Genome-wide measurement of DNA replication fork directionality and quantification of DNA replication initiation and termination with Okazaki fragment sequencing

Xia Wu, Yaqun Liu, Yves D'aubenton-Carafa, Claude Thermes, Olivier Hyrien, Chun-Long Chen, Nataliya Petryk

nature protocols

PROTOCOL

https://doi.org/10.1038/s41596-022-00793-5

Check for updates

Genome-wide measurement of DNA replication fork directionality and quantification of DNA replication initiation and termination with Okazaki fragment sequencing

Xia Wu[1,7], Yaqun Liu[2,7], Yves d'Aubenton-Carafa[3], Claude Thermes[3], Olivier Hyrien[4]✉, Chun-Long Chen[2]✉ and Nataliya Petryk[5,6]✉

Studying the dynamics of genome replication in mammalian cells has been historically challenging. To reveal the location of replication initiation and termination in the human genome, we developed Okazaki fragment sequencing (OK-seq), a quantitative approach based on the isolation and strand-specific sequencing of Okazaki fragments, the lagging strand replication intermediates. OK-seq quantitates the proportion of leftward- and rightward-oriented forks at every genomic locus and reveals the location and efficiency of replication initiation and termination events. Here we provide the detailed experimental procedures for performing OK-seq in unperturbed cultured human cells and budding yeast and the bioinformatics pipelines for data processing and computation of replication fork directionality. Furthermore, we present the analytical approach based on a hidden Markov model, which allows automated detection of ascending, descending and flat replication fork directionality segments revealing the zones of replication initiation, termination and unidirectional fork movement across the entire genome. These tools are essential for the accurate interpretation of human and yeast replication programs. The experiments and the data processing can be accomplished within 6 d. Besides revealing the genome replication program in fine detail, OK-seq has been instrumental in numerous studies unravelling mechanisms of genome stability, epigenome maintenance and genome evolution.

Introduction

DNA fiber autoradiographic studies of mammalian cells showed long ago that eukaryotic DNA replication origins are spaced at 20–400 kb intervals and fire at different times in S phase[1]. However, mapping origins in metazoan cells has been historically challenging, due to the lack of workable genetic assays and the difficulties in purifying sufficient amounts of intact DNA replication initiation intermediates (for reviews, see refs. [2–4]).

In the pre-genomic era, early studies of the highly amplified Chinese Hamster Ovary DHFR locus identified a few specific initiation sites downstream of the DHFR gene. However, more extensive studies demonstrated that replication could initiate at any of a large number of sites over a broad (55 kb) zone downstream of the gene, at a global rate lower than one initiation event per cell cycle, even in cells with only a single copy of the locus[2]. Depending on the technique(s) used to purify initiation intermediates from cell populations, site-specific or dispersed initiation was also reported at a few other model loci[3]. Direct visualization of replication fork progression at the single DNA molecule level using DNA combing[5] or single molecule analysis of replicated DNA[6] revealed broad (3–100 kb) initiation zones (IZs), although site-specific origins were also reported[7]. It was unclear whether these variable results reflected the true genomic diversity of replication origins or different technical biases.

The advent of DNA microarrays and high-throughput sequencing has allowed much broader and more systematic scrutiny of origins. Crucially, different pictures were obtained depending on the

[1]Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China. [2]Institut Curie, Université PSL, Sorbonne Université, CNRS UMR3244, Dynamics of Genetic Information, Paris, France. [3]Institute for Integrative Biology of the Cell (I2BC), Université Paris-Saclay, CEA, CNRS, Gif-sur-Yvette, France. [4]Institut de Biologie de l'École Normale Supérieure (IBENS), École Normale Supérieure, CNRS, Inserm, Université PSL, Paris, France. [5]Epigenetics & Cell Fate CNRS UMR7216 Université Paris-Cité, Paris, France. [6]Present address: Institut Gustave Roussy, Université Paris-Saclay CNRS UMR9019, Genome Stability and Cancers, Villejuif, France. [7]These authors contributed equally: Xia Wu; Yaqun Liu. ✉e-mail: olivier.hyrien@bio.ens.psl.eu; chunlong.chen@curie.fr; nataliya.petryk@gustaveroussy.fr

technique used to purify initiation intermediates. Small nascent strands (SNS) synthesized at origins were purified by size selection, followed by λ-exonuclease digestion of the contaminating broken DNA strands lacking a protecting 5′ RNA primer (λ-SNS)[8,9], or by briefly labeling newly synthesized DNA with 5-bromo-2′-deoxyuridine (BrdU) or digoxigenin-dUTP, followed by size selection and immunoprecipitation[10–12]. Replication bubble-containing restriction fragments were purified by trapping in gelling agarose and electrophoretic elimination of bubble-devoid fragments[13,14]. These independent approaches to purify initiation intermediates often gave poorly concordant origin locations. Furthermore, SNS tended to highlight site-specific origins whereas bubbles revealed broad IZs. Lastly, no information about fork progression and termination could be obtained by these approaches.

Replication timing (RT) and replication fork directionality (RFD) profiling are orthogonal approaches to study DNA replication. They do not depend on isolating initiation intermediates but on analysis of the replication behaviours of all investigated loci. In Repli-seq, after pulse labeling with BrdU for 30–120 min, S-phase cells are sorted by FACS into two to six fractions based on total DNA content. Next, BrdU-DNA is immunoprecipitated and hybridized to microarrays or sequenced, allowing to generate global replication timing profiles[15,16]. In a recent improvement called high-resolution Repli-seq, up to 16 fractions of S-phase cells were used[17]. A second approach for measuring replication timing is based on assaying DNA copy number by sequencing sorted S and G1 cells, or even unsorted asynchronously proliferating cells[18]. Importantly, Repli-seq and copy number profile-based methods are highly consistent with each other and extremely reproducible between laboratories[19]. They identify peaks and valleys of early- and late-replicating DNA, respectively, but unlike in yeast, their spatial and temporal resolution (~2 h and ~100 kb, respectively) is insufficient to precisely map origins in mammals. In high-resolution Repli-seq, however, the resolution was improved to 50 kb, allowing the identification of some isolated IZs[17].

Genome-wide RFD profiles were first obtained by analysis of nucleotide strand compositional asymmetries defined as the skew $S = (G - C)/(G + C) + (T - A)/(T + A)$ (i.e., the relative excess of G over C and T over A on a single DNA strand), following the seminal observation that S sign abruptly changes at bacterial origins and termini[20]. Analysis of mammalian genomes revealed ~1,000 abrupt S upshifts similar to those at bacterial origins, separated by megabase-sized segments of progressive S decrease, tracing N-shaped domains of S (refs. [21,22]). S accumulates during evolution due to mutational asymmetries between the leading and lagging strands[23]. S amplitude and sign, therefore, reflect the average fork direction in the germline. Comparison with replication timing profiles of somatic cells corroborated that replication progresses from the borders to the centers of N domains, suggesting developmental and evolutionary conservation of these replication patterns[24–26]. However, many more origins than S upshifts were found in mammalian genomes; the missing origins must therefore be flexible enough or located within regions frequently rearranged to leave no evolutionary stable imprint on S profiles. Furthermore, the resolution was limited to ~20 kb and analysis of gene-rich regions was complicated by the added effect of transcription-associated mutational asymmetries[27]. These limitations called for a genome-wide, direct experimental determination of RFD at high resolution in mammalian genomes, which was first achieved by sequencing of purified Okazaki fragments[28].

Development and overview of OK-seq

At the replication fork, the leading strand is replicated continuously whereas the lagging strand is synthesized discontinuously, in the form of ~200 nt RNA-primed fragments (Okazaki fragments) that grow in the direction opposite to fork progression. Okazaki fragments are joined together one after another to build an elongating lagging strand. Okazaki fragments mapping to the Watson and Crick strands are generated by leftward- (L) and rightward- (R) moving forks, respectively (Fig. 1a). Therefore, strand-oriented sequencing of Okazaki fragments isolated from a cell population reveals the proportions of R and L forks at any locus, allowing quantitative analyses of replication fork initiation, progression and termination. Isolation and sequencing of Okazaki fragments were first achieved in ligase- and checkpoint-deficient mutants of *Saccharomyces cerevisiae*, which allowed continued DNA synthesis despite the accumulation of unligated Okazaki fragments behind the forks[29,30]. We independently developed a procedure for isolating and sequencing Okazaki fragments from mammalian cells that did not require the introduction of such mutations. In this method, asynchronously growing cells are briefly pulsed with 5-ethynyl-deoxyuridine (EdU) to label newly synthesized DNA, total DNA is denatured and fractionated by size, and the <200 nt EdU-labeled

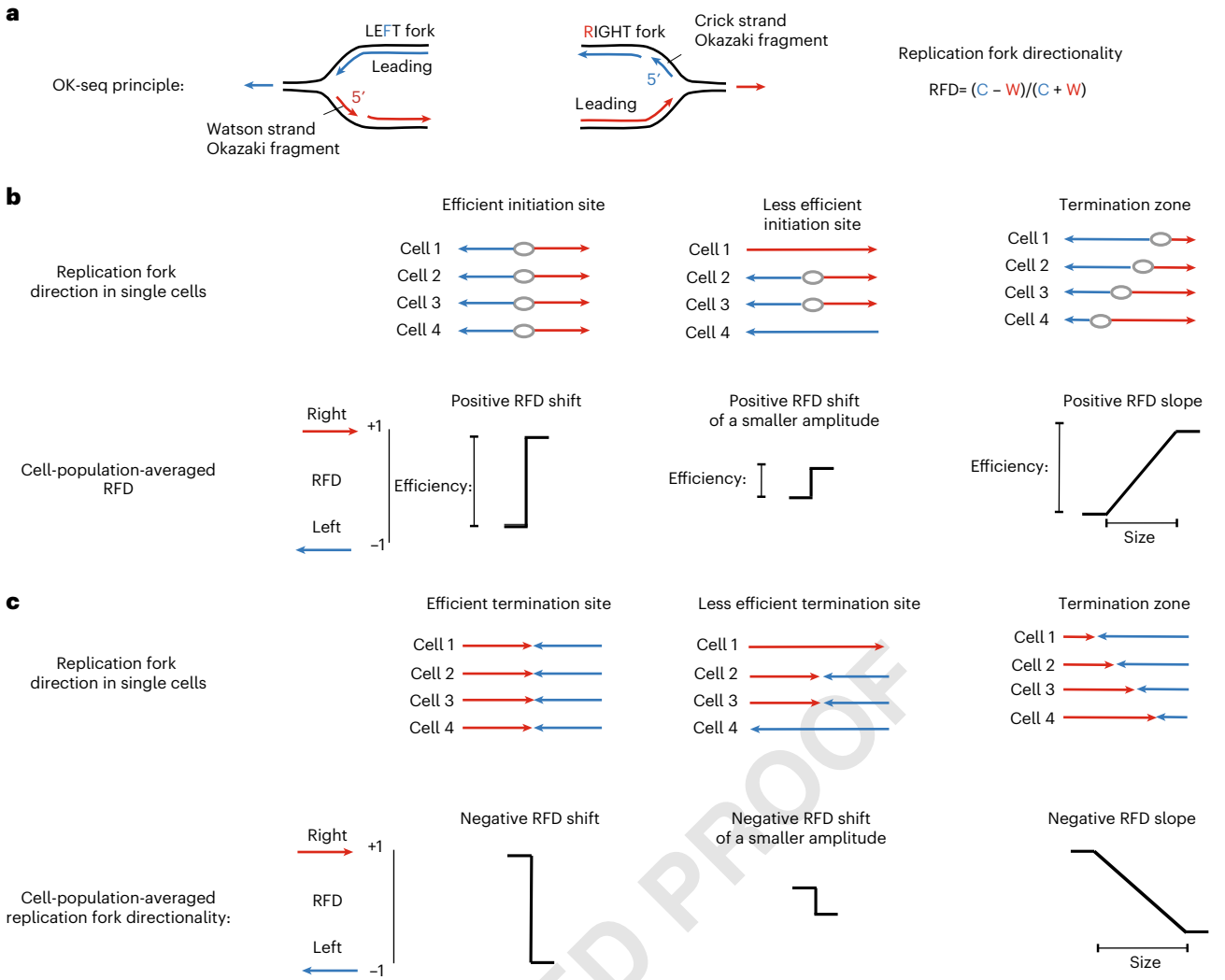


Fig. 1 | Detection of replication initiation and termination events by OK-seq. a, Okazaki fragment strandedness indicates the direction of ongoing replication forks. Watson strand Okazaki fragments (red) are generated from leftward-oriented forks. Crick strand Okazaki fragments (blue) are generated from rightward-oriented forks. RFD, the population-averaged fork directionality is computed as a proportion of reads from Crick and Watson strands. **b**, The RFD profile reflects the location, nature and efficiency of replication initiation. Site-specific initiation (left and center) results in an abrupt positive shift of RFD whereas IZ results in a progressive positive shift of RFD (right) (IZ). The amplitude of the RFD shift reflects the initiation efficiency. **c**, Negative shifts of RFD reflect the sites and zones of predominant fork merging (termination zones).

single DNA strands are click-labeled with biotin, captured on streptavidin beads and ligated to sequencing adapters. This procedure was dubbed Okazaki fragment sequencing (OK-seq)[28] (Fig. 2).

The RFD (RFD = $R - L$) profiles thus obtained had a high resolution (~1 kb for human cells and ~50 bp for yeast) and were informative. RFD at position x is mathematically linked to the mean replication timing (MRT) and to the speed of forks (v), such that $\mathrm{dMRT}/\mathrm{d}x = \mathrm{RFD}/v$ (refs. [25,26]). In other words, steep MRT slopes correspond to unidirectionally replicating regions, flat MRT zones are replicated equally in both directions, and intermediate MRT slopes are replicated by unequal proportions of R and L forks. Indeed, the human OK-seq RFD profiles were found to be extremely consistent with RFD profiles derived from skew and MRT data, but had higher resolution. In yeast, RFD upshifts, where fork direction switches from L to R, span 1 kb or less, identifying site-specific origins (Figs. 1b and 3b) at locations highly consistent with previous origin mapping studies[30,31]; see below for details). However, a completely different RFD pattern is observed along the human genome[28]. Most loci show a mixture of R and L forks, and changes in RFD are progressive rather than abrupt, spanning tens or hundreds of kb (Figs. 1b and 3a). These results imply an extensive cell-to-cell variability in replication patterns. An automated procedure based on a hidden Markov model (HMM)[28,32] was developed to objectively detect ascending, descending and flat RFD segments across the entire genome. Extended flat segments with extreme RFD values (close to ±1), which reveal

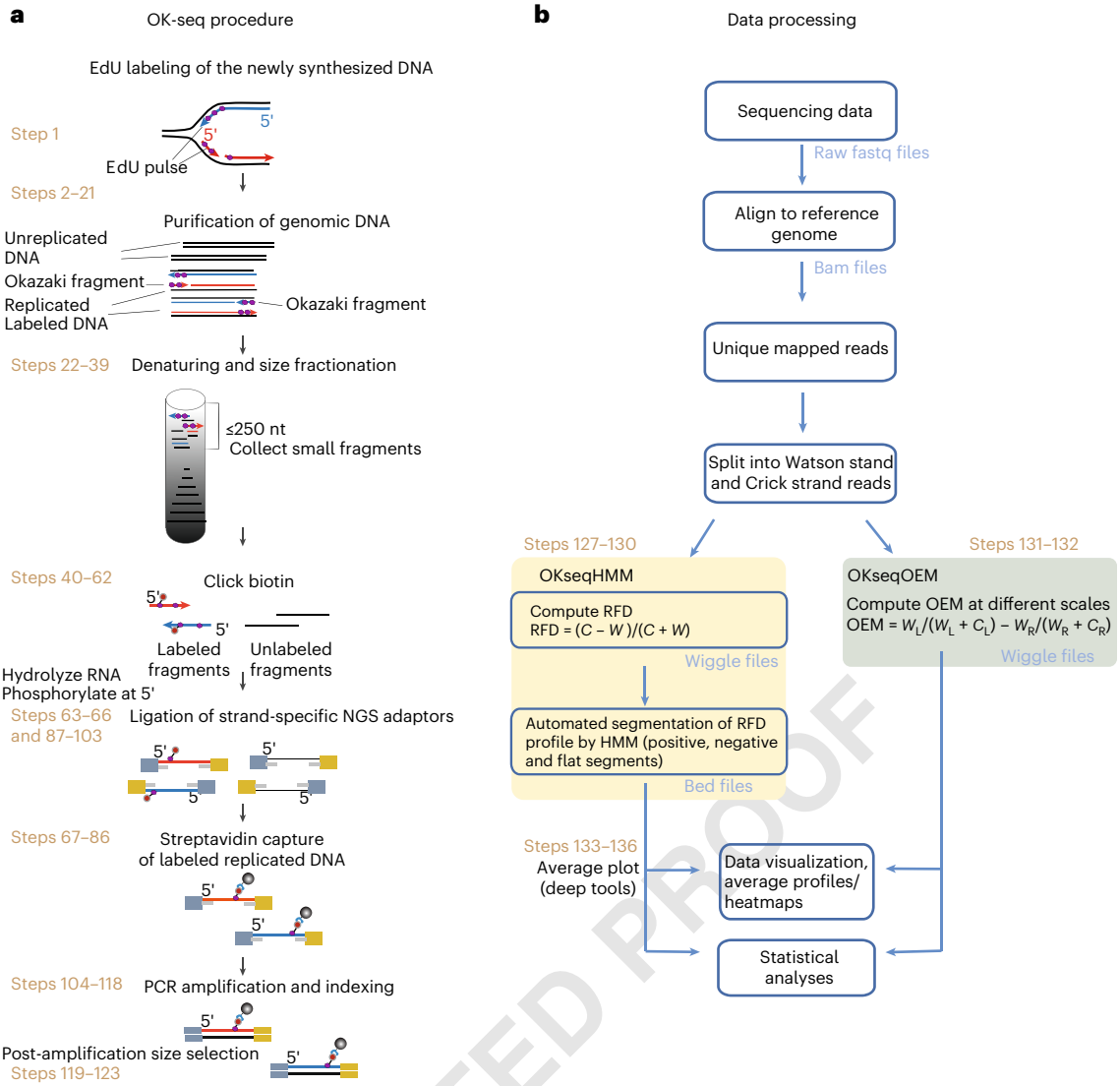


Fig. 2 | Experimental workflow and data processing pipelines of OK-seq. a, Illustration of the key experimental steps. Unreplicated DNA is in black and the replicated DNA strands are in red and in blue. Watson and Crick strand Okazaki fragments are shown as red and blue arrows; EdU (violet dots), biotin (red dots), streptavidin magnetic beads (black) and double-stranded adapters (gray and yellow). **b**, Flowchart representing the data analysis pipeline. OKseqHMM allows to split Watson and Crick strand reads and to compute the RFD values at defined bin sizes. Further, the automated detection of zones of replication initiation, termination and unidirectional fork movement is achieved by segmentation of the RFD profile into upward, downward and flat segments by HMM. The OKseqOEM tool computes OEM at different genomic scales. Average plot allows creating the heatmaps and linear plots to explore RFD patterns around genomic features of interest.

unidirectionally replicating regions, only cover 5–10% of the genome. Segments of ascending RFD, where fork direction progressively shifts from *L* to *R* (IZs), typically span 10–100 kb. They reveal 4,000–10,000 IZs that support a low and homogeneous rate of initiations over their entire length. The amplitude of the shift reveals the global efficiency of each zone (i.e., the fraction of molecular copies which support an initiation event), which ranges from <10% to >90%. Abrupt upshifts such as those found at yeast origins are extremely rare. Descending RFD segments between consecutive IZs reveal extended (10–1,000 kb) zones of replication termination (TZs), even broader than the IZs. Finally, extended segments of null RFD reveal randomly replicating regions, mostly in late-replicating heterochromatin[28].

Importantly, when OK-seq was adapted to purify EdU-labeled Okazaki fragments from *S. cerevisiae*, very similar profiles to those reported for ligase- and checkpoint-deficient *S. cerevisiae* mutants were obtained, consistent with the site-specific nature of yeast origins[31]. Therefore, the much broader RFD upshifts observed in mammalian genomes reflect the different biology of yeast and mammalian

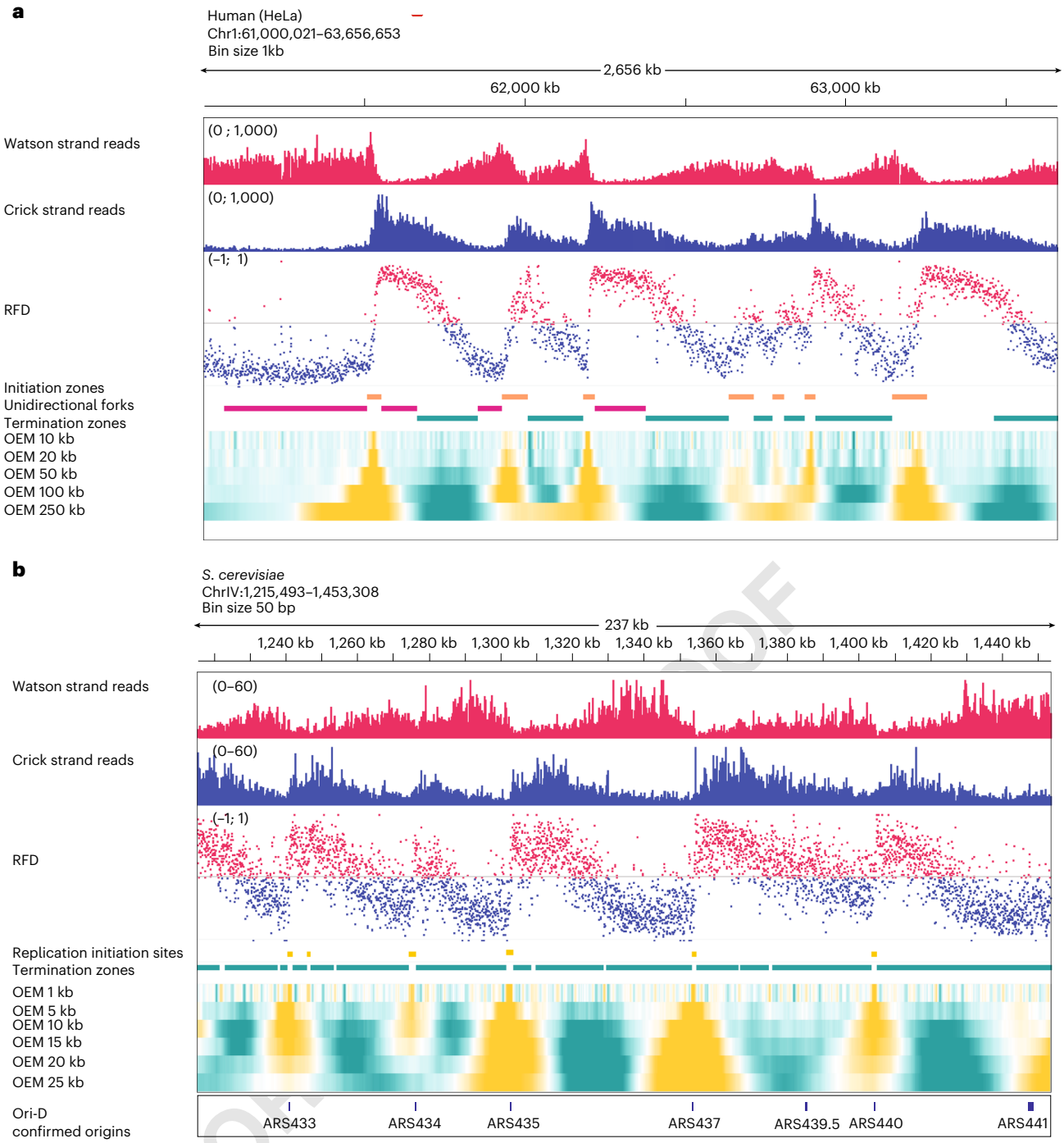


Fig. 3 | Representative results for OK-seq. Okazaki fragment Watson stand (red) and Crick strand (blue) read counts, RFD computed in 1 kb windows and OEM at indicated scales. IZs (yellow) and termination zones (teal blue), flat segments of unidirectional replication (pink), detected by OKseqHMM. Panel **a** shows data for HeLa cells[28] and panel **b** shows data for yeast *S. cerevisiae*[31].

cells and not an inability of OK-seq to reveal abrupt RFD upshifts, characteristic of site-specific origins (Figs. 1 and 3).

Given the cell-to-cell variability and dispersed nature of replication initiation and termination events, particularly in mammalian cells, caution is required to interpret changes in RFD along the profiles. Strictly speaking, the ΔRFD between two genomic positions is equal to twice the difference between the number of initiation and termination events in the considered interval. For example, a segment across which the RFD continuously decreases from +1.0 to −1.0 may simply be invaded by outer forks that merge at variable positions, resulting in a single, delocalized termination event

(Fig. 1c). However, a similar decreasing RFD segment may also arise if one internal, delocalized initiation event emits two diverging forks that meet at random positions with the two outer invading forks, resulting in two delocalized termination events. More generally, scenarios, where multiple delocalized initiation events take place between outer invading forks, can result in a decreasing RFD pattern. Similarly, an ascending RFD segment may in principle arise from multiple delocalized initiation events resulting in the net emission of outward-oriented forks. However, ascending RFD segments are markedly smaller than descending ones, so the scenario with at most one initiation event and no termination event, as first demonstrated for the DHFR IZ[2], is by far the most likely explanation. Single-molecule replication analyses of the budding yeast genome[31] and two chicken chromosome fragile sites[33] recently confirmed that a minor fraction of initiation and termination events occur in negative and positive RFD slopes, respectively. In addition, recent high-throughput single-molecule optical replication mapping (ORM) of early initiation events of human cells[34] also confirmed that a minor fraction of early initiation events occurs in negative RFD slopes as well as within late randomly replicating regions. Therefore, the positive or negative slope of an RFD segment reveals whether initiation or termination predominates, but a mixture of both, on different molecules or on the same molecule, cannot be excluded. Given that the number of ascending RFD segments in mammalian cells (4,000–10,000) is lower than the estimated number of initiation events per S phase (20,000–50,000) and that most IZs support at most one initiation event per cell cycle, the simplest model to reconcile these numbers is that many initiation and termination events occur within TZs and null RFD regions but in a manner that is too dispersed to leave an imprint on population RFD profiles. Such dispersed events can be detected only by single-molecule techniques[31,33,34]

Applications of OK-seq

OK-seq was used to obtain high-resolution, genome-wide RFD profiles of many types of cultured cancer and immortalized metazoan cells[28,33,35–38] and even in primary cells[35,39]. With the continuing development of novel origin mapping techniques, it should be noted that OK-seq IZs have been recently confirmed by EdU-seq HU[39], by high-resolution Repli-seq[17] and by ORM[34].

The HMM automated analysis of the RFD slope presented here allowed mapping of IZs and TZs and measuring of their efficiencies[28]. Alternatively, IZs and TZs can be automatically detected in OK-seq profiles by wavelet-transform analysis[40]. IZs often abut active genes but are not transcribed, consistent with reports that licensed origins are eliminated from transcribed genes[2,41–44]. Due to the different strengths of the 5′ and 3′ IZs, however, active genes tend to be replicated in the same direction as transcription, although the RFD tends to invert over long active genes such that their 3′ end is often replicated in the direction opposite to transcription[45,46].

IZs remote from active genes fire later than gene-bordering IZs. Finally, the HMM model can also detect extended segments of null RFD corresponding to randomly replicating heterochromatin and extended segments of high RFD corresponding to unidirectionally replicating regions[28]. A detailed analysis of RFD profile variability between multiple cell lines has been reported[35].

Besides replication program characterization of normal and cancer cells[28,35,36,39] and of cells subjected to replication stress[37], OK-seq has become very useful in a broad range of genomic studies. First, the inability to initiate replication within transcribed genes has been proposed as a mechanism for causing DNA breaks at common chromosomal fragile sites harboring long genes due to delayed replication[46–48]. The identification of unidirectionally replicated regions by OK-seq, combined with MRT analysis, allowed to predict chromosomal fragile sites genome-wide[46]. Second, the high probability of initiating replication between active genes in early-replicating domains was confirmed by EdU-seq HU[39]. Third, OK-seq data have been used to compare the density of MCM proteins, which mark potential replication origins, to the probability of initiation along the genome. The lack of initiation within transcribed genes was explained by a depletion of MCM proteins within gene bodies. However, ascending and descending RFD segments of similar replication timing and transcription status did not show different MCM densities, suggesting that additional factors to MCM density act to determine the probability of initiation along the genome[40]. Fourth, OK-seq data revealed that active genes tend to replicate codirectionally with transcription[28]. Later studies employing OK-seq data further revealed that head-on, but not codirectional, collisions between replication and transcription lead to the accumulation of potentially deleterious RNA–DNA hybrids (R-loops)[49], that replication stress markers accumulate at transcription termination sites, where forks progress head-on to transcription, but not at transcription start sites, where forks progress codirectionally with transcription[45] and that numerous factors, such as topoisomerase 1 (refs. [45,50]), the SAMHD1

ribonuclease[51] and the SWI/SNF chromatin remodeling complex[52] process R-loops and help resolve transcription–replication conflicts. Fifth, mapping RFD by OK-seq has contributed to revealing that leading and lagging strands are prone to different mutational rates across evolution and during cancer transformation, and have helped to deconvolve the strand-asymmetrical production of mismatches by leading- and lagging-strand DNA polymerases from their strand-asymmetrical removal by mismatch repair[28,53–56]. OK-seq data have also contributed to reveal the strand-biased integration preferences of LINE-1 retrotransposons[57,58]. Sixth, combining OK-seq with strand-specific profiling of replicated chromatin demonstrated that inheritance of parental modified histones proceeds by distinct mechanisms at the leading and the lagging strands[36,38], and combining OK-seq with the analysis of postreplicative DNA methylation maintenance revealed that nascent leading and lagging strands acquire DNA methylation with slightly different kinetics[59].

In sum, OK-seq is a quantitative method to reveal the genome replication dynamics and the impact of DNA replication on genome and epigenome function and evolution.

Comparison with other methods

Other direct and indirect methods for measuring replication directionality have been developed by different groups. As discussed above, nucleotide compositional skew analysis[21,22] and spatial derivation of MRT profiles[25,26] gave RFD profiles highly consistent with, but at lower resolution than OK-seq[28]. The enrichment of Okazaki fragments for direct sequencing was first achieved in *S. cerevisiae* through ligase and checkpoint inactivation[29]. While yeast RFD profiles obtained by this method and by OK-seq are extremely similar[31], the ligase-inactivation approach predominantly enriches for mature Okazaki fragments while the EdU-mediated purification enriches for growing Okazaki fragments, which is important to keep in mind when analysing Okazaki fragment processing and nucleosome phasing.

Recent indirect methods to map RFD are based on the fact that the leading (Pol ε) and lagging (Pols α and δ) strand replicative polymerases incorporate ribonucleotides into genomic DNA at different rates. Ribonucleotide excision repair mutants are viable, and polymerase mutants that incorporate ribonucleotides at higher rates than wild-type have been obtained. Four methods (dubbed EmRiboSeq[60], Pu-Seq[61], HydEn-Seq[62] and Ribose-Seq[63]) were reported to determine the genome-wide distribution of embedded ribonucleotides, and infer RFD, across the genome of ribonucleotide excision repair and polymerase mutants in *S. cerevisiae* and *S. pombe*. They also identified regions in which ribonucleotide incorporation deviates from lagging/leading strand expectations, such as at replication origins, which were proposed to result from leading strand initiation by Pol δ followed by an exchange with Pol ε[61], and at termini, suggesting a reciprocal switch from Pol ε to Pol δ[64]. A recent preprint reported the extension of Pu-seq to human cells[65].

A new method for strand-specific sequencing of SNS revealed that SNS are distributed with a sharp strand-specific asymmetry around the peak summits[66]. This finding is surprising as, during origin firing, SNS are expected to grow in both directions by leading and lagging strand synthesis from two forks.

Novel methods for mapping DNA breaks were reported to indirectly reveal RFD, suggesting that the frequency and/or kinetics of nick repair is distinct between the leading and lagging strands. The GLOE-seq method, which maps single-strand breaks in a strand-specific manner, also provided high-resolution RFD profiles in mammalian and yeast cells. GLOE-seq uses a reduced input cell number compared with OK-seq, yet it requires ligase inactivation[67]. A conceptually similar method that differs in library preparation strategy, TrAEL-seq, allows to map the 3′ ends of double-strand breaks and provides RFD information[68].

Recently, the population-averaged RFD profiles were assembled from the replication profiles of long single DNA molecules obtained by DNA combing in chicken cells[33], ORM based on Bionano high-throughput imaging in human cells[34] and nanopore sequencing in yeast cells (FORK-seq)[31], and all were in excellent agreement with OK-seq RFD profiles. In yeast cells, nanopore sequencing is now a faster and easier method than OK-seq to obtain RFD profiles, but in metazoan cells, the throughput of nanopore sequencing is still limiting.

Although the OK-seq approach is now well established, so far, there was no available bioinformatics protocol to fully explore the data. A recently published *Nature Protocols* paper[69] provided an approach to profile RFD around aggregate genomic features (such as transcription start sites), but no method to call IZ and TZ. Here we provide a complete protocol for using an R-based toolkit, OKseqHMM (https://github.com/CL-CHEN-Lab/OK-Seq), to process and analyse OK-seq data,

along the genomes of different species (human, mouse and yeast)[32]. Following the current protocol, we can (1) visualize high-resolution RFD profiles (1 kb for human/mouse cells and 50 bp for yeast) and detect the IZs and TZs by using a four-state HMM, (2) calculate the origin efficiency metric (OEM)[30] and visualize RFD changes at different scales, and (3) visualize the RFD and OEM profiles over genomic features of interest. This toolkit provides a useful resource for the broad scientific community working on DNA replication, genomic instability and epigenetics.

Limitations

One limitation of OK-seq is that, as any cell population method, it averages cell-to-cell variability. As with other next-generation sequencing (NGS)-based replication origin mapping approaches, rare events cannot be directly seen. Although cell-to-cell variability remains visible since most loci show a mixture of *R* and *L* forks, dispersed initiation and termination events may go undetected even if they represent the majority of events. For example, long segments of null RFD can only be explained by random initiation and termination, but the density of these events cannot be measured. The change in RFD across a segment is equal to twice the difference between the number of initiation and termination events within the segment[70]. Therefore, a minority of termination events may occur within ascending RFD segments. Similarly, a minority of initiation events may occur within descending RFD segments. Only single-molecule methods may directly reveal these events[31,33,71]. The OK-seq results thus led us to propose that replication of mammalian genomes combines predominant initiation within 'master' IZs detected as ascending RFD segments, with more dispersed, less efficient initiation elsewhere.

OK-seq relies on metabolic labeling with nucleotide analogs (EdU) and we anticipate that it may be used in any proliferating cells or even model organisms able to efficiently uptake EdU. OK-seq requires a significant amount of starting material since the half-life of Okazaki fragments is very short. Furthermore, the library preparation step may benefit from future improvements, for example, inspired from single-stranded library preparation from ancient genomes[72], although optimization will be required.

Expertise needed to implement OK-seq

OK-seq requires strong skills in molecular and cell biology. The protocols are accessible to most molecular biology laboratories and rely on common laboratory equipment. Bioinformatic analysis with prebuilt pipelines requires strong computational skills and experience with R.

Experimental design

Here we present some critical considerations and the key steps of the experimental and analytical workflows of OK-seq (Fig. 2).

Cell culture and starting cell number

Since we purify Okazaki fragments from unperturbed asynchronously growing cells, the amount of fragments is expected to be tiny, around hundreds of picograms per million asynchronous cells. Therefore, Okazaki fragment isolation requires a large number of input cells ($3–10 \times 10^8$). This requires setting up large-scale cell cultures, which needs to be carefully planned. Cell numbers may be optimized depending on genome size and a fraction of cells in S phase. For example, a lymphoblastoid cell line of nearly normal karyotype with ~20% of cells in S phase (GM06990) required $8–10 \times 10^8$ cells per biological replicate, whereas hyperploid cancer cell lines with 30–35% of cells in S phase, such as HeLa or K562, required 3×10^8 cells per replicate. Cell cultures should be split 1 or 2 d before the experiment, to ensure small colonies and uniform EdU labeling. For each experiment, two independent biological replicates are desired.

EdU labeling and cell harvesting

In this step, newly synthesized DNA strands are briefly labeled with ethynyl-containing nucleotide EdU[73]. The Okazaki fragments are transient, with a half-life shorter than 10 s, and are immediately ligated to the elongating nascent lagging strands[74,75]. We set the EdU pulse for 2 min because it was easy to keep consistent between experiments at a comfortable working pace. Yet, in theory, the pulse could be shortened since thymidine analogs are almost instantly assimilated. In contrast, longer pulses will increase the proportion of nascent labeled DNA of higher molecular weight that could contaminate the Okazaki fragment preparation. In any case, the duration of the pulse needs to be precisely controlled and stopped abruptly by adding ice-cold PBS. It is, therefore, preferable to treat a

small number of dishes (two or three) at the same time. Option A of this section explains how to label and harvest adherent cells (HeLa), and option B explains how to treat the cells growing in suspension (Epstein–Barr virus-immortalized lymphoblastoid GM06690). For labeling, we have also previously used a cytidine analog EdC[76], which in HeLa cells gave an identical result to EdU[28]. However, the use of EdC has limitations, as EdC assimilation efficiency varies in different cell types and depends on cytidine deaminase activity[77,78].

Nucleic acid extraction

Nucleic acids are extracted with the proteinase K/phenol–chloroform method[79], which allows inexpensive milligram-scale preparation of pure high-molecular-weight genomic DNA. At this step, it is critical to avoid pipetting and vortexing to minimize DNA breakage and potential contamination of Okazaki fragment preparation with fragments of elongating nascent strands. After ethanol precipitation, we typically leave the DNA pellet in TE buffer for 3–7 d at 4 °C to allow it to dissolve without pipetting. We omit RNAse A digestion and use intracellular RNAs as molecular cargo during subsequent purification steps.

Size fractionation and recovery of small single-stranded fragments

To release Okazaki fragments, genomic DNA is heat denatured and size-fractionated on neutral linear 5–30% sucrose gradients[80]. The number of required gradients (typically six to ten) depends on the starting cell number; we fractionate <500 µg of genomic DNA per gradient (from 1×10^8 to 1.5×10^8 of starting cells). Sucrose gradients are unstable and should be handled with care during preparation. After overnight centrifugation, the small fragments (<250 nt) contained in the upper fractions of gradients are collected, concentrated and purified.

Biotinylation by click reaction

For isolation of EdU-labeled replicated DNA, EdU is coupled with biotin-TEG-azide in a click reaction (copper-catalyzed azide-alkyne cycloaddition (CuAAC) click chemistry)[81–83]. Afterward, cellular RNAs, including the RNA portions of Okazaki fragments are hydrolyzed with alkali and 5′ extremities of DNA fragments are phosphorylated with T4 PNK.

Sequencing adapter ligation and streptavidin capture of biotinylated fragments

In OK-seq, it is critical to prepare strand-oriented libraries from single-stranded DNA with minimal technical bias, to achieve uniform coverage of reads over the genome. In library preparation, double-stranded DNA ligation with T4 DNA ligase is used since it has lower sequence preference compared to single-stranded DNA ligation[84]. Two different double-stranded adapters with a single-stranded random hexanucleotide overhang are hybridized to the ends of the purified fragments. To reduce self-complementary interactions of the 5′ adapter (A1) and 3′ adapter (A2), the standard Illumina sequence of 5′ adapter was shortened by five bases[85]. To prevent self-ligation, adapter A2 contains 3′-terminal dideoxy-modifications (Table 1). After the ligation step, the library fragments containing nascent biotinylated molecules are captured with streptavidin-coated magnetic beads. We perform an additional step of hybridization and ligation of adapters on beads to increase the chance of successful recovery of Okazaki fragments into the library. Each step is followed by stringent high-salt washes to remove the nonspecifically bound DNA molecules and unligated adapters.

Library amplification and sequencing

Libraries are amplified by PCR with indexing primers (Table 1). The template library fragments remain attached to the beads during PCR and may be recovered, washed and reused for an additional round of amplification. In our hands, this additional amplification step resulted in a much higher yield of the final amplified library with nearly identical library complexity, without a strong increase in PCR duplicates[28]. PCR products containing >30 bp inserts are size-selected and eluted from agarose gels. Illumina sequencing is performed following standard protocols but replacing the sequencing primer of the first read by the shortened primer[85].

Data processing

The raw sequencing data (fastq files) need to be preprocessed and aligned to a reference genome using standard bioinformatics procedures. With this protocol, we could obtain high-quality RFD profiles and call replication initiation and termination zones with as few as 50 millions of deduplicated uniquely mapped reads in the human genome[32,35]. In our toolkit, the first function (OKseqHMM)

Table 1 | Oligonucleotides used in the study

Oligo name	Sequences (5′ to 3′)
$A1_{top}$ (R1)	ACACTCTTTCCCTACACGACGCTCTTCC
$A1_{bottom}$	NNNNNNGGAAGAGCGTCGTGTAGGGAAAGAGTGT
$A2_{top}$	[Phos]-AGATCGGAAGAGCACACGTCTGAACTCCAGTCA[ddC]
$A2_{bottom}$	TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNN[ddC]
PEM_1.0	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC
TruSeq_Index 1	CAAGCAGAAGACGGCATACGAGATcgtgatGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
TruSeq_Index 2	CAAGCAGAAGACGGCATACGAGATacatcgGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
TruSeq_Index 3	CAAGCAGAAGACGGCATACGAGATgcctaaGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
TruSeq_Index 4	CAAGCAGAAGACGGCATACGAGATtggtcaGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
TruSeq_Index 5	CAAGCAGAAGACGGCATACGAGATcactgtGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
TruSeq _Index 6	CAAGCAGAAGACGGCATACGAGATattggcGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
TruSeq_Index 7	CAAGCAGAAGACGGCATACGAGATtcaagtGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
TruSeq_Index 8	CAAGCAGAAGACGGCATACGAGATctgatcGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

automatically detects whether the input-aligned sequencing data are single-end or paired-end reads, then splits reads into Watson and Crick strands and calculates the RFD values within adjacent windows (by default 1 kb) along the reference genome; $RFD = \frac{C-W}{C+W}$, where C and W correspond to the number of reads mapped on the Crick and the Watson strands respectively. Next, an HMM algorithm allows segmentation of the RFD profile into upward, downward and flat segments to predict the location of initiation, termination and unidirectional fork movement zones respectively. The second function of the tool kit, OKseqOEM, uses the Watson and Crick strand-aligned reads to compute the OEM at multiple scales defined by the user; $OEM = \frac{W_L}{W_L+C_L} - \frac{W_R}{W_R+C_R}$ (where W_L and W_R are the numbers of reads in the left and right quadrants of the Watson strand, while C_L and C_R refer to the read numbers in the left and right quadrants of the Crick strand). Finally, the function AveragePlot generates average metagene profiles and heatmaps to analyze the distribution of RFD and OEM around genomic features of interest.

Materials

Biological materials

▲ **CRITICAL** For the yeast *S. cerevisiae*, please refer to Supplementary Protocol 1.

Human cell lines

- HeLa (clone MRL2, a kind gift from Dr. Olivier Bensaude, IBENS)
- Immortalized lymphoblasts GM06990 (Coriell Institute, RRID: CVCL_9587) **! CAUTION** The cell lines used in your research should be checked regularly to ensure they are authentic and mycoplasma-free.

▲ **CRITICAL** Use the appropriate medium and supplements for the cell type of interest.

Reagents

Cell culture reagents for HeLa cells

- DMEM (Gibco, cat. no. 31966-021)
- Fetal bovine serum (FBS; Sigma-Aldrich, cat. no. F2442)
- Penicillin–streptomycin, 10,000 U/mL (Gibco, cat. no. 15140-122) **! CAUTION** Irritant upon contact with skin. Wear gloves and a lab coat.
- Trypsin–EDTA, 0.25% (Gibco, cat. no. 25200-056)

Cell culture reagents for GM06990

- 1× PBS (Thermo Fisher, cat. no. 14200083)
- 2-Mercaptoethanol (Sigma, cat. no. M6250) **! CAUTION** Toxic if swallowed or if inhaled. It may cause skin irritation. Work under a chemical hood and wear gloves and a lab coat when handling.
- FBS (Sigma-Aldrich, cat. no. F2442)
- Penicillin–streptomycin, 10,000 U/mL (Gibco, cat. no. 15140-122) **! CAUTION** Irritant upon contact with skin. Wear gloves and a lab coat.
- RPMI1640 (Thermo Fisher, cat. no. 61870127)

Common reagents

- 5 M betaine (Sigma-Aldrich, cat. no. B0300-5VL)
- (Optional) 5-ethynyl-2′-deoxycytidine (EdC; Jena Bioscience, cat. no. CLK-N003-10) (see 'Experimental design')
- 5-Ethynyl-deoxy-uridine (Jena Bioscience, cat. no. CLK-N001-25)
- 50% PEG8000 (Jena Bioscience, cat. no. CSS-256)
- Absolute ethanol (Sigma-Aldrich, cat. no. 1117272500) **! CAUTION** Ethanol is flammable and an irritant. Wear disposable gloves when handling.
- Acetic acid (Sigma-Aldrich, cat. no. 33209) **! CAUTION** Flammable, volatile and irritative. Work under a chemical hood and wear gloves and a lab coat when handling.
- Agilent High Sensitivity DNA Kit (Agilent, cat. no. 5067-4626)
- Ammonium acetate (VWR, cat. no. 21200.297) **! CAUTION** Work under a chemical hood while wearing a lab coat and disposable gloves.
- AMPure beads (Beckman, cat. no. A63881)
- ATP, 100 mM (Thermo Fisher, cat. no. R0441) **▲ CRITICAL** Aliquot into 20–50 µL aliquots, store at −20 °C and avoid multiple freeze–thaw cycles.
- Biotin-TEG azide (Berry & Associates, cat. no. BT1085)
- Bromophenol blue (Sigma-Aldrich, cat. no. 32712-5G) **! CAUTION** Work wearing a lab coat and disposable gloves.
- Chloroform (VWR, cat. no. BDH83627.400) **! CAUTION** Toxic and corrosive. Work under a chemical hood while wearing a lab coat and disposable gloves.
- Copper (II) sulfate ($CuSO_4$; Jena Bioscience, cat. no. CLK-MI004-50) **! CAUTION** Is irritant to the skin and eyes and is toxic if swallowed. Work wearing a lab coat and disposable gloves.
- Dimethyl sulfoxide (DMSO; Sigma-Aldrich, cat. no. D2650) **! CAUTION** DMSO is harmful to the skin and is combustible. Work wearing a lab coat and disposable gloves.
- Distilled deionized water (ddH_2O) or UltraPure DNase/RNase-Free Distilled Water (Thermo Fisher, cat. no. 10977035)
- dNTPs, 10 mM each (Thermo Fisher, cat. no. R0192) **▲ CRITICAL** Prepare 5–10 µL aliquots, store at −20 °C and avoid freeze–thawing.
- Dynabeads MyOne streptavidin T1 (Thermo Fisher, cat. no. 65601)
- EB buffer (Qiagen, cat. no. 19086)
- EDTA Ultrapure, 0.5 M, pH 8.0 (Life Technologies, cat. no. 15575-038) **! CAUTION** Toxic if swallowed. Work wearing a lab coat and disposable gloves.
- Gel loading buffer II, 2×, for urea PAGE (Thermo Fisher, cat. no. AM8546G) **! CAUTION** Contains formamide and is toxic. Work under a chemical hood while wearing a lab coat and disposable gloves.
- Gel loading dye, purple, 6×, for PAGE and agarose gels (NEB, cat. no. B7024s)
- KAPA HiFi HotStart DNA Polymerase (Roche, cat. no. 07958889001)
- Low molecular weight DNA ladder (NEB, cat. no. N3233S)
- MinElute Gel extraction Kit (Qiagen, cat. no. 29604)
- MinElute PCR Purification Kit (Qiagen, cat. no. 28004)
- 1× PBS (Thermo Fisher, cat. no. 14200083)
- 10× PBS (Thermo Fisher, cat. no. 70011044)
- Phenol chloroform isoamyl alcohol, 25:25:1 (Thermo Fisher, cat. no. 15593-049) **! CAUTION** Toxic and corrosive. Work under a chemical hood while wearing a lab coat and disposable gloves.
- Potassium acetate (CH_3COOK; Calbiochem, cat. no. 529553)
- Primers for sequencing adapters and library construction (common supplier, Table 1)
- Proteinase K (Roche, cat. no. 3115879001)
- Qubit dsDNA BR Assay Kit, 2–1,000 ng/µl (Thermo Fisher, cat. no. Q32853)
- Qubit ssDNA HS Assay Kit, 0.05–100 ng/µl (Thermo Fisher, cat. no. Q10212)
- Small fragments agarose (Eurogentec, cat. no. EP-0020-10)
- Sodium acetate (Merck, cat. no. 1.06268.0250).
- Sodium ascorbate (Jena Bioscience, cat. no. CLK-MI005-50)
- Sodium chloride (NaCl; Sigma-Aldrich, cat. no. S7653)
- Sodium dodecyl sulfate (SDS) solution 20% (wt/vol) (Sigma-Aldrich, cat. no. 05030-500ML-F) **! CAUTION** SDS is corrosive to the skin and a respiratory irritant. Work wearing a lab coat and disposable gloves. Thoroughly wash with water skin or eyes exposed to this chemical.
- Sodium hydroxide (NaOH; Sigma-Aldrich, cat. no. 1.06469.1000) **! CAUTION** NaOH is corrosive. Wear gloves and a lab coat when handling.

- Sucrose (Sigma-Aldrich, cat. no. 1.07687.5000)
- SYBR Gold Nucleic Acid Gel Stain, 10,000× concentrate in DMSO (Thermo Fisher, cat. no. S11494) **! CAUTION** Is a potential cancer hazard. Work wearing a lab coat and disposable gloves.
- SYBR Green I Nucleic Acid Gel Stain 10,000× (Thermo Fisher, cat. no. S7585) **! CAUTION** Is a potential cancer hazard. Work wearing a lab coat and disposable gloves
- T4 DNA ligase (Thermo Fisher, cat. no. EL0014) **▲ CRITICAL** Aliquot the ligase buffer into 20–50 µL aliquots. Store at −20 °C and avoid exceeding three freeze–thaw cycles.
- T4 polynucleotide kinase, T4 PNK (Thermo Fisher, cat. no. EK0031)
- TAE buffer (Thermo Fisher, cat. no. 15558026)
- Taq DNA polymerase (NEB, cat. no. M0273)
- TBE buffer (Thermo Fisher, cat. no. B52) **! CAUTION** Harmful if swallowed or inhaled. Work wearing a lab coat and disposable gloves.
- TBE gels, 10% (Thermo Fisher, cat. no. EC62752BOX) **! CAUTION** Acrylamide is a potential cancer hazard. Work wearing a lab coat and disposable gloves.
- TBE–urea gels, 10% (Thermo Fisher, cat. no. EC68752BOX) **! CAUTION** Acrylamide is a potential cancer hazard. Work wearing a lab coat and disposable gloves.
- Tris–HCl buffer, 1 M, pH 7.5 (Thermo Fisher Scientific, cat. no. 15567027)
- Tris–HCl buffer, 1 M, pH 8.0 (Thermo Fisher Scientific, cat. no. 15568025)
- Tris (3-hydroxypropyl-triazolyl methyl) amine (THPTA; Sigma-Aldrich, cat. no. 762342) **! CAUTION** - Skin and eye irritant. Work wearing a lab coat and disposable gloves.
- Triton X-100, molecular-biology grade (Sigma-Aldrich, cat. no. T8787-100ml) **! CAUTION** Skin and eye irritant. Work wearing a lab coat and disposable gloves.
- Tween 20 (Sigma-Aldrich, cat. no. P1379)
- HiSeq 3000/4000 SBS Kit, 50 cycles (Illumina, cat. no. FC-410-1001)

Equipment

- 0.2 mL PCR tube (Eppendorf, cat. no. 0030124332)
- 1.5 mL Eppendorf tube (Eppendorf, cat. no. 33290)
- 2100 Bioanalyzer Instrument (Agilent, cat. no. G2939BA)
- Allegra 64R High-Speed Centrifuge (Beckman, 367588) with fixed angle rotor JLA-10.500 (Beckman, cat. no. 369681)
- Amicon Ultra-15 centrifugal filter unit (Millipore, cat. no. UFC901024)
- ART wide bore filtered pipette tips, 1 mL (Thermo Fisher, cat. no. 2079G)
- Beckman Coulter 25 × 89 mm ultraclean tube (Beckman, cat. no. 344058)
- Benchtop centrifuge, refrigerated fixed angle rotor (Eppendorf, model no. 5424R)
- Benchtop centrifuge, swing bucket (Eppendorf, model no. 5910)
- Blades (Sigma-Aldrich, cat. no. Z290947)
- Cell culture incubator, 37 °C, 5% CO_2
- Cell scrapers (Duscher, cat. no. 010155)
- Counting chambers: KOVA Glasstic Slide 10 with Counting Grids (KOVA International, cat. no. 87144) (alternatively a hemacytometer or cell counter can be used)
- Dark Reader Non-UV Transilluminator (Clare Chemical, cat. no. DR-22A)
- Falcon tissue culture dishes 150 mm (VWR, cat. no. 25383-103)
- Falcon Petri flasks 175 cm^2 (Corning, cat. no. 353112)
- Falcon conical tubes 50 ml Cellstar (Greiner Bio-One, cat. no. 227-261)
- Falcon conical tubes 15 ml Cellstar (Greiner Bio-One, cat. no. 188-271)
- 500 mL centrifuge bottles (Beckman, cat. no. 361691)
- DiaMag Rotator (Diagenode, cat. no. B05000001)
- DNA LoBind Tubes, 1.5 mL (Eppendorf, cat. no. 022431021)
- DynaMag-2 Magnet (Thermo Fisher, cat. no. 12321D)
- Eppendorf ThermoMixer C (Eppendorf, cat. no. EP5382000023)
- Evaporator (Eppendorf, model no. 5301)
- Glass Pasteur pipettes (VWR, cat. no. 14673-043; clean and autoclaved)
- Gradient maker (Hoefer, cat. no. SG50) or Gradient Master (Biocomp, cat. no. 108)
- Electrophoresis system, vertical (Hoefer, model no. SE260-10A-1.5)
- Electrophoresis system, horizontal (Bio-Rad, model no. Sub-Cell Model 96)
- HiSeq 3000 System (Illumina, cat. no. SY-401-3001) or equivalent

- Integra Biosciences Pipetboy Accu 2 Pipette Controller (Fisher Scientific, cat. no. 10798252)
- Laminar flow hood (ESCO, Model No. LVG-4AG-F8)
- Safe Imager 2.0 Blue-Light Transilluminator (ThermoFisher, cat. no. G6600)
- Phase lock gel light 50 mL (5 Prime, cat. no. 713-2539) or MaXtract High-Density 50 mL (Qiagen, cat. no. 129073) or equivalent
- 50 mL plastic pipettes (Corning, cat. no. 07-200-17)
- 25 mL plastic pipettes (Corning, cat. No. 07-200-15)
- 10 mL plastic pipettes (Corning, cat. no. 07-200-12)
- ProFlex PCR System (Thermo Fisher, cat. no. 4484073)
- Qubit 4 fluorometer (Thermo Fisher, cat. no. Q33238)
- Qubit assay tubes (Thermo Fisher, cat. no. Q32856)
- Sorenson low-binding aerosol barrier tips, MicroGuard G, maximum volume 10 µL (Sigma-Aldrich, cat. no. Z719374)
- Sorenson low-binding aerosol barrier tips, MultiGuard, maximum volume 200 µL (Sigma-Aldrich, cat. no. Z719447)
- Sorenson low-binding aerosol barrier tips, MultiGuard, maximum volume 20 µL (Sigma-Aldrich, cat. no. Z719412)
- Sorenson low-binding aerosol barrier tips, MultiGuard, maximum volume 100 µL (Sigma-Aldrich, cat. no. Z719463)
- Micro Bio-spin columns P30 (Bio-Rad, cat. no. 732-6250)
- Optima XE-100-IVD Ultracentrifuge (Beckman, part no. A99836) with swinging rotor SW28 (Beckman, part no. 369650) or SW32 (Beckman, part no. 342207)
- Vortex-Genie 2 (Scientific Industries, cat. no. SI-A256)
- 250 mL glass beaker, clean and autoclaved (Fisher Scientific, cat. no. FB101250)
- 600 mL glass beaker, clean and autoclaved (Fisher Scientific, cat. no. FB101600)

Software

- deepTools (https://deeptools.readthedocs.io/en/develop/index.html)[86]
- IGV (https://software.broadinstitute.org/software/igv/)[87]
- OKseqHMM (https://github.com/CL-CHEN-Lab/OK-Seq)
- R (https://www.r-project.org/)[88]
- R package 'HMM'[89]
- R package 'Rsamtools'[90]
- R package 'GenomicAlignments'[91]
- RStudio[92]
- wigToBigWig (http://hgdownload.soe.ucsc.edu/admin/exe/)

Reagent setup

▲ **CRITICAL** For the common stock solutions, please refer to standard molecular biology recipes[79] and http://cshprotocols.cshlp.org/site/recipes/nav_s.dtl.

DMEM-serum medium for HeLa cells

Mix 500 mL of DMEM medium with 50 mL of FBS and 5 mL of 100× penicillin–streptomycin. The medium can be stored at 4 °C for up to 2 weeks. Prewarm to 37 °C in a water bath before use.

RPMI 1640-serum medium for GM06990 cells

Mix 500 mL of RPMI medium with 75 mL of FBS, 5 mL of 100× penicillin–streptomycin, and 3.5 µL of β-mercaptoethanol. The medium can be stored at 4 °C for up to 2 weeks. Prewarm to 37 °C in a water bath before use.

100 mM biotin-TEG azide

Add 0.562 mL of DMSO to a vial containing 25 mg of biotin-TEG azide. Mix by vortexing until dissolved. Quick spin and store at 4 °C for up to 1 year.

100 mM $CuSO_4$

Add 6.27 mL of ddH_2O to a vial containing 100 mg of $CuSO_4$. Aliquot 500 µL per tube and store at 4 °C for up to 1 year.

1 M sodium ascorbate

Add 1.01 mL of ddH_2O to a vial containing 200 mg of sodium ascorbate. Mix by vortexing until dissolved. Quick spin and store at −20 °C for up to 1 year. ▲ **CRITICAL** Discard the solution if it has turned yellow and prepare a fresh one.

20 mM EdU

Dissolve 25 mg in 4.956 mL of DMSO. Aliquot and store at −20 °C for up to 1 year.

2 × BWT

Prepare following the recipe listed below. Store at room temperature (RT, 22 °C) for up to 6 months.

Reagent	Final	Stock	Volume (mL) for 50 mL
Tris-HCl pH 7.5	10 mM	1 M	0.5
EDTA pH 8.0	1 mM	0.5 M	0.1
NaCl	2 M	5 M	20
Tween 20	0.1% (vol/vol)	10% (vol/vol)	0.5
ddH_2O			Up to 50 mL

1× BWT

Mix 25 mL of 2× BWT with 25 mL of ddH_2O. Store at RT for up to 6 months.

TE

Prepare following the recipe listed below. Store at RT for up to 6 months.

Reagent	Final	Stock	Volume (mL) for 50 mL
Tris-HCl pH 8.0	10 mM	1 M	0.5
EDTA pH 8.0	1 mM	0.5 M	0.1
ddH_2O			Up to 50 mL

500 mM THPTA

Add 460.3 µL of ddH_2O to a vial containing 100 mg of THPTA. Mix by vortexing until dissolved. Quick spin and store at 4 °C for up to 1 year.

80% (vol/vol) ethanol

Mix 8 mL of absolute ethanol with 2 mL of ddH_2O. ▲ **CRITICAL** Prepare freshly each time.

AMPure XP beads

Aliquot the bead solution in 2 ml tubes and store at 4 °C. ▲ **CRITICAL** The AMPure XP beads need to be equilibrated at RT (≥22 °C) for at least 30 min before use.

DNA lysis buffer

Prepare following the recipe listed below. Autoclave and store at RT for up to 1 year.

Reagent	Final	Stock	Volume (mL) for 500 mL
Tris-HCl pH 8.0	10 mM	1 M	5 mL
EDTA pH 8.0	25 mM	0.5 M	25 mL
NaCl	100 mM	5 M	10 mL
ddH_2O			Up to 500 mL

EB buffer

Mix 0.5 mL of 1 M Tris–HCl pH 8.0 with 49.5 mL of ddH_2O. EB buffer can be stored at RT for up to 6 months.

5% (wt/vol) TEN-sucrose buffer

Prepare following the recipe listed below. Autoclave and store at RT for up to 6 months.

Reagent	Final	Stock	Volume (mL) for 1 L
Tris-HCl pH 8.0	10 mM	1 M	10 mL
EDTA pH 8.0	1 mM	0.5 M	2 mL
NaCl	100 mM	5 M	20 mL
Sucrose	5% (wt/vol)	50% (wt/vol)	100 mL
ddH_2O			Up to 1,000 mL

30% (wt/vol) TEN-sucrose buffer

Prepare following the recipe. Add several crystals of bromophenol blue. Autoclave and store at RT for up to 6 months.

Reagent	Final	Stock	Volume (mL) for 1 L
Tris-HCl pH 8.0	10 mM	1 M	10 mL
EDTA pH 8.0	1 mM	0.5 M	2 mL
NaCl	100 mM	5 M	20 mL
Sucrose	30% (wt/vol)	50% (wt/vol)	600 mL
ddH_2O			Up to 1,000 mL

▲ **CRITICAL** The bromophenol blue is optional but is very useful for gradient visualization.

1× TE-Tween

Prepare following the recipe. Store at RT for up to 1 year.

Reagent	Final	Stock	Volume (mL) for 50 mL
Tris-HCl pH 8.5	10 mM	1 M	0.5
Tween 20	0.05% (vol/vol)	10% (vol/vol)	0.25
EDTA	1 mM	0.5 M	0.1
ddH_2O			Up to 50 mL

Oligonucleotides

Order the primers listed in Table 1 from a standard lab supplier. Adapters should contain the indicated modifications and be ordered in HPLC-grade, PCR primers can be ordered in a standard purification grade. Dissolve the oligonucleotides in EB buffer to the final concertation of 100 µM. Prepare working solutions of PCR primers by further diluting with nuclease-free H_2O to 10 µM. Store at −20 °C for up to 2 years. ▲ **CRITICAL** The index sequences in the TruSeq Primers (lowercased) can be substituted with any other index sequences. Dual indexing can be included in the primer sequences if desired.

Procedure

Cell culture, EdU labeling and cell harvesting ● Timing 2–7 d for cell culture, 2 h for labeling and harvesting.

1 Follow option A for adherent cells (HeLa) and option B for suspension cells GM06990.
▲ **CRITICAL** For the yeast *S. cerevisiae*, please refer to Supplementary Protocol 1.

(A) Cell culture, EdU labeling and harvesting of adherent cells (HeLa)

(i) Culture adherent HeLa cells in 15 cm dishes with 20 mL of DMEM-serum medium.

(ii) Seed 4 × 10^6 cells per dish and grow them for ~48 h at 37 °C, 5% CO_2 to reach 70-80% confluency. Prepare enough plates to harvest at least 300 million cells per replicate (~20 of 150 mm plates for HeLa cells).

▲ **CRITICAL STEP** Respect the optimal cell culturing conditions to maintain exponential cell growth.

▲ **CRITICAL STEP** The cell number may need to be optimized depending on the fraction of cells undergoing S phase in the population and the cell ploidy. For the details, see 'Experimental design'.

(iii) Transfer 10 mL of the medium from the plate with a 10 mL pipette to a 50 mL tube and add 20 µL of 20 mM EdU stock solution. Mix and pour the EdU-containing medium back to the plate. The final EdU concentration in the plate is 20 µM. Return the plates to the incubator for exactly 2 min.

▲ **CRITICAL STEP** To keep the labeling time consistent between the plates, the EdU-containing medium has to be added and removed exactly in the same order and at a fixed time interval (30 s to 1 min) between plates. For convenience, we do not recommend handling more than two or three plates at the same time.

(iv) Remove the plates from the incubator. Immediately aspirate the medium and add 10 mL of ice-cold 1× PBS to stop the labeling. Store the plates at 4 °C until all plates are processed.

(v) Scrape the plates with a clean cell scraper and transfer the cell suspension to 50 mL conical centrifuge tubes chilled on ice. Rinse each plate with 5–10 mL of ice-cold 1× PBS, and combine the suspension in the same 50 mL conical tubes. Centrifuge for 10 min at 4 °C, 300*g* and discard the supernatant.

■ **PAUSE POINT** Cell pellets can be snap-frozen in liquid nitrogen and stored at −80 °C for up to 1 year.

(B) Cell culture, EdU labeling and harvesting of suspension cells (B-lymphoblasts GM06990)

▲ **CRITICAL STEP** Lymphoblastoid cells make clumpy colonies at the bottom of the flasks. To maintain healthy cultures, resuspend the clumpy colonies during passaging to achieve a single-cell suspension between the passages.

(i) Culture cells in 175 cm^2 flasks with 50 mL of RPMI1640-serum medium at 0.8–1 million cells per milliliter.

(ii) Seed 2–2.5 × 10^7 cells in a 175 cm^2 flask with 100 mL of medium and incubate for ~48 h at 37 °C, 5% CO_2 to reach 0.8–1 million per milliliter. Keep the flasks upright during incubation. Prepare enough flasks to harvest at least 800 million cells per replicate (eight to ten flasks of 175 cm^2 for GM06990 cells).

(iii) Carefully remove 80 mL of medium from the top of the flask using a pipette without disturbing cell clumps formed at the bottom of the flask. Save 20 mL of the medium in a 50 mL conical tube.

▲ **CRITICAL STEP** This step allows reducing the volume of the labeling medium. Lymphoblastoid cells form clumpy colonies on the bottom of the flask and the excess medium can be removed by aspirating from the top. For cell types growing in spinning flasks, cells can be centrifuged before the labeling and resuspended in a smaller volume of prewarmed medium.

(iv) Add 40 µL of 20 mM EdU stock solution to the 20 mL of medium. Mix and pour the EdU-containing medium back into the flask containing 20 mL of cell suspension. The final EdU concentration is 20 µM. Incubate flasks in the cell culture incubator at 37 °C for exactly 2 min.

(v) Immediately immerse the flasks in an ice-cold water bath and add 40 ml of ice-cold 1× PBS and 250 µL of 0.5 M EDTA and mix well by shaking by hand, to stop the labeling. Store the flasks in the ice-cold water bath until all flasks are processed.

▲ **CRITICAL STEP** Respect the exact labeling time and immediately cool the flasks to quickly terminate the labeling.

(vi) Transfer cells to 50 mL Falcon tubes and centrifuge for 10 min at 4 °C, 300*g*. Discard the supernatant.

(vii) Resuspend all the pellets with 20 mL ice-cold 1× PBS in one 50 mL Falcon tube. Centrifuge at 300*g* for 10 min at 4 °C. Discard supernatant.

■ **PAUSE POINT** Cell pellets can be snap-frozen in liquid nitrogen and stored at −80 °C for up to 1 year.

Extraction of genomic DNA ● Timing 2 h with overnight incubation

▲ **CRITICAL** For *S. cerevisiae* cells, follow the 'Extraction of genomic DNA' section in Supplementary Protocol 1.

2 Thaw the cell pellets from Step 1A(v) or 1B(vii) on ice.

3 Resuspend cells in Lysis buffer to 1 million cells per milliter. Distribute 10 mL aliquots of cell suspension to 50 mL Falcon tubes. Place the tubes on a rack at RT.
▲ **CRITICAL STEP** Gently resuspend cells by pipetting up and down with a 10 mL pipet to minimize cell rupture and DNA shearing. A homogeneous cell suspension is necessary to ensure complete lysis and optimal DNA extraction.

4 Add 250 µL of 20% (wt/vol) SDS to the cell suspension. The final SDS concentration is 0.5% (wt/vol). Tightly close the cap and mix by gently inverting the tubes five to ten times.
▲ **CRITICAL STEP** Keep the tubes at RT during SDS addition. Invert the tubes gently to minimize DNA breaks.

5 Add 50 µL of proteinase K 20 mg/mL to the cell lysate. The final concentration of proteinase K is 0.1 mg/mL. Close the cap and mix by gently inverting the tube.
▲ **CRITICAL STEP** At this stage, the lysates will appear very viscous.

6 Incubate the tubes at 42 °C for 4 h or overnight (16 h).
▲ **CRITICAL STEP** After cell lysis is complete, the solution should appear homogeneous and transparent.
? TROUBLESHOOTING

7 In a chemical hood, add to each tube 10 mL of phenol–chloroform isoamyl alcohol mix solution pre-equilibrated at RT. Tightly close the cap and mix gently by inverting the tube until obtaining an entirely homogeneous mixture.
▲ **CRITICAL STEP** Bring the phenol–chloroform isoamyl alcohol solution to RT in advance.
▲ **CRITICAL STEP** Gently invert the tubes to allow the liquid to move between the cap and the bottom. Due to the high viscosity of the DNA solution, this step may require up to 10 min.
! CAUTION Perform the DNA extraction inside a chemical hood. Wear a lab coat and disposable gloves.

8 Centrifuge a 50 mL MaXtract High-Density tube at 1,500*g* at RT for 2 min, and pour the mixture from Step 7 into the tube.

9 Centrifuge for 4 min at 1,500*g* at RT with a swing rotor. This will separate the aqueous solution containing DNA while the organic phase will remain locked under the solid MaXtract gel phase.
▲ **CRITICAL STEP** Use of MaXtract High-Density tubes (or equivalent) is strongly recommended for achieving high-quality DNA preparation.

10 In a chemical hood, add to each tube 10 mL of phenol–chloroform–isoamyl alcohol mix. Tightly close the cap and mix gently by inverting the tube until full homogenization is achieved.
▲ **CRITICAL STEP** Ensure that the organic fraction from Step 9 remains locked under the MaXtract gel phase during this step.

11 Centrifuge for 4 min at 1,500*g* at RT. This will separate the aqueous solution containing DNA while the organic phenol phase will remain locked under the solid MaXtract gel phase.
▲ **CRITICAL STEP** If the aqueous phase after this step is not clear, perform an additional phenol-chloroform extraction by repeating Steps 7–9.

12 In the chemical hood, add to each tube 10 mL of chloroform. Tightly close the cap and mix gently by inverting the tube until full homogenization is achieved. Centrifuge for 4 min at 1,500*g* at RT.

13 Transfer the upper aqueous phase containing genomic DNA from all tubes by pouring into a clean 200 mL glass beaker.
▲ **CRITICAL STEP** Discard the organic fraction and the tubes in the appropriate chemical waste.

14 Add 2 mL of 7.5 M ammonium acetate per each 10 mL of lysate and mix gently with a Pasteur pipette.

15 Add 25 mL of absolute ethanol per each 10 mL of lysate and swirl gently with the same glass Pasteur pipette until the DNA precipitates.

16 Spool the precipitated DNA fibers with the Pasteur pipette and carefully transfer all the DNA precipitate into a clean 200 mL glass beaker containing 100 mL of 75% (vol/vol) of ethanol. Leave the DNA precipitate immersed for ~3–5 min. Repeat this step twice.
▲ **CRITICAL STEP** It may be convenient to recover the DNA precipitate using two Pasteur pipettes as chopsticks.

17 Place the DNA precipitate with the Pasteur pipettes inside a new 15 mL Falcon tube.

18 Remove any residual ethanol with a pipette fitted with a 1 mL tip.
19 Transfer the DNA precipitate to a new 15 mL tube and add 6 mL of TE.
▲ CRITICAL STEP Ensure the entire DNA precipitate is dipped in TE buffer. Do not pipette.
20 Leave the tube open for 30 min at 37 °C in a dry oven to allow the evaporation of residual ethanol.
21 Remove the Pasteur pipette and close the cap.
■ PAUSE POINT Store the DNA solution at 4 °C for at least 3–7 d to allow the complete dissolution of the DNA precipitate. Dissolved DNA can be stored for up to 1 month at 4 °C.

Size fractionation of denatured genomic DNA on neutral sucrose gradients ● Timing 3.5 h of handling and 17 h of centrifugation

▲ CRITICAL As the centrifugation lasts 17 h, it is convenient to start this step in the late afternoon.
22 Incubate the DNA solution from Step 21 at 37 °C for 1 h to diminish the viscosity.
23 Measure the DNA concentration with Qubit ds DNA BR Kit according to the manufacturer's protocol. Typically, a yield of ~2–3 mg of total DNA is expected.
? TROUBLESHOOTING
24 Split the volume into six equal aliquots of ~1–1.2 mL into 1.5 mL tubes using a 1 mL wide-bore tip.
▲ CRITICAL STEP If the yield of total DNA is higher than 3 mg, it is recommended to scale up the number of aliquots and gradient centrifugations accordingly.
▲ CRITICAL STEP The DNA solution is viscous and hard to pipette at this stage. Pipette slowly with a 1 mL wide-bore tip to minimize DNA shearing.
25 Prepare six linear sucrose gradients in Beckman Coulter Ultra clear tubes 25 × 89 mm by mixing 18 mL of 5% (wt/vol) TEN–sucrose and 18 mL of 30% (wt/vol) TEN–sucrose using a gradient master and following the gradient manufacturer's instructions (program 'Long_Sucr_05-30% _wv_1St' for SW32 rotor).
26 Place each tube containing the gradients in a centrifuge tube adapter (Beckman Ultra-high-speed centrifuge, Rotor SW28 or SW32) and keep them undisturbed.
▲ CRITICAL STEP Due to the bromophenol blue in 30% TEN–sucrose, a gradient of blue shade from the bottom to the top should be visible in the tube. If the blue gradient is not visible, discard the tube. Both Hoefer SG50 Gradient Maker and Gradient Master (Rotor: SW28; Program: Long_Sucr_05-30%_wv_1St) result in similar and acceptable size fractionation. We prefer Gradient Master as up to six highly uniform gradients can be simultaneously prepared within 15 min. Handle the gradients with care.
27 Heat DNA aliquots from Step 24 for 5 min at 94 °C to denature double-stranded DNA and chill immediately in an ice-cold water bath for 10 min.
28 Very carefully layer one aliquot of DNA from Step 27 on the surface of one gradient from Step 26 using a wide-bore tip. Load all gradients the same way.
29 Adjust the weight of the tubes (with adapter) at symmetric positions on the rotor (1 and 4; 2 and 5; 3 and 6). Balance the weight by adding the necessary amount of 5% TEN-sucrose to achieve the exact (≤0.1 g) weight balance. Pipette slowly drop by drop along the inner wall of the tube.
▲ CRITICAL STEP Any minor imbalance may lead to the tube or the rotor breaking.
▲ CRITICAL STEP Proceed immediately to the next step to avoid diffusion of the gradient.
30 Carefully close the caps, attach the adapters to the SW28/SW32 rotor and insert the rotor inside of the Beckman ultracentrifuge. Spin under vacuum for 17 h at 26,000 rpm at 20 °C, with acceleration and deceleration speed set on 'High'.
▲ CRITICAL STEP Keep an eye on the centrifuge for ~15 min after the program starts to display that the desired centrifuge speed has been achieved.
31 The next day once the centrifugation is finished, switch off the vacuum and open the lid.
32 Carefully transfer the adapters with the tubes to the rack. Open the adapter lids carefully.
▲ CRITICAL STEP Before collecting fractions, check the tube integrity. If the tube was broken during centrifugation the gradient should be discarded.
33 Number 18 15 mL Falcon tubes from 1 to 18.
34 Start collecting 1 mL fractions with a 1 mL wide-bore tip from the top of each gradient by slowly aspirating from the surface of the gradient. Combine fractions of the same order from all six gradients into a single 15 mL tube.
▲ CRITICAL STEP To collect the fractions, place a wide-bore tip vertically against the gradient surface and pipette slowly. Only pipette up from the surface of the gradient and never pipette down.
▲ CRITICAL STEP Usually the top eight 1 mL fractions contain DNA fragments of the desired size

Box 1 | Quality control of DNA size fractionation ● Timing 1 h

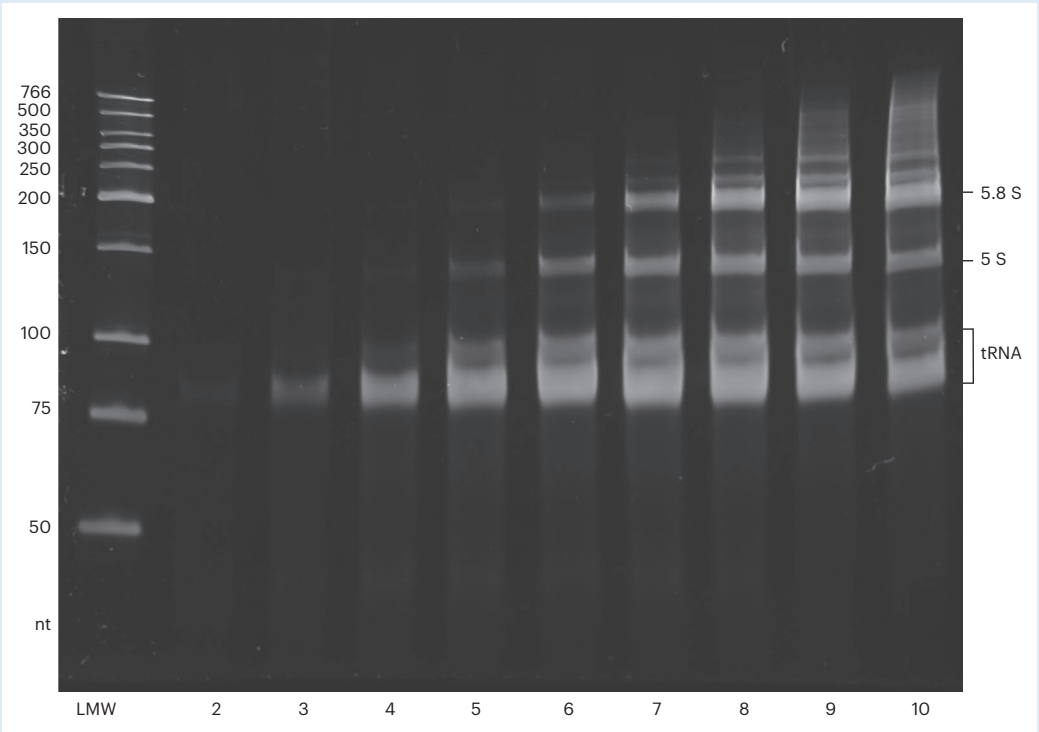
Procedure

1 Mix 10 µL of each gradient fraction from fraction 2 to 10 with 10 µL gel loading buffer II in a 1.5 mL tube.
2 Heat the tubes at 94 °C for 5 min.
3 Chill the tubes on ice for 5 min.
4 Set up a TBE–urea gel (10%, 1 mm, Thermo Fisher) on the vertical electrophoresis system with 1× TBE buffer. Flush carefully each well with 1× TBE buffer.
5 Prewarm the gel by running empty for 10 min at 400 V.
6 Quick spin the samples and load the entire volume to the wells.
7 Run at 180 V until the bromophenol blue reaches the bottom of the gel (usually 30–40 min).
8 Stain the gel by immersing in 20 mL of freshly prepared 1× Sybr Gold stain in TBE.
9 Visualize at a UV transilluminator.
10 Determine the gradient fractions containing the fragments of interest (≤250 nt).

▲ **CRITICAL STEP** The DNA size is increasing in the fractions from top to bottom. The tRNA and 5S rRNA serves as internal size markers. Typically, fractions 1–8 are combined to collect Okazaki fragments.

▲ **CRITICAL STEP** The quality control of gradient fractionation may also be assessed using 3% (wt/vol) TBE–agarose gels.

[Box 1 Figure legend] Quality control for DNA size fractionation. Representative electrophoresis in 10% urea PAGE. 2–10, second to tenth 1 mL gradient fractions; LMW, NEB low-molecular-weight marker.



(≤250 nt), but we suggest collecting more fractions to check the size distribution and linearity of the gradient fractionation (Box 1).

▲ **CRITICAL STEP** Observe the color of the fractions. As bromophenol blue distributes to the dense sucrose solution, the top fractions should be lighter, and the bottom fractions should appear progressively more colored.

▲ **CRITICAL STEP** If wide-bore tips are not available, cut the tips of 1 mL tips with clean scissors. Make sure the cut end is smooth and flat.

■ **PAUSE POINT** The fractions can be stored at 4 °C for 1–3 d or at −20 °C for up to 6 months.

35 Pool the fractions from Step 34 containing fragments smaller than 200–250 nt (typically the first one to eight fractions).

36 Concentrate the pooled fractions (48–80 mL) on a Millipore Amicon Ultra Centrifugal Filter, 15 mL, 10K. Add 15 ml of fractions to a centrifuging filter and centrifuge at 4,000*g* at RT for 10–15 min.

37 Discard the flowthrough and load the next 15 ml of the sample to the filter. Repeat centrifugations until the entire volume of fractions is concentrated to ~300 µL.

38 Buffer exchange by adding 5 ml of ultrapure water and centrifuge at 4,000*g* for 10 min. Discard the flowthrough. Repeat two more times.

39 Transfer the concentrated solution from the filter (~300 µL) to a new 1.5 mL tube. Measure the volume carefully with the pipette tip and note it on the tube.

PAUSE POINT The concentrated fractions can be stored at −20 °C for 2 weeks.

? TROUBLESHOOTING

Click biotinylation ● Timing 2 h

40 Add the following reagents in the specified order to the tubes containing purified gradient fractions from Step 39.

Reagent	Volume (µL)	Final
DNA	≤375 µL	
10× Click-it buffer (or 10× PBS pH 7.4)	50 µL	1×
100 mM biotin-TEG-azide	5 µL	1 mM
500 mM THPTA	10 µL	10 mM
100 mM $CuSO_4$	10 µL	2 mM
100 mM sodium ascorbate	50 µL	10 mM
ddH_2O	Up to 500 µL	

▲CRITICAL STEP If the volume of the concentrated fractions from Step 39 is >375 µL, scale up the volumes of all reagents accordingly.

▲CRITICAL STEP The THPTA and $CuSO_4$ should be premixed and added in a single pipetting step.

41 Mix by pipetting with a low-binding tip, quick spin and incubate the click reaction for 45 min in a thermoblock at 25 °C without mixing.

▲CRITICAL STEP Use freshly prepared sodium ascorbate.

42 Quick spin and split the reaction into two equal aliquots of 250 µL in two 1.5 mL Eppendorf tubes. Add 750 µL of absolute ethanol to each tube to precipitate DNA, close the caps and mix by inverting.

43 Chill the tubes at −80 °C for 15 min.

44 Spin for 30 min at ≥15,000*g* at 4 °C and decant the supernatant.

▲CRITICAL STEP The pellet can appear blue or brownish, probably due to the copper residue, which does not interfere with the experiment.

45 Add 500 µL of 75% (vol/vol) ethanol to the pellet, spin for 5 min at full speed at 4 °C. Decant the supernatant.

46 Quick spin and carefully remove the residual ethanol with a 200 µL tip without disturbing the pellet. Keep the tube open and air dry briefly (usually 2–5 min).

47 Dissolve each pellet in 45 µl of nuclease-free water and combine into a single 1.5 mL tube.

RNA Hydrolysis ● Timing 20 min

48 Add 10 µL 2.5 M NaOH into the 90 µL DNA from Step 47 to a final concentration of 250 mM, mix by pipetting, quick spin and incubate for 30 min at 37 °C

49 Quick spin and add 10 µL 2.5 M acetic acid to neutralize the pH and mix by pipetting.

50 Purify the DNA with 2× Bio-Rad Micro Biospin P-30 columns according to the manufacturer's instructions.

51 Combine the purified flowthrough from the two columns in one 1.5 mL tube.

52 Measure the volume of the solution using a 200 µL pipette tip. Place the tube on ice.

▲CRITICAL STEP Usually 120–150 µL DNA solution is recovered after purification.

DNA phosphorylation and precipitation ● Timing 1.5 h

53 Set up the phosphorylation reaction by adding the following reagents to the tubes containing purified DNA from Step 52. Mix by pipetting with a low-binding tip, quick spin, and incubate at 37 °C for 20 min.

Reagent	Volume (µL)	Final
DNA	≤117 µL	
10× T4 PNK buffer A	15 µL	1×
10 mM ATP	15 µL	1 mM
T4 PNK (10 U/µL)	3 µL	0.2 U/µL
ddH_2O	Up to 150 µL	

▲ **CRITICAL STEP** If the volume of the DNA from Step 52 is >117 µL, scale up the volumes of all reagents accordingly.

▲ **CRITICAL STEP** Use a fresh aliquot of ATP and avoid freezing–thawing cycles.

54 Incubate the tubes for 10 min at 75 °C to inactivate the T4 PNK enzyme.

55 Quickly spin the tubes and chill on ice.

56 To precipitate DNA, add 15 µL of 3 M sodium acetate (pH 5.2) and 415 µL of −20 °C chilled absolute ethanol, mix by inverting. Incubate for 15 min at −80 °C.

57 Centrifuge for 20 min at 4 °C ≥17,000*g*. Discard the supernatant.

58 Wash the pellet by adding 500 µL of 75% (vol/vol) ethanol without disturbing the pellet.

59 Centrifuge for 2 min at 4 °C ≥17,000*g*. Discard the supernatant.

60 Quick spin and remove all residual ethanol without disturbing the pellet.

61 Leave the tube open for 5 min to allow the residual ethanol to evaporate.

62 Dissolve the pellet in 20 µL of nuclease-free water and transfer to a 200 µL PCR tube. Place the tube on ice.

▲ **CRITICAL STEP** If the solution appears very viscous, add 60 µL of nuclease-free water to the DNA solution and transfer it to a 0.5 mL PCR tube. Scale up the volumes of all subsequent Steps 63–65 accordingly.

Hybridization and ligation of adapters, round 1 ● Timing 30 min to overnight

▲ **CRITICAL** For instructions on reannealing the adapters, see Box 2. Avoid the freeze–thaw cycles for the reannealed adapters.

63 Set up the reaction by adding the following reagents sequentially to the tube containing the purified phosphorylated DNA from Step 62. Use adapters reannealed as outlined in Box 2. Mix by pipetting and perform a quick spin.

Reagent	Volume (µL)
Phosphorylated DNA (Step 62)	20 µL
40 mM adapter A1 (Table 1 and Box 2)	2 µL
40 mM adapter A2 (Table 1 and Box 2)	2 µL

64 Incubate in a thermocycler using the following program:

Step	Temp	Time
Hybridization	65 °C	10 min
	16 °C	5 min

65 Take the tubes out of the thermocycler. Set up the ligation reaction by adding the following reagents to the tube:

Reagent	Volume (µL)	Final
10× T4 ligase buffer	4 µL	1×
50% PEG 8000 (wt/vol)	4 µL	5%
5 M betaine	4 µL	0.5 M
T4 DNA ligase 5 U/µL	4 µL	0.5 U/µl

Box 2 | Adapter preparation

Procedure

▲ **CRITICAL** To obtain double-stranded adapters A1 and A2 with single-stranded random hexamer overhangs, anneal the Adapter oligonucleotide 'top' with the adapter oligonucleotide 'bottom'.

1 Dissolve the adapter oligomers (Table 1) to 100 μM with nuclease-free H_2O and vortex to achieve complete dissolution.
2 Prepare two 200 μL PCR tubes labeled A1 and A2 for adapter 1 and adapter 2, respectively.
3 Assemble each adapter reannealing reaction in a PCR tube on ice by adding in the following order:

Reagent	A1	A2	Volume (μL)
Top strand 100 μM	$A1_{top}$ (R1)	$A2_{top}$	20 μL
Bottom strand 100 μM	$A1_{bottom}$	$A2_{bottom}$	20 μL
NaCl 5M			0.5 μL
Water			9.5 μL

4 Mix well by pipetting, quick spin and place the hybridization reactions in a thermal cycler: cool down from 94 °C to 16 °C at 0.1 °C/s.
5 Chill on ice and aliquot the annealed adapters into 5 μL aliquots.

■ **PAUSE POINT** Keep at −20 °C for up to 6 months. Avoid thaw-freezing to preserve the phosphorylation modifications on the oligomers.

▲ **CRITICAL STEP** Thaw on ice a fresh aliquot of 10× T4 ligase buffer. Avoid freeze–thawing the aliquots.

66 Mix by pipetting, quick spin and incubate at 16 °C in a thermocycler for 16 h.
▲ **CRITICAL STEP** The incubation can last overnight.

Streptavidin capture of biotinylated library fragments ● Timing 1 h

67 Vortex gently the stock of MyOne T1 streptavidin Dynabeads.
68 Pipette 20 μL of the bead suspension into a 1.5 mL low-binding tube. Capture the beads by placing the tube on the magnet for 1 min.
69 Remove and discard the supernatant with a 200 μL filter tip.
70 Remove the tube from the magnet and wash the beads by adding 200 μL of 1× BWT buffer, and mix by pipetting.
71 Place the tube on the magnet to pellet the beads. Incubate for 1 min or until all the beads are captured on the magnet.
72 Carefully remove and discard the supernatant with a 200 μL filter tip without disturbing the beads.
73 Repeat Steps 70–72 two more times.
74 Remove the tube from the magnet and resuspend the beads in 40 μL of 2× BWT buffer.
75 Add 40 μl of the bead suspension from Step 74 into the tube containing the ligation reaction from Step 66. Mix by pipetting with a low-binding tip.
76 Incubate the tube on a rotating platform at 20 rpm for 20 min at RT.
▲ **CRITICAL STEP** Resuspend the beads by gently flicking the tube every 5 min. Because of the small volume, sideways rotation of the tube is preferred rather than inversion.
77 Spin the tube briefly in a microcentrifuge and place the tube on the magnet to capture the beads. Transfer the supernatant to a new 1.5 mL tube labeled 'Supernatant 1' and keep it at −20 °C for the library construction quality control (Box 3)
78 Remove the tubes with the beads from the magnet, wash the beads by adding 200 μL of 1× BWT. Mix by pipetting with a 200 μL low-binding filter tip and transfer to a new 1.5 mL low-binding tube.
79 Place the tube on the magnet to capture the beads. Incubate until the liquid is clear. Remove and discard the supernatant with a 200 μL tip.
80 Repeat 1× BWT washing twice (Steps 78–79) without transferring the beads to a new tube.
81 Remove the tube from the magnet and wash the beads by adding 200 μl 1× TE + 0.05% (vol/vol) Tween 20 and mix.
82 Capture the beads on the magnet and remove the supernatant with a 200 μL pipette tip.
83 Repeat Steps 81–82 one more time.
84 Remove the tube from the magnet and wash the beads by adding of 200 μL of ddH_2O. Mix by pipetting.

85 Capture the beads on the magnet and remove the supernatant with a 200 µL tip.
86 Resuspend the beads in 10 µL ddH_2O and transfer the entire volume to a new 200 µL PCR tube. Place on ice and proceed immediately to the next step.

Ligation of adapters, round 2 ● Timing 4 h to overnight

87 Set up the second-round ligation reaction by adding the following reagents:

Reagent	Volume (µL)
Library bead suspension (Step 86)	10 µL
40 mM adapter A1 (Table 1 and Box 2)	1 µL
40 mM adapter A2 (Table 1 and Box 2)	1 µL

88 Mix by pipetting, quick spin and incubate in a thermocycler with the following program:

Step	Temperature	Time
Hybridization	65 °C	10 min
	16 °C	5 min

89 Take the tubes out of the thermocycler. Add the following reagents to the tube:

Reagent	Volume (µL)	Final
10× T4 ligase buffer	2 µL	1×
50% PEG 8000 (wt/vol)	2 µL	5%
5 M betaine	2 µL	0.5 M
T4 DNA ligase 5 U/µL	2 µL	0.5 U/µL

90 Mix by pipetting and quick spin. Incubate at 16 °C in a thermocycler for ≥2 h or overnight.
91 Prepare 10 µL of fresh ligation mix by mixing the following reagents in a tube on ice:

Reagent	Volume (µL)
ddH_2O	7 µL
10× T4 ligase buffer	1 µL
10 mM ATP	1 µL
T4 DNA ligase 5 U/µL	1 µL

92 Take the tube (Step 90) from the thermocycler, quick spin and capture the beads with the magnet.
93 Remove 10 µL of the supernatant without touching the bead pellet. Label the supernatant as 'Supernatant 2' and keep it at −20 °C for the quality control of library construction (Box 3).
94 Take the tube off the magnet and add 10 µL of the fresh ligation mix from Step 91. Mix by pipetting and quick spin. Incubate in the thermocycler for 1 h at 16 °C.
95 Capture the beads on the magnet. Carefully remove the supernatant without disturbing the beads.
96 Remove the tubes from the magnet, add wash by adding 200 µL of 1× BWT. Mix thoroughly by pipetting with a 200 µL low-binding filter tip and transfer the bead suspension to a new 1.5 mL low-binding tube.
97 Capture the beads on the magnet. Remove and discard the supernatant with a 200 µL tip.
98 Repeat washing Steps 96–97 four more times without transferring the beads to a new tube.
99 Remove the tube from the magnet and wash the beads by adding 200 µL 1× TE + 0.05% (vol/vol) Tween 20 and mix by pipetting.
100 Capture the beads on the magnet and remove the supernatant with a 200 µL tip. Repeat Step 99 once.

101 Remove the tube from the magnet, and wash the beads with 200 µL of nuclease-free water.
102 Capture the beads on the magnet and remove the supernatant with a 200 µL tip.
103 Resuspend the beads in 20 µL of EB and proceed to the quality control of library construction (Box 3).

PAUSE POINT The bead-bound library can be stored at −20 °C for up to 6 months.

Okazaki fragment library amplification ● Timing 1.5 h

104 Prepare the library amplification reaction in a low-binding 200 µL PCR tube as follows:

Component	Stock	Volume	Final
PEM1 (Table 1)	10 µM	1 µL	0.2 µM
Truseq_Index with the desired barcode (Table 1)	10 µM	1 µL	0.2 µM
KAPA HiFi Fidelity Buffer	5 ×	10 µL	1×
Bead suspension (Step 103)		5–10 µL	
KAPA dNTP Mix	10 mM	1.5 µL	0.3 mM
Taq Kapa HiFi Hotstart Polymerase	1 U/µL	0.5 µL	0.1 U/µL
H_2O		Up to 50 µL	

Box 3 | Quality control of library construction

Procedure

1 Assemble four amplification reactions in 4 PCR tubes on ice as follows:

Component	Stock	Volume	Final
PEM1 (Table 1)	10 µM	0.2 µL	0.1 µM
Truseq_Index (Table 1)	10 µM	0.2 µL	0.1 µM
Taq DNA polymerase buffer	10×	2 µL	1×
Template		1 µL	
dNTP Mix	10 mM	0.4 µL	0.2 mM
Taq DNA polymerase	5 U/µL	0.2 µL	0.05 U/µL
H_2O		Up to 20 µL	

2 Add 1 µL of the following templates to each PCR reaction tube: (1) 1 µl of nuclease-free H_2O (negative control); (2) 1 µL of the bead suspension with bound adapter-ligated library (Step 103); (3) 0.2 µL of ligation supernatant 1 (Step 77) plus 0.8 µL nuclease-free H_2O; (4) 1 µL ligation supernatant 2 (Step 93).
3 Amplify using the following cycling protocol:

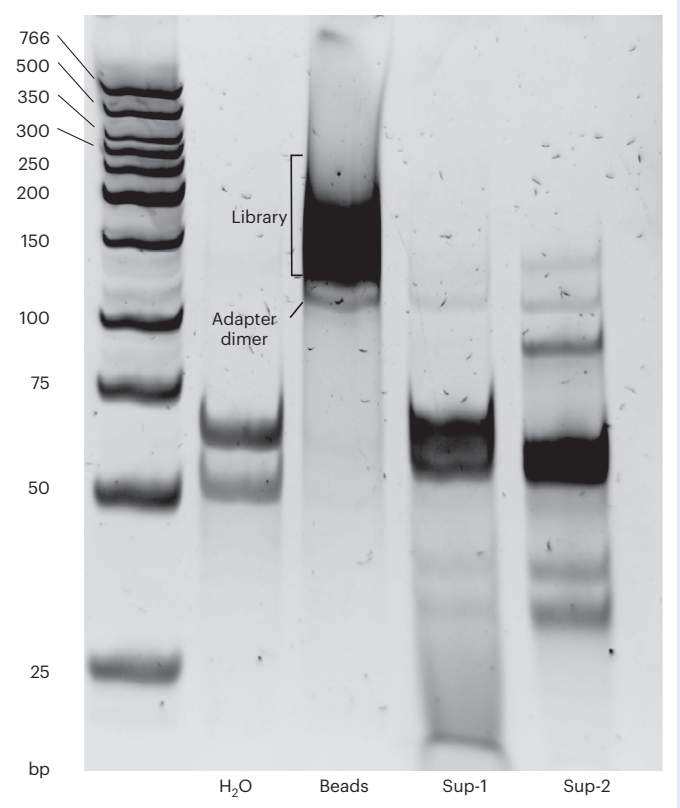
Step	Temp	Duration	Cycles
Initial denaturation	98 °C	45 s	1
Denaturation	98 °C	15 s	25–30
Annealing	60 °C	30 s	
Extension	72 °C	30 s	
Final extension	72 °C	1 min	1
Hold	4 °C	∞	

4 Prepare a 10% TBE PAGE gel.
5 Mix 10 µL of PCR product (Step 3) with 2 µL of 6× purple loading dye, and load the mix into the gel. Run the gel until the bromophenol blue reaches the bottom of the gel.
6 Stain the gel by immersing in 20 mL of freshly prepared 1× SybrGold for 5 min.
7 Visualize at a UV transilluminator and compare the lanes.

▲ CRITICAL STEP In the PCR reaction run with the bead-bound adapter-ligated library (lane 2), the 128 bp band corresponds to the self-ligated adapter dimers and the smear above contains the library with inserts. As an indicator of a successful library, the dimer band has to be visible but less prominent than the library smear. In PCR reactions run with the supernatants 1 and 2, no or very little smear above 128 bp is observed (lanes 3 and 4).

Box 3| (continued)

[Box 3 Figure legend] Quality control for the library construction. Representative electrophoresis in 10% TBE PAGE. 'LMW', NEB low molecular weight marker. 'H_2O', PCR reaction run without template (negative control). 'Beads', PCR reaction run with the bead-bound library. 'Sup 1' and 'Sup2', PCR reactions run with supernatants 1 and 2.



105 Incubate the PCR reaction in a thermocycler with the following program:

Step	Temp	Duration	Cycles
Initial denaturation	98 °C	45 s	1
Denaturation	98 °C	15 s	10
Annealing	60 °C	30 s	
Extension	72 °C	30 s	
Final extension	72 °C	1 min	1
Hold	4 °C	∞	

▲ **CRITICAL STEP** To minimize the generation of PCR duplicates, we do not recommend to exceed 12 amplification cycles. Usually, a ten-cycle library amplification synthesizes enough material for QC and sequencing.

106 Take out the tubes from the thermocycler, quick spin and place on the magnet to collect the beads.

107 Transfer the supernatant containing the amplified library into a new 1.5 mL low-binding tube without disturbing the beads.

108 Wash the streptavidin beads with 200 µL of EB + 0.05% (vol/vol) Tween 20 and resuspend in 20 µL of EB. Store at −20 °C for up to 1 year. These beads can be reused for an additional round of library amplification.

▲ **CRITICAL STEP** Okazaki fragment library amplification (Steps 104–107) can be performed once

more using the same beads as a template. Typically, this second amplification increases the final library yield without affecting the library complexity.

■ **PAUSE POINT** The beads can be stored at −20 °C for up to 1 year and the PCR product could be stored at 4 °C for 72 h or at −20 °C for up to 6 months.

Post-amplification clean-up ● Timing 1 h

109 Take the stock of AMPure XP beads out of the fridge 30 min in advance.

110 Perform cleanup of the PCR product (Step 107) by adding 75 µL of AMPure XP bead suspension (bead ratio 1.5×).

111 Vortex thoroughly. Incubate at RT for 10 min to bind DNA to the beads.

112 Capture the beads on the magnet. Carefully remove and discard the supernatant with a 200 µL filter tip.

113 Keeping the tubes on the magnet, wash the beads by adding 200 µL of freshly prepared 80% (vol/vol) ethanol and incubate at RT for at least 30 s. Carefully remove and discard the supernatant with a 200 µL filter tip.

114 Repeat Step 113 once. Remove all residual ethanol without disturbing the beads.

▲ **CRITICAL STEP** Do not overdry the beads as it will be difficult to elute the DNA.

115 Resuspend the beads in 10.5 µL of EB.

116 Incubate the open tubes in a thermomixer for 5 min at 37 °C to elute DNA. Cover the thermomixer with a clean lid or a piece of aluminum foil to protect the tubes from dust.

117 Capture the beads on the magnet.

118 Transfer 10 µL of the supernatant (containing the library) to a 1.5 mL low-binding tube without taking any beads.

■ **PAUSE POINT** The purified library can be stored at 4 °C overnight or −20 °C for up to 1 year.

Size selection on agarose gel ● Timing 2 h

▲ **CRITICAL** Size selection is a critical step for optimal sequencing results.

119 Prepare a 4% (wt/vol) agarose gel (15 cm × 15 cm) in 1× TAE buffer.

▲ **CRITICAL STEP** The electrophoresis tank should be rinsed with deionized water in advance, and fresh 1× TAE buffer should be used for electrophoresis.

120 Mix 10 µL eluted DNA (Step 118) with 2 µL 6× purple gel loading dye and 1 µL SYBR Green I (100×), and load the mix into the gel. Load a DNA ladder ranging between 20 bp and 1,000 bp (e.g., NEB low molecular weight ladder, or equivalent).

121 Run the gel until bromophenol blue reaches ¾ of the gel length.

122 Visualize the gel on a non-UV light transilluminator and cut the bands between 150 bp and 400 bp with a clean blade.

▲ **CRITICAL STEP** Do not use UV light as it damages DNA and may impact the sequencing quality.

▲ **CRITICAL STEP** A gap should be visible between the primer dimer (128 bp) and the shortest library fragments (135–140 bp). Do not touch the 128 bp band with the blade as it may lead to contamination with primer dimers.

123 Purify the DNA from the gel with the Qiagen Minelute Gel extraction kit according to the manufacturer's manual, but dissolve the agarose block at RT with gentle shaking instead of heating at 50 °C. Elute DNA with 10 µL EB buffer and proceed to the quality control of the library size selection (Box 4)

■ **PAUSE POINT** The size-selected and purified library could be stored at −20 °C for up to 1 year.

? TROUBLESHOOTING

Sequencing ● Timing variable

124 Prepare the library pool and dilution according to Illumina protocols.

125 Perform Illumina sequencing. During the run setup, load the custom sequencing primer for read 1 (Primer $A1_{top}$ (R1), Table 1).

▲ **CRITICAL STEP** Since the A1 adapter is shortened by 5 bp, the custom read 1 sequencing primer ($A1_{top}$ (R1), Table 1) must be loaded to the flowcell (following standard Illumina recommendations). Indicate to the sequencer program that a custom primer for read 1 was used before starting to run the program.

Box 4 | Quality control of the library size selection

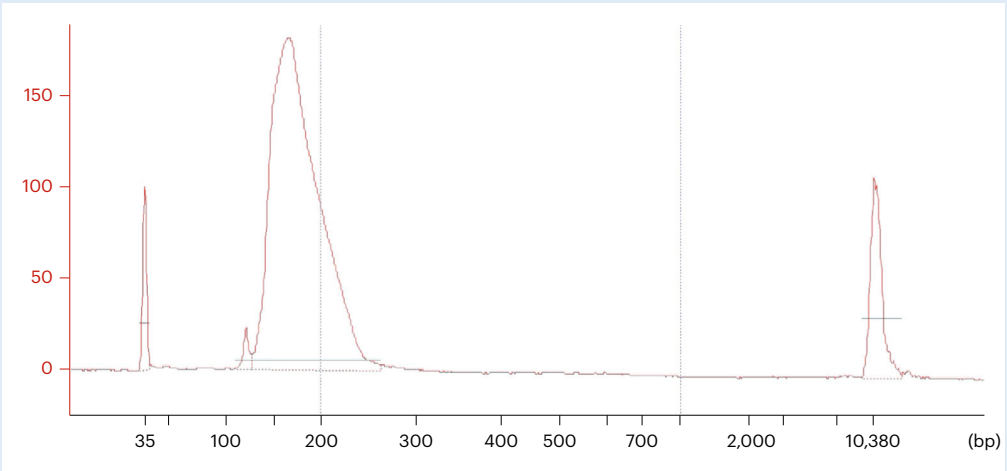
Procedure

1 Measure the library concentration of the size-selected and purified libraries using a Qubit dsDNA HS Kit following the manufacturer's recommendations. Typically, the library concertation ranges between 0.4 ng/μL and 2 ng/μL.

2 Check the fragment size distribution by running 1 μL on an Agilent Bioanalyzer High Sensitivity DNA Chip. A typical size-selected library ranges between 145 bp and 250 bp.

? TROUBLESHOOTING

[Box 4 Figure legend] Quality control for library size selection. Representative profile of OK-seq libraries obtained by Agilent Bioanalyzer. An average library size of 145–250 bp is expected.



Data processing ● Timing variable

▲**CRITICAL** Data processing typically takes ~12 h (tested with a classical desktop configuration: 3.5 GHz Intel Core i5 CPU with four cores for iMAC and 16 Go DDR4 2400 MHZ speed memory; for a dataset of ~300 million total reads).

126 Prepare/download the aligned sequencing data in .bam files.

▲**CRITICAL STEP** The current protocol starts from the aligned data, which can be processed following standard procedures and are frequently provided by sequencing facilities. Briefly, the raw sequencing data (.fastq) need to be preprocessed into genome-aligned files with the following major steps: fastqc[93] for checking the quality of reads, cutadapt[94]/Trim Galore[95]/Trimmomatic[96] for trimming adapters and low-quality reads, BWA[97]/Bowtie2[98] for read alignment, then Picard[99,100] for marking and deleting the duplicates, samtools[90] for sorting and indexing the aligned files.

127 Download the OKseqHMM toolkit from https://github.com/CL-CHEN-Lab/OK-Seq containing the necessary R scripts for the following analysis steps.

▲**CRITICAL STEP** The toolkit will count read matrices from aligned .bam files and calculate and output RFD and OEM profiles for a primary visualization (e.g., with IGV).

The R package OkseqHMM generates replication IZs (upward transitions of RFD profile), TZs (downward transitions of RFD profile), and two intermediate states (flat RFD profiles of low and high values, regions of leftward and rightward unidirectional replication, respectively) (MEP_L_fig3Fig. 3).

Generating the output files for visualization of RFD profile and IZ/TZ calling by a four-state HMM

▲**CRITICAL** Besides the aligned .bam files with the corresponding indexed file (.bai), the OKseqHMM function requires the annotation coordinates for all chromosomes and their lengths.

128 Download the annotation file containing all chromosomes and their lengths from the UCSC server (e.g., hg19.chr.sizes.txt for human hg19): ftp://hgdownload.cse.ucsc.edu/goldenPath/.

The program identifies automatically if the input.bam file is paired-end or single-end sequencing data, then splits the mapped reads within the.bam file into Watson (W) and Crick (C) strands, respectively, and calculates the read coverage and RFD along the reference genome. The bin size (with bin size parameter) can be defined by users depending on the data coverage and genome

size, and based on our experience, a 1 kb bin size is recommended for OK-seq data of human/mouse cells, and 50 bp bin size is recommended for budding yeast data.
After downloading the R scripts from GitHub, run this command line in the terminal:

```
source("PATH/OKseqHMM.R")
```

▲ **CRITICAL STEP** Before executing this function, make sure that R and the necessary R packages HMM, Rsamtools and Genomic Alignments are installed in your R working environment. Then you can either use the command line as a source ('`PATH/OKseqHMM.R`'), in which the PATH provides the PATH in your computer to the downloaded R package '`OKseqHMM.R`', or you can load the package directly into RStudio.
▲ **CRITICAL STEP** Make sure that the chromosome coordinates within the .bam file match the ones provided in the chromosome annotation file. Different sources of the reference genome having slightly different chromosome names may cause an error (e.g., sometimes '1-22, MT' is used in the .bam file while in the annotation file it is 'chr1-chr22, chrM' if you use the UCSC annotation).

129 Run OKseqHMM with the following options:

- For the human data:

```
OKseqHMM(bamfile = "my.bam", thresh = 10, chrsizes = "hg19.chr.size.txt", binSize=1000, winS=15, fileOut = "my_hmm"))
```

- For the yeast data:

```
OKseqHMM(bamfile = "my.bam", thresh = 1, chrsizes = " sacCer3.chrom.sizes.txt", binSize=50, winS=15, fileOut = "my_hmm"))
```

`'My.bam'` is your input path of the .bam file;
`'thresh'` is the threshold to eliminate the low read coverage bins;
`'chrsizes'` is your path linked with the annotation file containing the length of each chromosome;
`'binSize'` is the adjacent bin size in bp to calculate the read coverage and RFD;
`'winS'` is the smoothing window size for the HMM calling;
`'fileOut'` is the path of storage as well as the prefix of the name for your output files.
▲ **CRITICAL STEP** Bin size may need to be adjusted relative to the genome size of the analyzed species and the coverages of your data.

? TROUBLESHOOTING

130 After executing the OKseqHMM shown above, this function will automatically generate a series of output files including:

- Two .bam files, and their corresponding index .bai files, for the reads generated from the Watson and Crick strands, respectively
- Two bedgraph files containing RFD values in the adjacent windows defined by `'binSize'` and in the smoothed windows defined by `'winS'` (`'_RFD.bedgraph'`)
- log file (`'_log.txt'`) that records all of the parameters you use and also the default setting information
- HMM result in a text file (`'_HMM.txt'`) that records all of the global optimal hidden states calculated by the HMM Viterbi algorithm
- HMM result in a text file (`'_HMMpropa.txt'`) that records all of the previous state positions that caused the maximum local probability of a state by the HMM posterior algorithm
- Eight text files recording the genomic positions (.bed) and the corresponding probabilities (.txt) for the final identified optimal states:
- `'_HMMsegments_IZ.bed/txt'` contains the replication IZ calling result
- `'_HMMsegments_TZ.bed/txt'` contains the replication termination zone calling result
- `'_HMMsegments_highFlatZone.bed/txt'` and `'_HMMsegments_LowFlatZone.bed/txt'` are the results of two intermediate flat states (constant RFD transition regions)

▲ **CRITICAL STEP** RFD bedgraph files can be visualized directly in a genomic browser, e.g., IGV[87].
▲ **CRITICAL STEP** You can also further transform the bedgraphs into bigwig by the UCSC tool

Box 5 | Additional parameters of the OKseqHMM toolkit

To run the OKseqHMM function, one needs to predefine the initial start probabilities for the four states of HMM ('D' is downward state, 'L' is low-flat state, 'H' is high-flat state, 'U' is upward state) and fivr observation symbols ('sym'), including the transition matrix ('ptrans') containing the probabilities that the four states transit from one to another (e.g., the first four values in 'ptrans' matrix show that the transition probabilities of state 'U' turn into states 'U', 'H', 'L' and 'D', respectively), the emission probability matrix ('pem') between states and observations (the emission probability represents how likely RFD transitions between adjacent windows of a given region are to match a hidden state, e.g., the first five values in 'pem' matrix show that the emission probabilities of state 'D' are emitted from the five observations, respectively), and the five quantiles of RFD ('quant') as follows:

```
st=c("D", "L", "H", "U"), sym=c("V", "W", "X", "Y", "Z"), pstart=rep(1/4, 4),

pem=t(matrix(c(0.383886256, 0.255924171, 0.170616114, 0.113744076, 0.075829384,

.10,.20,.40,.20,.10,

.10,.20,.40,.20,.10,

0.022222222, 0.033333333, 0.066666667, 0.211111111, 0.666666667),

ncol=4)),

ptrans=t(matrix(c(0.9999,0.000020,0,0.000080,

0,0.999,0,0.001,

0.001,0,0.999,0,

0.000080,0,0.000020,0.9999),

ncol=4)).

quant=c(-1, -0.0082058939609862, -0.00141890249101162, 0.00103088286465956, 0.00800467305420799, 1))
```

These parameters and probabilities were validated with the OK-seq dataset of HeLa cells[28]. We have successfully applied them to different human, mouse and yeast OK-seq datasets, which all got satisfactory results with these presetting parameters. Therefore, the users should be able to use these default settings without modifications. However, users could modify these parameters to optimize the results for their dataset, for example, on the basis of the distribution of deltaRFD per chromosome of the corresponding dataset, one can adjust the 'quant' parameter as well as 'ptrans' and 'pem'.

bedGraphToBigWig (http://hgdownload.soe.ucsc.edu/admin/exe/) to get binary compressed files by running the command line in Shell:

```
bedGraphToBigWig in.bedGraph chrom.sizes out.bw
```

▲ **CRITICAL STEP** Additional details about the parameters are listed in Box 5.

Generating the output files for visualization of the RFD transitions

▲ **CRITICAL** The OKseqOEM function allows investigating the local origin efficiency metrics (i.e., deltaRFD)[30] at multiple scales.

131 Download the R script 'OKseqOEM.R' from GitHub and run this command line in the terminal:

```
source("PATH/OKseqOEM.R")
```

Use the following options:

- For the human data:

```
OKseqOEM(bamInF="path_to_bam_Forward_strand", bamInR="path_to_bam_
Reverse_strand", chrsizes="hg19.chr.size.txt", fileOut="path/name_
of_my_OEM",binSize=1000, binList=c(1,10,20,50,100,250,500,1000))
```

- For the yeast data:

```
OKseqOEM(bamInF="path_to_bam_Forward_strand", bamInR="path_to_bam_
Reverse_strand", chrsizes="sacCer3.chrom.sizes.txt", fileOut="path/
name_of_my_OEM", binSize=50, binList=c(1,20,100,200,300,400,500))
```

'bamInF' and 'bamInR' are the paths to the two .bam files of the Watson and Crick strand, respectively, generated by the OKseqHMM function.

'chrsizes' is the path to the annotation coordinates containing chromosome length information 'fileOut' is the path of storage as well as the prefix of the name given by the user (e.g., ~/Desktop/Okseq_results/my_HMM) for the output file.

'binsize' is to define the adjacent bin size in bp to calculate the read coverage for RFD 'binList' is to define a series of window sizes as different visualization scales that you would like to output the OEM results (e.g., for yeast, you will get OEM files at 50 bp, 1 kb, 5 kb, 10 kb, 15 kb, 20 kb and 25 kb window scales if you set binsize = 50 and binList = c(1, 20, 100, 200, 300, 400, 500) by multiplying 'binsize' with each element defined in 'binList').

? TROUBLESHOOTING

132 After executing OKseqOEM above, this function will automatically generate a series of wiggle (.wig) files calculated by using different sliding window sizes defined by 'binList'. Convert wiggle to bigwig format by executing in Shell the UCSC tool wigToBigWig (http://hgdownload.soe.ucsc.edu/admin/exe/) for the visualization:

```
wigToBigWig in.wig chrom.sizes out.bw
```

Generating the output files for the average profile and heatmap of RFD values around genomic regions of interest

▲ **CRITICAL** The shell-based script 'average_profile_heatmap.sh' contains the template on how to use deepTools[86] to generate the average profile and heatmap around genomic regions of interest (e.g., around transcription start sites, transcription termination sites, within annotated genes, around IZs) by using the 'computeMatrix' and 'plotProfile'/'plotHeatmap' functions. You can use these functions to define the upstream and downstream borders and the gene body length and to modify the other parameters indicated in the script.

▲ **CRITICAL** Since deepTools is a Python-based tool, the Python environment should be activated from Steps 133 to 136. Make sure that you have already installed deepTools[86] and the Python environment before running the scripts. The latest Python version could cause some incompatibility issues with deepTools[86]. Refer to the deepTools manual for different functions and set up the parameters (https://deeptools.readthedocs.io/en/develop/index.html).

133 Compute the matrix of values by running the following command line in the terminal or in built-in terminal of RStudio:

```
computeMatrix scale-regions --regionsFileName {your bed file of interested regions/genes PATH e.g.codingGenes.bed} --beforeRegionStartLength {e.g. 10000} --afterRegionStartLength {e.g. 10000} --regionBodyLength {e.g. 20000} --binSize {e.g. 1000} --scoreFileName {RFD bigwig file PATH e.g. Hela.EdC.Combined_OkaSeq.RFD.bw} --outFileName {e.g. "OUTPUT.matrix"} --missingDataAsZero -skipZeros
```

134 For obtaining the average RFD profile, run the 'plotProfile' function as follows:

```
plotProfile --matrixFile {e.g. "OUTPUT.matrix"} --outFileName {e.g. "RFD_averageProfile.stGeneLength.png"} --averageType mean --startLabel {e.g. start/TSS} --endLabel {e.g. end/TTS} --plotType se
```

135 For OEM, proceed following this example, which generates the matrix containing OEM values around the center of the IZ with the extension of ±100 kb in different scales (from 1 kb to 1 Mb). The bigwig files used in the example can be found at https://github.com/CL-CHEN-Lab/OK-Seq/tree/master/published_results/HeLa:

```
computeMatrix reference-point --regionsFileName {your IZ bed file PATH e.g. HeLa_hmm_HMMsegments_IZ.bed} --beforeRegionStartLength {e.g. 100000} --afterRegionStartLength {e.g. 100000} --binSize {e.g. 1000}
```

```
--scoreFileName {series of OEM bigwig file PATH e.g.
20130819CGM130726.Hela_OEM_10kb.bw 20130819CGM130726.He-
la_OEM_20kb.bw 20130819CGM130726.Hela_OEM_50kb.bw
20130819CGM130726.Hela_OEM_100kb.bw 20130819CGM130726.He-
la_OEM_250kb.bw 20130819CGM130726.Hela_OEM_500kb.bw
20130819CGM130726.Hela_OEM_1Mb.bw} --outFileName {e.g. "OUTPUT.ma-
trix"} --missingDataAsZero --skipZeros --referencePoint center
```

136 To plot the RFD profile and heatmap, use the matrix calculated by 'computeMatrix' and run 'plotHeatmap':

```
plotHeatmap --matrixFile {e.g. "OUTPUT.matrix"} --outFileName {e.g.
"OEM_sortbyLength.png"} --whatToShow "plot, heatmap and colorbar"
--refPointLabel center --samplesLabel {e.g. "HeLa 10kb" "HeLa 20kb"
"HeLa 50kb" "HeLa 100kb" "HeLa 250kb" "HeLa 500kb" "HeLa 1Mb"} --sortUs-
ing region_length --sortRegions ascend
```

Timing

Step 1, cell culture, EdU labeling and cell harvesting: 2–7 d of cell culture, 2 h of labeling and harvesting
Steps 2–21, extraction of genomic DNA: 2 h with overnight incubation
Steps 22–39, size fractionation of denatured genomic DNA on neutral sucrose gradients: 3.5 h of handling and 17 h of centrifugation
Steps 40–47, click biotinylation: 2 h
Steps 48–52, RNA hydrolysis: 20 min
Steps 53–62, DNA phosphorylation and precipitation: 1.5 h
Steps 63–66, hybridization and ligation of adapters, round 1: 30 min to overnight
Steps 67–86, streptavidin capture of biotinylated library fragments: 1 h
Steps 87–103, hybridization and ligation of adapters, round 2: 4 h to overnight
Steps 104–108, Okazaki fragment library amplification: 1.5 h
Steps 109–118, post-amplification cleanup: 1 h
Steps 119–123, library size selection: 2 h
Steps 124–125, sequencing: variable
Steps 126–127, data processing: variable
Steps 128–130, generating the output files for visualization of RFD profile and the initiation/termination zone calling by HMM: variable
Steps 131–132, generating the output files for visualization of the RFD transitions: variable
Steps 133–136, generating the output files for the generation of the average profile and heatmap of RFD values around the regions of interest: variable

Troubleshooting

Troubleshooting advice can be found in Table 2.

Q15

Table 2 | Troubleshooting table

Step	Problem	Possible reason	Solution
6	Nonhomogeneous or nontransparent solution	Cell aggregation formed before cell lysis, and/or inadequate proteinase K treatment	Thoroughly resuspend the cells before adding SDS. Add additional proteinase K to 0.1 mg/mL, invert gently to mix well, and incubate at 42 °C for an additional 2 h
23	Incomplete DNA dissolution	Ethanol residue and/or insufficient dissolution time	Incubate opened tubes with DNA solution at 37 °C for 1 h. Carefully resuspend with a wide-bore tip

Table continued

Table 2 (continued)

Step	Problem	Possible reason	Solution
39	Final volume >375 µL	Insufficient centrifugation and/or presence of polysaccharides	Spin for an additional 10 min in Step 39
			Scale up the reagents in Step 40
123	Prominent adapter-dimer peak	Low-complexity library and/or insufficient gel size selection	Amplify the library again with the beads from Step 108
			Perform a double-size selection of the library with Ampure beads, ratio 1:1.25 (if the total library amount is >10 ng)
	Smear containing libraries is absent or very weak	An insufficient number of starting cell	Increase starting cell number
			Use flow cytometry to ensure the cells are EdU labeled. Check the fraction of cells in S-phase and EdU-positive cells)
			For cell lines or conditions having less than 20% of cells in the S phase, increase the starting number of cells
			As a control, perform OK-seq on a well-proliferating cell line in parallel (HeLa)
129	Function execution interrupted by error	The prerequired R packages are not installed	Install all R packages and make sure all input fields are filled before execution
		Incomplete parameters	Check whether each parameter in the function has been defined with a proper value
		Inappropriate 'thresh' value	Check the statistical summary of the input mapping file to define a rational 'thresh' value
		Different annotations are used in the aligned files and 'chrsizes'	Check whether the chromosome names in your aligned files are consistent with your input annotation
		Inappropriate 'binSize'	Set a smaller 'thresh' or a larger 'binSize' if the sequencing depth is low
131	Function execution was interrupted by error	Incomplete setting of parameters	Complete all the required fields before execution
		Inappropriate 'binSize' and 'binList'	Modify the values of 'binSize' and test the scales of 'binList' based on the data

Anticipated results

DNA size fractionation

Genomic DNA preparation from 3×10^8 to 10×10^8 human cells typically yields 2–3 mg DNA, which is then denatured and size-fractionated on 4–6× sucrose gradients. When visualizing the DNA in each 1 mL fraction, the DNA size linearly increases in the fractions from top to bottom (Box 1). Typically, Okazaki fragments (<200 nt) are present in the top 1 mL fractions 1–8. It is important to avoid contamination from the lower fractions containing high molecular weight labeled nascent replicated strands.

Library size distribution

The library fragment size should range from 150 to 300 bp. To evaluate if the library preparation is successful a PCR control can be performed. A smear >140 bp containing the library with inserts should be more prominent than the adapter dimer (at 128 bp) (Box 3). After gel size selection, ideally no or very few adapter dimers should be present (Box 4). If the dimer peak is more abundant than the smear, this is an indication of a low-complexity library, which will require repeating the size-selection step and may impact the data quality.

Sequencing results

The examples of sequencing results of OK-seq in yeast and human cells are shown in Fig. 3. RFD profiles are calculated on the basis of the proportion of the read counts from the Crick and Watson genomic strands and reflect the locus-specific average fork direction (Fig. 3). HMM detection of RFD transitions identifies the initiation and termination zones. The automated approach used by OKseqHMM efficiently detects site-specific (yeast) and broad zones (human cells) of replication initiation events and the regions of predominantly unidirectional fork movement (flat segments). Applying OKseqOEM allows the assessment of local initiation efficiency at different scales (Fig. 3).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Published available OK-seq raw and processed datasets analyzed in this work are available in SRA: SRP065949 (HeLa cells)[28] and ENA: PRJEB36782 (*S. cerevisiae*)[31].

Code availability

The bioinformatics tool and all example datasets underlying this paper are available at the following GitHub page: https://github.com/CL-CHEN-Lab/OK-Seq with DOI number: https://doi.org/10.5281/zenodo.7056979.

References

1. Huberman, J. A. & Riggs, A. D. On the mechanism of DNA replication in mammalian chromosomes. *J. Mol. Biol.* **32**, 327–341 (1968).
2. Hamlin, J. L., Mesner, L. D. & Dijkwel, P. A. A winding road to origin discovery. *Chromosome Res.* **18**, 45–61 (2010).
3. Hyrien, O. Peaks cloaked in the mist: the landscape of mammalian replication origins. *J. Cell Biol.* **208**, 147–160 (2015).
4. Hulke, M. L., Massey, D. J. & Koren, A. Genomic methods for measuring DNA replication dynamics. *Chromosome Res.* https://doi.org/10.1007/s10577-019-09624-y (2019).
5. Lebofsky, R., Heilig, R., Sonnleitner, M., Weissenbach, J. & Bensimon, A. DNA replication origin interference increases the spacing between initiation events in human cells. *Mol. Biol. Cell* **17**, 5337–5345 (2006).
6. Demczuk, A. et al. Regulation of DNA replication within the immunoglobulin heavy-chain locus during B cell commitment. *PLoS Biol.* **10**, e1001360 (2012).
7. Anglana, M., Apiou, F., Bensimon, A. & Debatisse, M. Dynamics of DNA replication in mammalian somatic cells: nucleotide pool modulates origin choice and interorigin spacing. *Cell* **114**, 385–394 (2003).
8. Cadoret, J. C. et al. Genome-wide studies highlight indirect links between human replication origins and gene regulation. *Proc. Natl Acad. Sci. USA* **105**, 15837–15842 (2008).
9. Besnard, E. et al. Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat. Struct. Mol. Biol.* **19**, 837–844 (2012).
10. Karnani, N., Taylor, C. M., Malhotra, A. & Dutta, A. Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. *Mol. Biol. Cell* **21**, 393–404 (2010).
11. Mukhopadhyay, R. et al. Allele-specific genome-wide profiling in human primary erythroblasts reveal replication program organization. *PLoS Genet.* **10**, e1004319 (2014).
12. Langley, A. R., Gräf, S., Smith, J. C. & Krude, T. Genome-wide identification and characterisation of human DNA replication origins by initiation site sequencing (ini-seq. *Nucleic Acids Res.* **44**, 10230–10247 (2016).
13. Mesner, L. D. et al. Bubble-chip analysis of human origin distributions demonstrates on a genomic scale significant clustering into zones and significant association with transcription. *Genome Res.* **21**, 377–389 (2011).
14. Mesner, L. D. et al. Bubble-seq analysis of the human genome reveals distinct chromatin-mediated mechanisms for regulating early- and late-firing origins. *Genome Res.* https://doi.org/10.1101/gr.155218.113 (2013).
15. Hansen, R. S. et al. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl Acad. Sci. USA* **107**, 139–144 (2010).
16. Chen, C. L. et al. Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. *Genome Res.* **20**, 447–457 (2010).
17. Zhao, P. A., Sasaki, T. & Gilbert, D. M. High-resolution Repli-Seq defines the temporal choreography of initiation, elongation and termination of replication in mammalian cells. *Genome Biol.* **21**, 76 (2020).
18. Koren, A. et al. Genetic variation in human DNA replication timing. *Cell* https://doi.org/10.1016/j.cell.2014.10.025 (2014).
19. Hulke, M. L., Massey, D. J. & Koren, A. Genomic methods for measuring DNA replication dynamics. *Chromosome Res.* **28**, 49–67 (2020).
20. Lobry, J. R. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.* **13**, 660–665 (1996).
21. Touchon, M. et al. Replication-associated strand asymmetries in mammalian genomes: toward detection of replication origins. *Proc. Natl Acad. Sci. USA* **102**, 9836–9841 (2005).
22. Huvet, M. et al. Human gene organization driven by the coordination of replication and transcription. *Genome Res.* **17**, 1278–1285 (2007).
23. Chen, C. L. et al. Replication-associated mutational asymmetry in the human genome. *Mol. Biol. Evol.* **28**, 2327–2337 (2011).
24. Audit, B. et al. Open chromatin encoded in DNA sequence is the signature of 'master' replication origins in human cells. *Nucleic Acids Res.* **37**, 6064–6075 (2009).
25. Guilbaud, G. et al. Evidence for sequential and increasing activation of replication origins along replication timing gradients in the human genome. *PLoS Comput. Biol.* **7**, e1002322 (2011).

26. Baker, A. et al. Replication fork polarity gradients revealed by megabase-sized U-shaped replication timing domains in human cell lines. *PLoS Comput. Biol.* **8**, e1002443 (2012).
27. Green, P., Ewing, B., Miller, W., Thomas, P. J. & Green, E. D. Transcription-associated mutational asymmetry in mammalian evolution. *Nat. Genet.* **33**, 514–517 (2003).
28. Petryk, N. et al. Replication landscape of the human genome. *Nat. Commun.* **7**, 10208 (2016).
29. Smith, D. J. & Whitehouse, I. Intrinsic coupling of lagging-strand synthesis to chromatin assembly. *Nature* **483**, 434–438 (2012).
30. McGuffee, S. R., Smith, D. J. & Whitehouse, I. Quantitative, genome-wide analysis of eukaryotic replication initiation and termination. *Mol. Cell* **50**, 123–135 (2013).
31. Hennion, M. et al. FORK-seq: replication landscape of the *Saccharomyces cerevisiae* genome by nanopore sequencing. *Genome Biol.* **21**, 125 (2020).
32. Liu, Y., Wu, X., D'aubenton-Carafa, Y., Thermes, C. & Chen, C.-L. OKseqHMM: a genome-wide replication fork directionality analysis toolkit. Preprint at *bioRxiv* https://doi.org/10.1101/2022.01.12.476022 (2022).
33. Blin, M. et al. DNA molecular combing-based replication fork directionality profiling. *Nucleic Acids Res.* **49**, e69 (2021).
34. Wang, W. et al. Genome-wide mapping of human DNA replication by optical replication mapping supports a stochastic model of eukaryotic replication. *Mol. Cell* **81**, 2975–2988.e2976 (2021).
35. Wu, X. et al. Developmental and cancer-associated plasticity of DNA replication preferentially targets GC-poor, lowly expressed and late-replicating regions. *Nucleic Acids Res.* **46**, 10157–10172 (2018).
36. Petryk, N. et al. MCM2 promotes symmetric inheritance of modified histones during DNA replication. *Science* **361**, 1389–1392 (2018).
37. Chen, Y. H. et al. Transcription shapes DNA replication initiation and termination in human cells. *Nat. Struct. Mol. Biol.* **26**, 67–77 (2019).
38. Li, Z. et al. DNA polymerase alpha interacts with H3-H4 and facilitates the transfer of parental histones to lagging strands. *Sci. Adv.* **6**, eabb5820 (2020).
39. Tubbs, A. et al. Dual roles of poly(dA:dT) tracts in replication initiation and fork collapse. *Cell* **174**, 1127–1142.e1119 (2018).
40. Kirstein, N. et al. Human ORC/MCM density is low in active genes and correlates with replication time but does not delimit initiation zones. *eLife* https://doi.org/10.7554/eLife.62161 (2021).
41. Hyrien, O., Maric, C. & Méchali, M. Transition in specification of embryonic metazoan DNA replication origins. *Science* **270**, 994–997 (1995).
42. Dijkwel, P. A., Wang, S. & Hamlin, J. L. Initiation sites are distributed at frequent intervals in the Chinese hamster dihydrofolate reductase origin of replication but are used with very different efficiencies. *Mol. Cell Biol.* **22**, 3053–3065 (2002).
43. Powell, S. K. et al. Dynamic loading and redistribution of the Mcm2-7 helicase complex through the cell cycle. *EMBO J.* **34**, 531–543 (2015).
44. Gros, J. et al. Post-licensing specification of eukaryotic replication origins by facilitated Mcm2-7 sliding along DNA. *Mol. Cell* **60**, 797–807 (2015).
45. Promonet, A. et al. Topoisomerase 1 prevents replication stress at R-loop-enriched transcription termination sites. *Nat. Commun.* **11**, 3940 (2020).
46. Brison, O. et al. Transcription-mediated organization of the replication initiation program across large genes sets common fragile sites genome-wide. *Nat. Commun.* **10**, 5693 (2019).
47. Letessier, A. et al. Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature* **470**, 120–123 (2011).
48. Le Tallec, B. et al. Common fragile site profiling in epithelial and erythroid cells reveals that most recurrent cancer deletions lie in fragile sites hosting large genes. *Cell Rep.* **4**, 420–428 (2013).
49. Hamperl, S., Bocek, M. J., Saldivar, J. C., Swigut, T. & Cimprich, K. A. Transcription–replication conflict orientation modulates R-loop levels and activates distinct DNA damage responses. *Cell* **170**, 774–786.e719 (2017).
50. Manzo, S. G. et al. DNA topoisomerase I differentially modulates R-loops across the human genome. *Genome Biol.* **19**, 100 (2018).
51. Park, K. et al. Aicardi–Goutières syndrome-associated gene SAMHD1 preserves genome integrity by preventing R-loop formation at transcription-replication conflict regions. *PLoS Genet.* **17**, e1009523 (2021).
52. Bayona-Feliu, A., Barroso, S., Muñoz, S. & Aguilera, A. The SWI/SNF chromatin remodeling complex helps resolve R-loop-mediated transcription-replication conflicts. *Nat. Genet.* **53**, 1050–1063 (2021).
53. Andrianova, M. A., Bazykin, G. A., Nikolaev, S. I. & Seplyarskiy, V. B. Human mismatch repair system balances mutation rates between strands by removing more mismatches from the lagging strand. *Genome Res.* **27**, 1336–1343 (2017).
54. Jaksik, R., Wheeler, D. A. & Kimmel, M. Detection and characterization of replication origins defined by DNA polymerase epsilon. Preprint at *bioRxiv* https://doi.org/10.1101/2021.07.27.453931 (2021).
55. Shi, M. J. et al. APOBEC-mediated mutagenesis as a likely cause of FGFR3 S249C mutation over-representation in bladder cancer. *Eur. Urol.* **76**, 9–13 (2019).
56. DeWeerd, R. A. et al. Prospectively defined patterns of APOBEC3A mutagenesis are prevalent in human cancers. *Cell Rep.* **38**, 110555 (2022).
57. Flasch, D. A. et al. Genome-wide de novo L1 retrotransposition connects endonuclease activity with replication. *Cell* **177**, 837–851.e828 (2019).

58. Sultana, T. et al. The landscape of L1 retrotransposons in the human genome is shaped by pre-insertion sequence biases and post-insertion selection. *Mol. Cell* **74**, 555–570.e557 (2019).
59. Ming, X. et al. Kinetics and mechanisms of mitotic inheritance of DNA methylation and their roles in aging-associated methylome deterioration. *Cell Res.* https://doi.org/10.1038/s41422-020-0359-9 (2020).
60. Reijns, M. A. et al. Lagging-strand replication shapes the mutational landscape of the genome. *Nature* https://doi.org/10.1038/nature14183 (2015).
61. Daigaku, Y. et al. A global profile of replicative polymerase usage. *Nat. Struct. Mol. Biol.* https://doi.org/10.1038/nsmb.2962 (2015).
62. Clausen, A. R. et al. Tracking replication enzymology in vivo by genome-wide mapping of ribonucleotide incorporation. *Nat. Struct. Mol. Biol.* **22**, 185–191 (2015).
63. Koh, K. D., Balachander, S., Hesselberth, J. R. & Storici, F. Ribose-seq: global mapping of ribonucleotides embedded in genomic DNA. *Nat. Methods* **12**, 251–257 (2015).
64. Zhou, Z. X., Lujan, S. A., Burkholder, A. B., Garbacz, M. A. & Kunkel, T. A. Roles for DNA polymerase δ in initiating and terminating leading strand DNA replication. *Nat. Commun.* **10**, 3992 (2019).
65. Koyanagi, E. et al. Global landscape of replicative DNA polymerase usage in the human genome. Preprint at *bioRxiv* https://doi.org/10.1101/2021.11.14.468503 (2021).
66. Pratto, F. et al. Meiotic recombination mirrors patterns of germline replication in mice and humans. *Cell* **184**, 4251–4267.e4220 (2021).
67. Sriramachandran, A. M. et al. Genome-wide nucleotide-resolution mapping of DNA replication patterns, single-strand breaks, and lesions by GLOE-seq. *Mol. Cell* **78**, 975–985 e977 (2020).
68. Kara, N., Krueger, F., Rugg-Gunn, P. & Houseley, J., https://doi.org/10.1101/2020.08.10.243931 (2020).
69. Kit Leng Lui, S. et al. Monitoring genome-wide replication fork directionality by Okazaki fragment sequencing in mammalian cells. *Nat. Protoc.* **16**, 1193–1218 (2021).
70. Audit, B. et al. Multiscale analysis of genome-wide replication timing profiles using a wavelet-based signal-processing algorithm. *Nat. Protoc.* **8**, 98–110 (2013).
71. Muller, C. A. et al. Capturing the dynamics of genome replication on individual ultra-long nanopore sequence reads. *Nat. Methods* **16**, 429–436 (2019).
72. Gansauge, M. T. et al. Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res.* **45**, e79 (2017).
73. Salic, A. & Mitchison, T. J. A chemical method for fast and sensitive detection of DNA synthesis in vivo. *Proc. Natl Acad. Sci. USA* **105**, 2415–2420 (2008).
74. Burgers, P. M. J. & Kunkel, T. A. Eukaryotic DNA replication fork. *Annu. Rev. Biochem.* **86**, 417–438 (2017).
75. DePamphilis, M. L. *Genome Duplication*. (Garland Science/Taylor & Francis Group, New York, 2010).
76. Qu, D. et al. 5-Ethynyl-2′-deoxycytidine as a new agent for DNA labeling: detection of proliferating cells. *Anal. Biochem.* **417**, 112–121 (2011).
77. Ligasova, A. et al. Dr Jekyll and Mr Hyde: a strange case of 5-ethynyl-2′-deoxyuridine and 5-ethynyl-2′-deoxycytidine. *Open Biol.* **6**, 150172 (2016).
78. Manska, S., Octaviano, R. & Rossetto, C. C. 5-Ethynyl-2′-deoxycytidine and 5-ethynyl-2′-deoxyuridine are differentially incorporated in cells infected with HSV-1, HCMV, and KSHV viruses. *J. Biol. Chem.* **295**, 5871–5890 (2020).
79. Green, M. R. & Sambrook, J. *Molecular Cloning: A Laboratory Manual*. 4. edn (Cold Spring Harbor Laboratory Press, 2012).
80. Giacca, M., Pelizon, C. & Falaschi, A. Mapping replication origins by quantifying relative abundance of nascent DNA strands using competitive polymerase chain reaction. *Methods* **13**, 301–312 (1997).
81. Tornoe, C. W., Christensen, C. & Meldal, M. Peptidotriazoles on solid phase: [1,2,3]-triazoles by regios-pecific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem.* **67**, 3057–3064 (2002).
82. Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. A stepwise Huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem. Int. Ed.* **41**, 2596–2599 (2002).
83. Presolski, S. I., Hong, V. P. & Finn, M. G. Copper-catalyzed azide-alkyne click chemistry for bioconjugation. *Curr. Protoc. Chem. Biol.* **3**, 153–162 (2011).
84. Kwok, C. K., Ding, Y., Sherlock, M. E., Assmann, S. M. & Bevilacqua, P. C. A hybridization-based approach for quantitative and low-bias single-stranded DNA ligation. *Anal. Biochem.* **435**, 181–186 (2013).
85. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
86. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
87. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
88. Team, R. C. (2021).
89. Himmelmann, L. (2016).
90. Morgan, M., Pages, H., Obenchain, V. & Hayden, N. (2017).
91. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
92. Team, R. S. (PBC Boston, MA, 2020).

93. Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc
94. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* https://doi.org/10.14806/ej.17.1.200 (2011).
95. TrimGalore https://doi.org/10.5281/zenodo.5127899
96. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
97. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
98. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
99. Picard Toolkit. 2019. Broad Institute, GitHub Repository. https://broadinstitute.github.io/picard/; Broad Institute
100. Ma, E., Hyrien, O. & Goldar, A. Do replication forks control late origin firing in *Saccharomyces cerevisiae*? *Nucleic Acids Res.* **40**, 2010–2019 (2011).

Acknowledgements

X.W. is supported by The Young Scientists Fund of the National Natural Science Foundation of China (grant no. 31900415). Y.L. Thanks Agence Nationale pour la Recherche (ANR) for providing her PhD fellowship. C.T., Y.D.-C., C.-L.C., O.H. and N.P. thank the ANR grant BLAN2010–161501 (REFOPOL). Work in the O.H. lab is supported by the ANR grants 18-CE45-0002 (NanoPoRep) and 19-CE12-0028 (HUDROR). Work in the C.-L.C. lab is supported by the YPI program of I. Curie, the ATIP-Avenir program from Centre national de la recherche scientifique (CNRS) and Plan Cancer (grant number ATIP/AVENIR: N°18CT014-00); ANR grant 19-CE12-0016-02 (ReDeFINe) and 19-CE12-0020-02 (TELOCHROM); and Institut National du Cancer (INCa) grant PLBIO19-076. N.P. is the recipient of the CNRS-INSERM ATIP-Avenir grant and YPI funding from Institute Gustave Roussy; and was supported by LabEx 'Who Am I?' ANR-11-LABX-0071; the Université de Paris IdEx ANR-18-IDEX-0001 and ANR grant 19-CE12-0030-01 (INTEGER).

Author contributions

O.H., C.-L.C. and N.P. conceived and supervised the project. N.P. developed the OK-seq method in mammalian cells; X.W. adapted the method for yeast cells. Y.L. Y.D.-C., C.T. and C.-L.C. developed the bioinformatics approach and built the analysis pipeline. X.W., Y.L., O.H., C.-L.C. and N.P. wrote the manuscript with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41596-022-00793-5.

Correspondence and requests for materials should be addressed to Olivier Hyrien, Chun-Long Chen, Nataliya Petryk.

Peer review information *Nature Protocols* thanks Kuhulika Bhalla, Bruce Stillman and Zhiguo Zhang for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 23 March 2021; Accepted: 9 November 2022;

Related links

Key references using this protocol

Petryk, N. et al. *Nat. Commun.* **7**, 10208 (2016): https://doi.org/10.1038/ncomms10208

Wu, X. et al. *Nucleic Acids Res.* **46**, 10157–10172 (2018): https://doi.org/10.1093/nar/gky797

Hennion, M. et al. *Genome Biol.* **21**, 125 (2020): https://doi.org/10.1186/s13059-020-02013-3