MANUSCRIPT READY FOR PRODUCTION

DNA copy number measurement of genome replication dynamics by high-throughput sequencing – the sort-seq, sync-seq and MFA-seq family

Contributing authors: Dzmitry G Batrakou[1], Carolin A Müller[1], Rosemary H C Wilson[1] and Conrad A Nieduszynski

Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford, OX1 3RE, UK. [1]These authors contributed equally. Correspondence should be addressed to C.A.N. (conrad.nieduszynski@path.ox.ac.uk)

EDITORIAL SUMMARY This protocol describes a suite of related approaches for genome-wide profiling of DNA replication dynamics in unperturbed or synchronised cells by measuring DNA copy number changes using high-throughput sequencing.

Abstract: Genome replication follows a defined temporal programme that can change during cellular differentiation and disease onset. DNA replication results in an increase in DNA copy number that can be measured by high-throughput sequencing (HTS). Here we present a protocol to determine genome replication dynamics using DNA copy number measurements. Cell populations can be obtained in three variants of the method. First, sort-seq reveals the average replication dynamics across S phase in an unperturbed cell population; fluorescence activated cell sorting (FACS) is used to isolate replicating and non-replicating subpopulations from asynchronous cells. Second, sync-seq measures absolute replication time at specific points during S phase using a synchronised cell population. Third, marker frequency analysis (MFA) can be used to reveal the average replication dynamics using copy number analysis in any proliferating asynchronous cell culture. These approaches have been used to reveal genome replication dynamics in prokaryotes, archaea, and a wide range of eukaryotes, including yeasts and mammalian cells. We have found this approach straightforward to apply to other organisms and highlight example studies from across the three domains of life. Here, we present a *Saccharomyces cerevisiae* version of the protocol that can be performed in 7-10 days. It requires basic molecular and cellular biology skills, as well as a basic understanding of Unix and R.

INTRODUCTION

Complete and faithful DNA replication is essential for cell division, and, therefore, the development and maintenance of any organism. Bi-directional DNA replication forks initiate from origins of replication following their activation (or firing) during the synthesis or S phase of the cell cycle[1]. In eukaryotes and archaea, efficient replication of the whole genome is ensured by the presence of multiple origins throughout the genome, not all of which will fire in each S phase[2]. In organisms with multiple origins, the activity of individual origins or clusters of origins gives rise to a characteristic temporal pattern for genome replication[3].

Fundamentally, there are three different types of methods used to study replication timing: DNA copy number-based methods[4,5], which we describe here; incorporation of dense isotopes or nucleotide analogues that allow isolation of replicated DNA by density centrifugation or immunoprecipitation[6,7]; and immunoprecipitation of replication fork proteins[8]. We compare the advantages and limitations of these methods below.

DNA copy number as a proxy for DNA replication

During S phase the relative DNA copy number increases for each region as it is replicated; therefore, measuring DNA copy number can be used as a proxy for replication time[4,9]. For eukaryotes, DNA copy number increases from 1N to 2N in haploids and 2N to 4N in diploids (with a few notable exceptions[10], for example a greater upper limit for organisms undergoing multiple, concurrent rounds of replication) as the cell passes through S phase. Regions of DNA that are copied very early during S phase will be at their maximum copy number for most of S-phase, giving an average S phase copy number near to this maximum (e.g. 2N in haploids and 4N in diploids). Conversely, regions of DNA that are replicated very late have an average S phase copy number close to 1N in haploids and 2N in diploids. Regions of DNA that are replicated in the middle of S phase have an average S phase copy number that lies on a continuous scale between 1N and 2N in haploids or 2N and 4 N in diploids. Measuring the DNA copy number genome-wide at time points during S phase or in a population of cells that represent all stages of S phase can, therefore, serve as a proxy for the time at which a particular region was replicated. We note that, as these are population-based methods, these data are determined by a combination of both the *efficiency* of a replication origin being active (firing) within the population, or how likely an origin is to fire, and the *time* at which the origin fired within each cell. These two parameters can be interrogated mathematically[9].

The protocol we describe here measures DNA copy number as a proxy for replication timing by high-throughput sequencing. Different approaches for obtaining samples give rise to a family of closely related methods: sort-seq, sync-seq and marker frequency analysis (MFA-seq). Our protocol uses *Saccharomyces cerevisiae* as an example and we also show data from *Haloferax volcanii*. In addition, this family of methods has been used to determine replication timing for a wide variety of organisms (see box 1). Each method compares DNA copy number between replicating and non-replicating samples with the key difference being the manner in which the samples are obtained. In sort-seq, fluorescence-activated cell sorting (FACS) is used to obtain replicating cells from S phase and a non-replicating sample[11,12]. In sync-seq, cells are cell cycle synchronised and replicating samples are taken during S phase, with non-replicating samples taken before (or after) S phase[9]. MFA-seq measures DNA copy number in any rapidly proliferating cell sample, normalising to a stationary phase (non-replicating) control[13,14]. We detail consideration for which variant to use in 'Choice of Variant' below.

Applications of the method

An increasing number of studies, from a range of laboratories, have used this family of methods to study replication dynamics, in bacteria[13,14], archaea[14,15] and eukaryotes[16-18]. For the most straightforward application, these techniques allow replication timing to be determined for wild-type and genome-engineered cells, including cells with synthetic chromosomes[19,20]. Replication dynamics can also be measured following specific cell treatments, for example, hydroxyurea to interrogate the effect of replication stress[4]. When combined with experimentally induced genetic mutations, these techniques can be used to assess the impact of genetic inactivation of specific origins *in cis*, for example, to determine the effect of delaying the replication of specific genes or genomic loci[14,21-23]. These techniques can also be used to monitor the effect of naturally occurring (or disease related) polymorphisms on replication dynamics[17]. Alternatively, the *trans* effect of genetic alterations can be investigated, for example, the

genome-wide impact of altered replication protein expression either directly, or relative to the number of active origins[22,24-27]. At a broader level, a number of studies have used these methods to examine the evolution of replication origin location and activity, giving further insights into the control of replication timing[12,16,21]. The use of a standardised protocol that can be applied to a range of organisms is particularly important for these comparative genomic studies.

These techniques are also suitable for investigating non-typical DNA replication. For example, with slight adaptations, over-replication and under-replication can be investigated[28,29] as well as the identification of regions that replicate post-S phase, e.g. during G2 phase[24]. Additionally, these techniques have the potential to measure the DNA replication that occurs when cells repair DNA double strand breaks by Break-Induced-Replication[30]. Finally, data generated by this family of methods can be interrogated to reveal altered karyotypes, including large structural rearrangements[13,14] and aneuploidy (see step 61| A (iii) and Box 2|). Importantly, the protocol provides genome-wide datasets, which allow comparison with other datasets including protein binding sites (e.g. replication factors, transcription factors and chromatin marks), genetic elements (e.g. tRNA genes and autonomously replicating sequences) or structural features (e.g. G4 quadruplexes and topologically associating domains).

Alternative methods of measuring genome replication dynamics

A major advantage of the copy number-based approaches described here is that they do not require genetic manipulations or special treatments, allowing analysis of unperturbed wildtype cells. However, the dynamic range is limited (e.g. from one to two for most eukaryotes), thus requiring an accurate quantification of copy number. This requires high coverage sequencing and can prove challenging for large genomes if high spatial resolution is required[18]. This requirement can be mitigated by alternative approaches that utilize labelling of nascent DNA with nucleotide analogues, followed by enrichment of newly synthesised DNA. A variant of nascent DNA labelling and separation is the dense isotope transfer method similar to the Meselson–Stahl experiment; however, it is expensive, labour-intensive and has separation biases[31]. More recent variants include BrdU-IP-seq[7], EdUseq-HU[32,33] and repli-seq[34]. In these methods, nascent DNA containing an incorporated nucleotide analogue, such as BrdU or EdU, is enriched through immunoprecipitation or Click-chemistry and identified by microarrays or high-throughput sequencing. The enrichment for nascent DNA increases the dynamic range and thus makes these approaches particularly useful for organisms with large genomes. However, this qualitative separation of nascent from parental DNA comes at the expense of quantitative information, in particular biases from the enrichment and, consequently, the measurements are expressed in arbitrary units.

An alternative method for studying replication timing is through chromatin-immunoprecipitation sequencing (ChIP-seq) of replication fork proteins, for example, GINS components[8]. However, this approach is more technically challenging, requiring cell synchronisation and either tagging of proteins by genetic alteration or the availability of a suitable antibody to the endogenous protein.

All of these approaches to measure genome replication dynamics can be complemented with methods that identify replication origins[35,36], determine replication fork direction[37-39] or investigate the response to replication stress[40].

Limitations

The family of methods described here are powerful techniques to study replication timing in a variety of organisms. Nevertheless, there are, as always, some limitations of the techniques. Firstly, mapping of the high-throughput sequence data requires a reference genome, which is not available for some organisms. Second, the sort-seq and MFA-seq variants produce replication profiles by measuring the relative copy number of all genomic loci – i.e. there is no absolute measure of replication time. This is in contrast to sync-seq where DNA copy number is measured at specific time points to allow each locus to be assigned a median replication time relative to a cell cycle landmark, such as the start of DNA synthesis. Third, as mentioned above, the small dynamic range between one and two in eukaryotes requires an accurate quantification of DNA copy number, which can limit spatial resolution in organisms with large genomes. In this case, if only a few regions of the genome are of interest, loci specific copy number qPCR[41] or ddPCR[18] methods can be used. Such approaches can also be used to test for allele specific replication times[17,18]. Fourth, a non-replicating population provides a control to minimise sequence biases. However, replication profiles can be generated without a non-replicating control sample (step 61 A (v)), but they may be noisier. Finally, it is also important to reiterate that this protocol generates data from a population

of cells and as such reports the population average replication time. If relevant for the user, copy number approaches have recently been adapted to single cells[42] and labelled nascent DNA approaches - to single molecules[43].

Overview of the procedure

The first section of the protocol (steps 1| and 2|, options A, B, or C) describes the experimental setup and cell growth for each of the three variants of the family; sort-seq, sync-seq and MFA-seq (see Figure 1). The variants converge for the following experimental sections of the protocol - DNA extraction (steps 3| - 17|), library preparation (steps 18| - 52|) and high-throughput sequencing (steps 53| - 56|). The last section of the protocol details data analysis – firstly aligning and processing of sequencing data (steps 57| - 60), then normalisation and visualisation using Repliscope, our R package available from CRAN, either using the command line (step 61| option (A)) or a built-in graphical user interface (step 61| option (B)). We cover validation and quality control steps in both the experimental and data analysis parts of the protocol. At the end of the protocol users will have generated replication timing profiles for their strain(s) or conditions of interest, and compared them as appropriate. Finally, we suggest various bespoke analyses that users can consider undertaking with their data if desired.

Experimental Design

Choice of variant. The choice of variant depends on both the organism and experimental question. Sync-seq provides the highest temporal and spatial resolution of the three variants; it is the only variant that provides an absolute measure of replication time. However, this requires S phase cell cycle synchronisation from which samples can be taken as genome replication proceeds. Sort-seq alleviates the requirement for cell cycle synchronisation, but produces relative rather than absolute replication timing data. The main requirement for sort-seq is that replicating cells are clearly identifiable in a DNA-content flow cytometry profile, since this allows a representative replicating cell population to be obtained by FACS. Cell sorting may require optimisation of DNA staining conditions and cell sorter setup for the organism of interest; this may be problematic for organisms with a variable ploidy or that undertake cell septation during S phase. MFA-seq is technically the most straightforward method to apply to any organism, as it solely requires proliferating and stationary phase samples. However, the resulting dynamic range is dependent upon the fraction of replicating cells within the population. Therefore, MFA is particularly well suited to organisms for which the other variants are not technically applicable; for example, those that undergo concurrent rounds of replication, that cannot be synchronised or have unstable ploidy[14]. For organisms where multiple variants could be used, MFA-seq typically generates the smallest dynamic range and thus will be lower resolution. Box 1| provides further information for adapting this protocol to other organisms, as well as a list of published uses of this family of methods across the three domains of life.

Controls. Replicating samples should be normalised (in almost all cases) to a non-replicating sample to control for sequencing bias across the genome. The key requirement for the non-replicating control is that the whole genome is at a uniform copy number. Typically, the non-replicating control is DNA from G1 or G2 phase cells for sort-seq, DNA from cells arrested in G1 phase for sync-seq, and DNA from stationary cultures for MFA-seq. Sometimes the use of alternative non-replicating controls may be required – we detail these and other considerations pertaining to the users' choice of normalisation control in Box 3|. The non-replicating control sample should be prepared in parallel and sequenced on the same run as replicating samples.

Multiplex sequencing and sequence depth requirements. By using multiplexing barcodes in step 42| of the protocol, users can combine multiple samples on the same sequencing run. Reads are then demultiplexed at the end of the run using the differing barcode sequences. However, before multiplexing samples, users must consider the optimal number of reads per sample, which depends heavily on the genome size of the organism of interest and the chosen bin size. A simple formula to calculate number of

multiplexed samples per run is N=(t*b)/(g*d), where *t* is the estimated total yield of the run (reads), *b* is bin size (bp), *g* is haploid genome size (bp) and *d* is desired sequencing depth per bin (reads). We find that a depth of 1000-2000 reads/bin gives cost efficient sequencing depth without compromising the precision of the data. By default, we perform single-end sequencing, although the computational pipeline described here can be applied to paired-end sequencing data. A NextSeq run typically yields 400+ million single-end reads. Therefore, for samples from *S. cerevisiae* (~13 Mbp genome) we combine up to 24 samples per NextSeq run for sequencing depth of > 1250 reads per 1 kb bin. In contrast, for a human cell line we recently combined two samples on a NextSeq run to obtain an average depth of 2095 reads / 50 kb bin[18]. For more discussion of optimal read number for copy number-based methods see Müller *et al.*, 2014[4].

Required expertise. The protocol described here requires standard microbiology and molecular biology techniques. In addition, for in house high-throughput sequencing the experimentalist should be able to operate the sequencer (e.g. an Illumina NextSeq machine); an alternative is to use a high-throughput sequencing facility or company. The sort-seq protocol requires use of a FACS machine, either directly or through a facility. The data analyses require a basic familiarity with the Linux command line and the R programming language. For those with limited R experience, a web application provides a graphical user interface to R that allows the user to perform the standard data analyses described in this protocol. For more experienced users, the R Repliscope package permits an equivalent analysis using a command line interface and provides a starting point for more sophisticated custom analyses.

MATERIALS

REAGENTS

Biological Materials

Yeast strain of interest, taken from a fresh YPAD plate or appropriate Synthetic Complete (SC) medium (specific to *S. cerevisiae*). For the example data we used *S. cerevisiae* yeast strains of a W303 genetic background, for example strain BMA64-1A from Euroscarf (Fig. 1-6). Figure 5 also shows example MFA data for *Haloferax volcanii*[14].

▲ **CRITICAL** Yeast strains of mating type *a* are required for sync-seq.

▲ **CRITICAL** For sync-seq we used a *BAR1* strain. *BAR1* encodes a secreted protease that degrades alpha factor, therefore with *bar1* strains less alpha factor is required to arrest cells.

Chemicals

- 2-mercaptoethanol (Sigma-Aldrich, cat. no. M6250) **! CAUTION** Toxic and irritant, avoid inhalation and wear personal protective equipment (PPE)
- Agencourt AMPure XP magnetic beads (Beckman Coulter, cat. no. A63881)
- Alpha factor (specific to *S. cerevisiae*; GenScript, cat. no. RP01002)
- Chloroform (Sigma, cat. no. 24216) **! CAUTION** Harmful and irritant, avoid inhalation and wear PPE
- D-Sorbitol (Sigma-Aldrich, cat. no. S1876)
- Ethanol, absolute (Sigma-Aldrich, cat. no. 32221) **! CAUTION** Flammable
- EDTA (0.5 M sterile solution pH 8.0; VWR, cat. no. E177-500)
- Hydrochloric acid (HCl) (Sigma-Aldrich, cat. no. 30721) **! CAUTION** Corrosive, avoid inhalation and wear PPE
- NEBNext® Library Quant Kit for Illumina® (NEB, cat. no. E7630S)
- NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1) (NEB, cat. no. E7335S)
- NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 2) (NEB, cat. no. E7500S)
- NEBNext® Ultra II DNA Library Prep Kit for Illumina® (NEB, cat. no. E7645S)
- NextSeq 500/550 High Output Kit v2.5 (75 Cycles) (Illumina, cat. no. 20024906)
- Phenol:chloroform:isoamyl (Sigma, cat. no. 77617) **! CAUTION** Phenol is toxic and corrosive. Chloroform is harmful and an irritant. Wear PPE and avoid inhalation.
- Pronase (specific to *S. cerevisiae*; Calbiochem/MERCK, cat. no. 53702-250KU)
- Proteinase K (MP Biomedicals, cat. no.193504) ▲ **CRITICAL** We have found that proteinase K from other suppliers can produce artefacts in flow cytometry cell cycle analysis.
- RNase A (MP Biomedicals, cat. no. 219398050)
- Sodium dodecyl sulfate, 20% (wt/vol) solution (Fisher Scientific UK; cat. no. 10607633)
- Sodium azide (Sigma-Aldrich, cat. no. S8032) **! CAUTION** Toxic, handle with care and wear PPE
- Sodium chloride (Sigma-Aldrich, cat. no. S9888)
- SYTOX™ Green Nucleic Acid Stain, (5 mM solution; Life Technologies, cat. no. S7020) ▲ **CRITICAL** SYTOX Green generates superior cell cycle profiles compared to propidium iodide for *S. cerevisiae*[44].
- Tris base (Trizma® base; Sigma-Aldrich, cat. no. T6066)
- Tri-sodium citrate (Fisher, cat. no. S/3320/53) ▲ **CRITICAL** We have found that tri-sodium citrate from other suppliers can produce artefacts in flow cytometry cell cycle analysis.
- Qubit dsDNA HS Assay Kit (Life Technologies, cat. no. Q32854)
- Tapestation High Sensitivity D1000 Reagents (Agilent, cat. no. 5067-5585)
- Tapestation High Sensitivity D1000 Screentape (Agilent, cat.no. 5067-5584)
- YPAD broth or appropriate Synthetic Complete (SC) broth (specific to *S. cerevisiae*; ForMedium, cat. no. CCM1005)
- YPAD agar or appropriate Synthetic Complete (SC) agar (specific to *S. cerevisiae*; ForMedium, cat. no. CCM0510)
- Zymolyase T20 (specific to *S. cerevisiae*; Arthrobacter luteus; Seikagaku, cat. no. 120491)

EQUIPMENT

- 1 l glass conical (e.g. VWR, cat. no. 214-1134)
- 50 ml centrifuge tubes (e.g. Greiner Bio One, cat. no. 227261)
- 100 ml glass conical (e.g. Fisher, cat. no. FB33131)
- 4200 TapeStation System with PC (Agilent, cat. no. G2991AA) or alternate DNA fragment length measurement (e.g PAGE gels)
- Autoclave (e.g. Rodwell, Phoenix)
- Automatic pipette controller (e.g. Star Lab, ErgoOne Fast)
- Balance (e.g. Sartorius cat. no. BP1200)
- Basespace (Illumina):
  Set up a Basespace account https://basespace.illumina.com/home/index
- Bioruptor® Standard (Diagenode, cat. no. UCD-200) or alternate sonicator suitable for DNA shearing
- Benchtop centrifuge for 15 ml and 50 ml centrifuge tubes (e.g. Thermo Scientific, Heraeus Multifuge X3R)
- Cuvettes, (e.g. Fisher Scientific, Semi-micro cuvettes 2.5ml cat. no. 10781791)
- Filter (e.g. Sartorius, cat. no. 16555-K)
- Freezer, -20°C
- Fridge, 4°C
- Fume hood
- Haemocytometer (e.g. Hawksley, cat. no. Z4+AC1000)
- Heat block (e.g. Techne, Dri-block DB-2D)
- Ice bucket
- Ice machine (e.g. Scotsman, cat. no. AF206)
- Incubator shaker (e.g. New Brunswick Scientific, Innova 40)
- Light microscope (e.g. Olympus, cat. no. CKX41)
- Magnetic microcentrifuge tube rack (e.g. Invitrogen, DynaMag-2, cat. no. 12321D)
- Microcentrifuge capable of chilling to 4°C (e.g. Eppendorf, Centrifuge 5424R)
- Microcentrifuge tubes (e.g. Star Lab, cat. no. E1415-1500)
- Micropipettes (e.g. Star Lab, ErgoOne series)
- Micropipette tips (e.g. Star Lab, TipOne series)
- Microscope slides (e.g. Thermo Scientific, Superfrost Plus cat. no. J1800AMNZ)
- Milli-Q water purifier (e.g. Veolia, Purelab® Flex)
- NanoDrop™ spectrophotometer (NanoDrop, cat. no. ND-1000) or another DNA concentration quantification
- NextSeq 500 (Illumina)
- PCR tubes or strips (e.g. Greiner Bio One, cat. no. 683201)
- Petri dishes (e.g. Sarstedt, cat. no. 82.1473)
- pH meter (e.g. Jenway, cat. no. 3510)
- qPCR tubes (e.g. StarLab, Rotor-Gene Style 4-Strip Tubes and Caps, 0.1 ml cat. no. I1402-0400)
- Quantitative PCR machine (e.g. Qiagen, Rotor-Gene Q 2plex Platform cat. no. 9001550)
- Qubit Assay Tubes (Invitrogen, cat. no. Q32856)
- Qubit Fluorometer (e.g. Invitrogen, cat. no. Q32857) or another high sensitivity DNA concentration quantification
- Rotating wheel (e.g. Stuart, Rotator SB3)
- Screw capped microcentrifuge tubes (e.g. Star Lab, cat. no. E1415-2231)
- Serological pipettes (e.g. Greiner Bio One)
- Shaking heat block for microcentrifuge tubes (e.g. Eppendorf, Thermomixer Comfort)
- Spectrophotometer (e.g. ThermoScientific, Biomate 3 cat. no. 335905P-02)
- Syringe (e.g. Terumo, cat. no. SS+20ES1)
- Thermal Cycler (e.g. Biorad, Tetrad 2)
- Tip Sonicator (e.g. Branson, Digital Sonifier)
- Timer (e.g. Fisher Scientific, Traceable Nano Timer)
- Vortex (e.g. Fisher Scientific, TopMix cat. no. FB15013)

- Water bath shaker (e.g. New Brunswick Scientific, Innova 3100)

Flow Cytometry

- BD LSRFortessa X-20 flow cytometer (Becton Dickinson) equipped with 488 nm laser (Coherent), 505 nm long pass and 530/30 nm band pass filters
- PC running FACSDiva software
- FlowJo 10.5.3 software (FlowJo) or alternative software to analyse flow cytometry data
- MoFlo XPD cell sorter (Beckman-Coulter) equipped with 488 nm laser (Saphire, Coherent), 505 nm short pass and 530/40 nm band pass filters or another comparable cell sorter
- PC running Summit software

Software and hardware

- Workstation with 64-bit processor and at least 4 Gb of RAM and latest Ubuntu OS with the following installed software: FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), Bowtie 2[45] (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml), SAMtools[46] (http://www.htslib.org/), bedtools[47] (https://bedtools.readthedocs.io/en/latest/), Picard (https://broadinstitute.github.io/picard/), R (https://www.r-project.org/).
  R package Repliscope (https://cran.r-project.org/web/packages/Repliscope/)
- Additionally, a web browser is required for the interactive analysis using the app. We recommend Google Chrome (https://www.google.com/chrome) or the RStudio built-in browser (https://www.rstudio.com/).
  The analysis presented here has been performed with the following software versions: Ubuntu 18.04.01 LTS, fastqc 0.11.5, bowtie2 2.3.4.1, samtools 1.7, bedtools 2.26.0, picard-tools 2.8.1, R 3.4.4, Repliscope 1.0.0, ggplot2 3.0.0, shiny 1.1.0, colourpicker 1.0, Google Chrome 70.0.3538.77 (available as a bootable Ubuntu disc image at https://ln1.path.ox.ac.uk/groups/nieduszynski/Replibuntu/Replibuntu-18.04.0-amd64.iso.gz).

REAGENT SETUP

CRITICAL All solutions are prepared with Milli-Q water and stored at room temperature (RT, 20°C) for a maximum of 12 months unless otherwise indicated.

- **3 mM Alpha factor** Dissolve aseptically 10 mg alpha-factor in 2 ml sterile water. Store in 0.5 ml aliquots at -80°C. Avoid excessive freeze/thawing.
- **1 M D-Sorbitol** Dissolve 18.2 g D-sorbitol in 100 ml water. Sterilize by autoclaving.
- **2 M D-Sorbitol** Dissolve 36.4 g D-sorbitol in 100 ml water. Sterilize by autoclaving.
- **0.2 M EDTA, 0.1% (wt/vol) sodium azide** Combine 40 ml of 0.5 M EDTA pH 8.0, 1 ml of 10% (wt/vol) sodium azide and 59 ml of water.
- **70% (vol/vol) ethanol** Dilute 70 ml of 100% (vol/vol) ethanol to 100 ml with deionized H_2O.
- **80% (vol/vol) ethanol** Dilute 80 ml of 100% (vol/vol) ethanol to 100 ml with deionized H_2O. ▲ **CRITICAL** For AMPure XP bead clean up steps it is important that 80% (vol/vol) ethanol is made fresh on the day.
- **20 mg/ml Pronase** Dissolve 200 mg of pronase in 10 ml of water. Filter-sterilise using a 0.2 µm filter and syringe and store in 1 ml aliquots at -20°C.
- **20 mg/ml Proteinase K** Dissolve whole 100 mg bottle in 5 ml water, store in 0.5 ml aliquots at − 20°C.
- **10 mg/ml RNase A** Dissolve whole 50 mg bottle in 5 ml water, store in 1 ml aliquots at − 20°C.
- **10% (wt/vol) Sodium azide** Dissolve 10 g of sodium azide in 100 ml water. **! CAUTION** Sodium azide is toxic. Wear personal protective equipment.
- **3 M Sodium chloride** Dissolve 17.53 g of sodium chloride in 100 ml water.
- **5 M Sodium chloride** Dissolve 29.2 g of sodium chloride in 100 ml water.
- **1 M Sodium citrate pH 7.4** Dissolve 294.1 g of sodium citrate in 500 ml water. Adjust the pH to 7.4 with HCl. Add additional water to a total volume of 1 l. Sterilize by autoclaving.
- **50 mM Sodium citrate pH 7.4** Dilute 1 M sodium citrate pH 7.4 to 50 mM (1:20) with water. Filter to remove aggregates immediately before use using 0.45 µm filter.

- **Solution 1**

Component	Amount (ml)	Final concentration
Sorbitol (2 M)	30	1.2 M
Tris-HCl pH 8.0 (1 M)	10	0.2 M
EDTA pH 8.0 (0.5 M)	2	0.02 M
Water	To 50 ml total	
Total	50	

- **Solution 2**

Component	Amount (ml)	Final concentration
Tris-HCl pH 8.0 (1 M)	2.5	0.05 M
EDTA pH 8.0 (0.5 M)	10	0.1 M
NaCl (3 M)	1.7	0.1 M
SDS (20% wt/vol)	1.25	0.5% (wt/vol)
Water	To 50 ml total	
Total	50	

- **1× SYTOX™ Green in 50 mM sodium citrate** Dilute SYTOX™Green (1:5000) in 50 mM sodium citrate, pH 7.4.
- **1x SYTOX™ Green, 0.1% (wt/vol) sodium azide in 50 mM sodium citrate** Dilute SYTOX™Green (1:5000) and 10% (wt/vol) sodium azide (1:100) in 50 mM sodium citrate, pH 7.4.
- **1 M Tris-HCl pH 8.0** Dissolve 60.6 g Tris base in 400 ml water. Adjust the pH to 8.0 with HCl. Add additional water to a final volume of 500 ml. Sterilize by autoclaving.
- **10 mM Tris-HCl pH 8.0** Dilute 1 M Tris-HCl pH 8.0 (1:100) in water.
- **10 mM Tris-HCl pH 8.0, 10 mM NaCl** Dilute 1 M Tris-HCl pH 8.0 (1:100) and 5 M NaCl (1:500) in water.
- **TE buffer** (10 mM Tris, 1 mM EDTA) Dilute 1 M Tris-HCl pH 8.0 (1:100) and 0.5 M EDTA pH 8.0 (1:500) in water.
- **YPAD agar** Dissolve 56 g of YPAD agar powder in 800 ml deionized water. Sterilize by autoclaving and cool to 60°C. Pour 20 ml aseptically into plates and let the agar set for 10-20 min. Store hermetically sealed at 4°C.
- **YPAD broth** Dissolve 40 g of YPAD broth powder in 800 ml deionized water. Sterilize by autoclaving.
- **10 mg/ml Zymolyase T20** Dissolve 100 mg in 10 ml water, store in 1 ml aliquots at − 20°C.

EQUIPMENT SETUP

Sequencing analysis software

All the software can be installed from Ubuntu's Universe repository using the following commands:

```
$ sudo add-apt-repository universe
$ sudo apt update
$ sudo apt install fastqc bowtie2 samtools bedtools picard-tools
r-base
```

Repliscope

Installing Repliscope from CRAN will also install its dependencies: ggplot2, shiny and colourpicker. Repliscope can be installed with the following commands:

```
$ sudo R
> install.packages('Repliscope')
```

PROCEDURE

Growing primary cell culture •TIMING 17 h (hands-on time: 10 min)

1| Inoculate a single colony of the yeast strain into 5 ml of liquid medium (YPAD or appropriate SC selection) in a 50 ml centrifuge tube and incubate overnight in a shaking incubator at 30°C (or 23°C for sync-seq) with 200 rpm shaking.

Obtaining replicating and non-replicating cells •TIMING 1.5-3 days (see step 2A, B or C)

2| Obtain samples of replicating cells and a non-replicating control. Follow option A for sort-seq (FACS enrichment of cells in S or G2 phase), option B for sync-seq (replicating samples from a synchronous S phase population) or option C for MFA-seq (proliferating and stationary phase cells).

(A) sort-seq •TIMING 2 days (hands-on time: 4 h)

(i) Measure cell density by spectrophotometry or a haemocytometer and inoculate 3×10^7 cells from the overnight culture into 25 ml of YPAD medium in a 100 ml conical; allow to grow to $0.7 – 1.5 \times 10^7$ cells/ml at 30°C in a shaking incubator with 200 rpm shaking (~4.5 h). Check cell density by spectrophotometry or with a haemocytometer. For a wild-type W303 strain the required cell density corresponds to an optical density (measured at 660 nm) of 0.5 – 0.9. The conversion between optical density and cell density is non-linear and strain dependent, therefore we advise use of a haemocytometer to determine this relationship.

(ii) At a density of $0.7 – 1.5 \times 10^7$ cells/ml, transfer culture into a 50 ml centrifuge tube and harvest cells by centrifugation at 2,000*g* for 5 min at 20°C.

(iii) Discard the media. Resuspend the cell pellet in 25 ml of Milli-Q water. Pellet cells by centrifugation at 2,000*g* for 5 min at RT and discard supernatant. Repeat this wash and centrifugate again to pellet cells.

(iv) Fix cells by resuspending in 10 ml of 70% (vol/vol) ethanol and incubate for at least 1 h at room temperature (20°C) or store at 4°C overnight.

■ **PAUSE POINT** Cells can be stored at 4°C for up to one year.

(v) Pellet cells by centrifugation at 2,000*g* for 5 min at RT and discard supernatant. To wash cell pellet with sodium citrate, resuspend cells in 25 ml of freshly filtered 50 mM sodium citrate and pellet by centrifugation at 2,000*g* for 5 min at RT. Repeat this wash, pellet by centrifugation as before and resuspend cells in 1 ml 50 mM sodium citrate.

(vi) Add 250 µl of 10 mg/ml RNase A and incubate for 1 h at 37°C in a water bath.

(vii) Add 100 µl of 20 mg/ml Proteinase K and incubate for 1 h at 55°C in a water bath. Gently tap tube every 20 min to resuspend cells. Pellet cells by centrifugation at 2,000*g* for 5 min at RT.

(viii) Stain DNA by adding 2.5 ml of 50 mM sodium citrate, 25 µl of 10% (wt/vol) sodium azide and 5 µl SYTOX™ Green. Incubate in the dark for at least 1 h at RT or at 4°C overnight.

■ **PAUSE POINT** SYTOX™ Green stained cells can be stored in the dark at 4°C for up to one year.

(ix) Immediately prior to FACS, visually inspect cell aliquot by light microscopy. If cells are sticking to each other, sonicate sample twice for 3 s using a tip sonicator set to 20 W of sonic power or vortex vigorously until cells are separate. ▲ **CRITICAL STEP** Avoid excessive sonication as it damages cells and causes debris in the sample.

(x) On a MoFlo XFP cell sorter (or another comparable cell sorter), use the signal height and area information from the 488 nm laser (as a measure of DNA content) to select the desired population of cells (S and G2 phase, or 'ALL' population). Exclude cell doublets or debris, as shown in Figure 2a. Create a histogram with cell count on the *y* axis and signal area of DNA content on the *x* axis. **? TROUBLESHOOTING**

(xi) Set sorting gate to include single cells from all cell cycle stages and collect 300,000 cells into a 50 ml centrifuge tube as the 'ALL' control sample, as shown in Figure 2b.

(xii) Adjust sorting gates to two populations representing S and late G2 phase cells as shown in Figure 2b. For both fractions collect at least 5 million cells into separate 50 ml centrifuge tubes.

(xiii) Add 1/3 volume of 100% (vol/vol) ethanol to the FACS-enriched cells and mix by inversion.

■ **PAUSE POINT** Samples can be stored at 4°C overnight.

(xiv) For each sorted sample (i.e. there are now three sorted samples per strain, 'ALL', 'S', and 'G2'), the sorted cells are pelleted in a microcentrifuge tube. For each sorted sample, transfer 1.5 ml into a fresh 1.5 ml microcentrifuge tube and collect cells by centrifugation at 20,000*g* for 10 min at RT. Carefully remove and discard 1.4 ml of supernatant without disturbing the cell pellet.

(xv) Transfer another 1.4 ml of the same sample into the same microcentrifuge tube, centrifuge and remove the supernatant as before. Repeat these steps until all FACS-enriched cells have been pelleted into one microcentrifuge tube per sorted sample. Carefully remove the supernatant after the final centrifugation without disturbing the cell pellet.

(xvi) Resuspend the S and G2 phase samples in 500 µl Milli-Q water and transfer 5 µl of each into a fresh microcentrifuge tube for later flow cytometer analysis. Add 200 µl of 100% (vol/vol) ethanol to the remaining S and G2 phase samples to give samples for DNA extraction.

(xvii) Add 500 µl of 1× SYTOX™ Green in 50 mM sodium citrate to the control 'ALL' sample (from step 2 A (xv)) and the 5 µl S and G2 phase sample aliquots (from step 2 A (xvi)). Incubate these samples in the dark for at least 1 h at RT or at 4°C overnight to give 1x SYTOX™ Green re-stained cell samples for flow cytometry.

■ **PAUSE POINT** Samples for DNA extraction and flow cytometry from steps 2 A (xvi and xvii) can be stored at 4°C for up to 1 month.

(xviii) Check the purity of FACS-enriched 'S' and 'G2' cell fractions and compare to the 'ALL' fraction by running the 1x SYTOX™ Green re-stained cell samples on a BD LSRFortessa X-20 flow cytometer (see Figure 2). Gate events as in step 2 A (x) and acquire a minimum of 5,000 events gated to assess the sorted samples' purity.

(xix) Pellet cells in the remaining S and G2 samples by centrifugation at 20,000*g* for 10 min at RT. Remove supernatant carefully without disturbing the cell pellet. Proceed to step 3.

(B) sync-seq •TIMING 1.5 days (hands-on time: 3 h)

(i) Measure cell density by spectrophotometry or a hemocytometer and inoculate 10^7 cells from the overnight culture into 25 ml of YPAD medium in a 100 ml conical flask; allow to grow to 2 – 6×10^7 cells /ml at 23°C in a shaking incubator with 200 rpm shaking (~8 h). Check cell density by spectrophotometry or a hemocytometer.

(ii) Inoculate 1.5×10^5 cells / ml into 200 ml of YPAD medium in a 1 l conical flask. Allow to grow to 1×10^7 cells / ml at 23°C in a shaking incubator with 200 rpm shaking (~17 h).

(iii) Prepare 13 × 2 ml aliquots of 0.2 M EDTA, 0.1% (wt/vol) sodium azide in 50 ml centrifuge tubes and place tubes tilted at -20°C overnight.

(iv) Aliquot 10 µl of 10% (wt/vol) sodium azide into 28 microcentrifuge tubes and place on ice.

(v) At 1×10^7 cells / ml, transfer 1 ml of culture into a microcentrifuge tube containing 10 µl of 10% (wt/vol) sodium azide, briefly vortex and place on ice. This is the asynchronous flow cytometry sample.

(vi) Add 33 µl of 3 mM alpha factor (final concentration 0.5 µM) to the culture from step 2 B (ii) and start a timer. (Use a final concentration of 0.2 µM for *BAR1* mutant strains). ▲ **CRITICAL STEP** Strains with altered growth rates will need optimisation of alpha factor addition times.

(vii) After 2.5 h, remove 10 µl of culture and inspect cells by light microscopy to monitor cell cycle arrest by counting the fraction of budded cells.

(viii) When > 95% of cells are unbudded, take the 0 min flow cytometry timepoint by transferring 1 ml of culture into a microcentrifuge tube containing 10 µl of 10% (wt/vol) sodium azide, vortex and place on ice. **? TROUBLESHOOTING**

(ix) Take the 0 min (non-replicating control sample) high throughput sequencing (HTS) sample - relocate a pre-prepared 2 ml EDTA/sodium azide aliquot (step 2| B (iii)) from -20°C to ice, add 100 µl of 10% (wt/vol) sodium azide and 10 ml of the yeast culture to the tube. Immediately shake the tube vigorously until the frozen pellet dislodges, and then keep on ice until the final HTS sample has been collected.

(x) Add 2 ml of 20 mg/ml Pronase to the culture and start a timer immediately.

(xi) Take flow cytometry samples (as step 2| B (viii)) every 2.5 min from 5 - 55 min and then at 60, 70, 80, 90, 120 and 180 min after Pronase addition.

(xii) Take HTS samples (as step 2| B (ix)) every 5 min from 5 min to 55 min and 90 min after Pronase addition.

(xiii) After taking the 90 min timepoint, pellet cells from HTS samples by centrifugation at 2,000*g* for 5 min at RT.

(xiv) Resuspend cells in 25 ml of Milli-Q water and pellet by centrifugation at 2,000*g* for 5 min at RT.

(xv) Remove supernatant, resuspend pellet in 1 ml Milli-Q water and relocate to a screw-capped microcentrifuge tube.

(xvi) Pellet cells by centrifugation at 13,000*g* for 30 s at RT, remove supernatant and store tubes at -20°C.

(xvii) After the final (180 min) flow cytometry timepoint, pellet cells from flow cytometry samples by centrifugation at 13,000*g* for 30 s at RT, remove supernatant and wash cells in 1 ml of Milli-Q water.

(xviii) Pellet cells as before (13,000*g* for 30 s at RT), remove supernatant and resuspend cells in 1 ml of 70% (vol/vol) ethanol.

■ **PAUSE POINT** HTS samples can be stored at -20°C for up to a year. Flow cytometry samples can be stored at 4°C indefinitely.

(xix) Pellet cells from flow cytometry samples from steps 2| B (v), (viii) and (xi) by centrifugation at 13,000*g* for 1 min at RT and discard the supernatant. Resuspend cells in 1 ml of freshly filter-sterilized 50 mM sodium citrate and pellet by centrifugation at 13,000*g* for 1 min at RT. Repeat this wash, pellet by centrifugation as before and resuspend cells in 1 ml 50 mM sodium citrate.

(xx) Add 25 µl of 10 mg/ml RNaseA. Incubate for 1 h at 37°C in a water bath.

(xxi) Add 20 µl of 20 mg/ml Proteinase K and incubate for 1 h at 55°C in a water bath.

(xxii) Pellet cells by centrifugation at 13,000*g* for 1 min at RT. Resuspend cells in 500 µl of 1x SYTOX™ Green, 0.1% (wt/vol) sodium azide in 50 mM sodium citrate. Incubate for at least 1 h at RT or store at 4°C overnight, in the dark.

■ **PAUSE POINT** SYTOX™ Green stained cells can be stored at 4°C for up to one year.

(xxiii) Run flow cytometry samples on a BD LSRFortessa X-20 flow cytometer (or another comparable flow cytometer). Gate out doublets based on the forward scatter information as in Figure 2a. Acquire signal excitation intensity from the 488 nm laser to measure DNA content following SYTOX™ Green staining. ▲ **CRITICAL STEP** High quality data is key to calculate bulk replication in the following step. Acquire 100,000 gated events to assess cell cycle progression of each sample.

(xxiv) For each time point, extract the median DNA content intensity and normalise the values to have a minimum of 0% (from the 0 min timepoint) and maximum of 100% (the late time points) DNA content. Plot the normalised values against the time after release that the samples were taken; this shows the increase in bulk genome content and should fit a sigmoid curve (Figure 3). (Our online tool may be used for the steps described above – https://dnareplab.shinyapps.io/fitsigmoid/.) Select time points for HTS based on their genome content, for example 10% - 25% bulk genome content for early S phase, 40% - 60% mid S phase and 75% - 90% late S phase (Figure 3). The bulk replication percentage of the timepoint samples taken forward for sequencing will be required in steps 61| (A) (viii) or 61| (B) (ii).

(xxv) Proceed to step 3 with the selected HTS samples (step 2| B (xvi)), as chosen based on bulk genome content and the non-replicating control.

(C) MFA-seq •TIMING 3 days (hands-on time: 2 h)

(i) Prepare two 2 ml aliquots of 0.2 M EDTA and 0.1% (wt/vol) sodium azide in 50 ml centrifuge tubes and place tubes tilted at -20°C overnight.

(ii) Measure the cell density of a 5 ml overnight culture using a spectrophotometry or a hemocytometer. Inoculate 6×10^6 cells from the overnight culture into 25 ml of YPAD medium in a 100 ml conical and allow to grow to 1.0×10^7 cells / ml at 30°C in a shaking incubator with 200 rpm shaking (~5 h).

(iii) At 1.0×10^7 cells / ml remove 1 ml of culture for flow cytometry (process this sample as described in steps 2 B (xvii) – (xxiii)). Relocate a 2 ml EDTA/sodium azide aliquot from -20°C to ice, add 100 µl of 10% (wt/vol) sodium azide and transfer 10 ml of culture into the tube. Immediately shake the tube vigorously until the frozen pellet dislodges, and then place the tube onto ice. This is the replicating sample.

(iv) Pellet cells from the replicating sample by centrifugation at 2,000*g* for 5 min at RT.

(v) Remove the supernatant, resuspend cells in 25 ml of Milli-Q water and pellet by centrifugation at 2,000*g* for 5 min at RT.

(vi) Remove the supernatant, resuspend the pellet in 1 ml Milli-Q water and relocate to a screw-capped microcentrifuge tube.

(vii) Pellet cells by centrifugation at 13,000*g* for 30 s at RT, remove supernatant and store tube at -20°C until the non-replicating sample has been collected and both are ready to proceed to step 3.

(viii) Allow the remaining overnight culture to reach stationary phase at 30°C with 200 rpm shaking (~48 h).

(ix) Measure cell density of the overnight culture by spectrophotometry or a hemocytometer. When cell number stops increasing, relocate a 2 ml EDTA/sodium azide aliquot from -20°C to ice, add 50 µl of 10% (wt/vol) sodium azide and transfer the cell culture into the tube. Immediately shake the tube vigorously until the frozen pellet dislodges, and then place the tube onto ice. This is the non-replicating sample. Process as for the replicating sample, steps 2| C (iv) - (vii).

■ **PAUSE POINT** Samples can be stored at -20°C for up to a year.

Genomic DNA extraction •**TIMING** **3.5 h (hands-on time: 1.5 h)**

3| Thaw cell pellets (if required), resuspend cells in 500 µl of solution 1 and add 50 µl of 10 mg/ml zymolyase T20 and 0.5 µl β-mercaptoethanol.

4| Incubate samples at 37°C for 30 min in a water bath.

5| For sort-seq samples (from step 2 A (xix)) follow option A, and for sync-seq samples (from step 2 B (xvi)) or MFA-seq samples (from steps 2 C (vii) and (ix)) follow option B.

(A) Sort seq samples

(i) Set up the following reactions for each of the samples, incubate at 55°C for 2 h in a water bath, then proceed to step 6.

Component	Amount (µl)	Final concentration
Sort-seq sample from step 4|	555.5	
NaCl (3 M)	17	85 mM
SDS (20% wt/vol)	12.5	0.42% (wt/vol)
Proteinase K (20 mg/ml)	10	0.33 mg/ml
RNase A (10 mg/ml)	5	83.3 µg/ml
Total	600	

(B) Sync-seq and MFA-seq samples

(i) Pellet cells by centrifugation at 16,000*g* for 2 min at RT, remove supernatant and set up the following reactions for each sample. Incubate at 55°C in a water bath for 2 – 6 h, then proceed to step 6.

Component	Amount (µl)	Final concentration
Sync-seq or MFA sample	25	
Sorbitol (1 M)	50	83.3 mM
Solution 2	500	
Proteinase K (20 mg/ml)	10	0.33 mg/ml
RNase A (10 mg/ml)	5	83.3 µg/ml
Total	600	

6| Add 0.5 ml of phenol:chloroform:isoamyl alcohol (25:24:1) and invert vigorously to mix.

7| Centrifuge samples at 16,000*g* for 5 min at RT and transfer upper aqueous layer into a fresh microcentrifuge tube. **! CAUTION** Phenol is toxic and corrosive. Wear PPE and work in a fume hood.

8| Add another 0.5 ml of phenol:chloroform:isoamyl alcohol (25:24:1), mix well by inversion, centrifuge samples at 16,000*g* for 5 min at RT and transfer upper phase into a fresh microcentrifuge tube.

9| Add 0.5 ml of chloroform, mix well by inversion, centrifuge samples at 16,000*g* for 5 min at RT and transfer upper phase into a fresh microcentrifuge tube. ▲ **CRITICAL STEP** Ensure complete removal of residual phenol. If required, perform a second chloroform extraction step.

10| Add 1 ml of 100% (vol/vol) ethanol, and invert tube 10 times. Incubate sort-seq samples at -20°C for 1 h (this incubation is unnecessary for sync-seq and MFA-seq samples).

11| Centrifuge samples at 16,000*g* for 10 min at RT and remove supernatant carefully without disturbing the DNA pellet.

12| Add 200 µl of 70% (vol/vol) ethanol and invert tube carefully.

13| Remove supernatant carefully without disturbing the DNA pellet.

14| Pulse-spin the sample and remove the remaining ethanol.

15| Leave tube open at room temperature for 2 min to allow DNA pellet to dry.

16| Resuspend DNA pellet in 100 µl of TE.

17| Measure DNA concentration using Qubit™ dsDNA HS assay. The yield should be 50 – 100 ng of DNA for sort-seq samples and 1 – 10 µg of DNA for sync-seq or MFA-seq samples. Assess DNA

purity by NanoDrop™ spectrophotometer. Pure DNA samples should have absorbance ratios of ~1.8 for 260/280 and 2.0-2.2 for 260/230. **? TROUBLESHOOTING**

■ **PAUSE POINT** Store DNA at 4 °C for up to 1 month or at -20°C for longer storage.

Fragmentation of DNA by sonication •TIMING 1.5 h (hands-on time: 45 min)

▲ **CRITICAL** For sequencing using the NextSeq 500 (Illumina) platform, DNA should be fragmented to 300-550 bp. If using other sequencing platforms, the recommended insert size should be checked before fragmentation. We fragment DNA using sonication (Bioruptor Standard), however, fragmentation of DNA could also be achieved using other types of mechanical or enzymatic fragmentation.

▲ **CRITICAL** If other fragmentation methods are used, the conditions for the fragmentation procedure must be optimised to generate 300-550 bp fragments.

18| Decide on the input amount of DNA (5 ng – 1 µg) for library making. It is best to sonicate more than the minimum amount as some DNA may be lost during sonication and precipitation.

19| Dilute your chosen amount of DNA into TE in a total volume of 400 µl, in a 1.5 ml microcentrifuge tube. We have not encountered problems when using DNA concentrations lower than recommended by the manufacturer of the sonicator.

▲ **CRITICAL STEP** The volume is critical to ensure homogeneous shearing. Take care to dilute samples in TE and not water.

20| Insert sample tubes into the microcentrifuge tube holder of a Bioruptor standard, filling empty holes with blank tubes containing 400 µl liquid.

21| Sonicate for 20 min, on setting L, 30 s on, 30 s off, in ice cold water.

22| Precipitate DNA by adding 2x volumes of 100% (vol/vol) ethanol and mix by inversion. Centrifuge for 10-30 min at 13,000*g*, 4°C.

23| Carefully discard supernatant, wash pellet in 500 µl 80% (vol/vol) ethanol and air dry. Resuspend pellet in 29 µl 1x TE.

24| Check fragmentation is predominantly between 300-550bp using 2 µl of your sample with a TapeStation (see Figure 4a; alternatively use Bioanalyzer, Agilent or a PAGE gel), and determine concentration by Qubit (or another high-sensitivity DNA concentration measurement). **? TROUBLESHOOTING**

■ **PAUSE POINT** DNA can be stored overnight at 4°C. We do not recommend leaving fragmented DNA for longer than this.

End Prep •TIMING 1.5 h (hands-on time: 10 min)

25| If you have multiple samples, the amount of DNA carried forward into library preparation is determined by the sample with the lowest yield. Take maximal, but equal amounts of DNA for each sample by diluting samples as appropriate in TE.

26| For each sample, add the following to a PCR tube or strip, then mix and pulse spin:

Component	Amount (µl)	Final concentration
500 pg – 1 µg Fragmented DNA in TE	25	
NEBNext Ultra II End Prep Enzyme Mix	1.5	0.05X
NEBNext Ultra II End Prep Reaction Buffer	3.5	0.12X
Total	30	

27| In a thermal cycler with heated lid at 100°C, incubate the PCR tubes for the following amount of time:

20°C	30 min
65°C	30 min
4°C	hold

▲ **CRITICAL STEP** Continue to next step ideally straight away, or within 30 min.

Adaptor ligation. •TIMING 1 h (hands-on time: 20 min)

28| Depending on the amount of DNA carried forward into library preparation (step 25), prepare a dilution (1.25 µl per sample) of NEBNext Adaptor for Illumina in 10 mM Tris-HCl, pH 8.0, 10 mM NaCl. Use the following table as a guide:

Amount of DNA	Dilution of adaptor
101 ng - 1 µg	no dilution
5 ng - 100 ng	1:10
<5 ng	1:25

29| Add the following to each End Prep tube from step 27|, then mix and pulse spin:

Component	Amount (µl)	Final concentration
End prep sample (step 27|)	30	
NEBNext Ultra II Ligation Master Mix	15	0.32X
NEBNext Ligation Enhancer	0.5	0.01X
NEBNext Adaptor for Illumina dilution from step 28|	1.25	0.027X
Total	46.75	

30| In a thermal cycler with the heated lid off, incubate tubes for 25 min at 20°C.

31| Add 1.5 µl USER enzyme (NEBNext Multiplex Kit) to each PCR tube and mix.

32| Incubate 15 min at 37°C in a thermal cycler with the heated lid at 50°C.

■ **PAUSE POINT** Samples can be stored overnight at -20°C, however, we prefer to pause after bead cleanup.

Clean up •TIMING **1.5-2.5 h (hands-on time: 40 min)**

33| Equilibrate AMPure XP beads to RT, 30 min. Vortex 30 s.

34| Prepare 400 µl per sample of fresh 80% (vol/vol) ethanol.

35| Add 43.5 µl (0.9X) beads to USER reactions from step 32|, transfer to a 1.5 ml microcentrifuge tube. Vortex tubes and incubate on a rotating wheel at room temperature for 10-60 min.

36| Pulse spin tubes and place on magnetic rack for 5 min until beads have separated. Discard supernatant without disturbing the beads.

37| Pipette 200 µl freshly made 80% (vol/vol) ethanol into the tube on the side away from the beads. Incubate for 30 s and remove ethanol. Repeat this 80% (vol/vol) ethanol wash.

38| Close tubes and pulse spin. Pipette off any remaining ethanol. Air dry for ~30 s or until the beads are still glossy but starting to dry at the edges.

▲ **CRITICAL STEP** Both excess ethanol and overly dried beads can lead to poor DNA recovery. If getting poor DNA recovery, we recommend that the air dry time is optimised.

39| Elute DNA by adding 8 µl of 10 mM Tris HCl pH 8.0 and vortex well, place in a shaking heat block at 37°C for 10 min.

40| Pulse spin tubes and place on a magnetic rack for 5 min or until separated (the beads may not separate quite as well as in step 36|). Transfer 7.5 µl to new PCR tube.

■ **PAUSE POINT** Samples can be left overnight at 4°C.

PCR enrichment of adaptor ligated DNA •TIMING **1 h (hands-on time: 20 min)**

41| Decide on a suitable number of PCR cycles for your starting amount of DNA. We use 3 cycles (the minimum number of cycles) for 450 ng, 4 cycles for 230 ng, 7 cycles for 45 ng.

42| Decide on the index primers to use for your samples taking account of NEB recommended combinations of primers for pools of 6 samples or fewer.

43| To the PCR tubes with your samples from step 40|, add the following (total 25 µl), then mix and pulse spin:

Component	Amount (µl)	Final concentration
Adaptor ligated samples from step 40|	7.5	
NEBNext Ultra II Q5 Master mix	12.5	0.5X
Index Primer (NEBNext Multiplex Kit)	2.5	0.1X
Universal primer (NEBNExt Multiplex Kit)	2.5	0.1X
Total	25	

44| In a thermal cycler with heated lid at 100°C amplify with the following PCR conditions:

Cycle number	**PCR conditions**	
1	Denature	98°C 30 s
2-4*	Denature Anneal/Extension	98°C 10 s 65°C 75 s
5	Extension	65°C 5 min
6	Hold	4°C

***** Repeat 3+ cycles depending on the amount of input DNA (see step 41).

Clean up •TIMING 1.5-2.5 h (hands-on time: 40 min)

45| Repeat steps 33|- 40| except add 22.5 µl (0.9X) beads and 33 µl 0.1X TE for elution. Transfer final library to a new tube.

Quality control checks •TIMING 3 h (hands-on time: 1 h)

46| Check read fragment profile of final library by Tapestation using 2 µl of final library (see Figure 4). Alternatively, a Bioanalyzer or PAGE gel can be used. **? TROUBLESHOOTING**

47| Check concentration of correctly adaptor-ligated fragments for each library with NEBNext Library Quant Kit for Illumina using a quantitative PCR instrument and appropriate tubes. We use a Qiagene Rotor Gene and PCR tubes, using 20 µl reactions, without ROX, as recommended by the NEBNext Library Quantification protocol. We routinely just check 1:100,000 dilutions of final libraries in triplicate, however, initially we recommend running both 1:10,000 and 1:100,000 dilutions of final libraries in triplicate. These dilutions should be made in 1X NEBNext Library Quant Dilution Buffer using a series of 1:100 and 1:10 dilutions.

48| Prepare sufficient NEBNext Library Quant Master Mix (with primers) for 15 control reactions, 3 reactions per library and one spare reaction, where each reaction requires 16 µl of the mix. The mix is made by adding a 1/16 volume of NEBNext Library Quant Primer Mix to a 15/16 volume of NEBNext Library Quant Master Mix. For example, for 8 libraries prepare 640 µl NEBNext Library Quant Master Mix (with primers) by adding 40 µl NEBNext Library Quant Primer Mix to 600 µl NEBNext Library Quant Master Mix.

49| For a no template control (1X Library Dilution Buffer), each NEBNext Library Quant Kit DNA standard (10, 1, 0.1, 0.01 pM) and each library sample prepare qPCR reactions in triplicate:

Component	Amount (µl)
NEBNext Library Quant Master Mix (with primers) from step 48	16
No template control*, DNA standard or library dilution	4
Total	20

* The Ct value from the no template control allows detection of reagent contamination and it is not used in subsequent quantitation analysis.

50| Run qPCR programme, and acquire signal:

Cycle number	**qPCR conditions**	
1	Denature	95°C 60 s
2-36	Denature Anneal/Extension Acquisition	95°C 15 s 63°C 45 s (470 nm / 510 nm)

51| For each DNA standard, determine the average Ct value from the triplicates. Use these values to plot a standard curve of average Ct against log(pM) and fit a linear regression from which the gradient (m) and intercept with the y-axis (c) can be determined. Calculate the pM concentration for all library dilution triplicates from the Ct value and standard curve gradient (m) and intercept (c) values, using the formula:

$10^{((c - \text{Ct value})/m)}$

Average across triplicates, and adjust for average fragment size for library (from Tapestation) using the formula:

(Triplicate average x 399) / average library fragment size (from step 46)

52| Multiply the corrected dilution concentration by the dilution factor e.g. x 10,000 for a 1:10,000 dilution, and average the concentration across both 1:10,000 and 1:100,000 dilutions if both dilutions were performed. This gives the final concentration of the correctly adaptor ligated fragments of the library. **? TROUBLESHOOTING**

Steps 51 and 52| can be carried out using NEB's NGS Library Quant web tool: http://nebiocalculator.neb.com/#!/qPCRlibQnt

■ **PAUSE POINT** Libraries can be stored at 4°C for two weeks or at -20°C for 12 months.

Preparation for, and loading sequencing run •TIMING 13 h (hands on time: 1 h)

53| Set up a NextSeq run using Basespace (Illumina, https://basespace.illumina.com/home/index). Fill in sample names, barcode numbers and cycle number settings. By default we use single-end sequencing, but the downstream computational pipeline can also be used with paired-end sequencing data.

54| The evening before the NextSeq run: Follow instructions provided with NextSeq 500/550 High Output Kit v2.5 for cartridge preparation.

55| Pool libraries to obtain >5 µl at 4 nM. Optionally verify the pooled library concentration using the NEBNext Library Quant Kit as described in steps 47-52.

56| Denature, dilute and load library as recommended for NextSeq500/550 by Illumina. A 75 cycle run takes approximately 13 h from loading the library to downloading data from Basespace.

▲ **CRITICAL STEP** The optimal concentration of library to load for sequencing using a NextSeq 500/550 should be optimised by individual laboratories. We have found loading a 2.2 pM pooled library generates a high read number without compromising read quality. We expect to obtain 400-500 million reads where >85% of reads have a quality score ≥Q30 per NextSeq run.

■ **PAUSE POINT** Undiluted pooled libraries can be stored at 4°C for two weeks or at -20°C for 12 months.

Sequencing read alignment and generation of raw read counts •TIMING 2-8 h (hands-on time: 30 min)

57| Login to your Illumina account. Check run quality information and download the fastq sequencing files. Alternatively, follow instructions provided by your sequencing facility to download fastq sequencing files. **? TROUBLESHOOTING**

58| Use FastQC to make sure that the sequencing files are intact and to check the quality of the data generated from the sequencing run. See troubleshooting Table 1| for more information. **? TROUBLESHOOTING**

59| If using a reference genome for the first time, it needs to be indexed with Bowtie 2. Obtain a multiple sequence fasta file containing the genome contigs (for example, sacCer3 is available from https://www.ncbi.nlm.nih.gov/assembly/GCF_000146045.2/) and use the following command to build a Bowtie 2 index:

```
$ bowtie2-build -f /path/to/genome.fasta genomeName
```

This will create indexed genome files with bw2 or bw2l extensions in the directory where the command was executed.

60| To align the reads to the index genome (created in step 59|) and count the reads in genomic windows (bins), download the localMapper script from gitHub (https://github.com/DNAReplicationLab/localMapper) and make executable:

```
$ sudo chmod +x filePath/localMapper.sh
```

Navigate to the directory containing the sequencing files and follow option A for single-end sequencing and Option B for paired-end sequencing. Use -h argument to print more information about available options. By default, the script will create a subdirectory with the provided sample name. Once the script has finished, the sample_name/processed directory will contain a non-empty bed file with read counts in genomic bins of the size specified by the -w argument. We typically use 1 kb bins for small genomes (e.g. yeasts) and 50 kb bins for larger genomes (e.g. mammals) – discussed further in the section Experimental Design.

:

(A) Single-end sequencing

(i) Run the script as follows:

```
$ filePath/localMapper.sh -g <genome name> -U <fastq file(s)> -s
<sample name> -c <number of threads> -m <memory per thread> -w
<window size in bp>
```

(B) Paired-end sequencing

(i) Run the script as follows:

```
$ filePath/localMapper.sh -g <genome name> -1 <first mate
fastq file(s)> -2 <second mate fastq file(s)> -s <sample
name> -c <number of threads> -m <memory per thread> -w
<window size in bp>
```

Generate, normalise and plot the ratios •TIMING 5-30 min

61| To obtain replication profiles, the count files from replicating samples have to be normalised using counts from non-replicating samples to produce copy number ratios. The ratios are normalised based on the sequencing depth, followed by a dynamic range adjustment. The adjusted ratios are visualised and may be compared using statistics. All these steps can be performed using the Repliscope package available from CRAN and gitHub (https://github.com/DNAReplicationLab/Repliscope). To perform the analysis via the command line follow option A. To perform the analysis interactively in a web browser follow option B.

(A) Data analysis using command line functions.

(i) Launch R and load the Repliscope package:

```
$ R
> library(Repliscope)
```

(ii) Load the bed file from the non-replicative sample produced in step 60|.

```
> nonRepSample <- loadBed('filePath/sample.bed')
```

(iii) Visualise the loaded read counts as a boxplot to check for aneuploidy and outliers. All chromosomes, except for mitochondria, should cluster around the genome wide median, as shown in Figure 5a. Aneuploidy can be identified if one or more chromosomes are higher or lower than the median line (Figure 6a). A group of outliers on a single chromosome above the median indicates a segmental duplication (Figure 6b).

```
> plotBed(nonRepSample)
```

If outliers are detected, use rmOutliers() function to remove them. As a standard, we remove bins that contain less than a quarter of genomic median counts value – these tend to produce noisier ratios:

```
> nonRepSample <- rmOutliers(nonRepSample, 'median', loLim=0.25)
```

? TROUBLESHOOTING

(iv) Visualise read counts as a scatterplot using the following command:

```
> plotCoverage(nonRepSample)
```

This provides a visual check that read count values are even across chromosomes (Figure 5c). Curvature on the plot may indicate that there was ongoing replication in the sample and it may be necessary to use a different non-replicative control. **? TROUBLESHOOTING**

The plot also allow aneuploidy to be identified if one or more chromosomes are higher or lower than the median line (Figure 6c), while a segmental duplication is seen as a region of a chromosome that is abruptly higher than the median (Figure 6d).

(v) Repeat steps 61 A (ii)-(iv) using the replicative sample counts file and replace 'nonRepSample' with 'repSample' in the commands (Figure 5b and d).

(vi) Create the sequencing depth-normalised count ratio:

```
> theRatio <- makeRatio(repSample,nonRepSample)
```

This creates a dataframe that contains a 'ratio' column with calculated values adjusted by total read number.

(vii) Visualise the calculated ratio values as a histogram (Figure 5e).

```
> plotRatio(theRatio$ratio)
```

The values should be mostly between 0.5 and 1.5 and centre around 1. Outliers may be removed using the trimRatio() function.

(viii) Normalise the ratio to fit a biologically relevant scale, either using the default autonormalisation feature or by providing a normalisation factor manually (Box 3|). In case of sync-seq, the normalisation factor is calculated from bulk replication values obtained in step 2| B (xxiv) using the formula 1 + [bulk genome replication as a decimal]. Use the following command to normalise a sync-seq sample with 35% of bulk genome replication:

```
> theRatio <- normaliseRatio(theRatio,rFactor=1.35)
```

In case of sort-seq, the normalisation factor is determined algorithmically, by minimising the sum of values outside of the 1 to 2 scale. To do so, run the function without additional arguments:

```
>theRatio <- normaliseRatio(theRatio)
```

For MFA-seq, the maximum relative copy number can be calculated as 1 + [proportion of cells in S phase]. To run the minimising function with the new maximum (e.g. 1.25), run the function like this:

```
> theRatio <- normaliseRatio(theRatio,upperLimit=1.25)
```

Visualise the normalised ratio using the command from step 61 A (vii) (Figure 5f).

(ix) Plot the resulting replication profile:

```
> plotGenome(theRatio)
? TROUBLESHOOTING
```

(x) To plot smoothed data, run the spline smoothing function:

```
> theRatio <- smoothRatio(theRatio)
```

Repeat step 61 A (ix) to plot smoothed ratio values (Figure 5g).

(xi) To compare two replication profiles, each generated as described above in steps 60 – 61 A (viii) and named theRatio1 and theRatio2, use the following command:

```
> comparedRatios <- compareRatios(theRatio1,theRatio2)
```

The resulting dataframe will contain both replication profiles and an additional ‘p.value’ column. It can be plotted as in step 61 A (ix).

(xii) To plot multiple replication profiles (each generated as described above in steps 60 – 61 A (viii) and named ratio1, ratio2, ratio3, etc) on the same plot (e.g. for sync-seq), combine them first into a single dataframe:

```
> combinedReplicationProfiles <- rbind(ratio1,ratio2, ratio3…)
```

These replication profiles can be plot using the command in step 61| (A) (ix). An example plot is shown in Figure 5h.

(B) Data analysis using our Shiny app.

(i) Launch R, load and launch the Repliscope in the interactive mode:

```
$ R
> library(Repliscope)
> runGUI()
```

(ii) Follow instructions on the webpages of the app to create, normalise and plot the replicating to non-replicating sample counts ratios (Supplementary Note). **? TROUBLESHOOTING**

62| Depending on the specific application users may then wish to perform custom analyses on the data generated, some of which we suggest in Box 4|.

TIMING

Step 1|, growing primary cell culture: 17 h (hands-on time: 10 min)
Step 2|, obtaining replicating and non-replicating cells: 1.5 - 3 days
Step 2A (i) - (xix), sort-seq: 2 days (hands-on time: 4 h)
Step 2B (i) - (xxv), sync-seq: 1.5 days (hands-on time: 3 h)
Steps 2B (xxiii) - (xxv), process flow cytometry samples to check cell cycle synchrony: 4h (hands-on time: 1.5h)
Step C (i) - (ix), MFA-seq: 3 days (hands-on time: 2 h)
Steps 3| - 17|, genomic DNA extraction: 3.5 h (hands-on time: 1.5 h)
Steps 18| - 24|, fragmentation of DNA by sonication: 1.5 h (hands-on time: 45 min)
Steps 25| - 27|, end Prep: 1.5 h (hands-on time: 10 min)
Steps 28| - 32|, adaptor ligation: 1 h (hands-on time: 20 min)
Steps 33| - 40|, clean up: 1.5-2.5 h (hands-on time: 40 min)
Steps 41| - 44|, PCR enrichment of adaptor ligated DNA: 1 h (hands-on time: 20 min)
Step 45|, clean up: 1.5-2.5 h (hands-on time: 40 min)
Steps 46| - 52|, quality control checks: 3 h (hands-on time: 1 h)
Steps 53| - 56|, preparation for, and loading sequencing run: 13 h (hands on time: 1 h)
Steps 57| - 60|, sequencing read alignment and generation of raw read counts: 2-8 h (hands-on time: 30 min)
Step 61|, generate, normalise and plot the ratios: 5-30 min
Step 62|, custom analysis: user defined

TROUBLESHOOTING

Table 1| Troubleshooting

Step	Problem	Possible reason	Solution
2| A (x)	Cell flocculation	Insufficient cell separation	Perform additional sonication, step 2| A (ix).
2| A (x)	Cell debris	Excessive sonication Bacterial infection in yeast culture	Repeat sample preparation and separate cells by vortexing Repeat sample preparation, ensure sterility of reagents and consider adding an antibacterial agent
2| A (x)	Insufficient discrimination of cell cycle stages by FACS	Incomplete Proteinase K or RNaseA digest Insufficient SYTOX™ Green labelling Misaligned FASC lasers	Check using correct PK and RNaseA digestion conditions. Check using correct SYTOX™ Green labelling conditions. Check alignment of FASC lasers
2| B (viii)	Incomplete alpha-factor arrest	Non-optimal arrest conditions for strain	Check *BAR1* status of yeast. Optimise alpha factor arrest for your strain of interest
17|	RNA contamination	Incomplete digestion by RNase A	Check using correct RNaseA digestion conditions, if necessary additional RNaseA digestions can be performed (step 5|)
17|	Low amount of extracted DNA	Incomplete cell recovery during centrifugation Lost pellets Incomplete DNA resuspension	Centrifuge at recommended speeds Take care at step 2 A (xix) and steps 11|, 13|, 14| Leave overnight in the fridge or heat at 60°C step 16|
17|	Low 260/230 ratio	Contaminants in DNA	Include additional chloroform step

24|	Lower than expected sized fragments	Sonicated in water Sonicated for too long or using too high power	Sonicate in TE (step 19|) Check using correct sonication conditions or optimise if using alternative sonicator (step 21|).
24|	Larger than expected sized fragments	Sonicated in smaller volume than 400 μl	Sonicate samples in 400 μl, the volume is more critical than the concentration (step 19|).
46|	Shorter than expected library size	Adapter dimer contamination Starting size of fragmented DNA too small	Repeat AMPure XP bead clean up, step 45| or repeat library making steps with higher concentration of DNA and correct adaptor concentration (step 28|) See troubleshooting for step 24|
52	Low library concentration	Overestimation of DNA input	Check 260/230 nm and 260/280 nm ratios (step 17|). Use higher starting concentration of DNA (step 18|). Increase the number of PCR cycles during library amplification (steps 41 – 44).
57|	Low read number sequenced but high Q30 score	Low concentration of pooled library loaded onto NextSeq	Optimise concentration of loaded pooled library (step 56|).
57| and 58|	High read number but low Q30 score and FastQC quality scores	Too high concentration of pooled library loaded causing over formation of clusters	Optimise concentration of loaded pooled library (step 56|).
58|	High amount of sequence duplication	Low sequencing complexity due to overamplification of limited library	Use higher starting concentration of DNA (step 18|)
58|	High adaptor content	Adaptor dimer contamination	See troubleshooting for step 46|
58|	Presence of index sequences seen in kmer content	Index sequences not removed appropriately	Manually trim reads to remove index sequences
61| **A** (iii) and 61 (B) (ii)	Different average read count for one or more chromosomes	Possible aneuploidy or segmental duplication	See Box 2|
61| (A) (iv) and 61| **B (ii)**	Discontinuity in genomic reads	Possible genomic rearrangement	See Box 2|
61| (A) (iv) and 61| (B) **(ii**)	Increased recovery of chromosome ends in non-replicating sample	Biases from DNA purification, a known problem with some DNA extraction kits	Purify DNA by Phenol-chloroform extraction and ethanol precipitation
61| (A) (iv) and 61| (B) **(ii**)	Non-replicating sample is not flat across the genome	The sample is not completely non-replicating and is not suitable to use for normalisation	Use a different non-replicating sample from the same run if possible or repeat sample collection
61| (A) (ix) and 61| (B) **(ii)**	Low dynamic range, i.e. data points do not reach the expected maxima	Low dynamic range in original cell sample	Poor sort purity or gate setting (steps 2| A (x) - (xii)) e.g. contamination of the S phase sample with G2 cells reduces the dynamic range as does the

			S phase gate not capturing the full range of S phase cells Poor cell cycle synchronisation – optimise (steps 2| B (vi) - (xi))
		Inter-cellular stochasticity of replication	Other techniques can be useful to address stochasticity, such as D-NAscent[43] which identifies DNA replication on single molecules.

Anticipated Results

Upon successful sequencing and read mapping, non-replicating samples should have a fairly flat profile, featuring a coefficient of variation (CV) of less than 20%. This is evidenced by similar distribution of read number between chromosomes (Figure 5a) and sequencing coverage within the chromosomes (Figure 5c). Some windows may contain fewer than expected reads due to filtering out reads mapping to non-unique regions. Replicating samples, on the other hand, should have a wider distribution of read numbers (CV around 100%) as evidenced by the increased inter-quartile range in box plots (Figure 5b) and curved coverage profiles across chromosomes (Figure 5d). Once adjusted by the sequencing depth ratio between replicating and non-replicating samples, data points should be distributed around 1 (Figure 5e). Further normalisation will adjust the ratio distribution to fit a biological relevant scale, usually to fit a 1 to 2 relative copy number scale (Figure 5f). Normalised ratios, when plotted as a function of genomic position, generate the replication profile for the sample (Figure 5g-i). The curvature of the profile should be similar to the replicating sample read coverage and inversely proportional to the replication time of the DNA region[4].

Boxes

Box 1| Adapting for other organisms

Although the protocol we describe here uses *Saccharomyces cerevisiae* as the example organism, this family of methods can be applied to virtually any organism. Some adaptation to the protocol may be necessary for other organisms, nevertheless, we have found it straightforward to apply this family of methods across the three domains of life. Below we provide general guidance for how to adapt the protocol and a list of organisms where this protocol has been successfully applied.

Steps 1 and 2 (obtaining replicating and non-replicating cell populations) require organism-specific adaptation. As detailed in 'Choice of Variant' above, this is most straightforward for sort-seq and MFA, where culture conditions should be optimised to maximise cell proliferation and thus obtain a high fraction of replicating cells. Note that for sort-seq the most appropriate DNA stain and cell sorting condition may need to be optimised, for example we have found propidium iodide to be optimal when sorting cultured mammalian cells[18]. For sort-seq, the non-replicating samples are obtained by cell sorting of either the G1 or G2/M population, whichever is most numerous for the organism and growth conditions employed. For example, for mammalian cells we have used sorted G1 cells as the non-replicating control sample. For sync-seq a critical requirement is to have a cell population pass synchronously through S phase. While this generally requires an organism-specific protocol, elutriation has been successfully applied to a wide range of yeast species for which mating pheromone-based arrest and release approaches are not available[16,48].

The DNA extraction steps (3 – 17) should also be optimised to suit the organism of choice; we recommend following published protocols. However, we have observed an extraction bias in DNA samples that were prepared using DNA purification columns, therefore we recommend phenol:chloroform extraction and ethanol-precipitation. Once DNA has been obtained from the samples, library making and sequencing can be performed as described here (steps 18 – 55), with consideration of the number of reads required based upon genome size (see 'Multiplex Sequencing and Sequence Depth Requirements' section in Experimental Design). Finally, some organisms may benefit from optimisation of the computational protocol, particularly, the read mapping steps (58 and 59). For example, we have found STAR[49] to be a more efficient read mapper than BOWTIE2 when working with the human genome[18]. More details on adapting the protocol to other organisms are available from the studies listed below.

Organism	Example reference
Fission yeast species	Sort-seq: *Schizosaccharomyces pombe*[48] Sync-seq: G2 synchronisation by centrifugal elutriation *Schizosaccharomyces pombe*[48] G2 synchronisation by centrifugal elutriation *Schizosaccharomyces pombe*, *S. octosporus*, *S. japonicus*[50] *Cdc25-22* temperature G2 arrest *Schizosaccharomyces pombe*[51]
Budding yeast species	MFA-seq: *Lachancea fantastica, L. meyersii, L. dasiensis, L. nothofagi, L. waltii, L. thermotolerans, L. mirantina, L. fermentati, L. cidri, L. kluyveri* [16] *L. kluyveri*[52] *L. waltii*[53] Sort-seq: *Kluyveromyces lactis, Lachancea kluyveri, Candida glabrata, Naumovozyma castellii, Tetrapisispora blattae, Zygosaccharomyces rouxii*[21] *Saccharomyces paradoxus, S. arboricolus, S. bayanus*,[12] Sync-seq: G1 synchronisation by centrifugal elutriation *Lachancea fantastica, L. meyersii, L. dasiensis, L. nothofagi, L. waltii, L. thermotolerans, L. mirantina, L. fermentati, L. cidri, L. kluyveri*[16]

Other unicellular eukaryotes	Sort-seq: *Trypanosoma brucei*[41,54,55] *Leishmania major, L. Mexicana*[56]
Archaea	MFA-seq: *Sulfolobus acidocaldarius, Sulfolobus solfataricus*[57] *Haloferax volcanii*[14] Sync-seq: Acetic acid arrest *Sulfolobus acidocaldarius*[57]
Bacteria	MFA-seq: *Escherichia coli*[26]
Homo sapiens	MFA-seq: 161 individuals sequenced by the 1000 Genomes Project[17] Sort-seq: CO202 (microarray)[58] HeLa[18] Six lymphoblastoid cell lines[17]

Box 2| Genotypic information from replication profiles

Genome-wide DNA copy number measurements can reveal (unanticipated) genotypic variations. ▲ **CRITICAL** Visualise the non-replicating sample without normalisation.

Aneuploidy: Most cells contain complete chromosome sets with each chromosome present at the same copy number. Ploidy differences result in individual chromosomes with higher (or lower) copy number than the remaining genome (see Figure 6a, c). Stable aneuploid karyotypes result in integer copy numbers, whereas fractional increases (or decreases) indicate heterogeneity in the population.

Segmental duplications: Chromosomal regions with a copy number greater than the flanking sequences and other chromosomes indicate localised sequence amplifications (see Figure 6b, d).

Chromosomal rearrangements: Adjacent chromosomal sequences replicate at very similar times in S phase, giving rise to smooth replication profiles. Steep jumps in the copy number of adjacent sequences often indicate chromosomal rearrangements relative to the reference sequence[14]. The karyotype of the cell population should be analysed further.

Box 3| Normalisation

All three variants of the method rely on a non-replicating control sample to normalise for coverage biases from the genomic sequence context, including non-unique sequences that are discarded during the analysis.

Sort-seq takes advantage of a sorted non-replicating fraction, either G1 or G2/M.

Sync-seq uses an arrested, unreleased G1 sample.

MFA-seq uses a stationary phase culture as a non-replicating sample.

While the above ways of obtaining non-replicative samples are commonly used with the respective methods, there is no strict requirement for a particular one. For example, an arrested sample may be used as a non-replicative control in either MFA-seq or sort-seq.

▲ **CRITICAL** Visualise the non-replicating sample without normalisation to make sure that the read counts profile is even across all regions.

The first normalisation step is to calculate ratios between read counts of a replicating sample (B) and a non-replicating sample (A), adjusted for the differences in total read number between the two samples. The resulting ratios have a mean of 1 (C).

The second normalisation involves multiplying the ratio values by a normalisation factor in order to fit the data to a biologically relevant relative copy number scale (in most cases from 1-2) (D). This is calculated differently for sync-seq, sort-seq and MFA-seq.

Sync-seq samples are normalised to their bulk genome content measured by flow cytometry. For accuracy the median DNA content for every timepoint should be plotted and fit to a sigmoid function, before determining the percentage of genome replication for relevant timepoints. This can be done for experimental flow cytometry data using our online tool: https://dnareplab.shinyapps.io/fitsigmoid/. The normalisation factor for individual timepoints is then calculated as 1 + [fraction replicated]. For example, for a sample with 50% bulk genome replicated, the normalisation factor would be 1.5.

For sort-seq samples relative copy number values range between 1 and 2. Therefore, the normalisation factor is calculated to maximise the number of data points lying within this range (D). Normalisation factors of 1.3-1.5 are usually appropriate. This is implemented in the 'Auto normalisation' function of the Repliscope package.

For samples containing both replicating and non-replicating cell populations (i.e. MFA-seq), a similar approach can be used, but with a different maximum value for relative copy number. The value can be calculated as 1 + [proportion of replicating cells]. For example, an asynchronous culture of *S. cerevisiae* contains ~25% replicating cells. Thus, the expected relative copy number range is 1 to 1.25.

Box 4| Data interpretation and custom analysis

Following our protocol users will generate replication timing profiles for their cell type(s) of interest, and can identify genomic regions with a statistically significant difference between two samples. To facilitate comparison of sync-seq datasets, median replication time, or Trep, can be calculated from a series of timepoints throughout S phase, as described previously[4,9]. We provide this functionality within the Repliscope package. However, depending on the specific research question there may be other custom analyses that the user wishes to perform, some of which we suggest here.

Replication profiles allow localisation of active replication origins. These are evident as peaks in replication profiles and therefore can be identified with peak calling tools, such as in Müller et al. 2014[4]. Motif searches can be performed as described previously[4], to look for sequence motifs associated with replication timing profile peaks and, therefore, origins. In addition, the proportion of forks moving in either direction genome-wide can be determined from the gradients of replication timing profiles, and therefore can reveal firing efficiencies for replication origins[9,59].

Evolutionary comparisons can be made between replication timing profiles from related species of various evolutionary distance, for example using the liftOver tool or considering regions of various levels of conservation[12,16,21]. Replication timing profiles can also be compared to genomic variation, such as SNPs and alleles identified that associate with replication timing differences[17]. Of note, replication timing profiles can generate genotype information such as large-scale rearrangements (see box 2 and [14]). Finally, as a genome-wide dataset, replication timing profiles can be compared with other genome-wide datasets of interest, such as ChIP-seq, RNA-seq or Hi-C.

Figure legends

Figure 1: Workflow overview

For each variant, the protocol starts with obtaining replicating (cyan) and non-replicating (grey) cells (step 1-2). Sort-seq uses FACS to enrich cells in S phase or late G2 phase as indicated. Sync-seq uses cell cycle arrest and release to collect a non-replicating G1 sample and replicating samples at timepoints during S phase. In MFA-seq log and stationary phase cultures provide the replicating and non-replicating samples, respectively. Criteria for selecting the most appropriate variant are discussed in the section 'Choice of Variant'. The DNA isolated from the samples (steps 3-17) is prepared for high-throughput sequencing using short-read technology (steps 18-56). The resulting reads are aligned and summed in genomic bins (steps 57-60); example replicating and non-replicating data are shown with the whole genome median read count represented by a horizontal pink line. Finally, the replicating data can be normalised using the non-replicating data to generate replication timing profiles (step 61); an example replication profile is shown with smoothing in orange.

Figure 2: Cell sorting gates

Cell sorting for S and G2 phases from *S. cerevisiae.* Cells were labelled with SYTOX™ Green as described in 2| A (viii). a) A typical ungated plot showing DNA content signal (FL1) Height/Area. Cells are gated to remove doublets and debris (step 2| A (x)). b) Histogram shows DNA content post-gating. Sorting gates for S and G2 phases of the cell cycle, as well as the 'ALL' control fraction are indicated (step 2| A (xi)-(xii)). c) Typical flow cytometry profiles of the DNA content from sorted cell populations to check sort purity, step 2| A (xviii).

Figure 3: Sync-seq timecourse

Flow cytometry analysis of sync-seq timepoints. Samples were prepared as described in steps 2 B (xix – xxiv). a) A typical increase in DNA content in a *S. cerevisiae* population released synchronously from alpha factor arrest. In b) the median DNA content of each timepoint was plotted against the time the sample was taken and a sigmoidal curve was fit. Timepoints for HTS are chosen based on the bulk genome replication - early S phase was captured between 35 – 40 minutes, mid-S phase between 45 – 50 minutes and late S phase 55 – 60 minutes. This quantification of bulk genome replication provides the normalisation values for steps 61|(A) (viii) and 61| (B) (ii).

Figure 4: Library preparation

Example data from Tapestation fragment analysis for key steps during the library making procedure (steps 24-46). Lane names in a) indicate the stage after which the sample was taken for Tapestation analysis. Intensities are scaled to normalise maxima across the samples; the purple and green lines indicate the upper (1.5 kb) and lower (25 bp) markers respectively. Note primer bands in Adaptor ligation and PCR lanes (marked *) which are removed by clean up steps. Intensity plot in b) shows the size distribution for a representative final library (as second clean up lane in a), and highlighted in green is the region for which average size is calculated and used for quantification purposes (step 51|).

Figure 5: Anticipated results

Anticipated results and graphs generated at key steps through the Repliscope R package. a) Typical reads per bin distribution of a non-replicating sample (step 61|(A) (iii) or 61|(B) (ii)), after removing values that are smaller than 0.25*median. Median is shown as a pink line across. Most bins have similar values and,

therefore, bins containing fewer reads are detected as outliers (highlighted in grey) using the interquartile range method. Usually, there is no need to remove them. b) Typical reads per bin distribution of a replicating sample (step 61|(A) (v) or 61|(B) (ii)) after applying 0.25*median filtering. A wider distribution of reads than in non-replicating samples is expected. c) ChrVIII sequencing coverage of the sample in a) (step 61|(A) (iv) or 61|(B) (ii)). Data clustering close to the genome-wide median (pink line) is indicative of a good non-replicating sample. d) ChrVIII sequencing coverage of the sample in b) (step 61|(A) (v) or 61|(B) (ii)). Peaks and valleys are indicative of a good replicative sample. e) Sequencing depth-adjusted ratio distribution centres around 1 (step 61|(A) (vii) or 61|(B) (ii)). Region 1-2 highlighted green. f) A typical sort-seq sample with ratio values auto-normalised to fit a biologically meaningful scale (step 61|(A) (viii) or 61|(B) (ii)). Region 1-2 highlighted green. g) Sort-seq replication profile of *S. cerevisiae* chromosome VIII (blue points) with fitted spline smoothing (grey line). h) Sync-seq replication profile of *S. cerevisiae* chromosome XI. i) MFA-seq replication profile of *H. volcanii* chromosome. g,h,i) Steps 61|(A) (ix)-(x) or 61|(B) (ii).

Figure 6: Chromosomal aberrations

Two examples of samples showing chromosomal aberrations. Affected chromosomes are labelled in red. a,c) Evidence of chrXI and chrXII aneuploidy from a sort-seq experiment on a diploid *S. cerevisiae* strain. Data plotted as reads per bin, as (a) boxplot view (step 61|(A) (iii) or 61|(B) (ii)) and (c) whole chromosome view (step 61|(A) (iv) or 61|(B) (ii)). The region with missing data on chrXII is the location of rDNA repeat array. b,d) Evidence of segmental duplication of a region of chrIV. Data plotted as reads per bin, as (b) boxplot view and (d) whole chromosome view (step 61|(A) (v) or 61|(B) (ii)). Pink line across shows whole genome median count. Grey rectangle in (b) highlights outliers using the interquartile range method.

Supplementary information
Supplementary Note: Shiny App walkthrough.

Data availability
All data generated or analysed during this study are publicly available from the NCBI GEO functional genomics data repository with the accession numbers listed below:

Figure 5: GSE42243, GSE48212

Figure 6: GSE135178

Code availability
The custom bash script required for the analysis, as well as the script to download and analyse example data, are available from GitHub (https://github.com/DNAReplicationLab/localMapper/). The R package Repliscope described here is available from CRAN (https://cran.r-project.org/web/packages/Repliscope/) and GitHub (https://github.com/DNAReplicationLab/Repliscope/). We have also provided an official Ubuntu Desktop 18.04 LTS installation disk image with all the software required for the analysis (https://ln1.path.ox.ac.uk/groups/nieduszynski/Replibuntu/Replibuntu-18.04.0-amd64.iso.gz). The code in this manuscript has been peer reviewed.

Author contributions statements
All authors wrote and edited the manuscript.

Acknowledgements We thank Amanda Williams and Becky Busby (Zoology Sequencing Facility) for help with the NextSeq 500, and Michal Maj and Line Eriksen (Sir William Dunn School Flow Cytometry Facility) for their help with FACS.

Competing interests
The authors declare no competing interest.

Related links
Key references using this protocol
Hawkins, M. et al. Nature 503, 544-547 (2013) https://doi.org/10.1038/nature12650
Müller, C. A. et al. Nucleic Acids Research 42, e3 (2014)
https://doi.org/10.1093/nar/gkt878
Müller, C. A. & Nieduszynski, C. A. J Cell Biol 216, 1907-1914 (2017)
https://doi.org/10.1083/jcb.201701061

References

1 Fragkos, M., Ganier, O., Coulombe, P. & Mechali, M. DNA replication origin activation in space and time. *Nat Rev Mol Cell Biol* **16**, 360-374, doi:10.1038/nrm4002 (2015).

2 Newman, T. J., Mamun, M. A., Nieduszynski, C. A. & Blow, J. J. Replisome stall events have shaped the distribution of replication origins in the genomes of yeasts. *Nucleic acids research* **41**, 9705-9718, doi:10.1093/nar/gkt728 (2013).

3 Rhind, N. & Gilbert, D. M. DNA replication timing. *Cold Spring Harb Perspect Biol* **5**, a010132, doi:10.1101/cshperspect.a010132 (2013).

4 Muller, C. A. *et al.* The dynamics of genome replication using deep sequencing. *Nucleic acids research* **42**, e3, doi:10.1093/nar/gkt878 (2014).

5 Yabuki, N., Terashima, H. & Kitada, K. Mapping of early firing origins on a replication profile of budding yeast. *Genes Cells* **7**, 781-789 (2002).

6 Raghuraman, M. K. *et al.* Replication dynamics of the yeast genome. *Science* **294**, 115-121, doi:10.1126/science.294.5540.115 (2001).

7 Peace, J. M., Villwock, S. K., Zeytounian, J. L., Gan, Y. & Aparicio, O. M. Quantitative BrdU immunoprecipitation method demonstrates that Fkh1 and Fkh2 are rate-limiting activators of replication origins that reprogram replication timing in G1 phase. *Genome Res* **26**, 365-375, doi:10.1101/gr.196857.115 (2016).

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

9 Hawkins, M. *et al.* High-resolution replication profiles define the stochastic nature of genome replication initiation and termination. *Cell Rep* **5**, 1132-1141, doi:10.1016/j.celrep.2013.10.014 (2013).

10 Claycomb, J. M. & Orr-Weaver, T. L. Developmental gene amplification: insights into DNA replication and gene expression. *Trends Genet* **21**, 149-162, doi:10.1016/j.tig.2005.01.009 (2005).

11 Koren, A., Soifer, I. & Barkai, N. MRC1-dependent scaling of the budding yeast DNA replication timing program. *Genome research* **20**, 781-790, doi:10.1101/gr.102764.109 (2010).

12 Müller, C. A. & Nieduszynski, C. A. Conservation of replication timing reveals global and local regulation of replication origin activity. *Genome research* **22**, 1953-1962, doi:10.1101/gr.139477.112 (2012).

13 Skovgaard, O., Bak, M., Lobner-Olesen, A. & Tommerup, N. Genome-wide detection of chromosomal rearrangements, indels, and mutations in circular chromosomes by short read sequencing. *Genome research* **21**, 1388-1393, doi:10.1101/gr.117416.110 (2011).

14 Hawkins, M., Malla, S., Blythe, M. J., Nieduszynski, C. A. & Allers, T. Accelerated growth in the absence of DNA replication origins. *Nature* **503**, 544-547, doi:10.1038/nature12650 (2013).

15 Samson, R. Y. *et al.* Specificity and function of archaeal DNA replication initiator proteins. *Cell Reports* **3**, 485-496, doi:10.1016/j.celrep.2013.01.002 (2013).

16 Agier, N. *et al.* The evolution of the temporal program of genome replication. *Nat Commun* **9**, 2199, doi:10.1038/s41467-018-04628-4 (2018).

17 Koren, A. *et al.* Genetic variation in human DNA replication timing. *Cell* **159**, 1015-1026, doi:10.1016/j.cell.2014.10.025 (2014).

18 Batrakou, D. G., Heron, E. D. & Nieduszynski, C. A. Rapid high-resolution measurement of DNA replication timing by droplet digital PCR. *Nucleic Acids Res* **46**, e112, doi:10.1093/nar/gky590 (2018).
19 Shen, Y. *et al.* Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. *Science* **355**, doi:10.1126/science.aaf4791 (2017).
20 Annaluru, N. *et al.* Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**, 55-58, doi:10.1126/science.1249252 (2014).
21 Muller, C. A. & Nieduszynski, C. A. DNA replication timing influences gene expression level. *J Cell Biol* **216**, 1907-1914, doi:10.1083/jcb.201701061 (2017).
22 Natsume, T. *et al.* Kinetochores coordinate pericentromeric cohesion and early DNA replication by Cdc7-Dbf4 kinase recruitment. *Molecular cell* **50**, 661-674, doi:10.1016/j.molcel.2013.05.011 (2013).
23 Kedziora, S. *et al.* Rif1 acts through Protein Phosphatase 1 but independent of replication timing to suppress telomere extension in budding yeast. *Nucleic Acids Res* **46**, 3993-4003, doi:10.1093/nar/gky132 (2018).
24 Foss, E. J. *et al.* SIR2 suppresses replication gaps and genome instability by balancing replication between repetitive and unique sequences. *Proc Natl Acad Sci U S A* **114**, 552-557, doi:10.1073/pnas.1614781114 (2017).
25 Voichek, Y., Bar-Ziv, R. & Barkai, N. Expression homeostasis during DNA replication. *Science* **351**, 1087-1090, doi:10.1126/science.aad1162 (2016).
26 Rudolph, C. J., Upton, A. L., Stockum, A., Nieduszynski, C. A. & Lloyd, R. G. Avoiding chromosome pathology when replication forks collide. *Nature* **500**, 608-611, doi:10.1038/nature12312 (2013).
27 Tanaka, S., Nakato, R., Katou, Y., Shirahige, K. & Araki, H. Origin association of Sld3, Sld7, and Cdc45 proteins is a key step for determination of origin-firing timing. *Current biology : CB* **21**, 2055-2063, doi:10.1016/j.cub.2011.11.038 (2011).
28 Green, B. M., Finn, K. J. & Li, J. J. Loss of DNA replication control is a potent inducer of gene amplification. *Science* **329**, 943-946, doi:10.1126/science.1190966 (2010).
29 Tanny, R. E., MacAlpine, D. M., Blitzblau, H. G. & Bell, S. P. Genome-wide analysis of re-replication reveals inhibitory controls that target multiple stages of replication initiation. *Mol Biol Cell* **17**, 2415-2423 (2006).
30 Kramara, J., Osia, B. & Malkova, A. Break-Induced Replication: The Where, The Why, and The How. *Trends Genet* **34**, 518-531, doi:10.1016/j.tig.2018.04.002 (2018).
31 Alvino, G. M. *et al.* Replication in hydroxyurea: it's a matter of time. *Mol Cell Biol* **27**, 6396-6406, doi:10.1128/MCB.00719-07 (2007).
32 Macheret, M. & Halazonetis, T. D. Monitoring early S-phase origin firing and replication fork movement by sequencing nascent DNA from synchronized cells. *Nat Protoc* **14**, 51-67, doi:10.1038/s41596-018-0081-y (2019).
33 Macheret, M. & Halazonetis, T. D. Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress. *Nature* **555**, 112-116, doi:10.1038/nature25507 (2018).
34 Marchal, C. *et al.* Genome-wide analysis of replication timing by next-generation sequencing with E/L Repli-seq. *Nat Protoc* **13**, 819-839, doi:10.1038/nprot.2017.148 (2018).
35 Siow, C. C., Nieduszynska, S. R., Muller, C. A. & Nieduszynski, C. A. OriDB, the DNA replication origin database updated and extended. *Nucleic acids research* **40**, D682-686, doi:10.1093/nar/gkr1091 (2012).

36 Prioleau, M. N. & MacAlpine, D. M. DNA replication origins-where do we begin? *Genes Dev* **30**, 1683-1697, doi:10.1101/gad.285114.116 (2016).
37 Jinks-Robertson, S. & Klein, H. L. Ribonucleotides in DNA: hidden in plain sight. *Nat Struct Mol Biol* **22**, 176-178, doi:10.1038/nsmb.2981 (2015).
38 Petryk, N. *et al.* Replication landscape of the human genome. *Nat Commun* **7**, 10208, doi:10.1038/ncomms10208 (2016).
39 McGuffee, S. R., Smith, D. J. & Whitehouse, I. Quantitative, genome-wide analysis of eukaryotic replication initiation and termination. *Molecular cell* **50**, 123-135, doi:10.1016/j.molcel.2013.03.004 (2013).
40 Feng, W. *et al.* Genomic mapping of single-stranded DNA in hydroxyurea-challenged yeasts identifies origins of replication. *Nat Cell Biol* **8**, 148-155 (2006).
41 Tiengwe, C. *et al.* Genome-wide analysis reveals extensive functional interaction between DNA replication initiation and transcription in the genome of Trypanosoma brucei. *Cell reports* **2**, 185-197, doi:10.1016/j.celrep.2012.06.007 (2012).
42 Donaldson, A. D. & Nieduszynski, C. A. Genome-wide analysis of DNA replication timing in single cells: Yes! We're all individuals. *Genome Biol* **20**, 111, doi:10.1186/s13059-019-1719-y (2019).
43 Muller, C. A. *et al.* Capturing the dynamics of genome replication on individual ultra-long nanopore sequence reads. *Nat Methods* **16**, 429-436, doi:10.1038/s41592-019-0394-y (2019).
44 Haase, S. B. & Reed, S. I. Improved flow cytometric analysis of the budding yeast cell cycle. *Cell Cycle* **1**, 132-136 (2002).
45 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
46 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
47 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
48 Daigaku, Y. *et al.* A global profile of replicative polymerase usage. *Nat Struct Mol Biol* **22**, 192-198, doi:10.1038/nsmb.2962 (2015).
49 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
50 Xu, J. *et al.* Genome-wide identification and characterization of replication origins by deep sequencing. *Genome Biol* **13**, R27, doi:10.1186/gb-2012-13-4-r27 (2012).
51 Heichinger, C., Penkett, C. J., Bahler, J. & Nurse, P. Genome-wide characterization of fission yeast DNA replication origins. *EMBO J* **25**, 5171-5179, doi:10.1038/sj.emboj.7601390 (2006).
52 Agier, N., Romano, O. M., Touzain, F., Cosentino Lagomarsino, M. & Fischer, G. The spatiotemporal program of replication in the genome of Lachancea kluyveri. *Genome Biol Evol* **5**, 370-388, doi:10.1093/gbe/evt014 (2013).
53 Di Rienzi, S. C. *et al.* Maintaining replication origins in the face of genomic change. *Genome Res* **22**, 1940-1952, doi:10.1101/gr.138248.112 (2012).
54 Kim, H. S. Genome-wide function of MCM-BP in Trypanosoma brucei DNA replication and transcription. *Nucleic Acids Res* **47**, 634-647, doi:10.1093/nar/gky1088 (2019).

55 Devlin, R. *et al.* Mapping replication dynamics in Trypanosoma brucei reveals a link with telomere transcription and antigenic variation. *Elife* **5**, doi:10.7554/eLife.12765 (2016).
56 Marques, C. A., Dickens, N. J., Paape, D., Campbell, S. J. & McCulloch, R. Genome-wide mapping reveals single-origin chromosome replication in Leishmania, a eukaryotic microbe. *Genome Biol* **16**, 230, doi:10.1186/s13059-015-0788-9 (2015).
57 Lundgren, M., Andersson, A., Chen, L., Nilsson, P. & Bernander, R. Three replication origins in Sulfolobus species: synchronous initiation of chromosome replication and asynchronous termination. *Proc Natl Acad Sci U S A* **101**, 7046-7051, doi:10.1073/pnas.0400656101 (2004).
58 Woodfine, K. *et al.* Replication timing of the human genome. *Hum Mol Genet* **13**, 191-202, doi:10.1093/hmg/ddh016 (2004).
59 de Moura, A. P., Retkute, R., Hawkins, M. & Nieduszynski, C. A. Mathematical modelling of whole chromosome replication. *Nucleic Acids Res* **38**, 5623-5633, doi:10.1093/nar/gkq343 (2010).

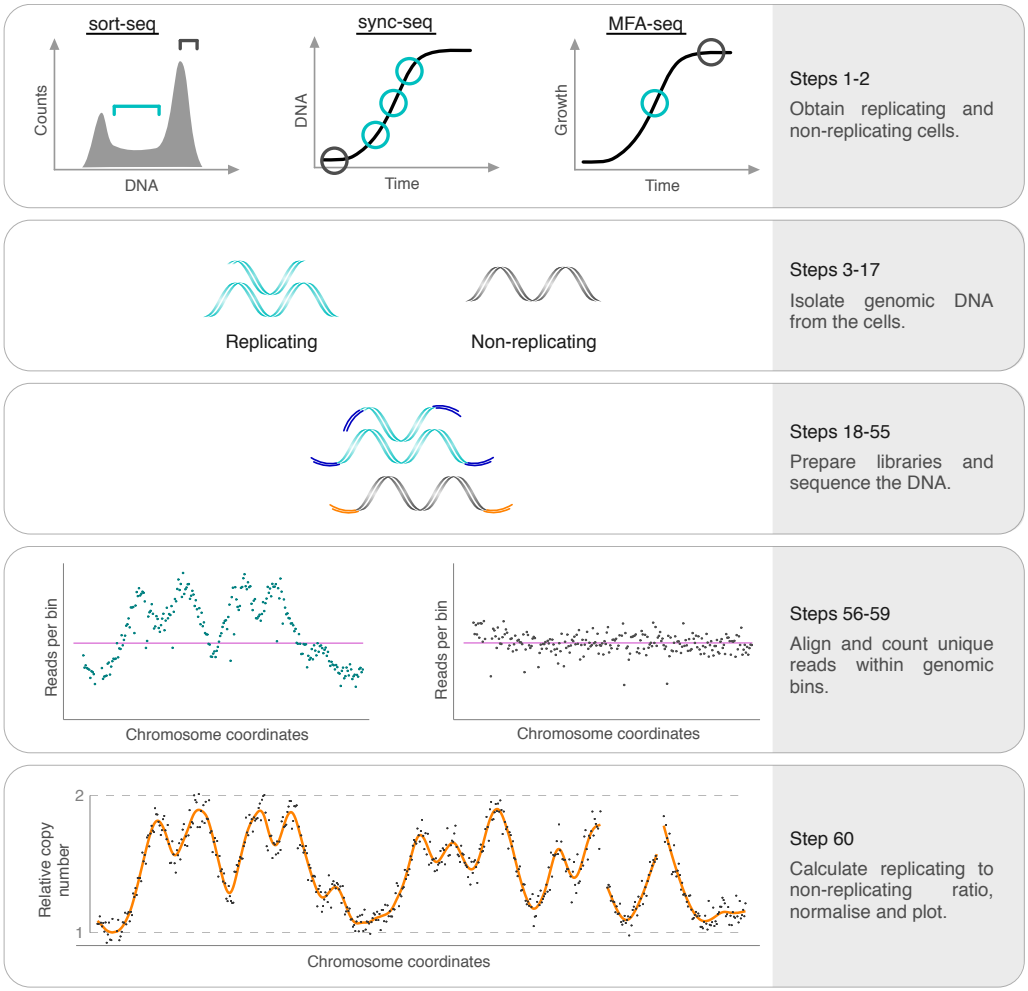

sort-seq
Counts
DNA
sync-seq
DNA
Time
MFA-seq
Growth
Time
Steps 1-2
Obtain replicating and non-replicating cells.
Replicating
Non-replicating
Steps 3-17
Isolate genomic DNA from the cells.
Steps 18-55
Prepare libraries and sequence the DNA.
Reads per bin
Chromosome coordinates
Reads per bin
Chromosome coordinates
Steps 56-59
Align and count unique reads within genomic bins.
Relative copy number
2
1
Chromosome coordinates
Step 60
Calculate replicating to non-replicating ratio, normalise and plot.

Figure 1

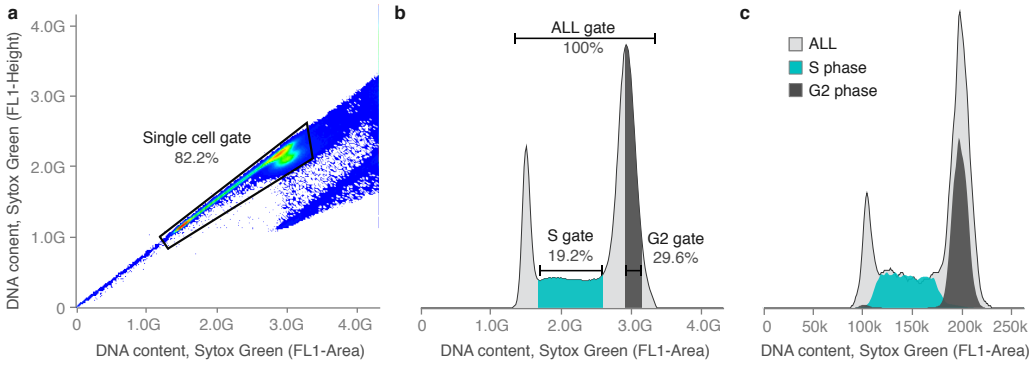

a
b
c
DNA conent, Sytox Green (FL1-Height)
4.0G
3.0G
2.0G
1.0G
0
Single cell gate
82.2%
0
1.0G
2.0G
3.0G
4.0G
DNA content, Sytox Green (FL1-Area)
ALL gate
100%
S gate
19.2%
G2 gate
29.6%
0
1.0G
2.0G
3.0G
4.0G
DNA content, Sytox Green (FL1-Area)
ALL
S phase
G2 phase
0
50k
100k
150k
200k
250k
DNA content, Sytox Green (FL1-Area)

Figure 2

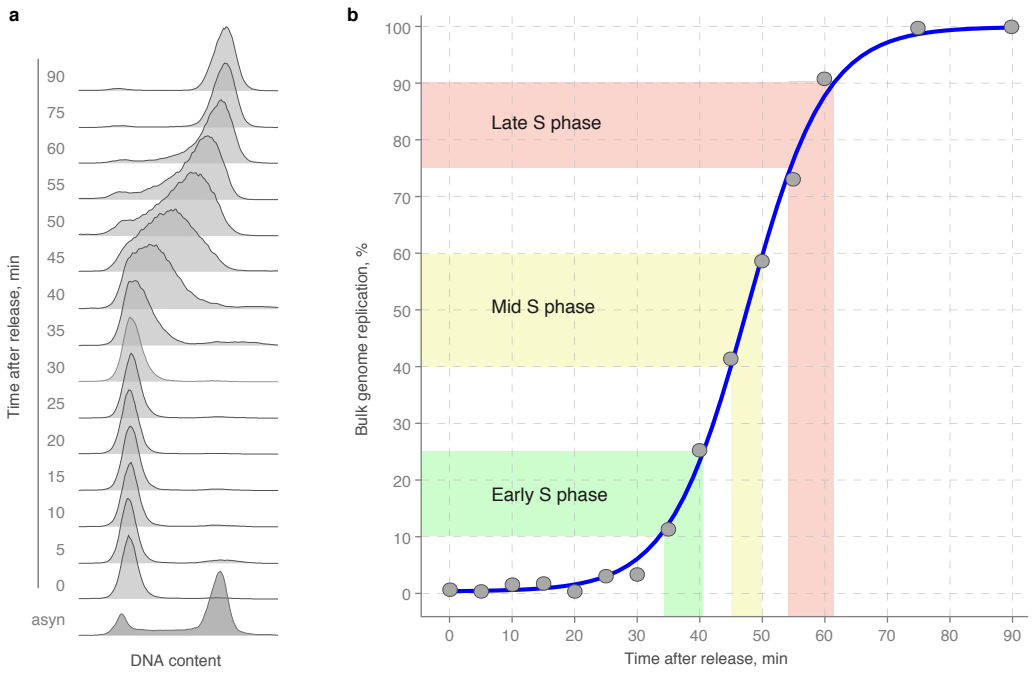

a
Time after release, min
90
75
60
55
50
45
40
35
30
25
20
15
10
5
0
asyn
DNA content
b
Bulk genome replication, %
Late S phase
Mid S phase
Early S phase
Time after release, min

Figure 3

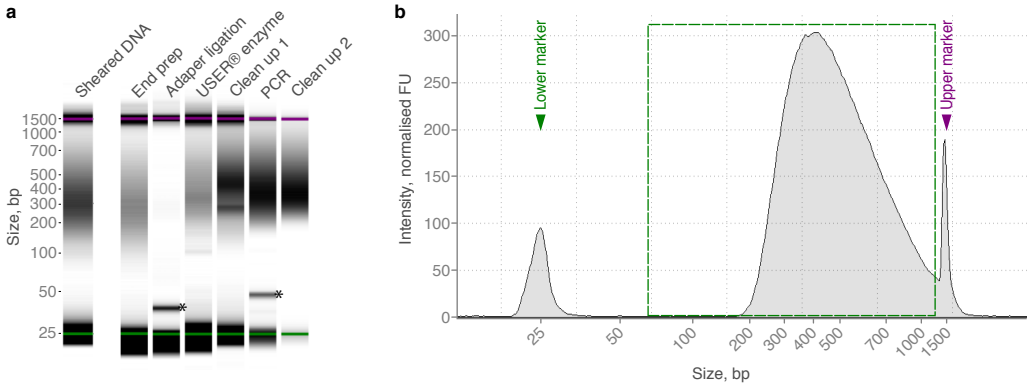

a
Sheared DNA
End prep
Adaper ligation
USER® enzyme
Clean up 1
PCR
Clean up 2
Size, bp
1500
1000
700
500
400
300
200
100
50
25
b
Lower marker
Upper marker
Intensity, normalised FU
300
250
200
150
100
50
0
25
50
100
200
300
400
500
700
1000
1500
Size, bp

Figure 4

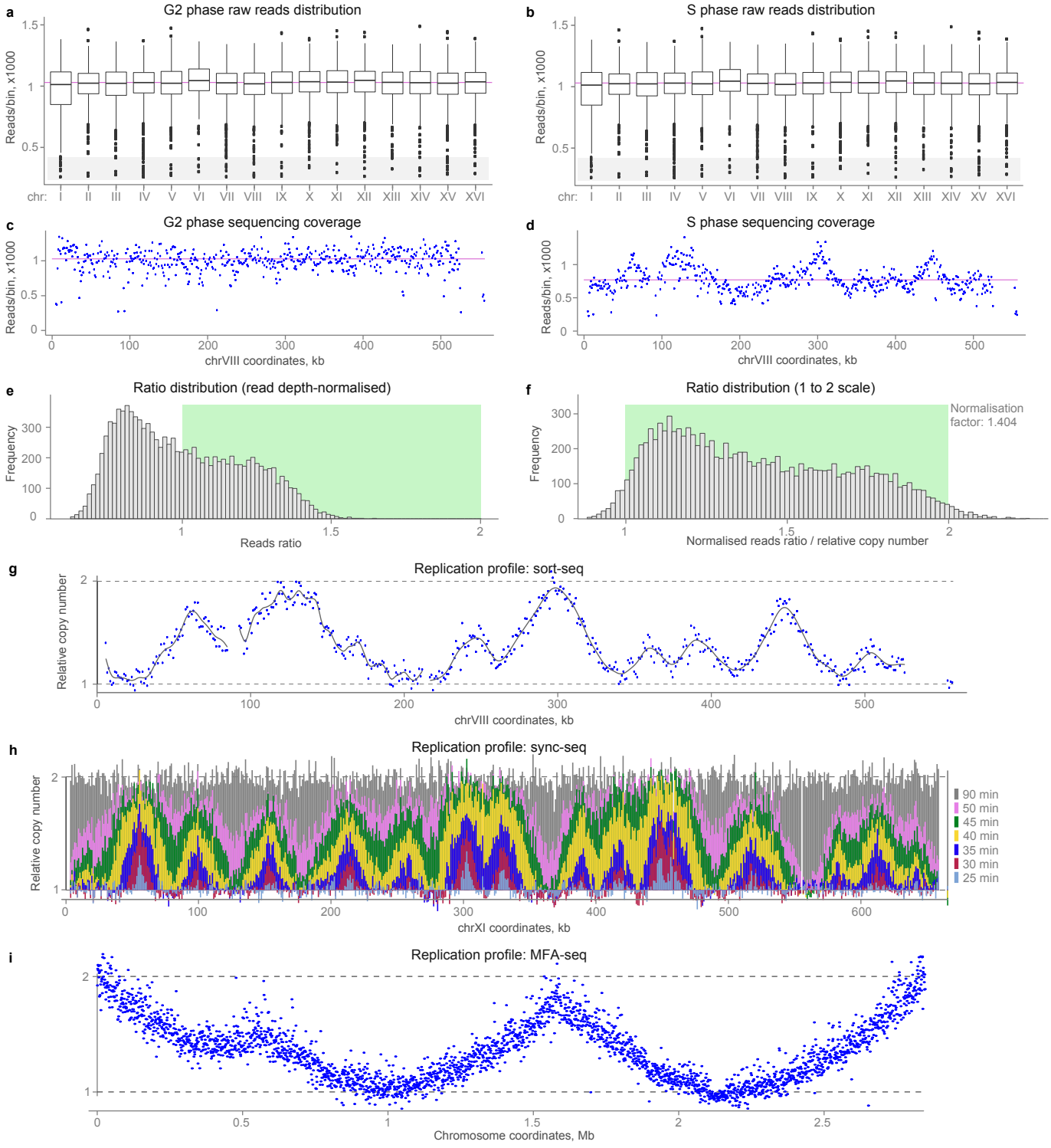

a
G2 phase raw reads distribution
Reads/bin, x1000
chr: I II III IV V VI VII VIII IX X XI XII XIII XIV XV XVI
b
S phase raw reads distribution
c
G2 phase sequencing coverage
chrVIII coordinates, kb
d
S phase sequencing coverage
e
Ratio distribution (read depth-normalised)
Frequency
Reads ratio
f
Ratio distribution (1 to 2 scale)
Normalisation factor: 1.404
Normalised reads ratio / relative copy number
g
Replication profile: sort-seq
Relative copy number
h
Replication profile: sync-seq
chrXI coordinates, kb
90 min
50 min
45 min
40 min
35 min
30 min
25 min
i
Replication profile: MFA-seq
Chromosome coordinates, Mb

Figure 5

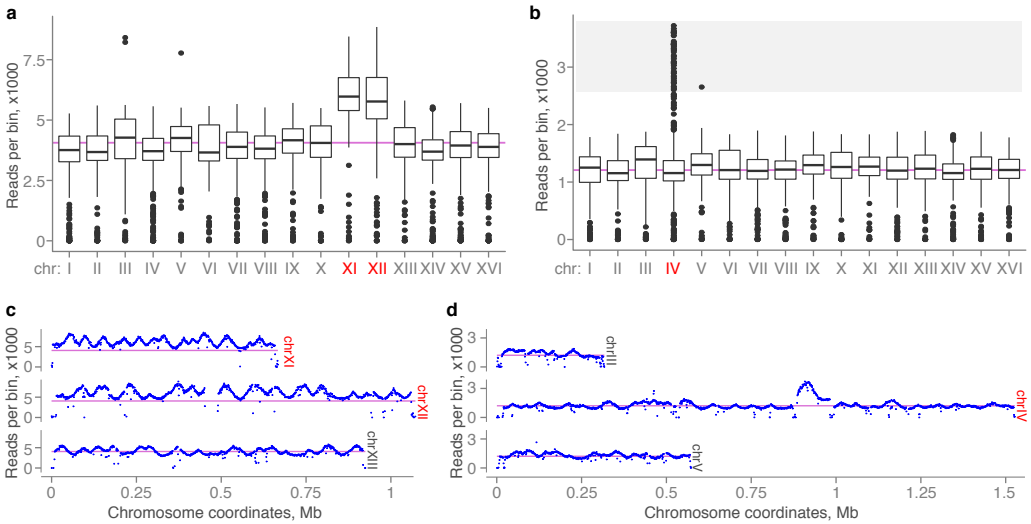
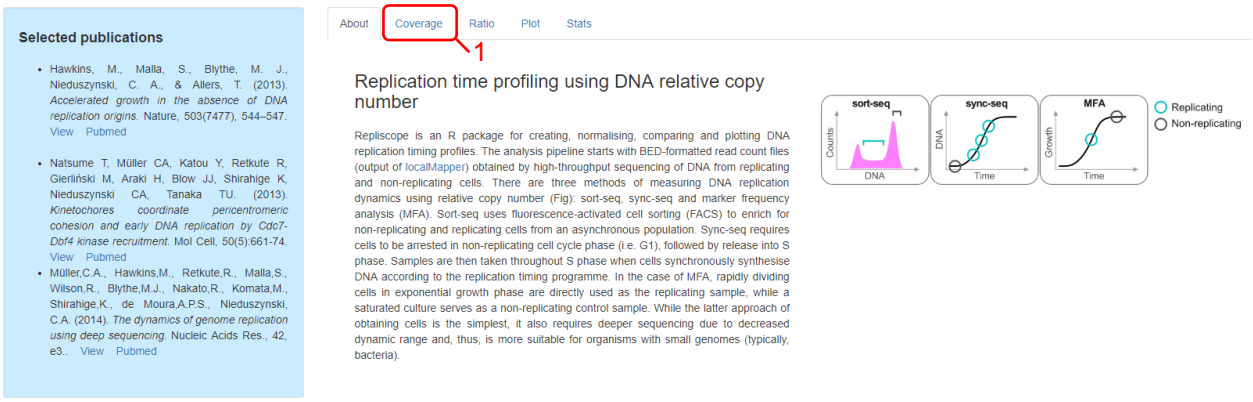

a
b
c
d
Reads per bin, x1000
chr: I II III IV V VI VII VIII IX X XI XII XIII XIV XV XVI
chrXI
chrXII
chrXIII
chrIII
chrV
Chromosome coordinates, Mb

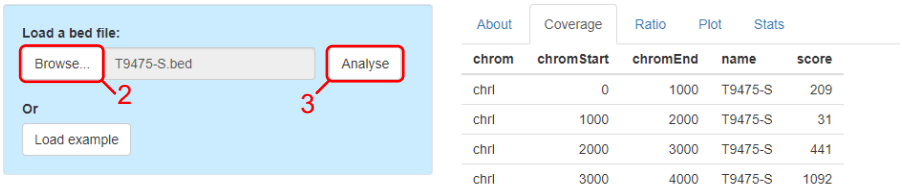
Figure 6

Supplementary Note: Shiny App walkthrough

a

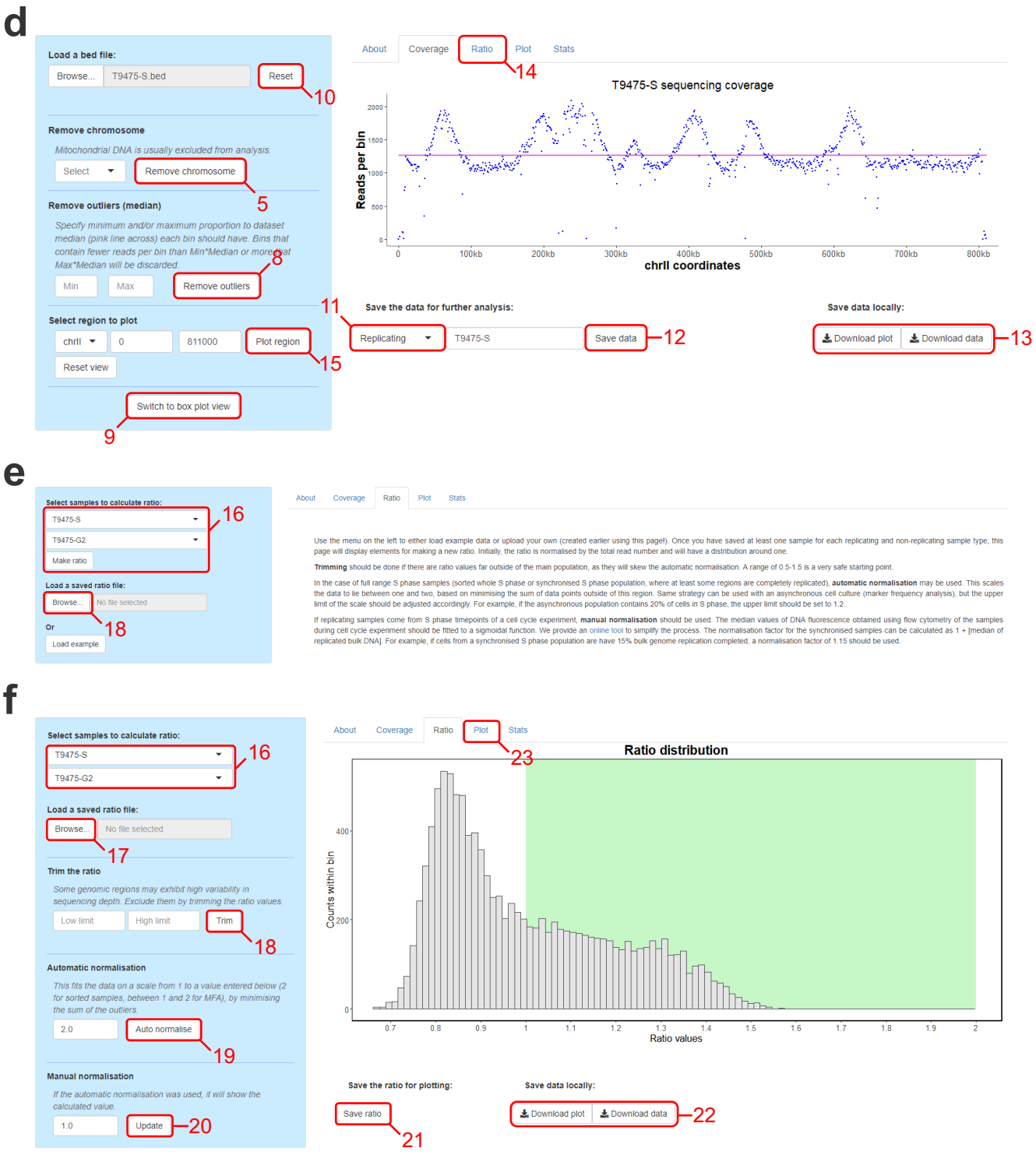


b



c

1. Start by clicking on the Coverage tab.
2. Load a bed file into the app using the "Browse..." button.
3. Once the bed file is loaded, a few lines of its content are displayed on the right - make sure it looks sensible. To initiate analysis, click the "Analyse" button.
4. Examine the boxplot for outliers. Hover over the points to display their properties in the box below.
5. Data for mitochondria can be deleted using the "Remove chromosome" button.
6. A low number of outstanding outliers may be removed using the "Remove max values" button.
7. [**Use only with very noisy data**] To remove all outliers highlighted in grey on the boxplot, the interquartile range (IQR) method may be used. Any data points that fall below Q1-3*IQR or above Q3+3*IQR are removed. Remove mitochondrial data and any outstanding outliers before using this method. **Do not use more than once.**
8. To remove outliers based on the median, provide a minimum and/or maximum threshold. Typically, it is a good idea to remove bins containing very low read numbers as they introduce noise into the ratio values later. We use 0.25 as a lower (Min) threshold.
9. Use this button to switch between boxplot and scatterplot view.
10. The "Reset" button may be used to reload the original bed file in case erroneous data manipulation has happened.
11. Mark the current sample as replicating or non-replicating.
12. Once the data has been pre-processed, it may be saved for further analysis within the app (the name may be changed).
13. The current plot and data may be saved locally.
14. Once at least one replicating and one non-replicating sample is saved for further analysis, proceed to the Ratio tab.



15. Region plotting controls are available in scatterplot view.

16. After selecting a replicating and a non-replicating sample, clicking the "Make ratio" button will calculate the ratio between the two samples. A histogram of the calculated values will be displayed on the right. Initially, the values are only normalised by sequencing depth and will center around 1.

17. A previously saved ratio table may be loaded into the app.

18. Ratio value outliers may be trimmed. 0.5-1.6 is a typical range of ratio values before further normalisation (for sort-seq or mid-S phase sync-seq samples).

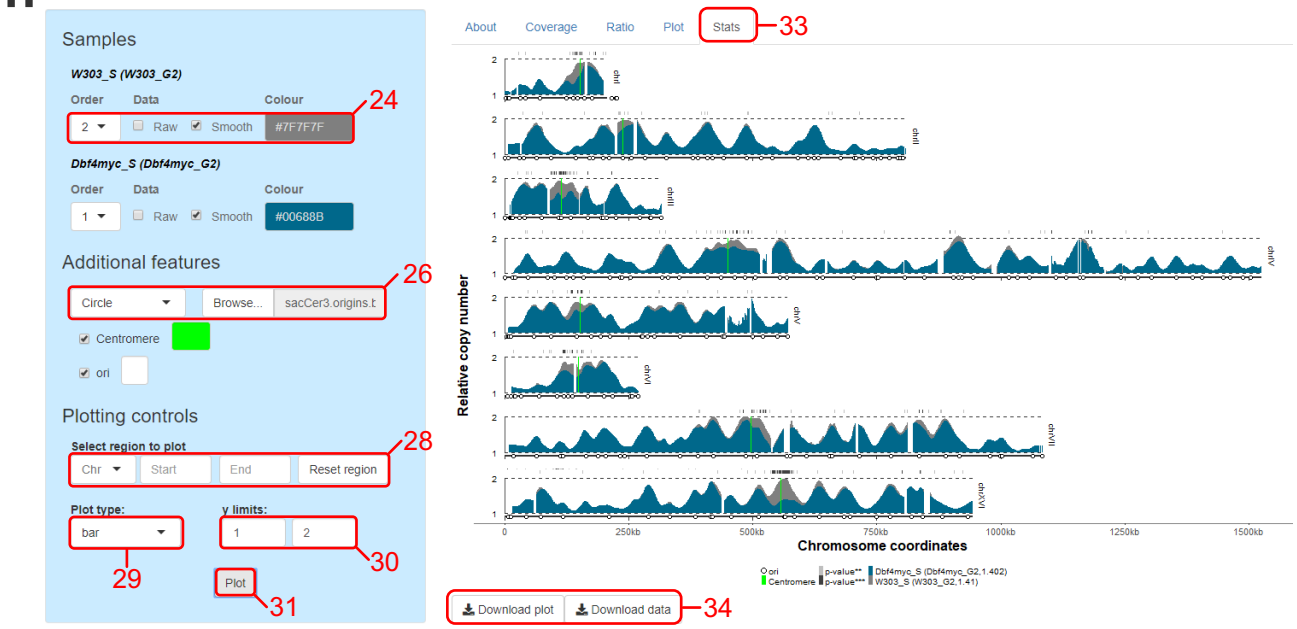
19. Best fitting normalisation may be used if the biological range of the relative copy number is known. Typically, a range of 1 to 2 is used for sort-seq or mid-S phase sync-seq samples and 1 to [1 + the proportion of cells in S phase] for MFA samples. Adjust the upper limit accordingly.

20. Manual normalisation may be used if means or medians of bulk DNA replication of sync-seq samples were calculated based on flow cytometry data. Use the online tool (https://dnareplab.shinyapps.io/fitsigmoid/) to fit a sigmoidal function into flow cytometry data.

21. Save the normalised ratio for plotting and further analysis in the app.

22. Download the plot and/or data locally.

23. Once at least one ratio is saved, it may be plotted in the Plot tab.



24. Individual ratio controls allow changing plotting order, whether to plot raw and/or smooth data, and colour of the sample. Smoothed data is available once smoothing has been applied.
25. A previously saved ratio may be loaded into the app.
26. Additional features may be plotted from bed files as different shapes/lines. The first value in the name field of the bed file will be used to name the feature on the plot.
27. All loaded ratios may be smoothed using cubic spline. Group size and split values control smoothed line grouping: the minimum number of data points in a group (at least 4) and number of bins with missing values to initiate a new group.
28. Region plotting controls.
29. Plot type - scatterplot, polygon or barplot.
30. y axis limits - data outside of the limits will not be displayed.
31. Refresh the plot on the right with current parameters.
32. Save plot/data locally.
33. To compare two replication profiles, use the Stats tab (at least two ratios must be saved in the app).
34. Save plot/data locally.