Ribose-seq: ribonucleotides in DNA to Illumina library

Jay Hesselberth (jay.hesselberth@gmail.com)
Department of Biochemistry and Molecular Genetics, University of Colorado Anschutz Medical School
Francesca Storici (storici@gatech.edu)
School of Biology, Georgia Institute of Technology
Kyung Duk Koh
School of Biology, Georgia Institute of Technology

Method Article

Keywords: Ribonucleotide; ribonucleotide in DNA; DNA sequencing.

Posted Date: May 19th, 2015

DOI: https://doi.org/10.1038/protex.2015.044

Abstract

Ribonucleotides are now widely considered as the most frequently incorporated non-canonical nucleotides in genomic DNA. Here, we describe an approach, ribose-seq, to prepare DNA libraries for next-generation sequencing to map ribonucleotide incorporation into DNA. By specifically targeting and capturing the unique ends of alkali-derived 2',3'-cyclic monophosphate or 2'-monophosphate from ribonucleotides embedded in DNA, ribose-seq libraries are constructed. Upon high-throughput sequencing and analysis, the distribution and identity of ribonucleotides in the genome could be determined as recently reported using yeast as a model organism1. Ribose-seq could potentially be applied to any cell type of any organism to allow profiling of ribonucleotide incorporation into genomic DNA.

Introduction

Ribonucleotides, also known as ribonucleoside 5'-monophosphates \(rNMPs), which are normally monomers of RNA, have been found to be the most abundant non-canonical nucleotides incorporated into DNA2,3. Sources of ribonucleotide incorporation into DNA include incomplete Okazaki fragment maturation, oxidative damage, and numerous DNA polymerases. Quantitative measures of rNMPs in DNA following alkali treatment of yeast genomic DNA derived from RNase H2-deficient cells estimated ~2,000 rNMPs per genome4,5 while similar measurement conducted for genomic DNA derived from RNase H2-deficient embryonic fibroblasts revealed the presence of more than one million rNMPs in the mouse genome6. Ribonuclease H type 2 \(RNase H2) is a major protein factor involved in repair of rNMPs in DNA and initiates ribonucleotide excision repair2,3. If not removed, rNMPs have been shown to have both positive roles, such as acting as strand discrimination signal during MMR, and negative consequences, including replication stress and genome instability2,3. Despite the numerous studies showing the presence of rNMPs in DNA, information about the distribution and identity of these rNMPs in genomic DNA was unknown till very recently. We developed ribose-seq to capture rNMPs in genomic DNA along with their upstream DNA sequence for library construction, next-generation sequencing, and analysis1. Implementation of ribose-seq to genomic DNA from RNase H2-deficient yeast Saccharomyces cerevisiae cells revealed widespread but non-random distribution of rNMPs in yeast genome, with specific base preferences, neighboring DNA sequence preferences, bias of leading and lagging strand, and hotspots. Ribose-seq allows us to explore rNMP incorporation into DNA potentially in any cell type of any organism and opens up a new direction to better comprehend the impact of rNMPs on genome integrity. Here, we detail the steps of ribose-seq library construction as shown in Figure 1. Ribose-seq employs the unique capacity of Arabidopsis thaliana tRNA ligase \(AtRNL) to ligate 2',3'-cyclic monophosphate or 2'-monophosphate termini of DNA and RNA to 5'-monophosphate ends of DNA and RNA7,8. A double-stranded \(ds) sequencing adaptor, which has both primer sequences for later PCR amplification step, is initially ligated to genomic DNA fragments as AtRNL prefers self-ligation or circularization. Alkali treatment exposes the rNMP-specific 2',3'-cyclic monophosphate or 2'-monophosphate termini. Upon ligation by AtRNL, the rNMP is captured along with its upstream DNA sequence while unligated linear single-stranded \(ss) DNA is degraded with T5 exonuclease. Yeast 2'-phosphotransferase Tpt1 then

removes the 2'-monophosphate at the ligation junction, allowing subsequent PCR amplification of the library. The details given in this protocol correspond to Illumina library to be sequenced on MiSeq with 50-cycle single-end reads.

Reagents

1. DNase/RNase-free water 2. Qubit dsDNA HS Assay Kit \(Life Technologies) 3. Adaptor.L oligonucleotide \(PAGE-purified, 5'-P-NNNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAG AGGGAGTTCAGACGTGTGCTCTTCCGATCTAGCCAGCGCAGACCGTGAGGT-3') \(Integrated DNA Technologies) 4. Adaptor.S oligonucleotide \(Desalted, 5'-P-CCTCACGGTCTGCGCTGGCT-Am-3') \(Integrated DNA Technologies) 5. Annealing Buffer \(500 mM Tris-HCl, pH 7.5, 2.5 M NaCl, 50 mM EDTA) 6. illustra MicroSpin G-25 Column \(GE Healthcare Life Sciences) 7. 10X NEBuffer 2 \(New England Biolabs) 8. 20U/uL DraI \(New England Biolabs) 9. 20U/uL EcoRV \(New England Biolabs) 10. 5U/uL SspI \(New England Biolabs) 11. PCR Purification Kit \(QIAGEN) 12. Experion DNA 12K Analysis Kit \(BIO-RAD) 13. 10 mM dATP \(Sigma-Aldrich) 14. 5U/uL Klenow Fragment \(3'→5' exo–) \(New England Biolabs) 15. 400U/uL T4 DNA Ligase with 10X T4 DNA Ligase Reaction Buffer \(New England Biolabs) 16. Agencourt RNAClean XP \(Beckman Coulter) 17. 2 M NaOH 18. 2 M HCl 19. pH Litmus Paper 20. 3.75 uM purified AtRNL protein 21. 10X AtRNL Reaction Buffer \(500 mM Tris-HCl, pH 7.5, 400 mM NaCl, 50 mM MgCl2, 10 mM DTT, 300 µM ATP) 22. 10U/uL T5 Exonuclease with 10X NEBuffer 4 \(New England Biolabs) 23. 37 uM purified Tpt1 protein 24. 10X Tpt1 Reaction Buffer \(200 mM Tris-HCl, pH 7.5, 50 mM MgCl2, 1 mM DTT, 4% Triton X-100) 25. 50 mM NAD+ \(Sigma-Aldrich) 26. PCR.1.Index oligonucleotide \(Desalted, 5'-CAAGCAGAAGACGGCATACGAGATcgtgatGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC-3') \(Eurofins MWG Operon) 27. PCR.2 oligonucleotide \(Desalted, 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3') 28. 2U/uL Phusion DNA Polymerase with 5X HF Buffer, 100% DMSO, and 10 mM dNTPs \(New England Biolabs) 29. 6% Non-denaturing Polyacrylamide \(29:1) Mini-Gel 30. 1X TBE Buffer 31. SYBR Gold Nucleic Acid Stain \(Life Technologies)

Equipment

1. Qubit 2.0 \(Life Technologies) 2. Experion Automated Electrophoresis System \(BIO-RAD) 3. NanoDrop Spectrophotometer \(Thermo Scientific) 4. Thermal Cycler for PCR 5. Mini-PROTEAN Tetra Cell \(BIO-RAD) 6. Gel Imaging System \(UV-light)

Procedure

A. Preparation of rNMPs-embedded genomic DNA 1. Extract a minimum of 10 ug of genomic DNA from the cells. Various methods of genomic DNA extraction could be used. Koh et al.1 used S. cerevisiae cells and the protocol “Preparation of Yeast Samples” in the QIAGEN Genomic DNA Handbook. 3. Use Qubit 2.0 \(dsDNA HS) to quantify the amount of extracted genomic DNA. **B. Preparation of ds sequencing adaptor** 1. Resuspend each of oligonucleotides \(oligos) Adaptor.L and Adaptor.S in

DNase/RNase-free water to a concentration of 50 uM and 500 uM, respectively. \(All subsequent steps of the protocol use DNase/RNase-free water). 2. Set up the mixture for annealing Adaptor.L/Adaptor.S as follows: Annealing Buffer: 3 uL 50 uM Adaptor.L: 25 uL 500 uM Adaptor.S: 12.5 uL H_2O: 9.5 uL Total: 50 uL 5-fold excess of Adaptor.S was added to the mixture to ensure that all Adaptor.L molecules are annealed to Adaptor.S. The remaining single-stranded Adaptor.S will be removed in subsequent purification steps. 3. Perform annealing by heating the mixture to 95–100 °C and gradually cooling to room temperature. The resulting ds Adaptor.L/Adaptor.S is at a concentration of 25 uM. 4. Desalt the mixture by using a spin column. Koh et al.1 used illustra MicroSpin G-25 Column. 5. Use NanoDrop to quantify the amount of desalted ds Adaptor.L/Adaptor.S. Typically, the yielded concentration ranges from 10 to 13 uM. A concentration of 10 uM will be assumed for subsequent steps of the protocol. **C. Fragmentation of rNMPs-embedded genomic DNA** 1. Set up 2 identical reactions of restriction enzyme digestion of genomic DNA as follows \(for S. cerevisiae genomic DNA as an example): 10X NEBuffer 2: 12 uL Genomic DNA: 5 ug \(x uL) 20U/uL DraI: 4 uL 20U/uL EcoRV: 4 uL 5U/uL SspI: 4 uL H_2O: 96-x uL Total: 120 uL Genomic DNA could be digested with a different set of blunt-end resulting restriction enzymes. However, one need to make sure that the restriction sites are well-distributed in the genome and that the digestion results in a population of fragments with an average size of 800–1,500 bp. 2. Incubate at 37 °C overnight. 3. Purify the fragmented DNA by using a spin column. Koh et al.1 used the QIAGEN spin column from their PCR Purification Kit. Both reactions can be purified using a single column with elution volume of 30 uL. 4. Use Qubit 2.0 \(dsDNA HS) to quantify the amount of fragmented DNA. Typically, the concentration of the resulting DNA is ~200 ng/uL, following the reaction conditions listed above. A concentration of 200 ng/uL will be assumed for subsequent steps of the protocol. 5. Check the size range of the fragmented DNA by using the Experion DNA 12K Analysis Kit. Typically, with the reaction conditions listed above, the fragmentation results in an average size of ~1,500 bp. **D. dA-tailing and ds sequencing adaptor-ligation of fragmented rNMPs-embedded DNA** 1. Set up a dA-tailing reaction as follows: 10X NEBuffer 2: 5 uL 10 mM dATP: 1 uL 200 ng/uL Fragmented DNA: 25 uL \(5 ug) 5U/uL Klenow Fragment \(3'→5' exo–): 3 uL H_2O: 16 uL Total: 50 uL 2. Incubate at 37 °C for 30 min. 3. Purify using a spin column. Koh et al.1 used the QIAGEN spin column from their PCR Purification Kit, with elution volume of 30 uL. 4. Set up a sequencing adaptor-ligation reaction as follows: 10X T4 DNA Ligase Buffer: 5 uL 10 uM Adaptor.L/Adaptor.S: 5 uL dA-tailed DNA: 30 uL 400U/uL T4 DNA Ligase: 5 uL H_2O: 5 uL Total: 50 uL 5. Incubate at 15 °C overnight. 6. Purify using Agencourt RNAClean XP with elution volume of 30 uL. **E. Alkali treatment of adaptor-ligated rNMPs-embedded DNA** 1. Set up an alkali-treatment reaction as follows: 2 M NaOH: 7.5 uL Adaptor-ligated DNA: 30 uL H_2O: 12.5 uL Total: 50 uL 2. Incubate at 55 °C for 2 h. 3. Neutralize with 2 M HCl to pH 7. Use pH Litmus Paper to check the pH. Typically, 7.5–8 uL is needed for neutralization. 4. Purify using Agencourt RNAClean XP with elution volume of 20 uL. 5. Heat the resulting solution at 95 °C for 3 min to ensure denaturation of dsDNA and immediately chill on ice. **F. Self-ligation \(circularization) of rNMP-terminating DNA by AtRNL** 1. Set up 2 reactions, one without AtRNL \(AtRNL–) and one with \(AtRNL+), as follows: AtRNL– 10X AtRNL Reaction Buffer: 2 uL Alkali-treated DNA: 10 uL H_2O: 8 uL Total: 20 uL AtRNL+ 10X AtRNL Reaction Buffer: 2 uL Alkali-treated DNA: 10 uL 3.75 uM AtRNL: 5.4 uL H_2O: 2.6 uL Total: 20 uL Final reaction

concentration of AtRNL is 1 uM. 3. Incubate all reactions at 30 °C for 1 h. 4. Purify each reaction using RNAClean XP with elution volume of 30 uL. **G. Removal of linear ssDNA** 1. Set up 4 reactions, one without \(AtRNL– T5Exo–; AtRNL+ T5Exo–) and one with T5 Exonuclease \(AtRNL– T5Exo+; AtRNL+ T5Exo+) for each of AtRNL– and AtRNL+ product, as follows: AtRNL– T5Exo– 10X NEBuffer 4: 5 uL AtRNL– DNA: 15 uL H_2O: 30 uL Total: 50 uL AtRNL– T5Exo+ 10X NEBuffer 4: 5 uL AtRNL– DNA: 15 uL 10U/uL T5 Exonuclease: 5 uL H_2O: 25 uL Total: 50 uL AtRNL+ T5Exo– 10X NEBuffer 4: 5 uL AtRNL+ DNA: 15 uL H_2O: 25 uL Total: 50 uL AtRNL+ T5Exo+ 10X NEBuffer 4: 5 uL AtRNL+ DNA: 15 uL 10U/uL T5 Exonuclease: 5 uL H_2O: 25 uL Total: 50 uL T5Exo– samples may not be necessary as dA-tailing and adaptor-ligation reactions are standard steps. However, they do act as positive control for the later PCR reaction. 2. Incubate all reactions at 37 °C for 2 h. 3. Purify each reaction using RNAClean XP with elution volume of 20 uL. **H. Removal of 2'-phosphate at the ligation junction** 1. Set up 4 reactions with Tpt1 for each of AtRNL– T5Exo–, AtRNL– T5Exo+, AtRNL+ T5Exo–, and AtRNL+ T5Exo+ products, as follows: Tpt1+ 10X Tpt1 Reaction Buffer: 4 uL 50 mM NAD+: 8 uL DNA: 20 uL 37 uM Tpt1: 1.1 uL H_2O: 6.9 uL Total: 40 uL Final reaction concentration of Tpt1 is 1 uM. DNA indicates either AtRNL– T5Exo–, AtRNL– T5Exo+, AtRNL+ T5Exo–, or AtRNL+ T5Exo+ product. 2. Incubate all reactions at 30 °C for 1 hr. 3. Purify each reaction using RNAClean XP with elution volume of 30 uL. **I. PCR amplification and library verification** 1. Set up 5 PCR reactions, one without any template \(Primers-only) and four with each of AtRNL– T5Exo– Tpt1+, AtRNL– T5Exo+ Tpt1+, AtRNL+ T5Exo– Tpt1+, or AtRNL+ T5Exo+ Tpt1+ products, as follows: Primers-only 5X HF Buffer: 10 uL 10 mM dNTPs: 2 uL 100% DMSO: 0.5 uL 40 uM PCR.1.Index: 1 uL 40 uM PCR.2: 1 uL 2U/uL Phusion DNA Polymerase: 0.5 uL H_2O: 35 uL Total: 50 uL Template+ 5X HF Buffer: 10 uL 10 mM dNTPs: 2 uL 100% DMSO: 0.5 uL 40 uM PCR.1.Index: 1 uL 40 uM PCR.2: 1 uL DNA: 20 uL 2U/uL Phusion DNA Polymerase: 0.5 uL H_2O: 15 uL Total: 50 uL DNA indicates either AtRNL– T5Exo– Tpt1+, AtRNL– T5Exo+ Tpt1+, AtRNL+ T5Exo– Tpt1+, or AtRNL+ T5Exo+ Tpt1+ product. 5–30 uL of each product could be used as template for PCR. Koh et al.1 used 20 uL. For products which were not treated with T5 Exonuclease \(AtRNL– T5Exo– Tpt1+ and AtRNL+ T5Exo– Tpt1+), 5 uL is sufficient to visualize non-specific amplification. 2. Run PCR with the following settings: 98 °C for 30 s 98 °C for 10 s, 65 °C for 20 s, and 72 °C for 30 s \(repeated for 30 cycles) 72 °C for 5 m 4 °C PCR could be run for 26–32 cycles. Koh et al. used 30 cycles. 3. Run 6% Non-denaturing PAGE with 10 uL aliquot of each sample. Koh et al. used 100 bp DNA Ladder \(NEB) as the ladder. 4. Stain the gel in 1X SYBR Gold \(Life Technologies) for 30–40 m. 5. Visualize under UV light. An exemplary gel image is shown in Figure 2. AtRNL+ T5Exo+ Tpt1+ sample will be your ribose-seq library while Primers-only and AtRNL– T5Exo+ Tpt1+ samples will be your controls where no amplification should be observed \(only primer dimers). 6. Purify PCR mixtures from Primers-only, AtRNL– T5Exo+ Tpt1+, and AtRNL+ T5Exo+ Tpt1+ using RNAClean XP with elution volume of 15 uL. Controls Primers-only and AtRNL– T5Exo+ Tpt1+ are also purified so that the amount of actual ribose-seq library can be determined and quantitatively confirmed. 7. Use Qubit 2.0 \(dsDNA HS) to quantify the amount of ribose-seq library. Confirm that the amount of purified Primers-only product is similar to the amount of AtRNL– T5Exo+ Tpt1+. The amount of the actual ribose-seq library can be calculated by subtracting the amount of AtRNL– T5Exo+ Tpt1+ \

(which should be just primer dimers) from AtRNL+ T5Exo+ Tpt1+. Typically, ~25 nM of the ribose-seq library is resulted.

Timing

This protocol takes ~3–4 working days.

Troubleshooting

Depending on the genotype of the cell, the cell type, and the organism, more genomic DNA and more reactions leading to more template for PCR may be necessary. The amount of template used and the number of cycles run during PCR can also vary. If using another DNA polymerase for PCR, different reaction conditions such as concentrations of DMSO and primers and annealing temperature need to be tested, minimizing the cycle number while minimizing the background amplification for AtRNL– T5Exo+ Tpt1+.

Anticipated Results

Application of ribose-seq to yeast genome was recently reported by Koh et al.1 Following this protocol, ribose-seq libraries from genomic DNA of various S. cerevisiae cells were constructed. The libraries were then sequenced on Illumina MiSeq, collecting 50-cycle single-end reads. Analysis of the sequencing data led to profiling of ribonucleotide incorporation in yeast genomic DNA.

References

1 Koh, K. D., Balachander, S., Hesselberth, J. R. & Storici, F. Ribose-seq: global mapping of ribonucleotides embedded in genomic DNA. Nat. Methods 12, 251–257 \(2015). 2 Williams, J. S. & Kunkel, T. A. Ribonucleotides in DNA: origins, repair and consequences. DNA Repair \(Amst.) 19, 27–37 \(2014). 3 Potenski, C. J. & Klein, H. L. How the misincorporation of ribonucleotides into genomic DNA can be both harmful and helpful to cells. Nucleic Acids Res. 42, 10226–10234 \(2014). 4 Lujan, S. A., Williams, J. S., Clausen, A. R., Clark, A. B. & Kunkel, T. A. Ribonucleotides are signals for mismatch repair of leading-strand replication errors. Mol. Cell 50, 437–443 \(2013). 5 Reijns, M. A. et al. Lagging-strand replication shapes the mutational landscape of the genome. Nature 518, 502–506 \(2015). 6 Reijns, M. A. et al. Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. Cell 149, 1008–1022 \(2012). 7 Schutz, K., Hesselberth, J. R. & Fields, S. Capture and sequence analysis of RNAs with terminal 2',3'-cyclic phosphates. RNA 16, 621–631 \(2010). 8 Remus, B. S. & Shuman, S. Distinctive kinetics and substrate specificities of plant and fungal tRNA ligases. RNA 20, 462–473 \(2014).

Acknowledgements

We thank S. Garrey for AtRNL and Tpt1 protein purification, F. Stewart for technical support, and all members of the Storici laboratory for experimental suggestions. This work was supported by US National Science Foundation award number MCB-1021763 \(to F.S.), Georgia Research Alliance award number R9028 \(to F.S.), an American Cancer Society Research Scholar Grant \(to J.R.H.), a Damon Runyon-Rachleff Innovation Award from the Damon Runyon Cancer Research Foundation \(to J.R.H.) and the University of Colorado Golfers Against Cancer \(to J.R.H.).

Figures

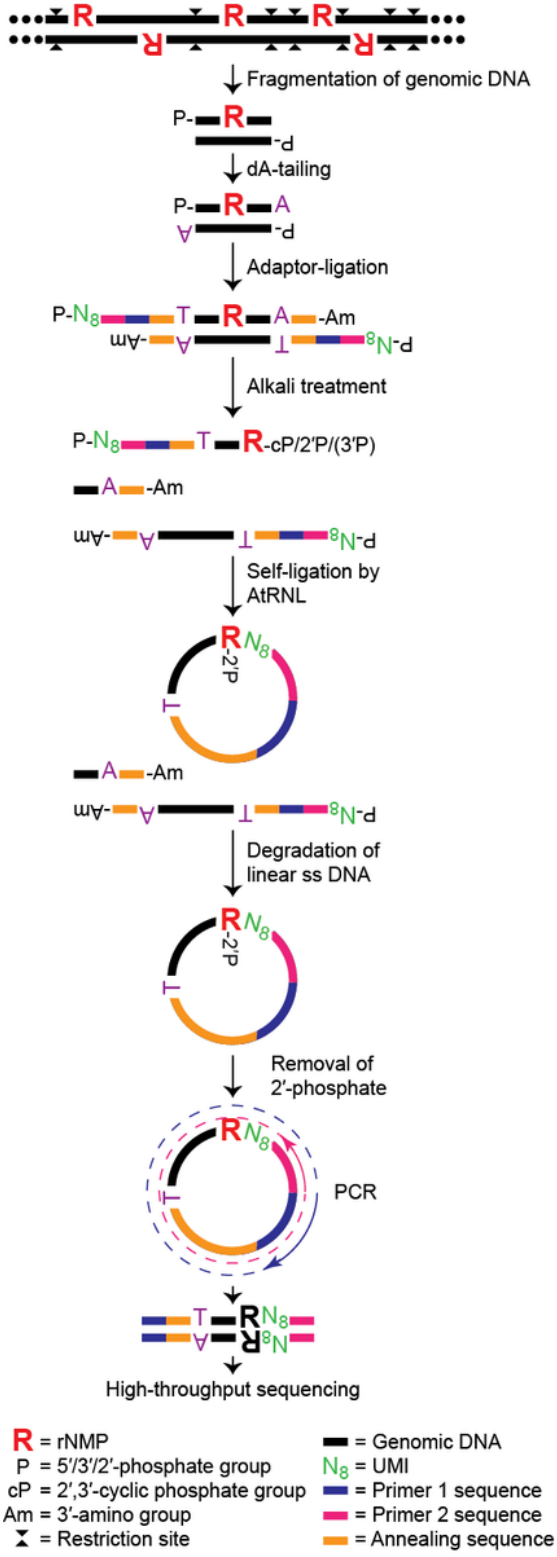

Figure 1

Scheme of ribose-seq ‘R’ in red denotes an rNMP while ‘P’ indicates a monophosphate group.

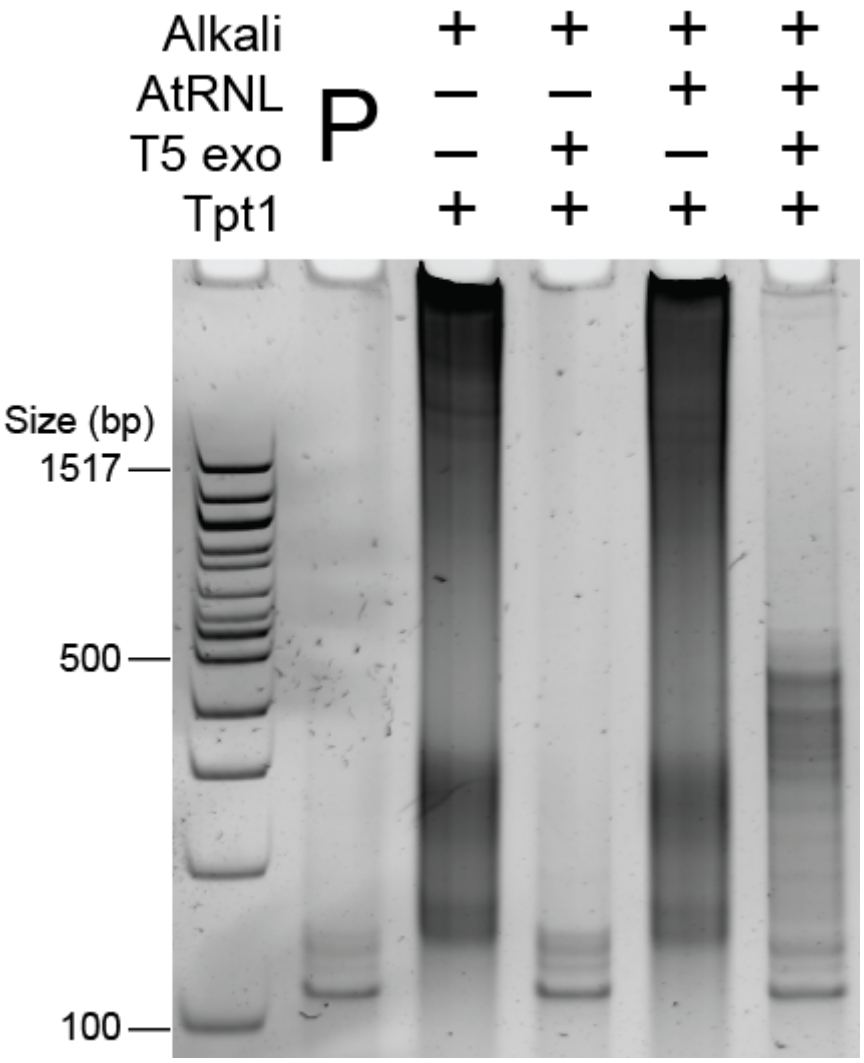

Figure 2

Ribose-seq library from genomic DNA of RNase H2-deficient S. cerevisiae cells Appropriate PCR products were analyzed by PAGE. ‘P’ indicates Primers-only.