STAR Protocols

Protocol

Experimental and computational workflow for the analysis of tRNA pools from eukaryotic cells by mim-tRNAseq

Quantifying tRNAs is crucial for understanding how they regulate mRNA translation but is hampered by their extensive sequence similarity and premature termination of reverse transcription at multiple modified nucleotides. Here, we describe the use of modification-induced misincorporation tRNA sequencing (mim-tRNAseq), which overcomes these limitations with optimized library construction and a comprehensive toolkit for data analysis and visualization. We outline algorithm improvements that enhance the efficiency and accuracy of read alignment and provide details on data analysis outputs using example datasets.

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

Andrew Behrens,
Danny D.
Nedialkova

nedialkova@biochem.
mpg.de

Highlights

mim-tRNAseq enables efficient quantitation and analysis of eukaryotic tRNA pools

Combined library generation and computational package ensures user-friendly workflow

Improved multiplexing enables more cost-effective sample analysis

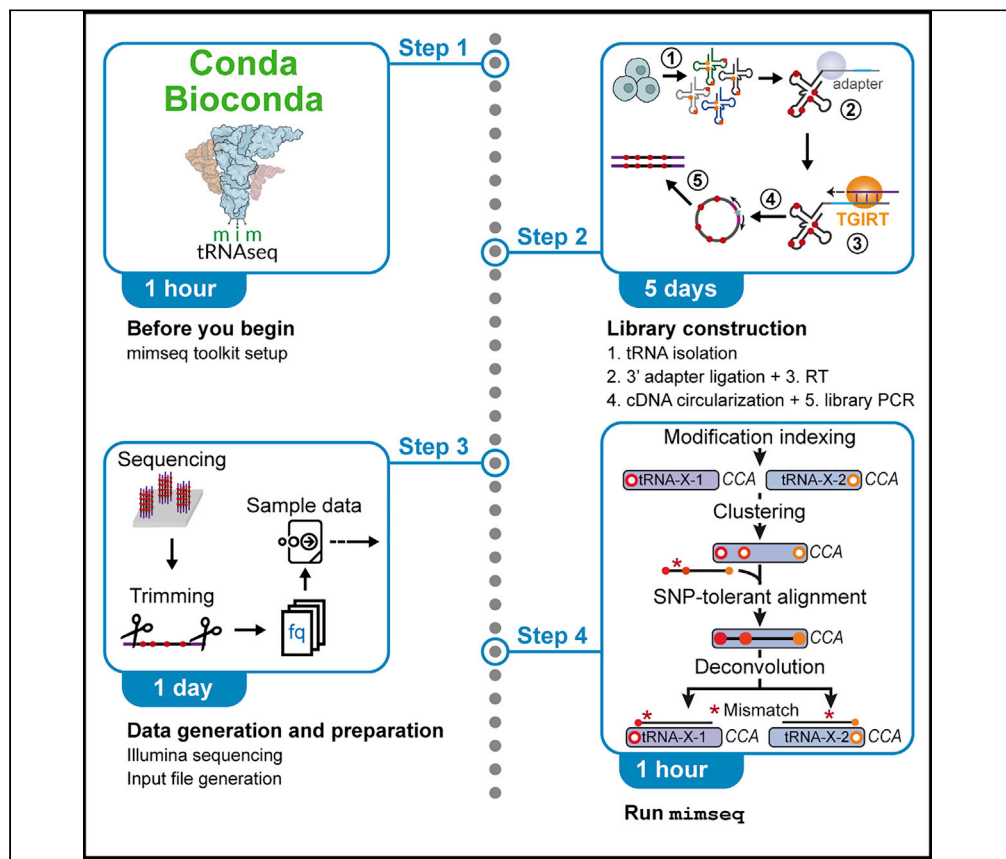Novel deconvolution algorithm enhances read alignment accuracy and sensitivity

Protocol

Experimental and computational workflow for the analysis of tRNA pools from eukaryotic cells by mim-tRNAseq

Andrew Behrens[1] and Danny D. Nedialkova[1,2,3,4,*]

[1]Mechanisms of Protein Biogenesis, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany

[2]Department of Chemistry, Technical University of Munich, 85748 Garching, Germany

[3]Technical contact

[4]Lead contact

*Correspondence: nedialkova@biochem.mpg.de
https://doi.org/10.1016/j.xpro.2022.101579

SUMMARY

Quantifying tRNAs is crucial for understanding how they regulate mRNA translation but is hampered by their extensive sequence similarity and premature termination of reverse transcription at multiple modified nucleotides. Here, we describe the use of modification-induced misincorporation tRNA sequencing (mim-tRNAseq), which overcomes these limitations with optimized library construction and a comprehensive toolkit for data analysis and visualization. We outline algorithm improvements that enhance the efficiency and accuracy of read alignment and provide details on data analysis outputs using example datasets.

For complete details on the use and execution of this protocol, please refer to Behrens et al. (2021).

BEFORE YOU BEGIN

mim-tRNAseq consists of an optimized protocol for cDNA library construction from eukaryotic tRNA pools and a suite of computational tools for the analysis, quantitation, and visualization of the resulting high-throughput sequencing data. It is based on the efficient synthesis of full-length cDNAs from tRNA transcripts with a thermostable group II intron reverse transcriptase (TGIRT) in reaction conditions that enable the readthrough of nearly all nucleotide modifications. To account for extensive sequence similarity among eukaryotic tRNA genes and the nucleotide misincorporations in cDNA introduced at modified tRNA sites by TGIRT, we developed the mim-tRNAseq computational toolkit, which includes multiple novel algorithms that are specifically tailored to the analysis of tRNA-derived sequencing data. Details of the development and optimization of mim-tRNAseq can be found in the original publication (Behrens et al., 2021) and in the package documentation (https://mim-trnaseq.readthedocs.io/en/latest/).

Here, we describe in detail the steps required to implement the library generation and data analysis workflow, starting with samples from cultured human cell lines. We also present an updated v1.1 of the mim-tRNAseq toolkit, which contains improvements to various aspects of the computational workflow that increase the accuracy and efficiency of data analysis.

> *Note:* For all computational steps throughout the protocol, commands to be entered on the command line are given in text boxes (starting with ">") along with comments describing the function of the command (starting with "#").

Preparation of the mim-tRNAseq toolkit environment

🕐 Timing: 1 h

Before use, the computational toolkit needs to be installed. The package (named mimseq) is available on bioconda (https://bioconda.github.io/recipes/mimseq/README.html), GitHub (https://github.com/nedialkova-lab/mim-tRNAseq), and archived on Zenodo (https://doi.org/10.5281/zenodo.6694873).

> *Note:* We strongly recommend using the conda package and environment manager to install the package with all of its dependencies from the bioconda channel.

> *Note:* The GitHub repository is a useful source of community discussion, issues and solutions, and requested upgrades and functionality (see Issues tab).

> *Note:* The toolkit has been tested extensively on multiple Linux-based servers and computing clusters. We recommend a Linux server with at least 32GB of memory and 8 CPU cores. These requirements will change depending on the number of samples and size of the datasets and tRNA reference, as will the processing time required by mimseq. Although running mimseq on a personal computer is theoretically possible, and multi-processing is customizable before runtime, we have not tested this.

1. Retrieve and install miniconda.

```
# Download miniconda

> wget https://repo.anaconda.com/miniconda/Miniconda3-py39_4.10.3-Linux-x86_64.sh

# Run the installation script

> bash Miniconda3-py39_4.10.3-Linux-x86_64.sh
```

After running the second command, follow the on-screen prompts and accept the defaults (unless otherwise required).

Following installation, restart the shell session by reconnecting to the remote server in order to activate the conda installation.

2. Initialize and configure the mimseq environment.

```
# Create the mimseq environment with the correct Python version

> conda create -n mimseq python=3.7

# Activate the environment

> conda activate mimseq

# Configure environment channels

> conda config –add channels conda-forge
```

3. Install mamba.

```
> conda install -c conda-forge mamba
```

4. Use mamba to install mimseq and all dependencies.

```
> mamba install -c bioconda mimseq
```

5. usearch is not available on conda and needs to be installed manually.

```
# Download and unzip usearch

> wget https://drive5.com/downloads/usearch10.0.240_i86linux32.gz

> gunzip usearch10.0.240_i86linux32.gz

# Make binary executable and rename

> chmod +x usearch10.0.240_i86linux32

> mv usearch10.0.240_i86linux32 usearch

# If root access is available then copy the binary into an accessible PATH location, for
example:

> cp usearch /usr/local/bin

# However, if this is not possible, add the path to the usearch binary to your PATH variable

> export PATH=$PATH:full/path/to/usearch

# where ''full/path/to/usearch'' should be replaced by the path to the usearch binary that was
just unzipped and modified.
```

Note: Exporting the usearch binary to a user-specific path as in the last command of step 5 is temporary and needs to be done for every new terminal session for that user. It is recommended to add the command to the user's .bashrc file (or similar), or preferably to have your system administrator add usearch to a global PATH location (see above, where usearch is copied to /usr/local/bin).

6. Test your mimseq installation by determining the version and printing the help. Output should look like the screenshots in Figure 1. Please ensure you have installed mimseq v1.1 or newer.

```
> mimseq –version

> mimseq –help
```

Preparation of cell or tissue samples for total RNA isolation

⏱ Timing: days (depending on the organism, developmental stage, and cell or tissue type)

7. Culture cells in an appropriate medium or collect tissue samples according to the purpose of the experiment.

⚠ CRITICAL: Extra care should be taken when working with animals, as post-mortem RNA degradation can occur rapidly in some tissue types (Richter et al., 2022). Dissection should be performed as quickly as possible, and tissue samples should be snap-frozen in liquid nitrogen and stored at −80°C.

Oligonucleotide ordering

⏱ Timing: 1 h–days (depending on supplier)

8. Order spike-in RNA, 3′ adapter, reverse transcription (RT), and library construction oligonucleotides (Table S1).

> △ CRITICAL: Since the library construction protocol is susceptible to the presence of ribonucleases, the spike-in, 3′ adapter, and RT oligonucleotides should be of RNase-free HPLC purity grade.

> *Note:* The *E. coli* tRNA-Lys-UUU tRNA is a suitable spike-in for eukaryotic samples as it differs sufficiently in sequence from eukaryotic tRNAs to avoid read misalignment. Other tRNA sequences can be substituted, provided that care is taken to ensure that no cross-mapping of sample reads to the spike-in reference will occur.

> *Note:* The custom 3′ adapter oligonucleotides should contain a 5′ phosphate, which is necessary for 5′ adenylation, and should be blocked with 3′ dideoxycytidine (or an alternative blocking group) to prevent self-ligation. The invariable 5′ GAT sequence of each adapter ensures an identical sequence context at the tRNA-adapter ligation junction (all mature tRNAs end with a single-stranded 3′ CCA, or a mixture of 3′ CCA and 3′ CC depending on their charging status after periodate oxidation and β-elimination). A 5-nt barcode (I1 – I8) is present at positions 4–8. This enables the pooling of up to eight samples before reverse transcription, which reduces sample input requirements and reagent costs.

> △ CRITICAL: The eight barcoded 3′ adapter oligonucleotides in Table S1 were specifically designed to minimize the potential for secondary structure formation, which interferes with ligation (Zhuang et al., 2012). The use of other 3′ adapters may result in decreased ligation efficiency to mature tRNAs and should be carefully evaluated first.

> *Note:* The RT oligonucleotide contains a 5′ phosphate necessary for cDNA circularization by CircLigase followed by an RN dinucleotide, which mitigates potential sequence biases in this reaction (Heyer et al., 2015; McGlincy and Ingolia, 2017).

> △ CRITICAL: Reverse primers for PCR library construction should contain a unique 6-nt index sequence (denoted with NNNNNN) to discriminate between different libraries loaded on the same flow cell. Indexes should differ by at least 2 nucleotides to avoid demultiplexing errors.

> △ CRITICAL: All primers for library construction PCR should contain a phosphorothioate bond between the last two nucleotides at the 3′ end to prevent their degradation by the KAPA HiFi Polymerase, which has a strong 3′-5′ exonuclease activity. The use of unmodified oligonucleotides results in poor cDNA amplification in step 131.

3′ adapter preadenylation

⏱ Timing: 2 h

9. Set up a 3' adapter preadenylation reaction as (McGlincy and Ingolia, 2017).

3' adapter preadenylation reaction

Reagent	Final concentration	Amount
10 × 5′ DNA Adenylation Reaction Buffer	1 ×	2 μL
ATP (1 mM)	0.1 mM	2 μL
Mth RNA Ligase (50 μM)	5 μM	2 μL

(Continued on next page)

Continued

Reagent	Final concentration	Amount
Barcoded 3′ adapter (one of I1-I8) (100 μM)	6 μM	1.2 μL
RNase-free water	N/A	12.8 μL
Total	N/A	20 μL

10. Incubate the reaction at 65°C for 1 h, followed by an incubation at 85°C for 5 min.
11. Add 30 μL RNase-free water and purify the pre-adenylated adapter with the Zymo Oligo Clean & Concentrator kit according to the manufacturer's protocol. Elute in 6 μL RNase-free water.
12. Preadenylated adapters can be stored at −20°C for several months.

   Note: To prepare a larger batch of a pre-adenylated 3′ adapter, set up several reactions and pool them prior to step 11.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
TRIzol Reagent	Thermo Fisher Scientific	Cat#15596026
1-bromo-3-chloropropane	Sigma-Aldrich	Cat#B9673
Formamide, deionized	Sigma-Aldrich	Cat# F9037
Acrylamide/Bis 19:1, 40% (w/v) solution	Thermo Fisher Scientific	Cat#AM9022
Urea	Sigma-Aldrich	Cat#U1250
UltraPure TBE buffer (10 ×)	Thermo Fisher Scientific	Cat#15581044
1,4-Dithiothreitol	Thermo Fisher Scientific	Cat#R0862
Sodium dodecyl sulphate	Sigma-Aldrich	Cat#74256
Sodium acetate	Sigma-Aldrich	Cat#S7545
Magnesium chloride	Sigma-Aldrich	Cat#M2670
Sodium chloride	Sigma-Aldrich	Cat# S7653
Potassoum chloride	Sigma-Aldrich	Cat#P9541
EDTA	Sigma-Aldrich	Cat#E5134
Sodium periodate	Sigma-Aldrich	Cat#311448
Sodium tetraborate decahydrate	Sigma-Aldrich	Cat#B3545
Ethanol	Sigma-Aldrich	Cat#51976
Glycogen	Thermo Fisher Scientific	Cat#AM9510
Low Range ssRNA ladder	New England Biolabs	Cat#N0364S
T4 Polynucleotide Kinase	New England Biolabs	Cat#M0201L
T4 RNA ligase 2, truncated KQ (200 U/μL)	New England Biolabs	Cat#M0373L
SUPERase·In RNase Inhibitor (20 U/μL)	Thermo Fisher Scientific	Cat#AM2694
TGIRT	InGex	Cat#TGIRT50
CircLigase ssDNA ligase	Lucigen	Cat#CL4115K
KAPA HiFi DNA Polymerase (1 U/μL)	Roche	Cat#KK2102
Critical commercial assays		
5′ DNA Adenylation Kit	New England Biolabs	Cat#E2610L
Micro Bio-Spin 30 Columns, RNase-free	Bio-Rad	Cat#326251
Costar Spin-X® Centrifuge Tube Filters, 0.22 μm	Corning Life Sciences	Cat#8160
SYBR Gold Nucleic Acid Gel Stain	Thermo Fisher Scientific	Cat# S11494
DNA Clean & Concentrator-5 kit	Zymo Research	Cat#D4013
Oligo Clean & Concentrator kit	Zymo Research	Cat#D4060
Qubit dsDNA HS Assay Kit	Thermo Fisher Scientific	Cat#Q32854
Deposited data		
mim-tRNAseq data from HEK293T and K562 cells	(Behrens et al., 2021)	GEO: GSE152621

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Oligonucleotides		
Oligonucleotides, RNA sequences and primers for library construction	(Behrens et al., 2021) and this paper; see Table S1	https://doi.org/10.17632/vy8z394gfh.1
Software and algorithms		
mim-tRNAseq v1.1.6	(Behrens et al., 2021)	GitHub: https://github.com/nedialkova-lab/mim-tRNAseq Zenodo: https://doi.org/10.5281/zenodo.6694873
cutadapt v3.5	(Martin, 2011)	https://cutadapt.readthedocs.io/en/stable/
GSNAP v2019-02-26	(Wu and Nacu, 2010)	http://research-pub.gene.com/gmap/
Samtools v1.14	(Li et al., 2009)	http://samtools.sourceforge.net/
Bedtools v2.30.0	(Quinlan and Hall, 2010)	https://bedtools.readthedocs.io/en/late
BLAST v2.10.1+	(Camacho et al., 2009)	https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&DOC_TYPE=Download
Infernal v1.1.4	(Nawrocki and Eddy, 2013)	http://eddylab.org/infernal/
usearch v10.0.240_i86linux32	(Edgar, 2010)	https://www.drive5.com/usearch/
R/DESeq2 v1.34.0	(Love et al., 2014)	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
R/ComplexHeatmap v2.10.0	(Gu et al., 2016)	https://www.bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html
Python/Biopython v1.79	(Cock et al., 2009)	https://biopython.org/
Other		
Low Range ssRNA ladder	New England Biolabs	Cat#N0364S
pBR322 DNA-MspI Digest	New England Biolabs	Cat#N3032S
Disposable homogenizer pestles for 1.5 mL tubes	VWR	Cat#47747-358

MATERIALS AND EQUIPMENT

Denaturing urea polyacrylamide gel, 10%

Reagent	Final concentration	Amount
TBE (10 ×)	1 ×	1 mL
Urea	7 M	4.2 g
Acrylamide/Bis 19:1, 40% (w/v) solution	10%	2.5 mL
RNase-free water	N/A	make up to 10 mL
APS (10%)	0.05%	50 μL
TEMED	0.5%	50 μL
Total	N/A	10 mL

Prepare freshly before use.

2 × RNA loading buffer

Reagent	Final concentration	Amount
TBE (10 ×)	1 ×	1 mL
Formamide	85%	8.5 mL
SDS (10%)	0.05%	50 μL
EDTA (0.5 M)	0.5 mM	100 μL
RNase-free water	N/A	350 μL
Total	N/A	10 mL

Store at ambient temperature for up to three months or at −20°C for up to two years.

RNA gel extraction buffer

Reagent	Final concentration	Amount
Sodium acetate, pH=4.5 (3 M)	300 mM	1 mL
SDS (10%)	0.25%	250 μL

(Continued on next page)

Continued

Reagent	Final concentration	Amount
EDTA (0.5 M)	1 mM	200 µL
RNase-free water	N/A	9.55 mL
Total	N/A	10 mL

Aliquot in small batches to minimize RNase contamination risk. Store at ambient temperature for up to a year.

TGIRT reaction buffer

Reagent	Final concentration	Amount
Tris-HCl, pH=8.3 (1 M)	50 mM	50 µL
KCl (3 M)	75 mM	25 µL
$MgCl_2$ (1 M)	3 mM	3 µL
RNase-free water	N/A	922 µL
Total	N/A	1 mL

Aliquot and store at −20°C for up to two years.

Sodium periodate solution

Reagent	Final concentration	Amount
Sodium periodate	500 mM	0.106 g
RNase-free water	N/A	1 mL
Total	N/A	1 mL

Prepare freshly before use.

Sodium tetraborate solution

Reagent	Final concentration	Amount
Sodium tetraborate decahydrate, pH=9.5	100 mM	0.38 g
RNase-free water	N/A	10 mL
Total	N/A	10 mL

Prepare freshly right before use. Dissolve in 7 mL water, adjust pH to 9.5 with 5 M NaOH, and make up to 10 mL with water.
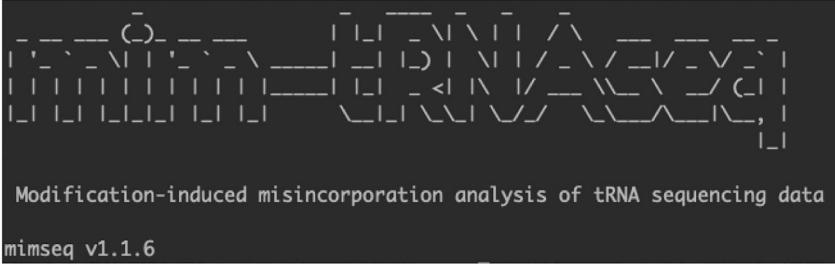
⚠ CRITICAL: To minimize the risk of RNase contamination, wear gloves during all experiments and change them frequently. Freshly purified Milli-Q water (Millipore) is sufficiently free of RNases for all steps of the protocol. Avoid using DEPC as it can inhibit some enzymatic reactions. Use RNase-free plasticware, low binding microtubes, and low-retention long filter pipette tips in all steps. Do not autoclave plasticware or buffers. RNases can be removed from glassware used for storing larger volumes of buffers (e.g., 1× TBE) by baking at 180°C–220°C for several hours.

⚠ CRITICAL: Sodium periodate, sodium tetraborate, acrylamide in solution, TRIzol, 1-bromo-3-chloropropane, formamide, DTT, and SDS are toxic and harmful. Wear appropriate personal protective equipment (goggles, gloves, lab coat) and follow institutional guidelines for handling and disposal.

Alternatives: TRIzol and acrylamide/bis (19:1, 40%) can be replaced with equivalent reagents from other commercial sources.

Alternatives: Handcast denaturing gels can be substituted with commercially available precast ones (e.g., 10% Novex™ TBE-Urea Gels from Thermo Fisher Scientific).
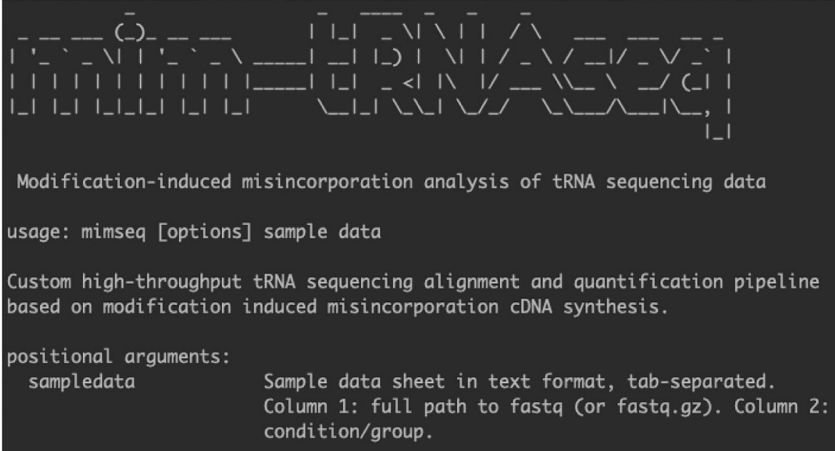
Figure 1. Testing mimseq environment and installation
(A) Running mimseq –version displays the mimseq logo and version number, which should be higher than v1.1.
(B) mimseq –help displays the help documentation on mimseq parameters.

STEP-BY-STEP METHOD DETAILS
Total RNA isolation under mild acidic conditions

⏲ Timing: 2–3 h

This section describes total RNA isolation from cultured mammalian cells under conditions that preserve tRNA charging.

⚠ CRITICAL: Working at low pH and maintaining samples at <4°C is essential for preserving tRNA charging. Avoid processing more than 8 samples at a time to decrease handling time.

Alternatives: If quantitation of tRNA charging is not of interest for the research question, total RNA can be extracted with the standard TRIzol protocol or other protocols or commercial kits that efficiently recover RNAs of <200 nt in length. In this case, proceed directly to step 46.

Note: RNA yield varies substantially depending on sample source, organism and cell type, and growth conditions. Commonly used human cell lines typically yield 15–30 μg of total RNA per 1×10^6 cultured human cells with TRIzol-based protocols.

Note: In our hands, RNA isolation with column-based commercial kits can lead to significant sample loss and poor recovery of small RNAs. We recommend avoiding the use of such kits, particularly when starting with low cell numbers.

1. Carefully aspirate all culture medium.
2. Add 1 mL of TRIzol Reagent per 1×10^5–1×10^7 cells directly to the culture dish.
3. Pipet the lysate up and down several times to homogenize and transfer to a 1.5 mL microfuge tube.
4. Snap-freeze samples on dry ice or in liquid nitrogen.

    ⏸ **Pause point:** Samples can be stored at −80°C for several months.

5. Thaw samples at room temperature (20°C–26°C).
6. Incubate at room temperature for 5 min to dissociate nucleoprotein complexes.
7. Add 0.2 mL of 1-bromo-3-chloropropane (BCP) per 1 mL of TRIzol Reagent used for lysis, then securely cap the tube.
8. Vortex briefly and incubate for 2 min at room temperature (20°C–26°C).
9. Centrifuge the sample for 15 min at 12,000 × g at 4°C in a pre-chilled centrifuge.
10. Place tubes on ice. Transfer the aqueous phase containing the RNA to a new 1.5 mL microfuge tube.
11. Add 0.5 mL of BCP per 1 mL of TRIzol Reagent used for lysis, then securely cap the tube. Vortex briefly.
12. Centrifuge the sample for 5 min at 12,000 × g at 4°C.
13. Place samples on ice. Transfer the aqueous phase containing the RNA to a new 2 mL microfuge tube.
14. Precipitate RNA by adding the following amounts per 1 mL of TRIzol Reagent used for lysis:
    a. 25 μg glycogen.
    b. 100 μL 3 M sodium acetate (pH=4.5).
    c. 1.25 mL of ice-cold 100% ethanol.
15. Vortex well and incubate at −20°C for at least 30 min.

    ⏸ **Pause point:** Samples can be stored at −20°C indefinitely.

    △ CRITICAL: The use of ethanol instead of isopropanol for RNA precipitation improves the recovery of RNAs <200 nt in length.

16. Centrifuge for 20 min at 12,000 × g at 4°C.
17. Carefully remove the supernatant with a 1-mL pipette tip.
18. Add 1 mL of ice-cold 80% ethanol containing 50 mM sodium acetate, pH=4.5.
19. Vortex the sample briefly, then centrifuge for 5 min at 7,500 × g at 4°C.
20. Carefully remove the supernatant with a 1-mL pipette tip.
21. Briefly spin down and remove all remaining liquid with a 10-μL pipette tip.
22. Air-dry the pellet for 2–3 min.
23. Resuspend in 30 μL 50 mM sodium acetate (pH=4.5), 1 mM EDTA per 1 mL of TRIzol Reagent used for lysis.
24. Measure RNA concentration on a NanoDrop or an equivalent UV spectrophotometer.

    ⏸ **Pause point:** Aliquot and store at −80°C for up to six months. Avoid repeated freeze-thaw cycles.

    Note: We recommend assessing the integrity of the extracted total RNA on a TapeStation system or by electrophoresis on a 10% denaturing polyacrylamide gel (steps 56–65). The tRNA cluster should be clearly visible at 60–100 nt and there shouldn't be any smearing that is indicative of RNA degradation.

Periodate oxidation and β-elimination

⏱ Timing: 4–6 h

The 3′ ends of tRNAs that carry an amino acid are protected from periodate oxidation. When followed by β-elimination, this treatment leads to the removal of the 3′ nucleotide in uncharged tRNAs. The proportion of transcripts that end with 3′-CCA versus those that end with 3′-CC after periodate oxidation and β-elimination can thus be used to quantify the fraction of aminoacylated tRNA molecules (Dittmar et al., 2005; Evans et al., 2017).

⚠ CRITICAL: This part of the protocol requires total RNA isolated under mild acidic conditions as input.

Note: The starting concentration of RNA in the oxidation reaction should be 0.25–1 μg/μL.

25. Assemble the oxidation reaction in 1.5-mL microfuge tubes.

Oxidation reaction		
Reagent	Final concentration	Amount
10 μg total RNA	0.5 μg/μL	X μL
10 mM sodium acetate, pH=4.5	N/A	20 - X μL
500 mM sodium periodate (freshly prepared)	50 mM	2.2 μL
Total	N/A	22 μL

26. Incubate for 30 min at 22°C in a Thermoblock.

Note: During the incubation time, prepare one Micro Bio Spin-30 column per sample according to the manufacturer's protocol.

27. Stop reaction by adding 2.4 μL 1 M glucose. Incubate for 5 min at 22°C in a Thermoblock.
28. Clean up reactions with Micro Bio Spin-30 columns according to the manufacturer's protocol. The recovered volume will be ~25 μL.
29. To each tube, add:
    a. 65 μL RNase-free water.
    b. 10 μL 3 M sodium acetate (pH=4.5).
    c. 25 μg glycogen.
    d. 300 μL ice-cold 100% ethanol.
30. Vortex well and incubate at −20°C for at least 30 min.

⏸ Pause point: Samples can be stored at −20°C indefinitely.

31. Precipitate RNA by centrifuging for 30 min at 16,000 × *g* at 4°C.
32. Carefully remove the supernatant with a 1-mL filter pipette tip. Briefly spin down and remove all residual liquid with a 10-μL filter pipette tip.
33. Resuspend pellets in 20 μL RNase-free water.
34. To perform β-elimination, mix the RNA with 30 μL freshly prepared 100 mM sodium tetraborate (pH=9.5).
35. Incubate at 45°C for 90 min in a Thermoblock.

Note: During the incubation time, prepare one Micro Bio Spin-30 column per sample according to the manufacturer's protocol.

36. Clean up reactions with Micro Bio Spin-30 columns according to the manufacturer's protocol. The recovered volume will be ~50 µL.
37. To each tube, add:
    a. 35 µL RNase-free water.
    b. 10 µL 3 M sodium acetate (pH=4.5).
    c. 25 µg glycogen.
    d. 300 µL ice-cold 100% ethanol.
38. Vortex well and incubate at −20°C for at least 30 min.

    ⫿⫿ **Pause point:** Samples can be stored at −20°C indefinitely.

39. Precipitate RNA by centrifuging for 30 min at 16,000 × *g* at 4°C.
40. Carefully remove the supernatant with a 1-mL filter pipette tip. Briefly spin down and remove all residual liquid with a 10-µL filter pipette tip.
41. Air-dry pellet for 2–3 min.
42. Resuspend pellets in 15 µL RNase-free water.
43. Measure RNA concentration on a NanoDrop or an equivalent spectrophotometer.

    ⚠ CRITICAL: The column purification and RNA precipitation steps are necessary to prevent carry-over of sodium periodate and sodium tetraborate. These reagents can inhibit the enzymatic removal of the RNA 3′ phosphate resulting from β-elimination with T4 polynucleotide kinase (PNK) in steps 49–56. This will lead to an underrepresentation of reads from uncharged tRNA in the final sequencing library and thus an overestimation of charged tRNA fractions.

Spike-in addition and tRNA deacylation

⧖ Timing: 1 h

The spike-in RNA added in this step is intended to be used as an internal control for 3′-CCA quantitation. This is achieved by adding two synthetic tRNAs that differ by a single nucleotide at the 3′ end (*E.coli* tRNA-Lys-UUU-CCA and tRNA-Lys-UUU-CC; for sequences, see Table S1) in a 3:1 ratio to total RNA.

Alternatives: Spike-in addition can be omitted if measurements of charging fractions are not performed.

Note: Following library construction and analysis by `mimseq`, the relative quantities of each spike-in RNA can be checked in the CCA plots to ensure the expected 3:1 ratio is recovered.

Note: We have successfully constructed tRNA sequencing libraries starting with 0.5 µg of total RNA. If input amounts are not limiting, we recommend starting with 2.5–5 µg of total RNA.

44. Dilute 2.5 µg total RNA in 18 µL RNase-free water in a 1.5-mL tube.
45. Add 0.75 µL of synthetic *E.coli* tRNA-Lys-UUU-CCA (7.5 ng/µL) and 0.75 µL of synthetic *E.coli* tRNA-Lys-UUU-CC (2.5 ng/µL).

    Note: If starting with RNA that has been subjected to periodate oxidation and β-elimination, the next step is not necessary; proceed directly to step 49.

46. Add 1.5 µL of 1 M Tris-HCl (pH=9.0) and 0.5 µL SUPERase·In.
47. Mix and incubate at 37°C for 45 min to deacylate tRNA.
48. Proceed to step 49.

Note: This step ensures the presence of a free 3′ OH group in tRNAs, which is a prerequisite for efficient ligation of a 3′ adapter in step 50. Although most tRNAs in total RNA samples isolated using standard protocols are likely to be deacylated, the stability of the acyl linkage can vary by an order of magnitude for different aminoacyl-tRNAs (Peacock et al., 2014).

RNA 3′ dephosphorylation

⏱ Timing: 1.5 h

This step entails the enzymatic removal of RNA 3′ phosphates resulting from β-elimination.

Note: We also perform this reaction when starting with total RNA that has not been subjected to oxidation and β-elimination. This serves to ensure that tRNAs with 3′ ends cleaved during stress (Czech et al., 2013) or as a result of ribosome-associated quality control (Yip et al., 2020) are also represented in sequencing libraries.

49. Assemble the following reaction in a 1.5-mL microfuge tube:

RNA 3′ end dephosphorylation reaction		
Reagent	Final concentration	Amount
Total RNA	0.025–0.1 μg/μL	X μL
10 × T4 PNK buffer	1 ×	10 μL
T4 PNK (10 U/μL)	5 U/μL	1 μL
SUPERase·In (20 U/μL)	5 U/μL	0.5 μL
RNase-free water	N/A	88.5 -X μL
Total	N/A	100 μL

50. Incubate at 37°C for 45 min.
51. To each tube, add:
    a. 10 μL 3 M sodium acetate (pH=4.5).
    b. 25 μg glycogen.
    c. 300 μL ice-cold 100% ethanol.
52. Vortex well and incubate at −20°C for at least 30 min.

⏸ Pause point: Samples can be stored at −20°C indefinitely.

53. Precipitate RNA by centrifuging for 30 min at 16,000 × *g* at 4°C.
54. Carefully remove the supernatant with a 1-mL filter pipette tip. Briefly spin down and remove all residual liquid with a 10-μL filter pipette tip.
55. Resuspend pellets in 5 μL RNase-free water.
56. Add 5 μL 2 × RNA loading buffer without dyes.

⚠ CRITICAL: Commonly used dyes such as xylene cyanol or bromophenol blue can co-migrate with RNA or DNA fragments of interest. Their co-purification with nucleic acids after gel extraction may interfere with downstream enzymatic reactions. Therefore, we recommend that these dyes be present only in the loading buffer added to marker samples or in empty wells.

Purification of tRNA from total RNA by gel size selection
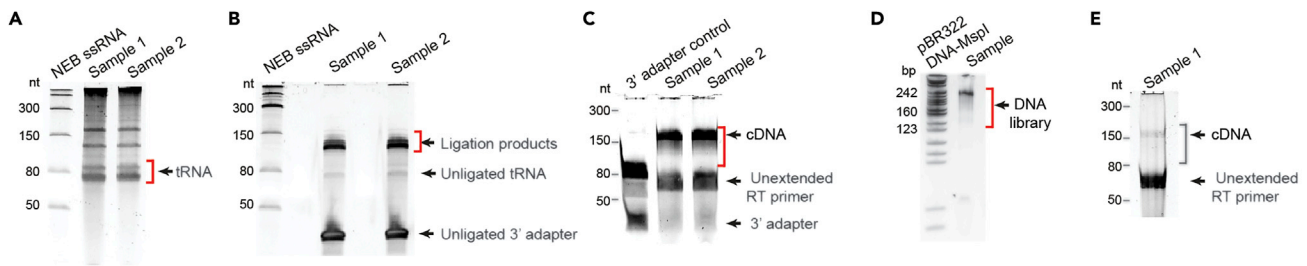
⏱ Timing: 30 min–1 h

Figure 2. Typical gel images of key steps during mim-tRNAseq library construction
Gel regions to be excised are indicated by a red bracket.
(A) Total RNA after step 66.
(B) Adapter-ligated tRNA after step 92.
(C) cDNA after successful reverse transcription in steps 105–113.
(D) Library DNA after step 135.
(E) cDNA after suboptimal reverse transcription in steps 105–113.

Transfer RNAs are 70–90 nt in length and run as a discrete cluster on denaturing 10% polyacrylamide gels. Gel size selection of 60–100 nt RNAs is a cost-effective approach to purify mature, intact tRNAs from total RNA preparations and to separate them from other highly abundant short RNAs with a similar size (e.g., 5S rRNA, 5.8S rRNA, snRNAs) or potential tRNA fragments and degradation intermediates, which would co-elute in small RNA fractions obtained by column-based commercial kits.

> *Note:* The presence of RNA transcripts with substantially fewer RT-blocking modifications that tRNAs may result in their preferential use as templates for cDNA synthesis in steps 102–107.

57. Briefly soak glass plates in 1% SDS and rinse with RNase-free water.
58. Cast a 10% denaturing polyacrylamide gel and let it polymerize for 30 min to 1 h.
59. Place gel in an appropriate tank and pre-run at 20 mA in 1 × TBE for at least 30 min.

> *Note:* Pre-running is critical for heating up urea-containing polyacrylamide gels, which aids RNA denaturation, and for removing excess urea from the wells, which can otherwise lead to band distortion.

60. Prepare gel size marker by mixing 1 µL NEB Low Range ssRNA ladder with 4 µL RNase-free water and 5 µL 2 × RNA loading buffer supplemented with xylene cyanol (0.005%) and bromophenol blue (0.01%).
61. Denature ladder and samples from step 30 at 90°C for 3 min and place immediately on ice.
62. Turn off the gel tank power supply.
63. Rinse gel wells to remove excess urea immediately before loading.
64. Load ladder and samples on gel and run at 20 mA until the bromophenol blue from the marker sample reaches the bottom of the gel.
65. Place gel in an RNAse-free container with SYBR Gold (1:10 000) in 1× TBE for 3 min on a platform shaker.

> *Note:* Longer staining periods do not improve signal substantially and may result in fuzzy bands and sample loss.

66. Place gel on a clean foil on top of a blue light transilluminator to visualize RNA (for an example, see Figure 2A).

> ⚠ CRITICAL: Visualization of total RNA on a denaturing gel is a useful indicator of RNA quality. Transfer RNAs, 5S rRNA, and 5.8S rRNA should be visible as discrete bands. Band

> smearing is indicative of RNA degradation, which can occur during sample collection and/
> or RNA isolation.

67. Use a clean scalpel blade to cut out the gel region corresponding to tRNA (∼60–100 nt).
68. Place the gel fragment in a 1.5-mL microfuge tube.
69. Elute RNA from gel slice (steps 70–83, resuspend the pellet in 8 μL RNAse-free water).

Elution of RNA from gel slices

⏱ Timing: 30 min–overnight

Inefficient recovery of RNA from gel slices is a major source of sample loss in sequencing library construction workflows that require gel size selection. We optimized this step to increase the yield of eluted RNA from ∼20% in the classical "crush and soak" method (Green and Sambrook, 2019) to ∼70%–80%.

70. Crush gel slice with a disposable 1.5-mL tube pestle.
71. Add 400 μL RNA gel extraction buffer.
72. Incubate tubes at 65°C for 10 min and 2000 rpm in a Thermoblock.

   Alternatives: Gel slurry can also be incubated at 65°C in a water bath. In this case, mix slurry by inverting tubes several times every 2–3 min.

73. Snap-freeze gel slurry on dry ice or in liquid nitrogen.
74. Thaw gel slurry at 65°C for 5 min and 2000 rpm in a Thermoblock.
75. Elute RNA overnight at room temperature with gentle mixing on a rotating wheel.

   ⚠ CRITICAL: Omitting the freeze-thaw cycle substantially decreases RNA elution efficiency.

   Alternatives: An incubation time of ∼2 h at room temperature in the final step is sufficient to achieve an elution efficiency of 40%–50%.

76. Remove gel pieces by centrifuging slurry through a SpinX filter at 10,000 × *g* for 30 s.
77. Transfer flow-through to a new 1.5-mL microfuge tube.
78. Add 25 μg glycogen and 1 mL ice-cold 100% ethanol.
79. Vortex well and incubate for at least 30 min at −20°C.

   ⏸ Pause point: Samples can be stored at −20°C indefinitely.

80. Precipitate RNA by centrifuging for 30 min at 16,000 × g at 4°C.
81. Carefully remove the supernatant with a 1-mL filter pipette tip. Spin down briefly and remove all residual liquid with a 10-μL tip.
82. Air-dry pellets for 2–3 min and resuspend in the indicated volume of RNase-free water.
83. Measure RNA concentration on a Nanodrop or an equivalent spectrophotometer.

   ⏸ Pause point: Samples can be stored at −80°C for several months.

Adapter ligation to tRNA 3′ ends

⏱ Timing: 4 h

This step adds a barcoded 3′ adapter oligonucleotide (Table S1) to tRNAs to serve as a priming site for cDNA synthesis by TGIRT.

⚠ CRITICAL: Select a distinct barcoded adapter (I1 to I8) for each sample that will be pooled prior to reverse transcription.

⚠ CRITICAL: Use block design (Auer and Doerge, 2010) to minimize the potential for differences between samples to be confounded by technical variation. For example, do not pool all replicates of a control condition in RT reaction 1 and all replicates of a treatment condition in RT reaction 2.

84. Assemble the 3′ adapter ligation reaction.

3′ adapter ligation reaction		
Reagent	Final concentration	Amount
Gel-purified tRNA (100 ng)	0.2 μM	5 μL
Pre-adenylated 3′ adapter (20 μM)	1 μM	1 μL
50% PEG-8000	12.5%	10 μL
10 × T4 RNA ligase buffer	1 ×	2 μL
T4 RNA ligase 2, truncated KQ (200 U/μL)	10 U/μL	1 μL
SUPERase·In (20 U/μL)	1 U/μL	1 μL
Total	N/A	20 μL

⚠ CRITICAL: The 50% PEG-8000 stock solution is very viscous, which impacts pipetting accuracy. Inaccurate pipetting can decrease 3′ adapter ligation efficiency, which strongly depends on the concentration of PEG-8000 (Heyer et al., 2015). For best results, equilibrate the 50% PEG-8000 stock solution at room temperature and pipet slowly and/or or use wide bore pipette tips. PEG-8000 should be added to each reaction tube separately rather than being included in a master mix.

85. Mix each reaction by carefully pipetting the entire volume up and down several times.
86. Incubate at 25°C for 3 h.
87. Pool reactions from samples with different barcodes and purify adapter-ligated tRNA with Zymo Oligo Clean & Concentrator kit according to the manufacturer's protocol.
88. Elute in 10 μL RNase-free water.
89. Add 10 μL 2 × RNA loading buffer without dyes.
90. Prepare a 10% denaturing polyacrylamide gel and a ladder sample as in steps 57–60.
91. Denature ladder and samples from step 88 at 90°C for 3 min and place immediately on ice.
92. Separate ligation products from excess adapter by running samples on the gel as in steps 62–65.
93. Cut out bands corresponding to adapter-ligated tRNA (∼110–125 nt, for example see Figure 2B) and put each gel fragment in a clean, low-biding RNase-free tube.
94. Recover adapter-ligated tRNA from gel slices as in steps 70–83.

⚠ CRITICAL: Reverse transcription by TGIRT is inhibited by trace amounts of ethanol. Ensure that all liquid is removed after precipitating adapter-ligated tRNA.

95. Resuspend pellets in 2.5 μL RNase-free water per sample pooled in step 87.
96. Measure RNA concentration on a Nanodrop or an equivalent spectrophotometer.

⏸ **Pause point:** Samples can be stored at −80°C for several months.

Alternatives: The gel purification step to remove excess 3′ adapter (steps 92–94) can be replaced by 5′-deadenylase treatment followed by digestion of the unligated adapter with the 5′–3′ ssDNA exonuclease RecJ (McGlincy and Ingolia, 2017). This alternative substantially increases the cost of library construction and is very inefficient in our hands.

Primer-dependent reverse transcription of tRNA pools with TGIRT

⏱ Timing: 30 min–overnight

In this step, adapter-ligated tRNA pools are used as templates for primer-dependent cDNA synthesis by TGIRT, an engineered version of a bacterial group II intron-encoded reverse transcriptase (Mohr et al., 2013; Qin et al., 2015), under reaction conditions that we optimized to favor tRNA modification readthrough.

> ⚠ CRITICAL: TGIRT cannot be substituted with commercially available reverse transcriptases of retroviral origin (e.g., ProtoScript II, SuperScript IV) since these enzymes have a much lower readthrough efficiency of modified sites in tRNAs.

Alternatives: Other group II intron- or retroelement-encoded enzymes such as MarathonRT (Zhao et al., 2018) and an engineered version of the *B. mori* R2 retroelement reverse transcriptase (Upton et al., 2021) have also been reported to efficiently read through RNA modifications (Guo et al., 2020; Gustafsson et al., 2022). These enzymes are not commercially available at this time, and protocols for their use may need to be further optimized for tRNA templates.

97. Prepare 100 mM DTT fresh from powder (0.0154 g/mL in RNase-free water).
98. Pre-warm a Thermocycler block at 82°C.
99. Assemble the primer hybridization reaction in a low-binding PCR tube:

RT reaction: primer hybridization

Reagent	Final concentration	Amount
Adapter-ligated tRNA (100 ng)	N/A	X μL
RT primer (1.25 μM)	N/A	2 μL
RNase-free water	N/A	10 – X μL
Total	N/A	12 μL

Note: Including a control reaction with only unligated 3′ adapter as a template can be useful to gauge RT efficiency, as the RT primer binds to the last 15 nucleotides of the 34 nt-long adapters. The product of this control reaction can be used as a marker during the size selection of cDNA products on the gel in step 113.

100. Denature at 82°C for 2 min and incubate at room temperature for 5 min.
101. In the meantime, pre-warm a Thermocycler block at 42°C (with lid temperature at 48°C).
102. Assemble the RT reaction by mixing the adapter-ligated tRNA:RT primer duplexes with the following components:

RT reaction

Reagent	Final concentration	Amount
Adapter-ligated tRNA:RT primer	125 nM	12 μL
5 × TGIRT Reaction Buffer	1 ×	4 μL
DTT (100 mM; freshly prepared)	0.5 mM	1 μL
SUPERase·In (20 U/μL)	1 U/μL	1 μL
TGIRT-III (10 μM)	500 nM	1 μL
Total	N/A	19 μL

103. Mix each reaction by pipetting the entire volume up and down several times.
104. Incubate at 42°C for 10 min in a Thermocycler.

105. Add 1 μL 25 mM dNTPs to each tube.
106. Mix each reaction by pipetting the entire volume up and down several times.
107. Incubate at 42°C for 16 h in Thermocycler.

⚠ CRITICAL: Incubation times as short as 1 h can yield a substantial proportion of full-length cDNA products from tRNA pools with lower modification frequency and complexity, e.g., from budding or fission yeast. An extended reaction time, however, is particularly important for the efficient recovery of cDNA from tRNAs with multiple RT-blocking modifications, which is characteristic of many human tRNAs. The integrity of tRNA templates is not compromised under these reaction conditions (Behrens et al., 2021).

⚠ CRITICAL: Efficient cDNA synthesis from endogenously modified tRNA requires a molar excess of TGIRT, so we highly recommend adhering to the template:TGIRT molar ratios described here.

⚠ CRITICAL: TGIRT is inhibited by trace amounts of ethanol. Ensure that all residual liquid is removed after precipitation of adapter-ligated tRNA in step 94.

108. Pre-warm a Thermocycler block at 95°C.
109. Add 1 μL 5 M NaOH to each reverse transcription reaction.
110. Incubate at 95°C for 3 min to hydrolyze the RNA template.
111. Add 20 μL 2 × RNA loading buffer to each reaction.
112. Denature cDNA at 95°C for 5 min and place samples immediately on ice.
113. Separate cDNA from unextended RT primer by running reactions alongside NEB Low Range ssRNA ladder (and optionally the 3′ adapter-only control reaction) on a 10% denaturing gel as in steps 57–65.
114. Excise the entire gel region above the unextended RT primer (or adapter-only RT product), which will correspond to all cDNA lengths derived from tRNA reverse transcription (for an example, see Figure 2C).
115. Put each gel fragment in a new 1.5 mL microfuge tube.
116. Crush gel slice with a disposable plastic pestle.
117. Add 400 μL 10 mM TE, pH=8.0, and snap-freeze gel slurry on dry ice or in liquid nitrogen.
118. Incubate gel slurry for 1 h at 70°C/2000 rpm in a Thermoblock to elute cDNA.

⚠ CRITICAL: Omitting the freeze-thaw cycle substantially decreases cDNA elution efficiency.

119. Remove gel pieces by centrifuging slurry through a SpinX filter at 10,000 × g for 30 s.
120. Transfer flow-through to a new 1.5-mL microfuge tube.
121. Add 25 μg glycogen, 40 μL 3 M NaCl, and 1 mL ice-cold 100% ethanol.
122. Vortex well and incubate for 30 min on ice.

⏸ **Pause point:** Samples can be stored at −20°C indefinitely.

123. Pellet DNA by centrifuging for 30 min at 4°C and 16,000 × g.
124. Carefully remove the supernatant with a 1-mL pipette tip. Briefly spin down and remove all remaining liquid with a 10-μL pipette tip.
125. Resuspend pellets in 5.5 μL water and proceed with cDNA circularization.

⏸ **Pause point:** Samples can be stored at −20°C for several weeks.

Circularization of cDNA

In this step, cDNA is circularized with CircLigase to provide a template for library construction by PCR.

126. Pre-warm a Thermocycler block at 60°C (lid temperature of 65°C).
127. Assemble cDNA circularization reaction in a low-binding PCR tube:

cDNA circularization reaction

Reagent	Final concentration	Amount
Gel-purified cDNA	n.d.	5.5 μL
Betaine (5 M)	1 M	2 μL
10 × CircLigase buffer	1 ×	1 μL
ATP (1 mM)	0.05 mM	0.5 μL
MnCl$_2$ (50 mM)	2.5 mM	0.5 μL
CircLigase ssDNA Ligase (100 U/ μL)	5 U/ μL	0.5 μL
Total	N/A	10 μL

128. Mix each reaction well by pipetting up and down.
129. Incubate at 60°C for 3 h in a Thermoblock.
130. Incubate at 80°C for 10 min to inactivate the enzyme.

> ⸏⸏ Pause point: Circularized cDNA can be stored at −20°C for several weeks.

> ⚠ CRITICAL: We do not recommend using CircLigase II as it has much lower ssDNA circularization efficiency in comparable library construction protocols (Heyer et al., 2015).

Note: Under the reaction conditions in step 102, TGIRT adds one to three non-templated adenosines to the 3′ ends of most cDNAs (Behrens et al., 2021), resulting in a nearly identical 5' – 3′ circularization sequence context.

Library construction PCR

> ⏱ Timing: 30 min–1 h

The circularized single-stranded cDNA is used as a template for the construction of a double-stranded DNA library with an appropriate structure for sequencing on Illumina platforms.

> ⚠ CRITICAL: Use reverse library PCR primers with distinct 6-nt indexes to amplify different libraries that will be sequenced on the same flow cell.

131. Set up a 50-μL PCR reaction to construct libraries from circularized cDNA.

Library construction PCR reaction

Reagent	Final concentration	Amount
Circularized cDNA	n.d.	2 μL
dNTPs (10 mM)	0.3 mM	1.5 μL
Forward library construction primer (10 μM)	0.5 μM	2.5 μL
Reverse library construction primer (10 μM)	0.5 μM	2.5 μL
5 × KAPA HiFi GC buffer	1 ×	10 μL
KAPA HiFi DNA Polymerase (1 U/μL)	0.05 U/μL	0.5 μL
PCR-grade water	N/A	31 μL
Total	N/A	50 μL

⚠ CRITICAL: The use of other DNA polymerases or buffers can result in the preferential amplification of cDNAs with a shorter length or lower GC content (Quail et al., 2011), which introduces bias in tRNA abundance measurements.

Note: Circularized cDNA does not need to be purified prior to PCR amplification if its volume does not exceed 10% of the final PCR reaction volume.

132. Amplify for 4–6 cycles at a ramp rate of 3°C/second with the following settings:

PCR cycling conditions

Steps	Temperature	Time	Cycles
Initial Denaturation	95°C	3 min	1
Denaturation	98°C	20 s	4–6 cycles
Annealing	62°C	30 s	
Extension	72°C	30 min	
Hold	16°C	forever	

Note: When performing RT of adapter-ligated tRNA and cDNA circularization according to steps 97–130, we find that 4–5 PCR cycles are sufficient to obtain DNA libraries of 2–5 nM, which is within the optimal concentration range for Illumina sequencing platforms. If more than 6 PCR cycles are necessary to achieve this yield, we recommend increasing the starting amount of circularized cDNA template in the PCR reaction or optimizing the reverse transcription reaction (see problem 1). Performing further PCR amplification will exacerbate bias as small differences in amplification efficiency of DNAs with different length and GC content will accumulate over multiple PCR cycles (Quail et al., 2011; van Dijk et al., 2014).

133. Purify PCR products with Zymo DNA Clean & Concentrator kit according to the manufacturer's instructions.
134. Elute in 12 μL 10 mM Tris-HCl, pH=8.0.
135. Use 2 μL to measure DNA concentration with the Qubit dsDNA HS kit.

Note: The DNA concentration of samples is too low to be accurately quantified by UV spectrophotometry.

Note: This step efficiently separates the double-stranded DNA libraries (210–225 bp) from PCR buffer components, excess dNTPs, and unused PCR primers (which are ~50 nucleotides and therefore below the retention cutoff). Primer removal will be less efficient if longer oligonucleotides are used for PCR. In this case, libraries should be purified by size selection on a non-denaturing polyacrylamide gel (McGlincy and Ingolia, 2017). Prior to electrophoresis, we recommend that DNA clean-up is performed as in step 133 since the presence of the KAPA HiFi GC buffer causes band distortion. For an example of a typical gel image of a DNA library, see Figure 2D.

⏸ **Pause point:** Purified DNA can be stored in low-binding tubes at −20°C indefinitely.

136. Assess DNA size and purity by non-denaturing polyacrylamide gel electrophoresis on an 8% non-denatring polyacrylamide gel.

Alternatives: Library size can be analyzed on a High Sensitivity D1000 ScreenTape on an Agilent 2200 TapeStation Nucleic Acid System.

⚠ CRITICAL: If purifying libraries by gel size selection, we recommend cutting out fragments of ~150 bp to ~230 bp in length. This will ensure the inclusion of fragments resulting from premature termination of cDNA synthesis at (hyper)modified sites. These fragments may not be visible as discrete bands after SYBR Gold staining of RT or library PCR samples (Figures 2C and 2D) since they originate from the very few remaining RT roadblocks in a small subset of tRNA transcripts (e.g., ms^{2}t^{6}A37 in human tRNA-Lys-UUU). Failure to include these fragments in the sequencing run, however, will lead to underestimating the abundance of the tRNAs that give rise to them.

Library sequencing on Illumina platforms

The libraries generated with this protocol should be sequenced on an Illumina platform with a single-end run of >=100 bp. The minimal sequencing read length can be determined by adding 12 nt to the longest predicted tRNA transcript in the organism of interest. This accounts for the supplementary sequences added during library construction and required for demultiplexing samples pooled prior to reverse transcription.

Note: Given the high quality scores of single-end reads with a length of 100–150 nucleotides on current Illumina platforms, paired-end sequencing is not required for the analysis of tRNA-derived libraries.

137. Pool and dilute sequencing libraries according to the requirements of your preferred Illumina sequencing provider.

Note: The common range of starting library concentration prior to cluster generation on Illumina sequencing platforms is 1 nM–4 nM.

Note: For NextSeq 550, the loading concentration of tRNA libraries to achieve an optimal cluster density (~220 k/mm^2) is 2 pM.

Note: Most organisms encode only a few hundred tRNA transcripts and the libraries generated with this protocol consist almost exclusively of tRNA-derived reads. We, therefore, find that a sequencing depth of 3–5 million reads per sample is sufficient for most analyses. This should be scaled by the number of individual samples pooled prior to reverse transcription, i.e., 40 million reads for a library that contains eight samples, each with a unique barcoded 3′ adapter.

Demultiplexing and adapter trimming of sequencing reads

⏱ Timing: 30 min–1 h

Using fastq files as an input, library demultiplexing and trimming extra sequences at the 3′ and 5′ added during library construction is performed with cutadapt v3.5 (or newer).

138. Prepare a file of barcodes for demultiplexing.

Note: barcodes.fa should be a fasta-formatted file with one user-selected header and sequence per barcoded adapter used in library construction. To ensure an efficient match to sequences in reads, we extend the sequence to 10nt (instead of only the 5nt barcode sequence itself). Therefore, each sequence in the file contains the common 5′-GAT sequence + unique barcode + CA-3'. Below is an example of *barcodes.fa*:

```
# Example barcodes.fa file for demultiplexing

>I1

GATATCGTCA

>I2

GATAGCTACA

>I3

GATGCATACA

>I4

GATTCTAGCA
```

Note: The names of the barcodes in *barcodes.fa* can be chosen freely and will be present in the names of the demultiplexed fastq files.

139. Demultiplex fastq files with the following command:

```
# Demultiplex each sample fastq file using barcodes.fa

> for i in *.fastq.gz

    do fn=$(basename $i .fastq.gz)

      cutadapt –no-indels -q 30,30 –trimmed-only -j 10 -a file:barcodes.fa -m 10 -o $fn'_
    {name}_trim.fastq.gz' $i 1> $fn'_log.txt'

    done
```

Note: This command uses a for loop to process each fastq file similarly. First, we create a filename variable $fn for more appropriate output file names lacking the file type extension and any other uninformative information (the basename command might need to be modified depending on the naming conventions of your files). Cutadapt is then used to quality-trim each read end with a threshold of 30 (-q 30,30) before demultiplexing according to *barcodes.fa*. Additionally, we discard trimmed reads shorter than 10 nt (-m 10) and prohibit indels in matches to barcode sequence (–no-indels).

Note: All reads of >= 100 nt should contain an adapter sequence since mature tRNAs are <=90 nts, and so only reads with a detected adapter which have been trimmed are kept (–trimmed-only).

⚠ CRITICAL: Adjust -j to a suitable number of processors for demultiplexing according to your system capabilities.

Note: An output log.txt file will be produced for each sample processed above (with 1> $fn'_log.txt'). We highly recommend assessing these files for trimming efficiency. Typically, the vast majority of reads should contain barcodes and adapters (as described above), and so the "Reads with adapters" value reported in the log should reflect that (i.e., at least 80% of reads containing adapters). Low values may suggest issues in library construction, because of e.g., the presence of no-insert libraries due to low RT efficiency.

140. Trim the two additional nucleotides introduced at the 5′ of reads following cDNA circularization.

```
# Trim 2× 5' random nucleotides
> for i in *trim*.fastq.gz
    do fn=$(basename $i trim.fastq.gz)
    cutadapt –j 10 –m 10 –u 2 –o $fn'_trimFinal.fastq.gz' $i
    done
```

Optional: Demultiplexing and trimming are now complete. However, you may wish to rename the "*trimFinal.fastq.gz*" files according to the samples they represent. For this, a table of which samples were multiplexed during library preparation, and which barcodes were used for each is useful to track which final trimmed fastq file represents each sample and/or condition.

Preparing mimseq input files

⏱ Timing: 15 min–days for custom references

Optional: Building custom references for mimseq.

mimseq contains pre-generated input files for multiple reference genomes. These are named using the first letter of the genus and the first three of the species. For example, *Homo sapiens* is given as Hsap. For a complete list of available species, please see the -s/–species section of the mimseq help page (mimseq –help) or the corresponding section in the documentation (https://mim-trnaseq.readthedocs.io/en/latest/start.html#pre-built-references). These references also contain the sequence for the synthetic *E. coli* tRNA-Lys-UUU spike-in that is added in step 45 (http://gtrnadb.ucsc.edu/genomes/bacteria/Esch_coli_K_12_MG1655/genes/tRNA-Lys-TTT-1-1.html).

Note: For species without pre-generated references, custom input files can be given. Ideally, the species will have predicted tRNAs on GtRNAdb (http://gtrnadb.ucsc.edu/). This data can be downloaded, and the respective tRNA fasta file and intron information file (.out file) can be specified to mimseq with -t and -o, respectively. If available, mitochondrial and/or plastid tRNA sequences (in the case of plant species) from the mitochondrial tRNA database (mitotR-NAdb; http://mttrna.bioinf.uni-leipzig.de/mtDataOutput/) and PtRNAdb (http://14.139.61.8/PtRNAdb/index.php) can be given in space-separated format with -m. Plastid sequences, in the case of plant species, can also be specified with -m.

Note: tRNAScan-SE2.0 (Chan et al., 2021) can be used to predict tRNA genes *de novo* and to generate the outputs used in mimseq.

⚠ CRITICAL: The formatting of output files and tRNA names must follow the convention as in GtRNAdb files and other mimseq prebuilt references. As an example, see the *S. cerevisiae* sacCer3 reference (https://github.com/nedialkova-lab/mim-tRNAseq/tree/master/mimseq/data/sacCer3-eColitK).

Note: The use of custom input files for species not currently included in mimseq has not been extensively tested and might lead to errors at runtime. See troubleshooting problem 3 for advice on how to overcome those.

Note: For species that have pre-built references in mimseq, or once the necessary inputs have been generated, the only input file needed for mimseq operation is one that specifies the

sample data. This is a tab-separated file with two columns; the first specifies the path to the trimmed fastq files from step 3, and the second specifies the condition or treatment group of the sample.

Note: From here on, protocol details are given in reference to running the example data present in the mimseq GitHub repository (https://github.com/nedialkova-lab/mim-tRNAseq; see *mimseq_hek_1.fastq.gz, mimseq_hek_2.fastq.gz, mimseq_k562_1.fastq.gz, mimseq_k562_2.fastq.gz and sampleData_HEKvsK562.txt*). These data are a subset of human HEK293T and K562 data generated (Behrens et al., 2021).

141. Create and save sample data tab-separated text file.

```
./mimseq_hek_1.fastq.gz    HEK293T

./mimseq_hek_2.fastq.gz    HEK293T

./mimseq_k562_1.fastq.gz    K562

./mimseq_k562_2.fastq.gz    K562
```

Running mimseq

⏲ Timing: 15 min–1 h

```
usage: mimseq [options] sample data
```

The mimseq pipeline and its most important parameters are schematically outlined in Figure 3.

Running the mimseq command only requires that the sample data input file is specified last. All other parameters can occur in any order.

142. Customize [optional] parameters.

Note: Listed below are the major customizable parameters that should be considered. This is not an exhaustive list; please see mimseq –help for all parameters. Parameters that are required are indicated.

    a. -n: Experiment name. Output files and indices will have this prefix. **REQUIRED.**
    b. –out-dir: Output directory name. This directory cannot exist. If it does, for example, from a previous mimseq run, please remove it and rerun mimseq. Default is the current directory.
    c. –species or -t, -o and -m: Select species/genome of interest (see https://mim-trnaseq.readthedocs.io/en/latest/start.html#pre-built-references), or specify input files (see preparing mimseq input files). **REQUIRED;** either –species, or -t and -o.
    d. –control-condition: Specify the condition you would like to use as the control. All comparisons in DESeq2 and differential modification analysis will be reported relative to the control. This must match one of the conditions listed in the sample data file exactly. **REQUIRED.**
    e. –threads: Total available processors to use during analysis (particularly alignment). Please scale according to resource availability on your server/computer.
    f. –cluster-id: Cluster identity threshold for tRNA clustering, between 0 and 1. This is a genome-specific parameter that needs to be determined for the specific needs of the user. Generally, a good choice will maximize uniquely mapping reads and minimize
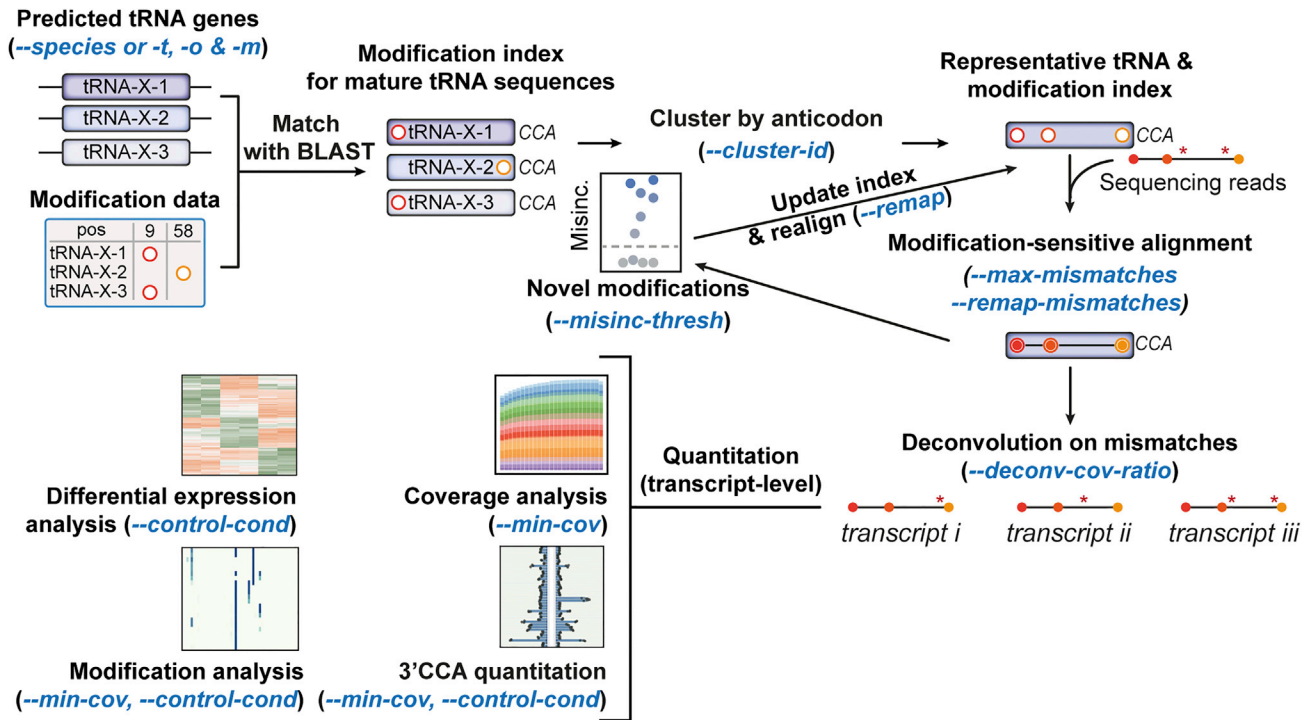
Figure 3. Schematic of the mimseq computational pipeline
Highlighted are the main customizable parameters and the analysis steps they affect. See text in running mimseq and mimseq –help for detailed parameter descriptions.

multi-mapping reads with little compromise to total deconvoluted sequences. In our experience, this value usually lies between 0.90 (e.g., yeast tRNAs) and 0.97 (human tRNAs).

Note: When analyzing your own data for the first time, we recommend performing several mimseq runs with –cluster-id set to 0.90, 0.93, 0.95, 0.97, or 1, and otherwise identical settings. Inspect alignment rate plots and deconvoluted transcript fractions printed in the log file to identify the optimal settings for your datasets, biological question, and organism of interest. A –cluster-id 1 run is equivalent to the commonly used tRNA read alignment strategy of collapsing multi-copy tRNA genes into a single reference. In our experience, this results in up to 25% of multi-mapping reads, depending on the organism.

g. –deconv-cov-ratio: Threshold of the required ratio between coverage at 3′ end and mismatch used for deconvolution. Coverage reductions greater than the threshold will result in non-deconvoluted sequences. This should be adjusted according to the general coverage quality of your libraries (see plots in the *cov/* output folder and expected outcomes). Low sequence coverage at 5′ ends with a high –deconv-cov-ratio will result in more non-deconvoluted sequences.

h. –max-mismatches and –remap-mismatches: Controls the proportion of mismatches allowed as a fraction of read length in the first and second round of alignment, respectively (excluding known and predicted modification sites). Due to new modification detection after the first round of alignment, it is generally advisable to reduce –remap-mismatches relative to –max-mismatches to reduce spurious and inaccurate read alignment in the second round. **Setting either parameter higher than 0.1 is not advisable!**

Note: When analyzing your own data for the first time, we recommend performing several mimseq runs with –max-mismatches set to 0.075 or 0.1 and –remap-mismatches set to

0.05 or 0.075 and otherwise identical settings. Inspect alignment rate plots and deconvo-luted transcript fractions printed in the log file to identify the optimal settings for your data-sets, biological question, and organism of interest.

    i. –min-cov: Minimum coverage per unique tRNA after <u>deconvolution</u> required for inclusion in coverage plots, modification analysis, and 3′-CCA analysis. This can be a fraction of total mapped reads between 0 and 1 or an integer representing absolute coverage. Note that all clusters are included for differential expression analysis with DESeq2. Default = 0.0005 (0.05% mapped reads).

    j. –max-multi: Maximum number of bam files to process simultaneously. Increasing this num-ber reduces processing time but increases total memory usage. Default is 3, maximum is the total number of samples.

Note: Processing too many files at once can cause termination of mim-tRNAseq due to insuf-ficient memory. If mim-tRNAseq fails during coverage calculation, lower this parameter.

    k. –misinc-thresh: Required fraction of reads per cluster or transcript containing a given mismatch to pass the novel modification detection threshold. Default is 0.1, i.e., 10% of cluster/transcript aligned reads must contain a given mismatch to call a new modification.

143. Run mimseq.

Note: An example of a mimseq command is given below using the example data in the GitHub repository. This utilizes the Hsap (hg38) reference with clustering at 97% sequence identity, allows 10% mismatches per read length on round 1 of alignment, and subsequently 7.5% on realignment, and filters all tRNAs with less than 0.05% total mapped reads from plots.

```
> mimseq –species Hsap –cluster-id 0.97 –threads 15 –min-cov 0.0005 –max-mismatches 0.1 –con-
trol-condition HEK293T –n hg38_test –out-dir hg38_HEK239vsK562 –max-multi 4 –remap –remap-
mismatches 0.075 sampleData_HEKvsK562.txt
```

Note: The above command will utilize the example data supplied in the GitHub repository. Please download the four fastq files and *sampleData_HEKvsK562.txt* file if you want to repro-duce this analysis.

Cluster deconvolution

By clustering tRNAs within an anticodon family by a sequence identity threshold, reads are aligned to a representative cluster parent, which substantially reduces multi-mapping for reads from nearly identical tRNA transcripts (Behrens et al., 2021). To restore single-transcript resolution for subse-quent analyses, we developed a cluster deconvolution algorithm which reassesses aligned reads and reassigns them to tRNA transcripts based on mismatch patterns to the cluster parent sequence.

Note: Since the original publication, we have introduced several major updates to the cluster deconvolution algorithm (Figure 4A). Since v0.3, the set of all mismatches between each unique tRNA transcript and the cluster parent are considered, rather than single mismatches as in earlier versions. From the full set of mismatches, unique 3′ subsets are determined, and aligned reads are searched for these mismatches to reassign them to unique transcripts within a cluster. This theoretically allows the distinction and deconvolution of all unique tRNA tran-scripts. When analyzing tRNA pools from an organism that encode a high number of highly similar tRNA transcripts, such as mammals, a small proportion of clusters, or particular tran-scripts within a cluster, cannot be deconvoluted due to three main reasons (Figure 4B):
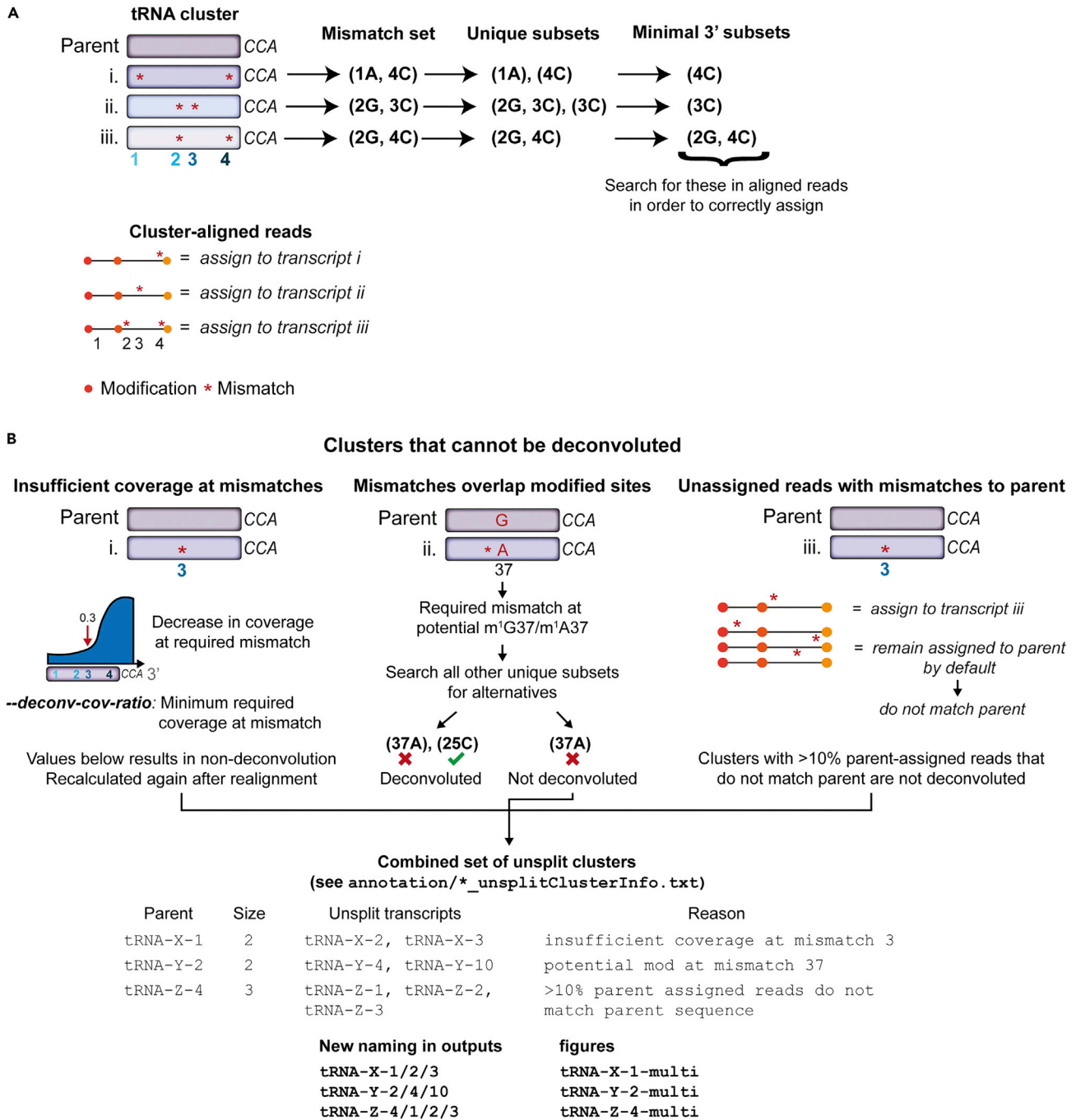
Figure 4. Fundamental principles of the new cluster deconvolution algorithm

(A) Schematic representation of deconvolution methodology. For each cluster, the set of mismatches distinguishing each transcript is found. For each set, the minimal unique subsets are found, from which the most 3′ subset is chosen. Reads are assessed individually for these mismatches in order to assign them to a member transcript within a cluster.

(B) Schematic representation of conditions when clusters or transcripts cannot be deconvoluted. Either coverage at a required mismatch is too low (left), the mismatch is a potentially modified site (middle), or >10% parent assigned reads contain mismatches to the parent.

A small subset of clusters may still exhibit 3′ coverage bias due to modifications that induce stops to RT, such as the rare ms^{2}t^{6}A/ms^{2}i^{6}A (Behrens et al., 2021). Complete deconvolution of reads for such clusters might not be possible if coverage is significantly lower at mismatches required for

deconvolution than at the 3′ end of the tRNA. To overcome this, the –deconv-cov-ratio parameter can be used to set a threshold for this difference in coverage (see running mimseq). Coverage at positions required for deconvolution not passing this threshold will result in the transcript being marked as not deconvoluted.

Some tRNA transcripts are only distinguishable from the parent by positions that might also be modified sites (for example, G26 or A58). In these cases, it is impossible to tell if a mismatch in a read is due to mismatches between cluster members or if it is due to misincorporations at the modified nucleotide. Such tRNA transcripts (and the parent of the cluster) are also labeled as not deconvoluted.

Thirdly, reads that cannot be assigned to a transcript within a cluster are by default left assigned to the parent sequence. In very rare cases, these parent-assigned reads also contain mismatches that do not pass our modification calling criteria, which indicates that they might not originate from the parent transcript either. If 10% or more parent-assigned reads contain such mismatches, the entire cluster is not deconvoluted.

Transcripts that are not deconvoluted are grouped and their count and modification data are aggregated. These are then renamed to provide information on which transcripts remain clustered, and are treated similarly to other single transcripts for differential expression analysis, modification profiling, and other downstream analyses (Figure 4B). For further information on non-deconvoluted transcripts and clusters, please refer to the *annotation/*unsplitClusterInfo.txt* output file.

EXPECTED OUTCOMES

There are a number of expected outcomes from the mimseq computational pipeline that can be used as quality control (QC) for your experiment or to guide further optimizations to the library construction protocol and/or the parameters used for running mimseq. Below are examples and details of these outputs. Please refer to the package documentation for a description of the full set of mimseq outputs (https://mim-trnaseq.readthedocs.io/en/latest/output.html).

Alignment (see *align/* folder)

An important QC step is ensuring that the majority of sequencing reads are aligned uniquely to the reference of choice, with minimal multi-mapping reads. To assess this, see *mapping_stats.txt*, and *Remap_alignstats.pdf* if –remap is enabled, or *Primary_alignstats.pdf* otherwise.

> *Note:* If –remap is enabled, each library will have two entries in the text file, the first representing the results from the first round of alignment, while those labeled "** NEW ALIGNMENT **" indicate statistics after realignment. Please only assess the new alignment results.

For eukaryotic tRNA libraries constructed with our workflow, there should be >70% uniquely mapped reads and <5% multi-mapped reads (Figure 5). Many factors can influence these fractions, including various aspects of library preparation (RNA integrity, 3′ adapter ligation, RT, and cDNA circularization efficiency), the organism and tRNA reference, and the stringency of alignment (mismatch allowance). Please consider all of these aspects when troubleshooting low read mappability.

Coverage and full-length transcript proportions (see *cov/* and *mods/* folders)

Our optimized RT protocol (steps 97–107) reduces RT stops at modified tRNA sites by promoting misincorporation. This should lead to a high proportion of full-length reads and reduced 3′ end bias in libraries. An easy way to assess this is by looking at the metagene plots for total normalized coverage per library scaled to the 3′ end in *cov/coverage_byaa_norm_scaled.pdf* (Figure 6A). For the test data, there is consistently more than 62.5% cumulative coverage at tRNA 5′ ends, indicating efficient readthrough of modifications and a majority of full-length transcripts. Full-length transcript
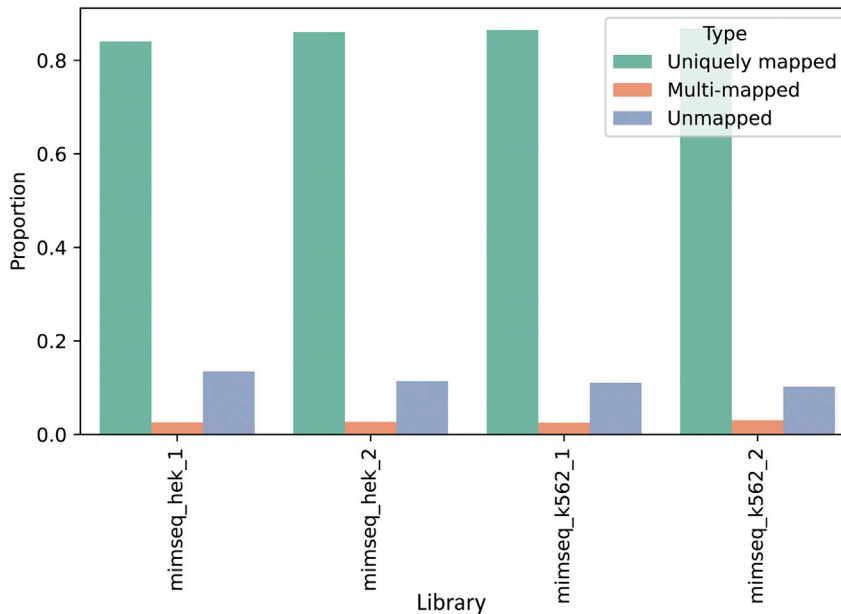
Figure 5. Alignment statistics for sample human dataset
Shown are the uniquely mapping, multi-mapped, and unmapped read proportions per library after realignment (if enabled with –remap as in step 5).

proportions can also be assessed by looking at *mods/RtstopTable.csv.* Here, you should see a high level of reads that stop (proportion column) for each tRNA transcript close to the 5′ end (canon_pos value close to 1, or 0 for tRNA-His). This is also visually represented in the "RT stops" heatmaps for each condition (top plot in *mods/*comb_heatmap.pdf*; Figure 6B).

> *Note:* The –min-cov parameter will influence the filtering of tRNA transcripts with low coverage. In the heatmap below (Figure 6B), only filtered transcripts are displayed. If you notice many transcripts in these plots with noisy stop and misincorporation data, please check the value used for –min-cov and the proportion of stops and misincorporations for these transcripts and their counts in the *counts/* output folder. Raising the –min-cov threshold might help filter out these noisy, low coverage transcripts.

Normalized tRNA counts and replicate clustering (see *DESeq2/*)
If your samples include replicates, DESeq2 is automatically run using your desired –control-condition as a basis for comparison. Ideally, replicates would show high similarity to each other with regards to DESeq2 variance stabilizing transformed tRNA count data (Love et al., 2014) to ensure high sensitivity for calling differentially expressed transcripts and anticodon pools (*isodecoder/* and *anticodon/* sub-directories, respectively; see *vst-transformedCounts.csv* for transformed count data). Here, it is useful to assess the Euclidean distance between the samples (*qc-sampledists.png;* Figure 7A), which should be low between replicates. Replicates should also cluster well, especially on principal component 1 in the PCA plots based on normalized count data for tRNA transcripts (*isodecoder/qc-pca.png;* Figure 7B). Poor clustering between samples might indicate biological variation in the samples (e.g., due to heterogeneity in cell-type composition of tissues) or large technical variation during library preparation and sequencing.

> ⚠ CRITICAL: We advise against using spike-in tRNAs for global-scaling data normalization and absolute tRNA quantification. The overall RNA content per cell, as well as the fraction of total RNA composed of tRNA, can vary substantially among different tissues, cell types, and growth conditions due to many biological factors (Coate and Doyle, 2015). These differences
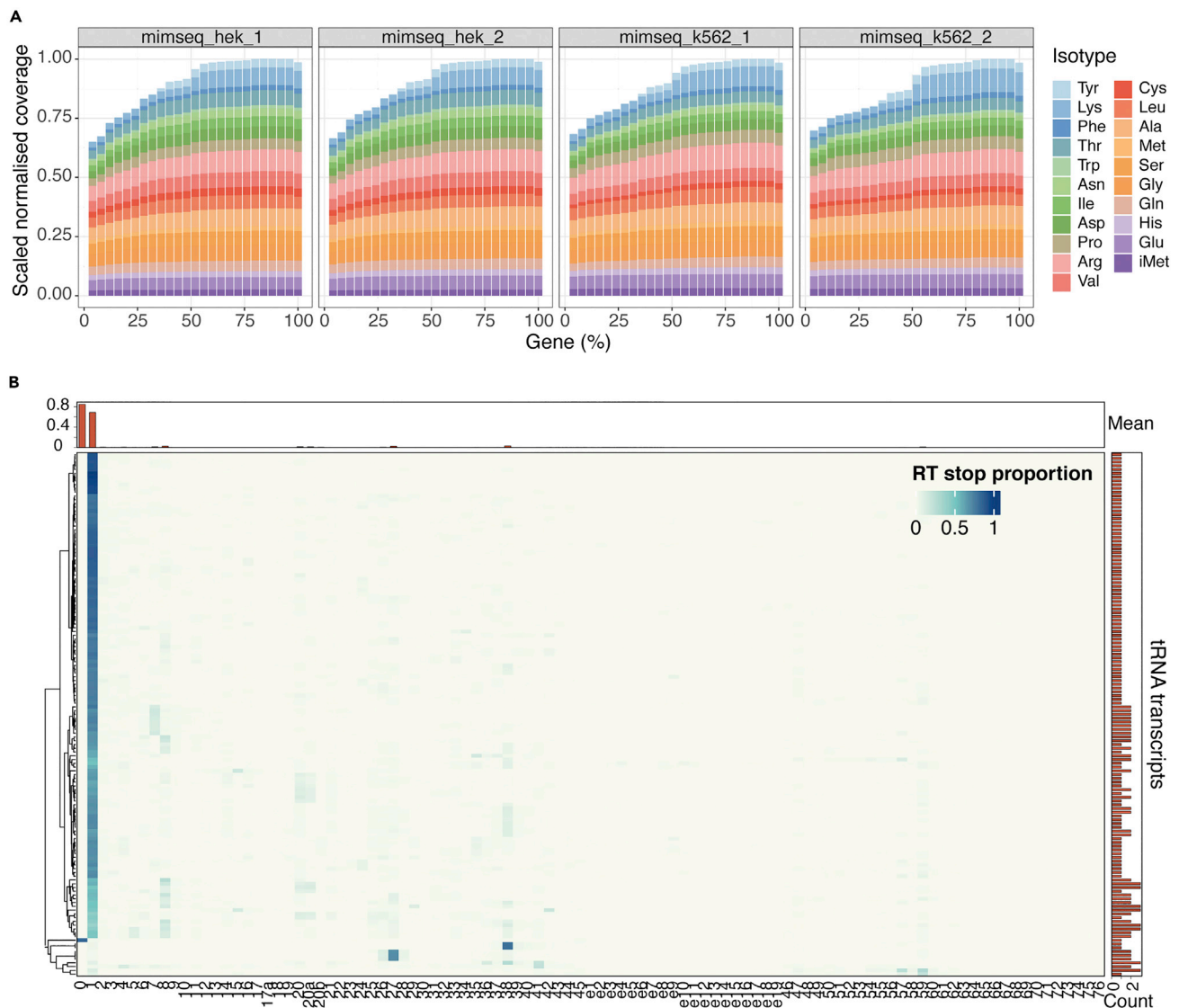
Figure 6. Quality control for tRNA coverage and full-length transcripts

(A) Metagene plots of coverage per nuclear-encoded tRNA isotype in each library specified in the sample data input file. Coverage is normalized to total mapped reads and scaled to the second last bin. See *cov/coverage_byaa_norm_scaled.pdf*.

(B) Global heatmaps of average proportions of stops to RT per canonical tRNA position for each unique tRNA transcript with coverage above 0.005 (as per –min-cov) in tRNA sequencing data from human K562 cells (n = 2).

can unduly influence the number of reads mapping to the spike-in in each library, as the spike-in is added to a fixed amount of total RNA rather than to a fixed number of cells.

Note: If statistical power is low, which may be the case for more heterogeneous samples such as tissues, we recommend increasing the number of biological replicates rather than the sequencing depth (Liu et al., 2014). Higher sequencing coverage may still be beneficial in specific cases, e.g., when analyzing misincorporation patterns at modified sites in low-abundance tRNA transcripts.

LIMITATIONS

We have successfully used mim-tRNAseq to profile tRNA abundance, aminoacylation, and modification status in samples from a wide range of eukaryotic organisms such as yeast, flies, plants, mouse
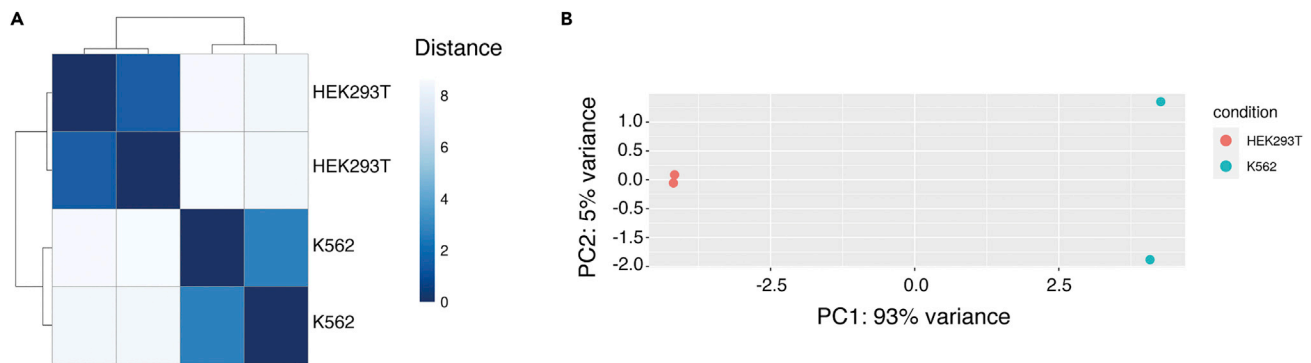
Figure 7. Assessing replicate similarity with variance stabilizing transformed (vst) tRNA count data from DESeq2
(A) Distance matrix representing pairwise Euclidean distance for each pair of samples.
(B) Principal component analysis (PCA) plot using the first two principal components from tRNA isodecoder analysis. Percent variance explained by each principal component is given in axis titles.

tissues, and human cell lines. The experimental workflow is applicable to any organism, though it may require some optimization depending on the type and frequency of tRNA modifications present in the sample. We have successfully generated high-quality libraries starting with as little as 0.5 μg of total RNA. The protocol is currently not compatible with ultra-low input samples or single-cell methodologies.

The computational pipeline has a degree of customization to allow the user to fine-tune the analysis to their particular needs and can also be used with tRNA sequencing datasets generated with other tRNA library construction protocols. It is designed for single-end sequencing reads and currently does not work on paired-end ones. However, since mature tRNAs are <= 100 nt in length, paired-end data are unnecessary. Finally, mim-tRNAseq is currently not designed for investigating pre-tRNAs, but this functionality is currently under development.

TROUBLESHOOTING
Problem 1
Low yield of full-length cDNA after reverse transcription.

Potential solution
Performing reverse transcription of adapter-ligated tRNA under low-salt conditions and for an extended time is critical for obtaining a high proportion of full-length cDNA molecules. Typically, more than 50% of the RT primer is extended, and the major cDNA products visible on gel are full-length (Figure 2C). Efficient reverse transcription is critically important for minimizing amplification bias during library construction PCR.

Poor reverse transcription manifests as the extension of only a small fraction of the RT primer (Figure 2E) or the presence of shorter cDNA fragments detectable as dominant and discrete bands on gel (Behrens et al., 2021). There are two potential causes for this: i) ethanol carry-over from template precipitation, which can inhibit reverse transcription, and ii) loss of TGIRT activity upon prolonged storage. We, therefore, recommend purifying adapter-ligated tRNA with the Zymo Oligo Clean & Concentrator after step 96. This can be done routinely if sample input amounts are not limited. In addition, we have observed that TGIRT can lose activity when stored for more than 3 months at −20°C. We recommend −80°C for prolonged enzyme storage while avoiding repeated freeze-thaw cycles.

Problem 2
When running mimseq, the step Analyzing misincorporations and stops to RT, and analyzing 3′ ends produces the following:

```
IndexError: Too many levels: Index has only 1 level, not 2
```

Potential solution

This error is often caused by an empty or sparse misincorporation table (*mods/mismatchTable.csv*). This might result from very poor alignment. First, check that adapter trimming has been done correctly by analyzing cutadapt logs (steps 1–3). Validate your trimming approach on another data-set, or by manually trimming reads at both ends to ensure some alignment.

Problem 3

When running mimseq, errors occur during tRNA processing (Processing tRNA sequences...) or shortly after. For example:

```
ID = re.search("tRNAscan-SE ID: (.*?)\).|\((chr.*?)-",seqIO_dict[seqIO_record].description).
groups() AttributeError: 'NoneType' object has no attribute 'groups'

tRNA_dict[seq]['anticodon'] = anticodon = re.search('.tR(NA|X)-.?-(.*?)-', seq).group(2)

AttributeError: 'NoneType' object has no attribute 'group'

anticodon = seq_parts[4] IndexError: list index out of range
```

Potential solution

This usually indicates problems with your input reference files (see preparing mimseq input files). You have most likely specified custom input files with -t, -o, and/or -m. If these files are not present on GtRNAdb (http://gtrnadb.ucsc.edu/) for your species of interest, please ensure that the header for each sequence in the fasta file of genomic tRNAs is formatted exactly the same as those from in the mimseq pre-built indices. An example can be found here; pay close attention to the order of information, the number of fields separated by spaces, the naming convention for tRNA genes, and the tRNAScan-SE ID given in parentheses that matches the corresponding entry in the out file (specified with -o).

> *Note:* If mitochondrial and/or plastid sequences are specified with -m, these also require specific formatting that is distinct from the nuclear genomic file as they match the format provided by the mitotRNAdb (http://mttrna.bioinf.uni-leipzig.de/mtDataOutput/). If formatting is incorrect, mimseq will produce an error similar to the third error above. See an example here for correct formatting; again, pay close attention to the number of fields per sequence header (i.e., 5) and the use of "|" as a field separator. In this case, the first field specifying the ID can be any user-chosen value. The third field is a unique species code, which is unused by mimseq but must be present.

Problem 4

Running mimseq fails at the alignment step with a non-zero exit status 9 error during GSNAP alignment:

```
subprocess.CalledProcessError: Command '['gsnap', '-gunzip', '-D', 'hg38_HEK239vsK562/
Hsap_tRNAgenome', '-d', 'Hsap_tRNAgenome', '-V', 'hg38_HEK239vsK562/Hsapsnp_index',
'-v', 'hg38_diff_modificationSNPs'...]' returned non-zero exit status 9.
```

Potential solution

These errors occur when mimseq tries to run GSNAP for read alignment. In this case, the align.log file in the *align/* folder can be very useful for debugging. Most commonly, there is an error in the path to

the trimmed fatsq files supplied in the sample data file (see preparing mimseq input files). Carefully check these paths to make sure they point to files that exist. In this case, the log file will show something such as the following:

```
Cannot open gzipped file ./mimseq_k562_.fastq.gz
```

Other possibilities include incorrect GSNAP version installations. Please ensure that GSAP version 2019-02-26 is installed by typing "gsnap –version" in the terminal within your mimseq environment. With newer versions, an error will be produced due to changes in parameters available in the gsnap command:

```
gsnap.avx512: unrecognized option '-ignore-trim-in-filtering'
```

Problem 5
Samples show low uniquely mapped read proportions in *align/mapping_stats.txt*.

Potential solution
This problem may arise for several reasons. First, reassess your read trimming step and ensure that the correct barcodes and adapter sequences were specified in *barcodes.fa*. Check that the proportions of reads trimmed in the log files are as expected (>80% reads with adapters) and that not too many reads were excluded because they were too short. A high proportion of short reads may indicate poor modification readthrough during RT or too many PCR cycles during library construction, resulting in overamplifying short cDNA fragments. Ensure that RT is performed with templates of high purity and with a fresh enzyme batch, and minimize PCR cycles.

Secondly, check that your alignment and realignment mismatch allowances (step 142) are not too stringent. Try raising these values and assessing how this impacts the alignment statistics for each alignment round. Be cautious not to raise these values too much as this may cause misalignment and spurious modification calling (see *mods/predictedMods.csv*).

Lastly, evaluate if you have contamination of other RNA types in your sample, such as rRNA or snoRNA. This may result from RNA degradation during sample collection and/or RNA isolation (steps 1–24) or from imprecise cutting out of tRNA-containing gel fragments in step 67 (Figure 2A). Align trimmed reads to the full genome of the species of interest and assess areas and gene features with high read coverage to identify potential contaminants and optimize RNA isolation and/or tRNA size selection accordingly.

Problem 6
Spike-in sequences other than *E. coli* tRNA-Lys-UUU (Table S1) are used during library generation and spike-in addition.

Potential solution
mimseq reference files can be easily edited to include new sequences of interest. The only requirement is that the formatting of the sequence header is maintained as in other tRNA references.

There are two methods to achieve this:

- The reference fasta files can be downloaded from GitHub (https://github.com/nedialkova-lab/mim-tRNAseq/tree/master/mimseq/data), edited, and specified to mimseq with -t. Please note that the corresponding intron information *.out* file also needs to be specified with -o and can also be downloaded from the link above.

- The reference files included in your local installation of mimseq can be directly modified. To find the location of these, activate your mimseq environment, determine the location of the mimseq executable file, and use this path to find the included *data/* folder. For example:

```
# find mimseq executable

> which mimseq

/home/drew/anaconda3/envs/mimseq/bin/mimseq
```

In this case, the prebuilt references will be found in.

```
/home/drew/anaconda3/envs/mimseq/lib/python3.7/site-packages/mimseq/data/
```

The appropriate reference folder can be found here, and the fasta file within can be edited and saved.

This is more problematic if something goes wrong as the mimseq master files would have been permanently changed. However, this method does allow you to specify the reference simply with the –species parameter.

In both cases, if an unspliced intron-containing spike-in is used, there will need to be a matching entry in the corresponding *.out* file with a matching tRNAScan-SE ID number to ensure correct splicing.

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Danny D. Nedialkova (nedialkova@biochem.mpg.de).

Materials availability
This study did not generate new unique reagents.

Data and code availability
The mim-tRNAseq computational pipeline is available under a GNU public License v3 on GitHub (https://github.com/nedialkova-lab/mim-tRNAseq), Zenodo (https://doi.org/10.5281/zenodo.6694873) and on Bioconda. The accession number for the sequencing data reported in the original publication is GEO: GSE152621. Example analyses presented here are based on a subset of the replicate HEK293T and K562 data (GEO: GSM4618859, GSM4618860, GSM4618861, GSM4618862) which are also available in the GitHub repository. A package description and installation guide are available at https://mim-trnaseq.readthedocs.io/en/latest/. Table S1 containing oligonucleotides, RNA sequences and primers for library construction can be found on Mendeley Data:https://doi.org/10.17632/vy8z394gfh.1.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xpro.2022.101579.

AUTHOR CONTRIBUTIONS

Conceptualization, D.D.N.; experimental methodology, D.D.N.; software, A.B.; writing, A.B. and D.D.N.; supervision and funding acquisition, D.D.N.

DECLARATION OF INTERESTS

A.B. and D.D.N. are inventors on a patent application filed by the Max Planck Society pertaining to the mim-tRNAseq technology.

REFERENCES

Auer, P.L., and Doerge, R.W. (2010). Statistical design and analysis of RNA sequencing data. Genetics *185*, 405–416.

Behrens, A., Rodschinka, G., and Nedialkova, D.D. (2021). High-resolution quantitative profiling of tRNA abundance and modification status in eukaryotes by mim-tRNAseq. Mol. Cell *81*, 1802–1815.e7.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinf. *10*, 421.

Chan, P.P., Lin, B.Y., Mak, A.J., and Lowe, T.M. (2021). tRNAscan-SE 2.0: Improved detection and functional classification of transfer RNA genes. Nucleic Acids Res. *49*, 9077–9096.

Coate, J.E., and Doyle, J.J. (2015). Variation in transcriptome size: are we getting the message? Chromosoma *124*, 27–43.

Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., and de Hoon, M.J.L. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics *25*, 1422–1423.

Czech, A., Wende, S., Mörl, M., Pan, T., and Ignatova, Z. (2013). Reversible and rapid transfer-RNA deactivation as a mechanism of translational repression in stress. PLoS Genet. *9*, e1003767.

Dittmar, K.A., Sørensen, M.A., Elf, J., Ehrenberg, M., and Pan, T. (2005). Selective charging of tRNA isoacceptors induced by amino-acid starvation. EMBO Rep. *6*, 151–157.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics *26*, 2460–2461.

Evans, M.E., Clark, W.C., Zheng, G., and Pan, T. (2017). Determination of tRNA aminoacylation levels by high-throughput sequencing. Nucleic Acids Res. *45*, e133.

Green, M.R., and Sambrook, J. (2019). Isolation of DNA fragments from polyacrylamide gels by the crush and soak method. Cold Spring Harb. Protoc. *2019*, 143–146.

Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847–2849.

Guo, L.-T., Adams, R.L., Wan, H., Huston, N.C., Potapova, O., Olson, S., Gallardo, C.M., Graveley, B.R., Torbett, B.E., and Pyle, A.M. (2020). Sequencing and structure probing of long RNAs using MarathonRT: a next-generation reverse transcriptase. J. Mol. Biol. *432*, 3338–3352.

Gustafsson, H.T., Galan, C., Yu, T., Upton, H.E., Ferguson, L., Kaymak, E., Weng, Z., Collins, K., and Rando, O.J. (2022). Deep sequencing of yeast and mouse tRNAs and tRNA fragments using OTTR. Preprint at bioRxiv. https://doi.org/10.1101/2022.02.04.479139.

Heyer, E.E., Ozadam, H., Ricci, E.P., Cenik, C., and Moore, M.J. (2015). An optimized kit-free method for making strand-specific deep sequencing libraries from RNA fragments. Nucleic Acids Res. *43*, e2.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

Liu, Y., Zhou, J., and White, K.P. (2014). RNA-seq differential expression studies - more sequence or more replication? Bioinformatics *30*, 301–304.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. J. *17*, 10.

McGlincy, N.J., and Ingolia, N.T. (2017). Transcriptome-wide measurement of translation by ribosome profiling. Methods *126*, 112–129.

Mohr, S., Ghanem, E., Smith, W., Sheeter, D., Qin, Y., King, O., Polioudakis, D., Iyer, V.R., Hunicke-Smith, S., Swamy, S., et al. (2013). Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. RNA *19*, 958–970.

Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics *29*, 2933–2935.

Peacock, J.R., Walvoord, R.R., Chang, A.Y., Kozlowski, M.C., Gamper, H., and Hou, Y.-M. (2014). Amino acid-dependent stability of the acyl linkage in aminoacyl-tRNA. RNA *20*, 758–764.

Qin, Y., Yao, J., Wu, D.C., Nottingham, R.M., Mohr, S., Hunicke-Smith, S., and Lambowitz, A.M. (2016). High-throughput sequencing of human plasma RNA by using thermostable group II intron reverse transcriptases. RNA *22*, 111–128.

Quail, M.A., Otto, T.D., Gu, Y., Harris, S.R., Skelly, T.F., McQuillan, J.A., Swerdlow, H.P., and Oyola, S.O. (2011). Optimal enzymes for amplifying sequencing libraries. Nat. Methods *9*, 10–11.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Richter, F., Plehn, J.E., Bessler, L., Hertler, J., Jörg, M., Cirzi, C., Tuorto, F., Friedland, K., and Helm, M. (2022). RNA marker modifications reveal the necessity for rigorous preparation protocols to avoid artifacts in epitranscriptomic analysis. Nucleic Acids Res. *50*, 4201–4215.

Upton, H.E., Ferguson, L., Temoche-Diaz, M.M., Liu, X.M., Pimentel, S.C., Ingolia, N.T., Schekman, R., and Collins, K. (2021). Low-bias ncRNA libraries using ordered two-template relay: serial template jumping by a modified retroelement reverse transcriptase. Proc. Natl. Acad. Sci. USA *118*. e2107900118.

van Dijk, E.L., Jaszczyszyn, Y., and Thermes, C. (2014). Library preparation methods for next-generation sequencing: tone down the bias. Exp. Cell Res. *322*, 12–20.

Wu, T.D., and Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics *26*, 873–881.

Yip, M.C.J., Savickas, S., Gygi, S.P., and Shao, S. (2020). ELAC1 repairs tRNAs cleaved during ribosome-associated quality control. Cell Rep. *30*, 2106–2114.e5.

Zhao, C., Liu, F., and Pyle, A.M. (2018). An ultraprocessive, accurate reverse transcriptase encoded by a metazoan group II intron. RNA *24*, 183–195.

Zhuang, F., Fuchs, R.T., Sun, Z., Zheng, Y., and Robb, G.B. (2012). Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. Nucleic Acids Res. *40*, e54.