RNA Modifications

DOI: 10.1002/ange.201408362

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

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

Abstract: *Ribose methylations are the most abundant chemical modifications of ribosomal RNA and are critical for ribosome assembly and fidelity of translation. Many aspects of ribose methylations have been difficult to study due to lack of efficient mapping methods. Here, we present a sequencing-based method (RiboMeth-seq) and its application to yeast ribosomes, presently the best-studied eukaryotic model system. We demonstrate detection of the known as well as new modifications, reveal partial modifications and unexpected communication between modification events, and determine the order of modification at several sites during ribosome biogenesis. Surprisingly, the method also provides information on a subset of other modifications. Hence, RiboMeth-seq enables a detailed evaluation of the importance of RNA modifications in the cells most sophisticated molecular machine. RiboMeth-seq can be adapted to other RNA classes, for example, mRNA, to reveal new biology involving RNA modifications.*

The major information-carrying macromolecules in the cell, DNA, RNA, and proteins, carry an additional layer of information on top of their sequence in the form of modifications of residues. The modifications provide functional groups and impact the structure and function of the molecules. Prominent examples are the the C-5-methylated cytosine modification (m^5C) in DNA, which is a key player in epigenetics, and phosphorylations of proteins. In contrast to the few modifications known from DNA, cellular RNA molecules contain more than 100 modifications of the four standard nucleotides and are found in eubacteria, archaea, and eukaryotes.[1] Approximately two-thirds of the types of modifications are methylations, and methylation of the 2′-O position in the ribose moiety of the nucleotide (2′-O-Me) is by far the most abundant.[1] Ribose methylations are found in all major classes of RNA in eukaryotic organisms. In ribosomal RNA (rRNA), most methylation sites are specified by guide RNAs and clustered in highly conserved and thus functionally important parts of the molecules.[2] The methylations are generally thought to be important for the folding of rRNA and the fidelity of the ribosome, but experiments are frequently enigmatic because ablation of individual methyl groups has little or no effect.[3] The lack of basic understanding impedes the analysis of the complex phenotypes that are characteristic of heritable human diseases that have defects in ribose methylation as the underlying molecular cause.[4] In other diseases, including cancer, the guide RNAs are found to be deregulated, and recently it was reported that cancer cells have alterations in ribose methylation at a subset of sites, brought about by deregulation of the methyltransferase fibrillarin.[5] It is clear that many aspects of ribose methylation are underexplored possibly because the existing methods for the detection of 2′-O-Me in RNA are laborious and can only address a few modification sites at a time.[6]

Here, we introduce RiboMeth-seq, a sequencing-based method for the determination of ribose methylations in RNA. RiboMeth-seq (Figure 1 and Figures S1–S5) is based on the activation of the 2′-OH of the ribose sugar by base for nucleophilic attack on the neighboring phosphodiester bond, resulting in strand cleavage with resulting 5′-OH and 2′,3′ cyclic phosphate ends. In contrast, the 2′-O-Me-modified ribose is resistant to alkaline cleavage (Figure 1A). The starting RNA is degraded under denaturing conditions into small fragments (20–40 nucleotides), long enough to be mapped by sequence alignment to the reference genome. The fragments are ligated to RNA oligomers using a tRNA ligase that has been mutated to remove its kinase activity, then reverse transcribed, and the cDNA is used as input for ion semiconductor sequencing. The first and last nucleotides of the inserts are recorded using the full sequence for mapping and the read-ends plotted against the sequence (Figure 1B). The nucleotides at the 3′ read-ends directly depend on their 2′-OH function, whereas the nucleotides at the 5′ read-ends result from the 2′-OH function of their 5′ neighbors. Thus, the 5′ read-ends are all shifted one position upstream in the data treatment such that reads corresponding to a methylated nucleotide will align in the two datasets. As 2′-O-Me nucleotides are resistant to alkaline degradation, these will

[*] U. Birkedal,[+] M. Christensen-Dalsgaard,[+] N. Krogh, Prof. H. Nielsen
Department of Cellular and Molecular Medicine
University of Copenhagen
Blegdamsvej 3B, 2200 Copenhagen N (Denmark)
E-mail: hamra@sund.ku.dk
R. Sabarinathan, J. Gorodkin
Center for Non-coding RNA in Technology and Health
Department of Veterinary Clinical and Animal Science
University of Copenhagen
Grønnegårdsvej 3, 1870 Frederiksberg C (Denmark)
[+] These authors contributed equally to this work.

[**] We thank Wayne A. Decatur for yeast strains; Jay Hesselberth for tRNA ligase expression constructs; Benoit Masquida for HDV plasmid; Stephen Bruce Levery and Adnan Halim for help with mass spectrometry; David Lilley for advice; and Anders Lund, Eric Westhof, and Jan Christiansen for critical reading of the manuscript. This research was supported by The Danish Council for Independent Research, Technology and Production Sciences (H.N.), The Lundbeck Foundation (grant R19A2306 to J.G.), and The Danish Council for Strategic Research (J.G.).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201408362.

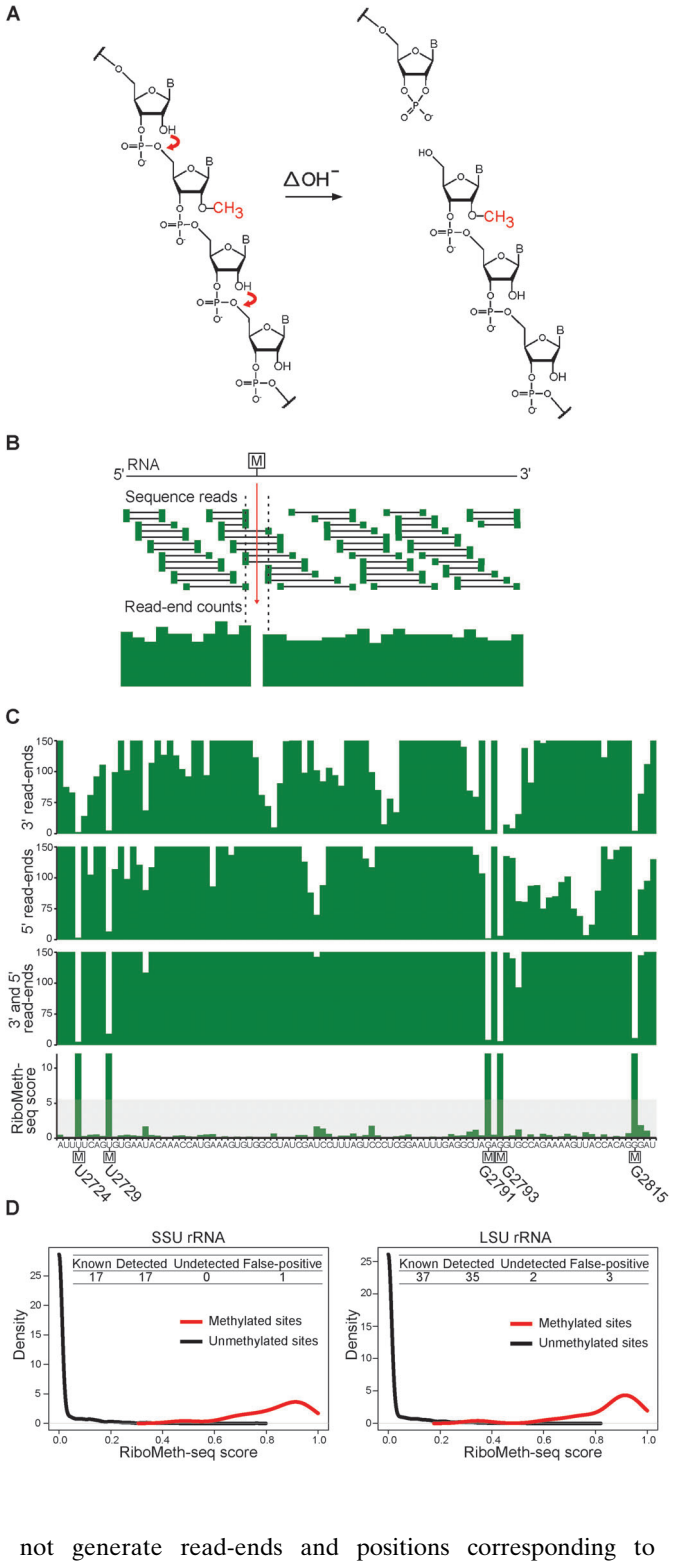

Figure 1. Detection of ribose methylations in RNA using RiboMeth-seq. A) The principle of RiboMeth-seq is that ribose methylation renders the neighboring phosphodiester bond resistant to (partial) alkaline degradation. B) The 20–40 nt fragments from partial alkaline hydrolysis are ligated to adaptors, reverse transcribed, and sequenced. The sequence is mapped to the reference sequence and the first and last nucleotide of the library fragment recorded. Read-ends at methylated residues will be underrepresented compared to flanking positions. C) Read-ends from either end of the fragments are recorded separately and combined, and a RiboMeth-seq score is calculated. The figure is based on a screenshot from the UCSC browser with a segment of LSU rRNA comprising five methylation sites. The gray-shaded area indicates the calculated optimal threshold. Links to the full dataset can be found in the Supporting Information. D) Distributions of scores of annotated unmethylated and methylated residues, respectively, showing that the two distributions are clearly separated. The inserts in the figures summarize the detection statistics for RiboMeth-seq analysis of yeast ribosomal RNA.

not generate read-ends and positions corresponding to a methylated nucleotide will thus be underrepresented creating a negative image of the methylation landscape. Finally, a scoring method is applied to convert this into a representation that normalizes the reads in relation to reads from 12 neighboring positions and show methylated nucleotides as peaks (Figure 1C). We have applied the method to budding yeast (*S. cerevisiae*) rRNA, which is currently the best-described model system for ribose methylation. The SSU (18S) and LSU (25S) rRNA contain 54 known 2′-O-Me modifications, and all but one are guided by 43 C/D box snoRNAs. The box C/D snoRNAs base pair extensively with the target RNA, and a methyltransferase (NOP1; fibrillarin in human) then modifies the fifth nucleotide counting from box D.[7] In a RiboMeth-seq analysis of gel-isolated rRNA, we detect 17 out of 17 positions in SSU with one potential false-positive and 35 out of 37 positions in LSU with three potential false-positives (Figure 1D, Table S1, and Figure S6).

RiboMeth-seq has the potential to provide quantitative data at single-nucleotide resolution because the method is based on a simple chemical principle, namely the difference in the nucleophilicity of a 2′-OH group and a 2′-O-Me group,

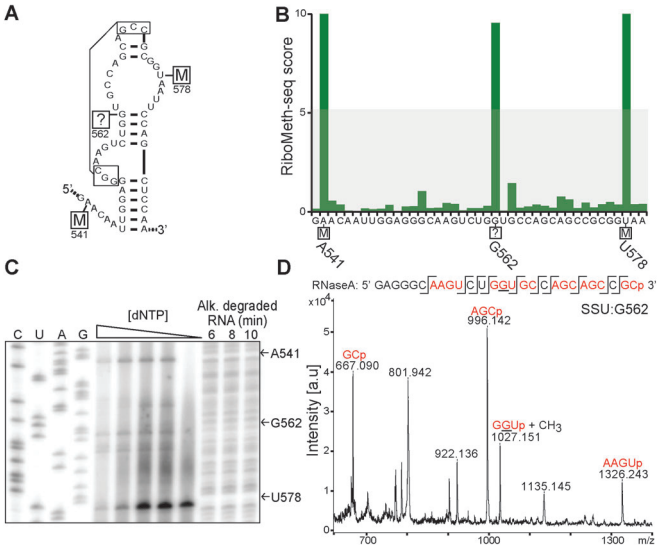

Figure 2. Discovery of a novel ribose methylation site. A) SSU-G562 is located in a conserved part of rRNA flanked by two known methylation sites. B) The RiboMeth-seq score of SSU-G562 is comparable to that of the flanking positions. C) Analysis of SSU-G562 by conventional primer extension methods produces only a faint stop signal at low dNTP concentration and a gap in the alkaline ladder. The sequencing ladder is labeled with the sequence corresponding to the RNA-like strand. D) MALDI mass spectrum of an RNase A digested fragment comprising SSU-G562. The fragment 5′ GGU has an additional mass corresponding to a methyl group.

which amounts to several orders of magnitude.[8] The average level of methylation calculated across all positions (Figure S7) was 0.94 compared to 0.82 ribose methylations per methylated residue obtained in an early study by calculation of the ratio of in vivo incorporation of ^{32}P and Me-^{14}C into methylated nucleotides.[9] The distribution of RiboMeth-seq fraction scores shows that the majority of positions (33/54) probably are close to fully methylated with scores >0.95 and that a small number of positions (8/54) have fraction scores below 0.9 (Figure S7). The degree of methylation at selected positions was verified using a quantitative real-time PCR-based method (Figure S7) and suggests that the two false-negatives in the analysis (Figure 1D) represent hypomethylated residues. The observed low fraction scores imply that a fraction of rRNA molecules are unmethylated at some positions and thus imply ribosome diversity at the level of ribose methylation.

We speculated that the false-positives in the RiboMeth-seq analysis could represent un-annotated modifications and subjected the four candidates to analysis by two conventional biochemical methods and mass spectrometry (Figure 2 and Figure S8). The biochemical methods provided supporting evidence for ribose methylation at SSU-G562 and LSU-G1142. SSU-G562 is located in Helix 18, which is involved in decoding. It is flanked by SSU-A541 and SSU-U578 (Figure 2A) and all three positions have high RiboMeth-seq scores (Figure 2B). The biochemical methods produced only faint signals (Figure 2C), but mass spectrometry revealed an additional mass signal corresponding to a methyl group on the expected RNase A fragment (5′ GGU) (Figure 2D). We conclude that SSU-G562 is ribose methylated and should be added to the list of modified nucleotides in *S. cerevisiae*. A comprehensive paper chromatography mapping of methylated nucleotides in *S. carlsbergensis* estimated that 55 ribose methylations should be found in yeast,[9] but so far only 54 have been annotated in the databases.[10] The missing modification was predicted to be at a GmU dinucleotide in SSU, and we conclude that we have mapped this missing modification.

In order to demonstrate the applicability of the method for the detection of changes in methylation patterns, we analyzed whole-cell (WC) RNA from wild-type (wt) yeast, a knock-out strain in which one of the guide RNAs (snR67) and its corresponding methyl group at position LSU-G2619 has been deleted,[11] and a knock-in strain that carries an artificial snoRNA which guides methylation at LSU-U2954[12] (Figure 3). LSU-G2619 is a universally conserved modification in the P-site and LSU-U2954 is a critical residue for P-site tRNA binding (Figure 3C). As shown in Figure 3A, the knock-out mutation results in complete loss of methylation of LSU-G2619. Residues in SSU are unaffected by the mutation, but, surprisingly, methylation at a few LSU sites, for example, LSU-U2347 and LSU-U2421, are reduced (Figure S9). The knock-in mutation results in prominent methylation at the target site (LSU-U2954) and reduced methylation at three neighboring sites (Figure 3B). This can be explained by an overlap between the artificial snoRNA guide target site and the three natural target sites (Figure 3D). In addition, methylation is affected at several distant sites in SSU and LSU (Figure S9). Thus, the RiboMeth-seq analysis reveals

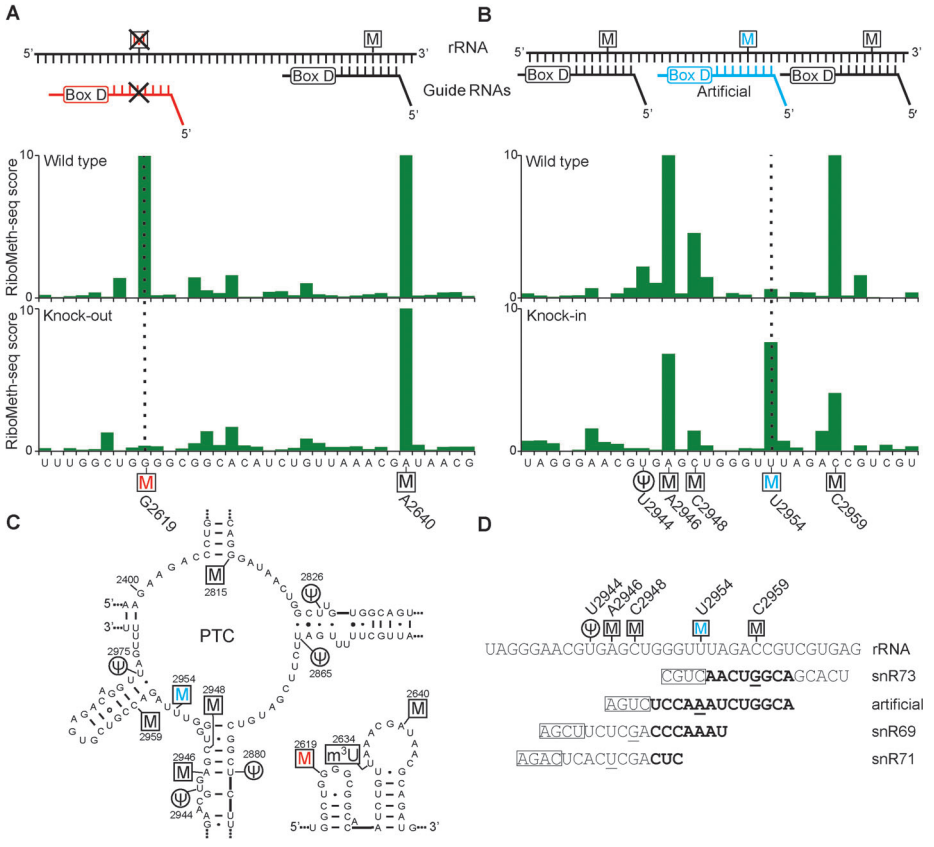

Figure 3. RiboMeth-seq analysis of strains with knock-out or knock-in of single ribose methylations. A) Deletion of LSU-G2619me was achieved by genetic deletion of snR67 (in red) resulting in complete loss of RiboMeth-seq signal at G2619. The signal at the neighboring A2640 was unaffected. B) Introduction of LSU-G2954me was done by expression of an artificial box C/D snoRNA targeting this site (in blue). The new methylation was readily detected by RiboMeth-seq. Methylation of flanking positions was reduced. C) Secondary structure diagrams of the positions with deletion (red) and insertion (blue) of methylations and their immediate surroundings in the PTC region of LSU rRNA. D) The artificial box C/D snoRNA overlaps with other target sites when base-paired to its target and thus has the potential to interfere with methylation at those sites. The D-boxes are indicated and the nucleotides base paired to the methylated nucleotides are underlined. Overlapping nucleotides are in bold.

interdependency between distant ribose methylation events and suggests that re-analysis of the many methylation mutants reported in the literature could provide important information on communication between different parts of the pre-ribosome.

A major unresolved issue is the order and timing of modification events during ribosome biogenesis. As a first step towards this, we compared the ribose methylation pattern in chromatin-associated and mature ribosomal RNA from *S. cerevisiae*. RNA was isolated according to a protocol aimed at isolation of nascent RNA. The degree of processing in this RNA can be deduced directly from the RiboMeth-seq data because ends that were generated prior to the alkaline treatment give rise to signatures that are shifted approximately 20 nt in the 5′ end and 3′ end datasets due to depletion of fragments <20 nt in the cloning procedure. Inspection of the processing sites in Figure 4A shows that this RNA comprises nascent RNA and processing intermediates up to the first steps after the cleavage that separates the two main parts of the transcript. It has previously been shown by kinetic labeling experiments that ribose methylation is predominantly co-transcriptional in SSU and co- and post-transcriptional in LSU.[13] RiboMeth-seq analysis confirms this overall picture at single-nucleotide resolution. Figure 4B shows a pair-wise comparison of the degree of methylation in chromatin-associated and mature RNA at all positions. Long bars connecting the data points reveal late modifications. In the SSU, all modifications except one (A100) appear to be early. In contrast, the modification pattern during biogenesis of LSU appears much more varied with several late modifications. This is consistent with the observations from kinetic labeling experiments and likely reflects the complex biogenesis of LSU that derive from a mixture of co- and posttranscriptionally cleaved precursors.[13] Figure 4C shows a detailed view of two selected positions on either side of the peptidyltransferase ring. LSU-G2922 is located in the A loop and is known to be modified late, during maturation of the 27S intermediate by the Spb1p methyltransferase.[14] Accordingly, this position is unmethylated in chromatin-associated RNA. Another late modification is LSU-U2421 which is located close to the P-site. Interestingly, methylation at this residue was down-regulated in the knock-out experiment (Figure S9)

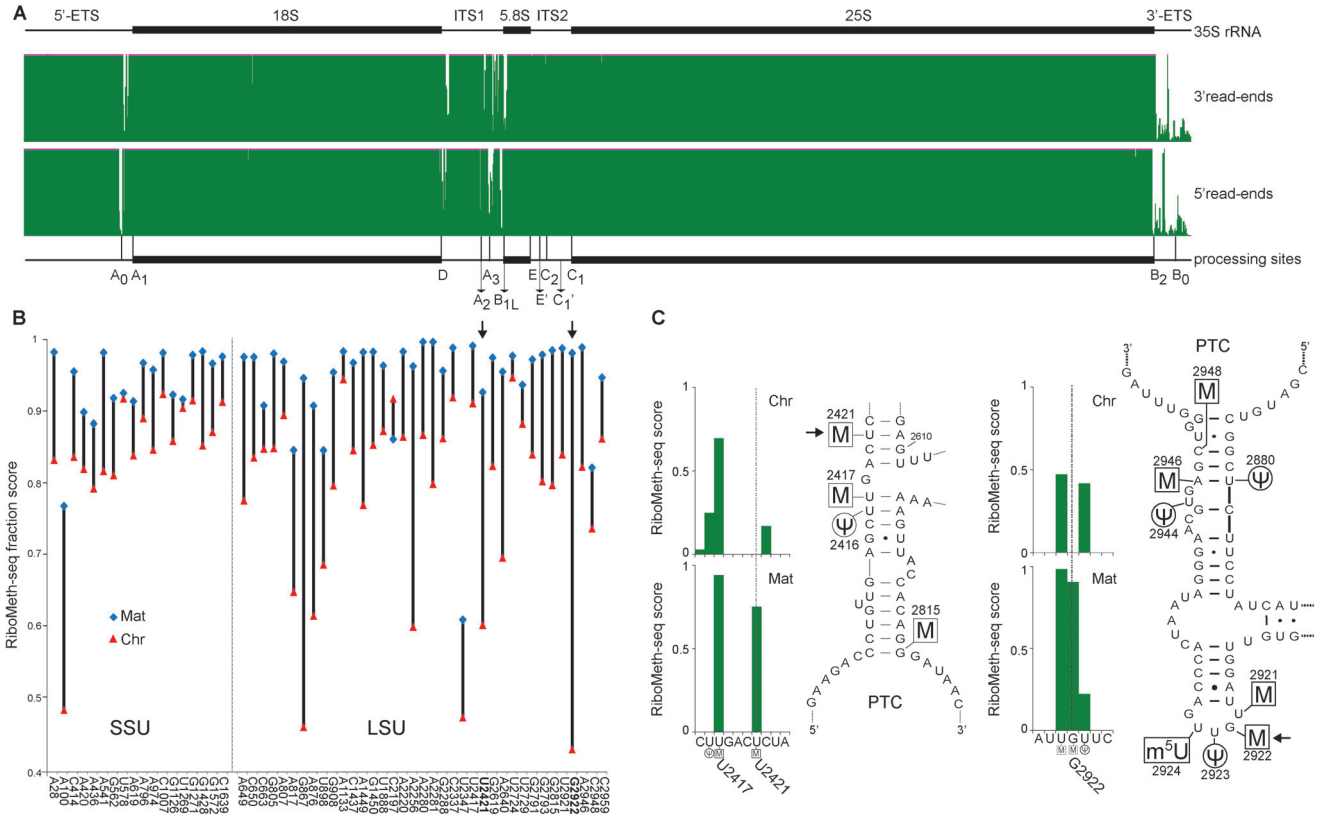

Figure 4. Ribose methylation in chromatin-associated and mature ribosomal RNA. A) Browser view of 3′ and 5′ read ends sandwiched between maps of the composition of yeast pre-rRNA and the processing sites (*21*). The read-end tracks are shown at low resolution and truncated at 500 reads to emphasize the signal produced by cleavage and trimming at the processing sites. This signal is shifted by 20 nt between the two panels corresponding to the lower cut-off limit during library production. The figure illustrates that almost all transcripts are cleaved at their 3′ end by Rnt1p at B_0, and in the 5′ ETS at A_0. A substantial fraction is also cleaved at A_1 and at A_2 which splits the transcripts into the two main parts. The 5′ fragment (20S) is normally exported to the cytoplasm for further processing and these processing steps are not represented in the chromatin-associated RNA, except for a small fraction cleaved at D. The 3′ fragment (27S-A_2) is further processed along two alternative pathways, and minor fractions are cleaved at the sites initiating these two pathways (A_3 and B_{1L}, respectively). B) Comparison of the fraction of molecules methylated in chromatin-associated (red) and mature (blue) rRNA across all 55 methylated positions. C) Two examples of nucleotides that are modified later in ribosome biogenesis than neighboring nucleotides.

suggesting that the knock-out specifically affects a late step in processing. Our data suggest as a general theme that a subset of late modification sites are hypomethylated and susceptible to changes upon perturbations of ribosome biogenesis.

In addition to ribose methylations, RiboMeth-seq provides information on a subset of base methylations (5/11 in yeast) that are found on the Watson–Crick edge of the base due to misincorporation in the cDNA synthesis step (Figure S10). Most of the pseudouridines are detected due to a ligation bias in the cloning step (Figure S10) and sequence variation among rDNA copies provide direct evidence for ribosome diversity. Hence, RiboMeth-seq enables a detailed evaluation of the biological importance of RNA modification and their subsequent significance for translational fidelity. Recently, it has become clear that the role of the ribosome in the synthesis of cellular proteins not is passive. Rather, specialized ribosomes exist that preferentially translate certain types of mRNAs.[15] The structural diversity is generally thought to be caused by transient association with protein factors or small RNAs but the molecular basis for their association is unclear and changes in modifications of the ribosome itself may play a role. RiboMeth-seq can be scaled up to analysis of full transcriptomes. Here, it will be of interest to find the targets of the many orphan guide RNAs that have been reported in humans.[16] Sequencing-based methods are required to map the epitranscriptome and probe its functions as exemplified by the burst of discoveries that has followed the recent development of a method for mapping of m^6A in transcriptomes.[17]

Experimental Section

The protocol for RiboMeth-seq uses as input 1 μg of 20–40 nt fragments gel-purified after partial alkaline hydrolysis. The library fragments have 5′ OH and 2′,3′ cyclic phosphate ends and were cloned using tRNA ligase. Thus, 5′ adaptors with 2′,3′ cyclic phosphate were generated by cleavage of an in vitro transcript incorporating an HDV ribozyme and a mutated version of *Arabidopsis* tRNA ligase without kinase activity was constructed and used for ligation. Prior to cDNA synthesis, the 2′-O phosphates left by the tRNA ligation step were removed in a standard alkaline phosphatase reaction. The adaptors incorporate sequences compatible with the Ion Torrent system that was used as sequencing platform. Reads were annotated to the precursor ribosomal RNA (35S rRNA) gene of *Saccharomyces cerevisiae* s288c (RDN37-1; SGD:S000006486, GeneID: 9164931)) and used to compute the number of 3′ read-ends and 5′ read-ends (Figures S4 and S5), 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 RiboMeth-seq scores. The scores are based on a comparison of the read number at each position to the average read number and standard deviation of six neighboring positions on either side. For detection of methylations, a Matthews's correlation coefficient is used to determine the optimal cut-off value.

Received: August 19, 2014
Published online: November 21, 2014

Keywords: ribosomes · ribozymes · RNA recognition · RNA structures · transferases

[1] Y. Motorin, M. Helm, *Wiley Interdiscip. Rev. RNA* **2011**, *2*, 611 – 631.
[2] a) J. P. Bachellerie, J. Cavaille, *Trends Biochem. Sci.* **1997**, *22*, 257 – 261; b) W. A. Decatur, M. J. Fournier, *Trends Biochem. Sci.* **2002**, *27*, 344 – 351.
[3] W. A. Decatur, X. H. Liang, D. Piekna-Przybylska, M. J. Fournier, *Methods Enzymol.* **2007**, *425*, 283 – 316.
[4] a) B. Gonzales, D. Henning, R. B. So, J. Dixon, M. J. Dixon, B. C. Valdez, *Hum. Mol. Genet.* **2005**, *14*, 2035 – 2043; b) E. F. Freed, F. Bleichert, L. M. Dutca, S. J. Baserga, *Mol. Biosyst.* **2010**, *6*, 481 – 493.
[5] V. Marcel, S. E. Ghayad, S. Belin, G. Therizols, A. P. Morel, E. Solano-Gonzalez, J. A. Vendrell, S. Hacot, H. C. Mertani, M. A. Albaret, J. C. Bourdon, L. Jordan, A. Thompson, Y. Tafer, R. Cong, P. Bouvet, J. C. Saurin, F. Catez, A. C. Prats, A. Puisieux, J. J. Diaz, *Cancer Cell* **2013**, *24*, 318 – 330.
[6] S. Kellner, J. Burhenne, M. Helm, *RNA Biol.* **2010**, *7*, 237 – 247.
[7] Z. Kiss-Laszlo, Y. Henry, J. P. Bachellerie, M. Caizergues-Ferrer, T. Kiss, *Cell* **1996**, *85*, 1077 – 1088.
[8] Y. Li, R. R. Breaker, *J. Am. Chem. Soc.* **1999**, *121*, 5364 – 5372.
[9] J. Klootwijk, R. J. Planta, *Eur. J. Biochem.* **1973**, *39*, 325 – 333.
[10] D. Piekna-Przybylska, W. A. Decatur, M. J. Fournier, *Nucleic Acids Res.* **2008**, *36*, 178 – 183.
[11] M. Saikia, Q. Dai, W. A. Decatur, M. J. Fournier, J. A. Piccirilli, T. Pan, *RNA* **2006**, *12*, 2025 – 2033.
[12] B. Liu, M. J. Fournier, *RNA* **2004**, *10*, 1130 – 1141.
[13] M. Koš, D. Tollervey, *Mol. Cell* **2010**, *37*, 809 – 820.
[14] B. Lapeyre, S. K. Purushothaman, *Mol. Cell* **2004**, *16*, 663 – 669.
[15] S. Xue, M. Barna, *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 355 – 369.
[16] J. Cavaille, K. Buiting, M. Kiefmann, M. Lalande, C. I. Brannan, B. Horsthemke, J. P. Bachellerie, J. Brosius, A. Huttenhofer, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 14311 – 14316.
[17] a) D. Dominissini, S. Moshitch-Moshkovitz, S. Schwartz, M. Salmon-Divon, L. Ungar, S. Osenberg, K. Cesarkas, J. Jacob-Hirsch, N. Amariglio, M. Kupiec, R. Sorek, G. Rechavi, *Nature* **2012**, *485*, 201 – 206; b) T. Pan, *Trends Biochem. Sci.* **2013**, *38*, 204 – 209.

