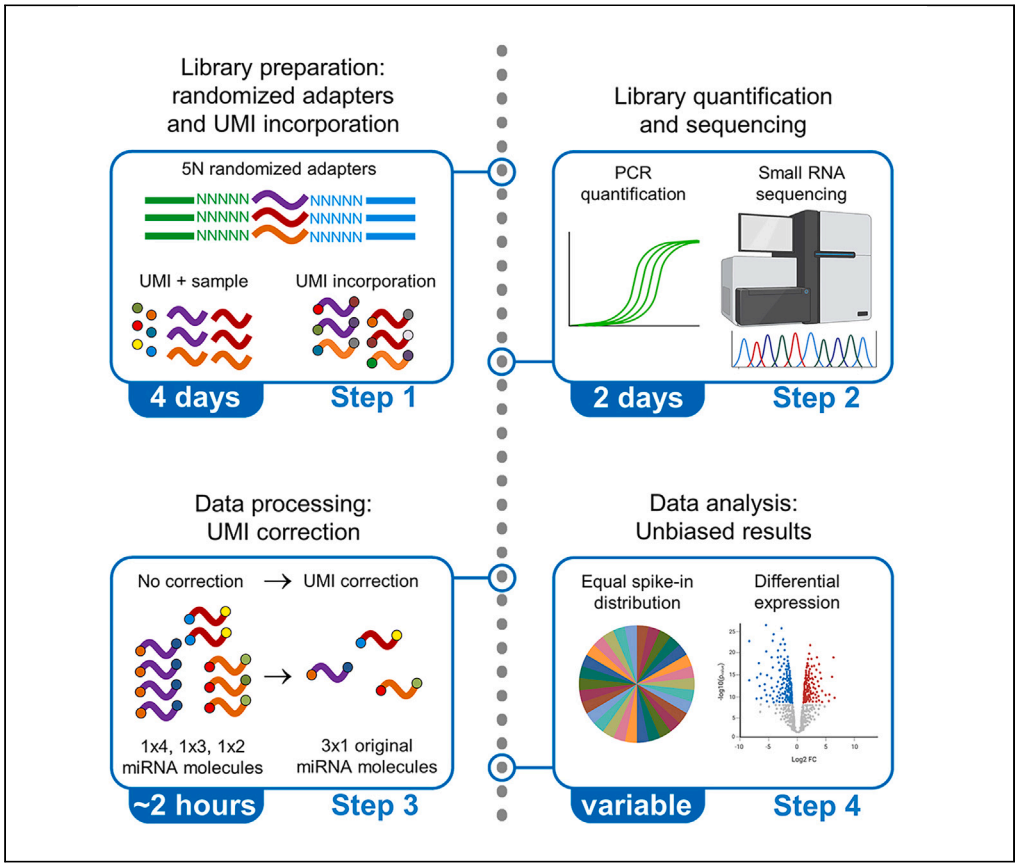
STAR Protocols

CellPress
OPEN ACCESS

Protocol

IsoSeek for unbiased and UMI-informed sequencing of miRNAs from low input samples at single-nucleotide resolution



Besides canonical microRNAs (miRNAs), sequence-based variations called isomiRs have biological relevance and diagnostic potential; however, accurate calling of these post-transcriptional modifications is challenging, especially for low input samples. Here, we present IsoSeek, a sequencing protocol that reduces ligation and PCR amplification bias and improves the accuracy of miRNA detection in low input samples. We describe steps for using randomized adapters combined with unique molecular identifiers, library quantification, and sequencing, followed by detailed procedures for data processing and analysis.

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

Monique A.J. van Eijndhoven, Chantal Scheepbouwer, Ernesto Aparicio-Puerta, Michael Hackenberg, D. Michiel Pegtel

m.vaneijndhoven@amsterdamumc.nl (M.A.J.v.E.)
d.pegtel@amsterdamumc.nl (D.M.P.)

Highlights

Accurate detection of miRNAs in low input samples

Accurate determination of isomiRs with single-nucleotide resolution

Reduction of adapter-miRNA ligation bias using 5N-randomized adapters

UMI-corrected reduction of PCR amplification bias for extracellular vesicle miRNAs

van Eijndhoven et al., STAR Protocols 4, 102645
December 15, 2023 © 2023 The Authors.
https://doi.org/10.1016/j.xpro.2023.102645

STAR Protocols

CellPress
OPEN ACCESS

Protocol

IsoSeek for unbiased and UMI-informed sequencing of miRNAs from low input samples at single-nucleotide resolution

Monique A.J. van Eijndhoven,[1,2,9,*] Chantal Scheepbouwer,[1,3,4] Ernesto Aparicio-Puerta,[5,6,7,8] Michael Hackenberg,[5,6,7,8] and D. Michiel Pegtel[1,2,10,*]

[1]Amsterdam UMC, Location Vrije Universiteit Amsterdam, Department of Pathology, 1081 HV Amsterdam, the Netherlands

[2]Cancer Center Amsterdam, Imaging and Biomarkers, 1081 HV Amsterdam, the Netherlands

[3]Amsterdam UMC, Location Vrije Universiteit Amsterdam, Department of Neurosurgery, 1081 HV Amsterdam, the Netherlands

[4]Cancer Center Amsterdam, Cancer Biology, 1081 HV Amsterdam, the Netherlands

[5]Department of Genetics, Faculty of Science, University of Granada, 18071 Granada, Spain

[6]Bioinformatics Laboratory, Biotechnology Institute, Centro de Investigacion Biomedica, PTS, Avda. del Conocimiento s/n, 18100 Granada, Spain

[7]Instituto de Investigacion Biosanitaria ibs.GRANADA, University of Granada, 18071 Granada, Spain

[8]Excellence Research Unit ''Modelling Nature'' (MNat), University of Granada, 18071 Granada, Spain

[9]Technical contact

[10]Lead contact

*Correspondence: m.vaneijndhoven@amsterdamumc.nl (M.A.J.v.E.), d.pegtel@amsterdamumc.nl (D.M.P.)
https://doi.org/10.1016/j.xpro.2023.102645

SUMMARY

Besides canonical microRNAs (miRNAs), sequence-based variations called isomiRs have biological relevance and diagnostic potential; however, accurate calling of these post-transcriptional modifications is challenging, especially for low input samples. Here, we present IsoSeek, a sequencing protocol that reduces ligation and PCR amplification bias and improves the accuracy of miRNA detection in low input samples. We describe steps for using randomized adapters combined with unique molecular identifiers (UMI), library quantification, and sequencing, followed by detailed procedures for data processing and analysis. For complete details on the use and execution of this protocol, please refer to C. Gómez-Martín et al. (2023)[1] and Van Eijndhoven et al. (2021).[2]

BEFORE YOU BEGIN

Extracellular vesicle-associated miRNAs are considered as a unique biosource for minimally-invasive diagnostics.[3–5] Small RNA sequencing is the method of choice to identify canonical miRNAs and its sequence variants, called isomiRs. However, most, if not all, sequencing strategies introduce bias especially when input material is low.[6,7] In the following section, we provide a step-by-step protocol that is specifically optimized for miRNA and isomiR detection in extracellular vesicles from plasma. Plasma extracellular vesicles (plasma EV) are isolated from 1 mL plasma using size exclusion chromatography (SEC),[8] followed by RNA isolation using the miRNeasy Serum/Plasma Kit (QIAGEN). ''IsoSeek'' diverges from conventional methods by using randomized 5′- and 3′-adapters to reduce ligation bias and bias towards GC-content. The randomized nucleotides (10N in total, giving more than 1 million different combinations) incorporated in the adapters are subsequently used as a unique molecular identifier (UMI), reducing the PCR amplification bias that is specifically problematic when using low input samples, e.g., plasma EV.[2,9]

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

This procedure was optimized for plasma EV but can be used for other low input samples, including EV derived from urine, peritoneal fluid and cell culture supernatant, total plasma and total urine. Additionally, with some adjustments mentioned below in the protocol, the method can be used for high input samples, e.g., cellular RNA, as well.

Institutional permissions

Institutional permission for obtaining and studying clinical samples is needed as well as informed consent from each individual patient or healthy donor. For the results described in this protocol, samples were collected through biobanking in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines.

Prepare adapters

Timing: 30 min

1. Pre-cool a microcentrifuge at 4°C.
2. Work on ice.
3. Dissolve the lyophilized 5′- and 3′-randomized-adapters (from here on referred to as 5′-5N-adapter and 3′-5N-adapter) in Nuclease free water to a concentration of 100 μM.
   a. Spin the tube containing the lyophilized adapters briefly at 4°C to collect them at the bottom of the tube.
   b. Add Nuclease free water to a concentration of 100 μM.
   c. Incubate for a few minutes at 20°C–22°C.
   d. Vortex for 15 s.
   e. Store adapter stocks at −80°C and/or continue with steps 4 and 5.
4. Dilute the 5′-5N-adapter to the appropriate concentration. For low input samples, e.g., plasma EV, this is 225 nM.
   a. Prepare a 11.25 μM dilution: add 11.25 μL 100 μM 5′-5N-adapter to 88.75 μL Nuclease free water.
   b. Prepare a 225 nM dilution: add 5 μL 11.25 μM 5′-5N-adapter to 245 μL Nuclease free water.
   c. Store aliquots at −80°C.
5. Dilute the 3′-5N-adapter to the appropriate concentration. For low input samples, e.g., plasma EV, this is 100 nM.
   a. Prepare a 5 μM dilution: add 5 μL 100 μM 3′-5N-adapter to 95 μL Nuclease free water.
   b. Prepare a 100 nM dilution: add 5 μL 5 μM 3′-5N-adapter to 245 μL Nuclease free water.
   c. Store aliquots at −80°C.

⚠ CRITICAL: For the random nucleotides in the 5N-adapters, the amidites proportions were taken into consideration to limit bias introduced by differences in coupling efficiencies of each amidite. Furthermore, to prevent synthesis errors that can negatively affect the sequencing results,[1] both adapters were dual HPLC (RP + IEX) purified followed by a quality control using MALDI-TOF MS. See key resources table for sequences and modifications.

Note: Adapter stocks and dilutions are preferably prepared in a PCR cabinet to avoid contamination.

Note: Adapter stocks can be stored at −80°C for at least 5 years, but freeze-thawing steps should be minimized.

Note: Prepare aliquots for single use to avoid freeze-thawing, e.g., when preparing 24 libraries make aliquots of 26 μL.

Optional: When preparing libraries from cellular RNA or other high input samples, use 5.63 μM 5′-5N-adapter and 2.5 μM 3′-5N-adapter. Prepare a 5.63 μM dilution of the 5′-5N-adapter by

adding 50 μL 11.25 μM 5′-5N-adapter to 50 μL Nuclease free water. Prepare a 2.5 μM dilution of the 3′-5N-adapter by adding 50 μL 5 μM 3′-5N-adapter to 50 μL Nuclease free water. Store aliquots at −80°C.

Prepare IsomiR spike-in pool

Timing: 2 h

6. Pre-cool a microcentrifuge at 4°C.
7. Work on ice.
8. Dissolve all lyophilized isomiR spike-ins (individually) in Nuclease free water to a concentration of 100 μM.
   a. Spin the tubes briefly at 4°C to collect the RNA at the bottom of the tube.
   b. Add Nuclease free water to a concentration of 100 μM.
   c. Incubate for a few minutes at 20°C–22°C.
   d. Vortex for 15 s.
   e. Store stocks at −80°C and/or continue with step 9.
9. Prepare a pool with all 26 isomiR spike-ins at a final concentration of 100 pM for each isomiR spike-in.
   a. Add 5 μL of each individual 100 μM isomiR spike-in to 120 μL Nuclease free water. This pool contains 2 μM of each isomiR spike-in.
   b. Prepare a 1:20 dilution: add 20 μL of the 2 μM pool to 380 μL Nuclease free water. This diluted pool now contains 100 nM of each isomiR spike-in.
   c. Prepare a 1:10 dilution: add 10 μL of the 100 nM pool to 90 μL Nuclease free water. This diluted pool now contains 10 nM of each isomiR spike-in.
   d. Prepare a 1:10 dilution: add 10 μL of the 10 nM pool to 90 μL Nuclease free water. This diluted pool now contains 1 nM of each isomiR spike-in.
   e. Prepare a 1:10 dilution: add 25 μL of the 1 nM pool to 225 μL Nuclease free water. This diluted pool now contains 100 pM of each isomiR spike-in.
   f. Store aliquots at −80°C.

⚠ CRITICAL: To prevent synthesis errors that can negatively affect the sequencing results,[1] all isomiR spike-ins were dual HPLC (RP + IEX) purified followed by a quality control using MALDI-TOF MS. See key resources table for sequences and modifications.

Note: IsomiR spike-in stocks and dilutions are preferably prepared in a PCR cabinet to avoid contamination.

Note: IsomiR stocks can be stored at −80°C for at least 5 years, but freeze-thawing steps should be minimized.

Note: Prepare aliquots for single use to avoid freeze-thawing, e.g., when preparing 24 libraries make aliquots of 26 μL.

Prepare SR RT primer

Timing: 10 min

10. Work on ice.
11. Dilute the SR RT primer (part of NEBNext Multiplex Small RNA Library Prep Set for Illumina) 1:50 for low input samples, e.g., plasma EV.
    a. Prepare a 1:50 dilution by adding 5 μL SR RT primer to 245 μL Nuclease free water.
    b. Store aliquots at −20°C.

Figure 1. Dried PEG 8000 tube

Note: Prepare aliquots for single use to avoid freeze-thawing, e.g., when preparing 24 libraries make aliquots of 26 μL.

Optional: When preparing libraries from cellular RNA or other high input samples, use 1:2 diluted SR RT primer. Prepare a 1:2 dilution of the SR RT primer by adding 50 μL SR RT primer to 50 μL Nuclease free water.

Prepare PEG 8000

Timing: 2.5 h

12. Add 4 μL 50% PEG (Polyethylene Glycol) 8000 to a PCR tube.
13. Dry using a SpeedVac for 2 h at 35°C (see Figure 1).
14. Store at 20°C–22°C until further use.

CRITICAL: The PEG 8000 stock solution is very viscous, which may affect the pipetting accuracy and lead to a subsequent decrease of adapter ligation efficiency.[10]

Note: Once dried, the PEG pellet at the bottom of the tube will take up 1 μL of the volume in the first step of the protocol (Library Preparation).

Note: Store dried PEG tubes for a maximum of 1 month.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
RNA from plasma extracellular vesicles	N/A	N/A
Chemicals, peptides, and recombinant proteins		
RNaseZAP	Sigma-Aldrich	Cat#R2020-250ML
PEG 8000 (part of T4 RNA ligase 2, truncated K227Q)	New England Biolabs	Cat#M0351S (B1004SVIAL)

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Novex TBE Running Buffer (5×)	Thermo Fisher Scientific	Cat#LC6675
Ethidium bromide	Merck	Cat#1116080030
GlycoBlue	Thermo Fisher Scientific	Cat#AM9515
Sodium acetate (3 M), pH 5.5, RNase-free	Thermo Fisher Scientific	Cat#AM9740
100% ethanol	Merck	Cat#1009832500
Nuclease-free water	New England Biolabs	Cat#B1500S
LightCycler 480 SYBR Green I Master	Roche	Cat#04887352001
Critical commercial assays		
miRNeasy Serum/Plasma Kit	QIAGEN	Cat#217184
TaqMan MicroRNA Reverse Transcription Kit	Thermo Fisher Scientific	Cat#4366597
NEBNext Multiplex Small RNA Library Prep Set for Illumina	New England Biolabs	Cat#E7300/E7580/E7560
Monarch PCR & DNA Cleanup Kit	New England Biolabs	Cat#T1030 S/L
KAPA Library Quantification Kit (KK4854) qPCR master mix optimized for LightCycler 480	Roche	Cat#07960298001
TaqMan Universal PCR Master Mix, no AmpErase UNG	Thermo Fisher Scientific	Cat#4364343
Deposited data		
miRXplore Universal Reference Pool, spike-in and plasma EV data	This paper	SRA PRJNA867189
Oligonucleotides		
QC-PCR RNA hsa-miR-486-5p (RT and PCR primers)	Thermo Fisher Scientific	Cat#4440887 (assay ID 001278)
QC-PCR RNA hsa-miR-21-5p (RT and PCR primers)	Thermo Fisher Scientific	Cat#4440887 (assay ID 000397)
QC-PCR RNA hsa-miR-10b-5p (RT and PCR primers)	Thermo Fisher Scientific	Cat#4440887 (assay ID 002218)
5′-5N-adapter (RNA)	This paper (Eurogentec)	5′-GUUCAGAGUUCUACAGUC CGACGAUCNNNNN-3′
3′-5N-adapter (DNA) with 5′ end adenylation and 3′-end amino modifier C6 modification	This paper (Eurogentec)	5′-rAppNNNNNAGATCGGAAG AGCACACGTCT-NH2-3′
miRXplore Universal Reference Pool	Miltenyi Biotec, sequences available in the study by Giraldez et al.[7]	Cat#130.094.407
IsomiR spike-in 1 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-UACCCGUAAUCUUCAUAAUCCGAG-3′
IsomiR spike-in 2 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-UACCCGUAAUCUUCAUAAUCCGAGA-3′
IsomiR spike-in 3 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-UACCCGUAAUCUUCAUAAUCCGAGAA-3′
IsomiR spike-in 4 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-UACCCGUAAUCUUCAUAAUCCGAGAAA-3′
IsomiR spike-in 5 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-UACCCGUAAUCUUCAUAAUCCGAGU-3′
IsomiR spike-in 6 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-UACCCGUAAUCUUCAUAAUCCGAGUU-3′
IsomiR spike-in 7 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-UACCCGUAAUCUUCAUAAUCCGAGUUU-3′
IsomiR spike-in 8 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-UACCCGUAAUCUUCAUAAUCCGAGAGC-3′
IsomiR spike-in 9 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-UACCCGUAAUCUUCAUAAUCCGAGUGC-3′
IsomiR spike-in 10 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-ACCCGUAAUCUUCAUAAUCCGAG-3′
IsomiR spike-in 11 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-UACCCGUAAUCUUCAUAAUCCG-3′
IsomiR spike-in 12 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-UACCCGUAAUCUUCAUAAUCCGAGGCG-3′
IsomiR spike-in 13 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-UACCCGUAAUCUUCAUAAUCCGAGCGC-3′
IsomiR spike-in 14 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-AAUACCCGUAAUCUUCAUAAUCCGAG-3′
IsomiR spike-in 15 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-UUUACCCGUAAUCUUCAUAAUCCGAG-3′
IsomiR spike-in 16 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-CCUACCCGUAAUCUUCAUAAUCCGAG-3′
IsomiR spike-in 17 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-GGUACCCGUAAUCUUCAUAAUCCGAG-3′
IsomiR spike-in 18 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-CACCCGUAAUCUUCAUAAUCCGAG-3′
IsomiR spike-in 19 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-GACCCGUAAUCUUCAUAAUCCGAG-3′
IsomiR spike-in 20 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-AGUACCCGUAAUCUUCAUAAUCCGAG-3′
IsomiR spike-in 21 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-UGUACCCGUAAUCUUCAUAAUCCGAG-3′
IsomiR spike-in 22 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-CUUACCCGUAAUCUUCAUAAUCCGAG-3′

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
IsomiR spike-in 23 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-CCCGUAAUCUUCAUAAUCCGAGAA-3′
IsomiR spike-in 24 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-CCCGUAAUCUUCAUAAUCCGAGUU-3′
IsomiR spike-in 25 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-ACCCGUAAUCUUCAUAAUCCGAGCC-3′
IsomiR spike-in 26 (RNA) with 5′-phosphate modification	This paper (Eurogentec)	5′-GGUACCCGUAAUCUUC AUAAUCCGAGAC-3′
Library quantification PCR forward primer (part of KAPA Library Quantification Kit)	Roche	5′-AATGATACGGCGACCACCGA-3′
Library quantification PCR reverse primer (part of KAPA Library Quantification Kit)	Roche	5′-CAAGCAGAAGACGGCATACGA-3′
QC-PCR library hsa-miR-486-5p forward primer	This paper (Eurogentec)	5′-TCCTGTACTGAGCTGCCCGGA-3′
QC-PCR library 3′-adapter reverse primer	This paper (Eurogentec)	5′-CAAGCAGAAGACGGCATACGA-3′
hY4 depletion oligo	This paper	5′-SpC3-ATCGGACCAGCCCGTAC TCTGGACTCTAGTCAGTAGCACG ACTAGAGTCCAGAGTACG-3′
hsa-miR-486-5p depletion oligo	This paper	5′-SpC3-CTCAGTACAGGACGTACT CTGGACTCTAGTCAGTAGCACGA CTAGAGTCCAGAGTACG-3′
hsa-miR-92a-3p depletion oligo	This paper	5′-SpC3-ACAAGTGCAATACGTAC TCTGGACTCTAGTCAGTAGCACG ACTAGAGTCCAGAGTACG-3′
Software and algorithms		
sRNAbench	Aparicio-Puerta et al.[11]	https://github.com/bioinfoUGR/sRNAtoolbox
Other		
PCR tubes	VWR	Cat#53509-304
XCell SureLock Mini-Cell Electrophoresis System	Life Technologies	Cat#EI0001
6% Novex TBE PAGE gel 1 mm, 10 well	Life Technologies	Cat#EC6265BOX
BD Microlance 3 hypodermic needle, 21G×1-1/2″, 0.8 mm × 40 mm	Becton Dickinson	Cat#304432
2 mL tubes	Eppendorf	Cat#0030120.094
1.5 mL tubes	Eppendorf	Cat#0030123.328
0.5 mL tubes	Eppendorf	Cat#0030121.023
Carbon steel surgical blades	Swann-Morton	Cat#0105
Spin-X centrifuge tube filters, 0.45 μm pore CA membrane	Costar	Cat#8162
DNA LoBind tubes, 1.5 mL	Eppendorf	Cat#0030108051
Microcentrifuge	N/A	N/A
SpeedVac	N/A	N/A
Thermocycler	N/A	N/A
UV transilluminator	N/A	N/A
Thermoshaker	N/A	N/A
Illumina sequencing platform	N/A	N/A
LightCycler 480 multiwell plate, white	Roche	Cat#04729692001
LightCycler 480	Roche	N/A
PCR plates	VWR	Cat#732-2389
TaqMan 7500 fast	Applied Biosystems	N/A

MATERIALS AND EQUIPMENT

1× Novex TBE Running Buffer

Reagent	Final concentration	Amount
Novex TBE Running Buffer (5×)	1×	200 mL
Demineralized water	N/A	800 mL
Total	**N/A**	**1 L**

Prepare fresh.

80% Ethanol		
Reagent	Final concentration	Amount
Ethanol (100%)	80%	40 mL
Nuclease free water	N/A	10 mL
Total	**N/A**	**50 mL**

Prepare fresh.

STEP-BY-STEP METHOD DETAILS

Library Preparation

Timing: 1 day

During this step, the 3′-5N-adapter is ligated to the RNA molecules, followed by RT primer hybridization, ligation of the 5′-5N-adapter, RT-PCR, and PCR amplification of the cDNA products. At the end of the day, you will have purified PCR libraries, ready to use for size selection.

1. Clean all surfaces and pipets with RNase ZAP.
2. Prepare an ice bucket for all reagents and reaction mixtures, work on ice until the PCR amplification step (step 7).
3. 3′-5N-adapter ligation.
   a. Pre-heat a thermocycler to 70°C.
   b. Take a PCR tube with dried PEG 8000 and add 4 μL plasma EV RNA, resuspend carefully. The PEG pellet at the bottom of the tube will take up 1 μL of the total volume.
   c. Add 1 μL isomiR spike-in pool (100 pM = 100 amol each).
   d. Add 1 μL 3′-5N-adapter (100 nM).
   e. Mix by pipetting.
   f. The total volume is 7 μL.
   g. Incubate for 2 min at 70°C in a thermocycler.
   h. Add (pre-mixed):
      i. 10 μL 3′ ligation buffer (part of NEBNext Multiplex Small RNA Library Prep Set for Illumina).
      ii. 3 μL 3′ ligation enzyme mix (part of the NEBNext Multiplex Small RNA Library Prep Set for Illumina).
   i. The total volume is 20 μL.
   j. Incubate for 1 h at 25°C in a thermocycler.

Note: When preparing a pre-mix for multiple samples, prepare 10% more than necessary.

Note: When using less than 4 μL of plasma EV RNA, adjust the total volume with Nuclease free water (step 3f).

Note: Because RNA from low input samples, e.g., plasma EV, is difficult to accurately quantify, it's preferred to use equal volumes of input RNA (= maximum 4 μL). To determine the presence of amplifiable miRNAs prior to library preparation, a QC qRT-PCR can be performed. For this, 3 μL 1:10 diluted RNA is reverse transcribed using the TaqMan MicroRNA Reverse Transcription kit in a multiplex reaction containing RT-primers for hsa-miR-486-5p (high abundance in plasma EV), hsa-miR-21-5p (intermediate abundance) and hsa-miR-10b-5p (low abundance).[8] After cDNA synthesis Nuclease free water is added up to a final volume of 50 μL. 3 μL of cDNA is added to PCR primers for the miRNA targets and the TaqMan Universal PCR master mix. The mixture is subjected to 40 cycles of 95°C for 15 s and 60°C for 1 min on an ABI 7500 Fast system. All samples are measured in duplicate, and data is analyzed using 7500 Software v2.0.6.

Optional: When preparing libraries from cellular RNA or other high input samples, use 200 ng of RNA and 1 μL 2.5 μM 3′-5N-adapter.

4. RT primer hybridization.
   a. Add (pre-mixed):
      i. 4.5 μL Nuclease free water.
      ii. 1 μL SR RT primer (1:50) (part of NEBNext Multiplex Small RNA Library Prep Set for Illumina).
   b. Mix by pipetting.
   c. The total volume is 25.5 μL.
   d. Incubate for 5 min at 75°C, 15 min at 37°C, 15 min at 25°C in a thermocycler.

Note: When preparing a pre-mix for multiple samples, prepare 10% more than necessary.

Optional: When preparing libraries from cellular RNA or other high input samples, use 1:2 diluted SR RT primer.

5. 5′-5N-adapter ligation.
   a. Pre-heat a thermocycler to 70°C.
   b. Transfer 1 μL of 5′-5N-adapter (225 nM) to a clean PCR tube.
   c. Denature the 5′-5N-adapter for 2 min at 70°C.
   d. Place on ice.
   e. Add 1 μL of the denatured 5′-5N-adapter (225 nM) to the reaction mixture from step 4d.
   f. Add 1 μL 5′ ligation buffer (part of NEBNext Multiplex Small RNA Library Prep Set for Illumina).
   g. Add 2.5 μL 5′ ligation enzyme mix (part of NEBNext Multiplex Small RNA Library Prep Set for Illumina).
   h. Mix by pipetting.
   i. The total volume is 30 μL.
   j. Incubate for 1 h at 25°C in a thermocycler.

Note: When preparing 5′-5N-adapter for multiple samples, prepare and denature 10% more than necessary.

Optional: When preparing libraries from cellular RNA or other high input samples, use 1 μL 5.63 μM 5′-5N-adapter.

6. Reverse transcription.
   a. Add (pre-mixed):
      i. 8 μL First strand synthesis reaction buffer (part of NEBNext Multiplex Small RNA Library Prep Set for Illumina).
      ii. 1 μL Murine RNase inhibitor (part of NEBNext Multiplex Small RNA Library Prep Set for Illumina).
      iii. 1 μL ProtoScript II Reverse Transcriptase (part of NEBNext Multiplex Small RNA Library Prep Set for Illumina).
   b. Mix by pipetting.
   c. The total volume is 40 μL.
   d. Incubate for 1 h at 50°C in a thermocycler.

Note: When preparing a pre-mix for multiple samples, prepare 10% more than necessary.

7. PCR amplification.
   a. Add 5 μL Nuclease free water.
   b. Add 2.5 μL SR primer for Illumina (part of NEBNext Multiplex Small RNA Library Prep Set for Illumina).

c. Add 2.5 μL INDEX (X) primer (part of NEBNext Multiplex Small RNA Library Prep Set for Illumina).
d. Add 50 μL LongAmp Taq 2× master mix (part of NEBNext Multiplex Small RNA Library Prep Set for Illumina).
e. Mix by pipetting.
f. The total volume is 100 μL.
g. Incubate:

PCR cycling conditions			
Steps	Temperature	Time	Cycles
Initial Denaturation	94°C	30 s	1
Denaturation	94°C	15 s	20 cycles
Annealing	62°C	30 s	
Extension	70°C	15 s	
Final extension	70°C	5 min	1
Hold	4°C	forever	

⚠ CRITICAL: Decide in advance which samples will be sequenced simultaneously in a single sequence run to make sure each sample has a unique index that can be de-multiplexed afterwards. Depending on which NEBNext Multiplex Small RNA Library Prep Set for Illumina is purchased, a total of 48 INDEX primers are available, "X" refers to the number of the index primer.

Optional: When preparing libraries from cellular RNA or other high input samples, reduce the number of PCR cycles to 15.

8. PCR cleanup and concentration, according to the Monarch PCR & DNA Cleanup Kit.
   a. Transfer the 100 μL PCR amplified sample from step 7g to a 1.5 mL tube.
   b. Add 700 μL DNA Cleanup Binding Buffer and mix by pipetting.
   c. Insert the column into the collection tube.
   d. Load the 800 μL sample onto the column.
   e. Centrifuge for 1 min, discard the flow-through (this is waste).
   f. Re-insert the column into the collection tube.
   g. Add 200 μL DNA Wash Buffer.
   h. Centrifuge for 1 min, discard the flow-through (this is waste).
   i. Re-insert the column into the collection tube.
   j. Add 200 μL DNA Wash Buffer.
   k. Centrifuge for 1 min, discard the flow-through (this is waste).
   l. Transfer the column to a clean 1.5 mL tube.
   m. Add 25 μL Nuclease free water to the center of the matrix.
   n. Incubate at 20°C–22°C for 1 min.
   o. Centrifuge for 1 min to elute the cDNA.
   p. Store the DNA eluate at −80°C until further use.

⚠ CRITICAL: Make sure no residual buffer is left in or attached to the outlet of the column after the second centrifugation step with DNA wash buffer (step 8k). This buffer contains ethanol that may interfere with the sample loading into the gel in the next phase of the protocol. If residual buffer is present, add an additional centrifugation step.

Note: All centrifugation steps are performed at maximum speed at 20°C–22°C in a microcentrifuge.

Table 1. Example of sample loading using a Novex TBE gel

Gel 1	Marker	Sample 1	Sample 1	Marker	Sample 2	Sample 2	Marker	Sample 3	Sample 3	Empty
Gel 2	Empty	Sample 4	Sample 4	Marker	Sample 5	Sample 5	Marker	Sample 6	Sample 6	Marker

Note: Often, less than 25 μL sample will be recovered from step 8o. This doesn't have any negative effect on the subsequent steps of the protocol.

Pause point: Samples can be stored for 16 h or up to 1 week at −80°C.

Gel size selection and purification

Timing: 3 days

In this step the adapter-ligated miRNAs are separated from unwanted products, e.g., adapter-dimers, single adapters and/or larger RNA molecules. After elution and precipitation the libraries are ready for quantification, pooling and sequencing.

9. Gel size selection.
   a. Take 2 × 6% Novex TBE PAGE gel 1 mM 10-well, remove the white strip and comb, rinse the wells with demineralized water.
   b. Indicate the position of the different samples if preferred.
   c. Place the gels in an XCell SureLock Mini-Cell Electrophoresis System and fill both inner and outer chambers with 1× Novex TBE Running Buffer.
   d. Prepare samples:
      i. Use 25 μL sample from step 8p and add 5 μL Gel Loading Dye (part of NEBNext Multiplex Small RNA Library Prep Set for Illumina).
      ii. Use 12.5 μL Nuclease free water and add 2.5 μL Gel Loading Dye for empty wells.
      iii. Use 5 μL of Quick-Load pBR322 DNA-MspI Digest (ready-to-use, part of NEBNext Multiplex Small RNA Library Prep Set for Illumina) as a molecular weight marker to guide size selection.
   e. Load samples and marker (see Table 1).
      i. Each sample is divided in 2 wells, to prevent overloading and obtain a better size separation.
      ii. Different samples are separated by either an empty well or the molecular weight marker to prevent cross-contamination during the excision of the correct sized bands.
   f. Run the gels for 1 h and 15 min at 120 V.
   g. Carefully remove the gels from the cassette and place them in 1× Novex TBE Running Buffer.
   h. Incubate the gel for 2–3 min in 1× Novex TBE Running Buffer (50 mL) with 1:20000 Ethidium Bromide (2.5 μL). Leave the second gel in buffer without Ethidium Bromide until ready for processing.
   i. Image the gel using a UV transilluminator.
   j. Excise the band corresponding to adapters containing 22 nt miRNA inserts (expected size ~151 nt) as tight as possible using a carbon steel surgical blade (see Figure 2). Avoid including adapter-dimers as much as possible (expected size of adapter-dimers 129 nt). The sample that was divided in 2 wells can be excised as 1 band and processed as 1 sample in the subsequent steps.
   k. For documentation image the gel again after excision of the correct band.
   l. Place the gel piece in a gel breaker tube; this is a 0.5 mL tube punctured at the bottom with a 21 gauge needle and subsequently placed in a 2 mL tube (see Figure 3).

⚠ CRITICAL: caution is warranted when using Ethidium Bromide (potentially carcinogenic) and UV light. Wear protective clothing, gloves, cover all exposed skin and use a face shield.

CellPress
OPEN ACCESS

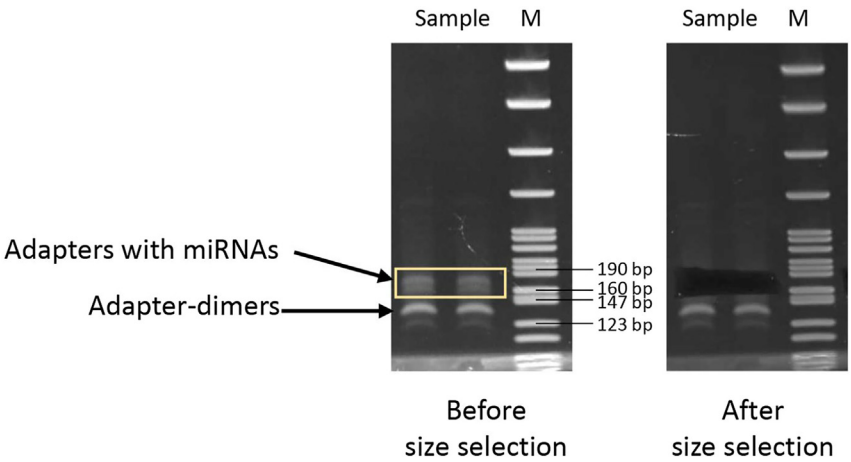


Figure 2. Example of a Novex TBE gel before and after size selection of the adapter-ligated miRNA products

Note: Running the gels for 1 h and 15 min at 120 V gives the most optimal separation between adapter-ligated 22 nt miRNA inserts and unwanted adapter-dimers.

Note: The presence of PEG affects the sample running through the gel, the correct size bands are present slightly higher than expected (~160–165 nt instead of the expected ~151 nt).

Note: Longer incubation times of the gels with Ethidium Bromide will not increase signal intensity for low input samples, but may increase sample loss.

Optional: Instead of performing a manual gel size selection, an automated size selection with the BluePippin can be used (Sage Science, 3% Agarose gel cassette), following the manufacturer's instructions.

Optional: Different electrophoresis systems and corresponding gels may be used, make sure to optimize the running time to get the most optimal separation between adapter-ligated 22 nt miRNA inserts and unwanted adapter-dimers.

Troubleshooting: problem 1; High level of adapter-dimers.

10. Elution of the libraries from the gel.
    a. Centrifuge the sample from step 9L for 2 min at maximum speed in a microcentrifuge at 20°C–22°C.
    b. Make sure all gel pieces passed the punctured 0.5 mL tube into the 2 mL tube.
    c. Add 250 μL DNA Gel Elution buffer (part of NEBNext Multiplex Small RNA Library Prep Set for Illumina).
    d. Shake for 16 h at 1400 rpm at 20°C–22°C in a thermoshaker.

Note: Do not incubate longer than 24 h.

11. Precipitation of the libraries.
    a. Transfer the eluate and gel debris from step 10d to a Spin-X Centrifuge Tube Filter with 0.45 μm Pore CA Membrane, use filter-tips with a small piece removed from the end of the tip.
    b. Centrifuge for 2 min at maximum speed at 20°C–22°C.
    c. Transfer the DNA containing eluate to a DNA LoBind Eppendorf tube.

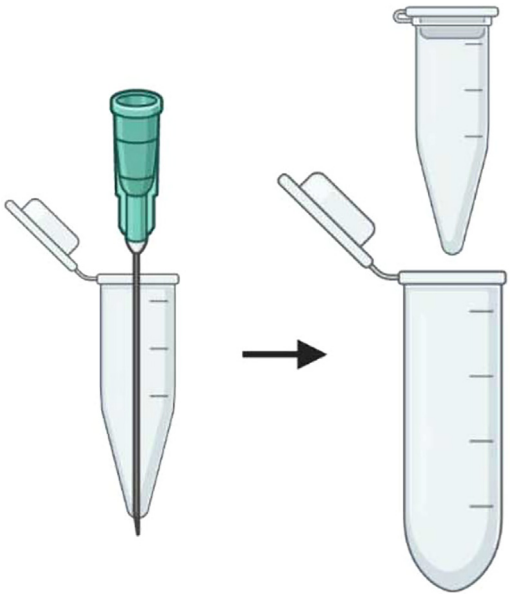


Figure 3. Gel breaker tube; 0.5 mL tube punctured at the bottom with a 21 gauge needle and subsequently placed in a 2 mL tube

d. Add 1 μL Linear Acrylamide (part of NEBNext Multiplex Small RNA Library Prep Set for Illumina), mix by flicking the tube.
e. Add 1 μL GlycoBlue, mix by flicking the tube.
f. Add 25 μL 3 M NaAc (pH 5.5, from −20°C), mix by flicking the tube.
g. Add 750 μL 100% Ethanol (from −20°C), mix by inverting the tube.
h. Precipitate the sample for 16–24 h at −80°C.
i. Pre-cool a microcentrifuge to 4°C.
j. Centrifuge for 30 min at 14000 g at 4°C.
k. Remove the supernatant as much as possible without disturbing the pellet and discard.
l. Wash the pellet with 1 mL cold freshly prepared 80% Ethanol, vortex briefly until the pellet starts "floating".
m. Centrifuge for 30 min at 14000 g at 4°C.
n. Remove the supernatant as much as possible without disturbing the pellet and discard.
o. Centrifuge for an additional 2 min at 14000 g at 4°C.
p. Remove the remaining supernatant and discard.
q. Air dry the pellet for 5–10 min at 20°C–22°C with open lid.
r. Resuspend the pellet in 10 μL Nuclease free water.
s. Make a 1:10 dilution in Nuclease free water.
t. Store the undiluted libraries at −80°C until preparation of the library pool.
u. Store the 1:10 diluted libraries at −80°C until library quantification and quality controls.

⚠ CRITICAL: caution is warranted when using Linear Acrylamide and NaAc. Wear protective clothing and gloves, work in a fume hood when possible.

Pause point: The precipitation step at −80°C (step 11h) can be extended up to 3 days.

Library quantification and quality control

Timing: 5 h

It is recommended to use the KAPA Library Quantification method (Roche) for quantification of low input plasma EV libraries. Quantification with bioanalyzer, TapeStation or fragment-analyzer results in 3 to 10-fold higher library concentrations compared to the KAPA kit, which may overstate the true library yield. More importantly, using the KAPA Library Quantification method (Roche) to prepare equimolar library pools ultimately results in less variation in sequencing yield between libraries. A quality control PCR using a forward primer directed against a miRNA (in this case hsa-miR-486-5p) and a reverse primer against the 3′-adapter is additionally used to determine the presence of adapter-ligated miRNAs.

12. Library quantification using KAPA PCR.
    a. Take the 1:10 diluted libraries from step 11u.
    b. Prepare 10-fold dilutions in Nuclease free water:
        i. Add 2 μL of the 1:10 diluted libraries to 18 μL Nuclease free water. This is now a 1:100 dilution.
        ii. Add 2 μL of the 1:100 diluted libraries to 18 μL Nuclease free water. This is now a 1:1000 dilution.
        iii. Add 2 μL of the 1:1000 diluted libraries to 18 μL Nuclease free water. This is now a 1:10000 dilution.
    c. The 1:1000 and 1:10000 dilutions are used for quantification using the KAPA Library Quantification Kit optimized for LightCycler 480.
    d. Prepare the Master Mix for the KAPA PCR:
        i. Add 1 mL Quantification Primer Premix for Illumina (10×) to 5 mL KAPA SG Fast.
        ii. Make 1 mL aliquots.
        iii. Store at −20°C.
    e. Prepare the PCR master mix:

PCR reaction master mix

Reagent	Amount
DNA template	2 μL
Master Mix for KAPA PCR	6 μL
Nuclease free water	2 μL

    f. Prepare the LightCycler 480 Multiwell plate:
        i. The DNA standard 1–6 (from the kit) is measured in triplicate.
        ii. The samples are measured in duplicate, both 1:1000 and 1:10000 dilutions.
    g. Run the PCR using "absolute quantification" and according to the following settings:

PCR cycling conditions

Steps	Temperature	Time	Cycles
Initial Denaturation	95°C	5 min	1
Denaturation	95°C	30 s	35 cycles
Annealing+Extension	60°C	45 s	
Hold	4°C	Forever	

    h. Calculation of the library yield is done according to the manufacturer's procedure: https://elabdoc-prod.roche.com/eLD/api/downloads/934c8a1e-670b-ee11-1d91-005056a71a5d?countryIsoCode=pi.

Optional: Use a different qPCR machine and a corresponding KAPA PCR equivalent. Additionally, smaller sized kits are available. Besides the commercially available KAPA Library Quantification kits, it's possible to design/make a standard curve yourself and use this in

combination with the Library quantification PCR Forward and Reverse primers (sequences are enlisted in the key resources table).

Optional: The KAPA PCR products can be subjected to a quality control using agarose gel analysis to check for the presence of adapter-dimers.

13. Quality control PCR.
    a. Take the 1:1000 diluted libraries from step 12b.
    b. Prepare the PCR master mix:

PCR reaction master mix	
Reagent	Amount
DNA template	2 μL
QC-PCR library hsa-miR-486-5p Forward primer (10 μM)	1 μL
QC-PCR library 3′-adapter Reverse primer (10 μM)	1 μL
LightCycler 480 SYBR Green I Master	5 μL
Nuclease free water	1 μL

    c. Prepare the LightCycler 480 Multiwell plate.
    d. The samples are measured in duplicate, use the 1:1000 dilution.
    e. Run the PCR using "absolute quantification" and according to the following settings: PCR cycling conditions

Steps	Temperature	Time	Cycles
Initial Denaturation	95°C	10 min	1
Denaturation	95°C	10 s	40 cycles
Annealing	60°C	15 s	
Extension	72°C	15 s	
Melting curve (optional)	95°C	10 s	1
Melting curve (optional)	55°C	30 s	1
Melting curve (optional)	95°C	N/A	1
Hold	4°C	Forever	

Optional: Use a different qPCR machine and a corresponding SYBR Green equivalent.

Optional: Use additional or other miRNA targets.

Preparation of the library pool and small RNA sequencing

Timing: 1 h (plus turnover time of the sequence run)

In this step equimolar library pools are prepared, followed by sequencing, data processing and analysis.

14. Preparation of an equimolar library pool.
    a. Take the undiluted libraries from step 11t and use the molarities calculated in step 12h.
    b. Dilute the libraries if necessary.
        i. Make sure not to pipette volumes less than 1 μL.
    c. Prepare the pool by adding 1.5–2 fmol of each library.
    d. Add Nuclease free water to reach a final concentration of 1.5–2 nM.
    e. If necessary use a SpeedVac to reach the desired concentration.

15. Small RNA sequencing.

The library pool can be sequenced on an Illumina NovaSeq 6000 platform, or any patterned flow cell (HiSeq 4000, NextSeq 2000), at Paired-End 50 bp. A maximum concentration is loaded in the presence of 10% PhiX to compensate for the low complexity of the plasma EV samples. Aim for 10 million reads per sample, the number of samples in the pool can be adjusted depending on the sequencing platform.

16. Data processing and UMI correction using sRNAbench.[11]
    a. Follow the instructions to install and run the sRNAtoolbox locally https://github.com/bioinfoUGR/sRNAtoolbox.
    b. Run the following command for each of the sequenced samples:

```
sRNAbench input={input}.fq.gz output={output_directory} microRNA=hsa adapter=AGATCGGAA-
GAGCACACGT           tRNA=GRCh38_p10_genomic_tRNA.fa#2#19#v#-a           libs=hsa_yRNA
libs=GRCh38_p10_ncRNA libs=GRCh38_p10_RNAcentral libs=GRCh38_p10_cdna mm=1 alignType=n
isoMiR=true plotMiR=false spikeExact=true spikeExact=true dbPath=/opt/sRNAtoolboxDB spi-
keIn=spike-in_sequences.fa
```

17. Generate miRNA and isomiR summary and expression files:

```
sRNAde   input={input_directory}   output={output_directory}   grpString=sample1,sam-
ple2,[...]sampleN  matrixDesc=condition1,condition2,[...]conditionN  dbPath=/opt/sRNA-
toolboxDB/ iso=true

sRNAde   input={input_directory}   output={output_directory}   grpString=sample1,sam-
ple2,[...]sampleN  matrixDesc=condition1,condition2,[...]conditionN  dbPath=/opt/sRNA-
toolboxDB/ isoSummary=true diffExpr=false

sRNAde   input={input_directory}   output={output_directory}   grpString=sample1,sam-
ple2,[...]sampleN  matrixDesc=condition1,condition2,[...]conditionN  dbPath=/opt/sRNA-
toolboxDB/ isoComparison=true readIso=true diffExpr=false minRCexpr=0 minRCiso=0

sRNAde   input={input_directory}   output={output_directory}   grpString=sample1,sam-
ple2,[...]sampleN  matrixDesc=condition1,condition2,[...]conditionN  dbPath=/opt/sRNA-
toolboxDB/ minRCexpr=0 stat=true statFiles=isomiR_summary.txt colData=2 minRCdata=0 dif-
fExpr=false folderData=stat
```

18. For more information, check sRNAbench's manual: https://arn.ugr.es/srnatoolbox/.

⚠ CRITICAL: Prepare a library pool consisting only of similar or comparable low input, low complexity samples. When mixing low input samples (e.g., plasma EV) with high input samples (e.g., cellular RNA) this may lead to a phenomenon called "index hopping" (Troubleshooting: problem 2; Index hopping).

Note: Sequencing on a non-patterned flow cell (MiSeq, NextSeq 500) or using different setting (Paired-End 100/150 or Single-Read) is possible but it is preferred not to use different platforms or settings when sequencing large patient sample cohorts.

Optional: To determine the quality of the sequence run and possibly correct for batch effects a control library can be added to each library pool. Prepare a large amount of a control small

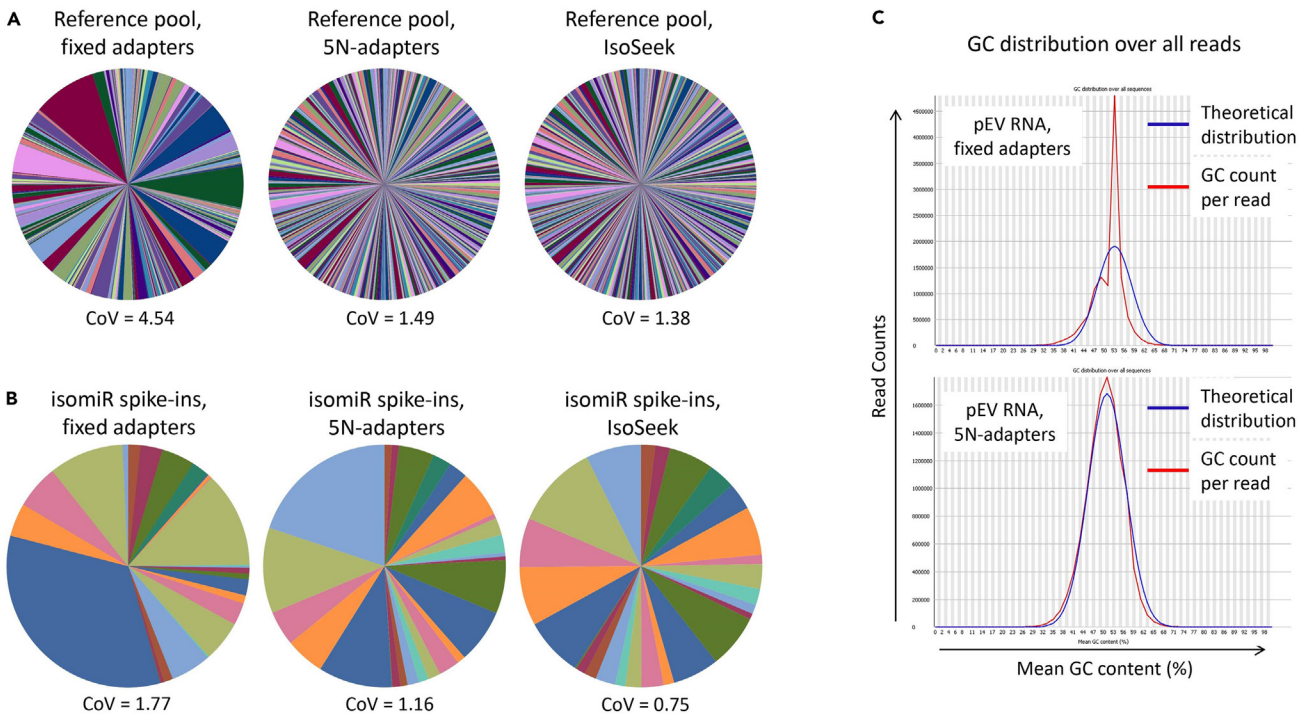


Figure 4. IsoSeek shows reduced ligation bias, PCR amplification bias and bias towards GC content

(A) Distribution of >950 mature miRNAs (miRXplore reference pool) after library preparation using a conventional protocol with fixed adapters (left) and 5N-adapters without (middle) and with UMI correction (IsoSeek, right). Representative data is shown for each procedure including the coefficient of variation (CoV).

(B) Distribution of 26 isomiR spike-ins added to plasma EV RNA prior to library preparation using a conventional protocol with fixed adapters (left) and 5N-adapters without (middle) and with UMI correction (IsoSeek, right). Representative data is shown for each procedure including the coefficient of variation (CoV).

(C) FastQC analysis of the GC content per read after sequencing plasma EV libraries prepared using a conventional protocol with fixed adapters (upper panel) or 5N-adapters (lower panel). The theoretical distribution is shown in blue, the observed distribution in red.

RNA library (preferably from similar input material as the material of interest), that can be added to multiple sequence runs. Store the control library in aliquots at -80°C.

Optional: Data processed by sRNAbench can be downloaded and analyzed according to the user's preferences. One option for normalization and differential expression analysis is the use of NORMSEQ.[12]

Troubleshooting: problem 3; High abundance of specific RNA molecules that are not of interest.

EXPECTED OUTCOMES

IsoSeek was developed to reduce the sources of bias that arise in conventional small RNA sequencing protocols, especially for low input samples, e.g., RNA associated with plasma extracellular vesicles.[1,6,7]

To overcome adapter-miRNA ligation bias, which occurs most prominently when using fixed adapters, we designed adapters with 5 random nucleotides (5N-adapters). As a proof of principle, we prepared libraries from a commercially available miRXplore reference pool consisting of 962 synthetic canonical miRNAs of human, mouse, rat and viral origin that are pooled in equimolar amounts (Miltenyi Biotec). The distribution of the equimolar canonical miRNAs when using a conventional protocol with fixed adapters is skewed with a Coefficient of Variation (CoV) of 4.54 (Figure 4A left), whereas in libraries prepared with our 5N-adapters the distribution is much more equal and the

CoV decreases to 1.49 (Figure 4A middle), suggesting a strong reduction in adapter-miRNA ligation bias. When we subsequently use the 10 random nucleotides (10N) ligated to both sides of the miRNAs as a unique molecular identifier (UMI) we can correct for PCR duplicates. Upon this correction, the CoV decreases to 1.38 (Figure 4A right) suggesting that the amplification bias in high input, high complexity samples is relatively small compared to the ligation bias.

For low input samples, more than 15 amplification cycles are typically necessary to gain sufficient library yield for proper sequencing results. However, introducing more amplification cycles makes these low input libraries more prone to PCR bias.[2,9] Furthermore, not only canonical miRNAs but also isomiRs are biologically relevant and they have diagnostic potential.[13–15] To check the accuracy of IsoSeek in distinguishing isomiRs with single nucleotide resolution in low input samples, we designed a synthetic spike-in set based on a canonical *C. elegans* miRNA and its isomiRs. The isomiR spike-ins are added to a background of plasma EV RNA in equimolar amounts followed by library preparation. Upon preparing libraries with 5N-adapters the distribution of isomiRs improves compared to libraries prepared with fixed adapters, the CoV decreases from 1.77 to 1.16 (Figure 4B left and middle), indicating a reduction in ligation bias. Adding the UMI correction (IsoSeek) gives a further improvement of the distribution, with a CoV dropping to 0.75 (Figure 4B right), indicating that for low input samples it is crucial to correct for PCR amplification bias.

Another common source of bias is sequence composition, particularly a high percentage of GC. When preparing small RNA libraries of plasma EV, the usage of 5N-adapters fits the theoretical distribution better than a conventional protocol with fixed adapters, indicating a reduced bias towards GC content (Figure 4C).

In conclusion, when using IsoSeek for low input samples a reduced ligation bias, PCR amplification bias and bias towards GC content can be expected.

Overall, approximately 600 different miRNAs and on average 25000 isomiRs can be detected in plasma EV when using the described IsoSeek protocol.

LIMITATIONS

The manual gel size selection is time consuming and may depend on the experience of the operator. This step can be automated by using the BluePippin (Sage Science, 3% Agarose gel cassette). Size settings of 140–165 bp give promising results, but further optimization and determining the reproducibility of this method is required for specific needs.

In the current setup of the protocol the number of index primers is limiting (n = 48), so a maximum of 48 libraries can be pooled. This can be circumvented by designing and synthesizing customized index primers. Alternatively unique dual indices can be designed (incorporated in both 5′- and 3′-adapters). With these, the effects caused by index hopping can be filtered out computationally, see troubleshooting: problem 2; index hopping.

TROUBLESHOOTING

Problem 1

High level of adapter-dimers (related to step 9j).

Potential solution

One of the biggest concerns when using low input samples is the generation of adapter-dimers during library preparation steps. When the level of adapter-dimers surpasses the level of adapter-ligated miRNAs, one may adjust the concentration of the 5N-adapters and SR RT primer. To get an optimal ratio of adapters vs. RNA input, we tested a dilution range of the 5N-adapters and SR RT primer: undiluted (5′-5N-adapter = 11.25 μM, 3′-5N-adapter = 5 μM, SR RT primer = undiluted)

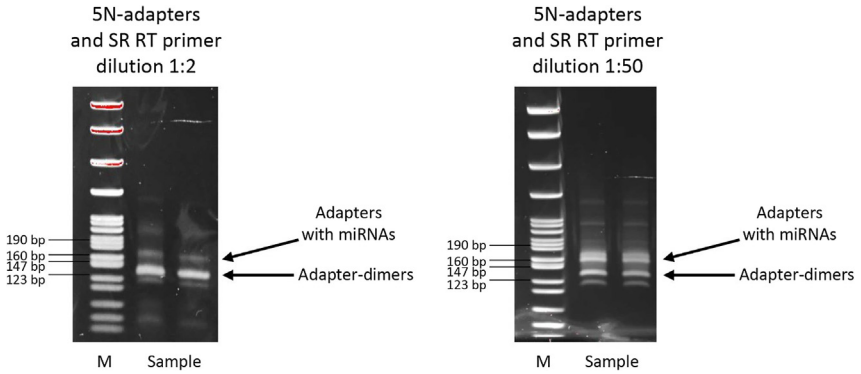


Figure 5. Novex TBE gel of a library prepared with 1:2 diluted 5N-adapters and SR RT primer (5′-5N-adapter 5.63 μM, 3′-5N-adapter 2.5 μM, SR RT primer 1:2 dilution; left) or 1:50 diluted 5N-adapters and SR RT primer (5′-5N-adapter 225 nM, 3′-5N-adapter 100 nM, SR RT primer 1:50 dilution; right)

Using 1:50 diluted 5N-adapters an SR RT primer leads to a reduction of adapter-dimers and an increase of the relative proportion of adapter-ligated miRNAs.

or diluted 1:2, 1:4, 1:10, 1:50 and 1:100. With each dilution-step of the 5N-adapters and SR RT primer the level of adapter-dimers decreases, but with the 1:100 dilution the library yield is borderline sufficient for reliable sequencing results. We concluded that in our hands a 1:50 dilution of the 5′-5N-adapter, 3′-5N adapter and SR RT primer was optimal for both library quality and yield (5′-5N-adapter = 225 nM, 3′-5N-adapter = 100 nM, SR RT primer = 1:50 dilution) (see Figure 5).

Alternatively, the input material can be increased. With patient samples however this is often not possible, we therefore tried the possibility to increase the sequence quality of low input plasma EV libraries and/or enhance the library preparation efficiency by adding yeast "filler" RNA to mimic samples of higher RNA complexity.[6] We used *Saccharomyces cerevisiae* yeast RNA (Thermo Fisher Scientific, Cat#AM7120G)) and either treated it with Shrimp Alkaline Phosphatase (rSAP, New England Biolabs, Cat#M0371S) or we enriched for >200 nt size using the RNA Clean & Concentrator Kit (Zymo Research, Cat#R1015). We added 1 ng of the rSAP treated or size enriched yeast RNA to plasma EV RNA before library preparation. Adding yeast RNA however does not affect the sequence quality, sequence depth, the level of miRNA reads or the number of miRNAs. Because we did not observe any effect (positive or negative) we exclude this from the final protocol.

As an alternative strategy to reduce adapter-dimer formation we employed enzymatic removal of the excess free 3′-5N-adapters after ligation and before hybridization of the RT-primer.[7] We subsequently added 1 μg extreme thermostable single-stranded DNA-binding protein (ET SSB) for 10 min at 25°C, 1 μL 5′-deadenylase for 1 h at 30°C and 1 μL single-stranded DNA-specific exonuclease (RecJf) for 1 h at 30°C (all enzymes from New England Biolabs, Cat#M2401S, M0331S, M0264S). However, in the gel size selection process and after sequencing we did not observe any reduction in adapter-dimer formation, therefore we excluded this step from our final protocol.

Problem 2

Index-hopping (related to step 17).

Potential solution

To determine the reproducibility of our IsoSeek procedure and possible 'batch effects', we re-sequenced several libraries and calculated the log 2-fold change between the technical replicates for each individual miRNA. We noticed that for some samples the reproducibility was high (Figure 6 upper panels) but for some other samples there was a difference in normalized reads for a substantial number of miRNAs (Figure 6 lower panels) (log 2 fold change >2).

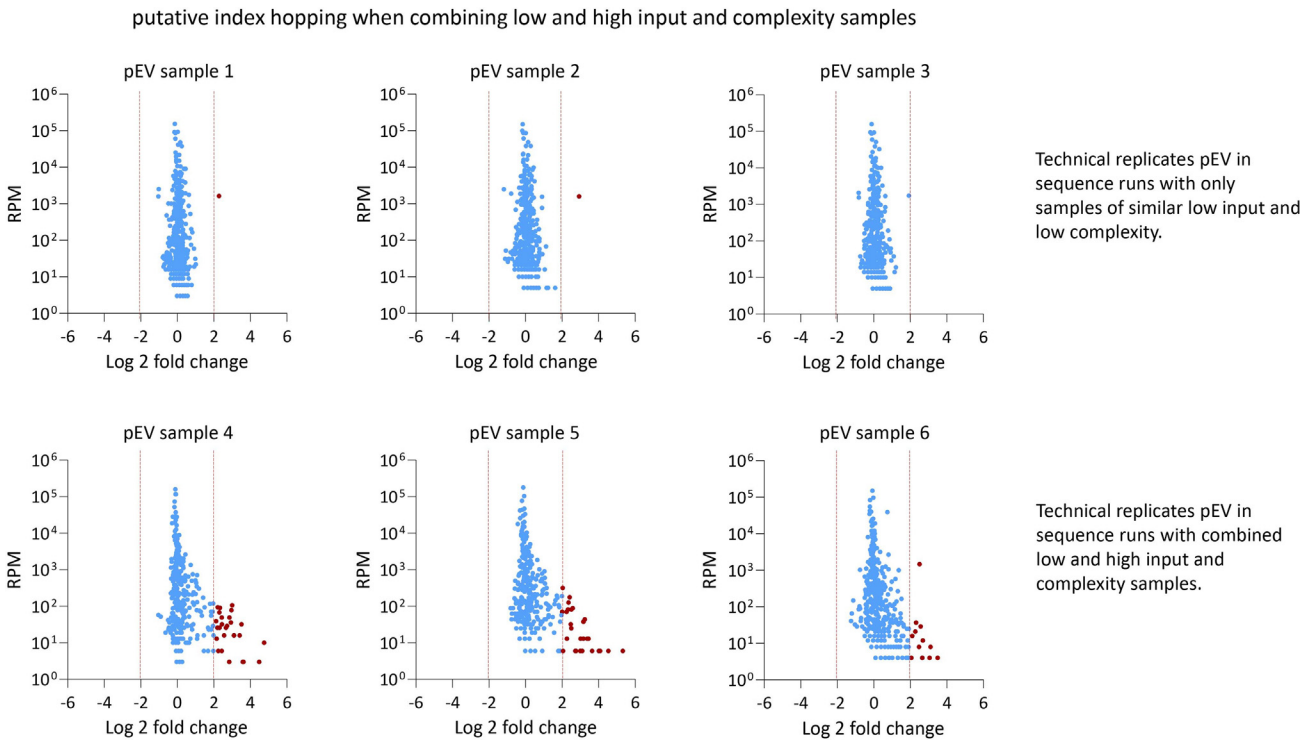


Figure 6. Putative index hopping when combining low and high input and complexity samples

Differential expression of miRNAs of plasma EV libraries prepared with IsoSeek after re-sequencing. In the upper panels both sequenced pools contained plasma EV samples exclusively. In the lower panels the sequenced library pool contained either only plasma EV samples or included libraries from the miRXplore reference pool and/or cellular RNA. Every dot depicts a canonical miRNA. Analysis only includes miRNAs detected in both sequence runs.

Intriguingly, the plasma EV samples with low reproducibility were sequenced in library pools that contained a combination of libraries from low input plasma EV samples and libraries from high input samples, e.g., the miRXplore reference pool and/or cellular RNA. However, from the samples with high reproducibility the technical replicates were sequenced in library pools with exclusively low input plasma EV samples. When mixing samples of low and high input and complexity, we detected considerable levels of miRXplore reference pool reads assigned to plasma EV samples that may be due to a phenomenon called index hopping. This artefact is inherent to Illumina sequence platforms with patterned flow cells, where excess indices from one library are incorporated to another co-sequenced library during the initial exclusion amplification steps prior to adhering to the flow cell, leading to miss assignment of the reads.[16] This is particularly important for biological low input samples that may have high levels of excess free adapters, which can subsequently "pick up" RNA molecules from high input libraries in the same pool, leading to miss assignment of reads. Therefore, we recommend that low input samples with low complexity should never be co-sequenced with high input and high complexity samples.

Alternatively unique dual indices can be designed (incorporated in both 5′- and 3′-adapters). By inclusion of these sequences index hopping errors can be filtered out computationally after sequencing.

Problem 3

High abundance of specific RNA molecules that are not of interest (related to step 17).

Potential solution

If RNA molecules are extremely abundant, yet not of interest, one may opt to deplete these molecules specifically prior to sequencing. We designed and tested oligos against the 5′ end of

hY4, hsa-miR-486-5p and hsa-miR-92a-3p (all highly abundant in plasma EV, at least with Illumina TruSeq library preparation[8]) to preclude them from adapter ligation.[17] The depletion oligos (see key resources table for sequences and modifications) were incubated for 5 min at 95°C, 65°C, 55°C, 45°C, 35°C and 25°C subsequently and cooled down to 4°C, enabling the formation of a hairpin structure. 1 μL of 1.5 μM pre-incubated oligos was added during the library preparation procedure after ligation of the 3′-5N-adapter in the presence of 10 mM ATP (New England Biolabs, Cat#P0756S) and T4 DNA ligase (New England Biolabs, Cat#M0202T). Samples were incubated for 1 h at 30°C and 10 min at 65°C, followed by the rest of the library preparation procedure. The ligation of the depletion oligos to the 5′-end of the targeted small RNA molecules, prevents the subsequent ligation of the 5′-5N-adapter. Using depletion oligos against hY4, hsa-miR-486-5p and hsa-miR-92a-3p leads to a sequence specific reduction of these transcripts but also other RNA molecules with high sequence homology at the 5′ end. However, when comparing libraries with and without depletion of specific abundant RNA transcripts in our hands we did not observe any effect on the overall miRNA profile of plasma EV. This step may thus be included or omitted depending on the biosource and need.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, D. Michiel Pegtel (d.pegtel@amsterdamumc.nl).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The accession number for the datasets reported in this paper is SRA: PRJNA867189.

ACKNOWLEDGMENTS

This work was supported by multiple grants awarded to D.M.P. including NWO Perspectief Cancer-ID, TKI-Health Holland AQrate, and Cancer Center Amsterdam Foundation. We thank E.J. Bradley for technical advice. Figure 3 and part of the graphical abstract were created with BioRender.com.

AUTHOR CONTRIBUTIONS

Conceptualization, M.A.J.v.E. and D.M.P.; methodology, M.A.J.v.E., M.H., and D.M.P.; investigation, M.A.J.v.E.; writing - initial draft, M.A.J.v.E. and D.M.P.; reviewing and editing, M.A.J.v.E., C.S., E.A.-P., M.H., and D.M.P.; technical advice, C.S.; data analysis, M.A.J.v.E., E.A.-P., and M.H.; funding acquisition, D.M.P.

DECLARATION OF INTERESTS

D.M.P. and M.H. are co-founders of ExBiome BV and D.M.P. is CSO of ExBiome BV. D.M.P. and M.H. are inventors on a patent related to this work (WO2015165779A3). M.A.J.v.E. received travel compensation from QIAGEN.

REFERENCES

1. Gómez-Martín, C., Aparicio-Puerta, E., van Eijndhoven, M.A.J., Medina, J.M., Hackenberg, M., and Pegtel, D.M. (2023). Reassessment of miRNA variant (isomiRs) composition by small RNA sequencing. Cell Rep. Methods *3*, 100480. https://doi.org/10.1016/j.crmeth.2023.100480.

2. Eijndhoven, M.A.J. van, Aparicio-Puerta, E., Gómez-Martín, C., Medina, J.M., Drees, E.E.E., Bradley, E.J., Bosch, L., Scheepbouwer, C., Hackenberg, M., and Pegtel, D.M. (2021). Unbiased and UMI-informed sequencing of cell-free miRNAs at single-nucleotide resolution. Preprint at bioRxiv. https://doi.org/10.1101/2021.05.04.442244.

3. Srinivasan, S., Treacy, R., Herrero, T., Olsen, R., Leonardo, T.R., Zhang, X., DeHoff, P., To, C., Poling, L.G., Fernando, A., et al. (2020). Discovery and Verification of Extracellular miRNA Biomarkers for Non-invasive Prediction of Pre-eclampsia in Asymptomatic Women. Cell Rep. Med. *1*, 100013. https://doi.org/10.1016/j.xcrm.2020.100013.

4. Drees, E.E.E., Roemer, M.G.M., Groenewegen, N.J., Perez-Boza, J., van Eijndhoven, M.A.J., Prins, L.I., Verkuijlen, S.A.W.M., Tran, X.-M., Driessen, J., Zwezerijnen, G.J.C., et al. (2021). Extracellular vesicle miRNA predict FDG-PET status in patients with classical Hodgkin

Lymphoma. J. Extracell. Vesicles *10*, e12121. https://doi.org/10.1002/jev2.12121.

5. Max, K.E.A., Bertram, K., Akat, K.M., Bogardus, K.A., Li, J., Morozov, P., Ben-Dov, I.Z., Li, X., Weiss, Z.R., Azizian, A., et al. (2018). Human plasma and serum extracellular small RNA reference profiles and their clinical utility. Proc. Natl. Acad. Sci. USA *115*, E5334–E5343. https://doi.org/10.1073/pnas.1714397115.

6. Heinicke, F., Zhong, X., Zucknick, M., Breidenbach, J., Sundaram, A.Y.M., T Flåm, S., Leithaug, M., Dalland, M., Farmer, A., Henderson, J.M., et al. (2020). Systematic assessment of commercially available low-input miRNA library preparation kits. RNA Biol. *17*, 75–86. https://doi.org/10.1080/15476286.2019.1667741.

7. Giraldez, M.D., Spengler, R.M., Etheridge, A., Godoy, P.M., Barczak, A.J., Srinivasan, S., De Hoff, P.L., Tanriverdi, K., Courtright, A., Lu, S., et al. (2018). Comprehensive multi-center assessment of small RNA-seq methods for quantitative miRNA profiling. Nat. Biotechnol. *36*, 746–757. https://doi.org/10.1038/nbt.4183.

8. van Eijndhoven, M.A., Zijlstra, J.M., Groenewegen, N.J., Drees, E.E., van Niele, S., Baglio, S.R., Koppers-Lalic, D., van der Voorn, H., Libregts, S.F., Wauben, M.H., et al. (2016). Plasma vesicle miRNAs for therapy response monitoring in Hodgkin lymphoma patients. JCI Insight *1*, e89631. https://doi.org/10.1172/jci.insight.89631.

9. Kivioja, T., Vähärautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S., and Taipale, J. (2011). Counting absolute numbers of molecules using unique molecular identifiers. Nat. Methods *9*, 72–74. https://doi.org/10.1038/nmeth.1778.

10. Kim, H., Kim, J., Kim, K., Chang, H., You, K., and Kim, V.N. (2019). Bias-minimized quantification of microRNA reveals widespread alternative processing and 3′ end modification. Nucleic Acids Res. *47*, 2630–2640. https://doi.org/10.1093/nar/gky1293.

11. Aparicio-Puerta, E., Gómez-Martín, C., Giannoukakos, S., Medina, J.M., Scheepbouwer, C., García-Moreno, A., Carmona-Saez, P., Fromm, B., Pegtel, M., Keller, A., et al. (2022). sRNAbench and sRNAtoolbox 2022 update: accurate miRNA and sncRNA profiling for model and non-model organisms. Nucleic Acids Res. *50*, W710–W717. https://doi.org/10.1093/nar/gkac363.

12. Scheepbouwer, C., Hackenberg, M., van Eijndhoven, M.A.J., Gerber, A., Pegtel, M., and Gómez-Martín, C. (2023). NORMSEQ: a tool for evaluation, selection and visualization of RNA-Seq normalization methods. Nucleic Acids Res. *51*, W372–W378. https://doi.org/10.1093/nar/gkad429.

13. Valdmanis, P.N., Kim, H.K., Chu, K., Zhang, F., Xu, J., Munding, E.M., Shen, J., and Kay, M.A. (2018). miR-122 removal in the liver activates imprinted microRNAs and enables more effective microRNA-mediated gene repression. Nat. Commun. *9*, 5321. https://doi.org/10.1038/s41467-018-07786-7.

14. Yang, A., Bofill-De Ros, X., Shao, T.J., Jiang, M., Li, K., Villanueva, P., Dai, L., and Gu, S. (2019). 3′ Uridylation Confers miRNAs with Non-canonical Target Repertoires. Mol. Cell *75*, 511–522.e4. https://doi.org/10.1016/j.molcel.2019.05.014.

15. Juzenas, S., Venkatesh, G., Hübenthal, M., Hoeppner, M.P., Du, Z.G., Paulsen, M., Rosenstiel, P., Senger, P., Hofmann-Apitius, M., Keller, A., et al. (2017). A comprehensive, cell specific microRNA catalogue of human peripheral blood. Nucleic Acids Res. *45*, 9290–9301. https://doi.org/10.1093/nar/gkx706.

16. Larsson, A.J.M., Stanley, G., Sinha, R., Weissman, I.L., and Sandberg, R. (2018). Computational correction of index switching in multiplexed sequencing libraries. Nat. Methods *15*, 305–307. https://doi.org/10.1038/nmeth.4666.

17. Roberts, B.S., Hardigan, A.A., Kirby, M.K., Fitz-Gerald, M.B., Wilcox, C.M., Kimberly, R.P., and Myers, R.M. (2015). Blocking of targeted microRNAs from next-generation sequencing libraries. Nucleic Acids Res. *43*, e145. https://doi.org/10.1093/nar/gkv724.