Establishing MapR: A Genome-wide R-loop Mapping Strategy

A Thesis

SUBMITTED TO THE FACULTY OF THE

UNIVERSITY OF MINNESOTA

BY

Zijun Zhang

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

Adviser: Dr. Hai Dang Nguyen

Co-adviser: Dr. Anja K. Bielinsky, Dr. Justin Drake

June 2021

Dedications

To the almighty god and my savior Jesus Christ. May all glory be to you. Thank you for always guiding me in both my schoolwork and my life.

To my dear parents Minshun Zhang and Hongtao Zhang and my dear grandparents Shaochun Zhang and Shufang Wang. Thank you for fully supporting what I would like to do all the time and always being confident in me.

To my dear cousin Zihan Zhang, Xueli Zhang and my grandmother Huixian Tian. Thank you for being part of my life.

To all my dearest friends who always support and understand me.

Acknowledgements

I would like to thank my mentor Dr. Hai Dang Nguyen for all the help provided for in-lab research and thesis writing, thank you for training me in these 2 years and teaching me how to be a scientist.

In addition, I would like to appreciate all members in HDN lab: Maxwell Banister, Victor Corral, Cassandra Leibson, Zhiyan (Silvia) Liu, Youngjay Park and Dawei Zong (ranked in alphabetical order according to surnames) for all the help provided in lab and all the suggestions provided on this thesis; Dr. Anja K. Belinsky and Dr. Justin Drake, for all the work done on reviewing this thesis; Dean of University of Minnesota Pharmacology Department Colin Campbell, for giving me the chance to join the MS Pharmacology program and get educated at U of M.

Abstract

R-loops are transcription intermediates containing an RNA:DNA hybrid and a displaced single-strand DNA. R-loops are important regulators of various cellular processes such as transcription and DNA repair. However, when R-loop levels and/or distributions in the genome are dysregulated, they act as endogenous sources of genomic instability, a hallmark of cancers. Mutations in genes encoding for R-loop regulators are observed in various cancers and diseases, linking R-loop dysregulation to disease pathogenesis. As the number of known R-loop regulators increases, this raises an intriguing question on how different factors contribute to R-loop homeostasis. Furthermore, whether various R-loop regulators function at the same R-loop regions in different diseases is largely unknown. In order to answer these questions, a proper mapping method to analyze R-loop landscapes in different diseases will be necessary. My thesis project focuses on establishing an R-loop mapping strategy called MapR. MapR was reported to produce a high signal-to-noise ratio in genome-wide next-generation sequencing (NGS). However, it lacks a quality control step prior to NGS. In this study, I successfully established and optimized the MapR method to quantify R-loop levels by quantitative PCR (MapR-qPCR) as the quality control step. A significantly higher level of R-loop enrichment was observed at the R-loop positive RPL13A locus compared to that at the R-loop negative SNRPN locus by the MapR-qPCR analysis. Furthermore, treating MapR samples with RNaseH, an enzyme that hydrolyzes the RNA moiety within RNA:DNA hybrids, partially suppressed R-loop enrichment at RPL13A locus, suggests that MapR enriches for R-loops. My project will allow us to integrate the MapR-qPCR analysis as quality control checkpoints prior to genome-wide sequencing.

Table of Contents

List of Tables

List of Figures

Introduction

R-loop is a transcription intermediate containing an RNA:DNA hybrid and a displaced single-strand DNA[1]. R-loops participate in several cellular events including transcription regulation, transcription termination, class-switch recombination, DNA repairing, and mitochondrial replication[2–6]. However, dysregulated R-loops were shown to promote transcription-replication collisions (TRCs) and transcription stalling, which induce DNA double-strand breaks, a source of genomic instability[7,8]. In bacteria, this was caused by co-directional collisions and RNA polymerase backtracking, and it was suspected that eukaryotes have a similar mechanism [9]. Mutations in genes encoding R-loop factors were observed in several diseases and cancers such as myelodysplastic syndromes (MDS), Burkitt's lymphoma, and Fanconi anemia, indicating a potential high association between R-loop dysregulation and disease pathogenesis[10–12].

Mechanisms of R-loop regulation

R-loops are regulated by 5 different processes (Figure 1). First, the topoisomerase and RNA binding proteins (RBPs) can prevent R-loop formation during transcription. Supercoiling of unwound DNA promotes R-loop formation during transcription, which is resolved by Topoisomerase I (TopI)[13]. At the same time, RBPs (i.e., RNA splicing factor SRSF1, RNA exporting factor THO complex) physically bind to the pre-mRNA synthesized by RNA polymerase II, thereby preventing RNA:DNA hybrid formation[14,15]. Second, if R-loops are formed, the RNA:DNA hybrids can be unwound by RNA:DNA hybrid helicases such as SETX, DHX9, and several members of the DDX family including DDX5 and DDX21[16–19]. The unwound RNA is then digested by the RNA exonuclease XRN2[17]. Third, stable RNA:DNA hybrids can also be resolved by the ribonucleases H1 and H2 (RNaseH1/2). These enzymes hydrolyze the RNA components within the RNA:DNA hybrids of R-loops[20,21]. Finally, accumulating evidence also revealed several DNA repair factors such as BRCA1/2, Fanconi anemia proteins FANCM and FANCD2, and ATR checkpoint kinase protect stalled DNA replication forks from collapse when colliding with unscheduled R-loops during S-phase to prevent genomic instability[12,22–24]. Finally, epigenetic regulators such as BAF and BRD4 have been reported to prevent unscheduled transcription during DNA replication to prevent TRCs. Overall, the increasing number of R-loop regulators that were revealed in the past several years brought new scientific challenges: It is still not clear whether all the R-loops in the genome are the same or

different and whether the factors involved in the R-loop resolution have any preferences or bias at specific genomic loci or regions.

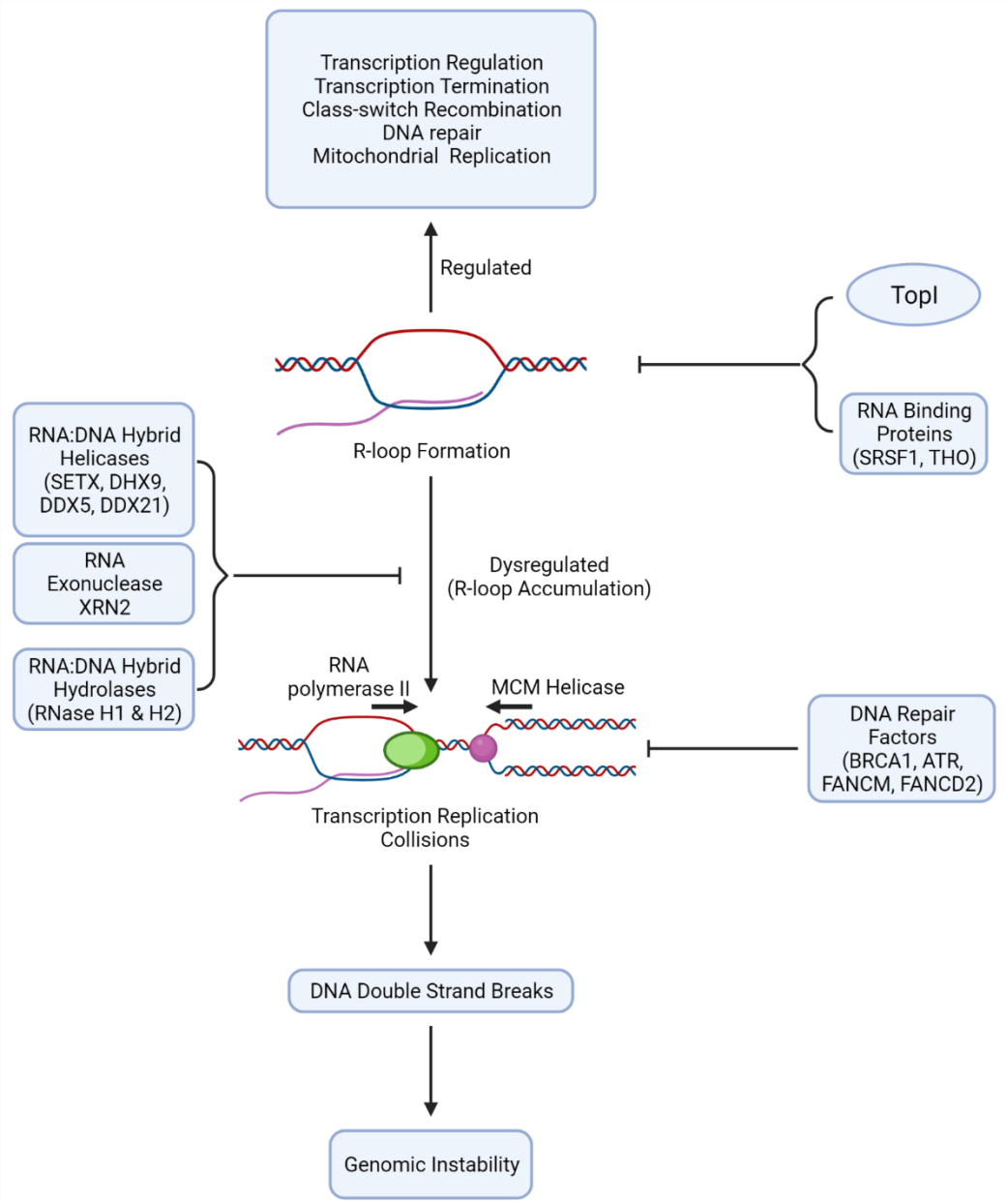

Figure 1: **Mechanisms of R-loop resolution and functions.** The R-loop contains an RNA:DNA hybrid and a displaced single-strand DNA. R-loops are normally regulated by topoisomerase I, RBPs, helicases, RNA:DNA hybrid hydrolases, DNA repair factors and epigenetic factors. Regulated R-loops participated in several important cellular events, but when dysregulated will cause genomic instability.

R-loop mapping strategies

In order to have a better understanding of R-loops, a proper R-loop mapping method is needed. Current R-loop mapping methods are grouped into two main categories (Table 1). The first category uses the S9.6 antibody to map R-loops. The S9.6 is an antibody that has specificity towards RNA:DNA hybrids[25,26]. Most of the mapping methods in this category are the derivative of DNA:RNA immunoprecipitation followed by sequencing (DRIP-seq)[27]. In DRIP-seq, the genomic DNA is first purified and fragmented by restriction enzymes. Fragmented DNA containing RNA:DNA hybrids are immunoprecipitated using the S9.6 antibody, followed by NGS. Some of the variations of DRIP include DRIPc-seq, which purifies RNA instead of DNA; bis-DRIP, which treats samples with bisulfite and converts only the cytosines on the displaced DNA strand in R-loops to uracils to generate strand specificity; and RDIP, which pre-treats samples with RNase I to eradicate ssRNA and uses sonication instead of enzymes for DNA fragmentation to increase mapping resolution[28–30]. Methods in the S9.6 antibody category share some of the same problems. First, the specificity of the S9.6 antibody towards RNA:DNA hybrid is questionable. It was observed in the fission yeast system that they have some weak affinity towards double-strand RNAs[31]. Although this was only observed in fission yeasts, it generally raises a concern as to whether the detected R-loop genome-wide has similar problems in human cells. Second, the fragmentation of genomic DNA by restriction enzymes may result in R-loop site bias and large fragment size, which reduces the mapping resolution in the NGS analysis except for RDIP[32].

The second category uses either *E. coli* RNaseH or human RNaseH1 that are catalytically inactive by introducing point mutations within its active site. Here, the modified RNaseH or RnaseH1 are referred to as RHΔ or hRHΔ, respectively. The RHΔ and hRHΔ retain their high binding affinity but lose their catalytic activity towards RNA:DNA hybrids[33,34]. Assays under this category include DNA:RNA in vitro enrichment followed by NGS (DRIVE-seq), expression for RHΔ in cells followed by chromatin immunoprecipitation (R-ChIP), and MapR[27,35,36]. The DRIVE-seq uses the purified catalytically dead human RNaseH1 combined with maltose-binding protein (MBP), called MBP-hRHΔ to target R-loops, followed by amylose affinity pull-down and NGS. However, the MBP-hRHΔ is not very efficient in affinity pull down, which makes DRIVE-seq even less sensitive compared with DRIP-seq[27]. The R-ChIP method uses a genetically modified cell-line expressing the catalytically dead RNaseH1 fused with a V5 epitope tag (RHΔ-V5)[35]. DNA is fragmented

by sonication, and the R-loops are recovered by V5 immunoprecipitation. The R-ChIP method requires time to generate stable genetically mutated cell lines expressing hRHΔ-V5. Moreover, the expression of hRHΔ-V5 may result in a dominative negative effect, thereby stabilizing transient R-loops that may skew the relevance of R-loops in different biological contexts. [35,36].

MapR is a new mapping method derived from the CUT & RUN assay[36,37]. The MapR uses the GST-RHΔ-MNase, a fusion protein with RHΔ and the micrococcal nuclease enzyme, MNase. The D10R and E48R mutations in *E. coli* RNaseH eradicate RNaseH activity, but this mutated RNaseH retains a high binding affinity towards RNA:DNA hybrids [33]. MapR uses the purified protein GST-RHΔ-MNase on cells bound with concanavalin A beads under native conditions (non-fixed cells). R-loops were targeted and bounded with the RHΔ module in permeabilized cells. The MNase enzyme, which is fused to RHΔ is activated by the addition of calcium ions and cleaves the nucleosome-wrapped DNA region flanking R-loops. The fragments are released and purified to construct a DNA library. The MapR workflow can be completed in 2 days and it only requires ~100000 cells whereas S9.6 antibody-based methods need at least 1 million cells as input material for them to get similar strength of sequence readings.

Table 1: Summary of different R-loop mapping strategies.

Categories	Examples	Advantages	Shortcomings
S9.6 antibody	DRIP-seq[27] DRIPc-seq[28] RDIP[30] bis-DRIP[29]	● Well understood ● Large amount of archived dataset ● Have strand specificity (DRIPc-seq, RDIP, bis-DRIP)	● Low resolution of Mapping ● Large amount of input required. ● Longer than DRIP-seq to finish (DRIPc-seq, RDIP, bis-DRIP)
Mutated RNaseH (Traditional)	DRIVE-seq[27] R-ChIP[35]	● High affinity towards R-loops (Both) ● RHΔ has higher sensitivity towards R-loops in vivo (R-ChIP) [35]	● Insufficient affinity pulldown (DRIVE) ● Low sensitivity (DRIVE) ● Need a stable genetically modified cell line (R-ChIP) ● Potential artificial R-loop induction (R-ChIP)

Project overview

The original MapR does not provide any pre-library sequencing quantification control. Therefore, it is difficult to know whether MapR assay works prior to NGS. This potentially makes DNA library construction and NGS more expensive and time-consuming. To overcome this, my goal is to modify the original MapR protocol, combining the general procedure of MapR with qPCR by the method of percent input analysis (MapR-qPCR) to evaluate MapR experimental and sample quality prior to NGS. In this study, I successfully developed the whole MapR-qPCR assay, as well as proof that the result obtained from MapR-qPCR analysis quantifies bona-fide R-loop accumulation level at certain loci.

Results

GST-MNase and GST-RHΔ-MNase protein expression and purification

Plasmids containing GST-MNase and GST-RHΔ-MNase genes were transformed into *E. coli* BL21(DE3) competent cells. Proteins were expressed by IPTG induction and purified using GST-agarose affinity beads (Figure 2A). Aliquots from pre- and post-purification were analyzed using SDS-PAGE stained with coomassie blue. GST-MNase (~44 kDa) and GST-RHΔ-MNase (~61 kDa) protein expressions were observed by the addition of IPTG (Figure 2B, compare lanes 2 versus 3, 3C, lanes 2 versus 3). Cells were then sonicated and bound to GST-agarose beads. The unbound fractions post-GST-binding and after each wash step was collected. As shown in Figure 2B and C lanes 5-7, the GST-MNase and GST-RHΔ-MNase proteins are reduced as they are bound to beads. After three washes, GST-MNase and GST-RHΔ-MNase proteins were eluted five times and collected into different fractions (Figure 2B and 2C, lanes 9 -13 and 8-12). Finally, all eluted fractions were pooled and concentrated (Figure 2B and 2C, lane 14 and 13, respectively Figure 2D). The molecular weight of GST-MNase and GST-RHΔ-MNase are ~40kDa and 60kDa, as previously reported[38]. Final purified proteins were analyzed on an SDS-PAGE gel. In order to purify sufficient proteins, I have conducted protein expression and purification twice. GST-MNase and GST-RHΔ-MNase protein concentrations were quantified by comparing to known BSA protein concentrations (Table 2).

Table 2: Concentrations of purified GST-MNase and GST-RHΔ-MNase proteins for 2 batches. The concentrations from the two independent purifications were evaluated using BSA standard protein (see Methods and Materials).

Protein		Concentration (µg/µL)
Batch 1	GST-MNase	2
	GST-RHΔ-MNase	1
Batch 2	GST-MNase	1
	GST-RHΔ-MNase	0.5

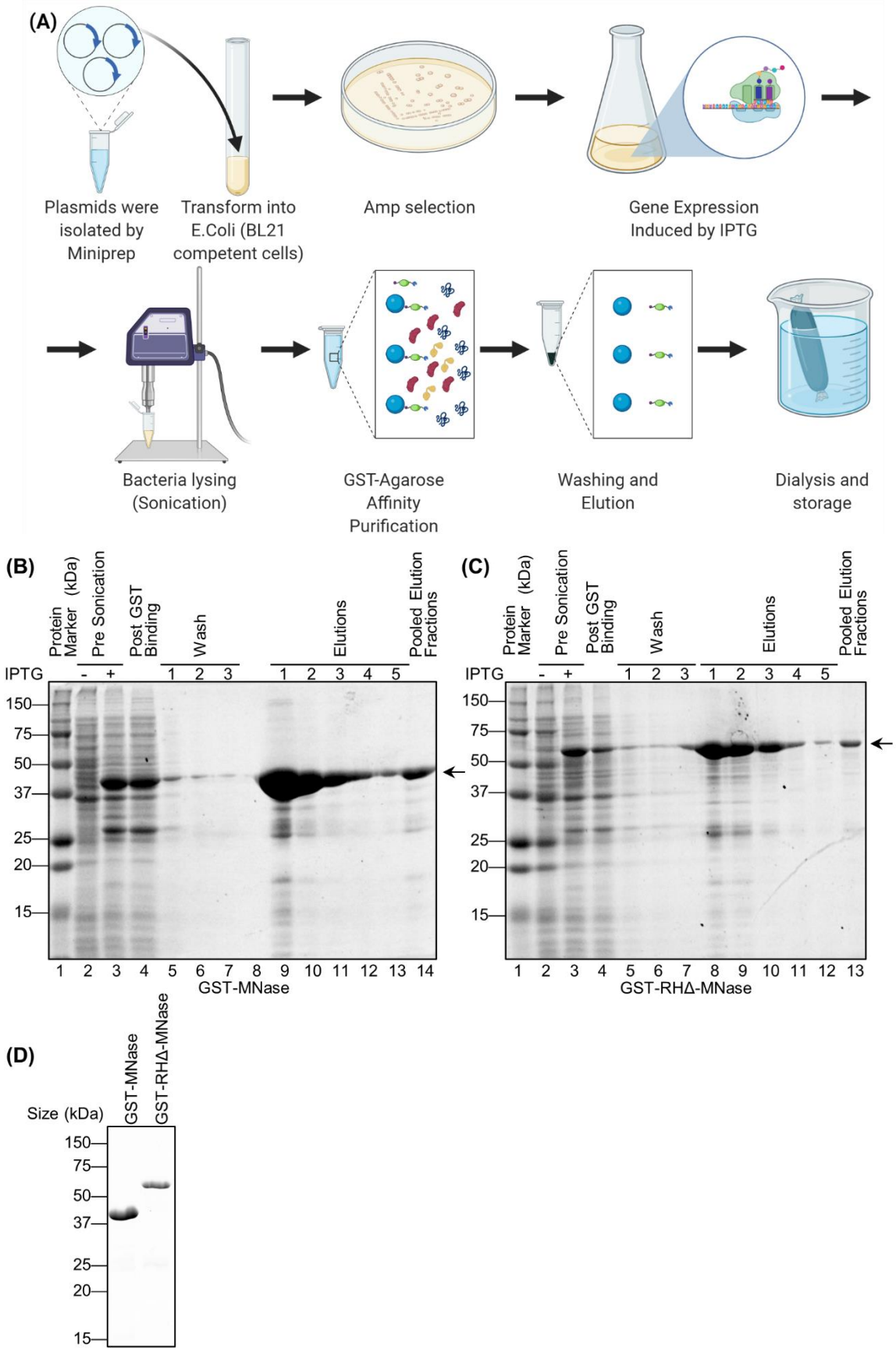
(A)
Plasmids were isolated by Miniprep
Transform into E.Coli (BL21 competent cells)
Amp selection
Gene Expression Induced by IPTG
Bacteria lysing (Sonication)
GST-Agarose Affinity Purification
Washing and Elution
Dialysis and storage
(B)
Protein Marker (kDa)
Pre Sonication
Post GST Binding
Wash
Elutions
Pooled Elution Fractions
IPTG
GST-MNase
(C)
GST-RHΔ-MNase
(D)
Size (kDa)
GST-MNase
GST-RHΔ-MNase

Figure 2: **GST-MNase and GST-RHΔ-MNase protein expression and purification. (A)** Protein Purification Workflow. **(B and C)** Protein purification quality check step by step. Each sample was taken from the supernatant after the step was done. IPTG (0.05mM) was induced 4 hours before sonication. The pre-sonication samples were taken at the same time in 2 groups: one group with IPTG induction and the other without IPTG induction. Pooled elution fraction stands for the sample taken after all 5 elutions were pooled together. **(D)** Protein gel electrophoresis of GST-MNase and GST-RHΔ-MNase. Protein was run on a 12% SDS-PAGE gel and stained by coomassie blue. kDa, kilodalton.

Examining MNase activity of purified GST-MNase and GST-RHΔ-MNase proteins

To verify that the MNase on the purified GST-MNase and GST-RHΔ-MNase are active, I performed the MNase activity assay using DNA extracted from U2OS cells. Equal numbers of cells from U2OS cells were treated with GST-MNase or GST-RHΔ-MNase to test the digestion rate of MNase on these proteins. At 0 min, both GST-MNase and GST-RHΔ-MNase group showed that most of the DNA materials were at the position with a mass above 10 Kbps, indicating intact genomic DNA wrapped in nucleosomes (Figure 3). After a 10-minute incubation, strong bands above 100 bp appeared, which is similar in size to DNA wrapped around one nucleosome (~145 base pairs, bp). Moreover, laddering band patterns were observed with DNA fragment sizes similar to multiples of 145 bp (Figure 3). This indicates that the MNase units on both our purified proteins: GST-MNase and GST-RHΔ-MNase are functioning normally, and they have a similar digestion efficiency.

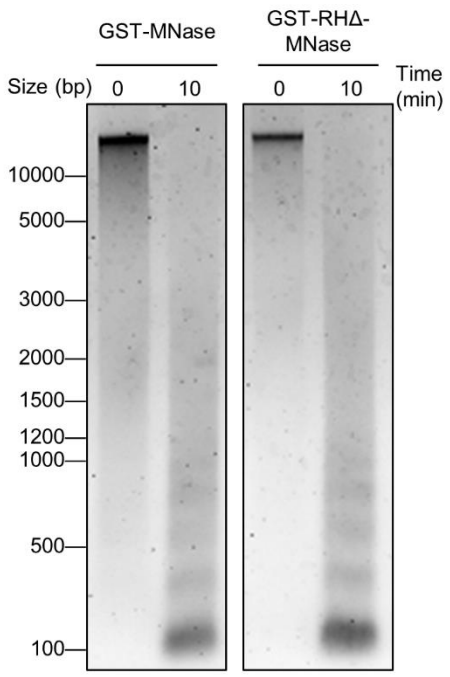

Figure 3: **Examining MNase activity of purified GST-MNase and GST-RHΔ-MNase proteins.** Samples were purified at different digestion time points and run together on 1% agarose gel. The gel was stained by ethidium bromide for imaging. bp, base pairs, min, minutes.

MapR-qPCR Design

MapR-qPCR follows the procedure of the original MapR protocol till the chromatin release and DNA purification step[38]. Then, the purified DNA was directly analyzed by qPCR with primers for loci of interest to be ready for the percent input method. An input sample is required for the percent input method as a reference to how much chromatin is used in each MapR experiment. The input samples were prepared separately from the MapR experimental groups. Only 10% of the number of the cells used in each MapR experiment were taken during cell collection for each different condition to make a 10% input sample. The input samples were only treated with MNase at 37°C for 30min for DNA fragmentation and then purified as MapR experimental samples. By applying the percent input method and using the 10% input sample as the standard, the relative cleaved DNA material by the 2 purified proteins can be directly quantified (Figure 4).

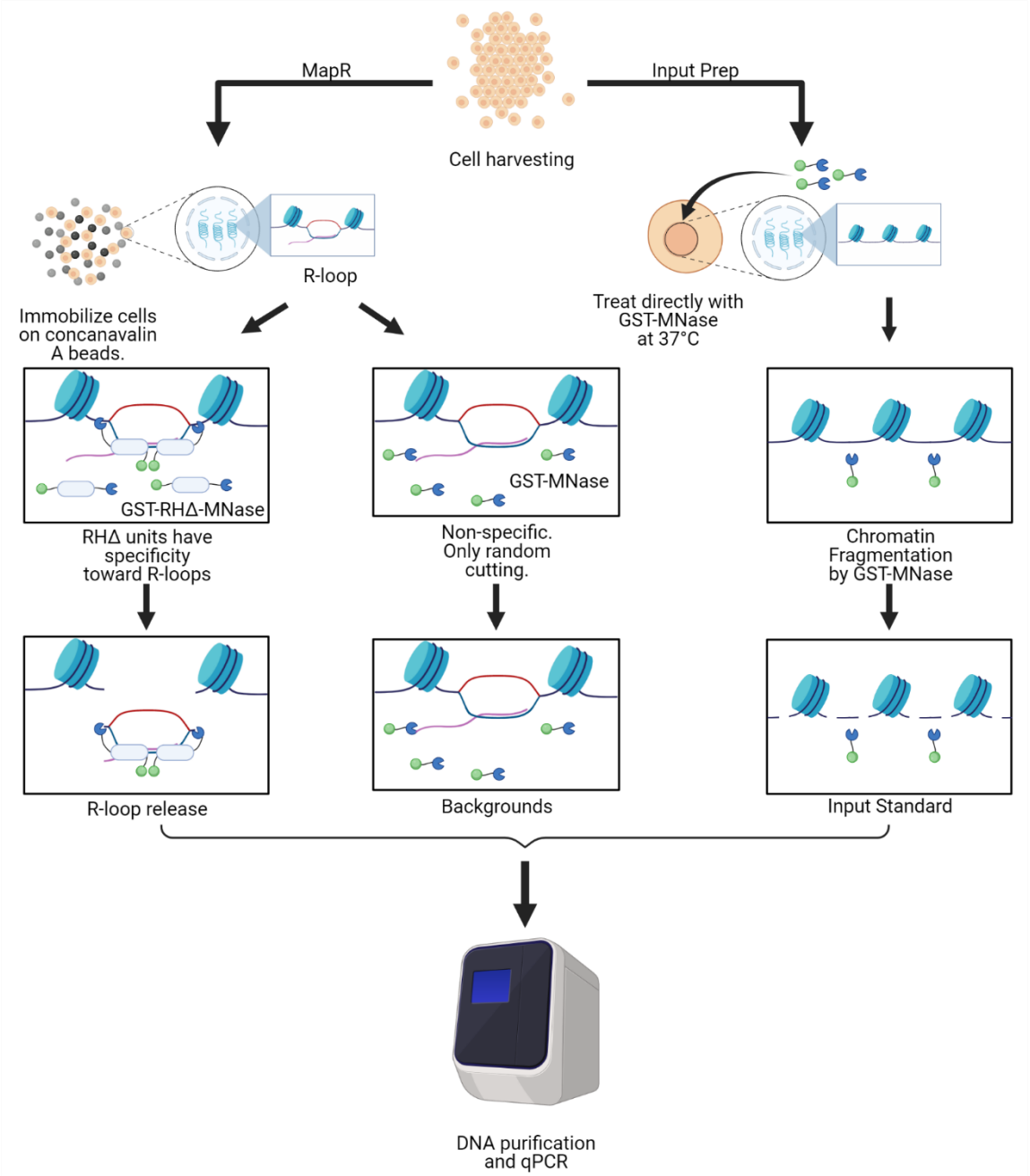

Figure 4: **MapR-qPCR design.** Cells were separated into 2 groups. One group was digested by GST-MNase at 37°C for 30 min for fragmentation and directly purified, adjusted to 100% as input. The other groups were first immobilized on concanavalin A beads, then separated into 2 different groups, treated with either GST-RHΔ-MNase or GST-MNase. These samples were purified and run on a qPCR machine with input and analyzed using the percent input method

Detection of R-loops by MapR-qPCR

Having the MapR-qPCR design with the percent input method, we performed MapR using U2OS cells. We chose RPL13A and SNRPN loci as positive and negative R-loop loci, respectively, as previously reported by DRIP-based methods[39]. Encouragingly, R-loops at RPL13A loci were highly enriched by the GST-RHΔ-MNase treated sample compared to the GST-MNase treated sample (Figure 5A, lanes 1 versus 2, lanes 3 versus 4). Moreover, R-loops detected in the GST-RHΔ-MNase treated samples were significantly higher than SNRPN loci (Figure 5A, lanes 2 versus 4). Although the increased enrichment at SNRPN loci in GST-RHΔ-MNase treated sample was observed compared to GST-MNase treatment, this might be attributed to the fact that active transcription does occur at this gene locus, thereby generating R-loops. Moreover, the difference in enrichment efficiency may be due to the differential technique used (DRIP vs MapR). To validate that the enrichment observed at the RPL13A locus is indeed R-loops, I pretreated additional samples using commercially available purified *E. coli* RNaseH, an enzyme that specifically hydrolyzes the RNA moiety within RNA:DNA hybrids[40]. Encouragingly, RNaseH pretreatment in GST-RHΔ-MNase treated sample reduced partial but significant, R-loop enrichment at RPL13A locus but not at SNRPN locus (Figure 5B, lanes 2 versus 4 under RPL13A locus and SNRPN locus). The suppression observed by RNaseH is around 8%. This partial suppression was observed in previous studies as well, indicating that R-loops formed at certain genomic regions have resistance towards RNaseH[28,41]. Altogether, the results obtained from MapR-qPCR quantified R-loop accumulation at different loci that were predicted to have different R-loop levels.

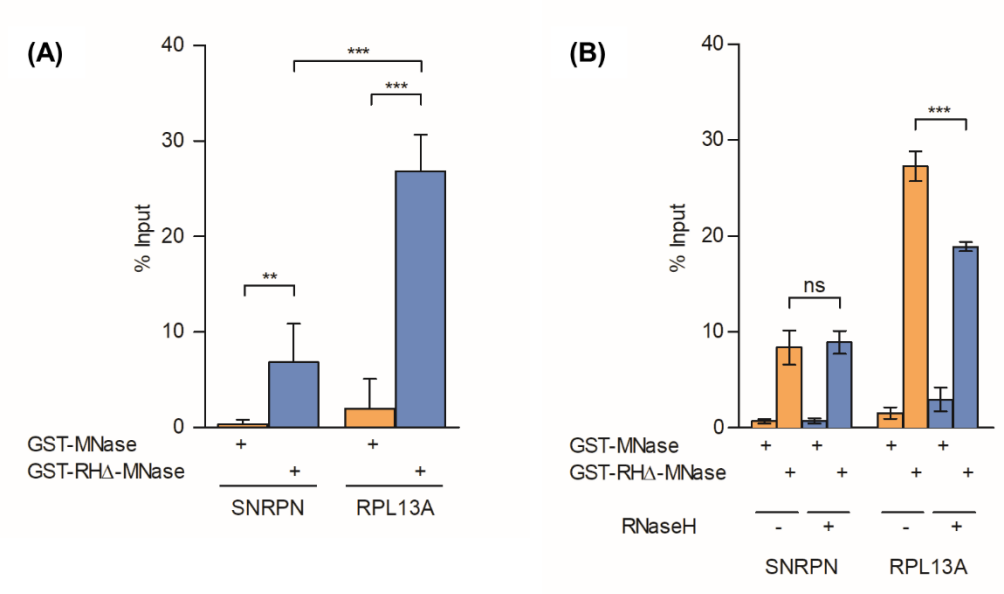

Figure 5: **Detection of R-loops by MapR-qPCR. (A)** GST-MNase and GST-RHΔ-MNase show significant affinity differences at positive (RPL13A) and negative (SNRPN) loci. The statistics were done by doing an unpaired T-test. **(B)** Input percentage of RNaseH treatment group and control group at positive (RPL13A) and negative (SNRPN) loci. The statistics were done by doing an unpaired T-test. ***, p≤0.001. **, p≤0.01. ns, not significant.

Discussion

In this study, I established the MapR workflow and developed a method to quantify R-loop enrichment by MapR-qPCR. I successfully purified the GST-RHΔ-MNase and GST-MNase proteins, showing similar MNase activities (Figure 2, 3). The GST-RHΔ-MNase treatment was able to significantly isolate known R-loops at the RPL13A positive locus more than that at the SNRPN negative locus (Figure 5). Finally, I also managed to partially suppress the R-loop enrichment at the RPL13A locus in the GST-RHΔ-MNase by RNaseH treatment. Several studies have also shown partial suppression of RNaseH treatment at R-loop levels [28,41]. Since MapR is done under native conditions, R-loop binding proteins may bind to RNA:DNA hybrids and protect them from being digested by RNaseH, causing this partial suppression[28]. Although the MapR-qPCR method was established, R-loop suppression at the positive locus in MapR-qPCR needs further optimization. If purified RNaseH from *E. coli* does not improve the suppression, one can consider suppressing R-loops in cells prior to the MapR process. For example, since R-loops are co-transcriptionally regulated, treating cells with transcription inhibitors such as actinomycin D or cordycepin to inhibit transcription would suppress R-loop formation[36,42]. This would theoretically result in a decrease of R-loop enrichment by the GST-RHΔ-MNase treatment compared to non-treated samples. Alternatively, overexpression of human RNaseH1 in cells to resolve R-loops in cells prior to MapR may show a similar effect. In Summary, the result has shown that GST-RHΔ-MNase strongly enriched R-loop level at the R-loop positive locus, and this enrichment can be partially suppressed by RNaseH treatment. As a result, MapR-qPCR is potent to map bona-fide R-loops in cells.

Methods and Materials

Cell line

The U2OS cells were grown in DMEM media supplemented with 10% FBS, 1%L-glutamate and 1% PEN/STREP. All cells were grown at a 37°C incubator for 20 passages before discarding.

Transformation and Expression

Plasmids (GST-MNase #136291, GST-RHΔ-MNase #136292) were ordered from Addgene, purified by a Miniprep kit (IBI Scientific 47101), and confirmed by Sanger sequencing. The DNA plasmids were then transformed into *E. coli* BL21(DE3) competent cells and selected on LB agar plates containing 100mg/mL ampicillin. Single bacterial colonies were first inoculated for at least 16 hours in 4mL LB broth with 100mg/mL ampicillin, then transferred into 500mL LB broth with 100mg/mL ampicillin. The growth of bacteria was monitored until the OD_{600} reached 2.0-2.5 using the instrument Cytation 5 (BioTek). IPTG (1M) was added into the LB broth and incubated for 3 hours. 4mL of bacterial cultures were taken separately and weren't treated with IPTG to serve as negative protein expression controls. After 3 hours, a 10μL aliquot of both negative control suspension and IPTG included suspension was taken and mixed with a 10μL sample buffer for SDS-PAGE gel analysis. The bacterial cultures were pelleted (5000g x 10 min) at 4°C and stored in -80°C freezer until protein purification.

Protein Purification

The cells were lysed by sonication. The supernatants were treated with GST-Agarose beads and incubated on a rotisserie at 4°C overnight. Beads were washed at least 3 times with 1x PBS (pH=7.4) at 4°C for 5 min per wash, then incubated in the GST-elution buffer (125 mM Tris·HCl, pH 8.0, 150 mM NaCl, 10 mM reduced L-glutathione) at 4°C for 30 min. Four additional elution steps were conducted. Purified proteins were pooled and dialyzed against sufficient BC100 buffer (50 mM Tris·HCl, pH 7.6, 2 mM EDTA, pH 8.0, 100 mM KCl, 10% volume by volume glycerol, 0.1 mM DTT and 0.2 mM PMSF) for at least 6 hours with changing buffer per 2 hours. Protein extracts after dialysis were stored in 25% glycerol in a -80°C freezer. A 10μL aliquot of supernatant was taken in each step of this section

and mixed with the 2x sample buffer, run on a 12% SDS-PAGE gel and stained using SimplyBlue™ SafeStain.

Protein Concentration Check

Protein concentrations were standardized by BSA standard proteins. 2μg/μL BSA standard protein were made into a serial dilution of 1μg/μL, 0.4μg/μL, 0.2μg/μL, 0.1μg/μL, 0.04μg/μL and 0.02μg/μL in the 1x sample buffer. Unknown concentrations of GST-RHΔ-MNase and GST-MNase were made into a serial dilution of 2μL protein mixture/5μL, 1μL protein mixture/5μL, 0.5μL protein mixture/5μL, 0.2μL protein mixture/5μL and 0.1μL protein mixture/5μL in the 1x sample buffer. 5μL of all samples were run on a 12% SDS-PAGE gel and stained using SimplyBlue™ SafeStain. The concentrations of purified proteins were measured against the BSA standards.

MNase Activity Analysis

The MNase activity assay was performed as previously reported [39]. Approximately 300,000 cells were collected, pelleted (1000xg, 10min) at 4°C, then washed twice with 1x PBS. Cells were resuspended in 400μL Buffer A (10 mM MES pH 6.5, 0.25M Sucrose, 60 mM KCl,15 mM NaCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 0.5% Triton X-100, 0.5 mM PMSF) and incubate on ice for 20 minutes. Cells were pelleted by cold centrifuge and resuspended in 160μL Buffer B (10 mM PIPES pH 6.8, 50 mM NaCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 0.1 mM PMSF). Cell suspensions were equally divided into 2 groups. 1μM GST-RHΔ-MNase or GST-MNase were added separately in each tube. 25μL cell suspension from each group were taken immediately into a new Eppendorf tube, 30μL MNase stop buffer (100mM EDTA, 10mM EGTA, pH=7.5) were added. These tubes were labeled as 0 min. The rest of the cells were incubated at 37°C for 10 minutes. 25μL cell suspension from each group was taken and 30μL of MNase stop buffer was added. These tubes were labeled as 10 min. DNA from all groups were purified using the Monarch PCR & DNA Cleanup Kit. Purified DNA was run on an agarose gel for analysis.

MapR

The MapR procedure was performed as previously reported with some modifications [39]. The Concanavalin A beads (10μL per experiment) were activated before the experiment by washing them with the binding buffer (20 mM HEPES-KOH, pH 7.9, 10 mM KCl, 1 mM

$CaCl_2$, 1 mM $MnCl_2$) 3 times. Beads were resuspended in the wash buffer (20 mM HEPES-NaOH, pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 1 mM cOmplete Protease Inhibitor Cocktail). Four million cells per condition were collected and ~200,000 cells per condition were collected for input preparation (see below). The cells were washed 3 times with the wash buffer. Concanavalin A beads were added into the cell suspension and the whole suspension was rotated at room temperature for 1 hour to immobilize cells to concanavalin A beads. Immobilized cells were resuspended in 400µL dig-wash buffer (20 mM HEPES-NaOH, pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 1 mM cOmplete Protease Inhibitor Cocktail, 0.2% digitonin) and split into two tubes (200µL each). 1µM of either GST-RHdelta-MNase or GST-MNase purified proteins were added into each experimental group. The beads were rotated on a rotisserie at 4°C. After the overnight incubation, the beads were washed three times with the dig-wash buffer and chilled in an ice water bath to reach 0°C. 2µL of 0.1M $CaCl_2$ were added to activate the MNase enzyme and incubated for 30 minutes. The 2x stop buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 0.02% Digitonin, 5 mg RNaseA, 5 mg linear acrylamide) was added to the bead suspension to stop the reaction and all tubes with beads were incubated in a 37°C water bath for 15 minutes to release the chromatin fragments that contain RNA:DNA hybrids. Suspensions with beads were pelleted down at high speed. The supernatant was transferred to a new tube and incubated with 2µL of 10% SDS and 2.5µL of proteinase K. RNA:DNA hybrids were purified using the Monarch PCR & DNA Cleanup Kit (BioLab T1030S).

RNaseH Treatment

In cases where RNaseH treatment is needed, cells were bound to the concanavalin A beads and before the purified proteins were added to the MapR protocol, all supernatants in the suspension were withdrawn for all groups including input samples. Cells were resuspended in 50µL of dig-wash buffer. 150 units of RNaseH (NEB M0297L) were added to each tube including input samples. Equal amounts of dig-wash buffer were added to all control groups, including samples. Both control and RNaseH treated groups were incubated at 25°C for 1 hour. All supernatants in the suspension were withdrawn for all groups including input samples. The cells were resuspended in 200µL dig-wash buffers and purified proteins were added. The rest of the MapR protocol was processed as normal.

Input Preparation

200,000 cells per condition were collected and washed with the wash buffer. Cells were resuspended in 200μL of the dig-wash buffer. 1μM GST-MNase was added to the cell suspension. 4μL 0.1M $CaCl_2$ were added into each group of cell suspensions and the suspensions were incubated in a 37°C water bath for 30 minutes to perform enzyme digestion. The 2x Stop buffer was added to the bead suspension to stop the reaction and all tubes with beads were incubated in a 37°C water bath to release the chromatin fragments. 2μL 10% SDS and 2.5μL proteinase K were added to each group of suspensions to remove protein. The DNA fragments in the suspension were purified using the Monarch PCR & DNA Cleanup Kit.

qPCR

20μL of reaction mixtures, including 2μL of purified DNA sample, 10μL of QuantaBio Perfecta, 7μL of nuclease-free water, and 1μL of 6μM forward and reverse primer mixture were set up for each qPCR reaction. Sequences of the primers used in the experiment were included in Table 3. Each sample was done in triplicates and averaged to minimize the error. For each of the qPCR runs, the activation step was set as 95°C for 3 min. Each cycle was performed in 2 steps: 95°C for 10 seconds as step 1, and 60°C for 30 seconds as step 2. 40 cycles in total were performed in one run and there was no cooling down phase. Ct values were obtained after qPCR runs for both input samples and experimental group samples. The adjusted Ct values for the input (Ct_{Ad}) were calculated by the following equation: $Ct_{Ad}=Ct_{input}-\log_2 \frac{100}{\text{input percentage}}$. The adjusted input value represents the value when the input percentage is 100%. The percent input values of each experimental group (Ct_{Sample}) were calculated by the following equation: $\%\ \text{input}=100*2^{Ct_{Sample}-Ct_{Ad}}$. The percent input for each sample was calculated and averaged to generate the final data representing the enrichment for a whole MapR run at a certain locus.

Table 3: Oligos used for qPCR

Primer (Loci) Name	Sequence	Vendor
RPL13A	Forward: 5' AAT GTG GCA TTT CCT TCT CG 3' Reverse: 5' CCA ATT CGG CCA AGA CTC TA 3'	IDT
SNRPN	Forward: 5' GCC AAA TGA GTG AGG ATG GT 3' Reverse: 5' TCC TCT CTG CCT GAC TCC AT 3'	IDT

Bibliography

1. Thomas, M., White, R. L. & Davis, R. W. Hybridization of RNA to double-stranded DNA: formation of R-loops. *Proceedings of the National Academy of Sciences* **73**, 2294–2298 (1976).
2. Grunseich, C. *et al.* Senataxin Mutation Reveals How R-Loops Promote Transcription by Blocking DNA Methylation at Gene Promoters. *Molecular Cell* **69**, 426-437.e7 (2018).
3. Skourti-Stathaki, K., Kamieniarz-Gdula, K. & Proudfoot, N. J. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* **516**, 436–439 (2014).
4. Chaudhuri, J. & Alt, F. W. Class-switch recombination: Interplay of transcription, DNA deamination and DNA repair. *Nature Reviews Immunology* vol. 4 541–552 (2004).
5. Yasuhara, T. *et al.* Human Rad52 Promotes XPG-Mediated R-loop Processing to Initiate Transcription-Associated Homologous Recombination Repair. *Cell* **175**, 558-570.e11 (2018).
6. Pohjoismäki, J. L. O. *et al.* Mammalian mitochondrial DNA replication intermediates are essentially duplex but contain extensive tracts of RNA/DNA hybrid. *Journal of Molecular Biology* **397**, 1144–1155 (2010).
7. Gan, W. *et al.* R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes and Development* **25**, 2041–2056 (2011).
8. Belotserkovskii, B. P. *et al.* Mechanisms and implications of transcription blockage by guanine-rich DNA sequences. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 12816–12821 (2010).
9. Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M. E. & Nudler, E. Linking RNA polymerase backtracking to genome instability in E. coli. *Cell* **146**, 533–543 (2011).
10. Chen, L. *et al.* The Augmented R-Loop Is a Unifying Mechanism for Myelodysplastic Syndromes Induced by High-Risk Splicing Factor Mutations. *Molecular Cell* **69**, 412-425.e6 (2018).
11. Yang, Y. *et al.* Arginine Methylation Facilitates the Recruitment of TOP3B to Chromatin to Prevent R Loop Accumulation. *Molecular Cell* **53**, 484–497 (2014).
12. Schwab, R. A. *et al.* The Fanconi Anemia Pathway Maintains Genome Stability by Coordinating Replication and Transcription. *Molecular Cell* **60**, 351–361 (2015).
13. el Hage, A., French, S. L., Beyer, A. L. & Tollervey, D. Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes and Development* **24**, 1546–1558 (2010).
14. Li, X. & Manley, J. L. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122**, 365–378 (2005).
15. Domínguez-Sánchez, M. S., Barroso, S., Gómez-González, B., Luna, R. & Aguilera, A. Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genetics* **7**, (2011).

16. Skourti-Stathaki, K., Proudfoot, N. J. & Gromak, N. Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. *Molecular Cell* **42**, 794–805 (2011).
17. Villarreal, O. D., Mersaoui, S. Y., Yu, Z., Masson, J. Y. & Richard, S. Genome-wide R-loop analysis defines unique roles for DDX5, XRN2, and PRMT5 in DNA/RNA hybrid resolution. *Life Science Alliance* **3**, 1–14 (2020).
18. Chakraborty, P. & Grosse, F. Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair* **10**, 654–665 (2011).
19. Song, C., Hotz-Wagenblatt, A., Voit, R. & Grummt, I. SIRT7 and the DEAD-box helicase DDX21 cooperate to resolve genomic R loops and safeguard genome stability. *Genes and Development* **31**, 1370–1381 (2017).
20. Parajuli, S. *et al.* Human ribonuclease H1 resolves R-loops and thereby enables progression of the DNA replication fork. *Journal of Biological Chemistry* **292**, 15216–15224 (2017).
21. Cornelio, D. A., Sedam, H. N. C., Ferrarezi, J. A., Sampaio, N. M. V. & Argueso, J. L. Both R-loop removal and ribonucleotide excision repair activities of RNase H2 contribute substantially to chromosome stability. *DNA Repair* **52**, 110–114 (2017).
22. Hatchi, E. *et al.* BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Molecular Cell* **57**, 636–647 (2015).
23. Bhatia, V. *et al.* BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* **511**, 362–365 (2014).
24. Matos, D. A. *et al.* ATR Protects the Genome against R Loops through a MUS81-Triggered Feedback Loop. *Molecular Cell* **77**, 514-527.e4 (2020).
25. Boguslawski, S. J. *et al.* Characterization of monoclonal antibody to DNA · RNA and its application to immunodetection of hybrids. *Journal of Immunological Methods* **89**, 123–130 (1986).
26. Hu, Z., Zhang, A., Storz, G., Gottesman, S. & Leppla, S. H. An antibody-based microarray assay for small RNA detection. *Nucleic Acids Research* **34**, e52–e52 (2006).
27. Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I. & Chédin, F. R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters. *Molecular Cell* **45**, 814–825 (2012).
28. Sanz, L. A. & Chédin, F. High-resolution, strand-specific R-loop mapping via S9.6-based DNA–RNA immunoprecipitation and high-throughput sequencing. *Nature Protocols* **14**, 1734–1755 (2019).
29. Dumelie, J. G. & Jaffrey, S. R. Defining the location of promoter-associated R-loops at near-nucleotide resolution using bisDRIP-seq. *eLife* **6**, (2017).
30. Nadel, J. *et al.* RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. *Epigenetics and Chromatin* **8**, 1–19 (2015).

31. Hartono, S. R. *et al.* The Affinity of the S9.6 Antibody for Double-Stranded RNAs Impacts the Accurate Mapping of R-Loops in Fission Yeast. *Journal of Molecular Biology* **430**, 272–284 (2018).
32. Halász, L. *et al.* RNA-DNA hybrid (R-loop) immunoprecipitation mapping: An analytical workflow to evaluate inherent biases. *Genome Research* **27**, 1063–1073 (2017).
33. Tsunaka, Y., Haruki, M., Morikawa, M. & Kanaya, S. Strong nucleic acid binding to the Escherichia coli RNase HI mutant with two arginine residues at the active site. *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology* **1547**, 135–142 (2001).
34. Wu, H., Lima, W. F. & Crooke, S. T. Investigating the Structure of Human RNase H1 by Site-directed Mutagenesis. *Journal of Biological Chemistry* **276**, 23547–23553 (2001).
35. Chen, L. *et al.* R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. *Molecular Cell* **68**, 745-757.e5 (2017).
36. Yan, Q., Shields, E. J., Bonasio, R. & Sarma, K. Mapping Native R-Loops Genome-wide Using a Targeted Nuclease Approach. *Cell Reports* **29**, 1369-1380.e5 (2019).
37. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, (2017).
38. Yan, Q. & Sarma, K. MapR: A Method for Identifying Native R-Loops Genome Wide. *Current Protocols in Molecular Biology* **130**, 1–12 (2020).
39. Wood, M. *et al.* TDP-43 dysfunction results in R-loop accumulation and DNA replication defects. *Journal of Cell Science* **133**, (2020).
40. Berkower, I., Leis, J. & Hurwitz, J. Isolation and characterization of an endonuclease from Escherichia coli specific for ribonucleic acid in ribonucleic acid deoxyribonucleic acid hybrid structures. *Journal of Biological Chemistry* **248**, 5914–5921 (1973).
41. Wang, K. *et al.* Genomic profiling of native R loops with a DNA-RNA hybrid recognition sensor. *Science Advances* **7**, 3516–3533 (2021).
42. Nascakova, Z. *et al.* Rad51 inhibition induces r-loop formation in early g1 phase of the cell cycle. *International Journal of Molecular Sciences* **22**, (2021).