



HAL
open science

High-Resolution Mapping of 3' Extremities of RNA Exosome Substrates by 3' RACE-Seq.

Hélène Scheer, Caroline de Almeida, Natalia Sikorska, Sandrine Koechler,
Dominique Gagliardi, Hélène Zuber

► **To cite this version:**

Hélène Scheer, Caroline de Almeida, Natalia Sikorska, Sandrine Koechler, Dominique Gagliardi, et al.. High-Resolution Mapping of 3' Extremities of RNA Exosome Substrates by 3' RACE-Seq.. *Methods in molecular biology* (Clifton, N.J.), 2062, 2020, 10.1007/978-1-4939-9822-7_8 . hal-02882621

HAL Id: hal-02882621

<https://hal.science/hal-02882621v1>

Submitted on 31 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

High-resolution mapping of 3' extremities of RNA exosome substrates by 3' RACE-seq

Hélène Scheer, Caroline De Almeida, Natalia Sikorska, Sandrine Koechler, Dominique Gagliardi*, and Hélène Zuber*

Institut de biologie moléculaire des plantes, Centre national de la recherche scientifique (CNRS), Université de Strasbourg, 12 rue du général Zimmer, 67000 Strasbourg, France.

* Authors for correspondence

Hélène Zuber

helene.zuber@ibmp-cnrs.unistra.fr

Dominique Gagliardi

dominique.gagliardi@ibmp-cnrs.unistra.fr

Running head

3' RACE-seq

Abstract

The main 3'-5' exoribonucleolytic activity of eukaryotic cells is provided by the RNA exosome. The exosome is constituted by a core complex of nine subunits (Exo9), which coordinates the recruitment and the activities of distinct types of co-factors. The RNA exosome co-factors confer distributive and processive 3'-5' exoribonucleolytic activities, endoribonucleolytic and RNA helicase activities. In addition, several RNA binding proteins and terminal nucleotidyltransferases also participate in the recognition of exosome RNA substrates.

To fully understand the biological roles of the exosome, the respective functions of its co-factors must be deciphered. This entails the high-resolution analysis of 3' extremities of degradation or processing intermediates in different mutant backgrounds or growth conditions. Here, we describe a detailed 3' RACE-seq procedure for targeted mapping of exosome substrate 3' ends. This procedure combines a 3' RACE protocol with Illumina sequencing to enable the high-resolution mapping of 3' extremities and the identification of untemplated nucleotides for selected RNA targets.

Keywords Exosome, rRNA maturation, rapid amplification of cDNA 3' end, 3' RACE-seq, 3' adapter ligation, Illumina sequencing, MiSeq, untemplated nucleotides

1. Introduction

The RNA exosome provides all eukaryotic cells with a 3'-5' exoribonucleolytic activity which plays a central role in the processing and the degradation of many nuclear and cytosolic

RNAs. Nine subunits compose the exosome core, which is also called Exo9. Exo9 is structurally related to bacterial polynucleotide phosphorylases (PNPases) and archaeal exosomes. These prokaryotic exoribonucleases are processive enzymes whose central channel accommodates three phosphorolytic active sites. By contrast, in mammals and yeast, Exo9 have lost the original phosphorolytic activity and the ribonucleolytic activity of the exosome relies on ribonucleases associated to Exo9 (*1–3*). In yeast, Rrp6 confers a distributive 3'-5' exoribonuclease activity while Dis3/Rrp44 provides both processive 3'-5' exoribonuclease and endonucleolytic activity (*1–3*). In addition to Rrp6, the human Exo9 associates with two Dis3 homologs, Dis3 and Dis3L, the latter lacking an endoribonucleolytic active site. In Arabidopsis, homologs of RRP6 and DIS3 also contribute to exosome activity (*4–6*). However the most striking difference with the mammalian and yeast Exo9 is that Arabidopsis Exo9 has retained a single site conferring a distributive and phosphorolytic activity (*6*).

A complex set of ribonucleolytic activities are therefore coordinated by Exo9 in eukaryotes. In addition, RNA helicases, various RNA binding proteins and terminal nucleotidyl transferases (TNTases) also assist the exosome in recognizing, degrading or maturing its RNA substrates. The respective functions of all these co-factors and the coordination of their associated activities must be determined to fully appreciate the biological functions of the RNA exosome.

One of the key aspects towards understanding the roles of each activities linked to exosome function is to analyze the degradation or trimming of exosome RNA substrates. One way to do that is to map 3' extremities of exosome RNA substrates at high density by determining their precise position at nucleotide level and by identifying eventual untemplated nucleotides added to processed 3' extremities. We present here a high throughput sequencing-based strategy called 3' RACE (3' Rapid Amplification of cDNA End)-seq. The 3' RACE-seq method combines a modified 3' RACE-PCR method for 3' end analysis of cDNA (*7, 8*) and

the Illumina sequencing technology. The classical 3' RACE-PCR is a low-throughput method that implies cloning PCR amplicons and Sanger sequencing of individual clones. By contrast, the 3' RACE-seq procedure allows for the simultaneous analysis of millions of amplicons for multiple samples. The 3' RACE-seq procedure is summarized in Fig. 1. It comprises the ligation of an adapter at the 3' end of each molecule in a total RNA sample and subsequent cDNA synthesis by using a reverse transcriptase (RT) primer complementary to the ligated 3' adapter. Importantly, the ligated 3' adapter, described in TAIL-seq protocol (9–12), contains a random region that allows for the removal of PCR duplicates during bioinformatics analysis and thus each final sequence corresponds to a unique RNA molecule. Chosen targets are then amplified using forward and reverse primers that bind to the target sequence and the 3' adapter, respectively, and that comprise the Illumina sequences required for flow cell hybridization and sequencing. Finally, amplicon libraries are sequenced using MiSeq paired-end sequencing for an average yield per run of 40 million of reads: 20 million of read 1 and 20 million of read 2. The use of barcoded Illumina adapters allows for sequencing in parallel more than 30 conditions or replicates in a single MiSeq run.

To illustrate the 3' RACE-seq procedure, we present the detailed protocol adapted for two classical RNA substrates of the exosome during ribosomal RNA (rRNA) maturation. In eukaryotes, three out of four ribosomal RNAs, the 18S, 5.8S and 25/28S rRNAs, are transcribed as a common polycistronic precursor. The 18S, 5.8S and 25/28S rRNAs are separated by internal transcribed spacers (ITS1 and 2) and flanked by two external transcribed spacers (5' and 3' ETS). The production of mature rRNAs requires endonucleolytic cleavages and exoribonucleolytic processing steps to remove internal and external transcribed spacers. Two of the archetypical RNA substrates of the exosome in eukaryotes are the 5' external transcribed spacer (5' ETS) of the 18S-5.8S-25/28S rRNA primary transcript and the 5.8S rRNA precursors. Using 3' RACE-seq, we have recently shown the complexity of

ribonucleolytic and tailing activities that contribute to these rRNA maturation steps in *Arabidopsis* (6). Of note, the 3' RACE-seq procedure can easily be adapted to any other RNA targets, such as poly(A) tailed transcripts or RISC-cleaved fragments, with slight modifications of the protocol.

2. Materials

2.1 RNA extraction

1. TRI Reagent® (Molecular Research Center).
2. Acid phenol (Biophenol water saturated, pH 4):chloroform:isoamyl alcohol solution (125:24:1).
3. Absolute ethanol.
4. 3 M sodium acetate, pH 5.2.
5. 20 mg/ml glycogen.
6. Refrigerated microcentrifuge reaching 16,000 g.
7. 75 % ethanol.
8. Nuclease-free water.
9. Ultraviolet (UV) spectrophotometer e.g. Thermo Scientific NanoDrop 2000.
10. Heating block that can heat to 65 °C.
11. 2 x RNA loading buffer: 95 % (v/v) formamide, 0.025 % (w/v) bromophenol blue, 0.025 % (w/v) xylene cyanol FF, 5 mM EDTA, 0.025 % (w/v) SDS, pH 8.5.
12. Agarose
13. 0.5 x TBE (10 x stock solution: 1 M Tris base, 1 M boric acid and 0.02 M EDTA).
14. Gel system for agarose electrophoresis (well combs, casting tray, gel box) and electrophoresis power supply.
15. 10 mg/ml ethidium bromide (EtBr)

16. (Optional) Agilent 2100 Bioanalyzer

2.2 3' adapter ligation

1. 3'-Adap oligonucleotide (Table 1).
2. Nuclease-free water.
3. Water bath or heating block for 37°C and 65 °C incubation.
4. 10x T4 RNA Ligase Reaction buffer (NEB): 10 mM MgCl₂, 50 mM Tris-HCl, 1 mM DTT, pH 7.5.
5. 10 000 U/ml T4 ssRNA Ligase 1 (NEB).

2.3 RNA separation by denaturing PAGE and isolation of RNA fragments from polyacrylamide gel

1. 2 x RNA loading buffer: 95 % (v/v) formamide, 0.025 % (w/v) bromophenol blue, 0.025 % (w/v) xylene cyanol FF, 5 mM EDTA, 0.025 % (w/v) SDS, pH 8.5.
2. 40 % acrylamide (19:1) solution.
3. 1 x TBE (10 x stock solution: 1 M Tris base, 1 M boric acid and 0.02 M EDTA).
4. Urea.
5. 10 % (w/v) ammonium persulfate solution (APS).
6. N,N,N',N'- Tetramethylethylenediamine (TEMED).
7. Gel system for PAGE (gel combs, gel cassettes and spacers) and electrophoresis power supply.
8. Water bath or heating block that can heat to 65 °C.
9. Syringe with a needle.
10. 10 mg/ml ethidium bromide (EtBr).
11. Scalpel.

12. Elution buffer: 500 mM ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA and 0.1 % (w/v) SDS.
13. Rotating wheel.
14. UV light Transilluminator.
15. Acid phenol (Biophenol water saturated, pH 4):chloroform:isoamyl alcohol solution (125:24:1).
16. Absolute ethanol.
17. 3 M sodium acetate, pH 5.2.
18. 20 mg/ml glycogen.
19. Refrigerated microcentrifuge reaching 16,000 g.
20. 75 % ethanol.
21. Nuclease-free water.
21. UV spectrophotometer *e.g.* Thermo Scientific NanoDrop 2000.
22. Dry ice.

2.4 cDNA synthesis

1. 3'-RT primer (Table 1).
2. 10 mM dNTP mix (dATP, dGTP, dCTP and dTTP, each at 10 mM).
3. Nuclease-free water.
4. PCR thermal cycler.
5. 0.2 ml strip PCR tubes.
6. 5 x SuperScript™ IV buffer (Invitrogen™).
7. 0.1 M DTT.
8. 40 U/μl RNaseOUT™ (Invitrogen™).
9. 200 U/μl SuperScript™ IV Reverse transcriptase (Invitrogen™).

2.5 PCR amplification and quality assessment by electrophoresis

1. Primers: forward PCR primer (target specific) and reverse PCR primer (TruSeq RNA PCR index primer, RPI, Table 1).
2. 5 U/ μ l DreamTaq DNA Polymerase (Thermo Fisher Scientific) supplied with 10 x DreamTaq buffer.
3. 10 mM dNTP mix (dATP, dGTP, dCTP and dTTP, each at 10 mM).
4. Nuclease-free water.
5. PCR thermal cycler.
6. 0.2 ml strip PCR tubes.
7. 6 x DNA loading buffer: 10 mM Tris-HCl (pH 7.6), 60 % (v/v) glycerol, 0.03 % (w/v) bromophenol blue, 0.03 % (w/v) xylene cyanol FF, 60mM EDTA.
8. 0.5 x TBE (10 x stock solution: 1 M Tris base, 1 M boric acid and 0.02 M EDTA).
9. Agarose
10. Gel system for agarose electrophoresis (well combs, casting tray, gel box) and electrophoresis power supply.
11. 10 mg/ml ethidium bromide (EtBr).

2.6 PCR product purification

1. AMPure XP beads (Agencourt).
2. Benchtop mini-centrifuge.
3. Magnetic stand compatible with 1.5 ml microtubes (e.g. DynaMag-2 InvitrogenTM).
4. 80 % ethanol.
5. Nuclease-free water.
6. Absolute ethanol.

7. Refrigerated microcentrifuge reaching 16,000 g.
8. 75 % ethanol.
9. 3 M sodium acetate, pH 5.2.
10. 20 mg/ml glycogen.
11. 21. UV spectrophotometer *e.g.* Thermo Scientific NanoDrop 2000.

2.7 Qubit and Bioanalyzer analysis of purified amplicons

1. Qubit fluorometric quantitation system (Invitrogen™).
2. Agilent 2100 Bioanalyzer.
3. DNA chip kit (*see Note 2*).
4. Microcentrifuge.

2.8 Preparing libraries for sequencing on MiSeq

1. Illumina MiSeq system.
2. 1.0 N NaOH.
3. PhiX control v3 library (Illumina, FC-110-3001).
4. HT1 (Hybridization Buffer provide by Illumina).
5. Benchtop microcentrifuge.

2.9 MiSeq run and analysis

1. MiSeq Reagent Kit v3 (Illumina, MS-102-3001) that contains:
 - Reagent Cartridge.
 - HT1 (Hybridization Buffer).
 - PR2 (Incorporation Buffer).
 - MiSeq Flow Cell.

3. Method

3.1 RNA extraction

1. Extract total RNA using TRI Reagent® (Molecular Research Center) according to the manufacturer's protocol.
2. (*Optional*). A second round of purification using acid phenol: chloroform: isoamyl alcohol and a subsequent RNA precipitation can be performed in order to remove residual contaminants:
 - a. Add 1 volume of acid phenol:chloroform:isoamyl alcohol solution (125:24:1) (*see Note 1*).
 - b. Vortex well and centrifuge 15 minutes at 16,000 g.
 - c. Transfer supernatant into new tube and precipitate RNA by adding 0.1 volume of 3 M sodium acetate pH 5.2, 0.5 µl of glycogen (20 mg/ml) and 2.5 volumes of absolute ethanol.
 - d. Mix by tube inversion.
 - e. Incubate at least 1 hour at -80 °C.
 - f. Centrifuge 30 minutes at 16,000 g (4 °C).
 - g. Discard supernatant.
 - h. Wash pellet with 75 % ethanol to remove residual salt.
 - i. Discard supernatant thoroughly and dissolve RNA pellet in 20 µl of nuclease-free water.
3. Measure the RNA quantity and purity of your samples with an UV spectrophotometer (e.g. Thermo Scientific NanoDrop 2000, *see Note 2*).
4. To assess the integrity of your total RNA preparation in a quick and cheap manner, you can check the profile(s) of your sample(s) on a 1 % agarose gel (*see Note 3*).

- a. Take a volume containing between 500 ng and 1 μg of RNA and add 1 volume of 2x RNA denaturing loading dye.
- b. Heat 5 minutes at 65 °C and chill on ice prior to loading into the wells of agarose gel.

Alternatively, RNA quality can be assessed on Agilent Bioanalyzer system.

3.2 3' adapter ligation

In order to analyze RNA 3' extremities, the 5'-riboadenylated DNA oligonucleotide (3'-Adap, Table 1) is ligated at the RNA 3' end. This primer is as described in (9) except that it is not biotinylated. The features of 3'-Adap are shown in Fig. 2.

1. Take 20 μg of total RNA for each sample.
2. Add 5 pmol of the 3'-Adap.
3. Add nuclease-free water to a final volume of 44 μl .
4. Denature sample for 3 minutes at 65°C.
5. Put on ice for at least 2 minutes.
6. Add 5 μl of 10 x T4 RNA Ligase Reaction buffer (NEB).
7. Add 1 μl of T4 ssRNA Ligase 1 (10 000 U/ml, NEB) (*see Note 4*).
8. Incubate 1 hour at 37 °C (*see Note 5*).

3.3 RNA separation by denaturing PAGE and isolation of RNA fragments from polyacrylamide gel.

Before proceeding to cDNA synthesis, the ligation reaction needs to be stopped and the ligation products purified from reagents and non-ligated adapter molecules. This can be achieved using RNA purification columns (*see Note 6*). Here, RNAs are separated by denaturing PAGE and RNA molecules of 100-400 nucleotides are eluted from the gel in order

to enrich for desired targets, *i.e.* 5.8S rRNA precursors and 5' ETS fragments in Arabidopsis (Fig. 3).

1. Cast a 6 % urea-polyacrylamide gel (6 % polyacrylamide, 7 M Urea, 1 x TBE), see example protocol in **(13)**.
2. Pre-run the gel in 1x TBE at 15 W for 15-20 minutes.
3. During the pre-run, add 1 volume of 2x RNA loading buffer to your samples. Heat the samples at 65 °C for 3 minute and chill on ice. Spin down briefly before loading.
4. Wash the residual urea from the wells using a syringe with a needle.
5. Load the samples on the gel and run at 15 W until the bromophenol blue tracking dye reaches three quarters of the gel.
6. Stain the gel in an ethidium bromide solution (0.5 µg/ml of EtBr in 1x TBE solution) for approximately 5 minutes and visualize using a UV transilluminator.
7. Excise RNA molecules of 100 to 400 nucleotides using a clean scalpel, put the gel slices in 1.5 ml tubes. Place tubes on dry ice to freeze the gel slices.
8. Fragment slices using a sterile 1 ml tip.
9. Elute RNA by adding 1:1 (v:w) volume of elution buffer to gel slices, for example 100 µl of elution buffer to 100 mg of gel, and incubate tubes overnight at 4 °C on a rotating wheel.
10. Centrifuge 10 minutes at 16,000 g to pellet the gel pieces.
11. Collect the supernatant and add 1 volume of acid phenol: chloroform: isoamyl alcohol solution (125:24:1).
12. Vortex well and centrifuge 15 minutes at 16,000 g.
13. Transfer supernatant into new tube and precipitate RNA by adding 0.1 volume of 3 M sodium acetate pH 5.2, 0.5 µl of glycogen (20 mg/ml) and 2.5 volumes of absolute ethanol (*see Note 7*).

14. Mix by tube inversion.
15. Incubate at least 1 hour at -80 °C.
16. Centrifuge 30 minutes at 16,000 g (4 °C).
17. Discard the supernatant.
18. Wash pellet with 75 % ethanol to remove residual salt.
19. Discard the supernatant thoroughly and dissolve the pellet in 20 µl of nuclease-free water.
20. Determine the RNA concentration and purity of your samples using UV spectrophotometer (*see Note 2*).

3.4 cDNA synthesis

cDNA synthesis is initiated using a primer complementary to the last 17 nucleotides of the adapter sequence ligated at the 3' end (3'-RT; Table 1, *see Note 8*). All steps of the cDNA synthesis are performed in a PCR thermal cycler using 0.2 ml strip PCR tubes.

1. Take 500 ng of adapter-ligated and size-selected RNAs and add 50 pmol of 3'-RT primer and 1 µl of 10 mM dNTP mix.
2. Add nuclease-free water to a final volume of 13 µl.
3. Denature 5 minutes at 65 °C.
4. Chill samples on ice for at least 2 minutes.
5. Add 7 µl of RT Mix comprising 4 µl of 5 x SuperScriptTM IV buffer, 1 µl of 0.1 M DTT, 1 µl of RNaseOUTTM (40 U/µl; InvitrogenTM) and 1 µl of SuperScriptTM IV (200 U/µl; InvitrogenTM).
6. Incubate 10 minutes at 50 °C (*see Note 9*).
7. Inactivate the reaction by incubating 10 minutes at 80 °C.

3.5 PCR amplification and quality assessment by electrophoresis

In order to prepare the libraries for MiSeq sequencing, the cDNA molecules of interest are amplified by PCR using forward PCR primers that comprise the Illumina P5 sequence and 21 nucleotides of the sequence of interest (here the 5' ETS downstream of the P processing site or 3' extended 5.8 S rRNA precursors) and a TruSeq RNA PCR index primer (RPI) complementary to the 3' end of the 3' adapter sequence that comprise the Illumina P7 sequence (*see* **Note 10**, Table 1). PCR reactions are performed in a PCR thermal cycler in 0.2 ml strip PCR tubes.

1. Set up PCR reaction by mixing the following components for each reaction:
 - a. 2.5 μ l of 10 x DreamTaq buffer (Thermo Fisher Scientific).
 - b. 0.5 μ l of 10 mM dNTP .
 - c. 0.5 μ l of 10 μ M forward primer (Table 1).
 - d. 0.5 μ l of 10 μ M reverse primer (Table 1).
 - e. 0.25 μ l DreamTaq Polymerase (5 U/ μ l; Thermo Fisher Scientific).
 - f. Add nuclease-free water to a final volume of 25 μ l.
2. Distribute PCR mix in strip PCR tubes and add 1 μ l of template cDNA.
3. Spin down and place reaction in thermal cycler.
4. Run PCR reaction with the following settings:
 - a. Initial denaturation step: 2 min at 94 °C.
 - b. 30 cycles composed of:
 - i. Denaturation step: 30 sec at 94 °C.
 - ii. Hybridization step: 30 sec at 55 °C.
 - iii. Elongation step: 30 sec at 72 °C.
 - c. Final extension: 2 min at 72 °C.

5. Visualize PCR products by loading a 3 μ l aliquot with DNA loading dye on a 2 % agarose gel (0.5 x TBE) (*see Note 11*).

3.6 PCR product purification

After amplification, the PCR products are purified using AMPure XP beads (Agencourt). This system uses magnetic beads that can bind PCR amplicons of at least 100 bp, thereby purifying amplicons from nucleotides, primer dimers, salts or other reagents. The protocol has been adapted from the manufacturer's protocol. AMPure XP beads purification is performed in 1.5 ml microtubes that are compatible with the magnetic stand.

1. Warm the AMPure XP beads to room temperature for at least 10 minutes and shake the Agencourt AMPure XP bottle before pipetting to resuspend magnetic particles.
2. Transfer each PCR reaction to individual 1.5 ml tubes.
3. For each tube, add 1 volume of beads to 1 volume of PCR reaction and mix well by pipetting or gentle vortexing (*see Note 12*).
4. Incubate the mixture 5 minutes at room temperature.
5. Transfer tubes to a magnetic stand.

For the next steps (5-10), tubes are kept on the magnetic stand.

6. Let sit for about 5 minutes or until solution appears clear.
7. Carefully discard the supernatant without disturbing the beads.
8. Keep the tubes on the magnetic stand and wash beads carefully with 200 μ l of 80 % ethanol.
9. Incubate 1 minute and carefully remove the ethanol.
10. Repeat the washing step.
11. Air-dry beads for a maximum of 3 minutes (*see Note 13*).

12. Add 100 μ l of nuclease-free water to beads, remove tubes from the magnetic stand and mix gently by pipetting.
13. Incubate 5 minutes at room temperature.
14. Put tubes back on stand and let sit for about 5 minutes or until solution is clear.
15. Transfer the eluate to a new, clean tube by paying attention not to take beads (*see Note 14*).
16. Perform a second elution with 100 μ l of nuclease-free water.
17. Precipitate purified amplicons with 5 volumes of absolute ethanol, 0.1 volume of sodium acetate 3 M, pH 5.2 and 0.5 μ l of glycogen (20 mg/ml).
18. Mix by tube inversion.
19. Incubate at least 1 hour at -80 °C.
20. Centrifuge 30 minutes at 16,000 g (4 °C).
21. Discard the supernatant.
22. Wash pellet with 75 % ethanol to remove residual salt.
23. Discard the supernatant thoroughly and dissolve pellet in 11 μ l of nuclease-free water.
24. Measure the DNA quantity and purity of your samples using UV spectrophotometer (*see Note 2*).

3.7 Qubit and Bioanalyzer analysis of purified amplicons.

Before Illumina sequencing, we quantify amplicon libraries by a fluorometric method, *i.e.* Qubit fluorometric quantitation system. We also control the size and the quality of our amplicon profiles after AMPure XP beads purification using Agilent 2100 Bioanalyzer, notably to check the complete removal of primer dimers (*see Note 15*).

1. Determine Qubit concentration of each library (Qubit fluorometric quantitation system; Invitrogen™, see manufacturer's protocol (*14*)) (*see Note 16*).

2. Check library profiles using Agilent 2100 Bioanalyzer according to manufacturer's instructions (**15**). Choose the reagent kit according to the range of concentration of your samples (*see Note 17*).
3. Use Qubit concentration and Bioanalyzer size estimation to calculate the molarity of each sample library (*see Note 18*).

3.8 Preparing libraries for sequencing on MiSeq

Here, we prepare libraries for sequencing with v3 MiSeq chemistry (*see Note 19*). You need to prepare a final amplicon library according to the depth wanted for each sample and to denature the library. We usually allocate from 0.5 to 3 % of the flow cell per condition/genotype and target. 10 to 20 % of PhiX control v3 library (Illumina) are also included to compensate for the low-diversity of the samples (*see Note 20, 21*).

1. Prepare at least 5 μ l of a final 4 nM amplicon library with all individual sample libraries to be sequenced.
2. Combine 5 μ l of 4 nM final amplicon library and 5 μ l of a fresh 0.2 N NaOH dilution.
3. Vortex briefly and centrifuge 1 minute at 280 g.
4. Incubate for 5 minutes at room temperature.
5. Stop denaturation reaction by adding 990 μ l pre-chilled HT1 to the 10 μ l denatured library. This results in a 20 pM denatured library.
6. Dilute the 20 pM amplicon library to 15 pM by adding 150 μ l pre-chilled HT1 to 450 μ l of the 20 pM denatured amplicon library.
7. Mix by inversion and quickly centrifuge the resulting 15 pM amplicon.
8. Final 15 pM library should be kept on ice.
9. Repeat step 2 to 8 with the 4 nM PhiX control library as described above for amplicon library to get a 15 pM PhiX library.

10. For MiSeq sequencing with 15 % of PhiX, combine 90 μ l of 15 pM denature PhiX control library and 510 μ l of 15 pM denature amplicon library (*see Note 20, 21*). Keep tubes on ice until loading on the reagent cartridge.

3.9 MiSeq run and analysis

Here we use a MiSeq Reagent kit v3 150 cycles. The final library is paired-end sequenced with a 76 x 76 bp cycle setting. Cycle setting may be adjusted according to the type of analyzed RNA target (*see Note 22*). Read 1 and read 2 will be used during bioinformatics analysis for RNA target identification and 3' end analysis, respectively.

1. Thaw Reagent Cartridge and mix according to Illumina manufacturer's instructions.
2. In the Illumina Experiment Manager (IEM) software, create a custom library prep kit as indicated in the IEM software guide (**16**). Take the "TruSeq Small RNA.txt" template as model (model with RPI barcodes) and change the setting section to allow for paired-end sequencing (*see Note 23*).
3. Use the IEM software to create sample sheet. Select "MiSeq", "other" and "fastq only" in the instrument, category and application sections, respectively. For workflow parameters, select as option the new custom library prep kit and set cycle setting as 76 x 76 bp (*see Note 22*).
4. Fill the sample sheet wizard as indicated in the IEM software guide.
5. Start MiSeq Control Software and follow steps indicated by the software to start MiSeq sequencing.
6. When asked by the MiSeq Control Software, load your sample (combination of PhiX and sample libraries as prepared in subsection 3.8) in the reservoir labeled "Load Samples" of the reagent cartridge.
7. Use the Sequence Analysis Viewer (SAV) to monitor sequencing during run.

8. When sequencing is finished, check quality control metrics using SAV and control quality of read 1 and read 2 fastq files using the quality control tool FASTQC (*17*).
9. Finally process fastq files using the pipeline available in (*6*). An overview of the bioinformatic workflow is shown in Fig. 4.

4. Notes

1. During phenol extraction of nucleic acid molecules, the partition between aqueous and organic phase is pH-dependent. At acidic pH conditions, RNA molecules are highly soluble and retained in the aqueous phase, while DNA molecules are retained in the organic phase and interphase. Acid phenol is thereby used for the isolation of RNA molecules, whereas DNA isolation is best performed with phenol buffer-saturated to $\text{pH} > 7.4$.
2. NanoDrop spectrophotometer measures the sample absorbance across a wide spectrum that spans UV and visible light. Nucleotides, RNA and DNA, have a absorbance peak at 260 nm. By contrast, proteins have a peak of absorbance at 280 nm, while other usual RNA contaminants, such as carbohydrates, EDTA and phenol have a absorbance maximum at 230 nm or less. 260/280 and 260/230 ratios can thus be used to assess RNA purity. Values around 2.0 are usually considered as acceptable for 260/280 and 260/230 ratios (*18*).
3. The assessment of RNA integrity by electrophoresis on agarose gel and ethidium bromide staining is a basic and cheap technique that gives a first indication of the quality of your RNA preparation. Sharp bands corresponding to rRNAs should be visible on the gel. Partially degraded RNA will appear as smeared bands.
4. T4 RNA ligase 1 catalyzes the ligation of 5'-phosphoryl terminated DNA or RNA to 3'-hydroxyl terminated single strand DNA or RNA. Here, we use a pre-adenylated adapter (3'-Adap) containing a 5',5'-adenyl pyrophosphoryl moiety,

which can be directly ligated to the RNA without the addition of ATP during the ligation reaction (Fig. 2). This strategy prevents that endogenous RNA with 5' phosphate extremities compete with the adapter for 3' ligation and ensures that only the pre-adenylated adapter is ligated to the 3' hydroxylated end of endogenous transcripts.

5. We usually use a water bath for the incubation at 37 °C during the ligation step. Water bath provide a better contact surface area for heat transfer as compared to dry bath, allowing for better reproducibility.
6. Separation and purification of ligated RNA on PAGE is not required for all applications. Here, the molecules of interest are small RNA fragments (< 400nt). Consequently, we purified ligated RNA on PAGE to remove larger RNAs and to enrich for desired targets before proceeding to cDNA synthesis. Alternatively, to purify the ligated RNAs from reagents and non-ligated adapter, you can perform fast purification of nucleic acids using RNA purification kits.
7. GlycoBlueTM (15 mg/ml, Thermo Fisher Scientific) can be used instead of glycogen to increase pellet visibility and is recommended if you are working with low amounts of RNA.
8. During the setting up of 3' RACE-seq procedure, we detected in our first amplicon libraries a strong proportion of sequences that did not contain the delimiter sequence (Fig. 2, 4) and thus did not correspond to real RNA 3' ends. These artefacts were likely caused by non-specific binding of the primer used in the reverse transcription reaction. Consequently, to specifically amplify cDNA 3' ends, we designed a shortened RT oligonucleotide (3'-RT, Table 1) that is complementary to the last 17 nt of the 3' ligated adapter and lacks five nucleotides

contained by both 3' ligated adapter and reverse PCR primers. Thus, cDNA resulting from non-specific reverse transcription cannot be amplified.

9. The reverse transcriptase SuperScript™ IV (Invitrogen™) is very effective and robust, allowing for efficient cDNA synthesis in only 10 minutes. If you use other reverse transcriptase, adapt the incubation time according to the manufacturer's protocol.
10. RPI primers used for cDNA amplification contain index sequences, also called barcodes, which allow for the sequencing of a large number of samples in a single run, *i.e.* multiplexing sequencing. Individual index sequences can be assigned to each genotype and/or conditions and then be used to distinguish and sort samples during data analysis. The use of individual index sequences for each different target, here 5.8S and P-P1 fragments, is not necessary as target sequences can be distinguished during bioinformatic analysis.
11. Separation of smaller DNA molecules and fragments is improved with TBE buffer, whereas TAE buffer is well suited for larger fragments (**19**).
12. The concentration of PEG and NaCl in the AMPure XP bead solution is crucial for size-selection of the DNA fragments that are purified. The size of purified fragments is determined by the ratio of beads/sample: the lower is the chosen ratio, the larger are the eluted fragments. To adapt the ratio according to the size of your amplicon library, refer to (**20**).
13. Take care to not over dry beads. This would significantly decrease the elution efficiency.
14. The carry-over of magnetic beads results in an additional peak in Bioanalyzer electropherograms and could lead to inaccurate estimation of the library size. Trace amounts of beads may also affect the performance of Illumina sequencing.

15. As we usually analyze a large number of samples, replicates and genotypes, we use to pool PCR products for each analyzed target prior to Qubit and Bioanalyzer analysis. We first adjust each sample to the same concentration, verify the dilution by checking the profiles on a 2 % agarose gel (at least 75 ng DNA for ethidium bromide staining), and then pool the samples.
16. The Invitrogen™ Qubit is a fluorometric quantitation system that allows for a more specific and sensitive quantification of RNA and DNA than using the NanoDrop spectrophotometer. Select the Qubit assay kit according to your sample. The dsDNA HS Assay kit is well suited for the quantification of the prepared libraries and is designed to measure sample concentration from 10 pg/μL to 100 ng/μL.
17. DNA kits for Bioanalyzer analysis may be chosen according to the size of the amplification product and the range of concentration of the sample. DNA 12000, DNA 7500 and DNA 1000 kits allow for the analysis of dsDNA fragments from 100 to 12000 bp, 100 to 7500 bp and 25 to 1000, respectively. All of them offer a 0.5 to 50 ng/μl sensitivity (21). For samples with a low concentration, DNA High Sensitivity kit allows for the analysis of 50 to 7000 bp fragments and offers a 5 to 500 pg/μl sensitivity (22).
18. The molarity in nM of amplicon library is calculated using the following formula $([\text{ng}/\mu\text{l conc.}] \times 10^6) / ([\text{bp length}] \times 607,4 + 157,9)$. See Illumina manufacturer's protocol for more information.
19. Here, we sequenced amplicon libraries using MiSeq Reagent Kit v3 (150 cycles). Reagent kits with v2 chemistry are also available for MiSeq. v2 chemistry is cheaper than v3 but enables smaller depth compared to v3: for single-end run, up to 15 and 25 millions of output reads are obtained for v2 and v3, respectively.

Final concentration of denatured library need to be adjusted according to the selected MiSeq chemistry. v2 and v3 chemistry support a maximum of 10 pM and 20 pM concentration, respectively.

20. To compensate for the low-diversity of amplicon libraries, PhiX Control v3 Library should be sequenced alongside samples. Illumina recommends spiking-in a minimum of 5% of PhiX control library. This percentage may be adjusted according to experiments and may be increased if the sample library clusters more efficiently than the PhiX library. When analyzing rRNA maturation intermediates, we usually spike-in from 10 to 20 % of PhiX control v3 library.
21. The procedure detailed here can be extended to other types of RNA targets, such as poly(A) tailed transcripts. Illumina sequencing of poly(A) stretches requires spiking-in a particular high amount of PhiX control library, at least 20% of the flow cell. Indeed, sequencing a highly diversified library alongside samples is necessary to counteract the strong negative impact of the base composition bias toward A of poly(A) tails on sequencing quality.
22. Amplicon libraries are paired-end sequenced: read 1 is use to identify target, whereas read 2 is used to determine RNA 3' end position and to identify added untemplate nucleotides (Fig. 4). Reagents provided in MiSeq 150-Cycle kit are sufficient to perform 152 sequencing cycles. The number of cycles for read 2 sequencing can be adjusted according to the type of analyzed RNA target and the expected length of 3' untemplated tails. A cycle setting of 76 x 76 nt enables the analysis of untemplated tails up to 56 nt (76 nt - 15 nt of the random sequence - 5 nt of the delimiter sequence). For longer 3' tail, such as mRNA poly(A) tail, cycle setting can be desynchronized and cycle number for read 1 sequencing can be reduced in favor of read 2 sequencing. For example in (6), we sequenced in

parallel amplicon libraries for rRNA intermediates and poly(A) tailed mRNA and we thus set cycle setting as 41 x 111 nt, 41 nt being sufficient for transcript identification.

23. In this protocol, we use the TruSeq RNA PCR index primers, classically used for the preparation of TruSeq Small RNA library. However, default options for TruSeq Small sequencing in the IEM software does not allow for paired-end sequencing. Therefore, a custom template, based on the “TruSeq Small RNA.txt” template, needs to be created.

5. References

1. Januszyn K, Lima CD (2014) The eukaryotic RNA exosome. *Curr Opin Struct Biol* 24:132–140 . doi: 10.1016/J.SBI.2014.01.011
2. Schneider C, Tollervey D (2014) Looking into the barrel of the RNA exosome. *Nat Struct Mol Biol* 21:17–18 . doi: 10.1038/nsmb.2750
3. Zinder JC, Lima CD (2017) Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes Dev* 31:88–100 . doi: 10.1101/gad.294769.116
4. Kumakura N, Otsuki H, Tsuzuki M, et al (2013) Arabidopsis AtRRP44A Is the Functional Homolog of Rrp44/Dis3, an Exosome Component, Is Essential for Viability and Is Required for RNA Processing and Degradation. *PLoS One* 8:e79219 . doi: 10.1371/journal.pone.0079219
5. Lange H, Holec S, Cognat V, et al (2008) Degradation of a polyadenylated rRNA maturation by-product involves one of the three RRP6-like proteins in Arabidopsis thaliana. *Mol Cell Biol* 28:3038–44 . doi: 10.1128/MCB.02064-07
6. Sikorska N, Zuber H, Gobert A, et al (2017) RNA degradation by the plant RNA

- exosome involves both phosphorolytic and hydrolytic activities. *Nat Commun* 8:2162 .
doi: 10.1038/s41467-017-02066-2
7. Sement FM, Ferrier E, Zuber H, et al (2013) Uridylation prevents 3' trimming of oligoadenylated mRNAs. *Nucleic Acids Res* 41:7115–7127 . doi: 10.1093/nar/gkt465
 8. Sement FM, Gagliardi D (2014) Detection of uridylated mRNAs. *Methods Mol Biol* 1125:43–51 . doi: 10.1007/978-1-62703-971-0_4
 9. Chang H, Lim J, Ha M, Kim VN (2014) TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. *Mol Cell* 53:1044–1052 . doi: 10.1016/j.molcel.2014.02.007
 10. Lim J, Ha M, Chang H, et al (2014) Uridylation by TUT4 and TUT7 marks mRNA for degradation. *Cell* 159:1365–1376 . doi: 10.1016/j.cell.2014.10.055
 11. Zuber H, Scheer H, Ferrier E, et al (2016) Uridylation and PABP cooperate to repair mRNA deadenylated ends in Arabidopsis. *Cell Rep* 14:2707–2717 . doi: 10.1016/j.celrep.2016.02.060
 12. Morgan M, Much C, DiGiacomo M, et al (2017) MRNA 3' uridylation and poly(A) tail length sculpt the mammalian maternal transcriptome. *Nature* 548:347–351 . doi: 10.1038/nature23318
 13. Summer H, Grämer R, Dröge P (2009) Denaturing Urea Polyacrylamide Gel Electrophoresis (Urea PAGE). *J Vis Exp* 3–5 . doi: 10.3791/1485
 14. Thermo Fisher Scientific (2017) Qubit™ 4 Fluorometer guide
 15. Agilent (2001) Agilent 2100 Bioanalyzer user guide
 16. Illumina (2018) Illumina Experiment Manager User Guide.
<https://support.illumina.com/downloads/illumina-experiment-manager-user-guide-15031335.html>. Accessed 19 Jan 2018
 17. (2018) Babraham Bioinformatics - FastQC A Quality Control tool for High

- Throughput Sequence Data.
<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Accessed 19 Jan 2018
18. Desjardins P, Conklin D (2010) NanoDrop microvolume quantitation of nucleic acids. *J Vis Exp*. doi: 10.3791/2565
 19. Sanderson BA, Araki N, Lilley JL, et al (2014) Modification of gel architecture and TBE/TAE buffer composition to minimize heating during agarose gel electrophoresis. *Anal Biochem* 454:44–52 . doi: 10.1016/j.ab.2014.03.003
 20. Bronner IF, Quail MA, Turner DJ, Swerdlow H (2009) Improved Protocols for Illumina Sequencing. In: *Current Protocols in Human Genetics*. p 2
 21. Agilent (2018) DNA Analysis Kits & Reagents - Details & Specifications.
<https://www.genomics.agilent.com/article.jsp?crumbAction=push&pageId=1636>.
Accessed 19 Jan 2018
 22. Agilent (2018) High Sensitivity DNA Analysis Kits - Details & Specifications.
<https://www.genomics.agilent.com/article.jsp?crumbAction=push&pageId=1300002>.
Accessed 19 Jan 2018
 23. Illumina (2018) Illumina Adapter Sequences Document.
<https://support.illumina.com/downloads/illumina-customer-sequence-letter.html>.
Accessed 26 Jan 2018

Acknowledgments

This work was supported by the Centre National de la Recherche Scientifique (CNRS, France) and research grants from the French National Research Agency as part of the ‘Investments for the Future’ program in the frame of LABEX ANR-10-LABX-0036_NETRINA and ANR-15-CE12-0008-01 to D.G, and in the frame of the IdEx Unistra to H.Z.

Figure Captions

Fig. 1 Flowchart steps for 3' RACE-seq procedure. After total RNA purification, the 3' hydroxyl (3' OH) end of each RNA molecule is ligated to the 3'-Adap adapter (see Fig. 2, Subheading 3.2 in the text). To remove large ribosomal RNAs and non-ligated 3' adapters, ligated RNAs are purified using RNA purification columns or separated on an acrylamide gel when size-selection is possible (Subheading 3.3 in the text, *see Note 6*). cDNA synthesis is then initiated using the 3'-RT primer complementary to the 3'-Adap adapter (Subheading 3.4 in the text, Table 1). To specifically analyze 3' regions of targets of interest, cDNA are PCR amplified using a forward primer that binds specifically to the regions of interest and a reverse primer that is complementary to the ligated adapter (Subheading 3.5 in the text, Table 1). Forward and reverse primers contain P5/Rd1 SP and P7/Rd2 SP Illumina sequences, respectively. P5 and P7 sequences is used for the hybridization to the flow cell. Rd1 SP and Rd2 SP correspond to the binding site of read 1 and read 2 sequencing primer. Reverse primer also contains an index sequence, which allows for multiplexed sequencing. In order to remove primer-dimers, salts and other PCR reagents, PCR products are then purified using AMPure XP beads (Subheading 3.6 in the text). Amplicon libraries are quantified and analyzed with Bioanalyzer for quality assessment (Subheading 3.7 in the text). Libraries are paired-end sequenced on the Miseq Illumina system. Read 1 allows for the identification of the target and read 2 for the identification of the RNA 3' end.

Fig. 2 Sequence details for the 3'-Adap adapter. 3'-Adap is a pre-adenylated oligonucleotide containing a 5',5'-adenyl pyrophosphoryl moiety (*see Note 4*). The delimiter sequence allows for the demarcation between the sequence corresponding to the 3' end of the RNA and the adapter sequence. This sequence is also used during bioinformatic analysis to discriminate

read 2 containing adapter sequence from other sequences that could arise from artifacts of reverse transcription priming (Fig. 4). The 15 random bases (15N) allows for deduplicating and therefore for eliminating PCR duplicates. The 3'-RT primer complementary region is used as template during reverse-transcription reaction. Finally, the 3'-dideoxynucleotide prevents the 5' adenylated oligo from self-ligation.

Fig. 3 Scheme of Arabidopsis 5' ETS maturation by-products and 5.8S precursors analyzed by 3' RACE-seq. 3' RACE-seq procedure described in this chapter was used to map at high density the 3' extremities of P-P1 fragments and of 5.8S precursors. P-P1 fragments result from endonucleolytic cleavage of the 5' ETS at P and P' sites followed by 3'-5' exonucleolytic degradation of the P-P' fragment by the RNA exosome. 5.8S precursors result from endonucleolytic cleavage at the 5' end of 5.8S and at the C2 site. This 5.8S-C2 fragment is then further processed by 3'-5' exoribonucleolytic activity, including by the RNA exosome. The shortest 5.8S rRNA precursor is extended by 11/12 nt in Arabidopsis. The 5' ETS-fw primer was used to analyze 3' extremities of P-P1 fragments. The 5.8S-fw primer was used to map 3' extremities of 5.8S rRNA precursors.

Fig. 4 Schematic overview of the bioinformatic pipeline for 3' RACE-seq analysis of 5.8S and P-P1 fragments. Color code as in Fig. 1. Scripts are available in (6).

Table Captions

Table 1 Oligonucleotides used in 3' RACE-seq procedure to analyze Arabidopsis 5' ETS P-P1 intermediates and 5.8S precursors.

All oligonucleotides are listed in the 5' to 3' orientation. 3'-Adap adapter contains two modifications: 5rApp = 5', 5'-adenyl pyrophosphoryl moiety, 3ddC=3'-dideoxy-C. 3'-Adap

should be HPLC purified in an RNase-free environment. For 5' ETS-fw and 5.8S-fw primers, bolded nucleotides correspond to the target specific sequence, while the rest of the sequence is used for hybridization to the Illumina flow cell and for sequencing. Red bold nucleotides in Illumina RPI PCR primer correspond to the index sequence (for further detail see Illumina manufacturer's instruction for TruSeq Small RNA RPI primers (23)).

Table 1

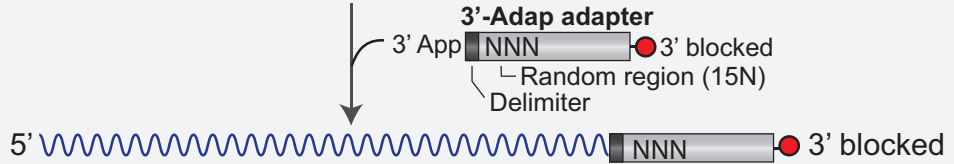
Oligonucleotide name	5'-3' sequence
3'-Adap	/5rApp/CTGACNNNNNNNNNNNNNNNTGGAATTCTCGGGTGCCAAGGC/3ddC/
3'-RT	GCCTTGGCACCCGAGAA
5' ETS-fw	AATGATACGGCGACCACCGAGATCTACACGTTTCAGAGTTCTACAGTCCGACGAT CATCTCGCGCTTGTACGGCTTTG
5.8S-fw	AATGATACGGCGACCACCGAGATCTACACGTTTCAGAGTTCTACAGTCCGACGAT CTCTGCCTGGGTGTCACAAATC
Illumina RPI	CAAGCAGAAGACGGCATAACGAGAT XXXXXX GTGACTGGAGTTCCTTGGCACCC GAGAATTCCA

Mortar and pestle + Leaf + TRI Reagent®

Total RNA

5' ~~~~~ 3'OH

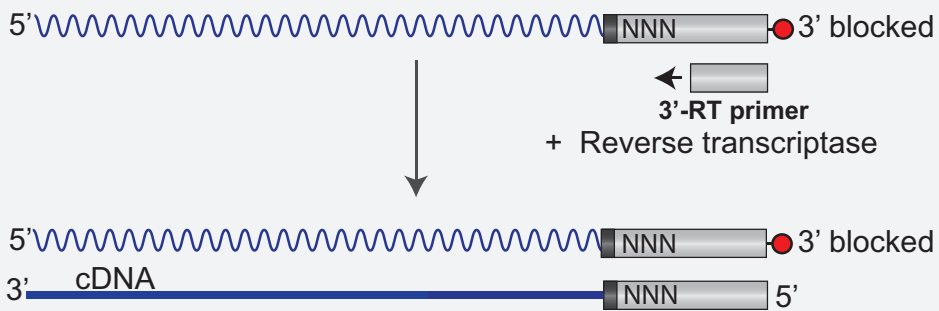
3' adapter ligation



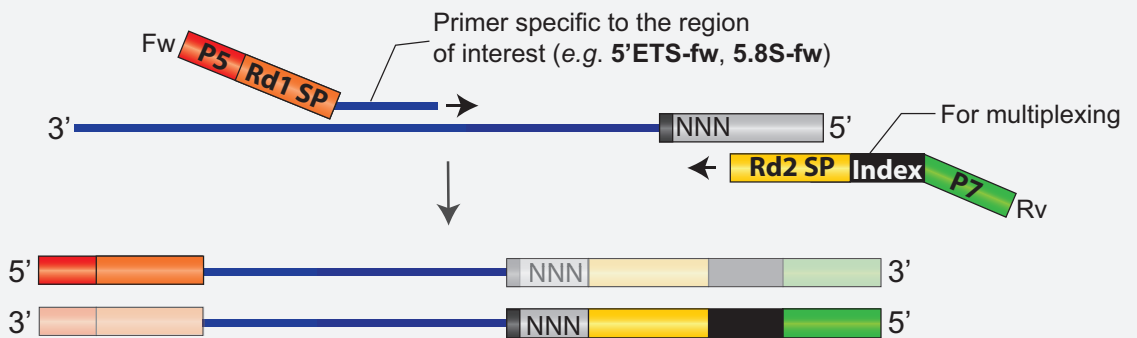
Purification from ligation reagents and non-ligated adapter molecules

RNA purification on columns
or
Separation on acrylamide gel and size selection

Reverse-transcription



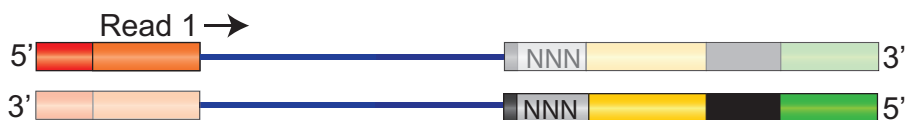
PCR amplification of target 3' region



PCR product purification using AMPure beads XP

Quantification & quality assessment of amplicon libraries

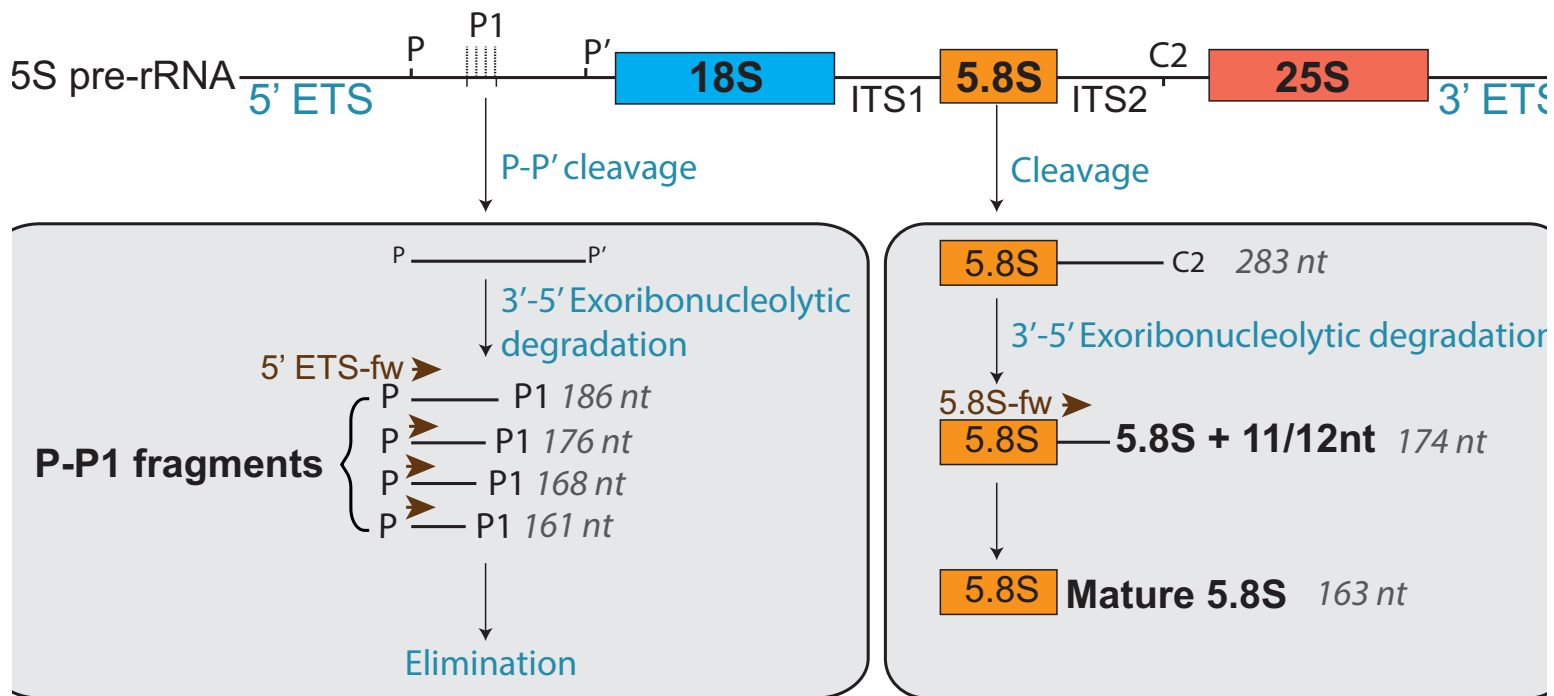
Illumina sequencing with MiSeq

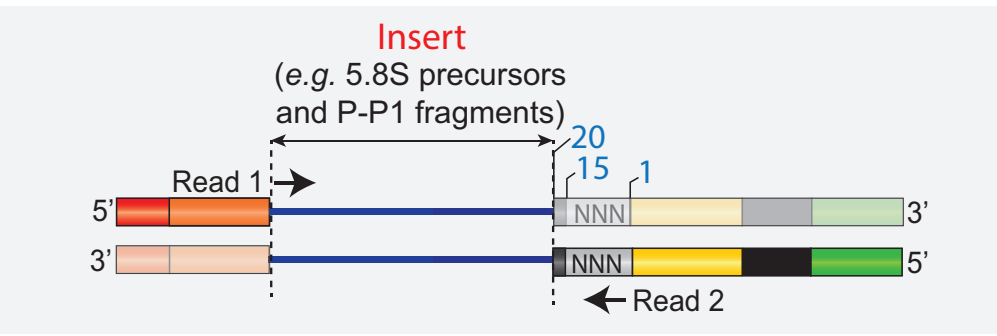


5',5'-adenyl
cyrophosphoryl
moiety



3'-dideoxy
nucleotide





Raw data

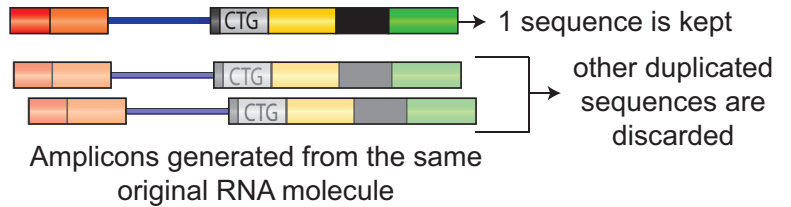
Read1.fastq
5' end of the **insert**

Read2.fastq

- Nucleotides **1** to **15**: random region (15N) of the adapter
- Nucleotides **16** to **20**: delimiter sequence
- From nucleotides **21** to the end: 3' end of the **insert**

DEDUPLICATION

Sequences with identical nucleotides in read 1 and the same degenerate base region in read 2 are removed.



TARGET RESEARCH

Identification of read 1 that correspond to the region of interest (e.g. 5.8S, P-P1 fragments).

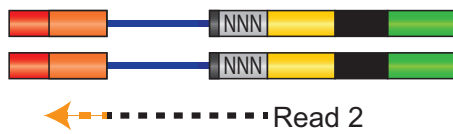
DELIMITER RESEARCH

Identification of read 2 containing delimiter sequence.

Read 2 without delimiter sequence are **discarded**

Read 2 TRIMMING

If the insert is short, read 2 can run into the 5' PCR primer sequence. In this case the unwanted nucleotides are trimmed from read 2.



3' END MAPPING AND ANALYSIS OF UNTEMPLATED TAIL

Read 2 is mapped to the RNA reference sequence to identify the 3' end position of 5.8S and P-P1 fragments. Potential untemplated nucleotides are analyzed.

