Methods 155 (2019) 30–40

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Genome-wide probing RNA structure with the modified DMS-MaPseq in Arabidopsis

Zhiye Wang[a,b,1], Meiyue Wang[c,d,1], Tian Wang[a,b,e], Yijing Zhang[c,d,*], Xiuren Zhang[a,b,*]

[a] Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843, USA
[b] Institute for Plant Genomics and Biotechnology, Texas A&M University, College Station, TX, 77843, USA
[c] National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences (CAS), 300 Fenglin Road, Shanghai 200032, China
[d] University of the Chinese Academy of Sciences, Beijing 100049, China
[e] College of Food Science and Nutritional Engineering, China Agricultural University, Beijing 100083, China

ARTICLE INFO

Keywords:
Arabidopsis
RNA structure
Genome-wide profiling
Dimethyl sulfate
Thermostable group II intron reverse transcriptase
Next-generation sequencing

ABSTRACT

Transcripts have intrinsic propensity to form stable secondary structure that is fundamental to regulate RNA transcription, splicing, translation, RNA localization and turnover. Numerous methods that integrate chemical reactions with next-generation sequencing (NGS) have been applied to study in vivo RNA structure, providing new insights into RNA biology. Dimethyl sulfate (DMS) probing coupled with mutational profiling through NGS (DMS-MaPseq) is a newly developed method for revealing genome-wide or target-specific RNA structure. Herein, we present our experimental protocol of a modified DMS-MaPseq method for plant materials. The DMS treatment condition was optimized, and library preparation procedures were simplified. We also provided custom scripts for bioinformatic analysis of genome-wide DMS-MaPseq data. Bioinformatic results showed that our method could generate high-quality and reproducible data. Further, we assessed sequencing depth and coverage for genome-wide RNA structure profiling in Arabidopsis, and provided two examples of in vivo structure of mobile RNAs. We hope that our modified DMS-MaPseq method will serve as a powerful tool for analyzing in vivo RNA structurome in plants.

1. Introduction

Once thought to be only a messenger bridge between DNA and proteins, RNA is now known to influence many aspects of biology through activities that are attributable to its secondary and tertiary structures. These structures are formed through intra- and/or intermolecular base pairing and can serve as catalytic bases for ribozymes and scaffold for assembling ribosomes. The structures can also provide new regulatory layers of gene expression in various steps of post-transcriptional regulation including RNA processing, modification, stability and translation [1–4].

Over the past decades, several methods based on in vivo chemical modifications coupling with NGS have been developed for probing RNA structure [5–15]. Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) reagents and dimethyl sulfate (DMS) are two major chemical adducts for RNA modifications. Because of its low cost, small size and high reactivity, DMS has been the most popular chemical for in vivo RNA structure analysis [16]. DMS alkylates the Watson-Crick face of unpaired adenosine (A) and cytosine (C). Modified A and C residues generally block reverse transcription (RT), resulting in truncations of complementary DNA (cDNA). These RT terminal sites are detected by gel electrophoresis or high-throughput sequencing [5,6,17,18]. One limitation for this method is that RT reactions of DMS-treated transcripts tend to stop near the primer-binding regions, resulting in a

Abbreviations: NGS, next-generation sequencing; DMS, dimethyl sulfate; DMS-MaPseq, dimethyl sulfate probing coupled with mutational profiling through next-generation sequencing; SHAPE, selective 2′-hydroxyl acylation analyzed by primer extension; RT, reverse transcription; cDNA, complementary DNA; TGIRT, thermostable group II intron reverse transcriptase; mRNA, messenger RNA; pri-miRNAs, primary miRNAs; rRNA, ribosomal RNA; MS, Murashige and Skoog; ORF, open reading frame; UTR, untranslated region

* Corresponding authors at: National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences (CAS), 300 Fenglin Road, Shanghai 200032, China (Y. Zhang). Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843, USA (X. Zhang).
E-mail addresses: zhangyijing@sibs.ac.cn (Y. Zhang), xiuren.zhang@tamu.edu (X. Zhang).
[1] These authors contributed equally: Zhiye Wang, Meiyue Wang.

https://doi.org/10.1016/j.ymeth.2018.11.018
Received 1 September 2018; Received in revised form 20 November 2018; Accepted 27 November 2018
Available online 29 November 2018

highly skewed distribution of reads near the primer-binding sites [13,19]. Recently, a new in vivo RNA structure probing technology called DMS mutational profiling with sequencing (DMS-MaPseq) has been developed [13]. Key to this method is the substitution of a regular reverse transcriptase with a special RT enzyme from group II self-splicing intron, a thermostable group II intron reverse transcriptase (TGIRT) [20]. TGIRT can read through the DMS-modified A and C of RNA templates, and insert mismatches in the corresponding cDNA, thereby offering an approach for mapping DMS modifications using full-length RT products. Moreover, TGIRT has higher processivity and fidelity, and can function at higher temperature (more than 60 °C) to overcome the effect of strong secondary structure of RNAs on the RT reaction [20]. Technically, the frequency of mutation per nucleotide (referred to as ratiometric DMS signal), representing unstructured regions, is counted through NGS and bioinformatic analysis. DMS-MaPseq provides high-quality RNA structure data with high signal and low background, allowing more precisely probing both global RNA structure and specific RNA targets, including low-abundant RNAs [13].

By contrast to bacteria, yeast and mammals [5,9,13,14,19,21–28], in vivo RNA structure probing methods coupled with NGS are less frequently applied in plants. Genome-wide profiling of in vivo RNA structure in Arabidopsis was performed based on DMS modification coupled with RT-truncation approach [6]. This study revealed several novel structural features of messenger RNA (mRNA). Later, the in vivo RNA structure in rice seedlings was also profiled using the same method, uncovering the conservation and divergence in RNA structure between dicotyledon and monocotyledon [29], and connecting heat stress and RNA structurome changes in the crop [30]. Recently, our lab adopted the modified target-specific DMS-MaPseq [31] to probe in vivo secondary structure of primary miRNAs (pri-miRNAs) in Arabidopsis plants, demonstrating that SWI2/SNF2 ATPase CHR2/BRM remodels secondary structure of pri-miRNAs to inhibit miRNA biogenesis at a post-transcriptional level. However, whether DMS-MaPseq can be optimized to detect global mRNA secondary structure in vivo remains unknown.

In this paper, we described a modified DMS-MaPseq approach for probing in vivo RNA structurome at a nucleotide resolution in Arabidopsis. We optimized the DMS treatment condition in both seedlings and adult plants. We also integrated TGIRT and a commercial kit, providing a user-friendly and time-saving protocol for library construction. Moreover, the custom scripts were developed for bioinformatic data processing. Later, we validated the quality and reproducibility of sequencing data and estimated sequencing depth and coverage for genome-wide RNA structure analysis in Arabidopsis. Using *U1* snRNA as an example, we found that the structure predicted through DMS-MaPseq data was consistent with the established reference folding. Finally, the secondary structures of two long-distance mobile RNAs were modeled based on our DMS-MaPseq data. It is our intention that this method will pave a road for plant researcher community to further explore RNA structure and study its biological function.

2. Method

2.1. Overview

The workflow of our method is shown in Fig. 1. Briefly, seedlings or adult plants are harvested and treated with DMS, either applied with vacuum or without vacuum. Treated samples are immediately frozen by liquid nitrogen (N_2) and ground to fine powder. Next, total RNA is extracted. After DNase treatment and depletion of ribosomal RNA (rRNA), cDNA library is generated using Illumina TruSeq® Stranded Total RNA Library Prep Plant kit, where RT is performed with TGIRT. Following assessment of concentration and size distribution of cDNA by an Agilent bioanalyzer, the cDNA library is sequenced in an NGS machine that is compatible with TruSeq® adapters. The raw sequencing data are filtered with quality, and mapped to the reference genome with TopHat pipeline (https://ccb.jhu.edu/software/tophat/index.shtml), and uniquely mapped reads are extracted. Finally, the ratiometric DMS signal per nucleotide is calculated by custom scripts to infer the in vivo RNA structure.

2.2. Detailed experimental protocol

This protocol describes the preparation of one sample for genome-wide in vivo RNA secondary structure profiling in Arabidopsis. Nuclear RNA is used for the library construction in this protocol. This method is also suitable for any other kinds of RNA, such as total RNA and chloroplast RNA.

2.2.1. In vivo DMS reaction

Since DMS is extremely toxic, this chemical should be always handled in a fume hood including grinding DMS treated samples. Non-latex gloves should be worn, and cautions should be taken to ensure there is no DMS contamination on clothing. After the DMS treatment, β-mercaptoethanol is added to react with excess soluble DMS. The resulting liquid waste must be pooled into a specific container for hazardous waste, and picked up by local Safety & Health Office.

1. Harvest 2–3 g 10-day-old seedlings growing on the Murashige and Skoog (MS) agar medium or three-week-old adult plants growing on soil.
   Note: The agar medium or soil should be avoided during harvest.
2. Immerse the sample in a 50 ml Falcon tube containing 25 ml DMS reaction solution (40 mM HEPES pH 7.5, 100 mM KCl and 0.5 mM $MgCl_2$). Pipet 250 µl of DMS (final concentration 1%, v/v) into the solution and mix well immediately. For a negative control, add the same volume of water into the solution.
3. Treat the sample for 30 min with or without vacuum (approximately 12 psi) at room temperature, with swirling the solution every 10 min.
4. Quench the DMS reaction by addition of β-mercaptoethanol (final concentration 20%, v/v) to react with excess soluble DMS. Place the tissue under vacuum for an additional 5 min.
5. Wash the tissue three times with 50 ml distilled water, following by blotting the tissue between paper towels to remove as much water as possible.
6. Freeze the samples with liquid N_2 and grind to fine powder.

2.2.2. Nuclear RNA extraction

7. Mix 0.3 g powder with 4 ml of pre-cooled lysis buffer (20 mM HEPES, pH 7.5, 20 mM KCl, 2.5 mM $MgCl_2$, 25% glycerol, 250 mM Sucrose, 5 mM DTT, 40 U/ml SUPERase·In™ RNase Inhibitor [Thermo Fisher] and 1 × cOmplete™ EDTA-free Protease Inhibitor Cocktail [Roche]).
8. Wet a double layer of Miracloth with 0.5 ml lysis buffer. Filter the homogenate through the double layer of Miracloth into a Falcon round-bottom tube that is placed on ice.
9. Spin the filtered solution at 1,500 *g* at 4 °C for 10 min in a swinging-bucket rotor.
10. Discard the supernatant. Gently resuspend the pellet with 4 ml of pre-cooled nuclei resuspension buffer (20 mM HEPES, pH 7.5, 25% glycerol, 2.5 mM $MgCl_2$, 0.2% Triton X-100 and 20 U/ml SUPERase·In™ RNase Inhibitor [Thermo Fisher]).
    Note: It is important to resuspend the pellet gently by pipetting to avoid breaking nuclei.
11. Spin the solution at 1,500 *g* at 4 °C for 10 min in a swinging-bucket rotor.
12. Repeat step 10 and 11 twice.
13. Discard the supernatant. Resuspend the pellet with 1 ml TRIzol® Reagent (Thermo Fisher) and isolate RNA according to the manufacturer's instructions.

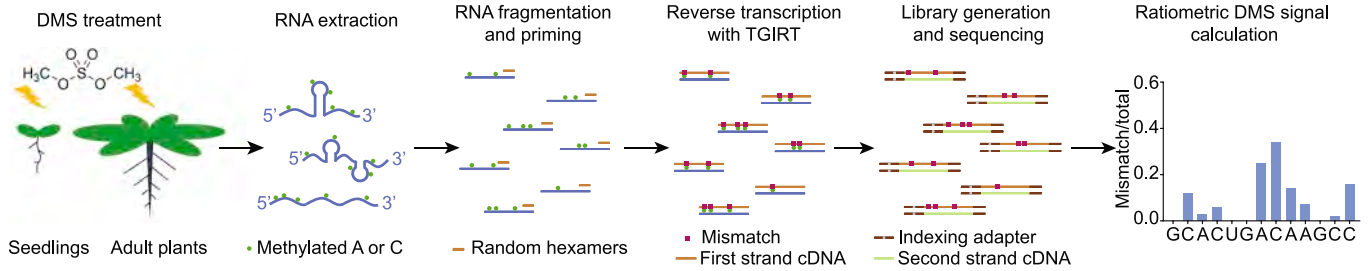


Fig. 1. Overview of the modified DMS-MaPseq method for genome-wide probing in vivo RNA secondary structure in plants.

14. Measure the concentration of isolated nuclear RNA using the NanoDrop.
15. To quickly assess the integrity of nuclear RNA, take one μg RNA from each sample into 10 μl of FA buffer (95% deionized formamide, 5 mM EDTA, 0.025% SDS, 0.025% bromophenol blue and 0.025% xylene cyanol); Heat the mixture at 95 °C for 3 min and quick chill on ice; Then run the mixture on a standard agarose gel stained with ethidium bromide (EtBr). The intact RNA sample will have sharp, and clear bands of 28S and 18S rRNA.
    Note: Ahead of assessment RNA integrity by gel electrophoresis, clean the trays, combs and electrophoresis tank with RNaseZAP™ to decontaminate potential RNases.
16. Normalize RNA concentrations between the samples according to image quantification of rRNA bands' signal intensity.

2.2.3. DNase treatment

17. Add 15 μg of nuclear RNA to a 50 μl 1x TURBO™ DNase Reaction buffer containing 4 U TURBO™ DNase (Thermo Fisher) and mix well by pipetting.
18. Incubate the mixture at 37 °C for 20 min.
19. Adjust the mixture to a volume of 100 μl with DNase/RNase-free water. Then the DNase-treated nuclear RNA is cleaned up using RNeasy Mini Kit (QIAGEN) according to the manufacturer's RNA cleanup protocol.
20. Measure the concentration of DNase-treated nuclear RNA using NanoDrop equipment.
21. Assessment of the integrity of DNase-treated nuclear RNA is described as step 15.

2.2.4. Library construction

DMS-MaPseq library is prepared using Illumina TruSeq® Stranded Total RNA Sample Prep Plant kit (Illumina) with some modifications in the RT step. These modifications enable the kit to be compatible with TGIRT. We typically start with 2 μg of DNase-treated RNA for each sample.

22. rRNA depletion is performed by an Ribo-Zero® rRNA removal kit for plant (Illumina) according to the manufacturer's manual.
23. Fragmentation and priming are performed by TruSeq® Stranded Total RNA Sample Prep Plant kit. In the RNA fragmentation step, heat the mixture at 95 °C for 3 min.
    Note: According to our experimental result, this fragmentation condition leads to RNA fragments of 100–200 nucleotides in length. Fragmentation time might need to be optimized for different RNA sample preparations.
24. Quick spin the PCR tube containing fragmented and primed RNA. Add 7.2 μl First Strand Synthesis Act D mix, 1 μl TGIRT™-III enzyme (InGex) and 0.5 μl 0.2 M DTT solution (fresh preparation) into the RNA solution. Gently mix by pipette.
25. Place the PCR tube on a thermal cycler and run the TGIRT RT program.
    a. Set the temperature of pre-heat lid to 100 °C.
    b. 25 °C for 10 min.
    c. 42 °C for 30 min.
    d. 60 °C for 1.5 h.
    Note: The step of 25 °C allows sufficient annealing between random hexamers and RNAs.
    Since the annealing temperature of random hexamers is low, directly changing RT reaction temperature from 25 °C to 60 °C might impair the annealing of random primers with templates. Therefore, cDNA is synthesized at a moderate temperature in first 30 min *(step c), then RT reaction is moved to high temperature (step d) to overcome the effect by high-order RNA structures. TGIRT has similar enzyme activity at both 42 °C and 60 °C* [20].
26. Transfer the reaction mixture to a new 1.5 ml Eppendorf tube. Adjust the sample to a volume of 140 μl with DNase/RNase-free water.
27. Add 140 μl phenol:chloroform:isoamyl alcohol (25:24:1, pH 8.0) and violent vortex for 30 s.
    Note: Violent vortex is important, because TIGRT binds templates tightly.
28. Centrifuge at 4 °C for 5 min at 12,000 rpm.
29. Transfer 110 μl upper solution to a new 1.5 ml Eppendorf tube. Add 75 μl chloroform:isoamyl alcohol (24:1) and vortex for 30 s.
30. Centrifuge at 4 °C for 5 min at 12,000 rpm.
31. Transfer 100 μl upper solution to a new 1.5 ml eppendorf tube. Add 10 μl 3 M Sodium acetate (pH 5.2), 1 μl GlycoBlue™ coprecipitant and 300 μl ethanol. Mix well and leave the sample at −80 °C for at least 1 h or −20 °C for overnight.
32. Centrifuge at 4 °C for 15 min at 12,000 rpm.
33. Discard the supernatant and wash the pellet with 500 μl 75% ethanol.
34. Centrifuge at 4 °C for 5 min at 12,000 rpm.
35. Decant the ethanol. Quick spin the eppendorf tube and remove the residue ethanol by pipetting. Air-dry the pellet for 5 min.
36. Resuspend the pellet in a 25 μl 1 × SuperScript® II first-strand buffer with 5 mM DTT.
37. Secondary strand cDNA synthesis, 3′ end adenylation, adapter ligation and DNA fragment enrichment are performed by using the TruSeq® Stranded Total RNA Sample Prep Plant kit (Illumina) according to the manufacturer's manual. The kit provides multiple indexing adapters. Unique barcodes for individual samples are introduced during the adapter ligation.
38. Assess the concentration and size distribution of library on an Agilent Bioanalyzer.
39. Qualify the library amount by NEBNext® Library Quant Kit for Illumina® (NEB) following the manufacturer's manual. According to the quantification result, libraries are pooled together with different barcodes.

2.2.5. High-throughput sequencing

Library can be sequenced on any sequencing platform that is compatible with TruSeq® adaptors. The length of sequencing reads is depended on the size range distribution of the cDNA library. We sequenced multiplexed library on an Illumina NextSeq 550 sequencer

with 75-bp single end reads.

2.3. Bioinformatics analysis

2.3.1. Demultiplexing

The raw sequencing data can be transferred from Illumina sequencing instruments to the BaseSpace Sequence Hub (https://basespace.illumina.com/home/index), where the raw data is demultiplexed by barcodes and converted to FASTQ file formats.

2.3.2. Cleaning sequencing reads

1. Assess the quality of the raw FASTQ files using FastQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).
2. According to the original DMS-MaPseq method [13], trim 2 nucleotides (nt) from 5′ end of each reads using FASTX-Toolkit Trimmer function (http://hannonlab.cshl.edu/fastx_toolkit/).
3. The reads are filtered based on sequencing quality using FASTX-Toolkit Quality Filter function, requiring that at least 80% of sequenced bases having quality score ≥30.
4. Keep reads with length ≥70 nt using FASTX-Toolkit Quality Trimmer function.
   Note: this parameter needs to be optimized for different types of libraries and the lengths of sequencing reads. Accordingly, the other parameter settings in the following computational part should be optimized.

2.3.3. Mapping to the reference genome

5. Map the cleaned reads to Arabidopsis thaliana reference genome (TAIR 10) using TopHat pipeline (https://ccb.jhu.edu/software/tophat/index.shtml). The parameter setting is - -library-type fr-firststrand - -no-novel-juncs -N 5 - -read-gap-length 7 - -read-edit-dist 7 - -max-insertion-length 5 - -max-deletion-length 5 -g 3. This setting allows final read alignments have up to 5 mismatches, total 7 nt gaps, 7 edit distances and 5 nt indel tolerance.
6. Extract uniquely mapped reads from bam files using Linux command grep with NH:I:1 tag. NH tag means the number of reported alignments that contain the query in the current record. For instance, NH tag is 1 for uniquely mapped reads.

2.3.4. Calculating ratiometric DMS signal

7. Counting mismatches is performed through a custom Python script (shown in http://bioinfo.sibs.ac.cn/arabRSS/mismatch_counting.html). Mismatches located within 3 nt of an indel are discarded.
8. The ratiometric DMS signal was calculated for each nucleotide as mismatch/sequencing depth using a custom bash script (shown in http://bioinfo.sibs.ac.cn/arabRSS/sequencing_depth_and_ratio.html). To reach reliable results, we kept bases with sequencing depth > 5 for calculation.

2.4. Troubleshooting

2.4.1. DMS treatment control

In contrast to truncation-based methods, DMS-MaPseq allows detection of multiple DMS-modification sites in a full-length RT product. The extent of DMS modifications could be assessed by gene-specific primer extension before library preparation. Compared to nontreatment, DMS treatment should result in dramatical decrease in full-length cDNA but increase in truncations in a primer extension assay. If the extent of DMS modifications is lower than desired, it might be due to DMS oxidation or insufficient vacuum pressure. The pressure of vacuum in our experiment is approximately 12 psi. Try to use a fresh bottle of DMS, or check pressure of bench vacuum.

2.4.2. RNA quality control

Low integrity of isolated RNA might result from frequent freezing and thawing samples without protection by a preservation reagent, RNase contamination in tips, tubes, reagents or gel electrophoresis accessories, or inappropriate practices during RNA extraction and handling. Besides, too harsh DMS treatment can cause RNA decay (see Section 3). If so, try to reduce DMS concentration and/or treatment time.

2.4.3. Library quality control

If the size of inserted cDNA fragments is shorter than 50 bp, optimization of RNA fragmentation condition is necessary. If the yield of library is lower than expected, try to increase initial amount of RNA or decrease DMS modification level by altering DMS concentration and/or treatment time.

3. Result and discussion

3.1. Optimize DMS treatment condition for plant materials

Genome-wide DMS-MaPseq for probing in vivo RNA structure has not been applied in plants yet. To optimize the DMS treatment condition for Arabidopsis, we treated 10-day-old seedlings and 3-week-old adult plants with different DMS concentrations (1%, 2% and 5%, v/v) and treatment time periods (10 min, 20 min and 30 min) with vacuum. We observed that high DMS concentrations (2% and 5%) or long treatment time (30 min) caused browning of plant samples (Fig. 2A). Like nucleotides, chlorophyll molecules have similar ring structures containing nucleophilic group, which might react with DMS at the high concentration levels and long-time exposure. Next, images of denaturing gel electrophoresis showed that RNA appeared to decay under 2% and 5% DMS treatment conditions, but not under 1% DMS treatment condition (Fig. 2B). The result suggests that treatment with high DMS concentration would cause RNA decay, although the precise mechanism is not clear.

To validate the extent of DMS modification on RNA, we did the primer extension assay for 18S rRNA in untreated and 1% DMS treated samples. The results showed that DMS treatment, compared with untreated, leaded to less yield of full-length cDNA but more truncations (Fig. 2C). Moreover, the imaging quantification showed that longer treatment resulted in a moderate increase in ratios of truncations relative to the full-length fragments (Fig. 2C).

We also studied the impact of vacuum treatment on DMS modifications. For this purpose, 10-day-old seedlings and 3-week-old adult plants were treated with different DMS concentrations (1%, 2% and 5%, v/v) for 30 min. Different from the vacuum treatment, DMS at low concentrations (like 1% and 2%) without vacuum did not easily cause browning of plant samples and severe RNA decay (Fig. S1A and B). However, the primer extension assay showed that the truncations of 18S rRNA cDNA had identical patterns with or without vacuum treatment. This result indicated that the vacuum treatment for 30 min does not alter 18S rRNA folding in vivo. Importantly, the extent of DMS modifications under the condition of 1% DMS with vacuum was higher than that of a comparable treatment without vacuum (Fig. S1C). This result suggested that vacuum treatment facilitated the penetration of DMS into plant cells. Notably, 2% DMS treatment without vacuum could also lead to sufficient DMS modifications (Fig S1C). Altogether, we recommended that the treatment with 1% DMS for 20–30 min under vacuum or with 2% DMS for 20 min without vacuum would be appropriate experimental conditions for both seedlings and adult plants.

3.2. Simplify library construction

In the original genome-wide DMS-MaPseq methods, the procedure of library construction is time-consuming, including gel-based size-selection. To simplify the library construction, we prepared library with

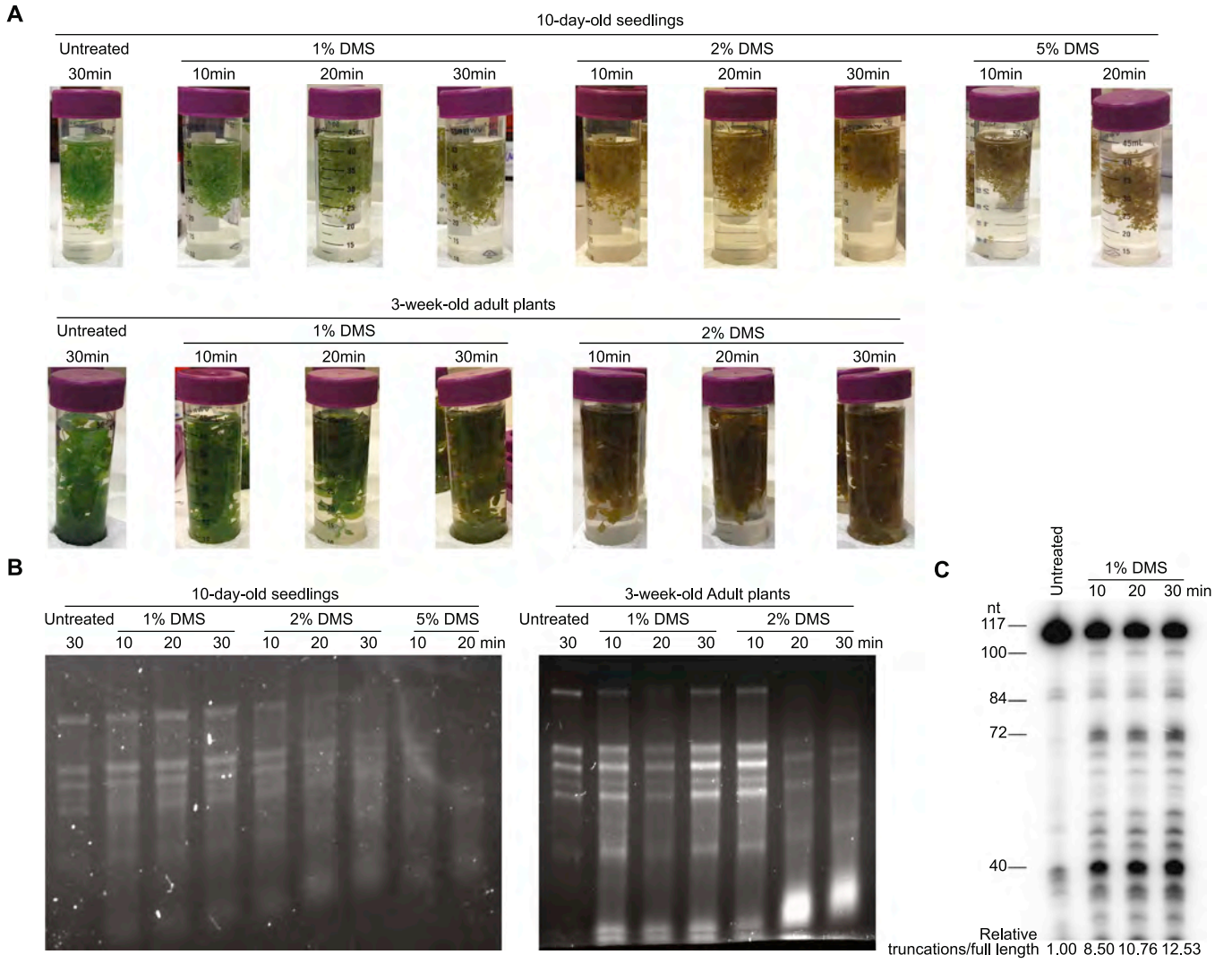


Fig. 2. Optimization of DMS treatment conditions for plant materials. (A) The phenotype of 10-day-old seedlings and 3-week-old adult plants treated with different DMS concentrations and different times under vacuum. (B) Denaturing gel electrophoresis of RNA samples showed that harsh DMS treatment caused RNA decay. (C) A primer-extension assay of 18S rRNA validated the extent of DMS modifications on RNA. The extent of DMS modification is indicated by the ratio of the signal intensity of total truncations to full-length cDNA (Relative truncations/full length). The ratio of truncations/full length was normalized to that of untreated samples, where the number was arbitrarily set to 1. SuperScript® IV reverse transcriptase was used for this assay.

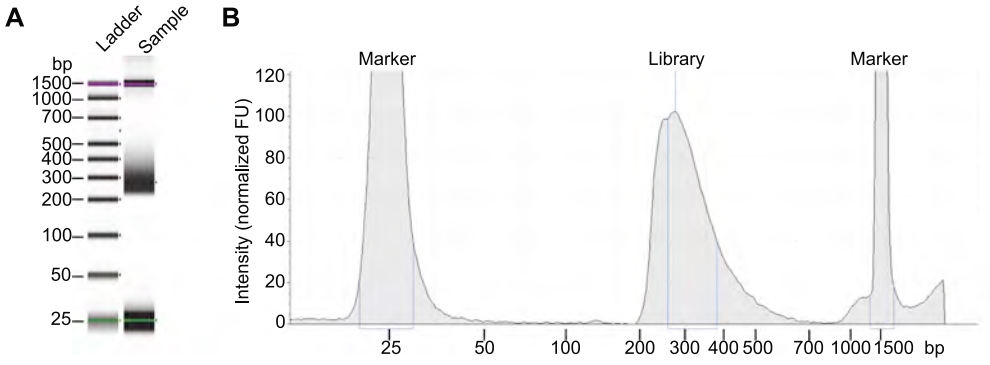


Fig. 3. Size distribution of libraries generated by the modified genome-wide DMS-MaPseq method. (A) Densitometry plot of a pooled DMS-MaPseq library sample run on a bioanalyzer shows a narrow smear DNA band between 200 and 400 bp. (B) Bioanalyzer results can also be visualized in an electropherogram that is plotted with fluorescent intensity (FU) on the Y axis and size (nt) on the X axis.

Illumina Truseq® Total RNA Prep Plant kit, which is gel-free. We modified experimental procedure of RT to synthesize first strand cDNA with TGIRT, instead of a regular reverse transcriptase. In addition, we minimized PCR cycle to 15 cycles for library construction to avoid otherwise potential PCR biases at a higher amplification number. To examine the library quality, the size distribution of the library was assessed by microfluidic capillary electrophoresis with the Agilent bioanalyzer. The result showed the length of the majority of cDNAs fell in the size range of 200–400 bp, validating the feasibility of the modified protocol (Fig. 3). It typically takes two days to construct the DMS-MaPseq library by using our protocol, and the starting RNA requires 1–2 μg.

Table 1
The numbers of reads in alignment.

TopHat v2.1.1

Sample Name	Clean reads	Mapped reads	Mapped/total clean reads	Uniquely mapped reads	Uniquely/total mapped reads
DMS treated Rep 1	43,778,079	41,019,748	93.70%	23,002,145	56.08%
DMS treated Rep 2	29,165,756	27,932,055	95.77%	8,937,169	32.00%
DMS treated Rep 3	36,839,131	34,308,549	93.13%	21,368,538	62.28%
Untreated	24,004,069	22,604,226	94.17%	21,115,688	93.41%

HISAT2 v2.1.0

Sample Name	Clean reads	Mapped reads	Mapped/total clean reads	Uniquely mapped reads	Uniquely/total mapped reads
DMS treated Rep 1	43,778,079	43,778,079	100.00%	20,339,324	46.46%
DMS treated Rep 2	29,165,756	29,165,756	100.00%	8,577,728	29.41%
DMS treated Rep 3	36,839,131	36,839,131	100.00%	18,500,069	50.22%
Untreated	24,004,069	24,004,069	100.00%	21,896,389	91.22%

3.3. Mapping to genome

Genome-wide DMS-MaPseq libraries were sequenced with three biological replicates for DMS treated adult plants and one replicate for untreated sample. After cleaning reads with sequencing quality, we obtained approximately 44, 29, 37 and 24 million reads for each replicate of DMS treated samples and untreated sample, respectively (Table 1). The resulting reads were initially aligned to the reference genome of Arabidopsis accession Columbia (Col-0, TAIR 10) using TopHat v2.1.1 with bowtie2, as has been widely used for DMS-MaPseq and SHAPE-MaPseq [13,19,24,25]. The setting of alignment allows 5 mismatches, total 7 nt gaps, 7 edit distances and 5 nt indel tolerance. More than 93% of reads were mapped (Table 1). Next, uniquely mapped reads were extracted with NH:I:1 tag. Notably, average half of the mapped reads were maintained in three replicates of DMS treated samples, whereas the majority of the mapped reads (93.41%) were maintained in the untreated sample (Table 1). These uniquely mapped reads represented 52.0%, 35.7%, 50.1% and 57.4% of the transcriptome in three replicates of DMS treated and untreated samples, respectively.

Next, we examined the multiple-mapped reads and observed that 91%, 97% and 88% of the reads in replicate #1, #2 and #3 were mapped at rRNA gene (rDNA) loci in DMS-treated samples, while only 12% was present in the untreated sample. The reason for this scenario is likely that DMS treatment causes insufficient rRNA depletion. In our study, rRNA is sequestered by biotin-labeled probes through complementary sequences, and subsequently removed by streptavidin beads. It appears that DMS treatment causes multiple modifications on rRNA, resulting in incomplete pairing of the probes with rRNA targets and insufficient rRNA depletion. Moreover, rDNA is located on the chromosomes in clusters of tandem repeats; insufficient rRNA depletion can further increase numbers of multiple-mapping reads.

HISAT2 v2.1.0 (https://ccb.jhu.edu/software/hisat2/index.shtml) is the successor of TopHat. We also performed the sequencing alignment using this new version of computational tool. In this case, alignment with HISAT2 appeared to obtain a similar amount of unique mapped reads compared to TopHat (Table 1). This result suggests that HISAT2 could also be used as an alternative approach for mapping of DMS-MaPseq sequence if one feels more comfortable with the new version of the tool.

3.4. High-quality DMS-MaPseq data

TGIRT decodes DMS lesions on RNA templates as mismatches on cDNA. To assess data quality, distribution of mutation types generated by TGIRT was analyzed. In both DMS treated and untreated samples, mismatches accounted for up to 96% of the total mutations, whereas insertions and deletions only accounted for less than 4% (Fig. 4A). These results indicate that TGIRT mainly produces mismatches at DMS modification sites during RT reactions using our modified protocol. This scenario is different from that of RT with SuperScript II (SSII) plus Mn^{2+}, which could also introduce a large portion of deletion mutations [10,13]. Therefore, only mismatches will be counted in DMS signal calculation in our method. To measure signal-to-noise ratio in our data, we examined the enrichment of mismatches on nucleotides. Compared to the untreated sample, the mismatch percentages of A and C, but not guanosine (G) and thymidine (T), are significantly increased in all three DMS treated replicates, indicating that our data have high signal-to-noise ratios (Fig. 4B). We also observed that the contribution of A on DMS-induced mismatches is slightly but consistently higher than that of C (Fig. 4B). This phenomenon is also in lines with in the original DMS-MaPseq data [13]. In addition, we found there is no strand bias of mismatches (Fig. 4C). Next, we assessed the reproducibility of our DMS-MaPseq data. Pearson's *r* and Gini index were used to describe the degree of similarity in DMS ratiometric signal pattern within a selected region between two replicates, where high *r* value or low Gini index means high similarity between two samples [5,13]. The custom script for Pearson's *r* value and Gini index calculation is shown in http://bioinfo.sibs.ac.cn/arabRSS/cal_gini_index_and_R_value.html. The data of three biological replicates for DMS treatment displayed high *r* value and small Gini index, indicating a strong reproducibility among replicates (Fig. 4D). These results demonstrated our modified DMS-MaPseq methods for plants produce high-quality and reproducible data.

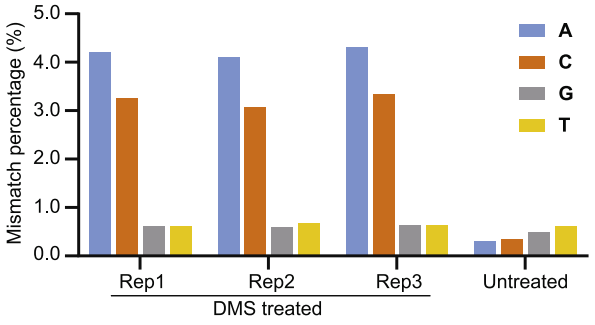
3.5. Sequencing depth and coverage for DMS-MaPseq in Arabidopsis

Sequencing depth and coverage are key considerations in NGS experiments [32]. To determine sequencing depth requirement for DMS-MaPseq in plants, we assessed the reproducibility of DMS signals between replicates with different coverage depths of mismatches. We observed that an average coverage depth of mismatches that exceeds 20x could greatly improve data reproducibility as the *r* value between replicates was greatly improved (Fig. 5A). The threshold of 20× mismatch coverage depth is also consistent with the previous result in yeast [13]. Next, we assessed the correlation between the coverages of the Arabidopsis reference genome and numbers of uniquely mapped reads. Arabidopsis has a transcriptome size of ~74 Mbp. We found that with 25 and 50 million uniquely mapped reads, only 1.5% and 2.5% of Arabidopsis genes could be recovered with more than 20× average mismatch coverage depth, respectively. When extrapolated to 100 million reads, coverage of only 4.4% of the reference genome at the minimum average depth can be achieved (Fig. 5B). This result is comparable with the observation from the genome-wide human DMS-MaPseq data [13], suggesting that DMS-MaPseq could typically detect the structural information of the abundant RNAs. For relatively low abundant RNA, intensive sequencing depth is required to obtain

A

Sample name	Percentage of total mutations		
	Mismatches	Deletions	Insertions
DMS treated Rep1	97.1%	2.8%	0.1%
DMS treated Rep2	96.9%	3.0%	0.1%
DMS treated Rep3	97.5%	2.4%	0.1%
Untreated	98.2%	1.2%	0.7%

B



C

Sample name	Fraction of total mismatches in each sample	
	Forward strand	Reverse strand
DMS treated Rep1	49.5%	50.5%
DMS treated Rep2	49.1%	50.9%
DMS treated Rep3	49.5%	50.5%
Untreated	48.2%	51.8%

D

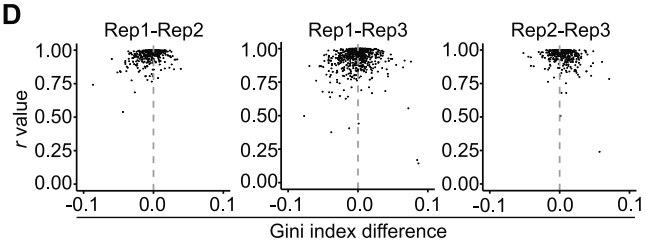
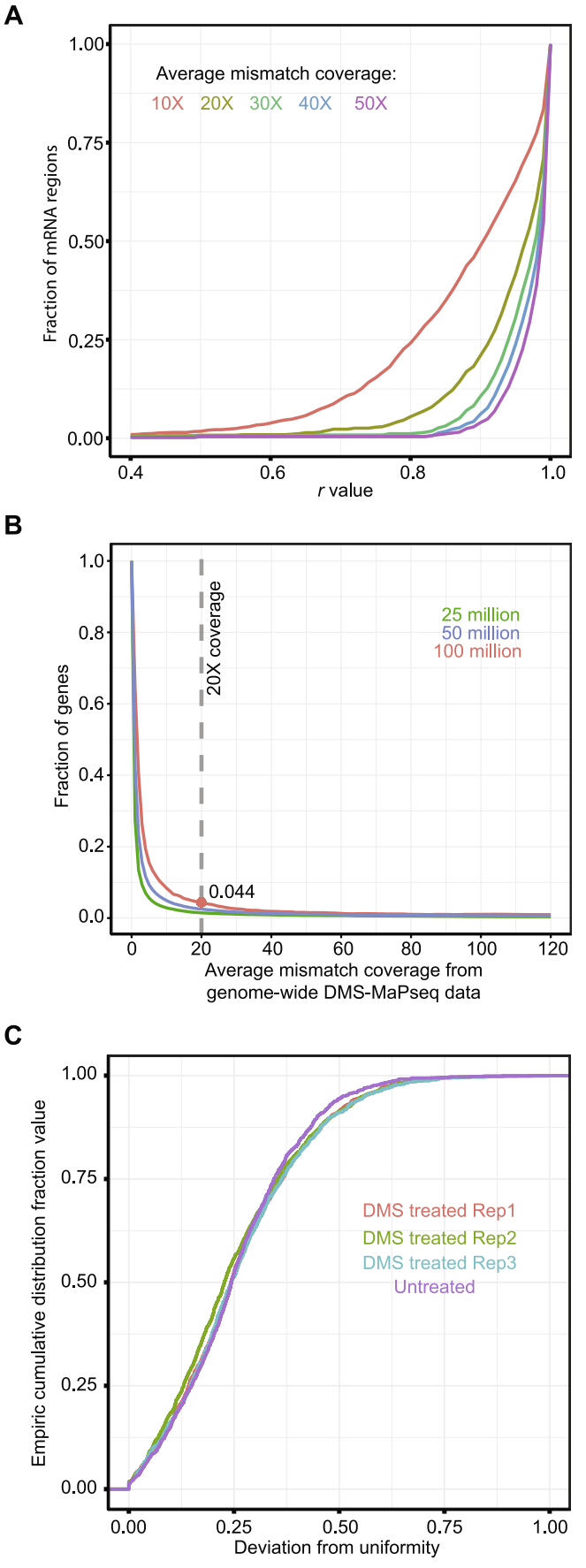


Fig. 4. The modified genome-wide DMS-MaPseq produces high-quality data for structural analysis. (A) Distribution of mutation types generated by RT with TGIRT in DMS treated and untreated samples. (B) Total mismatch percentage on each nucleotide in DMS treated and untreated samples. (C) Distribution of mismatches on forward and reverse strands. (D) Reproducibility between biological replicates for DMS treated were measured by Pearson's *r* value and Gini index.

sufficient coverage of 20× mismatch coverage. Such in-depth sequencing would cause a significant increase in sequencing cost, especially for plant species with large genome sizes. To address this issue, one way would be to, optimize or develop new approach(es) for rRNA depletion for DMS-MaPseq to decrease DMS treatment-elicited multiple mapping of sequencing reads. Alternatively, since chloroplast RNA accounts for a large proportion of total mRNA in plants, etiolated seedlings could be used for genome-wide RNA structure probing to minimize the chloroplast RNA contamination as shown before [6,29]. In addition, high-through sequencing cost has been dramatically reduced over the past decade [33], this trajectory would predict more cost-effective ways to obtain a substantial amount of reads to recover low-abundant RNA in the foreseeable future. Last but not least, in vivo secondary structure of low abundant transcripts, can be easily probed by target-specific DMS-MaPseq method [13,31].



(caption on next page)

Fig. 5. DMS-MaPseq requires high sequencing depth to detect low-abundant transcripts of Arabidopsis genome, and DMS treatment does not affect the uniformity of reads coverage. (A) Cumulative histogram of Pearson's *r* values for Arabidopsis RNA regions between DMS treated replicates with different average mismatch coverage depths. (B) Fraction of genes that pass the minimum average mismatch coverage threshold (20×) with varied sequencing depths. 0.015 and 0.025 are the fraction of genes achieving this threshold at 25 and 50 million uniquely mapped reads, respectively. (C) Uniformity of sequencing read distribution was validated on a subset of exons in the Arabidopsis genome that exceeded a minimal coverage threshold (> 20×). The coefficient of variation of exons in each gene was computed for a comparison of the read distribution among the samples. X-axis is the coefficient of variation, and Y-axis is the empirical cumulative distribution function (ECDF) of the coefficient of variation of exons from the exon collection. Lower coefficient of variation indicates a closer approximation to uniformity.

To assess the effect of DMS-stimulated mutations on genome-wide uniformity of reads coverage, we calculated the coefficient of variation of exons in each gene to compare the genome-wide uniformity of reads coverage among the untreated and treated samples. we did not observe clear difference in the genome-wide uniformity of reads coverage among samples, indicating that DMS treatment does not alter genome-wide uniformity of reads coverage (Fig. 5C).

It has been noted that random hexamers for RT priming might cause bias and uneven coverage along the transcriptome [34]. However, this might not be an issue for DMS-MaPseq method. First, high reproducible DMS signals require high sequencing depth that exceeds 20x mismatch coverage (Fig. 5A). With this depth, even uneven coverage does not affect the quality of DMS signal. Second, a read count reweighting scheme has been developed to mitigate the impact of the bias caused by random primers, in which the counts of each genomic position is re-weighted based on the first seven bases of each read [34]. For DMS-MaPseq method, the DMS signal is calculated as the ratio of mismatches to total reads, wherein the weight is neutralized during calculation. Thus, any potential bias that might result from by random primers should not affect the DMS signal. In fact, RT priming with random primers had been still widely used for genome-wide RNA secondary structure probing [6,25,29,30].

3.6. RNA secondary structure modeling

Based on DMS-MaPseq data, RNA structure can be modeled using RNApvmin and RNAfold in ViennaRNA webserver (http://rna.tbi.univie.ac.at/) [35,36]. The equivalent command line in our study is: RNApvmin -d2 -T 22 - -shapeConversion = C0.03 - -tauSigmaRatio = 1.0 - -minimizer = default - -initialVector = 0 - -samplesize = 1000 sequence1.shape < sequence1.fa > sequence1.sc 2 > sequence1.err; and RNAfold -d2 -T 22 - -shapeMethod = W - -shape = sequence1.sc < sequence1.fa > sequence1.fold. In this programming, a vector of pseudoenergy that minimized the discrepancies between the predicted and empirically-inferred pairing probabilities was computed using RNApvmin, then the resulting perturbation vector and the Washietl algorithm [37] were implemented in RNAfold to model the RNA structure. To benchmark our method on RNA structure prediction, we modeled the secondary structure of nuclear localized *U1* snRNA that with an established reference structure [38]. The predicted secondary structure of 4-way junction region of *U1-5* snRNA (AT1G08115) based on our DMS-MaPseq data is consistent with the known reference structure (Fig. 6), indicating that these constraints are feasible for RNA structure prediction.

To further investigate the feasibility of our DMS-MaPseq data, we selected some transcripts with high sequencing depth in our DMS-MaPseq data for structural profiling. It has been reported that tRNA-like RNA structure triggers long-distance mRNA transport in plants, although the underlying mechanism remains elusive [39]. Among top 10 high-coverage nuclear mRNAs in our data, *CHLOROPHYLL A/B*

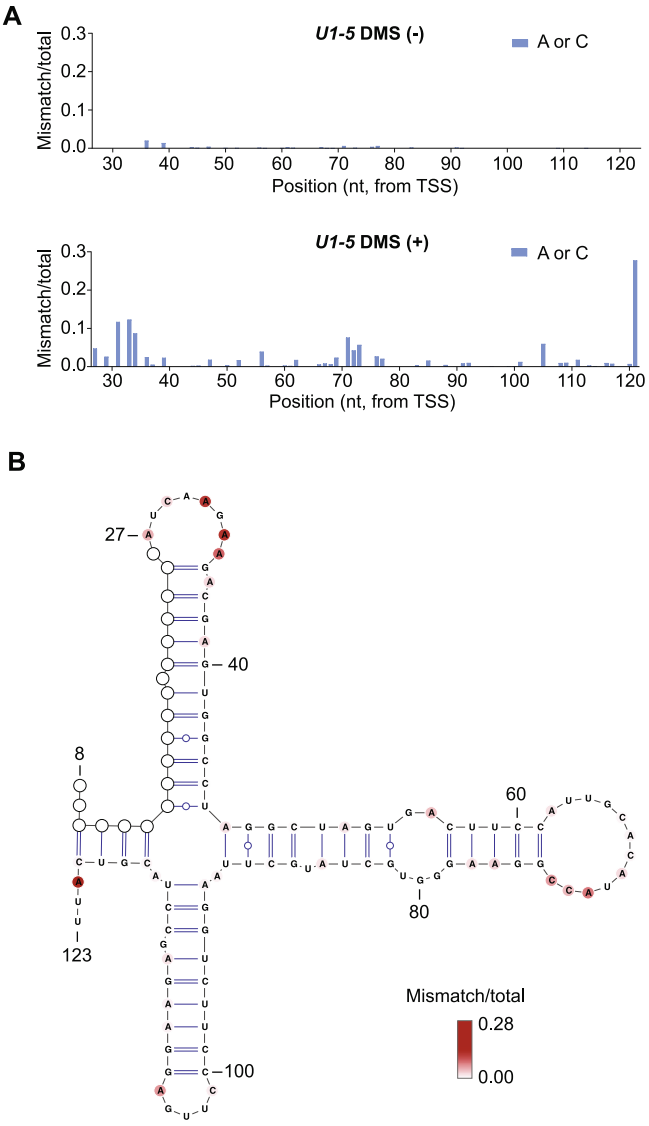


Fig. 6. Modeling secondary structure of *U1-5* snRNA. (A) Average ratiometric DMS signals of each A and C are plotted along *U1-5* snRNA sequence. Untreated and DMS treated samples are indicated by DMS (−) and DMS (+), respectively. TSS, Transcriptional start site, (B) The predicted secondary structure of 4-way junction region of *U1-5* snRNA, based on DMS-MaPseq data. Color-coded A and C residues display different ratiometric DMS signal. White cycles indicate nucleotides without read coverage.

BINDING PROTEIN 1 (*CAB1, AT1G29930.1*) and *LIGHT-HARVESTING COMPLEX GENE 3* (*Lhca3*, *AT1G61520.1*), both coding chloroplast-localized proteins, have been identified as long-distance mobile RNAs that bidirectionally moves between roots and shoots upon both normal and nutrient-limited conditions [40]. First, we measured the variability, by the Gini index, in ratiometric DMS signal of residues on a 1000-nt region of *CAB1* and *Lhca3,* which comprises full-length open reading frame (ORF) and adjacent untranslated regions (UTRs) (Fig. 7A). Relatively even DMS signal distribution results in low Gini index, indicating unstructured conformation, whereas uneven signal distribution results in high Gini index, indicating structured fold [5,13,26]. The Gini index of *CAB1* and *Lhca3* are relatively high, 0.453 for *CAB1* and 0.497 for *Lhca3*, suggesting that these transcripts are folded in nuclei.

Of note, there is a large outlier base with about 50% mismatch rate in *CAB1* mRNA (Fig. 7A), whereas the mismatch rate of this outlier base in untreated sample is less than 0.3% (Fig. S2A). This contrasted result indicated that the high DMS signal of the base in the DMS treated

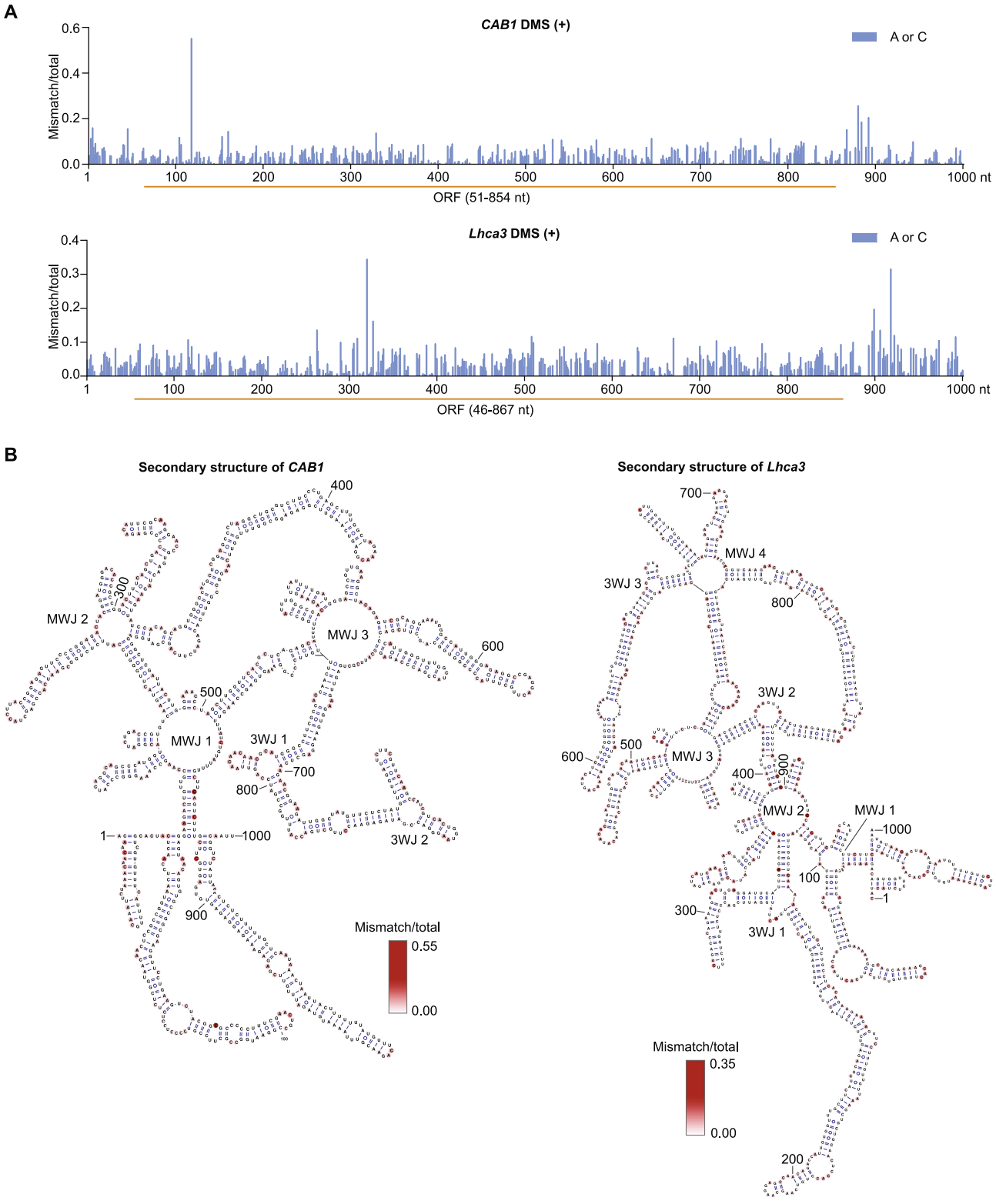


Fig. 7. Profiling secondary structure of *CAB1* and *Lhca3*. (A) Average ratiometric DMS signal of each A and C are plotted along *CAB1* and *Lhca3* sequences. DMS (+), DMS treated; ORF region is shown by an orange line. (B) The predicted secondary structure of *CAB1* and *Lhca3*, based on DMS-MaPseq data. The ratiometric DMS signal of A and C residues are color coded. 3WJ, three-way junction; MWJ, multiple-way junction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

samples does not result from endogenous modifications, nor the variation from insufficient coverages (Fig. S2). Since we used nuclear RNA as templates that are not protected and/or unwound by ribosomes in nuclei, the folding of nuclear RNA could be different or more structured than that of cytosol RNA, leading to more an uneven DMS signal pattern.

Finally, we used our DMS-MaPseq data to model the secondary structure of *CAB1* and *Lhca3* with RNApvmin and RNAfold in

ViennaRNA webserver (http://rna.tbi.univie.ac.at/) [35,36]. Interestingly, both *CAB1* and *Lhca3* mRNAs exhibit complicated structured conformation. *CAB1* contains 2 three-way junctions (3WJs), 3 multiple-way junctions (MWJs) and dozens of helixes (Fig. 7B). Similar with *CAB1*, *Lhca3* forms 3 3WJs, 4 MWJs and dozens of helixes. These cloverleaf-like RNA structures might be required for long-distance transport of *CAB1* and *Lhca3*. These examples suggest that our modified DMS-MaPseq can quantitatively analyze in vivo RNA structure at a single nucleotide resolution.

4. Conclusion

In this study, we modified the DMS-MaPseq method that is initially developed in yeast, human and *Drosophila* [13] for profiling genome-wide RNA secondary structure in plants. Our pilot study showed that the treatment conditions of 1% DMS and 20–30 min with vacuum or 2% DMS and 20 min without vacuum are feasible for plant materials. We have also optimized several important experimental protocols including the RT procedure, library construction, and simplified bioinformatics analysis. Our preliminary data showed that the protocol could yield high-quality RNA structure signal with high signal-to-noise ratio and reproducibility. We hoped that this powerful method will be useful to systematically investigate biological functions of RNA structure in plants.

Funding

This study was supported by National Science Foundation [grant number MCB-1716243] and Welch Foundation [grant number A-1973-20180324] to X. Z. T. W. was supported by the China Scholar Council Fellowship.

Acknowledgements

We thank Dr. Andrew Hillhouse for sequencing assistance.

Competing interests

The authors declare no competing interests

Appendix A. Supplies and equipment

Enzymes, kits and chemicals

TRIzol™ Reagent (Thermo Fisher, cat. No. 15596026)
TURBO™ DNase (Thermo Fisher, cat. No. AM2238)
TGIRT™-III (InGex, cat. No. TGIRT50)
SUPERase·In™ RNase Inhibitor (Thermo Fisher, cat. No. AM2694)
cOmplete™ EDTA-free Protease Inhibitor Cocktail (Roche, cat. No. 05056489001)
Miracloth (Millopore, cat. No. 475855-1R)
RNeasy® Mini Kit (QIAGEN, cat. No. 74104)
TruSeq® Stranded Total RNA Library Prep with Ribo-Zero Plant (Illumina, cat. No. 20020610)
TruSeq® RNA Single indexes Set A (Illumina, cat. No. 20020492)
TruSeq® RNA Single indexes Set B (Illumina, cat. No. 20020493)
NEBNext® Library Quant Kit for Illumina® (NEB, E7630S)
Agencourt® RNAClean® XP (Beckman, cat. No. A63987)
Agencourt® AMPure® XP (Beckman, cat. No. A63881)
DMS (Sigma, cat. No. D186309)
HEPES (Fisher, cat. No. BP310-1)
KCl (Sigma, cat. No. P3911)
$MgCl_2$ (Sigma, cat. No. M9272)
β-Mercaptoethanol (Sigma, cat. No. M6250)
EDTA (Sigma, cat. No. E5134)
Glycerol (Fisher, cat. No. BP229-4)
Sucrose (Fisher, cat. No. S5-3)
Triton® X-100 (Sigma, cat. No. T9284)
Formamide (Fisher, cat. No. BP228-100)
SDS (Sigma, cat. No. L3771)
Bromophenol Blue sodium salt (Alfa Aesar, cat. No. 32639)
Xylene Cyanol FF (Sigma, cat. No. X4126)
Agarose (Fisher, cat. No. BP1360-100)
EtBr (Sigma, cat. No. E1510)
UltraPure™ DNase/RNase-free distilled water (Thermo Fisher, cat. No. 10977-015)
DTT (Sigma, cat. No. D9779)
Phenol:chloroform:isoamyl alcohol (25:24:1) (Fisher, cat. No. BP17521-400)
Chloroform (Sigma, cat. No. C2432)
Isoamyl alcohol (Fisher, cat. No. A393-500)
Sodium acetate (Sigma, cat. No. S2889)
GlycoBlue™ Coprecipitant (Thermo Fisher, cat. No. AM9515)
200 proof pure ethanol (KOPTEC, cat. No. V1001)
5 X SuperScript® II first-strand buffer (Thermo Fisher, cat. No. 18064014)

Supplies and equipment

Desiccator (Scienceware, cat. No. 420270000)
Bench vacuum
Eppendorf centrifuge 5810 R
Eppendorf centrifuge 5424 R
Thermal cycler
NanoDrop® ND-1000 Spectrophotometer
Agilent 2200 TapeStation System
Mortar and pestle
14 ml polypropylene round-bottom tube (FALCON, cat. No. 352059)
RNase-free 1.5 ml Eppendorf tube
RNase-free tips

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ymeth.2018.11.018.

References

[1] C.K. Kwok, Y. Tang, S.M. Assmann, P.C. Bevilacqua, The RNA structurome: transcriptome-wide structure probing with next-generation sequencing, Trends Biochem. Sci. 40 (2015) 221–232.
[2] P.C. Bevilacqua, L.E. Ritchey, Z. Su, S.M. Assmann, Genome-wide analysis of RNA secondary structure, Annu. Rev. Genet. 50 (2016) 235–266.
[3] L.E. Vandivier, S.J. Anderson, S.W. Foley, B.D. Gregory, The conservation and function of RNA secondary structure in plants, Annu. Rev. Plant Biol. 67 (2016) 463–488.
[4] X. Yang, M. Yang, H. Deng, Y. Ding, New era of studying RNA secondary structure and its influence on gene regulation in plants, Front. Plant Sci. 9 (2018) 671.
[5] S. Rouskin, M. Zubradt, S. Washietl, M. Kellis, J.S. Weissman, Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo, Nature 505 (2014) 701–705.
[6] Y. Ding, Y. Tang, C.K. Kwok, Y. Zhang, P.C. Bevilacqua, S.M. Assmann, In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features, Nature 505 (2014) 696–700.
[7] J. Talkish, G. May, Y. Lin, J.L. Woolford Jr., C.J. McManus, Mod-seq: high-throughput sequencing for chemical probing of RNA structure, RNA 20 (2014) 713–720.
[8] N.A. Siegfried, S. Busan, G.M. Rice, J.A. Nelson, K.M. Weeks, RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP), Nat. Methods 11 (2014) 959–965.
[9] R.C. Spitale, R.A. Flynn, Q.C. Zhang, P. Crisalli, B. Lee, J.W. Jung, H.Y. Kuchelmeister, P.J. Batista, E.A. Torre, E.T. Kool, H.Y. Chang, Structural imprints in vivo decode RNA regulatory mechanisms, Nature 519 (2015) 486–490.
[10] M.J. Smola, G.M. Rice, S. Busan, N.A. Siegfried, K.M. Weeks, Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis, Nat. Protoc. 10 (2015) 1643–1669.
[11] Y. Ding, C.K. Kwok, Y. Tang, P.C. Bevilacqua, S.M. Assmann, Genome-wide profiling of in vivo RNA structure at single-nucleotide resolution using structure-seq,

Nat. Protoc. 10 (2015) 1050–1066.
[12] R.A. Flynn, Q.C. Zhang, R.C. Spitale, B. Lee, M.R. Mumbach, H.Y. Chang, Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE, Nat. Protoc. 11 (2016) 273–290.
[13] M. Zubradt, P. Gupta, S. Persad, A.M. Lambowitz, J.S. Weissman, S. Rouskin, DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo, Nat. Methods 14 (2017) 75–82.
[14] D. Incarnato, E. Morandi, F. Anselmi, L.M. Simon, G. Basile, S. Oliviero, In vivo probing of nascent RNA structures reveals principles of cotranscriptional folding, Nucleic Acids Res. 45 (2017) 9716–9725.
[15] M.J. Smola, K.M. Weeks, In-cell RNA structure probing with SHAPE-MaP, Nat. Protoc. 13 (2018) 1181–1195.
[16] M. Kubota, C. Tran, R.C. Spitale, Progress and challenges for chemical probing of RNA structure inside living cells, Nat. Chem. Biol. 11 (2015) 933–941.
[17] P. Tijerina, S. Mohr, R. Russell, DMS footprinting of structured RNAs and RNA-protein complexes, Nat. Protoc. 2 (2007) 2608–2623.
[18] C.K. Kwok, Y. Ding, Y. Tang, S.M. Assmann, P.C. Bevilacqua, Determination of in vivo RNA structure in low-abundance transcripts, Nat. Commun. 4 (2013) 2971.
[19] X. Wu, D.P. Bartel, Widespread influence of 3′-end structures on mammalian mRNA processing and stability, Cell 169 (2017) 905–917.
[20] S. Mohr, E. Ghanem, W. Smith, D. Sheeter, Y. Qin, O. King, D. Polioudakis, V.R. Iyer, S. Hunicke-Smith, S. Swamy, S. Kuersten, A.M. Lambowitz, Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing, RNA 19 (2013) 958–970.
[21] Z. Lu, Q.C. Zhang, B. Lee, R.A. Flynn, M.A. Smith, J.T. Robinson, C. Davidovich, A.R. Gooding, K.J. Goodrich, J.S. Mattick, J.P. Mesirov, T.R. Cech, H.Y. Chang, RNA duplex map in living cells reveals higher-order transcriptome structure, Cell 165 (2016) 1267–1279.
[22] E. Sharma, T. Sterne-Weiler, D. O'Hanlon, B.J. Blencowe, Global mapping of human RNA-RNA interactions, Mol. Cell 62 (2016) 618–626.
[23] J.D. Beaudoin, E.M. Novoa, C.E. Vejnar, V. Yartseva, C.M. Takacs, M. Kellis, A.J. Giraldez, Analyses of mRNA structure dynamics identify embryonic gene regulatory programs, Nat. Struct. Mol. Biol. 25 (2018) 677–686.
[24] U.P. Guenther, D.E. Weinberg, M.M. Zubradt, F.A. Tedeschi, B.N. Stawicki, L.L. Zagore, G.A. Brar, D.D. Licatalosi, D.P. Bartel, J.S. Weissman, E. Jankowsky, The helicase Ded1p controls use of near-cognate translation initiation codons in 5′ UTRs, Nature 559 (2018) 130–134.
[25] A.M. Mustoe, S. Busan, G.M. Rice, C.E. Hajdin, B.K. Peterson, V.M. Ruda, N. Kubica, R. Nutiu, J.L. Baryza, K.M. Weeks, Pervasive regulatory functions of mRNA structure revealed by high-resolution SHAPE probing, Cell 173 (2018) 181–195.
[26] Y. Zhang, D.H. Burkhardt, S. Rouskin, G.W. Li, J.S. Weissman, C.A. Gross, A stress response that monitors and regulates mRNA structure is central to cold shock adaptation, Mol. Cell 70 (2018) 274–286.
[27] E.M. Langdon, Y. Qiu, A.G. Niaki, G.A. McLaughlin, C.A. Weidmann, T.M. Gerbich, J.A. Smith, J.M. Crutchley, C.M. Termini, K.M. Weeks, S. Myong, A.S. Gladfelter, mRNA structure determines specificity of a polyQ-driven phase separation, Science 360 (2018) 922–927.
[28] D.H. Burkhardt, S. Rouskin, Y. Zhang, G.-W. Li, J.S. Weissman, C.A. Gross, Operon mRNAs are organized into ORF-centric structures that predict translation efficiency, eLife 6 (2017) e22037.
[29] H. Deng, J. Cheema, H. Zhang, H. Woolfenden, M. Norris, Z. Liu, Q. Liu, X. Yang, M. Yang, X. Deng, X. Cao, Y. Ding, Rice in vivo RNA structurome reveals rna secondary structure conservation and divergence in plants, Mol. Plant 11 (2018) 607–622.
[30] Z. Su, Y. Tang, L.E. Ritchey, D.C. Tack, M. Zhu, P.C. Bevilacqua, S.M. Assmann, Genome-wide RNA structurome reprogramming by acute heat shock globally regulates mRNA abundance, PNAS (2018) PMID: 30413617.
[31] Z. Wang, Z. Ma, C. Castillo-Gonzalez, D. Sun, Y. Li, B. Yu, B. Zhao, P. Li, X. Zhang, SWI2/SNF2 ATPase CHR2 remodels pri-miRNAs via Serrate to impede miRNA production, Nature 557 (2018) 516–521.
[32] D. Sims, I. Sudbery, N.E. Ilott, A. Heger, C.P. Ponting, Sequencing depth and coverage: key considerations in genomic analyses, Nat. Rev. Genet. 15 (2014) 121–132.
[33] J.A. Reuter, D.V. Spacek, M.P. Snyder, High-throughput sequencing technologies, Mol. Cell 58 (2015) 586–597.
[34] K.D. Hansen, S.E. Brenner, S. Dudoit, Biases in Illumina transcriptome sequencing caused by random hexamer priming, Nucleic Acids Res. 38 (2010) e131.
[35] R. Lorenz, S.H. Bernhart, C.H. Siederdissen, H. Tafer, C. Flamm, P.F. Stadler, I.L. Hofacker, ViennaRNA Package 2.0, Algorithms Mol. Biol. 6 (2011) 26.
[36] R. Lorenz, D. Luntzer, I.L. Hofacker, P.F. Stadler, M.T. Wolfinger, SHAPE directed RNA folding, Bioinformatics 32 (2016) 145–147.
[37] S. Washietl, I.L. Hofacker, P.F. Stadler, M. Kellis, RNA folding with soft constraints: reconciliation of probing data and thermodynamic secondary structure prediction, Nucleic Acids Res. 40 (2012) 4261–4272.
[38] D.A. Pomeranz Krummel, C. Oubridge, A.K. Leung, J. Li, K. Nagai, Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution, Nature 458 (2009) 475–480.
[39] W. Zhang, C.J. Thieme, G. Kollwig, F. Apelt, L. Yang, N. Winter, N. Andresen, D. Walther, F. Kragler, tRNA-related sequences trigger systemic mRNA transport in plants, Plant Cell 28 (2016) 1237–1249.
[40] C.J. Thieme, M. Rojas-Triana, E. Stecyk, C. Schudoma, W. Zhang, L. Yang, M. Minambres, D. Walther, W.X. Schulze, J. Paz-Ares, W.R. Scheible, F. Kragler, Endogenous Arabidopsis messenger RNAs transported to distant tissues, Nat. Plants 1 (2015) 15025.