Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

69451 Weinheim, Germany

Profiling of Ribose Methylations in RNA by High-Throughput Sequencing**

*Ulf Birkedal, Mikkel Christensen-Dalsgaard, Nicolai Krogh, Radhakrishnan Sabarinathan, Jan Gorodkin, and Henrik Nielsen**

anie_201408362_sm_miscellaneous_information.pdf

Supporting information

Table of contents:

Materials and Methods

Yeast strains and growth conditions

The strain used to represent wt *Saccharomyces cerevisiae* is BY4741[1] grown in YPD broth (yeast extract, peptone, dextrose; Sigma-Aldrich) with 2% glucose (Life Technologies). The strain referred to as "knock-out" is YWD486 that carries a plasmid harboring the entire snR67-snR53 region with a single substitution that blocks methylation at LSU-G2619.[2] YWD486 was grown in yeast nitrogen base (Sigma-Aldrich) supplemented with yeast synthetic dropout medium without leucine and tryptophan (Sigma-Aldrich) and with 2% glucose. The strain referred to as "knock-in" is YWD680 that carries a plasmid encoded artificial snoRNA targeting LSU-U2954.[3] YWD680 was grown in yeast nitrogen base supplemented with yeast synthetic dropout medium without tryptophan and with 2% galactose (Sigma-Aldrich). All strains were grown at 30°C with shaking at 180 rpm till they reached an OD_{600} = 0.7.

Isolation of whole cell RNA and purification of ribosomal RNA

Whole cell RNA was extracted using standard hot phenol method. In brief, 40 mL of cells, grown to OD_{600} = 0.7, were pelleted and resuspended in 500 µL TES buffer (Tris-HCl 10 mM pH 7.5, EDTA 10 mM; pH 8, 0.5% SDS) and 450 µL acidic phenol (pH 4.5). Samples were incubated at 65°C with maximum shaking for 1 h followed by phase separation by centrifugation. The aqueous phases were re-extracted once with 400 µL acidic phenol and twice with 400 µL Phenol:Chloroform:Isoamyl Alcohol (25:24:1, 10 mM Tris-HCl pH 8.0, 1 mM EDTA). RNA was precipitated in 0.3 M NaOAc pH 5.3 and 3 vols. of 96% ethanol and resuspended in appropriate amounts of DEPC treated H_2O. The large and small ribosomal subunits were purified from denaturing (formaldehyde) 1% agarose gels. Bands containing the appropriate subunit rRNAs were electrophoresed into dialysis chips (TaKaRa RECOCHIP) recovered by centrifugation and precipitated in 0.3 M NaOAc pH 5.3 and 3 vols. of 96% ethanol and resuspended in DEPC treated H_2O.

Isolation of chromatin-associated RNA

Chromatin was extracted according to.[4] The chromatin pellet was resuspended in 1X DNase I buffer and incubated with 5 U DNase I (Thermo Scientific) for 30 min at 37°C followed by addition of 2 µL of Proteinase K (20 mg/mL; Thermo Scientific) and further incubation for 1 hr at 37°C. The RNA was then extracted twice with acid phenol (pH 4.5) and twice with chloroform followed by ethanol precipitation. The integrity and composition of the RNA was assessed on an Agilent Bionanalyzer. The enrichment of ribosomal RNA precursors was assessed by qRT-PCR with three amplicons: Amplicon (A) spanning the 5'ETS and 18S, (B) located in 18S and (C) spanning 18S and ITS1. First-strand cDNA synthesis was made on 200 ng whole cell RNA and 50 ng chromatin-associated RNA using Superscript III Reverse-transcriptase (Life Technologies) with 200 ng random hexamer primers (dN6). 0.1% of the cDNA was used for qRT-PCR analysis on a LightCycler Nano System (Roche) using FastStart Essential DNA Green Master reaction mixture according to the manufacturer.

Oligos used for qRT-PCR analysis of chromatin associated RNA:

Amplicon	Oligo name	Sequence
A	C998	5'-cttctagcaagagggaataggtg
	C999	5'-atgagccattcgcagtttc
B	C1000	5'-tcttgtgaaactccgtcgtg
	C1001	5'-aattctccgctctgagatgg
C	C1002	5'-cgattgaatggcttagtgagg
	C1003	5'-ctctcatgctcttgccaaaac

Alkaline degradation and isolation of library fragments

RNA was resuspended in water at 1 µg/µL, mixed with an equal volume of 100 mM Na_2CO_3 (pH 9.0; adjusted with HCl) and distributed into several tubes that were incubated at 90°C in a PCR cycler for various times ranging from 1 to 10 min. After incubation, glycogen was added to a final concentration of 0.1 µg/µL and the samples transferred to tubes containing 3 vols. of 96% ethanol/ 2% KAc for precipitation of the RNA. After resuspension in water, the RNA was fractionated on a 10% denaturing polyacrylamide, the gel stained with Sybr®Gold (Invitrogen) and fragments in the desired size range (20-40 nt) excised and eluted o/n in 400 µL TEN buffer (1 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 10 mM NaCl)/ 400 µL Phenol:Chloroform:Isoamyl Alcohol (25:24:1, 10 mM Tris-HCl pH 8.0, 1 mM EDTA) at 4°C. A buffer at near neutral pH was used during elution to avoid opening of the cyclic phosphates. The aqueous phase was recovered and the RNA precipitated with 3 vols. of 96% ethanol after addition of 15 µg of glycogen as a carrier.

Construction of a cloning enzyme based on A. thaliana tRNA ligase

We used *Arabidopsis thaliana* tRNA ligase (AtRNL) to clone the library fragments. The enzyme catalyses a sequence of reactions during tRNA intron splicing consisting of opening of the 2', 3' cyclic phosphate on the 5' substrate RNA to form a 2'-phosphate, phosphorylation of 5' OH and subsequent adenylation at this position on the 3' substrate, and finally ligation of the two substrate RNAs.[5] The kinase activity is unwarranted for cloning purposes because it will phosphorylate the 5' ends of library fragments and result in concatemerization and circularization. To mutate this activity, we identified critical residues by alignment of kinase domains of tRNA ligases and other kinases (Figure S1). K700 and S701 in AtRNL were found to be highly conserved and were targeted by site-directed mutagenesis according to Zheng et al. [6] using plasmid pET28-AtRNL [7] as template and mutagenic primers (mutation underlined) C813: 5'TCTGCTGCATCTGCACTTTGCAAGGAGTTATT GAAC and reverse C814: 5'AAAGTGCAGATGCAGCAGATCCAGGAATTCCTG (lysine to alanine at K700A), C815: 5'TCTGCTAAAGCAGCACTTTGCAAGGAGTTATTGAAC and C816: 5'AAAGTGCTGCTTTAGCAGATCCAGGAATTCCTG (serine to alanine at S701A), and C817: 5'TCTGCTGCAGCAGCACTTTGCAAGGAGTTATTGAAC and C818: 5'AAAGTGCTGCTGCAGCAGATCCAGGAATTCCTG (lysine and serine to alanines at positions K700A and S701A). The ligase and kinase activities of the mutant enzymes were analysed (see below). S701A was selected for

further use and the corresponding plasmid named AtRNLΔPNK. After completion of our study, we learned that Remus and Shuman[4] had produced similar mutants in order to study enzymatic properties of AtRNL.

Construction of an RNA 5' adaptor with a 2',3' cyclic phosphate

The cloning protocol employs a 5' adaptor consisting of an RNA oligo with a 5' triphosphate and a 2', 3' cyclic phosphate (P1 IVT; 5'-GGCCACUACGCCUCCGCUUUCCUCUCUAUGGGCAGUCGGTGAU>P). This oligo was produced from an in vitro transcription reaction of a DNA template containing the T7 promoter, the adaptor sequence, and the HDV ribozyme sequence (Figure S2). The template was made in two subsequent PCR reactions using Phusion DNA polymerase (Thermo Scientific) in a two-step reaction first using the primers C894: 5'-CCTCCGCTTTCCTCTCTATGGGCAGTCGGTGATGCCGGCCATGGTCCCAGCCTC and C887: 5'-TGTCCCATTCG CCATTACCGAGGGGACGGTC, and a plasmid containing the HDV ribozyme sequence (gift from Benoît Masquida, University of Strasbourg). The PCR product was purified on a 2% agarose gel using GeneJET PCR Purification Kit (Thermo Scientific) and used as template for the second PCR using C895: 5'-TAATACGACTCACTATAGGCCA CTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTG and C887. This resulted in a template for in vitro transcription with the sequence: 5'-TAATACGACTCACTATAGGCCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGATGC CGGCCATGGTCCCAGCCTCCTCGCTGGCGGCCGGTGGGCAACATTCCGAGGGGACCGTCCCCTCGGTAATGGCGAATGG GACA. The PCR template was in vitro transcribed with the MEGAshortscript T7 Transcription Kit (Life Technologies). The 5' adaptor oligo is released during standard in vitro transcription (Figure S2) and was subsequently gel purified. The AtRNLΔPNK enzymes and the P1IVT oligo were then validated in a model experiment (Figure S3).

Library construction

The library fragments have 5' OH and 2', 3' cyclic phosphate ends. Thus, a new protocol was developed based on tRNA ligase to specifically clone these fragments inspired by Schutz et al.[7] (Figure S4). The 3' adaptor was a standard RNA oligo (CR30: 5'-CGCACUGAGUCGGAG or CR35: 5'-CGAGCUGAGUCGGAG) with a sequence corresponding to the Ion Torrent A sequence including a 4 nt barcode (#1: CGCA or #2 CGAG). The oligo was phosphorylated in a standard 40 µL T4 PNK reaction containing 2 nmol of oligo. The 3' adaptor was ligated to the library fragments in a 10 µL reaction composed of 1 µg of library fragments (20-40 nt), 100 pmol of phosphorylated CR30, ligase buffer and additives (final: 10 mM Tris-HCl pH 7.5, 100 mM KOAc, 0.3 mM spermine, 6 mM $Mg(OAc)_2$, 0.5 mM DTT, 0.5% Triton X-100, 14% PEG 8000, 1 mM ATP), 20 U of RNase inhibitor (RiboLock; Thermo Scientific), and AtRNLΔPNK. The reaction was incubated for 1 h at 30°C after which the RNA was PCI-extracted and ethanol precipitated. Next, the ligation product was phosphorylated and the 2', 3' cyclic phosphates of unligated library fragments removed in a single standard T4 PNK reaction relying on enzymes' kinase and 3' phosphatase activities. Finally, the ligation products were gel purified on a 10% denaturing (urea) PAG. As 5' adaptor we used an in vitro transcribed and ribozyme cleaved RNA oligo carrying a 2', 3' cyclic phosphate (P1 IVT; 5'-GGCCACUACGCCUCCGCUUUCCUCUCUAUGGGCAGUCGGUGAU>P) RNA oligo. The conditions for 5' adaptor ligation were similar to those described for 3' adaptor ligation and the ligation products were subsequently gel purified on a 10% denaturing (urea) PAG. Prior to cDNA synthesis, the 2'-O phosphates left by the tRNA ligation step were removed in a standard alkaline phosphatase reaction (FastAP;

Thermo Scientific). The cDNA synthesis was made using a commercial cDNA synthesis kit (First strand cDNA synthesis; Invitrogen) and a primer corresponding to the short version of the Ion Torrent adaptor A with a 4 nt barcode (#1; CGCA) complementary to the 3' adaptor CR30. The library was assessed by qPCR analysis following the guidelines outlined in the kit that was used for sequencing (Ion Library Quantitation Kit User Guide).

Sequencing

Sequencing was carried out as Ion semiconductor sequencing using Ion Torrent or Proton instruments and dedicated sequencing kits (Life Technologies). After determining the dilution factor, the libraries were used as templates for emulsion PCR (60 cycles). In our experience the recommended library input concentration is too high, and we diluted the libraries up to 5 times more. The sequencings were made running 250 or 500 sequencing flows. The base calling from the electrograms was performed on the Ion server and the data files downloaded in fastq format without further processing.

Reads filtering and mapping

First, reads without barcode sequence were discarded. Next, Cutadapt v1.2[8] was used to remove the 3' adaptor and discard reads shorter than 15 nucleotides. The trimmed reads were mapped using the Bowtie 2 aligner[9] with default settings to the precursor ribosomal RNA (35S rRNA) gene of *Saccharomyces cerevisiae* s288c (RDN37-1; SGD:S000006486, GeneID: 9164931)). Only reads mapping uniquely and with a minimum of three correct nucleotides at the 5' or 3' end were considered for the following data handling. These reads were used to compute the number of 3' read-ends (corresponding to the 5' nt in the sequence insert) and 5' read-ends (corresponding to the 3' nt in the sequence insert), respectively. In the latter, the counts were shifted one position upstream at the level of the sequence read in order to get the 3' and 5' read-ends to align to the methylated nucleotide in the reference sequence. Finally, the two datasets were combined for calculation of the RiboMeth-Seq score. The full dataset is available at:
www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE54655

Calculation of the RiboMeth-Seq score

Three different scores were calculated to emphasize various aspects of the analysis. Score A is used for detection of ribose methylated positions and is based on both the average and standard deviation of the neighboring positions. It scales between 0 and 1 with relatively low signal to noise ratio. A Matthews's correlation coefficient[10] is used to determine the optimal cut-off value. Score B is based on the weighted average of neighboring positions. It has a very high signal to noise ratio and is used for inspection of the data in genome browser format. This score is very sensitive to small differences at high-level of methylation and is not suitable for comparisons of these positions. Score C is a normalized version of score B and expresses the percent methylation at a given position using the flanking positions to estimate the value corresponding to 100% of molecules methylated at the position. In contrast to the other scores, score C is only calculated for the selected positions.

Score A:

$$S_i = \max\begin{cases} 1 - \dfrac{2n_i + 1}{½\left|\mu_l - \sigma_l\right| + n_i + ½\left|\mu_r - \sigma_r\right| + 1} \\ 0 \end{cases}$$

where n_i is the read count at position i, μ_l and μ_r are the mean values computed from the left and right flanking regions of i, respectively. Similarly, σ_l and σ_r are the corresponding standard deviations and δ is the length of the flanking region (default = 6). For example, the mean and standard deviation of the left flanking region is calculated by

$$\mu_1 = \frac{1}{\delta}\sum_{j=i-\delta}^{i-1} n_j \ , \ \sigma_1 = \sqrt{\frac{\sum_{j=i-\delta}^{i-1}\left(\mu_l - n_j\right)^2}{\delta - 1}}$$

Score B:

$$S_i = \frac{\left| n_i - ½\left(\dfrac{\sum_{j=i-\delta}^{i-1}\omega_j n_j}{\sum_{j=i-\delta}^{i-1}\omega_j} + \dfrac{\sum_{j=i+1}^{i+\delta}\omega_j n_j}{\sum_{j=i+1}^{i+\delta}\omega_j} \right) \right|}{n_i + 1}$$

where n_i = read count at position i and δ = 6 (flanking window length). The weight ω various with respect to distance:

j	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
ω	0.5	0.6	0.7	0.8	0.9	1	0	1	0.9	0.8	0.7	0.6	0.5

Score C:

$$S_i = \max \begin{cases} 1 - \dfrac{n_i}{½\left(\dfrac{\sum_{j=i-\delta}^{i-1} \omega_j n_j}{\sum_{j=i-\delta}^{i-1} \omega_j} + \dfrac{\sum_{j=i+1}^{i+\delta} \omega_j n_j}{\sum_{j=i-1}^{i-\delta} \omega_j} \right)} \\ 0 \end{cases}$$

Matthew's correlation coefficient:

For scores A and B, we calculate the Matthews correlation coefficient (MCC) for all scores ranging between the minimum and maximum score value by considering each score value as threshold and compute the respective True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) to compute MCC by:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

The score value corresponding to the higher MCC value was chosen as an optimal threshold. The MCC close to a value of +1 indicates a better classification between the observed and predicted candidates. The analysis was carried out using R package ROCR. [11]

Primer extension analysis of 2'-O-methylated nucleotides

Two different biochemical methods based on primer extension were used to determine 2'-O-Me at specific sites. The first method relies on pausing of RT on intact template RNA one nucleotide prior to the methylated site during reverse transcription at limited dNTP concentrations (1 mM, 0.1 mM, 0.04 mM, 0.01 mM, or 0.004 mM). The second method uses partially alkaline degraded RNA and scores methylated positions as absence of a run-off signal at the methylated position. For partial alkaline hydrolysis, yeast whole cell RNA was adjusted to 0.5 µg/µl in 50 mM Na_2CO_3 (pH 9.0). The hydrolysis was carried out by incubation for 6, 8 or 10 min at 90°C, and the partially degraded RNA was recovered by ethanol precipitation. Primer extension for both methods was carried out at 45°C for 30 min on 5 µg intact or partially degraded whole cell RNA in 10 µl reactions containing 1X AMV RT buffer (50mM Tris-HCl pH 8.0, 25 mM KCl, 25 mM $MgCl_2$, 5 mM DTT and 0.05 mg/mL gelatine) with 1 pmol of the appropriate 5' end-labeled primer and AMV reverse transcriptase (Promega). The extension products were analyzed on 8% denaturing (50% urea) polyacrylamide gels next to the appropriate Sanger sequencing reactions of PCR products.

qRT-PCR assay for quantitation of ribose methylation

The degree of methylation of a selection of sites was analyzed by qRT-PCR. Four methylation sites were selected for the analysis, two from the SSU and two from the LSU. 100 ng of whole cell yeast RNA (DNaseI treated) was reverse transcribed into cDNA, in reactions containing AMV-Reverse transcriptase (Promega), 1X AMV RT buffer (50mM Tris-HCl pH 8.0, 25 mM KCl, 25 mM $MgCl_2$, 5 mM DTT and 0.05 mg/mL gelatine), 1 µM of each reverse primer targeting a sequence upstream to a specific methylation site, and either 10 µM or 1 mM

dNTP. Reverse transcription was performed at 42°C for 60 min. 0.1% of the cDNA was used for qRT-PCR analysis on a LightCycler Nano System (Roche) using FastStart Essential DNA Green Master reaction mixture according to the manufacturer.

Oligos used for qRT-PCR:

Amplicon	Oligo name	Sequence
SSU-A100	C1076	5'- CACAGTTATACCATGTAGTAAAGG (reverse primer)
	C1077	5'- GCAATTTATACAGTGAAACTGC
SSU-1428	C1078	5'-AAGGTTAGACTCGCTGG (reverse primer)
	C1079	5'-GTTTCAAGCCGATGGAAG
LSU-A220	C1082	5'-AGAGTCATAGTTACTCCCG (reverse primer)
	C1083	5'-TGACGCAATGTGATTTCTG
LSU-U2347	C1084	5'-GCTCAACAGGGTCTTC (reverse primer)
	C1085	5'-GTCCCTATCTACTATCTAGCG

rRNA fragment isolation and mass spectrometry

For mass spectrometry analysis a defined rRNA sequence was isolated according to [12] by incubating 100 pmoles of total RNA with 1000 pmoles of a DNA oligo complementary to the rRNA region of interest in 100 µl of 60 mM HEPES (pH 7.5) and 125 mM KCL. The mixture was incubated for 5 min at 90°C and then cooled slowly during 3 h to 45°C. Mung bean nuclease buffer was added to a final concentration of 30 mM NaCl, 50 mM NaOAc, 1 mM $ZnSO_4$ (pH 5 at 25°C). The DNA:rRNA hybrid was then digested with 30 U of Mung Bean nuclease and 0.5 µg of RNase A for 50 min at 35°C. The hybrid was PCI-extracted and ethanol precipitated followed by separation on a 10 % denaturing (urea) PAG. The rRNA fragment was excised and eluted overnight in 400 µl of NH_4OAc (pH 5.3) at 4°C followed by ethanol precipitation with 2.5 vol. of 96 % ethanol. The purified rRNA fragment was split in two aliquots and treated with either 0.1 µg of RNase A or 20 U of RNase T1 in 2 µl of 50 mM 3-hydroxypicolinic acid (3-HPA) for 4 h at 37°C. The 2'-3' cyclic phosphate produced by RNase T1 was hydrolyzed by adding HCl to a final concentration of 0.2 M and incubated for 30 min at room temperature followed by the addition of 20 µl of H_2O, the sample was vacuum dried and then re-dissolved in 2 µl H_2O. MALDI-TOF mass spectrometry analysis was performed on an Autoflex Speed (Bruker Daltonics, Bremen) mass spectrometer. In brief, 1-2 µl of the digested RNA and 0.7 µl 0.5 M 3-HPA was mixed on the target plate and dried at room temperature. Spectra were recorded in reflector and positive mode.

Oligos used for isolating rRNA fragments for mass spectrometry analysis:

rRNA target	Oligo name	Sequence

SSU:G562	C1046	5’-GCGGCTGCTGGCACCAGACTTGCCCTC
LSU:G234	C1047	5’-GGCACTTTACAAAGAACCGCACTCCTCGCCACACGG
LSU:U862	C1048	5’-CCGCTACGAGCCTCCACCAGAGTTTCCTCTGGC
LSU:G1142	C1049	5’-CCTTAACTCTACGTTCGGTTCATCCCGCATCGCCAG

Supporting figures

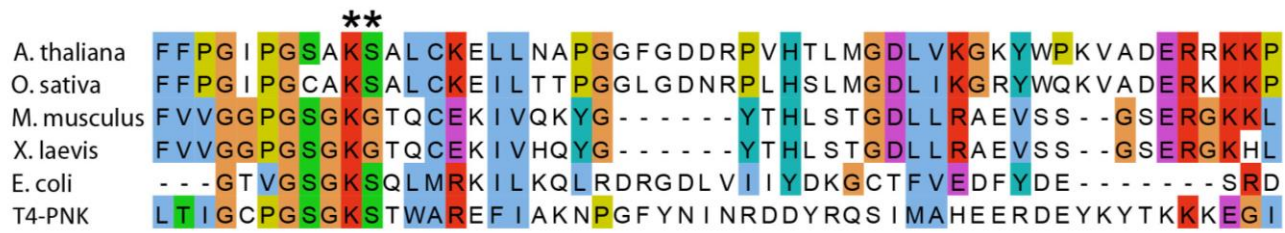


Figure S1. ClustalX alignment of *Arabidopsis thaliana* tRNL GI:145335250, *Oryza sativa* tRNL GI:4342624, *Mus musculus* adenylate kinase 1 GI:123858266, *Xenopus laevis adenylate* kinase 1 GI:147903781, *Escherichia coli* TraD GI:371606879 and Enterobacteria phage T4 PNK GI:9632811. The latter enzyme is well described and the lysine and serine residues highlighted with * are known to be critical for kinase activity and were the ones mutated to produce the kinase deficient AtRNLs.

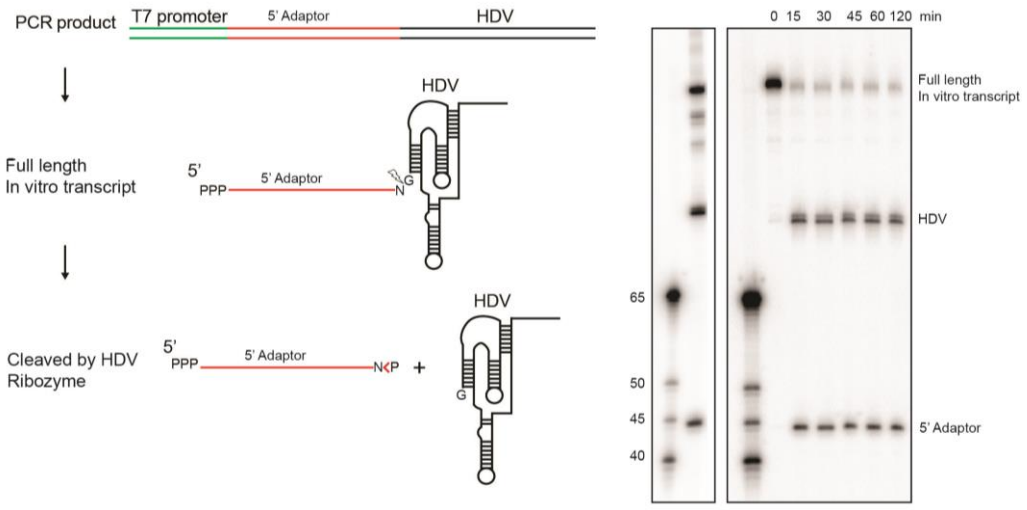


Figure S2. Construction of a 5' RNA adaptor with a cyclic phosphate end. To the left is shown an outline of the PCR construct that is in vitro transcribed into a precursor that undergo self-cleavage to produce a 5' product with a 2', 3' cyclic phosphate end. Note that the HDV ribozyme cleave immediately upstream of its own core sequence. To the right is shown a gel analysis of radiolabelled RNA. The second lane shows the co-transcriptional cleavage pattern. The precursor transcript was gel-purified and the remainder of the gel shows a time-course of cleavage of this transcript. The 5' fragment is 43 nt and was gel-purified and used as 5' adaptor in the RiboMeth-seq protocol.

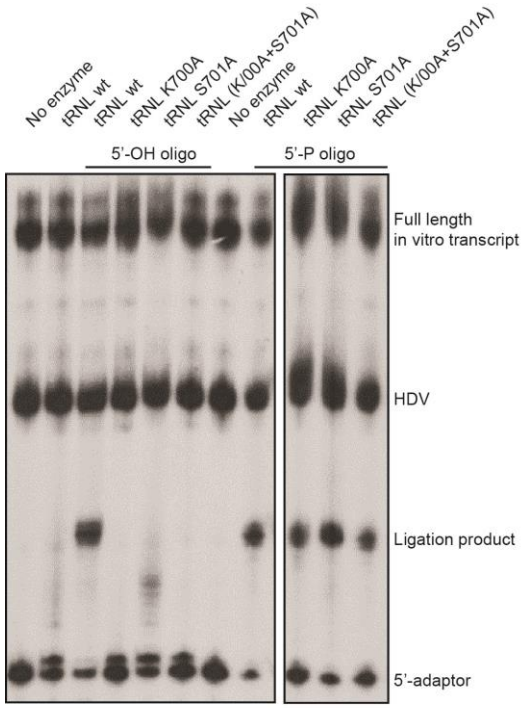


Figure S3. Assay of ligase and kinase activities of mutated tRNL enzymes. The three mutated forms of AtRNL were incubated with a radiolabelled in vitro transcription reaction containing the 5' RNA adaptor with a 2', 3' cyclic phosphate end and a cold test oligo with either a 5' OH or a 5' P. In reactions with 5' OH oligo, only the wt enzyme was able to ligate the oligo and the 5' adaptor. In contrast, both the wt and the three mutated enzymes were able to ligate the 5' adaptor and the 5' P oligo. This demonstrates that all mutated enzymes retain ligase activity, and that they most likely have lost their kinase. tRNLS701A was chosen for further applications.

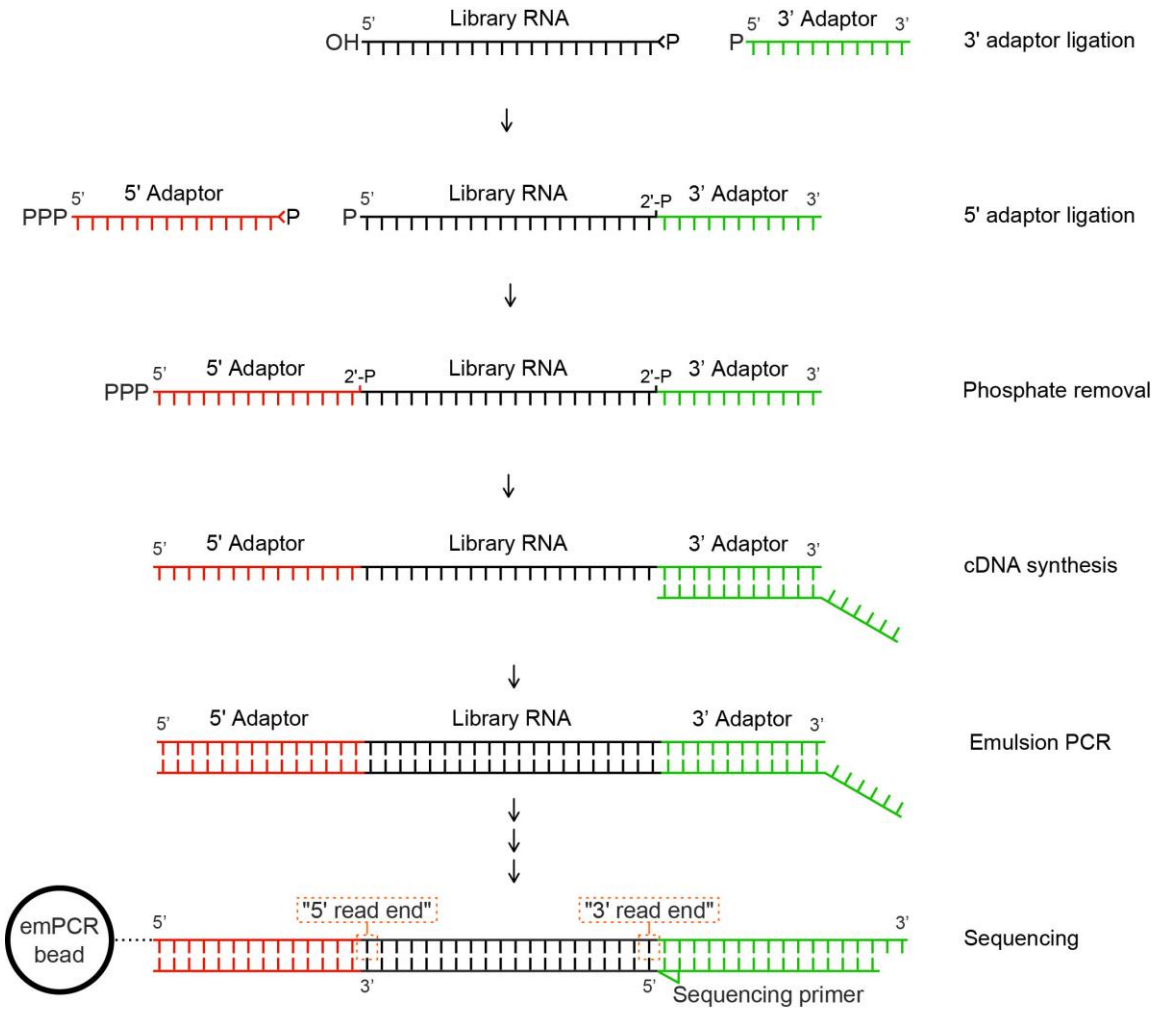


Figure S4. Library construction overview. The 3’ adaptor was an RNA oligo with a 5’ phosphate and was ligated to the partially alkaline degraded RNA containing 2’, 3’ cyclic phosphate ends. The product was gel-purified and ligated to the 5’ adaptor which also carried a 2’, 3’ cyclic phosphate end. This product was gel-purified as well, and the protruding 2’ phosphates resulting from the ligation reaction removed enzymatically. Finally, the RNA library was reverse transcribed and subjected to emulsion PCR and Ion semiconductor sequencing. The sequence of the library fragment was listed as the RNA-like sequence and the read-ends recorded as indicated in the figure.

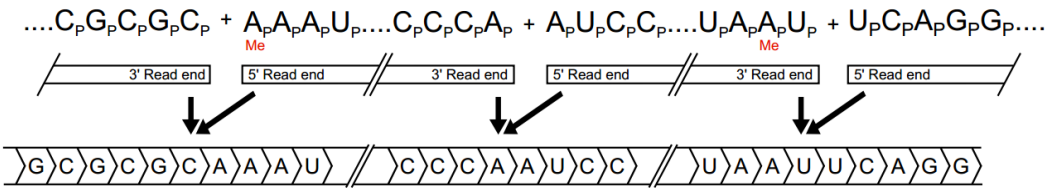


Figure S5. Read-end mapping. The read-ends derived from library sequencing (Figure S4) were mapped to the reference sequence according to the ribose that is being queried. Thus, 3' read-ends correspond directly to the nucleotide in the reference sequence, whereas 5' read-ends are shifted one nucleotide in the 5' directions such that they refer to the ribose 2'-O next to the phosphodiester bond that was cleaved. The three possible situations are depicted in the figure. The middle part shows cleavage in unmodified parts of the RNA. The left part shows the nearest possible cleavage 5' to a ribose methylation, and the right part the nearest possible cleavage 3' to a ribose methylation. At a methylated position, the nucleotide upstream of the resistant phosphodiester will not contribute to the 3' read-ends. Similarly, the nucleotide downstream of the resistant phosphodiester will not contribute to the 5' read-ends. Thus, a ribose methylation results in the absence of read-ends that are displaced by one nucleotide in the two datasets. By shifting the 5' read-ends one position upstream, the two datasets will align to the nucleotides that carry the ribose methylations.

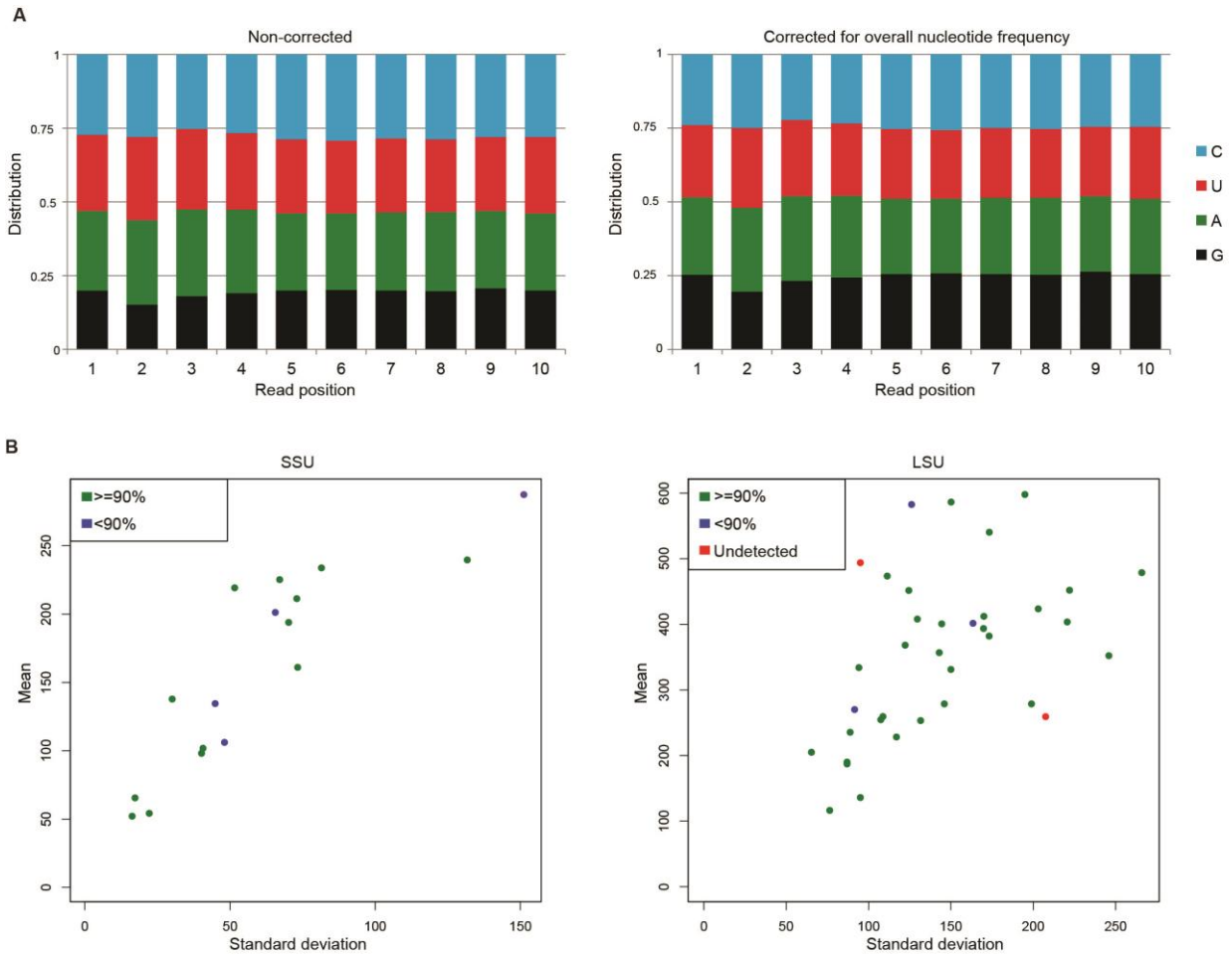
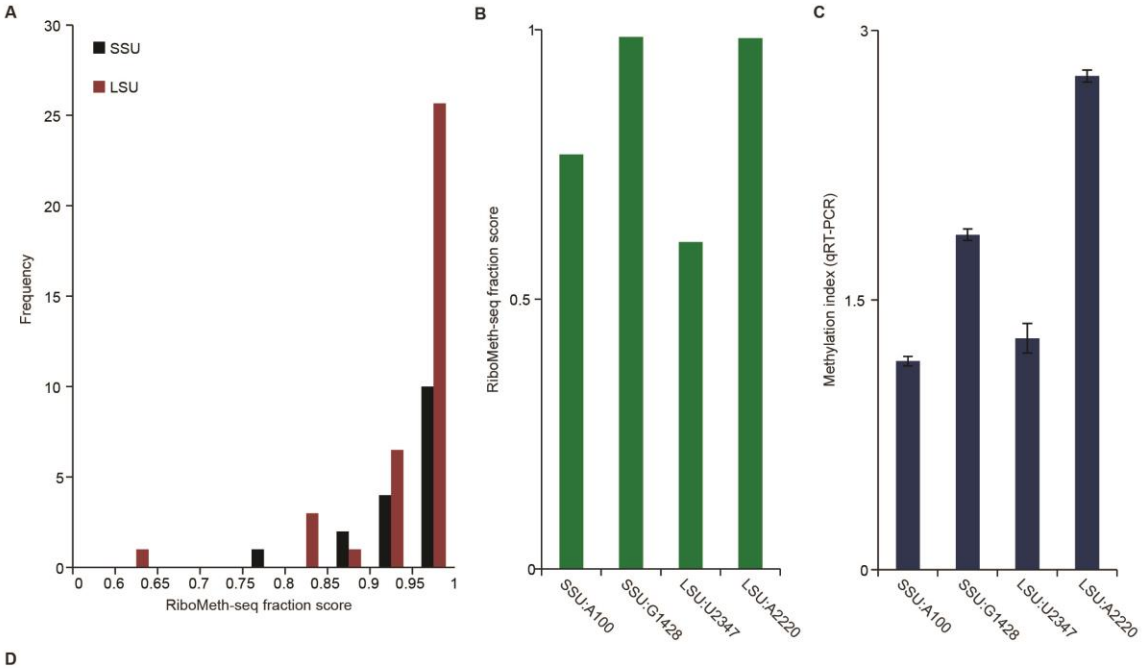


Figure S6. Analysis of bias in RiboMeth-seq. (A) The alkaline degradation step[13], the native structure of the rRNA, and the nucleotide frequencies of the first ten nucleotides in library fragments contribute little bias. Considerable bias in reads coverage is introduced in the ligation step, likely due to co-folding between the adaptor and the insert as has been observed in protocols using T4 RNA ligase. In the RiboMeth-seq protocol, combining data points from both ends of the inserts alleviates the problem. The map at the termini of the rRNA derives from reads from one end only due to the isolation of 20-40 nt fragments in the cloning procedure. Nevertheless, modifications close to the end, e.g. SSU-A28, are readily detected. (B) Scatter plot showing a comparison of standard deviation and mean values of counts of read-ends from the neighboring positions of true-positives (TP; black or blue) and false-negatives (FN; red). The low-scoring true-positives (score C < 0.9; in blue) and two false-negatives (in red) do not form outgroups suggesting that the low scores at these positions is not related to the quality of the data but rather reflect a low degree of methylation.



D

SSU	Fraction Score
A28	0.99
A100	0.77
C414	0.96
A420	0.90
A436	0.89
A541	0.99
G562*	0.92
U578	0.93
A619	0.92
A796	0.97
A974	0.96
C1007	0.98
G1126	0.93
U1269	0.92
G1271	0.98
G1428	0.99
G1572	0.97
C1639	0.98

LSU	Fraction Score
A649	0.98
C650	0.98
C663	0.91
G805	0.98
A807	0.97
A817	0.85
G867	0.95
A876	0.91
U898	0.85
G908	0.96
A1133	0.99
C1437	0.97
A1449	0.99
G1450	0.99
U1888	0.97
C2197	0.86
A2220	0.99
A2256	0.97
A2280	1.00

LSU	Fraction Score
A2281	1.00
G2288	0.96
C2337	0.99
U2347	0.61
U2417	0.99
U2421	0.93
G2619	0.98
A2640	0.96
U2724	0.98
U2729	0.94
G2791	0.98
G2793	0.98
G2815	0.99
U2921	0.99
G2922	0.98
A2946	0.99
C2948	0.82
C2959	0.95

Figure S7. Fractional modification. (A) The distribution of fraction scores in all SSU and LSU positions show that most positions are close to fully methylated in mature wt rRNA. Note that the two false-negatives in the overall analysis (Figure 1) are defined by calculation of a detection score (See Materials and Methods), which ranks the positions slightly differently than the fraction score used in this figure. (B) Histogram of the scores (score A) of positions selected for further analysis. (C) qRT-PCR analysis[14] of two residues with a high fractional RiboMeth-seq score (score C; SSU-G1428 (0.99), LSU-A2200 (0.99)) and two with a low score (SSU-A100 (0.77), LSU-U2347 (0.61). The high and low scores are significantly different (p-value < 0.0005, N =3 (unpaired t-test)). (D) Fraction scores (MethPercent) at all positions in wt yeast (BY4741). The fraction scores were calculated from the data sets on gel purified rRNA. (*) New modification identified in this study.

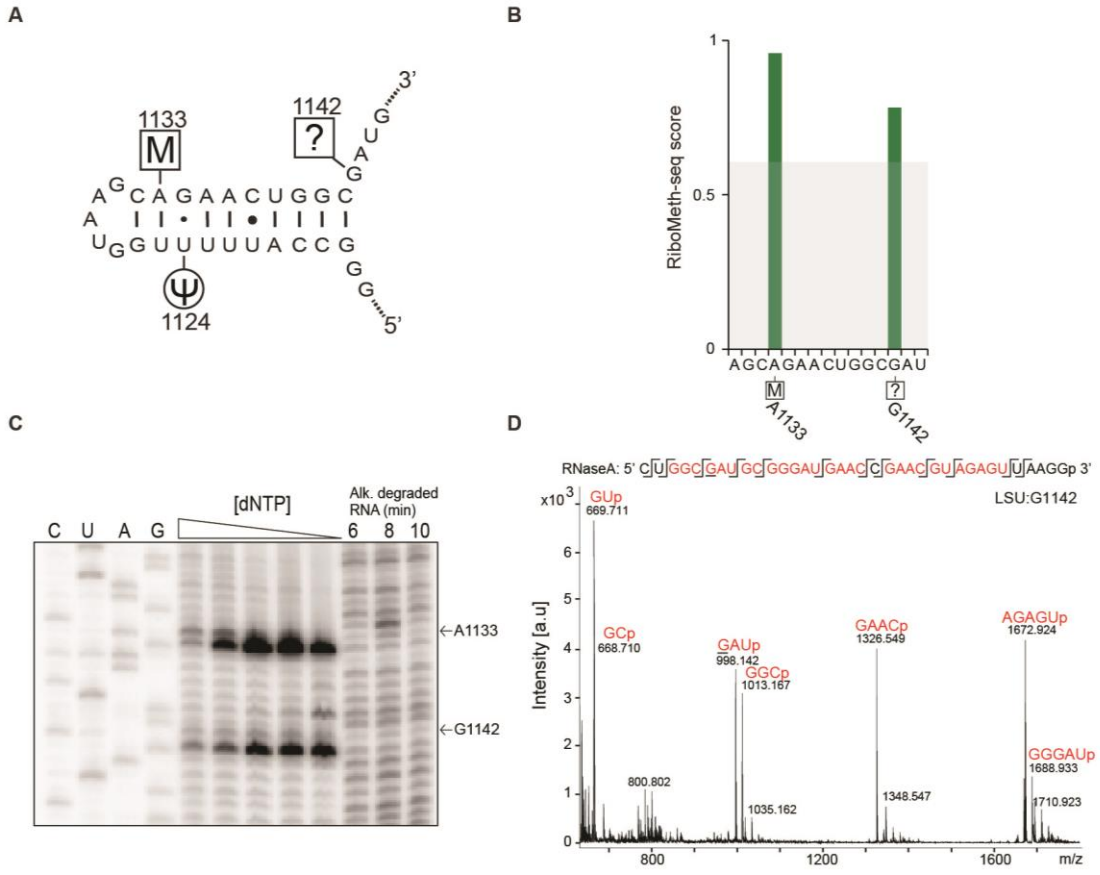


Figure S8. Analysis of false-positives. All four false-positive positions from the data sets were analyzed further. Ribose methylation at LSU-G234 and LSU-C1609 were not supported by two biochemical methods (not shown). (A) LSU-G1142 is located in a conserved part of rRNA close to other modifications. (B) The RiboMeth-seq signal is distinct and significantly above the cut-off value. (C) Analysis by primer extension resulted in a strong signal at limiting dNTP concentration and a pronounced gap in the alkaline ladder. However, these signals did not align as expected for ribose methylations. (D) Mass spectrometry did not reveal an additional mass suggesting that the modification is mass neutral or remains undetected for technical reasons. In light of the evidence provided by independent methods, LSU-G1142 remains a candidate for a new modification.

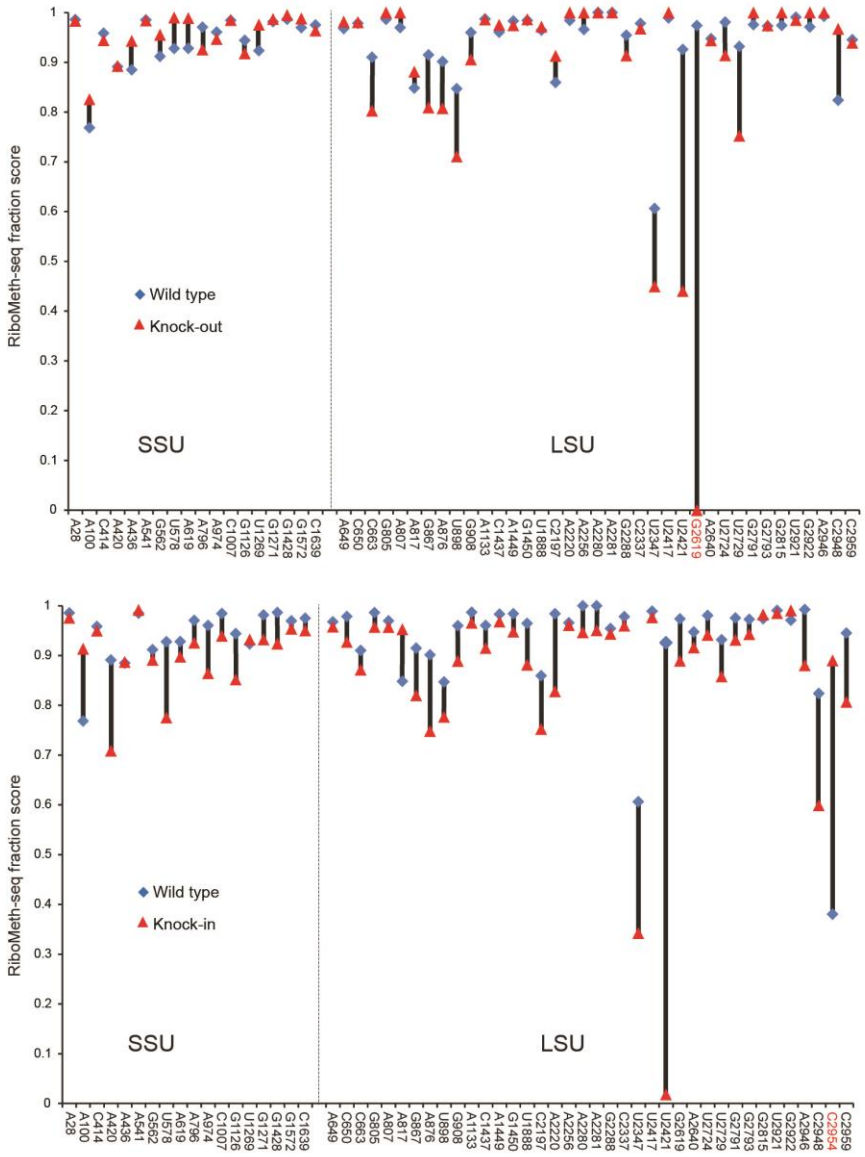
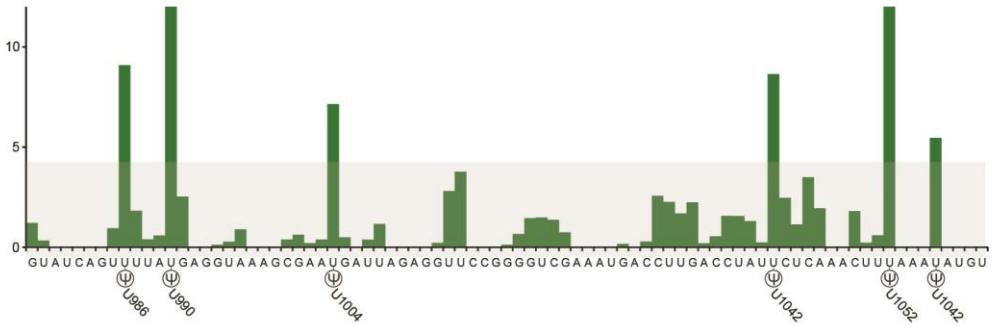


Figure S9. Global effects of knock-out and knock-in mutations. The fraction of molecules methylated in wt (blue) and in the knock-out (red; upper panel) and knock-in (red; lower panel) mutants were compared across all methylated positions. The knock-out of methylation at LSU-G2619 additionally affects a small number of positions in LSU, most notably LSU-U2421. The knock-in mutation at LSU-U2954 affects neighboring, as well as distant positions in both SSU and LSU. In general, it appears that late methylations (Figure 4) are most affected by the knock-out and knock-in mutations suggesting a processing defect in these strains. LSU-U2421 represents a special case showing a high degree of methylation (0.87) and no methylation, respectively, in the two datasets that were pooled for the figure. This observation suggests that slight changes in growth conditions in one of the two yeast cultures have enhanced the proposed processing defect.

A

	Position	Modification	Face	Total coverage	G%	A%	U%	C%
Mature	S1191U	$m^1acp^3\psi$	W	506	2.57*	0.99*	-	42.49*
	S1575G	m^7G	H	1064	-	0.09	0.28	0.19
	S1781A	m^6_2A	H	317	0.63	-	0.32	0.00
	S1782A	m^6_2A	H	305	0.00	-	0.00	0.00
	L645A	m^1A	W	3394	2.53*	-	2.98*	0.03
	L956U	m^5U	H	10726	0.00	0.00	-	0.01
	L2142A	m^1A	W	2608	2.11*	-	1.50*	0.08
	L2278C	m^5C	H	3930	0.00	0.00	0.05	-
	L2634U	m^3U	W	4512	14.18*	32.29*	-	0.31*
	L2843U	m^3U	W	2395	18.46*	31.15*	-	0.21*
	L2924U	m^5U	H	8272	0.00	0.00	-	0.01
Chromatin associated	S1191U	$m^1acp^3\psi$	W	38221	0.37*	0.40*	-	10.67*
	S1575G	m^7G	H	34266	-	0.05	0.24*	0.03
	S1781A	m^6_2A	H	64763	0.22*	-	0.04	0.01
	S1782A	m^6_2A	H	65402	0.04	-	0.26*	0.00
	L645A	m^1A	W	16507	2.13*	-	3.61*	0.13*
	L956U	m^5U	H	126388	0.00	0.00	-	0.02
	L2142A	m^1A	W	6020	1.41*	-	2.67*	0.07
	L2278C	m^5C	H	66474	0.01	0.00	0.04	-
	L2634U	m^3U	W	20406	6.24*	10.75*	-	0.28*
	L2843U	m^3U	W	31670	9.35*	16.87*	-	0.15*
	L2924U	m^5U	H	115390	0.00	0.00	-	0.03

B



MCC	Optimal threshold	True-positives	False-positives	False-negatives	True-negatives
0.5038	7.145	16	17	14	3312

Figure S10. Base modifications. (A) Misincorporation of nucleotides in cDNA template by base methylated positions in mature and chromatin-associated RNA was extracted from the sequence reads. Statistically significant misincorporation was found systematically at bases methylated at the Watson-Crick (W) face, but not at the Hoogsteen (H) face. When the misincorporation in mature and chromatin-associated RNA is compared, it is apparent that the $m^1acp^3\Psi$ modification at SSU-1191U and the m^3U modifications at LSU-2634U and LSU-2843U are introduced late in processing, as expected. P-value < 0.001 is indicated with * (B) Browser view showing pseudoU detection in the Helix 35 region of LSU. All six modifications are detected. The table shows the statistics for the entire LSU. The size of LSU RNA in yeast is 3396 nt. All 2'-O-Me modified residues were removed from the data set prior to the pseudoU analysis.

Supporting tables

Table S1

Library statistics	wt (SSU)	wt (LSU)	KO (WC)	KI (WC)	wt (Chromatin)	KI (WC1)	KI (WC2)	KI (LSU)
Total reads	357063	797976	417204	1340965	33590830	423173	917792	1198320
Reads with barcode	352775	788240	415511	1318917	32794029	420281	898636	1155223
Reads with barcode, mapped to RDN37-1	187395	655308	240967	700776	16793841	228996	471780	688492
Mapped reads with perfect alignment	158326	529650	215135	599722	14214599	213911	385811	464583
Mapped reads with errors but perfect 5' end match	22421	101154	18622	73247	1800176	11589	61658	176971
Total reads used for 5' read end analysis	180747	630804	233757	672969	16014775	225500	447469	641554
SSU	144384	-	108668	266558	-	78924	187634	-
LSU	-	614817	107127	370748	-	120649	250099	636263
Reads with barcode and adaptor, mapped to RDN37-1	183033	647332	239405	682565	16108374	228201	454364	651821
Mapped reads with perfect alignment	156109	526039	214277	589273	13804906	213331	375942	450266
Mapped reads with errors but perfect 3' end match	15791	90476	17387	53479	1379347	11107	42372	115427
Total reads used for 3' read end analysis	171900	616515	231664	642752	15184253	224438	418314	565693
SSU	135940	-	104942	243985	-	77541	166444	-
LSU	-	600751	106482	354177	-	120221	233956	560918
SSU total	280324	-	213610	510543	-	156465	354078	-
LSU total	-	1215568	213609	724925	-	240870	484055	1197181
Data points per residue	156	358	82	238	4549	77	161	353

	MCC	Optimal threshold	True-positives	False-positives	False-negatives	True-negatives
SSU rRNA						
Score A	0.971	0.482	17	1 (S562)	0	1756
Score B	0.971	3.214	17	1 (S562)	0	1756
LSU rRNA						
Score A	0.932	0.606	35	3 (L234, L1142, L1609)	2 (L2347, L2948)	3313
Score B	0.920	5.479	35	4 (L234, L862, L1142, L1609)	2 (L2347, L2948)	3314

Table S1. Library statistics. A complete data set for ribose methylations in rRNA can be obtained at less than 100 data points per residue from a sample of whole cell RNA. This makes RiboMeth-seq a low-cost technique when applied to low complexity transcriptomes (i.e. ribosomal RNA in the present case) and makes it feasible to conduct studies in which many barcoded samples are processed and sequenced in a single-experiment setting. The lower part shows the Matthews Correlation Coefficient (MCC)[10] and optimal threshold values used in discrimination of unmethylated and methylated positions in the two different scoring systems that were applied in the study (See Materials and Methods for details on calculation). The two scoring systems give very similar detection results.

Supporting references

[1] C. B. Brachmann, A. Davies, G. J. Cost, E. Caputo, J. Li, P. Hieter, J. D. Boeke, *Yeast (Chichester, England)* **1998**, *14*, 115-132.

[2] M. Saikia, Q. Dai, W. A. Decatur, M. J. Fournier, J. A. Piccirilli, T. Pan, *RNA (New York, N.Y.)* **2006**, *12*, 2025-2033.

[3] B. Liu, M. J. Fournier, *RNA (New York, N.Y.)* **2004**, *10*, 1130-1141.

[4] F. Carrillo Oesterreich, S. Preibisch, K. M. Neugebauer, *Mol Cell* **2010**, *40*, 571-581.

[5] a) B. S. Remus, S. Shuman, *RNA (New York, N.Y.)* **2014**, *20*, 462-473; b) M. Englert, H. Beier, *Nucleic acids research* **2005**, *33*, 388-399.

[6] L. Zheng, U. Baumann, J. L. Reymond, *Nucleic acids research* **2004**, *32*, e115.

[7] K. Schutz, J. R. Hesselberth, S. Fields, *RNA (New York, N.Y.)* **2010**, *16*, 621-631.

[8] M. Martin, *EMBnet.journal* **2011**, *17*, 10-12.

[9] B. Langmead, S. L. Salzberg, *Nat Methods* **2012**, *9*, 357-359.

[10] B. W. Matthews, *Biochim Biophys Acta* **1975**, *405*, 442-451.

[11] T. Sing, O. Sander, N. Beerenwinkel, T. Lengauer, *Bioinformatics* **2005**, *21*, 3940-3941.

[12] S. Douthwaite, F. Kirpekar, *Methods in enzymology* **2007**, *425*, 3-20.

[13] Y. Li, R. R. Breaker, *Journal of the American Chemical Society* **1999**, *121*, 5364-5372.

[14] S. Belin, A. Beghin, E. Solano-Gonzalez, L. Bezin, S. Brunet-Manquat, J. Textoris, A. C. Prats, H. C. Mertani, C. Dumontet, J. J. Diaz, *PLoS One* **2009**, *4*, e7147.