgin, M. I. Kuroda, V. Pirrotta, G. H. Karpen, P. J. Park, Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* **471**, 480–485 (2011).
40. G. dos Santos, A. J. Schroeder, J. L. Goodman, V. B. Strelets, M. A. Crosby, J. Thurmond, D. B. Emmert, W. M. Gelbart, Flybase Consortium, FlyBase: Introduction of the *Drosophila melanogaster* release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Res.* **43**, D690–D697 (2015).
41. J. C. Rivera-Mulia, D. M. Gilbert, Replication timing and transcriptional control: Beyond cause and effect—Part III. *Curr. Opin. Cell Biol.* **40**, 168–178 (2016).
42. D. M. MacAlpine, G. Almouzni, Chromatin and DNA Replication. *Cold Spring Harb. Perspect. Biol.* **5**, a010207 (2013).
43. S. P. Bell, A. Dutta, DNA replication in eukaryotic cells. *Annu. Rev. Biochem.* **71**, 333–374 (2002).
44. M. Fragkos, O. Ganier, P. Coulombe, M. Méchali, DNA replication origin activation in space and time. *Nat. Rev. Mol. Cell Biol.* **16**, 360–374 (2015).
45. J. van den Berg, V. van Batenburg, C. Geisenberger, R. B. Tjeerdsma, A. de Jaime-Soguero, S. P. Acebrón, M. A. T. M. van Vugt, A. van Oudenaarden, Quantifying DNA replication speeds in single cells by scEdU-seq. *Nat. Methods* **21**, 1175–1184 (2024).
46. L. Goehring, T. T. Huang, D. J. Smith, Transcription–replication conflicts as a source of genome instability. *Annu. Rev. Genet.* **57**, 157–179 (2023).
47. P. Rojas, J. Wang, G. Guglielmi, M. M. Sadurnì, L. Pavlou, G. H. D. Leung, V. Rajagopal, F. Spill, M. Saponaro, Genome-wide identification of replication fork stalling/pausing sites and the interplay between RNA Pol II transcription and DNA replication progression. *Genome Biol.* **25**, 126 (2024).
48. H. Kwak, N. J. Fuda, L. J. Core, J. T. Lis, Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**, 950–953 (2013).
49. D. K. Dimova, O. Stevaux, M. V. Frolov, N. J. Dyson, Cell cycle-dependent and cell cycle-independent control of transcription by the Drosophila E2F/RB pathway. *Genes Dev.* **17**, 2308–2320 (2003).
50. S. S. David, B. A. Pacheco, K. Kishimoto, S. Vantine, K. Hu, H. Liu, D. L. Davis, H. Tran, B. F. Sallis, L. Ali, C. M. Haynes, B. A. McCormick, L. J. Zhu, W. A. Flavahan, Multidimensional third-generation sequencing of modified DNA bases allows interrogation of complex biological systems. *Nat. Commun.* **16**, 5676 (2025).
51. A. F. Phillips, A. R. Millet, M. Tigano, S. M. Dubois, H. Crimmins, L. Babin, M. Charpentier, M. Piganeau, E. Brunet, A. Sfeir, Single-molecule analysis of mtDNA replication uncovers the basis of the common deletion. *Mol. Cell* **65**, 527–538.e6 (2017).
52. M. Meselson, F. W. Stahl, The replication of DNA. *Cold Spring Harb. Symp. Quant. Biol.* **23**, 9–12 (1958).

53. K. A. Gavin, M. Hidaka, B. Stillman, Conserved initiator proteins in eukaryotes. *Science* **270**, 1667–1671 (1995).
54. A. Correspondent, Eukaryotic DNA replication. *Nature* **241**, 10–11 (1973).
55. M. M. Shareef, C. King, M. Damaj, R. Badagu, D. W. Huang, R. Kellum, Drosophila heterochromatin protein 1 (HP1)/origin recognition complex (ORC) protein is associated with HP1 and ORC and functions in heterochromatin-induced silencing. *Mol. Biol. Cell* **12**, 1671–1685 (2001).
56. M. Gossen, D. T. S. Pak, S. K. Hansen, J. K. Acharya, M. R. Botchan, A Drosophila homolog of the yeast origin recognition complex. *Science* **270**, 1674–1677 (1995).
57. D. T. S. Pak, M. Pflumm, I. Chesnokov, D. W. Huang, R. Kellum, J. Marr, P. Romanowski, M. R. Botchan, Association of the origin recognition complex with heterochromatin and HP1 in higher eukaryotes. *Cell* **91**, 311–323 (1997).
58. A.-K. Bielinsky, H. Blitzblau, E. L. Beall, M. Ezrokhi, H. S. Smith, M. R. Botchan, S. A. Gerbi, Origin recognition complex binding to a metazoan replication origin. *Curr. Biol.* **11**, 1427–1431 (2001).
59. S. E. Celniker, L. A. L. Dillon, M. B. Gerstein, K. C. Gunsalus, S. Henikoff, G. H. Karpen, M. Kellis, E. C. Lai, J. D. Lieb, D. M. MacAlpine, G. Micklem, F. Piano, M. Snyder, L. Stein, K. P. White, R. H. Waterston, modENCODE Consortium, Unlocking the secrets of the genome. *Nature* **459**, 927–930 (2009).
60. A. Munden, M. L. Benton, J. A. Capra, J. T. Nordman, R-loop mapping and characterization during Drosophila embryogenesis reveals developmental plasticity in R-loop signatures. *J. Mol. Biol.* **434**, 167645 (2022).
61. D. Han, S. H. Schaffner, J. P. Davies, M. L. Benton, L. Plate, J. T. Nordman, BRWD3 promotes KDM5 degradation to maintain H3K4 methylation levels. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2305092120 (2023).
62. J. Sequeira-Mendes, R. Díaz-Uriarte, A. Apedaile, D. Huntley, N. Brockdorff, M. Gómez, Transcription initiation activity sets replication origin efficiency in mammalian cells. *PLOS Genet.* **5**, e1000446 (2009).
63. C. Cayrou, B. Ballester, I. Peiffer, R. Fenouil, P. Coulombe, J.-C. Andrau, J. van Helden, M. Méchali, The chromatin environment shapes DNA replication origin organization and defines origin classes. *Genome Res.* **25**, 1873–1885 (2015).
64. F. Comoglio, T. Schlumpf, V. Schmid, R. Rohs, C. Beisel, R. Paro, High-resolution profiling of drosophila replication start sites reveals a DNA shape and chromatin signature of metazoan origins. *Cell Rep.* **11**, 821–834 (2015).
65. X. Bi, Heterochromatin structure: Lessons from the budding yeast. *IUBMB Life* **66**, 657–666 (2014).
66. A. Panday, A. Grove, Yeast HMO1: Linker histone reinvented. *Microbiol. Mol. Biol. Rev.* **81**, e00037-16 (2016).
67. H. Merrikh, C. Machón, W. H. Grainger, A. D. Grossman, P. Soultanas, Co-directional replication–transcription conflicts lead to replication restart. *Nature* **470**, 554–557 (2011).
68. H. Stoy, K. Zwicky, D. Kuster, K. S. Lang, J. Krietsch, M. P. Crossley, J. A. Schmid, K. A. Cimprich, H. Merrikh, M. Lopes, Direct visualization of transcription-replication conflicts reveals post-replicative DNA:RNA hybrids. *Nat. Struct. Mol. Biol.* **30**, 348–359 (2023).
69. J. Malínsky, K. Koberna, D. Staněk, M. Masata, I. Votruba, I. Raska, The supply of exogenous deoxyribonucleotides accelerates the speed of the replication fork in early S-phase. *J. Cell Sci.* **114**, 747–750 (2001).
70. J. Poli, O. Tsaponina, L. Crabbé, A. Keszthelyi, V. Pantesco, A. Chabes, A. Lengronne, P. Pasero, dNTP pools determine fork progression and origin usage under replication stress. *EMBO J.* **31**, 883–894 (2012).
71. Y. Liu, X. Wu, Y. d'Aubenton-Carafa, C. Thermes, C.-L. Chen, OKseqHMM: A genome-wide replication fork directionality analysis toolkit. *Nucleic Acids Res.* **51**, e22 (2023).
72. I. Akerman, B. Kasaai, A. Bazarova, P. B. Sang, I. Peiffer, M. Artufel, R. Derelle, G. Smith, M. Rodriguez-Martinez, M. Romano, S. Kinet, P. Tino, C. Theillet, N. Taylor, B. Ballester, M. Méchali, A predictable conserved DNA base composition signature defines human core DNA replication origins. *Nat. Commun.* **11**, 4826 (2020).
73. E. Besnard, A. Babled, L. Lapasset, O. Milhavet, H. Parrinello, C. Dantec, J.-M. Marin, J.-M. Lemaitre, Unraveling cell type–specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat. Struct. Mol. Biol.* **19**, 837–844 (2012).
74. S. Thiyagarajan, A. M. Rogers, C. A. Müller, C. A. Nieduszynski, Single-molecule landscape of DNA replication pausing. bioRxiv 2025.08.14.670160 [Preprint] (2025). https://doi.org/10.1101/2025.08.14.670160.
75. H. Lee, C. J. McManus, D.-Y. Cho, M. Eaton, F. Renda, M. P. Somma, L. Cherbas, G. May, S. Powell, D. Zhang, L. Zhan, A. Resch, J. Andrews, S. E. Celniker, P. Cherbas, T. M. Przytycka, M. Gatti, B. Oliver, B. Graveley, D. MacAlpine, DNA copy number evolution in *Drosophila* cell lines. *Genome Biol.* **15**, R70 (2014).
76. H. Zhao, Z. Sun, J. Wang, H. Huang, J.-P. Kocher, L. Wang, CrossMap: A versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30**, 1006–1007 (2014).
77. The Galaxy Community, The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Res.* **52**, W83–W94 (2024).
78. W. J. Kent, A. S. Zweig, G. Barber, A. S. Hinrichs, D. Karolchik, BigWig and BigBed: Enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204–2207 (2010).
79. The pandas development team, pandas-dev/pandas: Pandas, version v2.2.2, Zenodo (2024); https://doi.org/10.5281/zenodo.10957263.
80. M. L. Waskom, seaborn: Statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).
81. J. T. Robinson, H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, J. P. Mesirov, Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
82. F. Ramírez, D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dündar, T. Manke, deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
83. A. Khan, A. Mathelier, Intervene: A tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics* **18**, 287 (2017).
84. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, C. K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

Acknowledgments: We thank D. Cortez and R. Bhowmick for critically reading the manuscript. We are indebted to R. Bhowmick for sharing results before publication. **Funding:** This research was supported by NIH grant 2R35GM128650 (to J.T.N.). D.S. was supported by F99AG083284 from the NIA. The initial phase of this work was made possible through the National Cancer Institute (NCI) Vanderbilt-Ingram Cancer Center Support Grant (P30CA068485). This research used the Delta advanced computing and data resource, which is supported by the National Science Foundation (award OAC 2005572) and the State of Illinois. Delta is a joint effort of the University of Illinois Urbana-Champaign and its National Center for Supercomputing Applications. **Author contributions:** Conceptualization: D.H. and J.T.N. Methodology: D.H. and C.S. Software: D.H., C.S., M.L.B. Validation: D.H. and C.S. Formal analysis: D.H., C.S., and M.L.B. Investigation: D.H. and C.S. Resources: D.H. Data curation: D.H. and C.S. Writing—original draft: D.H., C.S., and J.T.N. Writing—review and editing: D.H., C.S., and J.T.N. Visualization: D.H., C.S., and M.L.B. Funding acquisition: J.T.N. Supervision: J.T.N. Project administration: J.T.N. Funding acquisition: J.T.N. **Competing interests:** The authors declare that they have no competing interests. **Data, code, and materials availability:** All data and code needed to evaluate and reproduce the conclusions in the paper are present in the paper and/or the Supplementary Materials. All sequencing data have been deposited into the GEO under the accession number GSE309672. No new materials were generated in this study.

Submitted 22 October 2025
Accepted 27 March 2026
Published 1 May 2026
10.1126/sciadv.aed2806