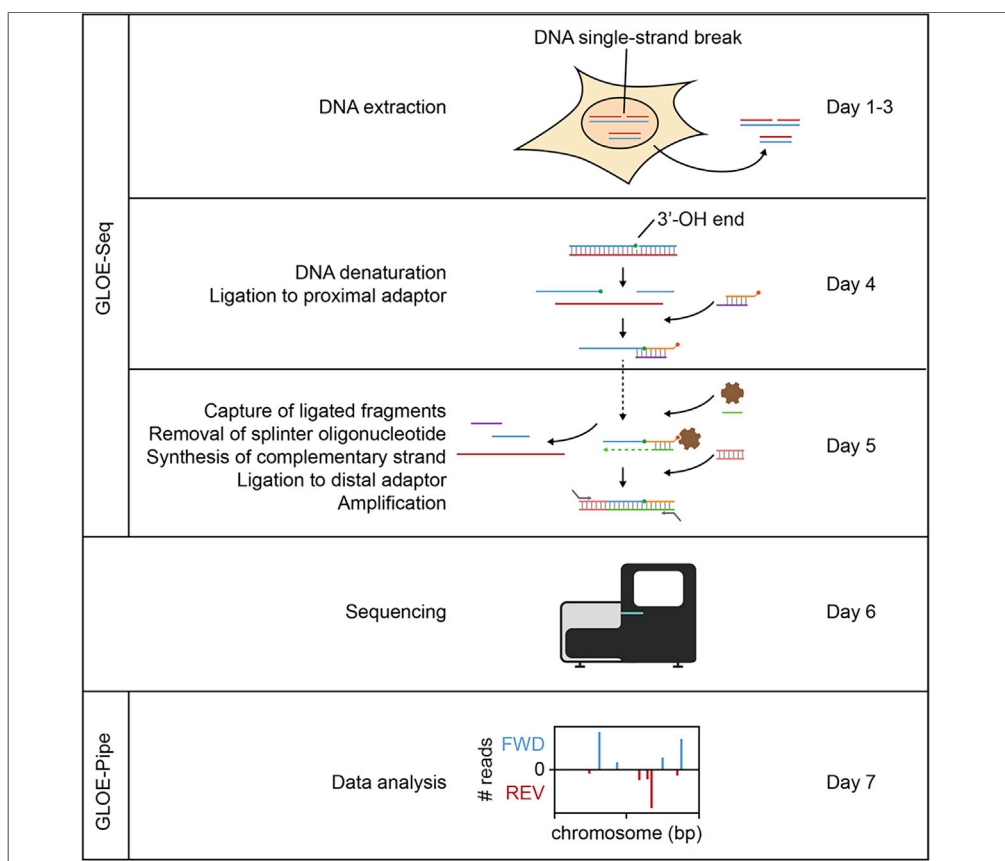Protocol

Preparation and Analysis of GLOE-Seq Libraries for Genome-Wide Mapping of DNA Replication Patterns, Single-Strand Breaks, and Lesions

Giuseppe Petrosino, Nicola Zilio, Annie M. Sriramachandran, Helle D. Ulrich

h.ulrich@imb-mainz.de

HIGHLIGHTS

GLOE-Seq is a sequencing-based method for the genome-wide mapping of 3′-OH termini

GLOE-Seq is useful for mapping single-strand breaks and base lesions or modifications

Protocols are provided for library preparation from budding yeast and mammalian cells

Data analysis is accomplished via a dedicated computational pipeline, GLOE-Pipe

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 instructions for isolation of genomic DNA from budding yeast or mammalian cells, preparation of libraries for sequencing, and data analysis by the associated computational pipeline, GLOE-Pipe. It is optimized for the Illumina next-generation sequencing platform and can be adapted to intact genomic DNA of any origin.

Protocol

Preparation and Analysis of GLOE-Seq Libraries for Genome-Wide Mapping of DNA Replication Patterns, Single-Strand Breaks, and Lesions

Giuseppe Petrosino,[1,2,3] Nicola Zilio,[1,2,3] Annie M. Sriramachandran,[1] and Helle D. Ulrich[1,3,4,*]

[1]Institute of Molecular Biology (IMB), Ackermannweg 4, D – 55128 Mainz, Germany

[2]These authors contributed equally

[3]Technical Contact

[4]Lead Contact

*Correspondence: h.ulrich@imb-mainz.de

https://doi.org/10.1016/j.xpro.2020.100076

SUMMARY

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 instructions for isolation of genomic DNA from budding yeast or mammalian cells, preparation of libraries for sequencing, and data analysis by the associated computational pipeline, GLOE-Pipe. It is optimized for the Illumina next-generation sequencing platform and can be adapted to intact genomic DNA of any origin.

For complete details on the use and execution of this protocol, please refer to Sriramachandran et al. (2020).

BEFORE YOU BEGIN

Preparation of Adaptors

⊙ Timing: 90 min

1. Proximal adaptor – prepare the following solution in a PCR tube:

Reagent	Final Concentration	Volume
Oligonucleotide 3898 (100 μM), splinter	40 μM	40 μL
Oligonucleotide 3899 (100 μM)	40 μM	40 μL
NaCl (5 M)	60 mM	1.2 μL
ddH₂O	N/A	18.8 μL
Total	N/A	100 μL

2. Distal adaptor – prepare the following solution in a PCR tube:

Reagent	Final Concentration	Volume
Oligonucleotide 3791 (100 μM)	40 μM	40 μL
Oligonucleotide 3792 (100 μM)	40 μM	40 μL

(Continued on next page)

Continued

Reagent	Final Concentration	Volume
NaCl (5 M)	60 mM	1.2 μL
ddH$_2$O	N/A	18.8 μL
Total	N/A	100 μL

Note: Oligonucleotides were purchased as HPLC-purified reagents. Values given in the tables above will yield stock solutions sufficient for 25 samples.

3. Anneal the oligonucleotides by placing both tubes containing the adaptors into a thermocycler and applying the following program:

Reaction Conditions					
Steps	Temperature	Time	ΔT (°C)	Ramp (°C s^{-1})	Cycles
Denaturation	95°C	5 min	-20	0.1	1
Annealing	75°C	1 s	-20	0.1	1
	55°C	1 s	-20	0.1	1
	35°C	1 s	-10	0.1	1
Hold	25°C	Forever			

Preparation of Low-Melting-Point Agarose and Casting Molds

⏱ Timing: 20 min

Note: This section only applies to the protocol for mammalian DNA isolation in agarose plugs.

Note: For the isolation of large intact DNA, for example from mammalian cultured cells, nuclei are embedded in agarose plugs for the lysis procedure. This is intended to prevent additional breaks and nicks during DNA isolation and is particularly important when analyzing native strand breaks as opposed to base modifications. A custom mold with pockets of ~45 μL is necessary to cast properly shaped plugs of 45 μL and needs to be produced in advance. Common commercially available molds have pockets of ~90 μL (e.g. Bio-Rad, 1703713), which produce unevenly shaped plugs that easily break during handling. A 3D printer template suitable for 45 μL molds and a dedicated extrusion tool is available as Supplemental Information.

4. Prepare aliquots of 1.2% (*w/v*) low-melting point agarose:
   a. Mix the following reagents in a 50 mL glass flask:

Reagent	Final Concentration	Volume/Amount
Low-melting point agarose	1.2 % (*w/v*)	0.12 g
PBS (5×)	1×	2 mL
EDTA, 0.5 M, pH 8.0	25 mM	0.5 mL
ddH$_2$O	N/A	7 mL
Total	N/A	10 mL

b. Heat the mixture in a microwave oven at 30% maximum power until it starts boiling.

c. Swirl the flask until the agarose has completely melted.

d. Allow the mixture to cool down for 10 min.

e. Prepare 1-mL aliquots in 1.5 mL tubes and store them at 4°C.

5. Prepare the custom-made casting molds:

a. Take a piece of aluminum foil sealing film and peel off the protective sheet to expose just enough adhesive surface to cover the bottom of an entire plug casting mold.

b. Push a dry mold onto the exposed adhesive side of the aluminum film.

c. With a pair of scissors, cut the aluminum film around the mold. A small excess of film can be left on one of the two shorter sides of the mold and folded over to facilitate peeling it off later.

d. Using your thumbs, firmly press the aluminum film against the mold to ensure that the seal is completely tight.

Preparation of Starting Material

⏱ Timing: variable

6. Obtain fresh cell cultures as a starting material for the GLOE-Seq procedure.

Note: The type and amount of starting material will vary depending on the experiment. In this protocol, we describe the procedure for samples of ca. 6×10^9 budding yeast cells (300 OD_{600}) or ca. 10^7 mammalian cultured cells, respectively. These amounts will be sufficient for most mapping purposes. However, the protocol is scalable and can be adapted for other species or tissues by modifying the DNA extraction method.

⚠ CRITICAL: If mapping of strand breaks rather than base lesions or modifications is planned, frozen cells should only be used in combination with a cryoprotectant to avoid the introduction of undesired strand breaks.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
2-mercaptoethanol	Sigma-Aldrich	Cat# M3148
Agarose, NuSieve GTG	Lonza	Cat# 859081
β-Agarase I, 1,000 U mL^{-1}	New England Biolabs	Cat# M0392
Deoxynucleotide triphosphate (dNTP) solution, 10 mM	New England Biolabs	Cat# N0447S
Ethylene diaminetetraacetic acid (EDTA)	Sigma-Aldrich	Cat# E6758
Ethanol, absolute, ≥99.8%	Fisher Scientific	Cat# 15643690
Glycerol	Sigma-Aldrich	Cat# G5516-500ML
Phenylmethanesulfonyl fluoride	Sigma-Aldrich	Cat# P7626
Polyethylene glycol (PEG) 8000	Sigma-Aldrich	Cat# 89510
Potassium acetate	VWR International	Cat# 236497
Potassium chloride	Sigma-Aldrich	Cat# P9541-1KG
Proteinase K	Roche	Cat# 3115801001
NextSeq PhiX Control Kit	Illumina	Cat# FC-110-3002
RNase A from bovine pancreas	Sigma-Aldrich	Cat# 10109169001
RNase H	New England Biolabs	Cat# M0288S
Sarkosyl	Sigma-Aldrich	Cat# 61743

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Sodium chloride	Sigma-Aldrich	Cat# S3014
Sodium citrate	Sigma-Aldrich	Cat# W302600
Sodium dodecyl sulfate (SDS), 20% (*w/v*) in ddH$_2$O	Sigma-Aldrich	Cat# 05030
Sodium hydroxide	Sigma-Aldrich	Cat# S8045
Sorbitol	Sigma-Aldrich	Cat# S1876
Sucrose	Sigma-Aldrich	Cat# S0389-1KG
T4 DNA ligase, 20,000,000 U mL^{-1}	New England Biolabs	Cat# B0202
T4 DNA ligase buffer	New England Biolabs	Cat# B0202S
Tris base	Sigma-Aldrich	Cat# T4661
Triton X-100	Sigma-Aldrich	Cat# T9284-500ML
Trypan blue solution, 0.4% (*w/v*)	Thermo Fisher Scientific	Cat# 15250061
Zymolyase 20T	AMS Biotechnology	Cat# 120491-1
Critical Commercial Assays		
Agilent Bioanalyzer High Sensitivity DNA Kit	Agilent Technologies	Cat# 5067-4646
AMPure XP beads	Beckman Coulter	Cat# A63881
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
NEBNext Ultra II DNA Library Prep Kit for Illumina	New England Biolabs	Cat# E7645
NextSeq 500/550 High Output Kit v2.5	Illumina	Cat# 20024906
PhiX Control v3	Illumina	Cat# FC-110-3001
Phusion Flash high-fidelity PCR master mix	Thermo Fisher Scientific	Cat# F-548
Qubit dsDNA HS Assay Kit	Invitrogen	Cat# Q32854
RNA ScreenTape	Agilent Technologies	Cat# 5067-5576
RNA ScreenTape sample buffer	Agilent Technologies	Cat# 5067-5577
RNA ScreenTape ladder	Agilent Technologies	Cat# 5067-5578
Deposited Data		
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/
GLOE-Pipe/test_data	Sriramachandran et al., 2020	https://gitlab.com/GPetrosino/GLOE-Pipe/-/tree/master/test_data
Oligonucleotides		
Primer #3898: CTACACGACGCTCTTCCGATCT NNNNN*N-NH$_2$ (*: phosphorothioate bond, IDT code *; NH$_2$: 3'-amino modification, IDT code /3AmMO/)	Integrated DNA Technologies	N/A
Primer #3899: PO$_4$-AGATCGGAAGAGCGTCGTGTAGG GAAAGAGTG TAGATCTCGTTTTT-Bio (PO$_4$: 5'-phosphorylation, IDT code /5Phos/; T-Bio: 3'-biotin-dT, IDT code /3BiodT/)	Integrated DNA Technologies	N/A
Primer #3790: CGAGATCTACACTCTTTCCCTACAC GACGCTCTTCCGATCT	Integrated DNA Technologies	N/A
Primer #3791: GACTGGAGTTCAGACG TGTGCTCTTCCGATCT	Integrated DNA Technologies	N/A

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Primer #3792: GATCGGAAGAGCACA CGTCTGAACTCCAGTC	Integrated DNA Technologies	N/A
Primer P5: AATGATACGGCGACCAC CGAGATCTACACTCTTTCCCTACACG ACGCTCTTCCGATCT	Integrated DNA Technologies	N/A
Primer P7: CAAGCAGAAGACG GCATACGAGAT(X)$_6$GTGACTG GAGTTCAGACGTGTGC TCTTCCGATCT	Integrated DNA Technologies	N/A
Software and Algorithms		
bcl2fastq, version 2.19	Illumina	https://support.illumina.com/sequencing/ sequencing_ software/bcl2fastq-conversion-software.html
bedGraphToBigWig, version 365	Kent et al., 2010	https://github.com/ENCODE-DCC/kentUtils
BEDTools, version 2.25.0	Quinlan and Hall, 2010	https://bedtools.readthedocs.io/en/latest/
Bowtie 2, version 2.3.4	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Bpipe, version 0.9.9.3	Sadedin et al., 2012	http://docs.bpipe.org/
Cairo, version 1.15-10	N/A	https://cran.r-project.org/web/packages/Cairo/index.html
ChIPseeker, version 1.14.1	Yu et al., 2015	https://bioconductor.org/packages/release/bioc/html/ChIPseeker.html
deepTools, version 3.1.0	Ramírez et al., 2014	https://deeptools.readthedocs.io/en/develop/
FastQC, version 0.11.5	Andrews, 2019	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
GLOE-Pipe, version 1.0	Sriramachandran et al., 2020	https://gitlab.com/GPetrosino/GLOE-Pipe
Lmod, version 6.6	Geimer et al., 2014	https://lmod.readthedocs.io
MACS2 version 2.1.1	Zhang et al., 2008	https://github.com/taoliu/MACS
Openxlsx, version 4.1.0	N/A	https://cran.r-project.org/web/packages/openxlsx/index.html
R, version 3.5.1	R Core Team, 2019	https://www.r-project.org/
regioneR, version 1.14.0	Gel et al., 2016	https://bioconductor.org/packages/release/bioc/html/regioneR.html
rtracklayer, version 1.42.2	Lawrence et al., 2009	https://bioconductor.org/packages/release/bioc/html/rtracklayer.html
Samtools, version 1.5	Li et al., 2009	http://samtools.sourceforge.net/
Trimmomatic, version 0.36	Bolger et al., 2014	http://www.usadellab.org/cms/?page=trimmomatic
Other		
3D model of a custom-made mold for agarose plugs	Sriramachandran et al., 2020	See Supplemental Information
3D model of a tool for extrusion of agarose plugs from custom-made mold	Sriramachandran et al., 2020	See Supplemental Information
Agilent 2200 TapeStation System	Agilent Technologies	Cat# G2964AA
Agilent 2100 Bioanalyzer Instrument	Agilent Technologies	Cat# G2939AA
Aluminum foil sealing film, e.g. AlumaSeal 96 film	Sigma-Aldrich	Cat# Z721549
Axygen 8-Strip PCR Tubes, 0.2 mL	Thermo Fisher Scientific	Cat# 14-222-252
Benchtop centrifuge, Heraeus Multifuge X3R	VWR International	Cat# 97040-234

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bioruptor Pico	Diagenode	Cat# B01060010
Centrifuge, Sorvall RC6 Plus	Thermo Fisher Scientific	Cat# 36-101-0816
Centrifuge Tubes, 12 mL	Sarstedt	Cat# 60.9922.937
Centrifuge Tubes, 50 mL	Corning	Cat# 430829
Counting chamber, Neubauer	neoLab	Cat# CE | C-1003
DNA LoBind Tubes, 1.5 mL	Eppendorf	Cat# 003010851
DynaMag-2 Magnet	Thermo Fisher Scientific	Cat# 12321D
Filter Tips, 10/20 μL, 20 μL, 200 μL and 1,000 μL	TipOne	Cat# S1120-3810, - 1810, -8810, -7810
Microcentrifuge Tubes, 2 mL	Eppendorf	Cat# 0030120094
Microtubes for Bioruptor Pico, 0.65 mL	Diagenode	Cat# C30010011
Qubit 2.0 Fluorometer	Thermo Fisher Scientific	Cat# Q32866
Qubit Assay Tubes	Life Technologies	Cat# Q32856
Refrigerated centrifuge, Heraeus Fresco 21	Thermo Fisher Scientific	Cat# 75002555
Rotating wheel	Stuart	Cat# SB2
Sequencer, Illumina NextSeq 500	Illumina	N/A
Stereomicroscope	Leica Biosystems	Cat# DM1000 LED
Thermocycler	Biometra	Cat# 070-723
Tube holders for Bioruptor, 0.5/0.65 mL	Diagenode	Cat# B01200043

Alternatives: Most chemicals and instruments can be purchased from alternative suppliers as long as the specifications match. However, we have not tested substitutes for most of them. We recommend using the indicated sources for all enzymes and critical commercial assay kits.

MATERIALS AND EQUIPMENT

Buffers and Solutions

Name	Reagents	Volume per sample
1× SSC	150 mM sodium citrate, 15 mM NaCl, pH 7.0	100 μL
5× PBS	50 mM Na_2HPO_4, 9 mM KH_2PO_4, 685 mM NaCl, 13.5 mM KCl, pH 7.4	50 mL (M)
Bind & Wash buffer	10 mM Tris-HCl, 2 mM NaCl, pH 8.5	1 mL
ddH_2O	sterile, deionized water	2 mL
$EDTA_{0.1}$	0.1 mM, pH 8.0 (prepare from $EDTA_{500}$)	1 mL
$EDTA_{500}$	500 mM, pH 8.0	2 mL
Ethanol	70% (*v/v*)	40 mL (Y), 7 mL (M)
NaCl	5 M	150 μL
NaOH	20 mM	25 μL
Nuclear isolation buffer	10 mM Tris-HCl, 50 mM NaCl, 50 mM EDTA, 340 mM sucrose, 10% (*w/v*) glycerol, 0.1% (*w/v*) Triton X-100	3 mL (M)
PBS + 5 mM EDTA	Prepare from 5× PBS and $EDTA_{500}$	100 mL (M)
PBS + 25 mM EDTA	Prepare from 5× PBS and $EDTA_{500}$	25 μL (M)
PEG 8000	50% (*w/v*)	100 μL
Plug wash buffer	10 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 8.0	40 mL (M)
PMSF solution	200 mM in 100% (*v/v*) ethanol	200 μL (M)

(Continued on next page)

Continued

Name	Reagents	Volume per sample
Potassium acetate	5 M	2 mL (Y)
Proteinase K solution	1 mg mL^{-1} Proteinase K, 1% (*w/v*) sarkosyl, 125 mM EDTA, pH 9.0	16 mL (M)
SDS	1% (*w/v*)	10 μL (Y)
TE	10 mM Tris-HCl, 1 mM EDTA, pH 8.0	250 μL (Y)
TE$_{0.1}$	10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0	3 mL (M)
Tris-HCl pH 8.5	10 mM Tris-HCl, pH 8.5	26 mL (Y)
Y1 buffer	1 M sorbitol, 100 mM EDTA, pH 8.0, 14 mM 2- mercaptoethanol (added immediately before use where noted)	8 mL (Y)
Yeast lysis buffer	50 mM Tris-HCl, 50 mM EDTA, pH 8.0	5 mL (Y)
Zymolyase solution	10 mg mL^{-1} Zymolyase 20T	400 μL (Y)

⚠ CRITICAL: Observe caution and wear gloves and eye protection when handling ethanol, NaOH or PMSF. Ethanol is a flammable, toxic chemical that can cause irritation upon inhalation or skin contact. NaOH is corrosive and causes severe skin burns and eye damage. PMSF is toxic if swallowed and causes severe skin burns and eye damage. See safety data sheets for further information.

Note: The volumes per sample given in the table are approximate values and may vary depending on the starting material and the number of samples processed in parallel. Values specific for the yeast (Y) and mammalian (M) protocols are noted separately.

STEP-BY-STEP METHOD DETAILS

Note: Unless otherwise noted, all steps should be performed at a temperature between 20 and 25°C.

Extraction of Genomic DNA from Budding Yeast

⏱ Timing: 3 days

This section describes the extraction of genomic DNA from budding yeast cells using a gentle lysis method based on spheroplasting in order to avoid as much as possible introducing additional nicks into the DNA.

⚠ CRITICAL: In order to maintain spheroplasts intact, do not freeze the cells prior to lysis.

Alternatives: This protocol is recommended for preparing genomic DNA for detection of strand breaks caused inside cells, as it minimizes the introduction of additional nicks and breaks during the extraction procedure. If a low background of nicks is unproblematic (for example when detecting base modifications not involving strand breaks) or if starting from frozen cells, any other method for the isolation of genomic DNA can be used alternatively.

1. Harvest ca. 6×10^9 cells (300 OD$_{600}$):
   a. Centrifuge the yeast culture in a 50 mL tube at $2,000 \times g$ for 5 min and discard the supernatant.
   b. Resuspend the pellet in 25 mL of 10 mM Tris-HCl pH 8.5.
   c. Centrifuge again as before and discard the supernatant.
2. Prepare spheroplasts:
   a. Resuspend the cells in 4 mL of Y1 buffer.

    b. Add 400 μL of Zymolyase 20T solution and incubate at 30°C.

Note: 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.

3. Lyse the spheroplasts:
    a. Pellet the spheroplasts at 2,000×*g* for 5 min and discard the supernatant.
    b. Carefully resuspend the pellet in 4 mL of Y1 buffer without 2-mercaptoethanol, centrifuge as before and discard the supernatant.
    c. Resuspend the pellet in 4.75 mL of yeast lysis buffer.
    d. Initiate lysis by adding 250 μL of 20% SDS and incubate at 37°C until the solution turns transparent.

Note: This should normally take about 45 min.

4. Clear the solution of debris by centrifugation at 5,000×*g* for 10 min and transfer the supernatant (by pouring) to a fresh 50 mL tube.

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

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

Note: 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.

7. Precipitate the DNA by transferring the supernatant to a fresh 50 mL tube, adding 19.5 mL of 100% ethanol and swirling the tube.
8. Collect the precipitated DNA by centrifugation 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 into two 2 mL microcentrifuge tubes.

⚠ CRITICAL: Use a wide-bore pipette tip (prepare for example by cutting the tip off a 1 mL pipette tip), pipet slowly and gently and limit the pipetting as much as possible to avoid shearing the DNA while still ensuring complete resuspension.

Note: If the DNA pellet does not resuspend well, it can be incubated in the 50 mL tube for 12–16 h and split into two 2 mL microcentrifuge tubes the next day before proceeding to the next step. This will add one day to the timing of the experiment.

Note: 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.

11. Collect the precipitated DNA by centrifugation at 21,100×*g* for 10 min and discard the supernatant.
12. Air-dry the pellets and gently resuspend in 245 μL of TE each by incubation at 4°C for 12–16 h.
13. On the next day, 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. Precipitate the DNA again:
    a. Add 1.5 mL of 100% ethanol to each tube and invert to mix.
    b. 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 at 4°C by gentle rocking in 250 μL of 10 mM Tris- HCl pH 8.5 for 12–16 h.
17. On the next day, prepare AMPure beads for capturing of the DNA:
    a. Add 250 μL AMPure beads to a fresh 1.5 mL DNA LoBind microcentrifuge tube.
    b. 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 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 for 10 min and collecting the supernatant using a magnetic rack.

> *Note:* A minimum of 10 μg genomic DNA should be recovered at this stage. Typical yields are higher (ca. 25–50 μg).

> ⏸ **Pause Point:** At this point, the extracted genomic DNA can be stored at 4°C for up to 6 months.

> ⚠ CRITICAL: Storage at 4°C rather than -20°C is recommended because freezing can introduce strand breaks.

> *Note:* When mapping base lesions or modifications rather than 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.:

> - abasic (AP) sites: AP endonuclease (APE1) (Sriramachandran et al., 2020)
> - pyrimidine dimers: T4 Endonuclease V + APE1 (Sriramachandran et al., 2020)
> - methylated purines: N-methylpurine DNA glycosylase (AAG/MPG) + APE1 (Sriramachandran et al., 2020)
> - ribonucleotides: RNase H (Ding et al., 2015)

Denaturation of Genomic DNA and Ligation of 3′-OH Termini (Yeast)

⏱ Timing: 3.5–4 h

This section describes the heat denaturation of the purified genomic DNA and the subsequent capture of the exposed 3′-OH termini by ligation to a biotinylated oligonucleotide.

> *Note:* The protocol can be applied to DNA preparations from any source, provided that they are available as a solution. For treatment of DNA embedded in agarose plugs, please refer to the protocol for mammalian DNA below.

23. Denature the purified genomic DNA:
    a. Incubate 2.5 μg of genomic DNA at 95°C for 10 min.

b. Incubate the suspension on slushy ice for 5 min.

24. Perform the ligation reaction to capture free 3′-OH termini:

   a. Set up the ligation reaction in a PCR tube as follows:

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	3.0 μL

ddH$_2$O up to a total volume of 65 μL.

   b. Incubate the reaction in a thermocycler, using the following program:

Reaction Conditions			
Steps	Temperature	Time	Cycles
Annealing	25°C	60 min	1
Ligation	22°C	120 min	1
Hold	16°C	Forever	

Note: The amount of adaptor required for the ligation 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). Incomplete removal of unligated adaptor can cause various problems during subsequent steps of the protocol, especially when processing native, undigested DNA exhibiting very few breaks. These will become apparent in steps 37, 39 and/or 49 (see Troubleshooting). If the problems cannot be overcome by the measures described in the Troubleshooting section, the ligation reaction would need to be repeated with a reduced amount of adaptor. The appropriate amount will have to be determined empirically in this case.

25. Prepare fresh AMPure beads by adding 100 μL of AMPure bead suspension to a 1.5 mL DNA LoBind microcentrifuge tube and removing the storage buffer, using a magnetic rack.

26. Add 35 μL of 5 M NaCl and 35 μL of ddH$_2$O to the ligation mix and transfer the entire mix to the tube containing the AMPure beads.

27. Allow the DNA to bind to the beads by incubation 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 ddH$_2$O and incubation 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.

   Note: This protocol continues with step 29 below.

Extraction of Genomic DNA from Mammalian Cultured Cells

© Timing: 3 days

This section describes the extraction of genomic DNA from mammalian cultured cells embedded in agarose plugs in order to avoid as much as possible introducing additional nicks into the DNA.

Alternatives: This protocol is recommended for preparing genomic DNA for detection of strand breaks caused inside cells, as it minimizes the introduction of additional nicks and breaks during the extraction procedure. If a low background of nicks is unproblematic (for example when detecting base modifications not involving strand breaks) or if starting from frozen cells or tissues, any other method for the isolation of genomic DNA can be used alternatively.

1. Harvest up to $\sim 10^7$ cells by trypsinization.

   Note: Using fewer cells is unproblematic, and the protocol can be scaled down considerably when only preparing a single plug.

2. Wash the cells twice in 25–50 mL of ice-cold PBS + 5 mM EDTA.
3. Lyse the cells by resuspending them in 1 mL of nuclear isolation buffer and incubating them on ice for 10 min to dissolve the plasma membrane.
4. Harvest the nuclei:
   a. Centrifuge in a swing-out rotor at 1,200×*g* for 3 min at 4°C. Discard the supernatant.
   b. Wash the pellet once by resuspension in 1 mL of nuclear isolation buffer and centrifugation as before. Discard the supernatant.

   ⚠ CRITICAL: Loosen the pellet by inversion or tapping of the tube, rather than pipetting.

5. Mix the nuclei with nuclear isolation buffer supplemented with RNase A at a final concentration of 100 μg μL^{-1} 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 hemocytometer in the presence of Trypan blue at a final concentration of 0.2% (*w/v*).

   Note: We do not recommend the use of automated cell counters because they do not produce reliable results for this type of analysis.

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.

   Note: The number of nuclei used at this step is flexible and can be increased up to a packed volume of ~20 μL per plug.

8. Remove all but ~10 μL of supernatant, taking care not to disturb the pellet of nuclei.
9. Prepare prewarmed solutions for agarose embedding of nuclei:
   a. Prewarm PBS + 25 mM EDTA to 50°C.
   b. Melt an aliquot of 1.2% (*w/v*) low-melting point agarose in PBS + 25 mM EDTA at 95°C for 5 min and then let it equilibrate to 50°C for 15 min.
10. Embed the nuclei in agarose:
    a. Resuspend the nuclei (by tapping) at a density of about 700,000 nuclei per 22.5 μL of suspension in the prewarmed PBS + 25 mM EDTA solution.
    b. Mix the suspension 1:1 (v/v) with the molten agarose solution.

   c.   Immediately pipette the cell-agarose mixture into the wells of a 45 μL plug mold (Methods Video S1).

   ⚠ CRITICAL: Since the precision of 3D printing can be inconsistent, it is important to check before use whether the wells of the agarose casting mold properly fit the dedicated extrusion tool. Use only those wells that snugly fit the tool.

11. Allow the agarose to solidify for 1 h at 4°C.
12. Combine the plugs into 12 mL tubes filled with 7–8 mL of Proteinase K solution (Methods Video S2) and incubate for 24 h at 42°C.

   ⚠ CRITICAL: Only combine plugs corresponding to the same DNA sample, and do not combine more than 3 plugs per tube. Handle them gently as shown in Methods Video S2, as they are extremely fragile.

13. On the next day, exchange the Proteinase K solution for a fresh aliquot of 7–8 mL and perform a second incubation at 42°C for 12–16 h (Methods Video S3).

   ⚠ CRITICAL: In order to avoid losing the plugs during buffer exchange, let them settle to the bottom of the tube and quickly decant as much of the liquid as possible by tipping the tube while holding a wide spatula tightly to it as shown in Methods Video S3. Pouring out the buffer through a sieve, e.g a cell strainer, can also be useful.

14. On the next day, wash the plugs three times with 7–8 mL of plug wash buffer + 1 mM PMSF for 10 min each.
15. Wash the plugs once with 7–8 mL of plug wash buffer + 1 mM PMSF for 1 h.
16. Wash the plugs once with 7–8 mL of plug wash buffer for 1 h.
17. Transfer each plug to a separate 1.5 mL DNA LoBind tube.
18. Wash each plug twice with 1 mL of $TE_{0.1}$.
19. Wash each plug once with 1 mL of $TE_{0.1}$ for 30 min.
20. Wash each plug twice with 1 mL of $EDTA_{0.1}$ for 15 min each.
21. Briefly rinse the plugs with ddH_2O and remove as much residual liquid as possible (Methods Video S4).

   ⏸⏸**Pause Point**: At this point, the plugs can be stored at 4°C for up to 3 days.

   Note: When mapping base lesions or modifications rather than strand breaks, additional treatments are applied at this stage, i.e. while the DNA is embedded in the agarose plugs, to generate the corresponding 3′-OH termini, e.g.:

   - abasic (AP) sites: AP endonuclease (APE1)
   - pyrimidine dimers: T4 Endonuclease V + APE1
   - methylated purines: N-methylpurine DNA glycosylase (AAG/MPG) + APE1
   - ribonucleotides: RNase H

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

Denaturation of Genomic DNA and Ligation of 3′-OH Termini (Mammalian Cells)

   ⏱ Timing: 30 min (day 1) and 3.5–4.5 h (day 2)

This section describes the heat denaturation of the agarose-embedded DNA and the subsequent capture of the exposed 3′-OH termini by ligation to a biotinylated oligonucleotide.

22. Denature the DNA:
    a. Incubate the tubes containing the agarose plugs at 95°C for 4 min.
    b. Incubate the suspension on slushy ice for 5 min.
23. Melt the agarose again at 65°C for 5 min and equilibrate to 37°C for 15 min. At the end of this incubation, spin the tube briefly to collect any condensate that may have collected in the lid.
24. In the meantime, prepare a ligation mix:
    a. Combine the following reagents per sample:

10× T4 DNA ligase buffer	65 μL
Proximal adaptor	35 μL
ddH$_2$O	7.0 μL

    b. Pre-warm the ligation mix to 37°C.

   Note: Typically, three plugs are generated for each preparation of genomic DNA. A master mix for several plugs can be made at this step. The ligase is then added separately to each tube at the beginning of the next step to avoid leaving it at 37°C for too long.

25. Set up the ligation reaction:
    a. Add 3 μL of T4 DNA ligase to the prewarmed ligation mix.
    b. Thoroughly mix the agarose, by tapping, with 20 μL of the prewarmed ligation mix now containing the ligase.
    c. Incubate the reaction at 16°C for 12–16 h.
26. On the next day, incubate the mixture at 65°C for 10 min to inactivate the ligase and melt the agarose.
27. Allow the solution 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 ddH$_2$O and transfer to a 0.5 mL Bioruptor tube.

Fragmentation and Capture of Biotinylated Single-Stranded (ss)DNA

   ⏱ Timing: 2 h (excluding quality control)

This section describes the fragmentation of the purified genomic DNA and the isolation of the ligated fragments by capture on streptavidin beads.

   Note: From this point onwards, the protocol is identical for yeast and mammalian DNA, independent of whether the ligation was performed in solution or in agarose plugs.

29. Sonicate the eluate from step 28 (100 μL) to a fragment size of 200 nt, using a Bioruptor Pico.

   Note: Make sure the Bioruptor Pico is cooled to 4°C before sonication. Incubation of the purified samples on ice is not required. The number of cycles (30 s on / 30 s off) depends on the size of the genomic DNA. As a guide, use 4 cycles for yeast or 16 cycles for mammalian DNA.

   △ CRITICAL: Make sure fragmentation yields a fragment size range of 150–450 nt, with an average size of 300 nt. This is best verified by analyzing a 1 μL aliquot of the fragmented
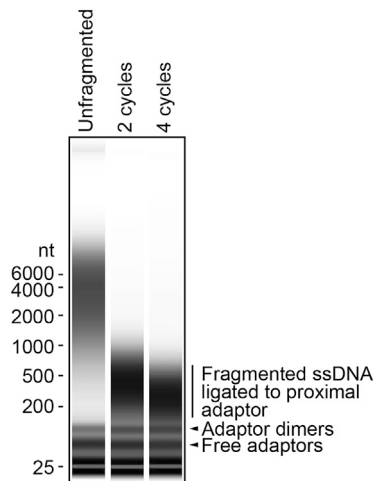
Figure 1. Quality Control for Efficient Fragmentation of Adaptor-Ligated Genomic DNA
Gel view of denatured, adaptor-ligated genomic yeast DNA before fragmentation and after 2 or 4 cycles of sonication, analyzed by Agilent RNA Screen Tape. The sample sonicated for 4 cycles shows an appropriate size distribution. The low molecular weight marker band is situated at 25 nt.

DNA either on an Agilent RNA ScreenTape (Agilent 2200 Tapestation) or an Agilent 2100 Bioanalyzer (using an RNA chip). A somewhat broader size distribution is not detrimental, but fragments should not be smaller than 150 bp as these will be depleted in the course of purification (Figure 1).

30. Clean up the DNA by two rounds of binding to an equivalent of 100 μL of AMPure beads and elution in 100 μL of ddH$_2$O:
    a. Prepare AMPure beads by adding 100 μL of AMPure bead suspension to a fresh 1.5 mL DNA LoBind microcentrifuge tube and removing the storage buffer, using a magnetic rack.
    b. Add NaCl and PEG 8000 to each sample of sonicated DNA at final concentrations of 0.9 M and 7.5% (*w/v*), respectively (i.e. add 27 μL of 5 M NaCl and 22.5 μL of PEG 8000 solution to a 100 μL sample).
    c. Allow the DNA to bind to the beads by incubation for 5 min.
    d. 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.
    e. Elute the DNA from the beads by adding 103 μL of ddH$_2$O and incubation for 5 min. Transfer 100 μL to a fresh tube, using a magnetic rack.

⚠ CRITICAL: Make sure you perform this purification step twice in order to remove excess unligated adaptor.

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 on a rotating wheel.
33. Pulse-spin the tube and collect the beads on a magnetic rack. Discard the supernatant.
34. Wash the beads twice with 50 μL of 1✕ SSC buffer each for 5 min on a rotating wheel.

Note: This step serves to remove unligated fragments of genomic DNA. When processing more than 5 samples in parallel, reduce the incubation time to 2.5 min to account for the increased processing time due to buffer addition and resuspension of the samples.

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

Note: This step serves to remove the splinter oligo by alkaline denaturation.

36. Wash the beads briefly with 100 μL ddH$_2$O.

37. Elute the DNA by adding 16 μL of ddH$_2$O, incubating at 95°C for 5 to 10 min and collecting the supernatant using the magnetic rack.

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

   ⚠ CRITICAL: Failure to remove unligated adaptors is one of the major sources of poor library quality. It is therefore recommended to verify efficient capture of biotinylated DNA and removal of unligated adaptor by analyzing a 1 μL aliquot of the eluted DNA either on an Agilent RNA ScreenTape (Agilent 2200 Tapestation) or an Agilent 2100 Bioanalyzer (using an RNA chip). The size range of the fragments should reflect the addition of the biotinylated adaptor (Figure 2).

Second Strand Synthesis, End Polishing, and Ligation of the Distal Adaptor

🕐 Timing: 3–4 h

This section describes the conversion of the captured DNA fragments to double-stranded (ds)DNA and the ligation of a second adaptor to their distal ends. This is required for the subsequent amplification in preparation for sequencing on the Illumina platform.

38. Perform the following reaction for second strand synthesis:
   a. Combine the following reagents in a PCR tube:

Eluate from step 37	14.85 μL
Phusion FLASH mix	15.0 μL
oHU3790 (100 μM)	0.15 μL

   b. Incubate the reaction mixture in a thermocycler with the following program:

Reaction Conditions			
Steps	Temperature	Time	Cycles
Denaturation	95°C	2 min	1
Annealing	60°C	30 s	1
Extension	72°C	2 min	1
Hold	4°C	Forever	

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

39. Purify the resulting dsDNA using AMPure beads:
   a. Pipette a 54 μL aliquot of AMPure bead suspension into a fresh 1.5 mL DNA LoBind microcentrifuge tube, keeping them in their storage buffer.

   Note: If a sample is expected to contain few nicks, the sample volume can be increased to 50 μL by addition of ddH$_2$O after completion of the reaction, and 90 μL of AMPure beads should then be used for purification. This minimizes loss of library.

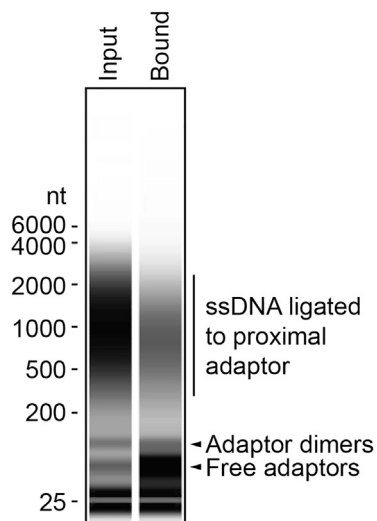   b. Allow the DNA to bind to the beads by incubation for 5 min.

Figure 2. Quality Control for Efficient Capture of Fragmented DNA
Gel view of genomic yeast DNA fragments before (Input) and after (Bound) capture on Streptavidin beads, analyzed by Agilent RNA Screen Tape. The low molecular weight marker band is situated at 25 nt.

c. 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.

d. Elute the DNA from the beads by adding 19 μL of ddH$_2$O and incubation for 5 min. Do not remove the eluate from the beads.

⚠ CRITICAL: The efficiency of second strand synthesis and ligation of the distal adaptor should be monitored after this step by analyzing a 2 μL aliquot of the eluate on an Agilent High Sensitivity D1000 ScreenTape (Agilent 2200 tape station). The DNA should be in a size range of 150–450 bp, with an average of 300 bp (Figure 3).

40. Perform an end polishing reaction:
    a. Set up the reaction by combining the bead suspension with the following reagents from the NEBNext Ultra II DNA Library Prep kit:
    Eluate (bead suspension) 17.0 μL End Prep Reaction Buffer2.3 μL End Prep Enzyme Mix 1.0 μL
    b. Incubate the reaction mixture in a thermocycler with the following program:

Reaction Conditions			
Steps	Temperature	Time	Cycles
End polishing	20°C	30 min	1
Enzyme inactivation	65°C	30 min	1
Hold	4°C	Forever	

41. Perform the ligation of the distal adaptor:
    a. Add the following reagents from the NEBNext Ultra II DNA Library Prep kit to the previous reaction to perform the distal adaptor ligation reaction:

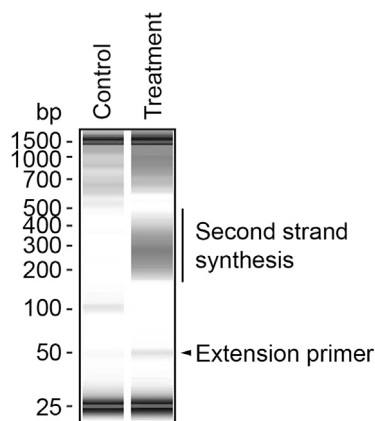|	
Ligation Master Mix	13.5 μL
Ligation Enhancer	0.45 μL
Distal adaptor	225 μL
ddH$_2$O	8.8 μL

Figure 3. Quality Control for Efficient Conversion of Captured ssDNA to dsDNA

Gel view of captured DNA fragments after second strand synthesis, analyzed on an Agilent High Sensitivity D1000 ScreenTape. The lane labeled "Control" shows the faint signal arising from a sample of undigested yeast genomic DNA and the lane labeled "Treatment" shows the abundant material derived from a sample of yeast genomic DNA digested with a restriction enzyme. The low and high molecular weight marker bands are situated at 25 and 1500 bp, respectively. Note that the original ssDNA present in the sample is not well detected on this type of dsDNA-selective ScreenTape.

   b. Incubate the reaction mixture in a thermocycler with the following program:

Reaction Conditions			
Steps	Temperature	Time	Cycles
Ligation	20°C	20 min	1
Hold	4°C	Forever	

42. Add the following reagents to the bead suspension in preparation for purification of the DNA:

5 M NaCl	13.5 μL
PEG 8000	11.25 μL
ddH$_2$O	5.25 μL

   Note: this step adjusts the conditions to re-bind the dsDNA fragments of the appropriate size range to the AMPure beads still present in the reaction from step 39.

43. Rinse the beads twice with 500 μL of 70% ethanol while keeping the tube on the rack.
44. Elute the purified DNA from the beads by incubation in 50 μL of ddH$_2$O for 5 min.
45. Perform a second round of purification on AMPure beads:
   a. Add the eluate (50 μL) to an 80 μL aliquot of AMPure bead suspension in the supplied buffer.
   b. Incubate the suspension for 5 min.
   c. Using a magnetic rack, remove the supernatant from the beads.
   d. Rinse the beads twice with 500 μL of 70% ethanol while keeping the tube on the rack.
46. Elute the DNA from the beads by adding 20 μL of ddH$_2$O and incubating for 5 min, then transfer the eluate to a fresh tube, using the magnetic rack.

  ⚠ CRITICAL: In order to ensure that the distal adaptor has been efficiently ligated (i.e. the size range of the library has shifted upwards by 32 bp) and the sample is free of unligated adaptors or adaptor concatemers, analyze a 1 μL aliquot of the purified product on a High Sensitivity D1000 Screen Tape (Agilent 2200 tape station) (Figure 4).

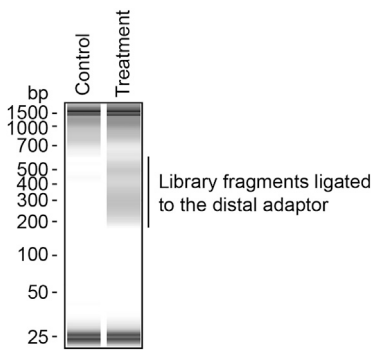 ⏸ Pause Point: At this point, the library can be stored at -80°C for months.

Figure 4. Quality Control for Efficient Ligation of the Distal Adaptor and Correct Size Range
Gel view of DNA libraries after ligation of the distal adaptor, analyzed on an Agilent High Sensitivity D1000 ScreenTape. As in Figure 3, the two lanes show material generated from a sample of undigested DNA (Control) and from a sample digested with a restriction enzyme (Treatment). The low and high molecular weight marker bands are situated at 25 and 1500 bp, respectively.

Library Amplification

⏱ Timing: 1.5 h

This section describes the amplification of the DNA library in preparation for sequencing. Primers P5 and P7 are compatible with the Illumina platform.

47. Amplify the library by the following PCR:
    a. Add the following reagents to a PCR tube, using the 2× Q5 Master Mix from the NEBNext Ultra II DNA Library Prep kit:

DNA (from step 46)	7.5 μL
2× Q5 Master Mix	12.5 μL
P5 (1 μM)	2.5 μL
P7 (1 μM)	2.5 μL

    b. Apply the following PCR program:

Reaction Conditions			
Steps	Temperature	Time	Cycles
Initial Denaturation	95°C	2 min	1
Denaturation	95°C	15 s	8
Annealing	60°C	30 s	
Extension	72°C	20 s	
Hold	4°C	Forever	

Note: Oligonucleotide P7 contains a barcode, $(X)_6$, for multiplexing, which should be applied according to the number of samples to be sequenced in parallel. A suitable set of indices for a specific number of samples and Illumina chemistry can be determined using a search tool at https://checkmyindex.pasteur.fr/ (Varet and Coppee, 2019).

48. Purify the amplified library using AMPure beads:
    a. Add the PCR reaction to a 25 μL aliquot of AMPure bead suspension in the supplied buffer.
    b. Incubate the suspension for 5 min.
    c. Using a magnetic rack, remove the supernatant from the beads.
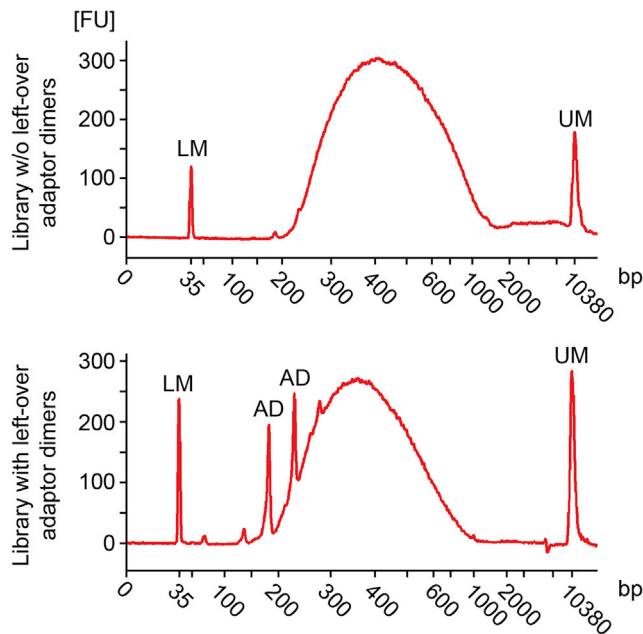   d. Rinse the beads twice with 500 μL of 70% ethanol while keeping the tube on the rack.

Figure 5. Quality Control for Efficient Amplification of the Library
The electropherogram shows GLOE-Seq libraries of high (top) and low (bottom) quality, analyzed by Agilent Bioanalyzer High Sensitivity DNA Kit. (AD: adaptor dimers; LM: lower marker; UM: upper marker). Note that the size range of the libraries shown here slightly exceeds the recommended value; however, a deviation of this magnitude is uncritical.

  e. Elute the DNA from the beads by adding 25 μL of ddH$_2$O and incubating for 5 min, then transfer the eluate to a fresh tube, using the magnetic rack.

49. Repeat the purification via AMPure beads, this time eluting in 20 μL of ddH$_2$O.

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

    △ CRITICAL: In order to ensure that the distal adaptor has been efficiently ligated (i.e. the DNA is in a size range of approximately 250–530 bp) and the sample is free of unligated adaptors or adaptor concatemers, analyze a 1 μL aliquot of the purified PCR product on an Agilent High Sensitivity DNA chip (Agilent 2100 Bionanalyzer) (Figure 5).

50. Determine the concentration of 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.

Sequencing and Demultiplexing

⏱ Timing: 12 h

This section describes the settings for sequencing GLOE-Seq libraries on the Illumina platform and the subsequent demultiplexing step that generates the input for the GLOE-Pipe routine.

51. Sequencing of GLOE-Seq libraries is carried out on an Illumina next-generation sequencer (e.g. NextSeq 500) with settings for single-ended 75 nt reads. For single indexed libraries use 7 nt indices; for dual indexed libraries use 8 + 8 nt indices.

Note: The necessary sequencing depth will depend on the genome size of the organism from which the samples are derived and on the purpose of the experiment. We aim to retrieve at least 3 million reads for yeast and at least 50 million reads for mammalian GLOE-seq libraries. We typically load a NextSeq 500 flow cell with 1.4 pM of a denatured library that includes 5% PhiX Control, as described in the NextSeq System Denature and Dilute Libraries Guide (https://www.well.ox.ac.uk/ogc/wp-content/uploads/2017/09/nextseq-denature-dilute-libraries-guide-15048776-02.pdf).

52. Upon completion of the run, perform demultiplexing with bcl2fastq and the relevant indices to generate FASTQ files for each library. These files will be used as input for subsequent data processing.

Data Processing Using GLOE-Pipe

⏱ **Timing**: variable (typically up to 1 day for human samples, depending on the specific workstation used for the analysis and the size of the libraries; ~3 min for the test dataset provided on GitHub)

This section describes a step-by-step walkthrough of GLOE-Pipe, an analysis pipeline that contains a set of modules for processing raw sequencing data and generating output files to detect, annotate, and visualize strand breaks or modified bases. In its default indirect mode, it identifies the position immediately upstream of a corresponding 3'-OH terminus. This mode should be used when analyzing strand breaks. Alternatively, in its direct mode it identifies the nucleotide immediately downstream of a 3'-OH terminus. This mode should be used when mapping the positions of modified bases by means of treatment with a nuclease that cleaves 5' to the affected nucleotide.

The pipeline is compatible with GLOE-Seq reads and can be accessed at https://github.com/helle-ulrich-lab/ngs-gloepipe. It is designed to run in a high-performance cluster environment with a Linux distribution, Bpipe (Sadedin et al., 2012) (version 0.9.9.3) as the domain specific language, and Lmod (https://lmod.readthedocs.io) (version 6.6) as the module system. In addition, GLOE-Pipe requires a set of additional software dependencies listed in the Key Resources Table that also need to be installed to run GLOE-Pipe successfully:

- The `gloeseq.pipeline.groovy` file contains the locations of the 12 modules included in GLOE-Pipe and how they are linked with each other in the direct and indirect modes.
- The `FastQC` and `Trimmomatic` modules are used by GLOE-Pipe to first report the quality of the raw data and then filter and trim the reads based on quality and adaptor inclusion, respectively. The `FastQC` module is used again to report the quality of the data after the filtering step.
- The `bowtie2` and `BAMindexer` modules are used to map the reads to the reference genome, generate BAM (Binary Alignment Map) files and create the index file for each one of them.
- The `bam2bedD` and `bam2bedI` modules convert the BAM files into BED (Browser Extensible Data) files for the direct and indirect modes, respectively. For the direct mode, each read represents one unit of signal that is positioned exactly at the 3'- terminal nucleotide of the original captured fragment. The indirect (default) mode positions the signal one nucleotide immediately upstream of the 5'-end of the original captured fragment, which corresponds to the nucleotide immediately 3' of the break.
- The `bedcoverage`, `bed2bw` and `rfd` modules convert BED files into normalized, non- normalized and replication fork directionality BIGWIG (Big Wiggle) files, respectively. The replication fork directionality (RFD) index reflects the relative strand bias of the break signals and is calculated as follows: RFD = (REV − FWD)/(REV + FWD).
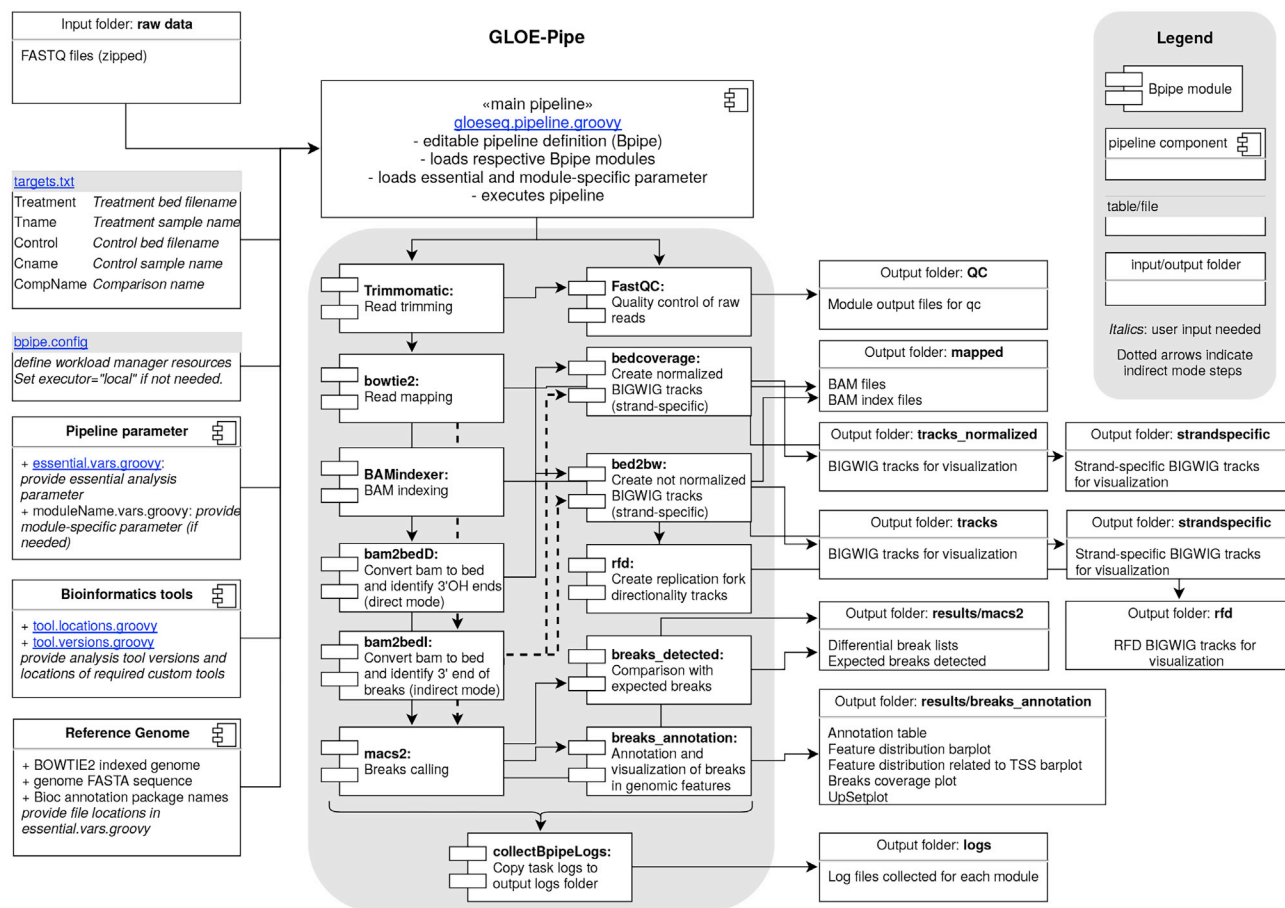
Figure 6. Flowchart of GLOE-Pipe

- The `macs2` module performs the breaks calling for the treatment sample by comparing the BED file for each strand to that of a control sample (for example, a digested sample to an undigested control).
- The `breaks_annotation` and `breaks_detected` modules perform the annotation of detected breaks and check the overlap between the detected and expected breaks.
- Finally, the `collectBpipeLogs` module creates a copy of all log files generated by the previous modules into the logs directory (Figure 6).

> *Note:* For demonstration of the pipeline and testing purposes, a small test dataset can be found in the `test_data` directory. This directory contains seven FASTQ files and the corresponding configuration file (`targets.txt`) that is needed to collect the file names and define the sample comparisons used to detect the breaks.

53. Clone or download the pipeline in your project directory using the link reported on GitHub:
```
$ git clone <git repository url> <project_dir>/GLOEPipe
```
54. Move the test dataset to the project directory and create symbolic links for the files that need to be edited:
```
$ cd <project_dir>
$ mv GLOEPipe/test_data .
$ ln -s test_data/targets.txt .
```

```
$ ln -s GLOEPipe/modules/GLOEseq/tool.locations.groovy .
$ ln -s GLOEPipe/modules/GLOEseq/tool.versions.groovy .
$ ln -s GLOEPipe/modules/GLOEseq/essential.vars.groovy .
$ ln -s GLOEPipe/pipelines/GLOEseq/gloeseq.pipeline.groovy .
$ ln -s GLOEPipe/pipelines/GLOEseq/bpipe.config .
```

Alternatives: Instead of using symbolic links, you can copy the `targets.txt` file into the project folder and specify the full path for the `gloeseq.pipeline.groovy` file to run the analysis:

```
$ cp test_data/targets.txt .
$ bpipe run GLOEPipe/pipelines/GLOEseq/gloeseq.pipeline.groovy
test_data/*.fastq.gz
```

In this case, the other 5 files (steps 55–59) need to be edited directly in the pipelines and modules directories.

55. Set the location and the version of the necessary tools in the `tool.locations.groovy` and `tool.versions.groovy` files, respectively. For example:
    `tool.locations.groovy`

    `TOOL_DEPENDENCIES="/fsimb/common/tools"` // Add the path to the folder which contains all the tools used by GLOE-Pipe

    `TOOL_TRIMMOMATIC=TOOL_DEPENDENCIES + "/trimmomatic/0.36/"` // Add the subfolder for Trimmomatic

    `tool.versions.groovy`

    `TRIMMOMATIC_VERSION="0.36"` // Add the version of Trimmomatic that should be loaded (e.g. `module load trimmomatic/0.36`)

56. Edit the `essential.vars.groovy` file to adapt it to your specific needs:
    a. Edit the file to include the full paths to:
       - the project directory (`ESSENTIAL_PROJECT`)
       - the adaptor FASTA file used by the trimmer (`ESSENTIAL_TRIMMOMATIC`)
       - the reference genome index files for Bowtie 2 (`ESSENTIAL_BOWTIE_REF`)
       - the reference genome chromosome sizes file (`ESSENTIAL_BOWTIE_GENOME_INDEX`). For example:

    `ESSENTIAL_TRIMMOMATIC=".../TruSeq3-SE.fa"` // Add the path to the adaptor FASTA file used by Trimmomatic
    b. If running GLOE-Pipe on a dataset not derived from budding yeast, change the information regarding the reference genome and corresponding annotations in the following statements:
       // Annotation packages used by the R scripts
       `ESSENTIAL_TXDB="TxDb.Scerevisiae.UCSC.sacCer3.sgdGene"` //
       http://bioconductor.org/packages/release/data/annotation/ - Annotation package for TxDb object(s)
       `ESSENTIAL_ANNODB="org.Sc.sgd.db"` //
       http://bioconductor.org/packages/release/data/annotation/ - Genome wide annotation
       // Mappable genome size for MACS2
       `ESSENTIAL_MACS2_GSIZE="1.20E+07"` //
       https://deeptools.readthedocs.io/en/develop/content/feature/effectiveGenomeSize.html
       // TSS region parameter from –200 bp to +200 bp
       `ESSENTIAL_TSS=200`

Note: The annotation packages contain genomic locations of the transcript-related features (5′ and 3′ untranslated regions (UTRs), protein coding sequences (CDSs) and exons) and genome-wide annotation databases (SYMBOL, GENENAME and ENSEMBL/ENTREZID) for the budding yeast genome (sacCer3).

Optional: The user can include up to 3 BED files containing coordinates for the expected breaks. As a default and for the purpose of the test dataset, GLOE-Pipe will use the expected breaks for *Bsr*DI, Nb.*Bsr*DI and *Not*I:

`ESSENTIAL_EB1=ESSENTIAL_PROJECT + "/GLOEPipe/tools/REs/BsrDI.bed" //` Expected BsrDI breaks in the yeast genome

`ESSENTIAL_EB2=ESSENTIAL_PROJECT + "/GLOEPipe/tools/REs/Nb_BsrDI.bed" //` Expected Nb.BsrDI breaks in the yeast genome.

`ESSENTIAL_EB3=ESSENTIAL_PROJECT + "/GLOEPipe/tools/REs/NotI.bed" //` Expected NotI breaks in the yeast genome

57. To perform pairwise comparisons between treatment and control samples on datasets other than the test dataset, edit the targets.txt file to match the relevant samples.

    Note: The `targets.txt` file contains the file names and the sample comparisons that are used by the MACS2 peak caller to detect significant breaks. For example:

File name (treatment sample)	Sample name	File name (control sample)	Sample name	Comparison name
BsrDI_rep1_trimmed.bed	BsrDI_rep1	Undigested_trimmed.bed	Undigested	BsrDI_rep1.vs.Undigested

The file is correctly set up for the test dataset. To run GLOE-Pipe on another dataset, the file names, sample names, and comparison name for the treatment and control samples (in the test dataset: *Bsr*DI-digested versus undigested) need to be included in the `targets.txt` file.

58. Edit the `gloeseq.pipeline.groovy` file to set GLOE-Pipe from the default indirect to the optional direct mode and adjust the set of individual modules as required. This is best done by commenting and uncommenting the relevant lines.
    a. If needed, set the pipeline to direct mode:

// MAIN PIPELINE TASK (DIRECT mode - optional)

```
run {
    "%.fastq.gz" * [ FastQC, Trimmomatic +
  [ FastQC, bowtie2 + BAMindexer + bam2bedD +
  [ bedcoverage, bed2bw + rfd ] +  macs2 ] ] +
  [ breaks_annotation, breaks_detected ] + collectBpipeLogs
}
```
    b. Remove unnecessary modules, for example:

// Portion of GLOE-Pipe with rfd module
```
… + [ bedcoverage, bed2bw + rfd ] +…
```

// Portion of GLOE-Pipe without rfd module
```
... + [ bedcoverage, bed2bw ] +...
```

Optional: new modules can be added to GLOE-Pipe by following the instructions reported in the documentation for Bpipe (http://docs.bpipe.org/).

59. Edit the `bpipe.config` file to adjust the resources (time limit, number of processors and amount of memory to reserve) required by each module included in the pipeline. For example:
```
Trimmomatic {
    walltime="02:00:00" // Time limit (in h)
    procs="1" // Number of processors
    memory="2" // Amount of memory (in GB)
}
```

Note: this step is optional when processing yeast samples.

60. To submit GLOE-Pipe jobs to the cluster, we recommend using Linux Screen (`screen`) to avoid network disruptions during data processing. After starting a new screen session, load the Bpipe module customized for the Slurm job manager:
```
$ screen
$ module load bpipe/0.9.9.3.slurm
```
61. To start running GLOE-Pipe on the test dataset, execute the following command:
```
$ bpipe run gloeseq.pipeline.groovy test_data/*.fastq.gz
```

Note: The pipeline should produce an output such as the following:

```
========================================================================
|                    Starting Pipeline at 2020-04-16 16:18
|
========================================================================
======================= Stage FastQC (BsrDI_rep1) ====================
======================= Stage FastQC (NbBsrDI_rep2) ==================
======================= Stage FastQC (Undigested) ====================
======================= Stage FastQC (NbBsrDI_rep1) ==================
================== Stage FastQC (NbBsrDI_NotI_1_10_rep1) =============
================== Stage FastQC (NbBsrDI_NotI_1_10_rep2) =============
======================= Stage FastQC (BsrDI_rep2) ====================
=========== Stage Trimmomatic (NbBsrDI_NotI_1_10_rep1) ===============
=========== Stage Trimmomatic (NbBsrDI_NotI_1_10_rep2) ===============
================== Stage Trimmomatic (BsrDI_rep2) ====================
================== Stage Trimmomatic (NbBsrDI_rep1) ==================
================== Stage Trimmomatic (NbBsrDI_rep2) ==================
================== Stage Trimmomatic (BsrDI_rep1) ====================
================== Stage Trimmomatic (Undigested) ====================
```

⚠ CRITICAL: If the pipeline finishes running without errors, it should output the following:

```
========================= Pipeline Succeeded =========================
16:21:03 MSG: Finished at Thu Apr 16 16:21:03 CEST 2020

16:21:03 MSG: Outputs are:

/home/project/mapped/NbBsrDI_NotI_1_10_rep2_trimmed.bam.bai
```

```
/home/project/tracks/strandspecific/rfd/NbBsrDI_NotI_1_10_rep2_trimmed.
1kb.rfd.bw

/home/project/tracks_normalized/NbBsrDI_NotI_1_10_rep2_trimmed.
normalized.bw

/home/project/results/macs2/NbBsrDI_NotI_1_10_rep2_trimmed_macs2.done

/home/project/results/macs2/Breaks_detected_re1.RData

... 32 more ...
```

EXPECTED OUTCOMES

When GLOE-Pipe finishes running without errors, the project directory should contain the following files:

```
├── rawdata_trimmed
├── qc
│   └── fastqc
├── mapped
├── bed
├── tracks
│   └── strandspecific
│       └── rfd
├── tracks_normalized
│   └── strandspecific
├── results
│   └── macs2
│   └── breaks_annotation
├── logs
└── commandlog.txt
```

- `rawdata_trimmed`: this directory contains the FASTQ files produced by the trimming step that filters and trims the reads based on quality and adaptor inclusion. These files are used as inputs by the mapping step.
- `qc/fastqc`: this directory contains the FastQC reports for each raw (`SAMPLE_NAME_-fastqc.html`) and trimmed FASTQ file (`SAMPLE_NAME_trimmed_fastqc.html`).
- `mapped`: this directory contains the BAM file (`SAMPLE_NAME_trimmed.bam`) and the BAM index file (`SAMPLE_NAME_trimmed.bam.bai`) for each sample. These BAM files contain the mapped reads on the appropriate genome, filtered by map quality $\geq$ 30.
- `bed`: this directory contains a BED file for each sample (`SAMPLE_NAME.bed`). In the indirect mode, these files list the coordinates of the nucleotide upstream and on the opposite strand of the 5′-end of each mapped read. In the direct mode, they list the first nucleotide on the strand opposite the 5′-end of each mapped read. The information in these BED files is further filtered to produce files that contain counts for breaks that map exclusively to the plus (`SAMPLE_NAME.fwd.bed`) or the minus (`SAMPLE_NAME.rev.bed`) strand.
- `tracks` and `tracks_normalized`: these directories contain non-normalized and BPM- (Bins Per Million mapped reads) -normalized BIGWIG files, respectively, for each sample (`SAMPLE_NA-ME.bw` and `SAMPLE_NAME.normalized.bw`). These files can be visualized in any commonly available genome browser, such as Integrative Genomics Viewer (IGV, http://software.broadinstitute.org/software/igv/).

- `strandspecific`: this subdirectory contains two BIGWIG files for each sample with strand- specific information (plus strand, `SAMPLE_NAME.fwd.bw`; minus strand, `SAMPLE_NAME.rev.bw`).
- `tracks/strandspecific/rfd`: this subdirectory contains a BIGWIG file for each sample with replication fork directionality (RFD) values. RFD is calculated as follows: RFD = (REV − FWD)/ (REV + FWD).
- `results/macs2`: this directory contains the following files:
  - Standard output files generated by MACS2 (see https://github.com/taoliu/MACS)
    - `TREATMENT.vs.CONTROL_macs2_peaks_narrowPeak`
    - `TREATMENT.vs.CONTROL_macs2_peaks.xls`
    - `TREATMENT.vs.CONTROL_macs2_summits.bed`
  - `TREATMENT.vs.CONTROL_EB.xls`: the list of detected breaks that match the position of the expected breaks (defined in the optional part of step 56) produced by each analyzed treatment (see `essential.vars.groovy`).
  - `EB_breaks_detected_table.csv`: summary table for each BED file (optional part of step 56), reporting the number and the percentage of expected and detected breaks for each comparison included in `targets.txt`.
- `results/breaks_annotation`: this directory contains the following files based on the annotation packages used in step 56:
  - `GLOEseq_Feature_Distribution_Barplot.png`: The barplot shows the percentage of detected breaks that overlap with genomic features, such as promoter, 5′ UTR, 3′ UTR, 1st Exon, Other Exon, 1st Intron, Other Intron, Downstream and Distal Intergenic.
  - `GLOEseq_Feature_Distribution_Related_to_TSS_Barplot.png`: The barplot shows the percentage of breaks that are detected upstream and downstream of the transcription start site (TSS) of the nearest genes.
  - `TREATMENT.vs.CONTROL_GLOEseq_UpSetplot.png`: The UpSet plot illustrates potential interactions between various genomic features and the detected breaks by showing how many breaks overlap with more than one genomic feature.
  - `TREATMENT.vs.CONTROL_GLOEseq_Breaks_Coverageplot.png`: The coverage plot shows the locations of all breaks over the entire genome.
  - `Breaks_Annotation.xls`: The annotation of detected breaks for each sample with their distance from the TSS of the nearest genes.
- `logs`: this directory contains all the log outputs generated by the modules included in GLOE-Pipe.
- `commandlog.txt`: this file includes all the commands submitted by the pipeline to the cluster.

Note: After successfully running GLOE-Pipe on the test dataset, we recommend to test the pipeline on a full dataset, which can be downloaded from the Gene Expression Omnibus repository, https://www.ncbi.nlm.nih.gov/geo, e.g. GSE134225, (digestion experiment: GSE134219). This allows the user to compare their results directly with those generated in our laboratory.

QUANTIFICATION AND STATISTICAL ANALYSIS

Each read produced by GLOE-Seq maps to the strand opposite that on which the break was originally situated, with its first (5′) nucleotide at a position one nucleotide upstream of the respective event. Therefore, each GLOE-Seq read represents one 3′-OH terminus per break.

Raw, non-normalized break counts are normalized for sequencing depth using the "Bins Per Million mapped reads" method in deepTools (−normalizeUsing BPM). This is equivalent to the TPM (Transcripts Per Million) method that has been developed for RNA-Seq data (Wagner et al., 2012).

If control and treatment DNA samples are available, GLOE-Pipe compares them to determine the statistical properties, p- and q- values, of the breaks in the treatment, using the callpeak function, without model building, in MACS2 (Zhang et al., 2008).

LIMITATIONS

GLOE-Seq is a powerful tool to map DNA strand breaks as well as modified or damaged bases with nucleotide resolution. Yet, the method has inherent limitations that need to be considered when designing experiments. The availability of free, ligatable 3′-OH ends is an obvious necessity for their capture. Thus, strand breaks with different termini, e.g. bearing a phosphate group or a protein adduct emerging from topoisomerase action, cannot be mapped without an appropriate pre-treatment to liberate the ends. Moreover, base lesions near the termini would interfere with second strand synthesis and thus prevent detection of the corresponding breaks. The need for size selection of captured fragments during removal of unligated adaptors imparts a limitation on the detection of multiple events in close proximity. Therefore, according to our analysis, breaks that are situated less than 100 nt downstream of another break are not efficiently mapped (Sriramachandran et al., 2020). Finally, we observed a measurable pattern of single-stranded breaks in untreated samples of genomic yeast and human DNA that we attribute to the ongoing removal of incorporated ribonucleotides after DNA replication, based on its bias towards leading strand sequences (Sriramachandran et al., 2020). This "background" has hampered the mapping of rare events with frequencies below 0.1%, and we have not examined whether increasing the number of reads would improve detection.

Beyond these inherent limitations, technical problems can interfere with key steps of the procedure. The quality of the proximal adaptor is critical, as insufficient biotinylation will prevent its capture.

Efficient sonication is important to generate fragments suitable for library amplification and sequencing. Conditions yielding an appropriate size range need to be established empirically, as they will depend on the origin of the DNA and the specifications of the instrumentation. Removal of free adaptors is crucial at multiple steps: during second strand synthesis, excess proximal adaptor can interfere with efficient extension, and after ligation of the distal adaptor, inefficient removal of unligated oligonucleotides will result in undesired PCR products during amplification of the library that may reduce the number of mappable reads during sequencing.

Although we have performed extensive proof-of-principle experiments to ensure the robustness of GLOE-Seq, many variations and potential adaptations remain untested. For example, we have not explored alternative sequencing chemistries or platforms or determined the minimal amount of input DNA for library preparation. At present, our procedure does not involve the use of Unique Molecular Identifiers (UMIs) to account for PCR duplicates during library amplification. Although absolute quantification of breaks should be possible by pre-digestion of the genomic DNA with a rare-cutting restriction enzyme (Zhu et al., 2019), we have not rigorously optimized this feature yet. Additional information about the properties of mapped breaks could further be derived from the capture of 5′-termini from the same sample. All these options would constitute valuable extensions of the method and will be developed in the course of future applications of GLOE-Seq.

TROUBLESHOOTING
Problem
Improper size range of DNA after fragmentation (visible upon quality control, step 29, Figure 1).

Potential Solution
Adjust and standardize the fragmentation conditions for the type of sample and the instrument used. If the user is unfamiliar with the specific properties of the samples or the instrument used for

sonication, the degree of fragmentation should best be determined empirically with a test sample before performing this step on valuable experimental samples.

Problem

No DNA in the sample of eluted material after capture (visible upon quality control, step 37, Figure 2).

Potential Solution

If the sample is expected to contain very few nicks, the concentration of captured DNA might be too low to detect at this step. In this case, no action is required. Alternatively, the problem may be due to inefficient capture of the biotinylated DNA. In this case, verify the quality of the Streptavidin beads using unligated proximal (biotinylated) adaptor as a control. Capture efficiency may also depend on the type of biotin label used. We use "3'-Biotin-dT" (from IDT) as a label, as the "Standard Biotin" label has caused problems in our hands.

Problem

Presence of large amounts of unligated adaptor in the eluted material after capture (visible upon quality control, step 37, Figure 2).

Potential Solution

The presence of unligated adaptors or other oligonucleotides is one of the most common problems arising in the course of this protocol. Unligated adaptor will be present at this step if it has not been efficiently removed during the clean-up steps before capture on Streptavidin beads. In this case, perform an additional clean-up of the captured fragments after step 37, using 1.6 volumes of AM-Pure beads and eluting in 16 μL of water.

Problem

Presence of unligated adaptor, but no DNA of larger size in the eluted material after capture (visible upon quality control, step 37, Figure 2).

Potential Solution

Excess unligated adaptor in the absence of ligation products can result from inefficient ligation of 3'- OH ends to the biotinylated adaptor. In that case, ligation efficiency should be verified by separating a small sample of the DNA on a TBE-urea gel, blotting onto a Nitrocellulose membrane and probing with an anti-Streptavidin antibody. If needed, ligation conditions can then be adjusted.

Problem

No product visible after second strand synthesis (visible upon quality control, step 39, Figure 3).

Potential Solution

Excess unligated biotinylated oligonucleotide may inhibit second strand synthesis because it might titrate away the oligonucleotide used for extension. Therefore, if large amounts of free adaptor are detected in the analysis, perform additional purification of the samples using AMPure beads to remove the excess adaptor and repeat the second strand synthesis reaction. If the reaction fails again or if there is no excess of unligated adaptor detected in the analysis, it is possible that the removal of the non-extendable oligonucleotide from the ligation products during the alkaline wash (step 35) was inefficient, which would impede efficient extension. In that case, repeat the streptavidin capture by re-starting the protocol from step 31.

Problem

Presence of adaptor concatemers (visible upon quality control, step 49, Figure 4).

Potential Solution

Adaptor concatemers can result from the presence of excess unligated proximal and distal adaptors. Extreme care must be taken to remove these low-molecular weight products as much as possible, as they will interfere with the efficiency of sequencing and greatly reduce the number of mappable reads. Additional purification using AMPure beads at a beads:sample ratio of 1:1 (v/v) will remove adaptor concatemers.

Problem

When configuring GLOE-Pipe, paths or the tools cannot be found (step 55).

Potential Solution

If full paths and the tools used by GLOE-Pipe have not been specified, add the required paths in `essential.vars.groovy`. Modify the versions and the locations for the required tools in `tool.versions.groovy` and `tool.locations.groovy`.

Problem

Low percentage of mapped reads during read mapping (step 61)

Potential Solution

Poor mappability of reads could result from the use of an incorrect reference genome. In this case, check that the organism specified is correct. Alternatively, poor sequencing quality could be the cause of poor mapping. In order to assess this, inspect the quality control (QC) reports for your data before and after the trimming step in the `qc/fastqc` directory.

Problem

Individual module(s) crash during GLOE-Pipe (step 61).

Potential Solution

Crashing of individual modules within GLOE-Pipe could result from insufficient time and/or memory allocation for the affected module. In this case, increase the resources required for running a module in `bpipe.config`.

Problem

R function cannot be found (step 61).

Potential Solution

Check the list of required R packages included in the Key Resources Table or at beginning of the R scripts (`GLOE-Pipe/tools/breaks_annotation`) and make sure they are installed.

RESOURCE AVAILABILITY

Lead Contact

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

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

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.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.xpro.2020.100076.

AUTHOR CONTRIBUTIONS

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

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Andrews, S. (2019). FastQC: A quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

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.

Geimer, M., Hoste, K., and McLay, R. (2014). Modern scientific software management using easybuild and Lmod. In Proceedings of the First International Workshop on HPC User Support Tools, pp. 41–51, https://doi.org/10.1109/HUST.2014.8.

Gel, B., Diez-Villanueva, A., Serra, E., Buschbeck, M., Peinado, M.A., and Malinverni, R. (2016). regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. Bioinformatics 32, 289–291.

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

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

Lawrence, M., Gentleman, R., and Carey, V. (2009). rtracklayer: an R package for interfacing with genome browsers. Bioinformatics 25, 1841–1842.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.

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

R Core Team (2019). R: A Language and Environment for Statistical Computing (Austria: Vienna), Available at: https://www.R-project.org/.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 42, W187–W191.

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

Sriramachandran, A.M., Petrosino, G., Mendez-Lago, M., Schafer, A.J., Batista-Nascimento, L.S., Zilio, N., and Ulrich, H.D. (2020). Genome-wide nucleotide-resolution mapping of DNA replication patterns, single-strand breaks, and lesions by GLOE-seq. Mol. Cell 78, 975–985.

Varet, H., and Coppee, J.Y. (2019). checkMyIndex: a web-based R/Shiny interface for choosing compatible sequencing indexes. Bioinformatics 35, 901–902.

Wagner, G.P., Kin, K., and Lynch, V.J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. 131, 281–285.

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., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137.

Zhu, Y., Biernacka, A., Pardo, B., Dojer, N., Forey, R., Skrzypczak, M., Fongang, B., Nde, J., Yousefi, R., Pasero, P., et al. (2019). qDSB-Seq is a general method for genome-wide quantification of DNA double-strand breaks using sequencing. Nat. Commun. 10, 2313.