Genome-wide Nucleotide-Resolution Mapping of DNA Replication Patterns, Single-Strand Breaks, and Lesions by GLOE-Seq

Graphical Abstract

Highlights

- GLOE-Seq detects 3’-OH ends with nucleotide resolution in purified genomic DNA
- GLOE-Seq maps single-strand breaks, lesions, and replication and repair intermediates
- GLOE-Seq reveals insight into the use of ligases 1 and 3 in human cells
- GLOE-Seq detects asymmetries in spontaneous nicks in yeast and human chromatin

Authors

Annie M. Sriramachandran, Giuseppe Petrosino, María Méndez-Lago, Axel J. Schäfer, Liliana S. Batista-Nascimento, Nicola Zilio, Helle D. Ulrich

Correspondence
n.zilio@imb-mainz.de (N.Z.), h.ulrich@imb-mainz.de (H.D.U.)

In Brief

We present a method for genome-wide, nucleotide-resolution mapping of DNA single-strand breaks in purified genomic DNA based on capture of 3’-OH ends followed by sequencing (GLOE-Seq). We validate the method and demonstrate its applicability to mapping of human and yeast Okazaki fragments, spontaneous single-strand breaks, and various other DNA lesions.
DNA single-strand breaks (SSBs) are among the most common lesions in the genome, arising spontaneously and as intermediates of many DNA transactions. Nevertheless, in contrast to double-strand breaks (DSBs), their distribution in the genome has hardly been addressed in a meaningful way. We now present a technique based on genome-wide ligation of 3’-OH ends followed by sequencing (GLOE-Seq) and an associated computational pipeline designed for capturing SSBs but versatile enough to be applied to any lesion convertible into a free 3’-OH terminus. We demonstrate its applicability to mapping of Okazaki fragments without prior size selection and provide insight into the relative contributions of DNA ligase 1 and ligase 3 to Okazaki fragment maturation in human cells. In addition, our analysis reveals biases and asymmetries in the distribution of spontaneous SSBs in yeast and human chromatin, distinct from the patterns of DSBs.

INTRODUCTION

Our understanding of the DNA damage response and the protective measures cells employ to resist genomic instability and malignant transformation relies to a large extent on analytical methods to detect and quantify DNA lesions in chromatin. Next-generation sequencing (NGS) tools for mapping DNA damage have transformed genome stability research by providing insight into the genome-wide distribution and dynamics of lesions at unprecedented resolution. A growing number of NGS protocols is already available for analysis of base lesions such as abasic sites, oxidative or methylation damage, and incorporated ribonucleotides (Clausen et al., 2015; Ding et al., 2015; Hu et al., 2017; Liu et al., 2019; Mao et al., 2017; Poetsch et al., 2018; Reijns et al., 2015; Wu et al., 2018) and for DNA double-strand breaks (DSBs) (Barnello et al., 2014; Canela et al., 2016; Crosetto et al., 2013; Hoffman et al., 2015; Hu et al., 2016; Lensing et al., 2016; Tsai et al., 2015; Vitelli et al., 2017; Yan et al., 2017). However, even with these established procedures, important information about the 3’ termini of DSBs remains untapped because of the need for end polishing during library preparation. In contrast to DSBs, single-strand breaks (SSBs), which are among the most common lesions and emerge in the genome spontaneously or as important intermediates of DNA replication and repair (Abbotts and Wilson, 2017; Caldecott, 2014), have, until recently, eluded high-resolution mapping by NGS approaches.

We developed a method based on capturing SSBs via genome-wide ligation of 3’-hydroxy (OH) ends followed by sequencing (GLOE-Seq). When applied directly to genomic DNA, this tool is capable of detecting nicks or gaps as well as the 3′ ends of DSBs. Alternatively, pre-digestion of the isolated DNA with a suitable lesion-specific endonuclease allows mapping of a variety of lesions, such as ultraviolet (UV) irradiation-induced pyrimidine dimers, abasic sites, or incorporated ribonucleotides. After assessing the accuracy and sensitivity of GLOE-Seq with in vitro-digested DNA, we benchmarked its performance by comparison with an established method for mapping of base lesions and demonstrated its applicability to physiological DNA damage and repair intermediates introduced in vivo by UV irradiation or an alkylating agent and by a site-specific endonuclease in budding yeast. We then explored the unique feature of GLOE-Seq, its ability to map pre-existing SSBs, by analyzing replication patterns as well as spontaneous breaks and nicks in budding yeast and human cells. We show that GLOE-Seq can accurately map Okazaki fragments without prior size selection, and we detect surprising biases in the distribution of spontaneous strand breaks, non-random and distinct from the pattern observed with DSB-selective methods. Our analysis provides insight into the relative contributions of DNA ligase 1 and ligase 3 to human Okazaki fragment maturation and validates GLOE-Seq as a versatile method for genome-wide mapping of a range of DNA lesions that promises to shed light onto the still poorly understood characteristics of SSBs in the genome.

SUMMARY

DNA single-strand breaks (SSBs) are among the most common lesions in the genome, arising spontaneously and as intermediates of many DNA transactions. Nevertheless, in contrast to double-strand breaks (DSBs), their distribution in the genome has hardly been addressed in a meaningful way. We now present a technique based on genome-wide ligation of 3’-OH ends followed by sequencing (GLOE-Seq) and an associated computational pipeline designed for capturing SSBs but versatile enough to be applied to any lesion convertible into a free 3’-OH terminus. We demonstrate its applicability to mapping of Okazaki fragments without prior size selection and provide insight into the relative contributions of DNA ligase 1 and ligase 3 to Okazaki fragment maturation in human cells. In addition, our analysis reveals biases and asymmetries in the distribution of spontaneous SSBs in yeast and human chromatin, distinct from the patterns of DSBs.
nick translation (Baranello et al., 2014) or polymerase tailing for adaptor ligation (Leduc et al., 2011) may obscure the original position of the terminus, reducing resolution. Direct ligation of short single-stranded DNA (ssDNA) fragments can be accomplished by RNA ligase and has been used for sequencing highly degraded DNA; for example, upon isolation from ancient samples (Gansauge and Meyer, 2013). Adaptor ligation was further improved by means of a splinter oligonucleotide harboring a stretch of random nucleotides that allows use of T4 DNA ligase (Gansauge et al., 2017). We explored whether an analogous approach could be applied to heat-denatured, intact DNA for genome-wide mapping of nicks and breaks with nucleotide resolution. In the optimized protocol (Figures 1A and S1), fragmentation of genomic DNA is delayed until after ligation of a biotinylated adaptor. To this end, all steps up to ligation, including cell lysis, are carried out in agarose plugs when applying GLOE-Seq to mammalian DNA. In this manner, SSBs are captured with minimal background. By applying an appropriate enzymatic treatment to the purified genomic DNA before thermal denaturation, nicks can also be generated at relevant positions to mark various types of base damage. For example, treatment with apurinic endonuclease (APE1) would introduce SSBs adjacent to abasic sites, whereas RNase H treatment would mark the positions of ribonucleotides in the genome. Thus, GLOE-Seq should be applicable to analysis of not only pre-existing SSBs but also any base lesion or modification for which a selective endonuclease is available. By developing a computational pipeline that allows mapping of $3\prime$ termini or base lesions, we assembled a ready-to-use platform for downstream data analysis.

RESULTS

Validation of the GLOE-Seq Protocol

The resolution, specificity, and sensitivity of GLOE-Seq were addressed in proof-of-principle experiments using purified S. cerevisiae genomic DNA. GLOE-Seq libraries were sequenced at a
depth of ~3 million reads in two replicates. To facilitate data analysis, we developed an easy-to-use, modular, and versatile computational pipeline called GLOE-Pipe. It detects, annotates, and visualizes strand breaks by assigning each uniquely mapping read to the corresponding original 3' terminus. Direct inspection of reads from samples digested in vitro with a restriction endonuclease, BsrDI, revealed precise assignment of 3' termini to the expected sequence (Figure 1B). Because BsrDI cleaves asymmetrically adjacent to its recognition sequence, digestion of genomic DNA yields two populations of termini with distinct complexities. Nevertheless, both populations exhibited an almost identical read distribution, confirming unbiased detection of termini (Figure 1C). Digestion with the analogous nicking enzyme Nb.BsrDI resulted in the expected strand-specific signals (Figure 1D), corresponding to more than 60% of total reads. Automatic peak-calling detected more than 90% of the predicted 6,271 sites in the yeast genome (Figure 1E, left). Among the undetected sites, more than 60% were either absent in our strain or resided in regions that did not sequence efficiently when randomly sheared DNA was used to prepare an equivalent library (Figure 1E, center). Moreover, Nb.BsrDI sites situated less than 100 nt downstream (3') of a neighboring nick were detected with poor efficiency, presumably because of size selection during library preparation (Figure 1F). Peak calling also resulted in a small number of unexpected signals (Figure 1E, left). More than 50% of these were detected within 5 nt of a predicted Nb.BsrDI site (Figure 1E, right), indicating that they likely resulted from imprecise nicking or erosion of the termini rather than sequencing or computational artifacts.

The sensitivity of the method was assessed by diluting Nb.BsrDI-treated with untreated DNA at defined ratios (Figures S2A–S2C). All samples had been digested with NsoI to generate a set of 80 defined peaks for standardization. As shown in Figure S2D, the number of detected Nb.BsrDI signals compared well with the numbers expected based on the dilution. At a sequencing depth of ~3 million reads, 1,000-fold dilution still resulted in a reproducible signal sufficient to provide good coverage of Nb.BsrDI sites (Figure 1G) and excellent correlation between replicates ($r^2 > 0.86$; Figure S2A). A higher sequencing depth could potentially provide additional sensitivity. These results confirm that GLOE-Seq is capable of selectively detecting even relatively rare events against the background of spontaneous SSBs in the genome.

For an assessment of GLOE-Seq in a physiological setting, we compared the break pattern in intact genomic yeast DNA with a sample that was heavily fragmented by sonication, representing a close-to-random distribution of breaks. In addition to...
excluding repetitive elements, reads were assigned with high stringency by disallowing any mismatches. Two salient features emerged from this analysis: a striking enrichment of break signals adjacent to centromere sequences (Figure 3A) and a number of well-defined peaks close to some chromosome ends, some of them displaying a high degree of strand bias (Figures 3B and S4). Although the centromere-associated breaks could result from topological stress or topoisomerase activity related to chromosome segregation, the subtelomeric patterns were unexpected and deserve further investigation.

To generate site-specific break signals in live cells, we used a strain carrying a galactose-inducible allele of the homothallic switching (HO) endonuclease (Lee et al., 1998). Induction for 1 h gave rise to prominent signals at the expected sequence on both strands (Figure 3C). Moreover, a population of 3’ termini was clearly detectable, revealing loss of a few nucleotides from each terminus. Thus, in contrast to DSB-selective methods, GLOE-Seq is capable of visualizing 3’ overhangs of DSBs with high precision.
Genome-wide Mapping of Base Lesions and Repair Intermediates in Budding Yeast

The potential to map base lesions by GLOE-Seq was explored by UV irradiation of live yeast and treatment of isolated genomic DNA with T4 endonuclease V, followed by APE1 endonuclease, to convert UV lesions to 3'-OH termini before adaptor ligation. As shown in Figures 3D and S5A, this treatment revealed a dose-dependent increase in the relative frequencies of break signals adjacent to pyrimidine dimers, with T-T representing the most frequent lesion, followed by T-C, C-T, and C-C.

In addition, we used exposure to the alkylating agent methyl methanesulfonate (MMS) to determine whether GLOE-Seq can detect base lesions and the resulting repair intermediates in a single experiment. MMS predominantly generates N7-methylmethanesulfonate (MMS) to determine whether GLOE-Seq can detect base lesions and the resulting repair intermediates. Consequently, GLOE-Seq analysis of AAG/APE1-treated genomic DNA showed strong MMS-induced enrichment of SSBs adjacent to G at the expense of C and T, likely representing m7dG (Figure 3E). This enrichment decreased in samples taken during a recovery period, suggesting in vivo repair activity. Deletion of the two major apurinic endonuclease genes, APN1 and APN2, strongly interfered with recovery (Figure 3E) and activated damage signaling (Figure S6B), indicating that BER was indeed responsible for removal of the damaged bases. GLOE-Seq samples without AAG/APE1 pre-treatment showed a similar skew of the nucleotide distribution toward G (Figure 3F). Importantly, this MMS-induced imbalance was almost absent in the apn1Δ apn2Δ mutant, consistent with failure of this strain to initiate BER. These results show that GLOE-Seq can detect biologically relevant base lesions as well as DNA repair intermediates with nucleotide precision.

Analysis of DNA Replication Patterns in Budding Yeast

An important physiological source of nicks are the termini of Okazaki fragments in replicating cells. Usually, these are quickly ligated, but depletion of the replicative DNA ligase Cdc9 has allowed their mapping in budding yeast (Smith and Whitehouse, 2012). To accomplish this, fragments of the expected size of around 200 nt had to be isolated for sequencing. We wanted to find out whether GLOE-Seq would enable us to map Okazaki fragments in an unbiased manner without prior size selection. Genomic DNA was isolated from cdc9AΔ cells harboring a degron-tagged Cdc9 allele (Kubota et al., 2013). GLOE-Seq analysis revealed the expected lagging-strand bias of break signals around replication origins (Figure 4A) and phasing of the reads correlating with nucleosome positioning around transcription start sites (Figure 4B), comparable with the results obtained by sequencing size-selected fragments (Smith and Whitehouse, 2012). Plotting of the replication fork directionality (RFD) index, representing the relative strand bias of the break signals, confirmed that GLOE-Seq is suitable for identifying the positions of replication initiation sites and termination zones in a genome-wide manner under conditions of ligase inhibition (Figure 4C). Intriguingly, RFD analysis of samples from ligase-proficient cells revealed a complementary but less prominent pattern with an excess of nicks in sequences corresponding to the leading rather than the lagging strand (Figure 4C). This pattern likely reflects processing of ribonucleotides naturally incorporated into newly synthesized DNA by replicative polymerases, considering that the leading-strand polymerase ε incorporates four times more...
To further test the efficiency of GLOE-Seq in the context of the human genome, we analyzed the distribution of human Okazaki fragments. In mammalian cells, such an analysis has been achieved by Okazaki fragment sequencing (OK-Seq) (Petryk et al., 2016), which uses large amounts of input material (2–3 × 10⁹ cells) and multi-step enrichment of relevant DNA fragments that involves labeling and affinity purification of nascent DNA in combination with size fractionation. We followed a strategy of inactivating Okazaki fragment ligation analogous to the yeast system by small interfering RNA (siRNA)-mediated depletion of ligase 1 and deletion of the nuclear form of ligase 3 in HCT116 cells (Oh et al., 2014; Figures 6A and S6B). GLOE-Seq analysis revealed a pattern similar to published OK-Seq profiles, although with lower efficiency (Figure 6B). The difference may well result from activation of the DNA damage response in ligase-deficient cells, which would suppress late-origin firing and eventually lead to triggering of dormant origins (Alver et al., 2014; Karhani and Dutta, 2011). Indeed, correlation of replication timing with RFD plots indicated that ligase inactivation predominantly affected the efficiency of late-firing origins. Moreover, accumulation of cells in G2 phase (Figure 6C) and phosphorylation of the checkpoint kinase CHK1 (Figure 6A) confirmed activation of ataxia telangiectasia-related protein (ATR)-dependent damage signaling. In contrast, inactivation of ligase 1 or ligase 3 separately caused at most marginal checkpoint activation and no significant disturbance of the cell cycle. We therefore wanted to find out whether Okazaki fragments would still be detectable under those milder conditions. As shown in Figure 6B, deletion of nuclear ligase 3 did not yield a discernible replication pattern in RFD plots, but depletion of ligase 1 resulted in a defined profile with higher amplitude peaks than upon complete ligase inactivation, resembling the published OK-Seq profiles. As observed in yeast, RFD plots from ligase-proficient HCT116 cells revealed a small but detectable signal bias toward the leading strand, suggesting that ribonucleotide incorporation significantly contributes to spontaneous SSBs in vertebrate cells as well (Figure 6D).

Importantly, these data were produced with 300–400 times fewer cells than the number required for OK-Seq (~700,000 versus 2–3 × 10⁹ cells) (Petryk et al., 2016) and without any size selection. Except for termination zones that show dispersed signals, progressive downsampling of a full dataset from ligase 1-inactivated cells (~310 million reads) revealed marginal loss of resolution in RFD profiles down to ~50 million reads (Figure S6C), showing that GLOE-Seq is a highly sensitive technique. In combination with ligase 1 depletion, it is a practicable and potentially less cumbersome alternative to OK-Seq.

**DISCUSSION**

**Unique Features of GLOE-Seq**

In this study, we present a versatile tool that complements other methods for mapping DNA damage and extends the range of structures accessible to analysis by NGS. The accompanying computational pipeline, GLOE-Pipe, allows straightforward

---

**Figure 5. Effects of Transcription on the Distribution of SSBs and DSBs in Human Cells**

Deviations of break signals from the genome-wide average reveal differences between transcribed and non-transcribed genes. Top: GLOE-Seq signals (representing SSBs and DSBs). Bottom: END-seq signals from Tubbs et al. (2018) (representing only DSBs). Break levels for TSSs, transcription termination sites (TTTs), and 5′ and 3′ exon junctions represent averages of 1-kbp-long sliding windows. Damage levels for 5′ and 3′ UTRs, exons, and introns represent averages of 1-kbp-long non-overlapping bins after concatenation of values (STAR Methods).

ribonucleotides than the lagging-strand polymerase δ (Nick McElhinny et al., 2010).

**Genome-wide Distribution of SSBs Relative to Transcription Units in Human Cells**

To expand the versatility of GLOE-Seq, we optimized the protocol for mammalian cells by performing all steps up to and including adaptor ligation on agarose-embedded nuclei (Figure S6A). This approach minimizes direct handling, and therefore potential breakage, of the very large genomic DNA of higher eukaryotes during extraction. Comparison of break distribution along coding regions using GLOE-Seq with published DSB-specific end sequencing (END-seq) data (Tubbs et al., 2018) from ligase-proficient HTC116 cells (Figure 5) showed that, unlike DSBs, reported to accumulate around the transcription start sites (TSSs) of transcribed genes, SSBs were underrepresented at this location, also in a transcription-dependent manner. Conversely, both SSBs and DSBs locally peaked around the transcription termination sites (TTSs) of transcribed genes. However, the overall enrichment along transcribed genes that is observable for DSBs was much less pronounced for GLOE-Seq signals. Considering that the latter arise from both SSBs and DSBs, the overall depletion of breaks around TSS illustrates the dominance of SSBs compared with DSBs even at sites where DSBs are enriched relative to other regions. At the same time, our data clearly show that the pattern of SSBs significantly differs from that of DSBs along a transcribed gene.
analysis of the data associated with this method. Unlike common procedures for analysis of DSBs, such as breaks labeling, enrichment on streptavidin and next generation sequencing (BLESS) or breaks labeling in situ and sequencing (BLISS) (Crosetto et al., 2013; Yan et al., 2017), END-Seq (Canela et al., 2016), or DSBCapture (Lensing et al., 2016), GLOE-Seq allows the detection of 3’ ends when applied to DSBs. It is flexible enough for application to base lesions and incorporated ribonucleotides. In this mode, it resembles established methods such as EndoSeq (Ding et al., 2015; Reijns et al., 2015) or Click-Code-Seq (Wu et al., 2018) but is more general than tools designed for specific lesions, such as ribonucleotides (Clausen et al., 2015) or abasic sites (Liu et al., 2019; Poetsch et al., 2018). Its most important and distinguishing characteristic, however, is its applicability to pre-existing SSBs. This feature has not only enabled us to map Okazaki fragments without prior size selection but it has also given insight into the overall distribution of spontaneous SSBs in the budding yeast and the mammalian genome, revealing their non-random nature.

Earlier reports of potentially SSB-selective analyses (Baranello et al., 2014; Leduc et al., 2011) have furnished no validation regarding resolution, sensitivity, or possible sequence bias. Although a strategy called damaged DNA immunoprecipitation (dDIP), involving end capture by TUNEL (terminal deoxy-nucleotidyl transferase-mediated biotin-deoxyuridine triphosphate labeling), could, in principle, provide nucleotide resolution for mapping of nicks and DSBs (Leduc et al., 2011), this approach has not been developed to the sequencing stage and cannot distinguish between SSBs and DSBs. An alternative method called single-strand break sequencing (SSB-Seq) (Baranello et al., 2014), based on labeling of SSBs via nick translation followed by immunoprecipitation, has generated low-resolution data for distribution of breaks in etoposide-treated cells but has not been able to yield strand-specific information.

Very recently, a promising method named SSiNGLe (SSB mapping at Nucleotide Genome Level) has been reported (Cao et al., 2019). It resembles GLOE-Seq in its use of 3’-OH ends to map SSBs but employs poly-dA tailing by terminal transferase for capture. Because of this feature, the procedure is particularly suited for single-molecule sequencing without amplification on a Helicos platform, which uses oligo-dT for immobilization of the templates. At the same time, the repetitive signals generated from the tails complicate use of the more common Illumina system. Moreover, fragmentation is achieved by means of micrococcal nuclease digestion at an early stage, producing unligatable 3’-PO4 termini. Unlike SSiNGLe, GLOE-Seq can map such structures by pre-treatment with phosphatase. SSiNGLe has been used to analyze the distribution of SSBs in response to a panel of anti-cancer drugs to follow DNA breakage in the early stages of apoptosis, and it has revealed correlations of the “SSB breakome” with transcribed and regulatory regions, evolutionarily less conserved regions, and topoisomerase IIA cleavage clusters.
GLOE-Seq, as a turnkey combination of molecular and computational pipelines, is an alternative approach we validated as particularly useful for straightforward and effective analysis of genome-wide replication patterns.

**GLOE-Seq Provides Insight into Replication-Associated SSBs in Yeast and Human Cells**

Despite their abundance in the genome, SSBs have, until recently, eluded systematic genome-wide analysis because of a lack of high-resolution methods for their analysis. The GLOE-Seq strategy of selectively capturing 3’-OH ends in thermally denatured but otherwise intact DNA fills this need. GLOE-Seq analysis of undamaged DNA from budding yeast and human cells has already provided us a glimpse into the genome-wide distribution of spontaneous SSBs as ubiquitous but understudied structures. Their non-random patterns highlight the importance of investigating the underlying mechanisms responsible for their formation and processing.

The most striking feature emerging from our analysis of unperturbed yeast and human cells is a clear bias of spontaneous SSBs toward the leading strand following the activity profile of polymerase ε (Figures 4C and 6D). Considering the propensity of this enzyme to incorporate ribonucleotides, the observed RFD profiles suggest that repair intermediates at ribonucleotides are a major source of spontaneous SSBs in yeast and human cells. Comparison of the approximate Okazaki fragment length (~200 nt) with the frequency of ribonucleotide incorporation by polymerase ε (~1 in 1,250 nt) (Nick McElhinny et al., 2010) implies that SSBs should emerge on the lagging strand with a more than 6-fold higher frequency than on the leading strand. In light of the leading strand bias in RFD plots from unperturbed cells, it follows that the nicks derived from processing of ribonucleotides must be significantly more persistent in the genome than unligated Okazaki fragments.

Further insight into Okazaki fragment maturation in human cells comes from our GLOE-Seq analysis under conditions of DNA ligase inactivation. The strong lagging-strand bias in RFD plots upon ligase 1 depletion (Figure 6B) clearly confirms this enzyme as the principal replicative DNA ligase. It also shows that ligase 3 acts in an important back-up function in the absence of ligase 1, maintaining cell proliferation and preventing major activation of the damage response despite significant accumulation of unligated Okazaki fragments. An alternative Okazaki fragment ligation pathway mediated by ligase 3 has been reported in chicken DT40 cells (Arakawa et al., 2012). From our results, we conclude that, in human cells, this pathway proceeds with significantly slower kinetics than the standard ligase 1-mediated reaction. However, the absence of a clear leading-strand bias in the Lig3Δ/Δ::Lig3 cell line suggests a non-negligible contribution of ligase 3 to Okazaki fragment maturation even when ligase 1 is functional.

**Potential Applications of GLOE-Seq**

For the future, we foresee numerous applications of GLOE-Seq to probe the effects of medically relevant factors or treatments associated with SSBs, such as PARP1, topoisomerases, proteins involved in homologous recombination or protein-DNA crosslink repair, as well as pertinent inhibitors of such factors; e.g., olaparib, camptothecin, or mirin (Abbotts and Wilson, 2017; Caldecott, 2014). The notion that etoposide-mediated inhibition of topoisomerase II predominantly results in SSBs rather than DSBs (Muslimovic et al., 2009) illustrates the necessity to differentiate between these two important lesions and their spatial distribution when assessing the actions of commonly used therapeutic agents. Such differentiation is made possible by combining GLOE-Seq with END-seq or BLISS analysis within the same experiment. In the context of replication, GLOE-Seq should be applicable to mapping not only Okazaki fragments but also the distribution of postreplicative daughter-strand gaps arising from replication of damaged templates (Wong et al., 2020). Pre-treatment with various enzymes should give insight into additional lesions in the genome. For example, signatures of covalent topoisomerase adducts could be traced by tyrosyl-DNA phosphodiesterases (Pommier et al., 2014) and 3’-PO₄ termi revealed by phosphatase treatment. Absolute quantification of break signals can be accomplished by pre-digestion with rare-cutting restriction endonucleases that would generate a defined set of breaks as internal standards (Figures S2C and S2D). Last but not least, the nucleotide precision of GLOE-Seq may be employed for evaluating the selectivity of emerging techniques in genome engineering not involving DSBs, such as Cas9 nickase-dependent base-editing systems (Eid et al., 2018).

**Limitations**

One of the limitations inherent in any method for mapping SSBs appears to be an inevitable background of spontaneous nicks in genomic DNA. In our validation experiments, this has obscured sequence-specific signals present at a frequency of lower than 0.1% (Figure 1G). Although a higher sequencing depth would likely improve the signal-to-noise ratio, detection of rare events by GLOE-Seq seems problematic. At the same time, however, RFD analysis has revealed that the distribution of spontaneous breaks detected in DNA from unperturbed yeast or human cells is, in fact, not random but follows the activity profile of the replicative polymerase ε (Figures 4C and 6D). Thus, the “background” signal present in our samples does not actually represent noise derived from spontaneous hydrolysis or shearing during preparation but, rather, a physiological feature of native DNA isolated from a natural source that is detectable by our protocol. Taking this pattern into account by careful comparisons of experimental conditions will be important for correct interpretation of SSB signals.

Because they represent a specialized form of 3’ ends in the genome, we used GLOE-Seq for mapping the 3’ termini of Okazaki fragments. In budding yeast, previously published methods have accomplished this by directly isolating Okazaki fragments by means of size fractionation followed by NGS (Smith and Whitehouse, 2012). Our data accurately reproduce the published patterns without employing size selection. In mammalian cells, sequencing of Okazaki fragments by OK-Seq has required not only size selection but also enrichment of replicating DNA via 5-Ethynyl-2’-deoxyuridine (EdU) pulse labeling (Petryk et al., 2016). In our procedure, we avoided these elaborate steps and were able to generate replication profiles from a much lower number of cells by abolishing replicative ligase activity,
analogous to the strategy applied in yeast (Figure 6C). Upon inactivation of both replicative ligases, this approach led to underrepresentation of late-firing origins accompanied by activation of the DNA damage response. Depletion of ligase 1 only by means of siRNA largely resolved this problem, resulting in highly defined replication profiles. Thus, although the OK-Seq protocol might still be preferable in cases where ligase cannot be inhibited or an absolutely undisturbed pattern of late-replicating regions is required, GLOE-Seq in combination with ligase 1 depletion is an attractive and practicable alternative to the potentially more cumbersome OK-Seq method. In conclusion, the strength of GLOE-Seq lies in analysis of structures not covered by other NGS methods; i.e., physiological DNA SSBs, which are some of the most abundant lesions and key metabolic intermediates in the genome.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Saccharomyces cerevisiae Strains
  - Cell Lines
- METHOD DETAILS
  - GLOE-Seq Library Preparation (S. cerevisiae)
  - Sequencing
  - Sequencing Data Analysis
  - Validation of the GLOE-Seq Method
  - Comparison of GLOE-Seq with EndoSeq
  - Analysis of Spontaneous SSBs in Yeast
  - Analysis of an H2O-Induced DSB in Yeast
  - Mapping of Pyrimidine Dimers in Yeast
  - Mapping of MMS Damage and Repair in Yeast
  - Mapping of Okazaki Fragments in Yeast
  - Ligase 1 Depletion in Human Cells
  - Preparation of Mammalian Genomic DNA
  - Application of GLOE-Seq to Human Genomic DNA
  - Distribution of SSBs over Gene Features
  - Modeling of Replication Timing
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES
  - Detailed Protocol

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.molcel.2020.03.027.

ACKNOWLEDGMENTS

Support by the IMB Genomics and Flow Cytometry Core Facilities and use of IMB’s NextSeq500 (INST 247/870-1 FUGG) is gratefully acknowledged. We thank Anke Ries from IMB’s Electronics Workshop for making the plug mold 3D model; Anne Donaldson, Susan Gasser, Jim Haber, and Eric Hendrickson for yeast strains, cell lines, and reagents; Chun-Long Chen and Olivier Hyrien for sharing datasets; and Nicola Crosetto for helpful advice in the early stages of the project. Funding for this work was provided by the European Research Council (ERC; AdG 323179) and the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation; project number 393547839 – SFB 1361, sub-project 07).

AUTHOR CONTRIBUTIONS

Conceptualization, A.M.S., N.Z., and H.D.U.; Methodology, A.M.S., N.Z., G.P., M.M.-L., and L.S.B.-N.; Investigation, A.M.S., A.J.S., and N.Z.; Software, Formal Analysis, and Data Curation, G.P. and N.Z.; Writing – Original Draft, N.Z. and H.D.U.; Writing – Review & Editing, N.Z., A.M.S., G.P., and H.D.U.; Project Administration and Funding Acquisition, H.D.U.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 7, 2019
Revised: February 13, 2020
Accepted: March 20, 2020
Published: April 21, 2020

REFERENCES

Abbotts, R., and Wilson, D.M., 3rd (2017). Coordination of DNA single strand break repair. Free Radic. Biol. Med. 107, 228–244.
Alver, R.C., Chadha, G.S., and Blow, J.J. (2014). The contribution of dormant origins to genome stability: from cell biology to human genetics. DNA Repair (Amst.) 19, 182–189.
Andrews, S. (2019), FastQC: A quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
Arakawa, H., Bednar, T., Wang, M., Paul, K., Mladenov, E., Bencsik-Theilen, A.A., and Illiakis, G. (2012). Functional redundancy between DNA ligases I and III in DNA replication in vertebrate cells. Nucleic Acids Res. 40, 2599–2610.
Baranello, L., Kozunie, F., Wotowicz, D., Cui, K., Przytycka, T.M., Zhao, K., and Levens, D. (2014). DNA break mapping reveals topoisomerase II activity genome-wide. Int. J. Mol. Sci. 15, 13111–13122.
Beranek, D.T. (1990). Distribution of methyl and ethyl adducts following alkylating agents in human lymphocytes. Mutat. Res. 231, 228–244.
Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.
Caldecott, K.W. (2014). DNA single-strand break repair. Exp. Cell Res. 329, 2–8.
Canela, A., Sridharan, S., Sciascia, N., Tubbs, A., Meltzer, P., Sleckman, B.P., and Nussenzweig, A. (2016). DNA breaks and end resection measured genome-wide by End Sequencing. Mol. Cell 63, 898–911.
Cao, H., Salazar-Garcia, L., Gao, F., Wahlestedt, T., Wu, C.L., Han, X., Cai, Y., Xu, D., Wang, F., Tang, L., et al. (2019). Novel approach reveals genomic landscapes of single-strand DNA breaks with nucleotide resolution in human cells. Nat. Commun. 10, 5799.
Clausen, A.R., Lujan, S.A., Burkholder, A.B., Oreaugh, C.D., Williams, J.S., Clausen, M.F., Malc, E.P., Mieczkowski, P.A., Fargo, D.C., Smith, D.J., and Kunkei, T.A. (2015). Tracking replication enzymology in vivo by genome-wide mapping of ribonucleotide incorporation. Nat. Struct. Mol. Biol. 22, 185–191.
Crosetto, N., Mitra, A., Silva, M.J., Bienko, M., Dojer, N., Wang, Q., Karaca, E., Chiarle, R., Skrzypczak, M., Ginalska, K., et al. (2013). Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. Nat. Methods 10, 361–365.
Ding, J., Taylor, M.S., Jackson, A.P., and Reijns, M.A.M. (2015). Genome-wide mapping of embedded ribonucleotides and other noncanonical nucleotides using emRiboSeq and EndoSeq. Nat. Protoc. 10, 1433–1444.
Eid, A., Alsharief, S., and Mahfouz, M.M. (2018). CRISPR base editors: genome editing without double-stranded breaks. Biochem. J. 475, 1955–1964.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74.

Finley, D., Ozkaynak, E., and Varshavsky, A. (1987). The yeast polyubiquitin gene is essential for resistance to high temperatures, starvation, and other stresses. Cell 48, 1035–1046.

Gansauge, M.-T., and Meyer, M. (2013). Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. Nat. Protoc. 8, 737–748.

Kryzhanovska, M., Eglinger, J., Holcman, D., Owen-Hughes, T., and Gasser, M.H., Seeber, A., Singh, V., Thierry, R., Sack, R., Amitai, A., Wellinger, R.J., and Boissonneault, G. (2011). Genome-wide mapping of crosstalk and high-resolution mapping of DNA breaks. Annu. Rev. Genomics Hum. Genet. 12, 87–113.

Hansen, R.S., Thomas, S., Sandstrom, R., Canfield, T.K., Thurman, R.E., Weaver, M., Dorschnier, M.G., Carter, S.M., and Stamatoyannopoulos, J.A. (2010). Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proc. Natl. Acad. Sci. USA 107, 139–144.

Hauer, M.H., Seeger, A., Singh, V., Thierry, R., Sack, R., Amitai, A., Kryzhanovska, M., Eglinger, J., Holcman, D., Owen-Hughes, T., and Gasser, S.M. (2017). Histone degradation in response to DNA damage enhances chromatin dynamics and recombination rates. Nat. Struct. Mol. Biol. 24, 99–107.

Hoffman, E.A., McCulley, A., Haarer, B., Arnak, R., and Feng, W. (2015). Break-seq reveals hydroxyurea-induced chromosomal fragility as a result of unscheduled conflict between DNA replication and transcription. Genome Res. 25, 402–412.

Hu, J., Meyers, R.M., Dong, J., Panchakshari, R.A., Alt, F.W., and Frock, R.L. (2016). Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. Nat. Protoc. 11, 853–871.

Hu, J., Selby, C.P., Adar, S., Adebali, O., and Sancar, A. (2017). Molecular mechanisms and genomic maps of DNA excision repair in Escherichia coli and humans. J. Biol. Chem. 292, 15588–15597.

Kamani, N., and Dutta, A. (2011). The effect of the intra-S-phase checkpoint on origins of replication in human cells. Genes Dev. 25, 621–633.

Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolichik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics 26, 2204–2207.

Kubota, T., Nishimura, K., Kanemaki, M.T., and Donaldson, A.D. (2013). The Elg1 replication factor C-like complex functions in PCNA unloading during chromosome replication. Mol. Cell 50, 273–280.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.

Leduc, F., Faucher, D., Bikond Nkoma, G., Grégoire, M.-C., Arguin, M., Wellinger, R.J., and Boissonneault, G. (2011). Genome-wide mapping of DNA strand breaks. PLoS ONE 6, e17353.

Lee, S.E., Moore, J.K., Holmes, A., Umezu, K., Kolodner, R.D., and Haber, J.E. (1998). Saccharomyces KU70, mre11/rad50 and RPA proteins regulate adaptation to G2/M arrest after DNA damage. Cell 94, 399–409.

Lensing, S.V., Marsico, G., Hänsel-Hertsch, R., Lam, E.Y., Tannahill, D., and Balasubramanian, S. (2016). DSBCapture: in situ capture and sequencing of DNA breaks. Nat. Methods 13, 855–857.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.

Liu, Z.J., Martinez Cuesta, S., van Delft, P., and Balasubramanian, S. (2019). Sequencing abasic sites in DNA at single-nucleotide resolution. Nat. Chem. 11, 629–637.

Maclsaac, K.D., Wang, T., Gordon, D.B., Gifford, D.K., Stormo, G.D., and Fraenkel, E. (2006). An improved map of conserved regulatory sites for Saccharomyces cerevisiae. BMC Bioinformatics 7, 113.

Mao, P., Brown, A.J., Malc, E.P., Mieczkowski, P.A., Smerdon, M.J., Roberts, S.A., and Wyrick, J.J. (2017). Genome-wide maps of alkylation damage, repair, and mutagenesis in yeast reveal mechanisms of mutational heterogeneity. Genome Res. 27, 1674–1684.

Musilovič, A., Nyström, S., Gao, Y., and Hammarsten, O. (2009). Numerical analysis of etoposide induced DNA breaks. PLoS ONE 4, e5859.

Nick McElinny, S.A., Watts, B.E., Kumar, D., Watt, D.L., Lundström, E.-B., Burgers, P.M.J., Johansson, E., Chabes, A., and Kunkel, T.A. (2010). Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. Proc. Natl. Acad. Sci. USA 107, 4949–4954.

Oh, S., Harvey, A., Zimbric, J., Wang, Y., Nguyen, T., Jackson, P.J., and Hendrickson, E.A. (2014). DNA ligase III and DNA ligase IV carry out genetically distinct forms of end joining in human somatic cells. DNA Repair (Amst.) 13, 97–110.

Petryk, N., Kahl, M., d’Aubenton-Caratya, Y., Jaszczyszyn, Y., Shen, Y., Silvain, M., Therms, C., Chen, C.-L., and Hyrien, O. (2016). Replication landscape of the human genome. Nat. Commun. 7, 10208.

Plosky, B., Samson, L., Engelward, B.P., Gold, B., Schlaen, B., Mills, T., Magnotti, M., Schor, J., and Scicchitano, D.A. (2002). Base excision repair and nucleotide excision repair contribute to the removal of N-methylpurines from active genes. DNA Repair (Amst.) 1, 683–696.

Poetsch, A.R., Boulton, S.J., and Luscombe, N.M. (2018). Genomic landscape of oxidative DNA damage and repair reveals reselective protection from mutagenesis. Genome Biol. 19, 215.

Pommier, Y., Huang, S.Y., Gao, R., Das, B.B., Murai, J., and Marchand, C. (2014). Tyrosyl-DNA-phosphodiesterases (TD1P and TD2P). DNA Repair (Amst.) 19, 114–129.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842.

Ramirez, F., Durand, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 42, W187–91.

Rejns, M.A.M., Kemp, H., Ding, J., de Procé, S.M., Jackson, A.P., and Taylor, M.S. (2015). Lagging-strand replication shapes the mutational landscape of the genome. Nature 518, 502–506.

Sadedin, S.P., Pope, B., and Oshlack, A. (2012). Biipe: a tool for running and managing bioinformatics pipelines. Bioinformatics 28, 1525–1526.

Smith, B.J., and Whitehouse, I. (2012). Intrinsic coupling of lagging-strand synthesis to chromatin assembly. Nature 483, 434–438.

Thomas, B.J., and Rothstein, R. (1989). Elevated recombination rates in transcriptionally active DNA. Cell 56, 619–630.

Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wywents, K., Khayter, C., Iafrate, A.J., Le, L.P., et al. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat. Biotechnol. 33, 187–197.

Tubbs, A., Sridharan, S., van Wietmarschen, N., Maman, Y., Callen, E., Stanlie, A., Wu, W., Wu, X., Day, A., Wong, N., et al. (2018). Dual roles of poly(dA:dT) tracts in replication initiation and fork collapse. Cell 174, 1127–1142.e19.

Vitelli, V., Galbiati, A., Ianneli, F., Pessina, F., Sharma, S., and d’Adda di Fagagna, F. (2017). Recent advancements in DNA damage transcription crosstalk and high-resolution mapping of DNA breaks. Annu. Rev. Genomics Hum. Genet. 18, 87–113.

Whitehouse, I., Rando, O.J., Delrow, J., and Tsukiyama, T. (2007). Chromatin remodelling at promoters suppresses antisense transcription. Nature 450, 1031–1035.

Wong, R.P., García-Rodríguez, N., Zilio, N., Hanulová, M., and Ulrich, H.D. (2020). Processing of DNA polymerase-blocking lesions during genome replication is spatially and temporally segregated from replication forks. Mol. Cell 77, 9–16.e4.
Wu, J., McKeague, M., and Sturla, S.J. (2018). Nucleotide-resolution genome-wide mapping of oxidative DNA damage by Click-Code-Seq. J. Am. Chem. Soc. 140, 9783–9787.
Xiao, W., and Chow, B.L. (1998). Synergism between yeast nucleotide and base excision repair pathways in the protection against DNA methylation damage. Curr. Genet. 33, 92–99.
Yan, W.X., Mirazadeh, R., Garnerone, S., Scott, D., Schneider, M.W., Kallas, T., Custodio, J., Wernersson, E., Li, Y., Gao, L., et al. (2017). BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. Nat. Commun. 8, 15058.
Yu, G., Wang, L.-G., and He, Q.-Y. (2015). ChiPseeker: an R/Bioconductor package for ChiP peak annotation, comparison and visualization. Bioinformatics 31, 2382–2383.
Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137.
## STAR METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE NAME | SOURCE | IDENTIFIER |
|--------------------------|--------|------------|
| Antibodies               |        |            |
| Mouse monoclonal anti-yeast Rad53 | Susan Gasser; Hauer et al., 2017 | N/A |
| Rat monoclonal anti-tubulin, clone YL1/2 | Sigma-Aldrich | Cat# 92092402-1VL, RRID: CVCL_J781 |
| Rabbit polyclonal anti-DNA Ligase 1 | Elabscience | Cat# E-AB-31210 |
| Mouse monoclonal anti-Chk1, clone 2G1D5 | Cell Signaling Technology | Cat# 2360, RRID: AB_2080320 |
| Rabbit polyclonal anti-phospho-Chk1 (Ser345) | Cell Signaling Technology | Cat# 2341, RRID: AB_330023 |
| Goat anti-mouse immunoglobulins, HRP | Dako | Cat# P0447, RRID: AB_2617137 |
| Goat anti-rabbit immunoglobulins, HRP | Dako | Cat# P0448, RRID: AB_2617138 |
| Goat anti-rat immunoglobulins, HRP | Dako | Cat# P0450, RRID: AB_2630354 |
| Goat anti-rat IgG Secondary Antibody, IRDye® 800CW | LI-COR | Cat# 926-32219, RRID: AB_1850025 |
| Chemicals, Peptides, and Recombinant Proteins |        |            |
| Zymolyase 20T | AMS Biotechnology | Cat# 120491-1 |
| T4 DNA ligase, 20,000,000 U/mL | New England Biolabs | Cat# B0202 |
| Q5® High-Fidelity DNA polymerase | New England Biolabs | Cat# M0491 |
| RNase A from bovine pancreas | Sigma-Aldrich | Cat# 10109169001 |
| Proteinase K | Roche | Cat# 3115801001 |
| Phenylmethanesulfonyl fluoride | Sigma-Aldrich | Cat# P7626 |
| β-Agarase I, 1000 U/mL | New England Biolabs | Cat# M0392 |
| Alpha-Factor peptide (WHWLQLKPGQPMY), > 95% | ProteoGenix SAS | Cat# GM-PT001 |
| BsrDI | New England Biolabs | Cat# R0574 |
| Nb.BsrDI | New England Biolabs | Cat# R0648 |
| NotI | New England Biolabs | Cat# R0189 |
| Nb.BtsI | New England Biolabs | Cat# R0707 |
| Antarctic Phosphatase | New England Biolabs | Cat# M0289 |
| T4 Endonuclease V (T4 PDG) | New England Biolabs | Cat# M0308 |
| APE1 | New England Biolabs | Cat# M0282 |
| hAAG | New England Biolabs | Cat# M0313 |
| Lipofectamine RNAiMAX Transfection Reagent | Thermo Fisher Scientific | Cat# 13778150 |
| Benzonase | Sigma-Aldrich | Cat# E1014-25KU |
| PhosSTOP Phosphatase Inhibitor | Sigma-Aldrich | Cat# 04906837001 |
| Opti-MEM® Reduced Serum Medium | Thermo Fisher Scientific | Cat# 11058021 |
| Critical Commercial Assays |        |            |
| AMPure XP beads | Beckman Coulter | Cat# A63881 |
| NEBNext Ultra II DNA Library Prep Kit for Illumina® | New England Biolabs | Cat# E7645 |
| Phusion Flash high-fidelity PCR master mix | Thermo Fisher Scientific | Cat# F-548 |
| Dynabeads MyOne Streptavidin C1 | Life Technologies | Cat# 65001 |
| High Sensitivity D1000 ScreenTape | Agilent Technologies | Cat# 5067-5584 |
| High Sensitivity D1000 ScreenTape reagents | Agilent Technologies | Cat# 5067-5585 |
| RNA ScreenTape | Agilent Technologies | Cat# 5067-5576 |
| RNA ScreenTape sample buffer | Agilent Technologies | Cat# 5067-5577 |
| RNA ScreenTape ladder | Agilent Technologies | Cat# 5067-5578 |
| Qubit dsDNA HS Assay Kit | Invitrogen | Cat# Q32854 |
| Agilent Bioanalyzer High Sensitivity DNA Kit | Agilent Technologies | Cat# 5067-4646 |

(Continued on next page)
**REAGENT or RESOURCE** | **SOURCE** | **IDENTIFIER**  
---|---|---  
NextSeq 550 System High-Output Kit | Illumina | Cat# 20024906  
NextSeq 550 System Mid-Output Kit | Illumina | Cat# 20024904  
NuSieve GTG Agarose | Lonza | Cat# 859081  
Yeast Genomic DNA Extraction Kit | QIAGEN | Cat# 10243  
Bio-Rad Protein Assay Dye Reagent | Bio-Rad | Cat# 500-0006  
NuPAGE LDS sample buffer | Thermo Fisher Scientific | Cat# NP0007  
NuPAGE 4-12% Bis-Tris Protein Gels | Thermo Fisher Scientific | Cat# NP0322  

**Deposited Data**  
Raw and analyzed data | This paper | GEO: GSE134225  
Human reference genome UCSC GRCh37/hg19 | UCSC Genome Browser | ftp://hgdownload.soe.ucsc.edu/goldenPath/hg19/chromosomes/  
Yeast reference genome UCSC sacCer3 | UCSC Genome Browser | ftp://hgdownload.soe.ucsc.edu/goldenPath/sacCer3/chromosomes/  
3’ end maps of Nb.BtsI-digested DNA, WT and pol1-L868M, Fastq (Figure 2A, reanalyzed with GLOE-Pipe) | Reijns et al., 2015 | GEO: GSM1573437, GSM1573438  
Motifs for Abf1, Reb1, Rap1 in yeast, GFF (Figure 4B) | MacIsaac et al., 2006 | https://downloads.yeastgenome.org/published_datasets/MacIsaac_2006_PMID_16522208/MacIsaac_high_confidence_with_sequence.gff  
Nucleosome positions in yeast, BED (Figure 4B) | Whitehouse et al., 2007 | https://static-content.springer.com/esm/art%3A10.1038%2Fnature06391/MediaObjects/41586_2007_BFnature06391_MOESM319_ESM.xls  
Okazaki fragment maps in Cdc9-depleted yeast, Fastq, rep 1 & rep2 (Figure 4C, reanalyzed with GLOE-Pipe) | Smith and Whitehouse, 2012 | GEO: GSM835650, GSM835651  
END-seq data in HCT116, Fastq (Figure 5, reanalyzed with GLOE-Pipe) | Tubbs et al., 2018 | GEO: GSM3227952  
DNase I hypersensitivity data in HCT116, BigWig, rep1 & rep2 (Figure 6B) | ENCODE Project Consortium, 2012 | GEO: GSM736493, GSM736600  
OK-Seq and S50 data in HeLa and GM06990, rep1 & rep2 (BedGraph, received from Chunlong Chen, chunlong.chen@curie.fr) (Figure 6B) | Petryk et al., 2016 | GEO SRA: SRP065949 (SRX1427549, SRX1427548, SRX1424659, SRX1424656)  

**Experimental Models: Cell Lines**  
HCT116 | Cancer Research UK London Research Institute Cell Services | N/A  
HCT116 LIG3<sup>-/-</sup> mL3 | Oh et al., 2014 | N/A  

**Experimental Models: Organisms/Strains**  
S. cerevisiae: strain W303 | ATCC | ATCC: 201238  
S. cerevisiae: strain JKM179 | Lee et al., 1998 | N/A  
S. cerevisiae: strain DF5 | ATCC | ATCC: 200912  
S. cerevisiae: strain DF5 <sup>apn1::KanMX apn2::KanMX</sup> | This paper | N/A  
S. cerevisiae: strain cdc9<sup>Δ9</sup> | Kubota et al., 2013 | N/A  

**Oligonucleotides**  
Primer #3899: CTACACGACGCTCTTCCGATCTNNN NNY’-NH<sub>2</sub> (phosphorothioate bond, IDT code *; NH<sub>2</sub>: 3’-amino modification, IDT code /3AmMO/) | Integrated DNA Technologies | N/A  
Primer #3899: PO<sub>4</sub>-AGATCGGAAGAGCGTCGTGTAG GGAAAGAGTG TAGATCTCGTTTT-Bio (PO<sub>4</sub>: 5’-phosphorylation, IDT code /5Phos/; T-Bio: 3’-biotin-dT, IDT code /3BiodT/) | Integrated DNA Technologies | N/A

(Continued on next page)
### Lead Contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by Helle D. Ulrich (h.ulrich@imb-mainz.de).

### REAGENT or RESOURCE SOURCE IDENTIFIER

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| Primer #3790: CCGAGATCTACACTTTCCCTACA CGACGCTCTCCGATCT | Integrated DNA Technologies | N/A |
| Primer #3791: GACTGGAGTTGACGTGGTGCTTCCGATCT | Integrated DNA Technologies | N/A |
| Primer #3792: GATCGGAGACACCGATCTG AACTCCAGTC | Integrated DNA Technologies | N/A |
| Primer P5: AATGATACGGCGACCACCGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT | Integrated DNA Technologies | N/A |
| Primer P7: CAAGCAGAAGACCGAGATCCTTCCGATCTX(X)6GTGACTGGAGTTCAGACGTGTGCTTCCGATCT | Integrated DNA Technologies | N/A |
| siRNA targeting sequence: Silencer Select Negative Control No. 1 | Thermo Fisher Scientific | Cat# 4390843 |
| siRNA targeting sequence: Ligase 1: Silencer Select s8174 | Thermo Fisher Scientific | Cat# 4390824 |

### Software and Algorithms

| Software and Algorithms | Source | Identifier |
|-------------------------|--------|------------|
| bcl2fastq, version 2.19 | Illumina | [https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html](https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html) |
| Bpipe, version 0.9.9.3 | Sadedin et al., 2012 | [http://docs.bpipe.org/](http://docs.bpipe.org/) |
| NGSpipe2go | Institute of Molecular Biology gGmbH | [https://github.com/imbforge/NGSpipe2go](https://github.com/imbforge/NGSpipe2go) |
| FastQC, version 0.11.5 | Andrews, 2019 | [https://www.bioinformatics.babraham.ac.uk/projects/fastqc/](https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) |
| Trimmomatic, version 0.36 | Bolger et al., 2014 | [http://www.usadellab.org/cms/?page=trimmomatic](http://www.usadellab.org/cms/?page=trimmomatic) |
| Bowtie 2, version 2.3.4 | Langmead and Salzberg, 2012 | [http://bowtie-bio.sourceforge.net/bowtie2/index.shtml](http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) |
| Samtools, version 1.5 | Li et al., 2009 | [http://samtools.sourceforge.net/](http://samtools.sourceforge.net/) |
| BEDTools, version 2.25.0 | Quinlan and Hall, 2010 | [https://bedtools.readthedocs.io/en/latest/](https://bedtools.readthedocs.io/en/latest/) |
| bedGraphToBigWig, version 365 | Kent et al., 2010 | [https://github.com/ENCODE-DCC/kentUtils](https://github.com/ENCODE-DCC/kentUtils) |
| MACS2 callpeak, version 2.1.1 | Zhang et al., 2008 | [https://github.com/taoliu/MACS](https://github.com/taoliu/MACS) |
| ChIPseeker package, version 1.14.1 | Yu et al., 2015 | [https://bioconductor.org/packages/release/bioc/html/ChIPseeker.html](https://bioconductor.org/packages/release/bioc/html/ChIPseeker.html) |
| deepTools, version 3.1.0 | Ramirez et al., 2014 | [https://deeptools.readthedocs.io/en/develop/](https://deeptools.readthedocs.io/en/develop/) |
| Replicon | Gindin et al., 2014 | [https://github.com/RepliconBioinfo/Replicon](https://github.com/RepliconBioinfo/Replicon) [downloaded 20.12.2019] |
| FlowJo, version 10.6.1 | FlowJo, LLC | [https://www.flowjo.com/](https://www.flowjo.com/) |
| GLOE-Pipe | This paper | [https://github.com/helle-ulrich-lab/ngs-gloepipe](https://github.com/helle-ulrich-lab/ngs-gloepipe) |

### Other

| Detailed protocol for the preparation of GLOE-Seq libraries | This paper | See Supplemental Information |
| 3D model of a custom-made mold for agarose plugs | This paper | See Data S1 |
| 3D model of a tool for extrusion of agarose plugs from custom-made mold | This paper | See Data S1 |
Materials Availability
This study did not generate new unique reagents.

Data and Code Availability
The datasets generated and analyzed during this study are available in the Gene Expression Omnibus repository, https://www.ncbi.nlm.nih.gov/geo, under accession number GEO: GSE134225. Published datasets used for this analysis and their accessibility are summarized in the Key Resources Table. GLOE-Pipe is publicly accessible, fully documented and regularly maintained by the developers at https://github.com/helle-ulrich-lab/ngs-gloepipe.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Saccharomyces cerevisiae Strains
W303: MATa, ade2-1 ura3-1 his3-11,15 trp1-1 leu2-3,112 can1-100 (Thomas and Rothstein, 1989)
JKM179: MATa, Δho Δhml::ADE1 Δhmr::ADE1 ade1-110 leu2-3,112 lys5 trp1::hisG ura3-52 ade3::GAL10:HO (Lee et al., 1998)
DF5: MATa, his3-Δ200 leu2-3,2-112 lys2-801 trp1-1(αm) ura3-52 (Finley et al., 1987)
DF5 apn1 Δ apn2 Δ: MATa, his3-Δ200 leu2-3,2-112 lys2-801 trp1-1(αm) ura3-52 apn1::KanMX apn2::KanMX (this study)
cdc9Δ: MATa, bar1::hisG ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 RAD5+ URA3::pRS306-PGAL1-10-OsTIR1 cdc9-Δ1Δ::kanMX (Kubota et al., 2013)

Yeast cultures were grown in YP medium containing 2% (w/v) glucose at 30°C unless otherwise noted.

Cell Lines
HCT116 (Cancer Research UK – London Research Institute's Cell Production Services)
HCT116 LIG3Δ/Δ/Δ: mL3 (Oh et al., 2014; received from Eric Hendrickson)

Cells were grown as monolayers in a humidified incubator at 37°C and 5% (v/v) CO₂ in DMEM supplemented with 10% (v/v) fetal bovine serum, penicillin (100 U/mL), streptomycin (100 μg/mL) and glutamine (300 μg/mL).

METHOD DETAILS

GLOE-Seq Library Preparation (S. cerevisiae)
Note that a detailed step-by-step protocol of the GLOE-Seq method is provided as Supplemental Information. Genomic DNA for validating the method, to detect chromosomal features and to map pyrimidine dimers and MMS damage was prepared using a spheroplast-based genomic DNA extraction kit (QIAGEN) according to the instruction manual. Genomic DNA for mapping Okazaki fragments and for capturing HO-induced DSBs was prepared by gentle lysis of spheroplasts, precipitation of the DNA, RNase A treatment and purification via AMPure beads (Beckman Coulter) as described in the step-by-step protocol (steps y1-28). Following additional treatment of the genomic DNA as specified for each experiment below, libraries were prepared from 2.5 μg of DNA by the core GLOE-Seq procedure (steps 29-50 of the step-by-step protocol). Briefly, this involved thermal denaturation, ligation of the proximal biotinylated adaptor, fragmentation to an average size of 200 nt, capture on streptavidin beads, second strand synthesis, end polishing and ligation of the distal adaptor. The resulting library was amplified using P5 and P7 primers of the Illumina system.

Sequencing
GLOE-Seq libraries were sequenced on an Illumina NextSeq 500 sequencer with High or Mid Output flow cells, depending on the number of libraries loaded at a time. All libraries were sequenced in single-end mode, with read lengths of 75, 84, 92, 150 or 160 bases (depending on the sequencing kit and the level of multiplexing per flow cell) plus 7 bases for single-indexed libraries or 8+8 bases for dual-indexed libraries. Sequencing depth varied with the type of library; for yeast GLOE-Seq libraries we aimed for 3 million reads per sample, while for mammalian libraries a single library was loaded per NextSeq 500 High Output flow cell. However, downsampling the reads to 50 million proved to yield sufficient depth for Okazaki fragment analysis via GLOE-Seq. Upon completion of the run, raw sequencing reads of pooled libraries were demultiplexed based on their index sequences by bcl2fastq (version 2.19, Illumina). The resulting Fastq files were used as input files for GLOE-Pipe.

Sequencing Data Analysis
To automate and standardize the processing and analysis of GLOE-Seq data, we developed GLOE-Pipe, a bioinformatics toolkit for the identification of strand breaks at nucleotide resolution in genomic DNA. GLOE-Pipe uses Bpipe (Sadedin et al., 2012) (version 0.9.9.3) as workflow manager and domain specific language (DSL). It represents a branch of NGSpipe2go, containing a set of modules that together process raw sequencing data and generate output files in order to detect, annotate and visualize breaks from raw data.
Validation of the GLOE-Seq Method

Purified genomic DNA from yeast strain W303 was treated with 1 U/μg DNA of the relevant restriction or nicking enzymes, BsrDI, Nb.BsrDI or NolI (New England Biolabs) for 90 min at 65°C (BsrDI and Nb.BsrDI) or 37°C (NolI), dephosphorylated with 2 U/μg Antarctic Phosphatase (New England Biolabs) for 30 min at 37°C and purified using AMPure beads (Beckman Coulter). The purified DNA was quantified (Qubit, Life Technologies), and 2.5 μg were used for GLOE-Seq library preparation (steps 29-50) and sequencing. Raw GLOE-Seq data from these experiments were processed with GLOE-Pipe in the indirect mode for downstream analysis, but visualized in the direct mode (Figure 1B). Significant peaks were assigned by comparing the BED file for each strand to an undigested sample, when possible, using MACS2 callpeak (Zhang et al., 2008) (version 2.1.1; parameters: --extsize 1, --nomodel, --shift 0, --keep-dup). Custom code based on ChIPseeker package (Yu et al., 2015) (version 1.14.1) was used to check the overlap between the detected and expected breaks.

Comparison of GLOE-Seq with EndoSeq

In order to benchmark GLOE-Seq against a published procedure, the GLOE-Seq protocol (steps 29-50) was applied to 2.5 μg of Nb.BstI-digested yeast genomic DNA. These data were analyzed in parallel with the corresponding published EndoSeq datasets (Reijns et al., 2015). In order to make the two datasets directly comparable, the EndoSeq datasets were downsampled to 5 million reads and the reads were trimmed to 75 nt. For direct comparison of the two protocols, Nb.BsrDI sites were mapped in genomic yeast DNA by either GLOE-Seq (steps 29-50) as described above or EndoSeq according to the published protocol (Ding et al., 2015). Both samples were sequenced on the same platform (Illumina NextSeq). In order to test the suitability of EndoSeq for capturing pre-existing nicks, genomic DNA was first treated with Nb.BsrDI and subsequently subjected to the EndoSeq protocol. Raw data from these experiments were processed with GLOE-Pipe in the indirect mode. Significant breaks for each protocol were called with MACS2 callpeak (Zhang et al., 2008) (version 2.1.1) without specifying the –control flag. Expected breaks were grouped into three categories (unique, common, neither) using BEDTools (Quinlan and Hall, 2010) intersect (version 2.25.0). The location of each break was combined with its normalized counts calculated by GLOE-Pipe and custom code was used to generate scatterplots showing the comparison between the two protocols.

Analysis of Spontaneous SSBs in Yeast

Genomic DNA prepared from exponential W303 cultures was subjected to GLOE-Seq (steps 29-50). As a control sample representing a random distribution of single-stranded breaks, an aliquot of the genomic DNA was first subjected to fragmentation by sonication to an average size of 300 bp before applying the GLOE-Seq protocol (steps 29-50). Raw GLOE-Seq data from these experiments were processed with GLOE-Pipe in the indirect mode. Reads were assigned with high stringency by disallowing any mismatches. For relevant samples, tracks based on the replication fork directionality (RFD) ratio [RFD = (REV – FWD)/(REV + FWD)] were generated using bigwigCompare [bigwigCompare [deepTools (Ramirez et al., 2014), version 3.1.0].

Analysis of an HO-Induced DSB in Yeast

JKM179 cells (Lee et al., 1998) harboring a galactose-inducible HO endonuclease construct were cultured in YP medium containing 2% (w/v) raffinose, and DSB formation at the mating-type locus was induced by growth in YP medium containing 2% (w/v) galactose for 1 h, as described previously (Lee et al., 1998). Control samples were prepared from cells grown in raffinose medium. Genomic DNA was prepared (steps y1-28), and GLOE-Seq was performed according to the step-by-step protocol (steps 29-50). Raw
Mapping of Pyrimidine Dimers in Yeast
Exponentially growing DF5 cells were exposed to UV irradiation at 20 and 120 J/m² (254 nm, Stratalinker, Stratagene) or left unirradiated. Total genomic DNA was prepared, and 2.5 μg of purified DNA was either treated with 5 U/μg T4 Endonuclease V (New England Biolabs, 30 min at 37°C), followed by 2 U/μg APE1 (New England Biolabs, 30 min at 37°C), or left untreated, before application of the GLOE-Seq protocol (steps 29-50). Raw GLOE-Seq data from these experiments were processed with GLOE-Pipe in the indirect mode. Custom code was used to extract dinucleotides corresponding to the reported damaged bases, in the downstream direction, and the percentage of each dinucleotide sequence was calculated.

Mapping of MMS Damage and Repair in Yeast
Yeast strains DF5 and an isogenic apn1Δ apn2Δ mutant were arrested in G1 phase with 5 μg/mL alpha-factor for 90 min, followed by a 30 min incubation with 0.02% MMS and release into S phase in the absence of MMS. For analysis of checkpoint activation, aliquots corresponding to 1 OD₆₀₀ were collected. Cell pellets were resuspended in 1.85 M NaOH, 7.4% β-mercaptoethanol and incubated on ice for 15 min. After addition of 75 μL of 55% (w/v) trichloroacetic acid, further incubation on ice for 10 min and centrifugation at 13,800 × g for 10 min at 4°C, pellets were resuspended in 40 μL of 100 mM dithiorthreitol in 1 × LDS sample buffer (Invitrogen) and incubated at 65°C for 20 min. Protein extracts were loaded onto 4%–12% Bis-Tris gradient gels (Invitrogen) and analyzed by western blotting using anti-Rad53 (Hauer et al., 2017) and anti-tubulin (YL1/2, Sigma-Aldrich) primary antibodies as well as anti-mouse IgG secondary antibody (Dako) and fluorophore-coupled secondary anti-rat antibody (IRDye® 800CW, LI-COR), respectively. Cell cycle phase was monitored by harvesting 1 mL samples, washing cells once with water and fixing in 1 mL of 70% (v/v) ethanol. Fixed cells were washed twice with 1 mL of 50 mM sodium citrate, pH 7.0, treated with 80 μg/mL RNaseA at 50°C for 1 h, followed by 80 μg/mL Proteinase K at 50°C for 1 h. Samples were stained with 32 mg/mL propidium iodide, sonicated and analyzed using flow cytometry (FACSVerse, BD Biosciences). Data were analyzed with FlowJo v10 software (FlowJo, LLC).

In order to detect strand breaks as repair intermediates, genomic DNA was prepared from cells harvested at the indicated time points after release (steps y1-28), followed by application of the GLOE-Seq protocol (steps 29-50). In order to map base lesions, the extracted genomic DNA was pre-treated with 3.75 U/μg hAAG (New England Biolabs, 90 min at 37°C), followed by 2.5 U/μg APE1 (New England Biolabs, 90 min at 37°C), purification via AMPure beads and application of the GLOE-Seq protocol (steps 29-50). In both cases, raw data were processed by GLOE-Pipe in the indirect mode.

Mapping of Okazaki Fragments in Yeast
Ligation of Okazaki fragments was inhibited by auxin-induced degradation of degron-tagged Cdc9 in a cdc9Δ strain as previously described (Kubota et al., 2013). Briefly, cells were arrested in G1 phase with 5 μg/mL alpha-factor for 90 min, followed by a 60 min incubation with 1 μM auxin and release into S phase for 60 min. Genomic DNA was prepared (steps y1-28), followed by dephosphorylation with Antarctic Phosphatase (2 U/μg DNA) for 30 min at 37°C, purification via AMPure beads and application of the GLOE-Seq protocol (steps 29-50). Raw GLOE-Seq data from these experiments were processed with GLOE-Pipe in the indirect mode. The distribution of breaks on each strand was compared with nucleosome locations (Whitehouse et al., 2007) and transcription factor binding sites (Maclsaac et al., 2006) (Abf1, Reb1, Rap1 occupancy) around TSSs using computeMatrix and plotProfile (deepTools; Ramírez et al., 2014; version 3.1.0). Tracks based on the RDF ratio were generated as described above.

Ligase 1 Depletion in Human Cells
Treatment of HCT116 and HCT116 LIG3−/−:mL3 cells with siRNA was carried out 24 h after seeding 200,000 WT cells in 6-cm plates (21.3 cm²) with 5.5 mL medium and 1.23 million LIG3−/−:mL3 cells in 10-cm plates (59 cm²) with 15.5 mL medium, to account for the slower growth rate of the latter and the consequent lower cell yield at the end of the procedure. For the 10-cm plates, 185 pmol of control (Silencer Select control no. 1, Thermo Fisher) or Ligase 1 (Silencer Select s8174, Thermo Fisher Scientific) siRNAs were mixed with 31 μL Lipofectamine RNAiMAX (Thermo Fisher Scientific) at a final volume of 3.1 mL OptiMEM (Thermo Fisher Scientific) and added to the cells. For the 6-cm plates, these amounts were scaled down according to surface area. Two days post siRNA transfection, cells were passaged into 10-cm plates and grown for two more days before being harvested and immediately processed for GLOE-Seq. In parallel, total protein extracts were prepared for western blot analysis by incubation of cells in 25 mM HEPES pH 7.5, 100 mM NaCl, 2 mM MgCl₂, 1% Triton X-100 and 0.5 U/L Benzonase (Sigma-Aldrich) at 4°C for 1-2 h. Samples of 50 μg (determined by Bio-Rad’s Protein Assay) were resolved on a NuPAGE 4%-12% Bis-Tris gel (Thermo Fisher Scientific) and transferred to a PVDF membrane. The membrane was blocked in 5% (w/v) non-fat milk in PBS with 0.1% (v/v) Tween-20 and incubated with antibodies against Ligase 1 (Elabscience), CHK1 (Cell Signaling Technology), phospho-CHK1 (Cell Signaling Technology) and tubulin (YL1/2, Sigma-Aldrich), which were detected by enhanced chemiluminescence. Nuclei left over from the preparation of genomic DNA (see below) were analyzed by flow cytometry (FACSVerse, BD Biosciences) in PBS with 80 μg/mL propidium iodide and the resulting data visualized in FlowJo 10. For UV irradiation, HCT116 cells were grown to a confluency of ~50%, washed twice in PBS, exposed to 25 J/m² UV (254 nm, Stratalinker 2400, Stratagene) and harvested after 4 h. Whole-cell extracts were prepared as described above with the addition of phosphatase inhibitors (PhosSTOP, Sigma-Aldrich) to the lysis buffer.
Preparation of Mammalian Genomic DNA

A detailed step-by-step protocol for the GLOE-Seq procedure, including the preparation of genomic DNA, is provided as Supplemental Information. Briefly, mammalian genomic DNA was isolated from agarose-embedded nuclei prepared from cultured cells as described in the step-by-step protocol (steps m1-21). Briefly, cells were harvested by trypsinization and nuclei were prepared by incubation with 0.1% (v/v) Triton X-100. Nuclei were washed, treated with RNase A and embedded in low-melting point agarose. After solidification, the agarose plugs were treated with Proteinase K solution containing 1% (w/v) Sarkosyl to induce lysis. Proteinase K was then inactivated by PMSF treatment, followed by extensive washing of the plugs. 3D models of the plug mold and an extrusion tool are available as Data S1.

Application of GLOE-Seq to Human Genomic DNA

All steps of the GLOE-Seq protocol up to the ligation of the proximal adaptor (steps m22-28, see Supplemental Information) were carried out in agarose plugs. Briefly, the agarose-embedded genomic DNA was denatured by heating to 95°C, followed by quenching on ice. Ligation was performed overnight at 16°C, followed by digestion of the agarose with β-agarase. The DNA was then fragmented by sonication in a Bioruptor Pico, purified with AMPure beads and processed using the core GLOE-Seq workflow (steps 29-50). Raw GLOE-Seq data from these experiments were processed with GLOE-Pipe in the indirect mode. RFD ratios were calculated within 1-kbp-stepped 10-kbp sliding bins.

Distribution of SSBs over Gene Features

RNA transcript levels (FPKM, Fragments Per Kilobase of transcript per Million mapped reads, as determined by RNA-Seq) for HCT116 cells were obtained from the ENCODE project (ENCODE Project Consortium, 2012) and supplemented with relevant additional information from UCSC’s annotation database. For genes with more than one reported isoform, the single largest FPKM value among all isoforms was chosen for downstream analysis. Genes were grouped into two categories: those without (FPKM = 0) and those with detectable transcription (FPKM > 0). For TSSs, TTSs, 5’ and 3’ exon junctions, the signal was extracted 5,000 bp upstream/downstream of the relevant reference point and processed as 1-bp-stepped moving averages of 1-kbp windows. For 5’ and 3’ UTRs, exons and introns, SSB data spanning the exact location of each individual feature were extracted, concatenated and divided into 1-kbp-long non-overlapping bins, which were used to compute averages. The resulting values were converted into percentage deviation from the “background” genome-wide level of signal, which was defined as the mean signal in 1-kbp non-overlapping bins along the entire length of the genome.

Multiple reads mapping to the same coordinate were assumed to derive from one unique break.

Modeling of Replication Timing

DNase I hypersensitivity data for HCT116 cells were obtained from the ENCODE project (ENCODE Project Consortium, 2012). Custom code was used to model the values for the six Repli-Seq fraction profile compartments (G1/G1b, S1, S2, S3, S4, G2) (Hansen et al., 2010), in 1 kbp bins, using Replicon (Gindin et al., 2014). S50 ratios, defined as the fraction of S phase (0 < S50 < 1) at which 50% of the DNA is replicated in a specific bin, were generated by linear regression of these data.
Supplemental Information

Genome-wide Nucleotide-Resolution Mapping of DNA Replication Patterns, Single-Strand Breaks, and Lesions by GLOE-Seq

Annie M. Sriramachandran, Giuseppe Petrosino, María Méndez-Lago, Axel J. Schäfer, Liliana S. Batista-Nascimento, Nicola Zilio, and Helle D. Ulrich
Figure S1. Related to Figure 1. GLOE-Seq workflow

A. Schematic step-by-step illustration of the GLOE-Seq procedure, indicating critical oligonucleotide sequences and modifications. Green circle: ligatable 3’-OH terminus; red circle (B): biotin-dT; -NH₂ (or x): 3’-amino modification; *: phosphorothioate bond; StrAv: Streptavidin.

B. Quality control samples were analysed on an Agilent TapeStation: QC1 – RNA ScreenTape of denatured genomic DNA (1 μL) after ligation of the proximal adaptor and fragmentation. QC2 – RNA ScreenTape of ligated single-stranded genomic DNA fragments (1 μL) captured on Streptavidin beads. QC3 – High Sensitivity DNA ScreenTape of captured DNA (2 μL) after second-strand synthesis. QC4 – High Sensitivity DNA ScreenTape of the libraries (1 μL) after ligation of the distal adaptor and amplification with Illumina primers P5 and P7.
Figure S2. Related to Figure 1. Assessment of the sensitivity of GLOE-Seq
A. GLOE-Seq profiles of Nb.BsrDI sites in independent replicates are highly correlated up to a dilution of 1:1000 (r²: Pearson correlation coefficients).
B. GLOE-Seq profiles of NotI sites in independent replicates are highly correlated across all samples.
C. SSBs can reliably be detected at high dilution. The distribution of GLOE-Seq signals is shown for a series of Nb.BsrDI/NotI-digested DNA diluted at the indicated ratios with NotI-digested DNA. The undiluted DNA was only treated with Nb.BsrDI.
D. The observed numbers of reads per Nb.BsrDI site correlate well with the expected numbers across all dilutions. The expected numbers were calculated as follows:

\[
\text{Expected} = \frac{\# \text{reads mapped to Nb.BsrDI & NotI sites} \times \# \text{Nb.BsrDI 3'-ends} \times \text{Dil. factor}}{\# \text{NotI 3'-ends} + \# \text{Nb.BsrDI 3'-ends} \times \text{Dil. factor}},
\]

where \# Nb.BsrDI 3'-ends = 6271, \# NotI 3'-ends = 80 and Dil. factor = [1, 0.1, 0.01, 0.001, 0.0001].
Figure S3. Related to Figure 2. Comparison of GLOE-Seq with a published EndoSeq data set

A. GLOE-Seq versus EndoSeq workflow. Yellow circle: base lesion; green circle: ligatable 3’-OH terminus; red circle: biotin.
B. A scatter plot shows the normalised numbers of reads at all Nb.BtsI sites for samples processed by GLOE-Seq in comparison to a published EndoSeq data set (Reijns et al., 2015), grouped by peak calling. C. GLOE-Seq and EndoSeq detect a similar percentage of Nb.BtsI sites.
D. Comparison of the percentage of total reads mapped to predicted Nb.BtsI sites by GLOE-Seq versus EndoSeq.
Figure S4. Related to Figure 3. Distribution of breaks at yeast chromosome ends
Breaks are enriched at several yeast chromosome ends. Strand-specific GLOE-Seq signals of untreated and randomly fragmented DNA are shown in the subtelomeric regions of all 16 chromosomes. Defined sequence elements (X, XCR, Y') are indicated below each plot. The telomeric tandem repeats (magenta) are indicated as far as they are covered by uniquely mapping sequencing reads. Breaks are shown at the same scale as in Figure 3b.
Figure S5. Related to Figure 3. GLOE-Seq analysis of DNA lesions and repair intermediates in yeast
A. GLOE-Seq detects UV-induced pyrimidine dimers in yeast. Plots show relative frequencies of dinucleotide sequences adjacent to the detected strand breaks. Data are identical to those shown in Figure 3D, but visualised in a different manner to show the percentages of individual sequences.
B. Checkpoint activation and cell cycle profiles in response to MMS treatment. Western blot images show Rad53 and its phosphorylated forms as well as tubulin (loading control) in total extracts of WT and apn1Δ apn2Δ cells from the experiment shown in Figures 3D and E. Corresponding flow cytometry profiles are shown beneath each lane.
Figure S6. Related to Figure 6. GLOE-Seq analysis in human cells
A. Workflow for the preparation of mammalian DNA from agarose-embedded nuclei (see step-by-step protocol for details). LMP: low-melting point.
B. Size distribution and amount of DNA recovered after second strand synthesis in the core GLOE-Seq workflow (see Figure S1A) from HCT116 WT and LIG3⁻/⁻:mL3 cells treated with either an unspecific (CTR) or a Ligase 1-specific siRNA (LIG1).
C. Effects of downsampling on RFD plots of GLOE-Seq data from HCT116 cells under conditions of DNA Ligase 1 inactivation (WT + siLIG1).
Methods S1. Related to Star Methods.

Step-by-step protocol for the preparation of GLOE-Seq libraries

Introduction

GLOE-Seq is a next generation sequencing method for the genome-wide mapping of 3'-OH termini, either resulting from single- or double-strand breaks or introduced by enzymatic conversion of lesions or modified nucleotides. This protocol provides step-by-step instructions starting with the isolation of genomic DNA up to the amplification of libraries in preparation for sequencing and includes explanatory notes, quality controls and a troubleshooting guide. Separate instructions are provided for isolation of DNA from budding yeast, accomplished by gentle lysis of spheroplasts, and from mammalian cultured cells, where lysis is induced in agarose-embedded nuclei (steps 1-28). The subsequent procedure (steps 29-50) follows a common protocol that is independent of the source of DNA.

Table of Contents

| Section | Page |
|---------|------|
| A. Reagents and consumables | 2 |
| Commercial reagents | 2 |
| Consumables | 3 |
| Buffers and solutions | 3 |
| Oligonucleotides | 4 |
| B. Equipment | 4 |
| C. Protocol | 5 |
| Steps y1 – 22: Extraction of genomic DNA from budding yeast | 5 |
| Steps y23 – 28: Denaturation and ligation of 3'-OH termini (yeast) | 6 |
| Steps m1 – 21: Isolation of genomic DNA from mammalian cell culture | 6 |
| Steps m22 – 28: Denaturation and ligation of 3'-OH termini (mammalian cells) | 7 |
| Steps 29 – 37: Fragmentation and capture of biotinylated ssDNA | 8 |
| Steps 38 – 46: Second strand synthesis, end polishing and ligation of the distal adaptor | 8 |
| Steps 47 – 50: Library amplification | 9 |
| D. Notes | 10 |
| E. Quality control | 11 |
| F. Troubleshooting | 12 |
A. Reagents and Consumables

Commercial reagents

PEG 8000 (Sigma-Aldrich, 89510)
Zymolyase 20T (AMS Biotechnology, 120491-1)
β-mercaptoethanol (Sigma-Aldrich, M3148)
AMPure XP beads (Beckman Coulter, A63881)
RNase A (Sigma-Aldrich, 10109169001)
RNase H (New England Biolabs, M0288S)
Sorbitol (Sigma-Aldrich, S1876)
Sodium chloride (Sigma-Aldrich, S3014)
Sodium hydroxide (Sigma-Aldrich, S8045)
Sodium citrate (Sigma-Aldrich, W302600)
Ethanol (Fisher Scientific, 15643690)
Tris base (Sigma-Aldrich, T4661)
EDTA (Sigma-Aldrich, E6758)
20% (w/v) SDS (Sigma-Aldrich, 05030)
10× T4 DNA ligase buffer (New England Biolabs, B0202S)
T4 DNA ligase, 20,000,000 U/mL (New England Biolabs, M0202T)
Q5® High-Fidelity DNA polymerase with reaction buffer (New England Biolabs, M0491)
NEBNext® Ultra™ II DNA Library Prep kit for Illumina® (New England Biolabs, E7645)
Phusion Flash high-fidelity PCR master mix (Thermo Fisher Scientific, F-548)
Deoxynucleotide (dNTP) Solution Mix, 10 mM (New England Biolabs, N0447S)
Dynabeads™ MyOne™ Streptavidin C1 (Life Technologies, 65001)
Potassium acetate (VWR International, 236497-500G)
High Sensitivity D1000 ScreenTape (Agilent Technologies, 5067-5584)
High Sensitivity D1000 ScreenTape reagents (Agilent Technologies, 5067-5585)
RNA ScreenTape (Agilent Technologies, 5067-5576)
RNA ScreenTape sample buffer (Agilent Technologies, 5067-5577)
RNA ScreenTape ladder (Agilent Technologies, 5067-5578)
Qubit™ dsDNA HS Assay Kit (Invitrogen, Q32854)
Agilent Bioanalyzer High Sensitivity DNA Kit (Agilent Technologies, 5067-4646)
NextSeq 550 System High-Output Kit (Illumina, 20024906)
NuSieve™ GTG™ Agarose (Lonza, 859081)
Proteinase K (Roche, 3115801001)
Sarkosyl (Sigma-Aldrich, 61743)
β-Agarase I, 1000 U/mL (New England Biolabs, M0392S)
Trypan blue solution, 0.4% (w/v) (Thermo Fisher Scientific, 15250061)
Phenylmethanesulfonyl fluoride (PMSF) (Sigma-Aldrich, P7626)
Consumables

- Microtubes for Bioruptor® Pico, 0.65 mL (Diagenode, C30010011)
- Filter Tips 10/20 µL, 20 µL, 200 µL, 1000 µL (TipOne, S1120-3810, S1120-1810, S1120-8810, S1126-7810)
- Centrifuge Tubes (Corning® 50 mL PP, 430829)
- Microcentrifuge Tubes, 2 mL (Eppendorf, 0030120094)
- DNA LoBind Tubes, 1.5 mL (Eppendorf, 0030108051)
- Axygen 0.2 mL 8-strip PCR Tubes (Thermo Fisher Scientific, 14-222-252)
- Qubit Assay Tubes (Life Technologies, Q32856)
- Tubes, 12 mL (Sarstedt, 60.9922.937)

Buffers and solutions

| Buffer/Tube Type                  | Composition                                                                 |
|-----------------------------------|-----------------------------------------------------------------------------|
| Water                             | sterile, deionized (e.g. Milli Q-purified)                                  |
| 0.5 M EDTA                        | pH 8.0                                                                      |
| Y1 buffer                         | 1 M sorbitol, 100 mM EDTA, 14 mM β-mercaptoethanol (added immediately before use) |
| Lysis buffer                      | 50 mM Tris-HCl, pH 8.0, 50 mM EDTA                                          |
| TE buffer                         | 10 mM Tris-HCl, pH 8.0, 1 mM EDTA                                           |
| 5 M potassium acetate             |                                                                             |
| 1% (w/v) SDS                      |                                                                             |
| 70% (v/v) ethanol                 |                                                                             |
| 10 mM Tris-HCl                    | pH 8.5                                                                      |
| Bind & Wash buffer                | 10 mM Tris-HCl, pH 8.5, 2 M NaCl                                            |
| 50% (w/v) PEG 8000                |                                                                             |
| 5 M NaCl                          |                                                                             |
| 20 mM NaOH                        |                                                                             |
| 1× SSC                            | 150 mM sodium citrate, pH 7.0, 15 mM NaCl                                   |
| Nuclear isolation buffer          | 10 mM Tris, pH 8.0, 50 mM NaCl, 50 mM EDTA, 0.34 M sucrose, 10% (v/v) glycerol, 0.1% (v/v) Triton X-100 |
| PBS, pH 7.4                       | 137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄                      |
| Proteinase K solution             | 1 mg/mL Proteinase K, 1% (w/v) sarkosyl, 125 mM EDTA, pH 9.0                |
| Plug wash buffer                  | 10 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA                              |
Oligonucleotides:

Proximal adaptor: 40 µM oHU3898 (annealing programme: see below)
40 µM oHU3899
60 mM NaCl

Distal adaptor: 40 µM oHU3791 (annealing programme: see below)
40 µM oHU3792
60 mM NaCl

Oligonucleotides were HPLC-purified (IDT).

| Name | Sequence (5’ – 3’) |
|------|-------------------|
| 3898 | CTACACGACGCTCTTCCGATCTNNNN*N-NH₂ (*: phosphorothioate bond, IDT code *; NH₂: 3’-amino modification, IDT code /3AmMO/) |
| 3899 | PO₄-AGATCGGAAGAGCGTGTAGGAAAGAATGTAGATCTCGTTTT-Bio (PO₄: 5’-phosphorylation, IDT code /5Phos/; T-Bio: 3’-biotin-dT, IDT code /3BiodT/) |
| 3790 | CGAGATCTACACTCITTTCCCTACACGACGCTCTTCCGATCT |
| 3791 | GACTGGAGTTAGACGTGCTCTTCCGATCT |
| 3792 | GATCGGAAGAGACGCTCTGAACTCCAGTC |
| P5   | AATGATACGCAACCGATCTACACTCTTTCCCTACAGACGCTCTTCCGATCT |
| P7   | CAACAGAAGACGCCCATACGAGAT(X)₆GTGACTGGAGTTAGACGCTGCTCTTCCGATCT |

Annealing Programme:

| T (°C) | mins:s | ΔT (°C) | Ramp rate (°C/s) |
|--------|--------|---------|-----------------|
| 95     | 05:00  | -20     | 0.1             |
| 75     | 00:01  | -20     | 0.1             |
| 55     | 00:01  | -20     | 0.1             |
| 35     | 00:01  | -10     | 0.1             |
| 25     | ∞      | -       | 6.0             |

B. Equipment

Stereomicroscope (Leica DM1000 LED, Leica Biosystems)
Bioruptor® Pico (Diagenode, B01060010)
Tube holders for Bioruptor®, 0.5/0.65 mL (Diagenode, B01200043)
Agilent 2200 TapeStation System (Agilent Technologies, G2964AA)
Agilent 2100 Bioanalyzer Instrument (Agilent Technologies, G2939AA)
Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, cat. #Q32866)
Thermocycler (Biometra TRIO, 070-723)
Magnetic rack (DynaMag™-2 Magnet, Thermo Fisher Scientific, 12321D)
Rotating wheel (Stuart, SB2)
Benchtop Centrifuge (Heraeus™ Multifuge™ X3R, VWR International, 97040-234)
Refrigerated Centrifuge (Heraeus Fresco 21, Thermo Fisher Scientific, 75002555)
Centrifuge (Sorvall™ RC 6 Plus, Thermo Fisher Scientific, 36-101-0816)
Illumina NextSeq 550 System
C. Protocol

Steps y1 – 22: Extraction of genomic DNA from budding yeast

1. Collect ca. 6×10⁹ cells (300 OD₆₀₀) by centrifugation in a 50 mL tube (2,000×g, 5 min), resuspend the pellet in 25 mL of 10 mM Tris-HCl pH 8.5, and centrifuge again. Discard the supernatants.

2. Resuspend the pellet in 4 mL of Y1 buffer and add 400 µL of Zymolyase 20T (10 mg/mL). Incubate at 30°C and monitor formation of spheroplasts under a microscope.

3. Pellet the spheroplasts at 2,000×g for 5 min, wash once with 4 mL of Y1 buffer without β-mercaptoethanol and resuspend in 4.75 mL of lysis buffer. Initiate lysis by adding 250 µL of 20% SDS and incubating at 37°C until the solution turns transparent.

4. Centrifuge the sample at 5,000×g for 10 min and transfer the supernatant (by pouring) to a fresh 50 mL tube.

5. Add 1.666 mL of 5 M potassium acetate, invert the tube to mix and incubate on ice for 45-60 min.

6. Centrifuge the sample at 17,000×g for 15 min at room temperature, transfer the supernatant (by pouring) to a fresh 50 mL tube and centrifuge again at 17,000×g for 15 min.

7. Transfer the supernatant to a fresh 50 mL tube, add 19.5 mL of 100% ethanol and swirl the tube to precipitate the DNA.

8. Centrifuge the sample at 8,000×g for 15 min and discard the supernatant.

9. Rinse the pellet by adding 25 mL of 70% ethanol and centrifuging again at 8,000×g for 10 min. Discard the supernatant.

10. Add 4 mL of 70% ethanol to the pellet and distribute the suspension with a wide-bore pipette (cut the tip off a 1 mL pipette tip) into two 2 mL microcentrifuge tubes.

11. Centrifuge at 21,100×g for 10 min and discard the supernatants.

12. Air-dry the pellets and resuspend in 245 µL of TE each by incubation overnight at 4°C.

13. Add 5 µL of 5 M NaCl and 2.5 µL of RNase A to each DNA sample and incubate at 37°C for 1 h.

14. Add 1.5 mL of 100% ethanol to each tube, invert to mix, centrifuge at 21,100×g for 10 min and discard the supernatant.

15. Rinse the pellets once with 1 mL of 70% ethanol.

16. Air-dry the pellets and resuspend overnight at 4°C by gentle rocking in 250 µL of 10 mM Tris-HCl pH 8.5.

17. Add 250 µL AMPure beads to a 1.5 mL DNA LoBind microcentrifuge tube and use a magnetic rack to remove 125 µL of the storage buffer.

18. Add the DNA sample (250 µL) to the bead suspension and incubate at room temperature for 10 min.

19. Using a magnetic rack, remove and discard the supernatant.

20. While keeping the tube on the magnetic rack, rinse the beads twice with 500 µL of 70% ethanol each.

21. Remove the supernatant and add 125 µL of 10 mM Tris-HCl pH 8.5.

22. Elute the DNA by incubating at room temperature for 10 min and collecting the supernatant using a magnetic rack.
#Pause point – At this point, the extracted genomic DNA can be stored at 4°C for up to 6 months. When mapping lesions not involving strand breaks, additional treatments (followed by renewed purification using AMPure beads) are applied at this stage to generate the corresponding 3′-OH termini, e.g.:

- pyrimidine dimers: T4 Endonuclease V + AP endonuclease (APE1)
- abasic (AP) sites: AP endonuclease (APE1)
- ribonucleotides: RNAse H

Steps y23 – 28: Denaturation and ligation of 3′-OH termini (yeast)

23. Incubate 2.5 µg of genomic DNA (treated as described above if relevant) at 95°C for 10 min, followed by incubation on slushy ice for 5 min to denature the DNA.

24. Set up the ligation reaction in a PCR tube in the specified order at room temperature:

| Component                      | Volume |
|--------------------------------|--------|
| Denatured DNA                 | 2.5 µg |
| 10× T4 DNA ligase buffer      | 6.5 µL |
| Proximal adaptor              | 3.55 µL|
| 50% PEG 8000                  | 19.5 µL|
| T4 DNA ligase                 | 3 µL   |

(Nota 7) Adjust the total volume to 65 µL with water and incubate in a thermocycler as follows:

| T (°C) | min:s |
|--------|-------|
| 25     | 60:00 |
| 22     | 60:00 |
| 22     | 60:00 |
| 16     | ∞     |

25. Add 100 µL of AMPure beads to a 1.5 mL DNA LoBind microcentrifuge tube and remove the storage buffer, using a magnetic rack.

26. Add 35 µL of 5 M NaCl and 35 µL of water to the ligation mix. Transfer the entire mix to the tube containing the AMPure beads.

27. Incubate at room temperature for 5 min. Using a magnetic rack, remove the supernatant from the beads and rinse the beads twice with 500 µL of 70% ethanol while keeping the tube on the rack.

28. Elute the DNA from the beads by adding 103 µL of water and incubation at room temperature for 5 min. Transfer 100 µL to a fresh tube, using a magnetic rack.

#Pause point – At this point, the ligated DNA fragments can be stored at -20°C for up to 3 days.

Steps m1 – 21: Isolation of genomic DNA from mammalian cell culture

1. Harvest up to 10 million cells by trypsinization.
2. Wash the cells twice in ice-cold PBS + 5 mM EDTA pH 8.0.
3. Resuspend cells in 1 mL of nuclear isolation buffer and incubate them on ice for 10 min to dissolve the plasma membrane.
4. Harvest nuclei by centrifugation in a swing-out rotor at 1,200×g for 3 min at 4°C and wash them once with 1 mL of nuclear isolation buffer. Loosen the pellet by inversion or tapping, rather than pipetting.
5. Mix the nuclei with nuclear isolation buffer supplemented with RNase A at a final concentration of 100 µg/µL at a ratio of 100 µL of buffer per 20 µL of nuclear pellet and incubate at 37°C for 15 min.

6. Count the nuclei with a haemocytometer in the presence of Trypan blue (0.2% (w/v) final).

7. To prepare agarose plugs, harvest the desired number of nuclei (~700,000 per plug) by centrifugation in a swing-out rotor at 1,200×g for 3 min.

8. Remove all but ~10 µL of supernatant, taking care not to disturb the pellet of nuclei.

9. Resuspend the nuclei (by tapping) at a density of about 700,000 nuclei per 22.5 µL of suspension in PBS + 25 mM EDTA pH 8.0, prewarmed to 50°C.

10. Mix the suspension 1:1 (v/v) with molten 1.2% (w/v) low-melting point agarose prepared in PBS + 25 mM EDTA pH 8.0, equilibrated to 50°C.

11. Immediately pipette the cell-agarose mixture into the wells of a 45 µL plug mould. (Note 8)

12. Allow the agarose to solidify for 1 h at 4°C.

13. Combine the plugs into 12 mL tubes filled with 7-8 mL of Proteinase K solution and incubate overnight at 42°C (no more than 3 plugs per tube).

14. Exchange the Proteinase K solution for a fresh aliquot of 7-8 mL and perform a second overnight incubation at 42°C.

15. Wash the plugs three times with 7-8 mL of plug wash buffer + 1 mM PMSF for 10 min each.

16. Wash the plugs once with 7-8 mL of plug wash buffer + 1 mM PMSF for 1 h.

17. Wash the plugs once with 7-8 mL of plug wash buffer for 1 h.

18. Transfer each plug to a separate 1.5 mL DNA LoBind tube and wash twice with 1 mL of 10 mM Tris-HCl pH 8.0 + 0.1 mM EDTA pH 8.0.

19. Wash each plug once with 1 mL of 10 mM Tris-HCl pH 8.0 + 0.1 mM EDTA pH 8.0 for 30 min.

20. Wash each plug twice with 1 mL of 0.1 mM EDTA pH 8.0 for 15 min each.

21. Quickly rinse the plugs with water and remove as much residual water as possible.

**Pause point** – At this point, the plugs can be stored at 4°C for up to 3 days. When mapping lesions not involving strand breaks, additional treatments could be applied at this stage to generate the corresponding 3’-OH termini, e.g.:

- pyrimidine dimers: T4 Endonuclease V + AP endonuclease (APE1)
- abasic (AP) sites: AP endonuclease (APE1)
- ribonucleotides: RNAase H

Inactivation of the enzymes after this step could be accomplished by heat denaturation in the presence of EDTA of by Proteinase K treatment.

**Steps m22 – 28: Denaturation and ligation of 3’-OH termini (mammalian cells)**

22. Denature the DNA by incubation at 95°C for 4 min and quenching on slushy ice for 5 min.

23. Melt the agarose at 65°C for 5 min and equilibrate to 37°C for 15 min.

24. Thoroughly mix the agarose, by tapping, with 17 µL of a prewarmed ligation mix consisting of:
10× T4 DNA ligase buffer 6.5 µL
Proximal adaptor 3.5 µL
Water 7 µL
and with 3 µL of T4 DNA ligase (added last).

25. After an overnight incubation at 16°C, supplement the ligation reaction with 2 µL of 0.5 M EDTA pH 8.0 and 30 µL of water.

26. Incubate the mixture at 65°C for 10 min to inactivate the ligase and melt the agarose.

27. Allow the agarose to equilibrate to 42°C for 20 min, supplement with 3 µL of β-agarase and incubate at 42°C for 3-4 h.

28. Adjust the sample volume to 100 µL with water and transfer to a 0.5 mL Bioruptor tube.

**Steps 29 – 37: Fragmentation and capture of biotinylated ssDNA**

29. Sonicate the eluate from step 28 (100 µL) to a fragment size of 200 nt, using a Bioruptor Pico (4 cycles for yeast DNA, 16 cycles for mammalian DNA, 30 s on/30 s off). (QC sample 1) (Note 9)

30. Clean up the DNA by two rounds of binding to an equivalent of 100 µL of AMPure beads, each time adding NaCl and PEG 8000 to final concentrations of 0.9 M and 7.5% (w/v), respectively, and elution in 100 µL of water as described in steps 27-28. (Note 10)

31. Transfer the sonicated DNA to a microfuge tube containing 20 µL of Streptavidin MyOne C1 Dynabeads (prewashed once with 1 mL Bind & Wash buffer).

32. Incubate for 15 min at room temperature on a rotating wheel.

33. Pulse-spin the tube and collect the beads on a magnetic rack. Remove the supernatant.

34. Wash the beads twice with 50 µL of 1× SSC buffer each for 5 min at room temperature on a rotating wheel. (Note 11)

35. Wash the beads once with 25 µL of 20 mM NaOH for 10 min at room temperature on a rotating wheel.

36. Wash the beads briefly with 100 µL water.

37. Elute the DNA by adding 16 µL of water, incubating at 95°C for 5 to 10 min and collecting the supernatant using the magnetic rack. (QC sample 2) (Note 12)

**Steps 38 – 46: Second strand synthesis, end polishing and ligation of the distal adaptor**

38. Combine the following reagents in a PCR tube to perform the second strand synthesis reaction:

| Eluate from step 37 | 14.85 µL |
| Phusion FLASH mix | 15 µL |
| 100 µM oHU3790 | 0.15 µL |

Incubate the reaction mixture in a thermocycler with the following programme:

| T (°C) | min:s |
|--------|-------|
| 95     | 02:00 |
| 60     | 00:30 |
| 72     | 02:00 |
| 4      | ∞     |

(Note 13)
39. Purify the resulting double-stranded DNA using 54 µL of AMPure beads in their supplied buffer as described in step y27-28. (Note 14)

40. Release the DNA by addition of 19 µL of water to the washed AMPure beads and incubation for 5 min at room temperature. Do not remove the eluate from the beads. (QC sample 3)

41. Add the following reagents (from the NEBNext® Ultra™ II DNA Library Prep kit) to perform the end polishing reaction:
   - Eluate (bead suspension) 17 µL
   - End Prep Reaction Buffer 2.3 µL
   - End Prep Enzyme Mix 1 µL

   Incubate the reaction mixture in a thermocycler with the following programme:

   | T (°C) | min:s |
   |--------|-------|
   | 20     | 30:00 |
   | 65     | 30:00 |
   | 4      | ∞     |

42. Add the following reagents from the NEBNext® Ultra™ II DNA Library Prep kit to perform the distal adaptor ligation reaction:
   - Ligation Master Mix 13.5 µL
   - Ligation Enhancer 0.45 µL
   - Distal adaptor 2.25 µL
   - Water 8.8 µL

   Incubate the reaction mixture in a thermocycler with the following programme:

   | T (°C) | min:s |
   |--------|-------|
   | 20     | 20:00 |
   | 4      | ∞     |

43. Add the following reagents:
   - 5 M NaCl 13.5 µL
   - 50% PEG 8000 11.25 µL
   - Water 5.25 µL

44. Rinse the beads as described in step y27.

45. Elute the purified DNA from the beads in 50 µL of water as described in step y28 and repeat the purification as described in steps y27-28, using 80 µL of fresh AMPure beads in their supplied buffer.

46. Elute the DNA from the beads in 20 µL of water by incubation for 5 min at room temperature.

**#Pause point** – At this point, the ligated DNA fragments can be stored at -80°C.

**Steps 47 – 50: Library amplification**

47. Add the following reagents to a PCR tube to amplify the libraries:
DNA (from step 46)  7.25 µL
5× Q5 reaction buffer  5 µL
1 µM P5  2.5 µL
1 µM P7  2.5 µL
10 mM dNTP solution  0.5 µL
Q5 DNA polymerase  0.25 µL
Water  7 µL

Perform 8 cycles of PCR amplification, using the following programme:

| T (°C) | min:s |
|--------|-------|
| 95     | 02:00 |
| 95     | 00:15 |
| 60     | 00:30 |
| 72     | 00:20 |
| 4      | ∞     |

48. Purify the amplified library using 25 µL of AMPure beads in their supplied buffer as described in step 27 and elute in 25 µL of water as described in step 28.

49. Repeat the purification (step 48) and elute with 20 µL of water.  (QC sample 4) (Note 16)

50. Determine the concentration the library sample using a Qubit fluorometer and reagents for quantifying dsDNA and prepare a sample pool with a total concentration of 4 nM for loading onto the sequencer.  (Note 17)

D. Notes

Note 1 – The protocol described here is recommended for preparing genomic DNA for detection of strand breaks, as it minimizes the introduction of additional nicks and breaks during the extraction procedure. For detection of DNA lesions involving an intact backbone, steps 1-22 can be replaced by the use of a standard genomic DNA extraction kit (e.g. Qiagen). The protocol can be scaled.

Note 2 – Spheroplast formation takes ~60-90 min. Its efficiency is monitored by adding 2 µL of 1% SDS to 2 µL of cell suspension on a microscope slide and observing the decrease in the number of intact cells under a stereomicroscope. Samples not treated with SDS can be used as a control. Efficient spheroplasting should result in close to 100% lysis within 2 min of incubation with SDS.

Note 3 – This should normally take about 45 min.

Note 4 – Repeat this step if the supernatant is not cleared of debris.

Note 5 – When handling more than 4 samples, pellets may detach from the wall of the tube over time. In this case, repeat the centrifugation to clear the supernatant of any visible debris.

Note 6 – If the DNA pellet does not resuspend well, it can be incubated overnight in the 50 mL tube and split into two 2 mL microcentrifuge tubes the next day before RNase A treatment. After splitting the original sample into two tubes, each of these is treated separately and in parallel. Hence, each sample of yeast cells eventually gives rise to two identical samples of genomic DNA. Alternatively, the initial steps can be scaled down to prepare a single sample of DNA.
Note 7 – The amount of adaptor can be calculated based on the approximate number of nicks expected in 2.5 μg of genomic DNA. A five-fold molar excess of adaptor should then be used. If the expected number of breaks is unknown, use 3.55 μL (resulting in a final concentration of 2.185 μM). If the excess of unligated adaptor presents a problem (e.g. in native, undigested DNA), this amount can be reduced.

Note 8 – A custom mould with pockets of ~45 μL is necessary to cast properly shaped plugs of 45 μL. Common commercially available moulds have pockets of ~90 μL (Bio-Rad, 1703713), which produce unevenly shaped plugs that easily break during handling. See Supplementary Information for a 3D printer template suitable for 45 μL moulds.

Note 9 – Make sure the Bioruptor Pico is cooled to 4°C before sonication. Incubation of the purified samples on ice is not required.

Note 10 – This additional clean-up step can be omitted in the yeast protocol, but at least one round of purification should be carried out if unligated adaptors present a problem (see Troubleshooting).

Note 11 – When processing more than 5 samples in parallel, reduce the incubation time to 2.5 min in order to account for the increased processing time due to buffer addition and resuspension of the samples.

Note 12 – Incubating the samples for 10 min does not affect the subsequent steps.

Note 13 – When performing more than one reaction, a master mix should be prepared using a 100 μM stock of oHU3790 in order to keep the total reaction volume as close to 30 μL as possible.

Note 14 – If a sample is expected to contain few nicks, sample volume can be increased to 50 μL by addition of water, and 90 μL of AMPure beads should then be used for purification. This minimises loss of library.

Note 15 – Oligonucleotide P7 contains a barcode, (X)6, for multiplexing, which should be applied according to the number of samples to be included in the experiment.

Note 16 – While removing the eluate, collect only 18 μL and pipette very slowly to avoid transferring any beads along with the supernatant.

Note 17 – Sequencing conditions will depend on the available platform (e.g. Illumina NextSeq) as well as the nature and number of the samples to be analysed. We recommend following the manufacturer’s specifications for single-end sequencing. For GLOE-Seq analysis in yeast, 3 million reads per sample are sufficient. For analysis of mammalian cells, a minimum of 50 million reads is recommended.

E. Quality Control

QC sample 1 – In order to verify appropriate fragmentation to an average fragment size of 200 nt, analyse 1 μL of the fragmented DNA either on an Agilent RNA ScreenTape (Agilent 2200 Tapestation) or an Agilent 2100 Bioanalyzer (using an RNA chip).

QC sample 2 – In order to verify efficient capture of biotinylated DNA and removal of excess unligated adaptor, analyse 1 μL of the eluted DNA either on an Agilent RNA ScreenTape (Agilent 2200 Tapestation) or an Agilent 2100 Bioanalyzer (using an RNA chip).

QC sample 3 – The efficiency of second strand synthesis and ligation of the distal adaptor should be monitored by analysing 2 μL of the eluate on an Agilent High Sensitivity D1000 ScreenTape (Agilent 2200 tape station).

QC sample 4 – In order to ensure that the sample is free of unligated adaptors or adaptor concatemers, analyse 1 μL of the purified DNA on an Agilent High Sensitivity DNA chip (Agilent 2100 Bionalyzer).
## Troubleshooting

| QC Sample | Expected Result | Possible Deviation | Possible Problem | Action |
|-----------|----------------|--------------------|-----------------|--------|
| 1         | ssDNA in a size range of 160-450 nt (average: 200) | Improper size range | Improper fragmentation conditions | Adjust/standardise fragmentation conditions. |
| 2         | ssDNA in a size range of 100-300 nt (average: 200), free of unligated adaptor | No product | Sample contains very few nicks | No action required. |
|           | Presence of unligated adaptor | Inefficient capture of biotinylated DNA | Verify the quality of Streptavidin beads using unligated proximal (biotinylated) adaptor as control. Make sure you use a “3’-Biotin-dT” label (IDT) on the adaptor. “Standard Biotin” has caused problems in our hands. |
| 3         | dsDNA in a size range of 150-450 bp (average: 350) | No product | Inefficient 2nd strand synthesis due excess unligated adaptor | Additional purification using AMPure beads will remove excess adaptor. |
| 4         | Library in a size range of 250-700 bp, free of adaptor concatemers | Presence of adaptor concatemers | Presence of excess unligated proximal and distal adaptors | Additional purification using AMPure beads at a beads: sample ratio of 1:1 (v/v) will remove adaptor concatemers. |