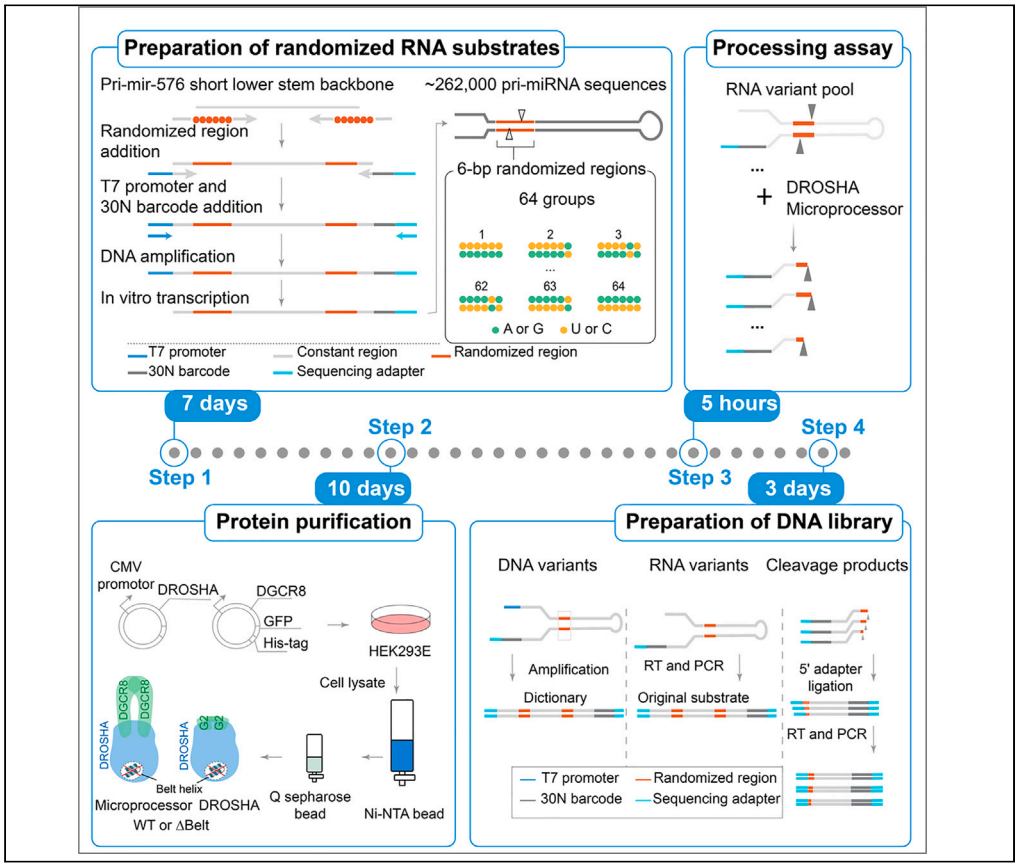
STAR Protocols

CellPress
OPEN ACCESS

Protocol

High-throughput protocol for studying pri-miRNA processing using randomized sequences



The Microprocessor complex is crucial in microRNA (miRNA) biogenesis, as it processes primary miRNAs (pri-miRNAs) into precursor miRNAs. Here, we present a high-throughput, radioisotope-free protocol for studying pri-miRNA processing using randomized sequences. We describe steps for randomized substrate preparation, protein purification, processing assays, and DNA library construction for sequencing. This technique explores pri-miRNA processing, uncovers key RNA elements, and illuminates gene expression regulation. However, its efficacy may be constrained by data analysis complexity and the requirement for specialized equipment.

Publisher's note: Undertaking any experimental protocol requires adherence to local institutional guidelines for laboratory safety and ethics.

Thi Nhu-Y Le, Tuan Anh Nguyen

tuananh@ust.hk

Highlights

Synthesis and assembly of randomized pri-miRNA sequences

Microprocessor expression and purification in human cells

Processing of randomized pri-miRNAs by purified Microprocessor

Construct and sequence DNA libraries from the processing assays

Le & Nguyen, STAR Protocols 5, 102782
March 15, 2024
https://doi.org/10.1016/j.xpro.2023.102782

STAR Protocols

CellPress
OPEN ACCESS

Protocol

High-throughput protocol for studying pri-miRNA processing using randomized sequences

Thi Nhu-Y Le[1,2] and Tuan Anh Nguyen[1,3,*]

[1]Division of Life Science, The Hong Kong University of Science & Technology, Hong Kong, China

[2]Technical contact

[3]Lead contact

*Correspondence: tuananh@ust.hk
https://doi.org/10.1016/j.xpro.2023.102782

SUMMARY

The Microprocessor complex is crucial in microRNA (miRNA) biogenesis, as it processes primary miRNAs (pri-miRNAs) into precursor miRNAs. Here, we present a high-throughput, radioisotope-free protocol for studying pri-miRNA processing using randomized sequences. We describe steps for randomized substrate preparation, protein purification, processing assays, and DNA library construction for sequencing. This technique explores pri-miRNA processing, uncovers key RNA elements, and illuminates gene expression regulation. However, its efficacy may be constrained by data analysis complexity and the requirement for specialized equipment.

For complete details on the use and execution of this protocol, please refer to Nguyen et al. (2023).[1]

BEFORE YOU BEGIN

The human Microprocessor complex, consisting of the RNase III enzyme DROSHA and its cofactor, DGCR8 (DiGeorge syndrome critical region 8), plays a crucial role in the biogenesis of microRNAs (miRNAs) by cleaving primary miRNAs (pri-miRNAs) in the nucleus.[2,3] This cleavage produces miRNA precursors, also known as pre-miRNAs. These pre-miRNAs are then transported to the cytoplasm for additional processing. Here, another RNase III enzyme, known as DICER, is responsible for generating the miRNA duplex.[2,4] The Argonaute protein then selects one strand from the duplex, which could either be the 5p or 3p strand, to form a mature miRNA. The mature miRNAs are essential for post-transcriptional gene regulation, impacting processes such as development, differentiation, and stress responses.

To cleave pri-miRNAs accurately and efficiently, the human Microprocessor depends on various *cis* RNA elements found within pri-miRNAs. These elements include the basal junction, UG, UGUG, mGHG, and midBWM motifs.[5–15] Similarly, the *Caenorhabditis elegans* (*C. elegans)* Microprocessor relies on the basal junction of pri-miRNAs to determine its cleavage sites.[16] In addition to the RNA elements, the human Microprocessor may require cofactors, such as hemin, SRSF3 and SRSF7, to enhance its interactions with these elements. For instance, hemin strengthens the human Microprocessor's interaction with the UGU motif while SRSF3 and SRSF7 fortify its interaction with the CNNC and CRC motifs.[5,17–19] These interactions promote accurate and efficient processing of pri-miRNAs, ultimately contributing to proper miRNA biogenesis and gene regulation.

The protocol below outlines the specific steps for conducting an *in vitro* high-throughput cleavage assays for the human Microprocessor using a randomized pri-miRNA based on one backbone, pri-mir-576. Pri-mir-576 has been observed to display unique processing characteristics that are not

canonical to most miRNAs. These distinctive features make it an excellent model for studying the intricacy of Microprocessor's noncanonical processing mechanisms, which, in turn, could highlight insights into the variability and complexity of miRNA biogenesis.[1] This protocol, while focusing on pri-mir-576, is also adaptable for different pri-miRNA backbones and Microprocessors from different species. This adaptability allows for the exploration of diverse miRNA backbones and the potential discovery of even more unique processing mechanisms.

This protocol offers a comprehensive, non-radioisotope method for investigating pri-miRNA processing by the Microprocessor complex using the *in vitro* high-throughput approach. It encompasses designing and preparing pri-miRNA substrates, conducting cleavage assays, and creating sequencing libraries for processed RNA fragments and input substrates. The objective is to map processing sites and calculate processing efficiency, providing valuable insights into pri-miRNA processing dynamics and gene expression regulation. Before initiating the protocol, researchers should design randomized pri-miRNA sequences, obtain essential reagents such as the Microprocessor complex and buffers, and prepare materials for DNA sequencing library construction. Ensuring access to a next-generation DNA sequencer is also vital. Lastly, it is important for researchers to prepare the experimental workspace by organizing reagents, setting up equipment, and maintaining a sterile and orderly environment to minimize contamination and maximize efficiency.

Institutional permissions

It is important to remind readers that they will need to acquire permissions from the relevant institutions. Ensure that you follow the proper safety protocols in line with the regulations of your university and institute regarding chemical and biological safety. These safety protocols should be obtained from your university or institution.

Preparation of randomized RNA substrates

Timing: 7 days

1. Design randomized pri-miRNA sequences.
   a. Determine the desired sequence length and complexity.
   b. Ensure sequences have appropriate features for studying cleavage dynamics.
2. Generate RNA substrates.
   a. Synthesize a double-stranded DNA (dsDNA) template containing a randomized region by 1st PCR (Ran-dsDNA: Randomized dsDNA).
   b. Add the barcode sequence and T7 promoter to the resulting Ran-dsDNA by 2nd PCR (BC-dsDNA: Barcode dsDNA).
   c. Dilute the BC-dsDNA and amplify it by 3rd PCR for generating DNA templates for *in vitro* transcription (IVT-dsDNA).
   d. Perform *in vitro* transcription for IVT-dsDNA to synthesize RNA substrates.
   e. Purify and pool synthesized RNA substrates.
   f. Quantify RNAs by NanoDrop 2000/2000c Spectrophotometer and analyze them on urea-PAGE.

Protein purification

Timing: 10 days

3. Express recombinant proteins using the human cell system.
   a. Construct DROSHA (WT or mutant) and DGCR8 in human expression plasmids.
   b. Co-transfect DROSHA and DGCR8-expressing plasmids into HEK293E cells.

c. Culture the transfected cells for a few days to express recombinant proteins. The use of a bacterial expression system is not recommended as soluble DROSHA is not successfully expressed in *E. coli*.

4. Purify recombinant proteins.
   a. Prepare cell lysates and perform protein purification using affinity and ion-exchange columns.
   b. Verify protein purity and concentration by Bradford assays and SDS-PAGE.

Preparation of processing assays

Timing: 5 h

5. Set up cleavage reactions.
   a. Mix randomized RNA substrates with a recombinant Microprocessor complex.
   b. Add appropriate buffers and cofactors for optimal cleavage conditions.
6. Incubate reactions.
   a. Allow cleavage reactions to proceed for a suitable time.
   b. Terminate reactions after optimal cleavage is achieved.
   c. Analyze the cleavage assays by urea-PAGE.

Preparation of DNA library construction

Timing: 3 days

7. Convert cleaved fragments and RNA control to cDNA.
   a. Purify the cleaved RNA from urea-PAGE.
   b. Perform adapter ligation for cleaved fragments and RNA substrates.
   c. Perform reverse transcription (RT) using appropriate primers to get cDNA products.
8. Prepare sequencing libraries.
   a. Perform PCR amplification.
   b. Purify and size-select final libraries.
   c. Confirm the quality of the libraries by bioanalyzer and qPCR.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Dulbecco's modified Eagle's medium (DMEM)	Gibco	Cat# 17192939
Fetal bovine serum (FBS)	Gibco	Cat# 16000044
Penicillin-Streptomycin	Gibco	Cat# 15140-122
Sodium bicarbonate	Sigma-Aldrich	Cat# 144-55-8
G418 disulfate salt	Sigma-Aldrich	Cat# 108321422
Linear polyethylenimine (PEI)	Sigma-Aldrich	Cat# 764965
Pierce protease inhibitor tablets	Thermo Fisher Scientific	Cat# A32963
β-Mercaptoethanol	Sigma-Aldrich	Cat# 60-24-2
1,4-Dithiothreitol (DTT)	Sigma-Aldrich	Cat# 3483-12-3
SYBR Green II RNA gel stain	Thermo Fisher Scientific	Cat# S7568
Ribonuclease H (RNase H)	Thermo Fisher Scientific	Cat# 18021071
SuperScript IV reverse transcriptase	Thermo Fisher Scientific	Cat# 18090010
MEGAscript T7 Transcription Kit	Invitrogen	Cat# AM1334
Phusion Hot Start II high-fidelity DNA polymerase	Thermo Fisher Scientific	Cat# 00882319
SUPERase In RNase inhibitor	Invitrogen	Cat# AM2698
HisPur Ni-NTA resin	Thermo Fisher Scientific	Cat# 88223
UNOsphere Q anion exchange media	Bio-Rad	Cat# 1560105

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
UNOsphere S cation exchange media	Bio-Rad	Cat# 1560113
Proteinase K	Thermo Fisher Scientific	Cat# EO0491
T4 RNA ligase I	NEB	Cat# M0204L
30% Acrylamide/bis-acrylamide solution, 29:1	Bio-Rad	Cat# 1610156
Ammonium persulfate (APS)	Sigma-Aldrich	Cat# 7727540
N,N,N′,N′-tetramethylethylenediamine (TEMED)	Sigma-Aldrich	Cat# 110-18-9
UltraPure TBE buffer, 10X	Invitrogen	Cat# 15581028
UltraPure 1 M Tris-HCl buffer, pH 7.5	Invitrogen	Cat# 15567027
0.5 M EDTA pH 8.0,	Invitrogen	Cat# AM9260G
1 M $MgCl_2$	Invitrogen	Cat# AM9530G
UltraPure urea	Invitrogen	Cat# 15505050
Critical commercial assays		
Qubit dsDNA HS Assay Kit	Invitrogen	Cat# Q32854
Other		
ProFlex PCR system	Thermo Fisher Scientific	Cat# 4484073
Eppendorf ThermoMixer	Thermo Fisher Scientific	Cat# 2231001005
PowerPac basic power supply	Bio-Rad	Cat# 1645050
Centrifuge	Eppendorf	5424R
Sonicator	Qsonica	Q125
High-speed refrigerated centrifuge	Hitachi	CR22N
NGC chromatography system	Bio-Rad	NGC Quest Plus
Qubit 4 fluorometer	Invitrogen	Cat# Q33238
Gel Doc XR+ gel documentation system	Bio-Rad	Cat# 1708195

MATERIALS AND EQUIPMENT

HEK293E culture medium

Reagent	Final concentration	Amount
Dulbecco's Modified Eagle Medium (DMEM)	13.37 *g*/L	13.37 *g*
Sodium bicarbonate	3.7 *g* /L	3.7 *g*
Fetal Bovine Serum (FBS)	5%	50 mL
Penicillin Streptomycin	1%	10 mL
ddH_2O	N/A	Up to 1 L
Total	**N/A**	**1 L**

Filter using 0.22 μm filter in the tissue culture hood and then store at 4°C for up to 1 month.

T500 stock buffer

Reagent	Final concentration	Amount
1 M Tris-HCl Buffer, pH 7.5	20 mM	20 mL
5 M NaCl	500 mM	100 mL
β-mercaptoethanol	4 mM	280 μL
ddH_2O	N/A	Up to 1 L
Total	**N/A**	**1 L**

Store at 4°C for up to 1 month.

T2000 stock buffer

Reagent	Final concentration	Amount
1 M Tris-HCl Buffer, pH 7.5	20 mM	20 mL
5 M NaCl	2,000 mM	400 mL

(*Continued on next page*)

Continued

Reagent	Final concentration	Amount
β-mercaptoethanol	4 mM	280 μL
ddH_2O	N/A	Up to 1 L
Total	**N/A**	**1 L**

Store at 4°C for up to 1 month.

T0 stock buffer

Reagent	Final concentration	Amount
1 M Tris-HCl Buffer, pH 7.5	20 mM	20 mL
β-mercaptoethanol	4 mM	280 μL
ddH_2O	N/A	Up to 1 L
Total	**N/A**	**1 L**

Store at 4°C for up to 1 month.

T100 stock buffer

Reagent	Final concentration	Amount
1 M Tris-HCl Buffer, pH 7.5	20 mM	20 mL
5 M NaCl	100 mM	20 mL
ddH_2O	N/A	Up to 1 L
Total	**N/A**	**1 L**

Store at 4°C for up to 1 month.

T150 stock buffer

Reagent	Final concentration	Amount
1 M Tris-HCl Buffer, pH 7.5	20 mM	20 mL
5 M NaCl	150 mM	30 mL
ddH_2O	N/A	Up to 1 L
Total	**N/A**	**1 L**

Store at 4°C for up to 1 month.

Protein stock buffer

Reagent	Final concentration	Amount
1 M Tris-HCl Buffer, pH 7.5	20 mM	20 mL
5 M NaCl	500 mM	100 mL
1 M 1,4-Dithiothreitol (DTT)	2 mM	2 mL
Glycerol	10%	100 mL
ddH_2O	N/A	Up to 1 L
Total	**N/A**	**1 L**

Store at 4°C for up to 1 month.

2X EMSA cleavage buffer

Reagent	Final concentration	Amount
1 M Tris-HCl Buffer, pH 7.5	100 mM	10 mL
1 M $MgCl_2$	4 mM	400 μL
1 M 1,4-Dithiothreitol (DTT)	2 mM	200 μL
Glycerol	10%	10 mL
ddH_2O	N/A	Up to 100 mL
Total	**N/A**	**100 mL**

Filter using 0.22 μm filter and then store at −20°C for up to 2 months.

8% native polyacrylamide stock solution		
Reagent	Final concentration	Amount
Acrylamide/Bis-acrylamide (29:1), 30% solution	8%	267 mL
10X TBE solution	1X	100 mL
ddH_2O	N/A	Up to 1 L
Total	**N/A**	**1 L**
Store at RT for up to 1 month.		

10% denaturing polyacrylamide stock solution		
Reagent	Final concentration	Amount
Urea	8 M	480 *g*
Acrylamide/Bis-acrylamide (29:1), 30% solution	10%	330 mL
10X TBE solution	1X	100 mL
ddH_2O	N/A	Up to 1 L
Total	**N/A**	**1 L**
Filter using 0.22 μm filter and then store at RT for up to 1 month.		

12% denaturing polyacrylamide stock solution		
Reagent	Final concentration	Amount
Urea	8 M	480 *g*
Acrylamide/Bis-acrylamide (29:1), 30% solution	12%	400 mL
10X TBE solution	1X	100 mL
ddH_2O	N/A	Up to 1 L
Total	**N/A**	**1 L**
Filter using 0.22 μm filter and then store at RT for up to 1 month.		

2XTBE-urea denaturing buffer		
Reagent	Final concentration	Amount
Urea	8 M	24 *g*
0.5 M EDTA (pH 8.0)	0.02 M	2 mL
1 M Tris-HCl Buffer, pH 7.5	2 mM	0.1 mL
ddH_2O	N/A	Up to 50 mL
Total	**N/A**	**50 mL**
Filter using 0.22 μm filter and then store at −20°C for up to 2 months.		

STEP-BY-STEP METHOD DETAILS

Preparation of randomized RNA substrates

Timing: 7 days

In this section, pri-miRNAs containing 6 randomized base pairs (bp) around cleavage using pri-mir-576 short lower stem (pri-mir-576-SLS) as substrate backbone are obtained, the representation of this scheme is summarized in Figure 1. A similar approach can be used for different pri-miRNA backbones containing randomized nt at different regions.

1. Design of randomized primers to yield dsDNA that incorporates randomized regions.

Note: Our objective is to generate six randomized bp, each of which can be paired as either RY or YR (where R represents G or A, and Y represents C or T). This requires a total of $2^6 = 64$ pairs of primers to produce 64 groups of pri-miRNAs. Each pair consists of a forward and a reverse

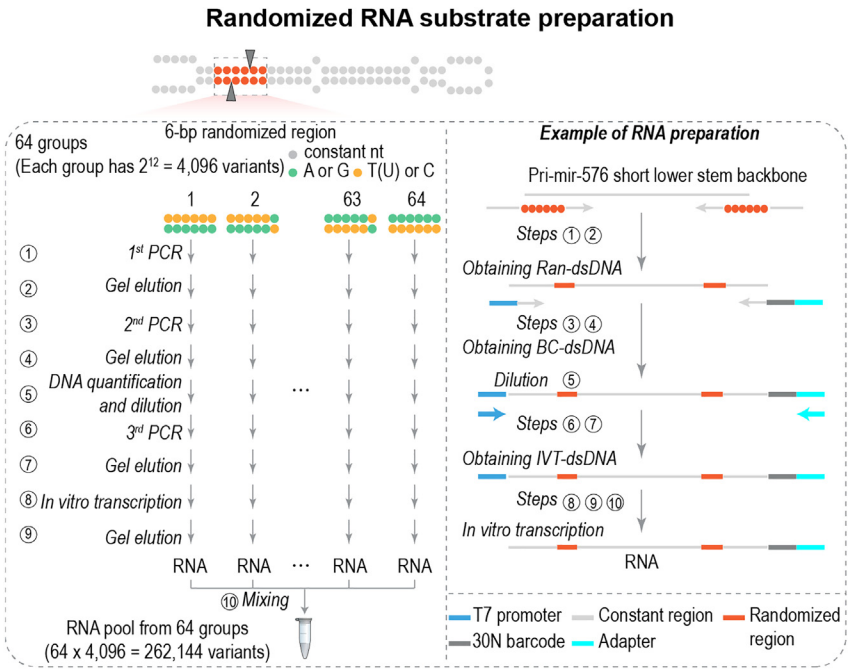


Figure 1. Preparation of randomized RNA substrates

The left panel illustrates the 9-step process of creating 64 groups of RNAs, with each group containing different combinations of A/G on one strand and U/C on the other strand within a 6-base pair window. The right panel demonstrates the step-by-step procedure for generating one RNA group. The randomized nt are incorporated into the DNA primers in step 1. These randomized nt are then converted into randomized DNA regions in the double-stranded DNA (dsDNA). Eventually, these randomized nt in dsDNA are converted into randomized ribonucleotides in RNA via *in vitro* transcription.

primer, each contains three segments: a constant sequence, a randomized sequence, and a complementary sequence. For instance, in the first group, F-sub-1 = CTG CCA TTT TAC AAT CCA AC RRR RRR TCT AAT TTC TCC ACG TCT, and R-sub-1 = TTC CGG TTG GGT TGA AAG AC RRR RRR TCC AAT TTT TCC ACA TCT, where RRR RRR is randomized sequences.

2. Obtain 64 pairs of primers, each contains 6 randomized nt, from Integrated DNA Technologies (IDT).
3. Use each primer pair containing a randomized sequence (for example: F-sub-1 and R-sub-1) to generate a dsDNA by PCR using pri-mir-576-SLS as a PCR template, resulting in 64 dsDNA containing 6-bp randomized regions (Ran-dsDNA).

Note: The number of PCR cycles should be adjusted for each primer pair by conducting a series of PCR tests with cycle numbers ranging from 20 to 35. This step is crucial to ensure the generation of sufficient PCR products.

a. PCR reaction mixture.

Reagent	Final concentration	Amount
Pri-mir-576-SLS DNA template	0.3 nM	1 μL (0.015 pmol)
Phusion HS II DNA polymerase	0.02 U/μL	0.5 μL (2 U/μL)
F-sub-x	1 μM	1 μL (50 pmol)
R-sub-x	1 μM	1 μL (50 pmol)
5X Phusion HF buffer	1X	10 μL
2 mM dNTPs	200 μM each	5 μL
ddH_2O		Up to 50 μL

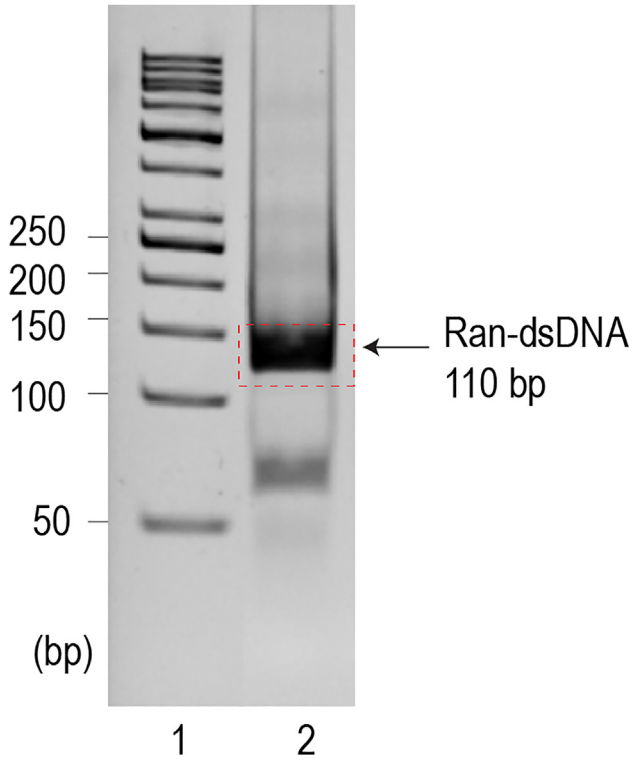


Figure 2. The figure depicts the Ran-dsDNA for group 64, analyzed using 8% native-PAGE

Lane 1 contains DNA ladders with the indicated sizes in bp. Lane 2 displays the Ran-dsDNA for group 64.

b. PCR cycling conditions.

Steps	Temperature	Time	Cycles
Initial denaturation	98°C	30 s	1
Denaturation	98°C	30 s	34 cycles
Annealing	60°C	15 s	
Extension	72°C	20 s	
Final extension	72°C	5 min	1
Hold	25°C	Forever	

4. Prepare the 8% native polyacrylamide gel (PAGE) and run the PCR product at 150 V for 45 min, stain the native PAGE gel with ethidium bromide and visualize it under UV transillumination. Figure 2 shows the PCR product of Ran-dsRNA for subgroup 64 using 8% native-PAGE.

Reagent	Final concentration	Amount
8% native polyacrylamide stock solution	8% acrylamide	15 mL
10% APS solution	N/A	150 μL
TEMED	N/A	15 μL
Total		15.165 mL

5. Excise the gel area containing Ran-dsDNA band of the correct size.
   a. Crush the excised gel into small pieces and transfer them into an Eppendorf tube.

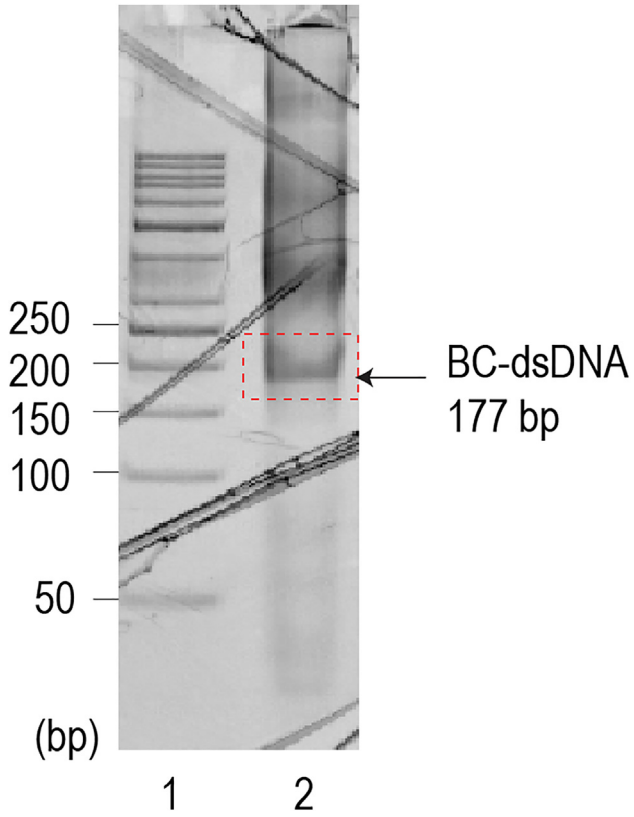


Figure 3. The figure depicts the BC-dsDNA of 177 bp for group 64, analyzed using 8% native-PAGE
Lane 1 contains DNA ladders with the indicated sizes in bp. Lane 2 displays the BC-dsDNA for group 64.

b. Add 500 μL of distilled water and shake the tube containing the gel pieces in a Thermomixer at 50°C for 1 h.
c. After shaking, centrifuge the tube at 20,000 × *g* for 10 min at 4°C and collect the supernatant, which contains the eluted Ran-dsDNA.

6. Recover dsDNA by isopropanol precipitation.
   a. Add sodium acetate (pH 5.5) to a final concentration of 0.3 M to the supernatant, followed by the addition of 2 volumes of isopropanol.
   b. Mix well and allow the Ran-dsDNA tube to be incubated at room temperature for 1 h.
   c. Centrifuge the Ran-dsDNA tube at 21,000 × *g* for 10 min at 4°C to pellet the DNA.
   d. Carefully remove the supernatant, leaving the Ran-dsDNA pellet.
   e. Add 500 μL of 80% ethanol to the pellet and centrifuge the tube again at 21,000 × *g* for 5 min at 4°C.
   f. Remove as much supernatant as possible and let the Ran-dsDNA pellet dry in a fume hood for about 15 min.
   g. The purified Ran-dsDNA is quantified using a NanoDrop spectrophotometer after dissolving in 20 μL of distilled water.
7. Obtain the F-T7-18nt-576v1s lib and R-RA3-30N-arti1-576-18comp primers to further amplify the Ran-dsDNA into BC-dsDNA. Figure 3 shows the PCR product of BC-dsRNA for subgroup 64 using 8% native-PAGE.

Note: Barcodes are incorporated into Ran-dsDNA by amplifying it using following primers. The forward primer contains the T7 promoter sequence (F-T7-18nt-576v1s lib, TAA TAC GAC TCA CTA TAG GGC TGC CAT TTT ACA ATC CA). The reverse primer contains 30

randomized nucleotides (30N) and a 3′-adapter region (R-RA3-30N-arti1-576-18comp, GTT CAG AGT TCT ACA GTC CGA CGA TCN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNC TGC CAT TTT ACA ATC CA), resulting in the generation of barcode-labeled dsDNA (BC-dsDNA). The 30N are incorporated into the reverse primers by the synthetic DNA company.

The optimal number of PCR cycles is determined by performing a series of PCR experiments with cycle numbers ranging from 10 to 25. This range helps to ensure that the PCR products are not over amplified.

a. PCR reaction mixture.

Reagent	Final concentration	Amount
Ran-dsDNA	1 nM	1 μL (0.05 pmol)
Phusion HS II DNA polymerase	0.02 U/μL	0.5 μL (2 U/μL)
F-T7-18nt-576v1s lib	1 μM	1 μL (50 pmol)
R-RA3-30N-arti1-576-18comp	1 μM	1 μL (50 pmol)
5X Phusion HF buffer	1X	10 μL
2 mM dNTPs	200 μM each	5 μL
ddH_2O		Up to 50 μL

b. PCR cycling condition.

Steps	Temperature	Time	Cycles
Initial denaturation	98°C	30 s	1
Denaturation	98°C	30 s	20 cycles
Annealing	60°C	15 s	
Extension	72°C	20 s	
Final extension	72°C	5 min	1
Hold	25°C	Forever	

c. Run the PCR reaction mixture on an 8% native PAGE at 150 V for 45 min. Stain the native-PAGE gel with ethidium bromide and visualize it under UV transillumination.

8. The BC-dsDNA is purified using a similar method for Ran-dsDNA. Quantify the BC-dsDNA, containing T7 promoter and barcodes, using the Qubit dsDNA HS Assay Kit.
9. Dilute the BC-dsDNA to obtain approximately 1,000 unique barcodes per a randomized sequence.

Note: For a specific group, we have 2^{12} variants (this number is subject to change depending on the design of the library). To achieve approximately 1,000 unique barcodes per randomized sequence, we need to obtain $2^{12} \times 1{,}000$ molecules of BS-dsDNA in 1 μL of PCR template. Given that 1 mol equals 6.02×10^{23} molecules, we need to dilute the BS-dsDNA to 6.8×10^{-3} fmol/μL and then use 1 μL of this for the PCR.

⚠ CRITICAL: This is an important step to get a prober number of barcodes for the sequencing library (see troubleshooting 4).

10. Amplify the diluted BC-dsDNAs using F-T7 (TAA TAC GAC TCA CTA TAG GG) and R-RA3 (TTG GCA CCC GAG AAT TCC A) primers to obtain IVT-dsDNAs. Figure 4 shows the PCR product of IVT-dsRNA for subgroup 64 using 8% native-PAGE.

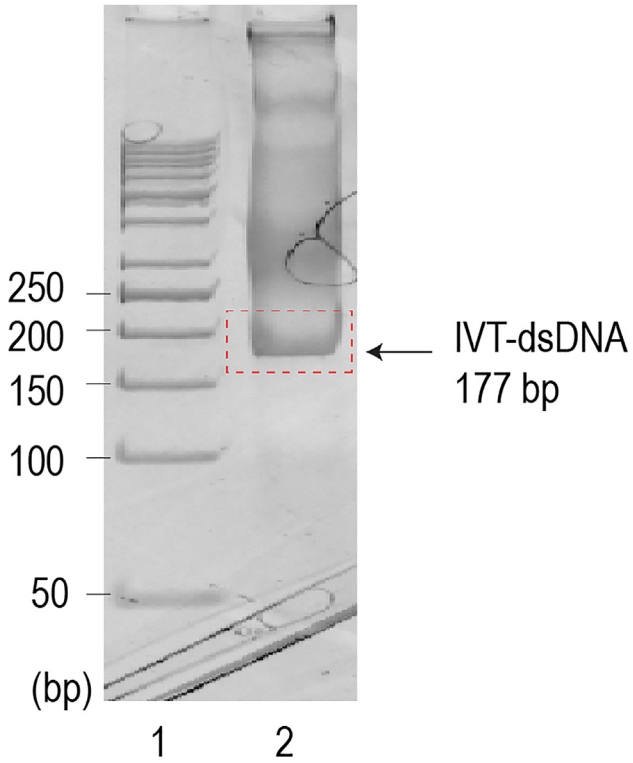


Figure 4. The figure depicts the IVT-dsDNA of 177 bp for group 64, analyzed using 8% native-PAGE
Lane 1 contains DNA ladders with indicated sizes in bp. Lane 2 displays the IVT-dsDNA for group 64.

Note: The number of PCR cycles is optimized by conducting a series of PCR experiments with number of cycle ranging from 10 to 25. This approach helps prevent the over-amplification of the PCR products.

a. PCR reaction mixture.

Reagent	Final concentration	Amount
BC-dsDNA	0.023 pM	1 μL (6.8×10^{-3} fmol)
Phusion HS II DNA polymerase	0.02 U/μL	3 μL (2 U/μL)
F-T7	1 μM	3 μL (50 pmol)
R-RA3	1 μM	3 μL (50 pmol)
5X Phusion HF buffer	1X	60 μL
2 mM dNTPs	200 μM each	30 μL
ddH_2O		Up to 300 μL

b. PCR cycling conditions.

Steps	Temperature	Time	Cycles
Initial denaturation	98°C	30 s	1
Denaturation	98°C	30 s	20 cycles
Annealing	60°C	15 s	
Extension	72°C	20 s	
Final extension	72°C	5 min	1
Hold	25°C	Forever	

11. The IVT-dsDNA is purified using gel electrophoresis, following the same procedure as described for the Ran-dsDNA. Use 200 ng IVT-dsDNA from the final PCR step for *in vitro* transcription reaction for 12 h to synthesize randomized pri-miRNAs.
12. Prepare 10% denaturing polyacrylamide gel (urea-PAGE) and pre-run the gel at 300 V for 45 min.

Reagent	Final concentration	Amount
10% denaturing polyacrylamide stock solution	10% acrylamide	15 mL
10% APS solution	N/A	150 μL
TEMED	N/A	15 μL
Total		15.165 mL

13. Load IVT-synthesized pri-miRNAs on 10% urea-PAGE and perform gel-purification following the same procedure as described for Ran-dsDNA. Obtain ~500 pmol of RNAs from each dsDNA.
14. Mix equal amounts of RNA from 64 randomized pri-miRNA groups to obtain a randomized pri-miRNA pool. Figure 5 shows the mixture of randomized pri-miRNAs using 10% urea-PAGE.

Preparation of recombinant proteins

Timing: 5 days

DROSHA in the complex with DGCR8 fragment is purified as following step

Note: The pXab-D3 plasmid containing the D3 fragment of DROSHA (amino acids 390–1365) fused with protein G. The pXC-G1 plasmid containing G1 (amino acids 728–750) and pXG-G2 plasmid containing G2 (amino acids 701–773) fragments of DGCR8 function to solubilize and stabilize DROSHA, although they do not enhance the RNA-binding affinity of the D3-G1 and D3-G2 complexes. As a result, the D3-G1 and D3-G2 complexes are functionally like DROSHA alone when it comes to substrate recognition. These two plasmids are fused with GFP and a 10x His-tag for protein purification step. The D3ΔBelt is a D3 fragment lacking amino acids 787–805. The pXab-D3ΔBelt plasmid is cloned from pXab-D3 by using in In-fusion mutagenesis method. Figure 6 shows the protein construct used in this protocol. The mutant plasmid is confirmed by Sanger sequencing before large scale purification. Purify the D3-G1 and D3-G2 complexes using the following method. The same protocol can also be applied to purify D3ΔBelt-G1 and D3ΔBelt-G2 complexes.

15. Large scale plasmid preparation.
    a. Use DH5α strain for plasmid transformation.
    b. Inoculate single colony of each plasmid is and grown in 300 mL LB medium for around 12 h at 37°C while shaking at 200 rpm.
    c. When the optical density (OD) of bacterial culture is 2–3 (OD600 = 2–3), harvest the cells by centrifugation at 5,000 × *g* for 10 min.
    d. Use the GeneJET Plasmid Maxi Prep Kit from Thermo Fisher Scientific for plasmid purification, measure plasmid concentration by NanoDrop and check plasmid quality on agarose gel.

⚠ CRITICAL: The plasmid quality can affect transfection in the next step. Prevent the denaturation of supercoiled plasmid DNA, in the lysis step do not incubate for more than 3 min.

16. The HEK293E cells are grown in a culturing medium at 37°C, supplemented with 5% CO2. When their confluence is 90%–100%, the cells are transferred to the shaking flash to get 10^5 cells per mL medium as the seeding cell density.

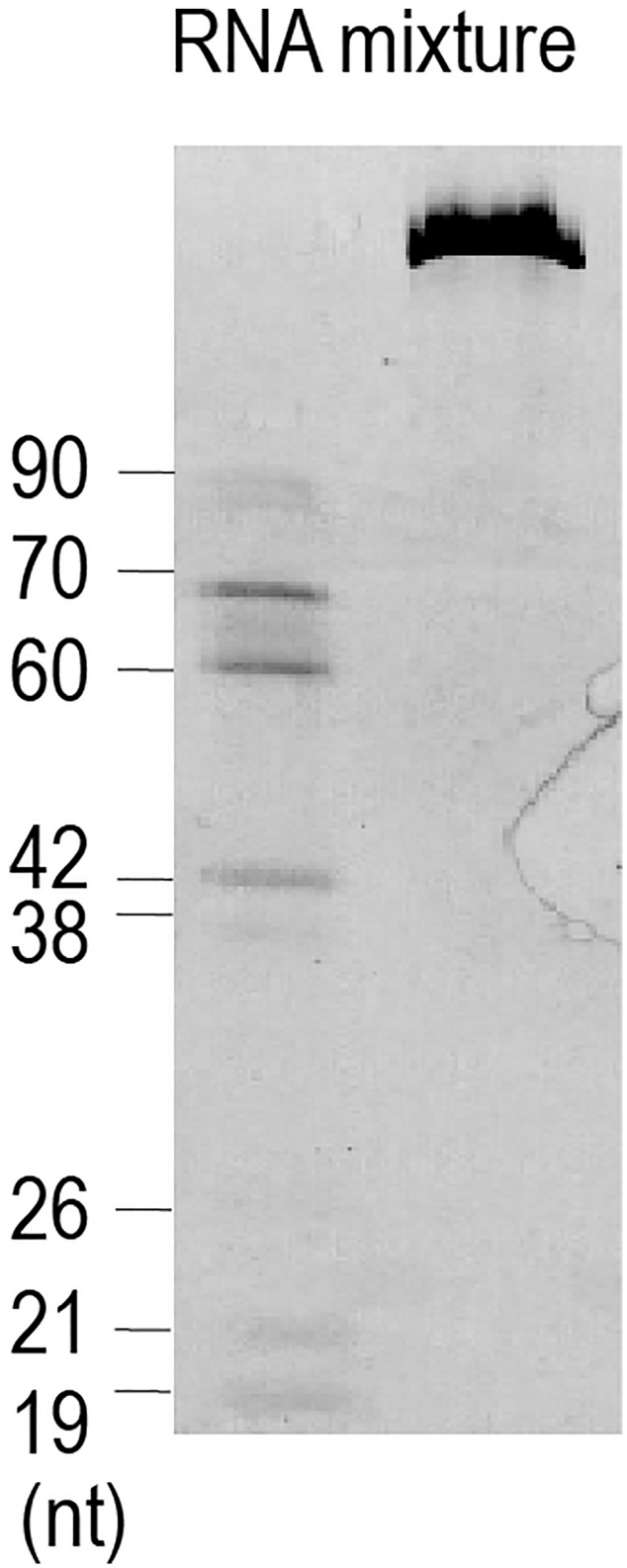


Figure 5. The figure depicts the mixture of randomized pri-miRNAs, analyzed using 10% urea-PAGE
Lane 1 contains ladders with indicated sizes in nt. Lanes 2 displays randomized pri-miRNAs for group 64.

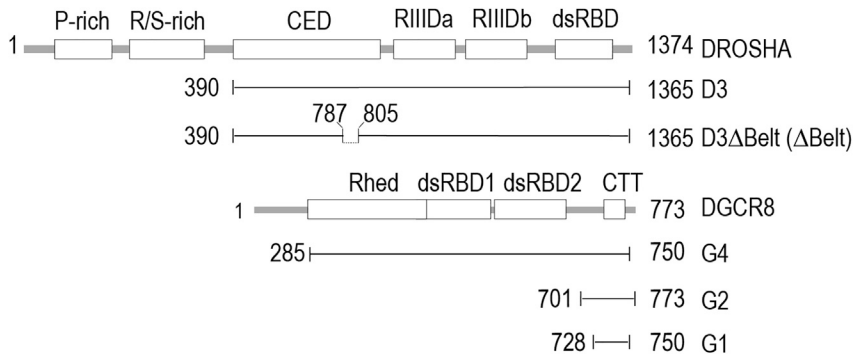


Figure 6. The protein constructs used in this study

The positions of the first and last amino acid residues for each construct are indicated.

17. Sixteen hours post-seeding, for 1 L of suspension cell medium, 0.5 mg of pXab-D3 plasmid is transfected. This is done along with either 0.2 mg of the pXC-G1 plasmid or the pXG-G2 plasmid, using 3 mg of linear polyethylenimine (PEI) and 100 mL of DMSO. Post-transfection, the cells are cultivated under the same conditions as those used during seeding.
18. Three days post-transfection, collect the transfected cells and lyse them by sonication in 100 mL lysis buffer containing T500 supplemented with 2 μg/mL RNase A (Thermo Scientific) and protease inhibitor tablet (Thermo Scientific).
19. Perform high-speed centrifugation for 40 min at 20,000 × *g* at 4°C to obtain a clear cell lysate and bind it to 2 mL of HisPur Ni-NTA resin. The Ni-NTA resin is pre-equilibrated with T500 buffer before binding.
20. Wash the protein-bound resin sequentially with three pre-chilled washing buffers: T2000, T0, and T500 supplemented 50 mM Imidazole, use 10 mL of washing buffer each time.
21. Elute the bound protein with T150 supplemented 250 mM imidazole.
22. Incubate the protein complex with Pierce HRV 3C Protease (Thermo Scientific) at 4°C for 16 h to cleave off the tags (protein G, GFP, and His-tag).
23. Equilibrate the UNOsphere Q strong anion exchange beads with T150 buffer, bind the tag-removed complex with 0.2 mL pre-equilibrate beads and wash it with T150 buffer supplemented 2 mM DTT.
24. Elute the complex in 1 mL elution buffer containing T500 supplemented 2 mM DTT. The eluted protein is concentrated by using 50 kDa MWCO centricon and centrifuging at 2,000 × *g* at 4°C until the total protein volume is 0.6 mL.
25. Run protein sample in gel filtration column.
    a. Connect the Superdex 200 Increase 10/300 GL gel filtration column (GE Healthcare) to the NGC Chromatography system and equilibrate it with 50 mL of enzyme stock buffer at a flow rate of 0.25 mL/min.
    b. Subsequently, activate the UV detector and set the detection wavelength. In this case, we have selected two main wavelengths: 260 nm and 280 nm.
    c. Load the concentrated protein onto the system through a 0.5 mL sample injection loop.
    d. Initiate the elution phase at a constant flow rate of 0.25 mL/min.
    e. Collect different elution fractions in collection tubes, allotting 400 mL for each fraction.
    f. The peak fraction containing the protein is aliquoted and stored at −80°C for future use.

The dimeric G4 (G4G4) is purified using the following procedure

26. The G4 fragment of DGCR8 (amino acids 285–750) is fused to an N-terminal His-sfGFP tag and cloned into the pET-28a vector. The resulting plasmid is transformed into BL21(DE3)-CodonPlus-RIPL and cultured in 500 mL of LB medium. When the OD600 of the cell culture reaches ∼0.5–0.6, isopropyl β-d-1-thiogalactopyranoside is added to a final concentration of 0.2 mM to induce gene expression. The cells are then cultured overnight at 16°C.

27. The cells are harvested and lysed by sonication in 100 mL T500 supplemented with 2 μg/mL RNase A (Thermo Scientific) and 1 mM phenylmethanesulfonyl fluoride.
28. The lysate is centrifuged at 20,000 × *g* for 40 min at 4°C, and the supernatant is collected and bound to 2 mL of pre-equilibrated HisPur Ni-NTA resin.
29. The resin is washed three times with 150 mL T500 buffer supplemented with 20 mM imidazole. The protein is eluted using the same buffer containing 200 mM imidazole.
30. Dilute the eluted protein with T0 buffer to get the final 150 mM NaCl in the protein mixture, after that, incubate with Pierce HRV 3C Protease (Thermo Scientific) at 4°C for 16 h to cleave off the His-sfGFP tag.
31. Bind the tag-free protein to 0.2 mL pre-equilibrated UNOsphere S bead and wash with T150 containing 2 mM DTT. The SP-bound protein is eluted with a 1 mL T500 buffer containing 2 mM DTT, then concentrated using a 50 kDa MWCO centricon.

⚠ CRITICAL: This step is crucial as it allows the removal of contaminants with a smaller molecular weight and concentrates the protein for subsequent procedures.

    a. Introduce the protein solution into the Centricon and centrifuge at a specified speed 2,000 × *g* for a certain duration 5 min at 4°C.
    b. Repeat the centrifugation step until the volume of the protein solution is reduced to 0.6 mL.
32. The 0.5 mL of concentrated protein is loaded onto a HiLoad 10/300 Superdex 200 column (GE Healthcare) and eluted with a protein stock buffer. The peak fraction, which contains the G4 dimer and exhibits a brown color, is aliquoted into 10 μL portions and stored at −80°C.

Note: Each aliquot is intended for single-use and should be fully utilized each time it is accessed.

Microprocessor reconstitution procedure

33. The Microprocessor is reconstituted by combining the equal amount of the D3-G1 complex and the G4 dimer in each reaction. Figure 7 shows the purified proteins using SDS-PAGE.

Note: G1 serves to solubilize and stabilize DROSHA. As G1 is more readily displaced by G4 dimer than G2, the D3-G1 complex is utilized in this reconstitution instead of the D3-G2 complex.

Preparation of processing assays

Timing: 5 h

34. Prepare the pri-miRNA cleavage assay materials.
    a. Gather 3 pmol of pri-miRNAs for use in the pri-miRNA cleavage assays.
    b. Obtain the D3-G4 complex by incubating 3 pmol of D3-G1 and 3 pmol of G4 dimer on ice for 10 min. In addition, thaw D3-G2 complex protein on ice.
35. Prepare the 10 μL reaction mixture for each sample.
    a. Combine the following components in the reaction mixture: 5 μL of 2X EMSA cleavage buffer and 2 U RNase Inhibitor.
    b. Add the appropriate 0.5–3 pmol of D3-G4 or D3-G2 complex and proper amount of protein stock buffer to get the final 150 mM NaCl.

Reagent	Final concentration	Amount
2X EMSA cleavage buffer	1X	5 μL
SUPERase In RNase Inhibitor (20 U/μL)	0.2 U/μL	0.1 μL

(*Continued on next page*)

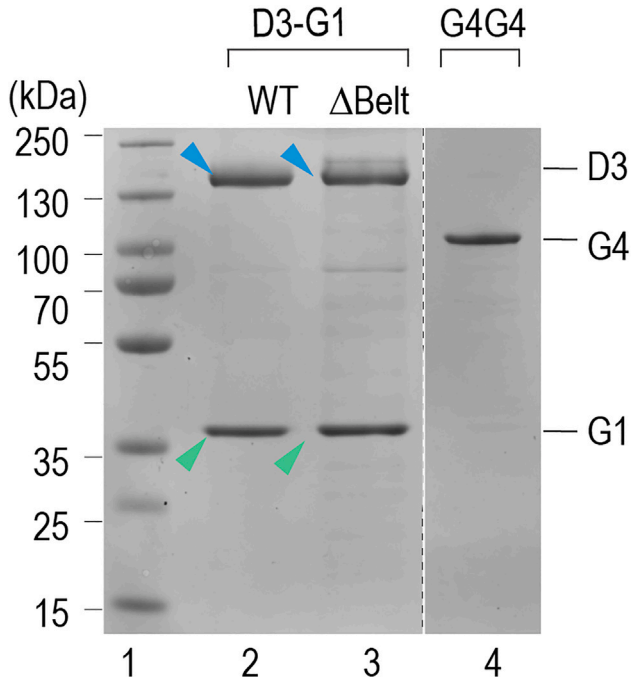


Figure 7. Analysis of purified proteins via SDS-PAGE

The WT and mutant D3-G1 complexes consist of a DROSHA fragment (D3, amino acids 390–1365) and the G1 fragment of DGCR8 (G1, amino acids 728–750). G4G4 represents the dimeric G4 fragment of DGCR8 (amino acids 285–750). The D3ΔBelt is D3 fragment lacking amino acids, 787–805.

Continued

Reagent	Final concentration	Amount
Pri-miRNA	0.3 pmol/μL	1 μL
D3-G4 or D3-G2 protein	0.3 pmol/μL	Total 3 μL
Protein stock buffer	N/A	
ddH_2O	N/A	0.9 μL
Total		10 μL

⚠ CRITICAL: The RNA and protein amounts should be tested by titration assay before conducting the processing assays.

36. Perform the pri-miRNA cleavage assay in the thermal cycler as below.
    a. Incubate the reaction mixture at 37°C for 2 h. This allows the proteins to recognize and cleave the RNA substrates.
    b. Stop the reaction by adding 10 μL of 2× TBE-urea denaturing buffer and 20 μg proteinase K.
    c. Incubate the reaction mixture at 37°C for 15 min, followed by 50°C for 15 min, and finally, 95°C for 5 min.
37. Analyze the cleavage mixture using urea-PAGE.
    a. Prepare 12% urea-PAGE.

Reagent	Final concentration	Amount
12% denaturing polyacrylamide stock solution	10% acrylamide	15 mL
10% APS solution	N/A	150 μL
TEMED	N/A	15 μL
Total		15.165 mL

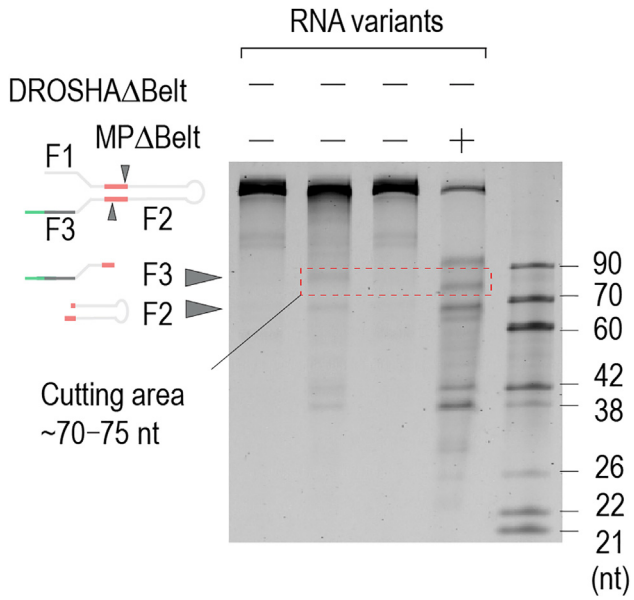


Figure 8. Analysis of pri-miRNA cleavage assays using 12% urea-PAGE

The gel region containing cleaved products, F3, resulting from the cleavages by DROSHA or Microprocessor (MP) is indicated within the dashed red box and excised from the gel for subsequent RNA purification.

b. Pre-run a 12% urea-PAGE in 1× TBE buffer for 45 min at 300 V.
c. Load the cleavage mixture onto the gel and continue to run for ~45 min at 300 V.

38. Stain the gel and visualize the RNA bands. Stain the gel with SYBR Green II RNA Gel Stain for 5 min.
39. Image the stained gel using the Bio-Rad Gel Doc XR + system with CCD detector, UV transilluminator 302 nm, standard emission filter, choosing SYBR Green as the application. Figure 8 shows the pri-miRNA cleavage assay using 12% urea-PARE, the cutting area is indicated in the red box.
40. Slice the cleaved RNA product bands from gel and perform gel purification as described above for Ran-dsDNA.

Preparation of DNA library construction

Timing: 3 days

41. Perform 5′ adapter ligation.
    a. Add the 5′ adapter for the purified F3 fragments.

Reagent	Final concentration	Amount
Purified F3 fragment	N/A	4 μL
5'-adapter (RA5-6N, GUU CAG AGU UCU ACA GUC CGA CGA UCN NNN NNN)	0.5 μM	1 μL (10 pmol)

b. Incubate the mixture at 70°C for 2 min, immediately transfer the tube to the ice box.
c. Add and mix the following components (see troubleshooting 2).

Reagent	Final concentration	Amount
SUPERase In RNase Inhibitor (20 U/μL)	1 U/μL	1 μL
10X T4 RNA ligase reaction buffer	1X	2 μL

(*Continued on next page*)

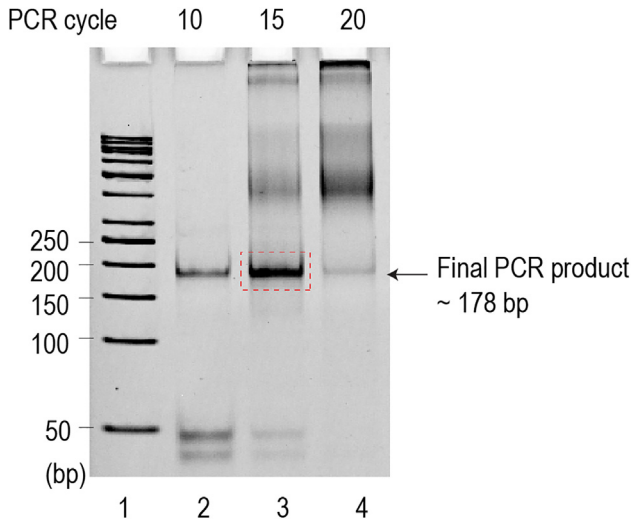


Figure 9. Analysis of the DNA library of cleaved products using 8% native-PAGE

The DNA libraries were amplified through 10, 15, and 20 PCR cycles. The resulting DNAs from the 15-cycle PCR were excised from the gel and subsequently purified.

Continued

Reagent	Final concentration	Amount
10 mM ATP	1 mM	2 μL
T4 RNA Ligase 1	1 U/μL	2 μL
50% PEG8000 (NEB)	20% PEG	8 μL

d. Incubate the reaction mixture at 25°C for 16 h.

42. Reverse-transcribe the RA5-ligated RNAs.

a. Add the R-RA3 primer (TTG GCA CCC GAG AAT TCC A) to the reaction mixture.

Reagent	Final concentration	Amount
RA5-ligated RNAs mixture	N/A	20 μL
R-RA3 primer (TTG GCA CCC GAG AAT TCC A)	0.4 μM	2 μL (20 pmol)

b. Incubate the mixture at 70°C for 2 min, immediately transfer the tube to the ice box.

c. Add and mix the following components.

Reagent	Final concentration	Amount
SUPERase In RNase Inhibitor (20 U/μL)	0.4 U/μL	1 μL
5× SSIV Buffer	1X	10 μL
2 mM dNTPs of each	0.5 mM of each	12.5 μL
100 mM DTT	5 mM	2.5 μL
SuperScript IV Reverse Transcriptase (200 U/μL)	8 U/μL	2 μL

d. Incubate the reaction mixture at 50°C for 1 h.

e. Stop the reaction by incubating it at 80°C for 10 min.

43. PCR-amplify the cDNA of cleavage product. Amplify the cDNA using a pair of sequencing primers RP1 and one of the RPIx primers. Figure 9 shows the DNA library of cleaved products using 8% native-PAGE.

Note: The number of PCR cycles should be firstly tested with 10 μL PCR reaction. After obtaining a suitable cycle, we can scale up to 50 μL PCR reaction.

a. PCR reaction mixture.

Reagent	Final concentration	Amount
cDNA of cleavage product	N/A	1 μL
Phusion HS II DNA polymerase	0.02 U/μL	0.1 μL (2 U/μL)
RP1	0.25 μM	0.25 μL (2.5 pmol)
RPIx	0.25 μM	0.25 μL (2.5 pmol)
5X Phusion HF buffer	1X	2 μL
2 mM dNTPs	200 μM each	1 μL
ddH_2O		Up to 10 μL

b. PCR cycling conditions.

Steps	Temperature	Time	Cycles
Initial denaturation	98°C	30 s	1
Denaturation	98°C	30 s	10, 15, 20 cycles
Annealing	60°C	15 s	
Extension	72°C	20 s	
Final extension	72°C	5 min	1
Hold	25°C	Forever	

44. Reverse-transcribe for original substrate.
    a. Add the R-RA3 primer (TTG GCA CCC GAG AAT TCC A) to the reaction mixture.

Reagent	Final concentration	Amount
Original RNA substrate	0.25 μM	0.5 μL (5 pmol)
R-RA3 primer (TTG GCA CCC GAG AAT TCC A)	1 μM	2 μL (20 pmol)
2 mM dNTP	0.4 mM each	4 μL
50 mM NaCl	3.5 mM	1.4 μL
ddH_2O		Up to 14.5 μL

    b. Incubate the mixture at 65°C for 5 min, immediately transfer the tube to the ice box.
    c. Add and mix the following components.

Reagent	Final concentration	Amount
SUPERase In RNase Inhibitor (20 U/μL)	0.5 U/μL	0.5 μL
5× SSIV Buffer	1X	4 μL
100 mM DTT	2.5 mM	0.5 μL
SuperScript IV Reverse Transcriptase (200 U/μL)	5 U/μL	0.5 μL

    d. Incubate the reaction mixture at 50°C for 15 min; stop the reaction by incubating the mixture at 80°C for 10 min.
45. Perform the first PCR amplification for the cDNA of the original substrate. Amplify the cDNA using a pair of primers F-RA5-18-comp (TTC AGA GTT CTA CAG TCC GAC GAT CCT GCC ATT TTA CAA TCC A) and R-RA3 primers to get the first original substrate dsDNA: FOS-dsDNA. Figure 10 shows the first PCR product for the cDNA from original substrate using 8% native-PAGE.

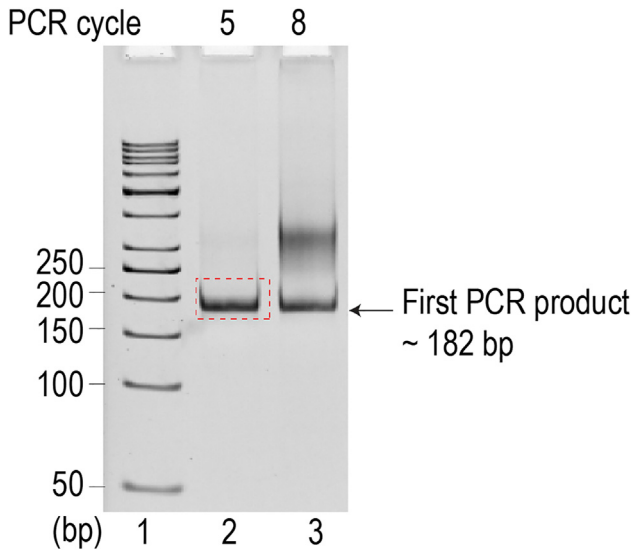


Figure 10. Analysis of the first PCR amplification for the cDNA of the original substrate using 8% native-PAGE

The DNAs underwent amplification through 5 and 8 cycles. The resulting DNA from the 5-cycle PCR was excised from the gel and subsequently purified.

Note: The PCR is conducted with two different cycles: 5 and 8. The PCR products of both are then examined. In the case of the 8-cycle PCR, larger DNA fragments are observed, apart from the expected DNA band of approximately 182 bp. These larger fragments may have been produced due to over-amplification. Consequently, the DNA band obtained from the 5-cycle PCR is selected for the subsequent steps.

a. PCR reaction mixture.

Reagent	Final concentration	Amount
cDNA of original substrate	N/A	1 μL
Phusion HS II DNA polymerase	0.02 U/μL	0.1 μL (2 U/μL)
RP1	0.25 μM	0.25 μL (2.5 pmol)
R-RA3	0.25 μM	0.25 μL (2.5 pmol)
5X Phusion HF buffer	1X	2 μL
2 mM dNTPs	200 μM each	1 μL
ddH_2O		Up to 10 μL

b. PCR cycling conditions.

Steps	Temperature	Time	Cycles
Initial denaturation	98°C	30 s	1
Denaturation	98°C	30 s	5, 8 cycles
Annealing	60°C	15 s	
Extension	72°C	20 s	
Final extension	72°C	5 min	1
Hold	25°C	Forever	

46. Perform the second PCR amplification for the resulting dsDNA. Amplify the resulting DNA using a pair of sequencing primers, RP1 and one of the RPIx primers. Figure 11 shows the DNA library of the original substrate using 8% native-PAGE.

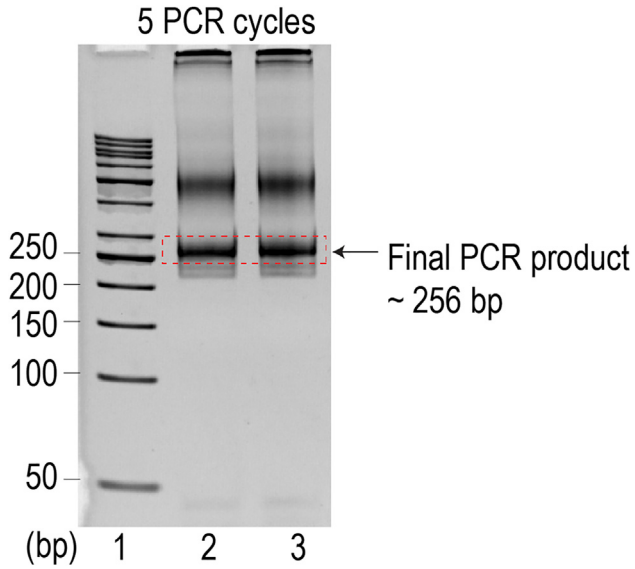


Figure 11. Analysis of the second PCR amplification for the cDNA of the original substrate using 8% native-PAGE
The DNAs were amplified through a 5-cycle PCR. The resulting DNAs from the 5-cycle PCR were excised from the gel and subsequently purified.

a. PCR reaction mixture.

Reagent	Final concentration	Amount
FOS-dsDNA	0.01 μM	1 μL (0.5 pmol)
Phusion HS II DNA polymerase	0.02 U/μL	0.5 μL (2 U/μL)
RP1	1 μM	1 μL (50 pmol)
RPIx	1 μM	1 μL (50 pmol)
5X Phusion HF buffer	1X	10 μL
2 mM dNTPs	200 μM each	5 μL
ddH_2O	N/A	Up to 50 μL

b. PCR cycling conditions.

Steps	Temperature	Time	Cycles
Initial denaturation	98°C	30 s	1
Denaturation	98°C	30 s	5 cycles
Annealing	60°C	15 s	
Extension	72°C	20 s	
Final extension	72°C	5 min	1
Hold	25°C	Forever	

47. Preparation the first PCR for DNA dictionary library. Amplify dsDNA from the IVT-dsDNA using a primer pair, F-RA5-18-comp (TTC AGA GTT CTA CAG TCC GAC GAT CCT GCC ATT TTA CAA TCC A), and R-RA3 to get the First dictionary dsDNA: FD-dsRNA.

Note: In the HT cleavage assays, randomized bp are incorporated into the RNA duplex region surrounding the cleavage sites of Microprocessor. This action results in the randomized pri-miRNA being divided into three distinct fragments: F1, F2, and F3, as depicted in Figure 8, with each fragment containing a selection of randomized nucleotides. Therefore, a means of

associating the cleaved product with its original substrate is necessary.

To facilitate this, a randomized 30 nt region, known as a 30N barcode, is integrated into the 3′ end region of the randomized IVT-dsDNA. Consequently, the 30N barcode is contained in both the synthesized randomized pri-miRNA and its cleaved F3 fragment.

A DNA dictionary library of IVT-dsDNA is then produced, enabling the sequencing of all barcode sequences present in the IVT-dsDNA. Following this, the sequences obtained from the pri-miRNA substrate and cleaved F3 fragments are analyzed, and their barcode sequences are mapped to those in the DNA dictionary. If the F3 and pri-miRNA share an identical barcode sequence with a DNA in the DNA dictionary, it indicates that the specific F3 fragment has been derived from that particular pri-miRNA substrate.

a. PCR reaction mixture.

Reagent	Final concentration	Amount
IVT-dsDNA	0.01 μM	1 μL (0.5 pmol)
Phusion HS II DNA polymerase	0.02 U/μL	0.5 μL (2 U/μL)
F-RA5-18-comp	1 μM	1 μL (50 pmol)
R-RA3	1 μM	1 μL (50 pmol)
5X Phusion HF buffer	1X	10 μL
2 mM dNTPs	200 μM each	5 μL
ddH_2O	N/A	Up to 50 μL

b. PCR cycling conditions.

Steps	Temperature	Time	Cycles
Initial denaturation	98°C	30 s	1
Denaturation	98°C	30 s	10 cycles
Annealing	60°C	15 s	
Extension	72°C	20 s	
Final extension	72°C	5 min	1
Hold	25°C	Forever	

48. Obtain the DNA library: further amplify the dsDNAs using a sequencing primer pair RP1 (AAT GAT ACG GCG ACC ACC GAG ATC TAC ACG TTC AGA GTT CTA CAG TCC GA) and one of RPIx.

a. PCR reaction mixture.

Reagent	Final concentration	Amount
FD-dsDNA	0.01 μM	1 μL (0.5 pmol)
Phusion HS II DNA polymerase	0.02 U/μL	0.5 μL (2 U/μL)
RP1	1 μM	1 μL (50 pmol)
RPIx	1 μM	1 μL (50 pmol)
5X Phusion HF buffer	1X	10 μL
2 mM dNTPs	200 μM each	5 μL
ddH_2O	N/A	Up to 50 μL

b. PCR cycling conditions.

Steps	Temperature	Time	Cycles
Initial denaturation	98°C	30 s	1
Denaturation	98°C	30 s	5 cycles
Annealing	60°C	15 s	
Extension	72°C	20 s	

(*Continued on next page*)

Continued

Steps	Temperature	Time	Cycles
Final extension	72°C	5 min	1
Hold	25°C	Forever	

49. Sequence the libraries. Run the libraries using the Illumina NovaSeq 6000 sequencer, paired end 150 bp and 200 million read for each sample.

EXPECTED OUTCOMES

The total IVT-dsDNA amount after PCR and gel purification is around 0.5 μM (~50 ng/μL) in 20 μL. The RNA amount after *in vitro* transcription and gel purification is around 10 μM in 50 μL. The concentration of D3-G1 or D3-G2 is about 2 μM in 400 μL. The concentration of dimeric G4 is around 10 μM in 400 μL. The dsDNA library prepared for sequencing has a concentration ranging from 4 to 40 nM in a volume of 10 μL, which is sufficient for Illumina sequencing. The sequencing analysis is expected to yield 99.99% of the anticipated randomized sequences. Additionally, in the control sample, more than 96% of the variants should be associated with at least 10 unique barcodes, and the average number of unique barcodes per variant should be approximately 100.

LIMITATIONS

The high-throughput processing assays use a recombinant protein complex purified from a human cell system, which contains all essential domains of the Microprocessor required for its catalytic function. However, it lacks the N-terminal region of DROSHA, and the role of this region in the enzyme's catalytic function is not fully known. It is possible that the N-terminal region may have roles in substrate recognition, binding, or regulation of the Microprocessor activity, and its absence could potentially impact the assay's outcomes. Future studies could investigate the importance of this region and its contribution to DROSHA function to gain a more comprehensive understanding of the Microprocessor's mechanism of action.

Furthermore, the study randomizes a 6 bp region in pri-miRNAs, intending to collect the bp from the randomized region, but mismatches may also contribute to the motifs, which are not considered in the protocol. These mismatches could introduce biases in the identified motifs and might affect the overall interpretation of the results. To address this limitation, researchers could develop a different randomized substrate system that includes mismatches by using random N-N pair instead of R-Y pair as in our library. This allows a more accurate representation of the sequence motifs critical for enzyme activity.

Additionally, the protocol is designed for human Microprocessor and pri-mir-576 as a backbone. While it can be adapted for Microprocessors from different species and various pri-miRNA backbones, this may require additional optimization and validation. This adaptation could involve identifying an optimal procedure for purifying Microprocessors from different species and optimizing reaction conditions for efficient processing.[16] The use of different pri-miRNA backbones may also reveal other regulatory mechanisms and sequence preferences that are not observed with pri-mir-576 as the sole substrate. By exploring these variations, researchers can gain valuable insights into the conservation and diversity of Microprocessor-mediated pri-miRNA processing across species.

TROUBLESHOOTING

Problem 1

Insufficient RNA yield after *in vitro* transcription or RNA degradation (step 16).

Potential solution

RNA is less stable than DNA, hence, all experiments should be conducted in RNase-free environment to prevent RNA degradation. Ensure that the bench and pipettes are thoroughly cleaned. For short transcripts, consider enhancing the *in vitro* transcription reaction yield by prolonging the incubation time and by increasing the quantity of the DNA template and RNA polymerase. Such alterations could potentially boost the number of transcription rounds, leading to an increased RNA yield.

Problem 2

Insufficient final DNA library concentration for Illumina sequencing (less than 4 nM) (steps 43–50).

Potential solution

The ligation of the 5′ adapter is a vital step in the DNA library preparation for the F3 cleavage product. Hence, it is critical to use the correct quantities of all components during this process. For instance, if the concentration of Polyethylene Glycol (PEG8000) drops below 15%, it can result in decreased ligation efficiency. Therefore, carefully monitoring and maintaining the appropriate concentrations of all components could potentially improve the final DNA library concentration.

Problem 3

Incomplete or biased representation of randomized sequences (steps 1–13). The protocol may not yield all expected randomized sequences due to RNA backbone usage or biases introduced during chemical ssDNA synthesis and PCR steps. This limitation could affect the overall diversity of the RNA substrates and the accuracy of the processing dynamics analysis.

Potential solution

- Optimize PCR conditions, specifically focusing on the number of PCR cycles. To achieve this, carrying out PCR reactions with varying cycle numbers ranging from 20 to 35 cycles. The objective of this process is to identify the optimal PCR cycle number that results in an expected PCR product free from nonspecific amplified DNA, thereby preventing DNA over-amplification. After determining the optimal PCR conditions for one group, apply the same method to process other groups.
- Validate the diversity of the DNA libraries at early steps through sequencing to guarantee an appropriate range of sequences is obtained. For instance, the DNA dictionary should be sequenced first to confirm that the randomized ssDNA ordered from the supplier and used to create the DNA dictionary contains a sufficient variety of randomized sequences.

Problem 4

Inaccurate barcode representation (steps 1–13). The protocol may not provide a proper number of barcodes for each sequence due to inaccurate calculations of DNA amounts by NanoDrop 2000/2000c Spectrophotometer. This could lead to an unequal representation of sequences in the final analysis, potentially skewing the results.

Potential solution

- It is crucial to obtain a sufficient number of barcodes per variant for data analysis. The barcode number per variant should be within the 50–1,000 range. The critical step to achieve this barcode number is in the PCR for IVT-dsDNA step, where BC-dsDNA is diluted to achieve 1,000 barcodes per variant. In this step, the amount of BC-dsDNA should be accurately calculated. Therefore, it is important to obtain highly purified and quantified dsDNA for this purpose. It is essential to quantify this DNA using at least two methods: Qubit dsDNA HS Assay Kit and gel electrophoresis, comparing the band density with DNA ladders of known amounts.

- Additionally, after obtaining the RNA substrates, they can be cloned into control samples and sent for sequencing to verify the barcode numbers. If the barcode numbers are satisfactory, these RNA substrates can be used for cleavage assays.

Problem 5

Protein impurities (steps 18–35). The presence of nucleases or other contaminants in the purified recombinant proteins might result in additional RNA cleavage, affecting the reliability and interpretation of the cleavage assays. Rigorous protein purification and quality control steps are essential to minimize this limitation.

Potential solution

- Assess the purity of the recombinant proteins using SDS-PAGE. Ensure that the proteins achieve high purity, over 90%, based on the band density in the gel. Additionally, the protein amount should be sufficient. The concentration of proteins should be at least 1 pmol/μL.
- Optimize protein purification and conduct experiments with the purified enzymes using known substrates, such as pri-mir-16-1 and pri-mir-30a, to ensure that the enzyme cleaves them normally, generating the expected RNA fragments. Additionally, the enzyme can be tested with other substrates to further confirm its activity and specificity.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Tuan Anh Nguyen (tuananh@ust.hk).

Technical contact

Technical questions on executing this protocol should be directed to and will be answered by the techinical contact, Thi Nhu-Y Le (tnyle@connect.ust.hk).

Materials availability

For DNA and RNA used in this protocol, if needed, please contact Tuan Anh Nguyen (tuananh@ust.hk).

Data and code availability

No datasets or codes were generated in this study.

ACKNOWLEDGMENTS

This work was supported by the Research Grants Council of the University Council Committee of Hong Kong (project number: 16102321). We extend our gratitude to members of our group for their critical discussions.

AUTHOR CONTRIBUTIONS

T.N.-Y.L. and T.A.N. thoroughly evaluated the protocol and wrote the manuscript and T.A.N. supervised the project. All authors reviewed and approved the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Nguyen, T.L., Nguyen, T.D., Ngo, M.K., Le, T.N.Y., and Nguyen, T.A. (2023). Noncanonical processing by animal Microprocessor. Mol. Cell *83*, 1810–1826.e8.

2. Bartel, D.P. (2018). Metazoan MicroRNAs. Cell *173*, 20–51.

3. Nguyen, H.M., Nguyen, T.D., Nguyen, T.L., and Nguyen, T.A. (2019). Orientation of Human Microprocessor on Primary MicroRNAs. Biochemistry *58*, 189–198.

4. Nguyen, T.D., Trinh, T.A., Bao, S., and Nguyen, T.A. (2022). Secondary structure RNA elements control the cleavage activity of DICER. Nat. Commun. *13*, 2138.

5. Auyeung, V.C., Ulitsky, I., McGeary, S.E., and Bartel, D.P. (2013). Beyond secondary structure: Primary-sequence determinants license Pri-miRNA hairpins for processing. Cell *152*, 844–858.

6. Dang, T.L., Le, C.T., Le, M.N., Nguyen, T.D., Nguyen, T.L., Bao, S., Li, S., and Nguyen, T.A. (2020). Select amino acids in DGCR8 are essential for the UGU-pri-miRNA interaction and processing. Commun. Biol. *3*, 344.

7. Partin, A.C., Zhang, K., Jeong, B.C., Herrell, E., Li, S., Chiu, W., and Nam, Y. (2020). Cryo-EM Structures of Human Drosha and DGCR8 in Complex with Primary MicroRNA. Mol. Cell *78*, 411–422.e4.

8. Fang, W., and Bartel, D.P. (2015). The Menu of Features that Define Primary MicroRNAs and Enable De Novo Design of MicroRNA Genes. Mol. Cell *60*, 131–145.

9. Nguyen, T.A., Jo, M.H., Choi, Y.G., Park, J., Kwon, S.C., Hohng, S., Kim, V.N., and Woo, J.S. (2015). Functional anatomy of the human microprocessor. Cell *161*, 1374–1387.

10. Kwon, S.C., Nguyen, T.A., Choi, Y.G., Jo, M.H., Hohng, S., Kim, V.N., and Woo, J.S. (2016). Structure of Human DROSHA. Cell *164*, 81–90.

11. Kwon, S.C., Baek, S.C., Choi, Y.G., Yang, J., Lee, Y.S., Kim, V.N., Woo, J.S., and Kim, V.N. (2019). Molecular Basis for the Single-Nucleotide Precision of Primary microRNA Processing. Mol. Cell *73*, 505–518.e5.

12. Le, C.T., Nguyen, T.D., Nguyen, T.L., and Nguyen, T.A. (2020). Human Disease-Associated Single Nucleotide Polymorphism Changes the Orientation of DROSHA on Pri-Mir-146a. Rna, Rna, pp. 077487.120.

13. Li, S., Nguyen, T.D., Nguyen, T.L., and Nguyen, T.A. (2020). Mismatched and wobble base pairs govern primary microRNA processing by human Microprocessor. Nat. Commun. *11*, 1926.

14. Li, S., Le, T.N.Y., Nguyen, T.D., Trinh, T.A., and Nguyen, T.A. (2021). Bulges control pri-miRNA processing in a position and strand-dependent manner. RNA Biol. *18*, 1716–1726.

15. Jin, W., Wang, J., Liu, C.P., Wang, H.W., and Xu, R.M. (2020). Structural Basis for pri-miRNA Recognition by Drosha. Mol. Cell *78*, 423–433.e5.

16. Nguyen, T.L., Nguyen, T.D., Ngo, M.K., and Nguyen, T.A. (2023). Dissection of the Caenorhabditis elegans Microprocessor. Nucleic Acids Res. *51*, 1512–1527.

17. Kim, K., Nguyen, T.D., Li, S., and Nguyen, T.A. (2018). SRSF3 recruits DROSHA to the basal junction of primary microRNAs. Rna *24*, 892–898.

18. Kim, M.J., Hwang, E.S., Kim, K.J., Maeng, S., Heo, H.J., Park, J.H., Kim, D.O., Kim, K., Baek, S.C., Lee, Y., et al. (2021). Resource A quantitative map of human primary microRNA processing sites Resource A quantitative map of human primary microRNA processing sites. Mol. Cell *11*, 1–18.

19. Le, M.N., Nguyen, T.D., and Nguyen, T.A. (2023). SRSF7 and SRSF3 depend on RNA sequencing motifs and secondary structures to regulate Microprocessor. Life Sci. Alliance *6*, e202201779.