nature protocols

https://doi.org/10.1038/s41596-025-01158-4

Protocol

Check for updates

Optimized identification and characterization of small RNAs with PANDORA-seq

Junchao Shi[1] ✉, Yunfang Zhang[2,3] ✉, Yun Li[2], Liwen Zhang[1], Xudong Zhang[4,5,6], Menghong Yan[7], Qi Chen[4,5,6] & Ying Zhang[8] ✉

Abstract

Small noncoding RNAs (sncRNAs) are a diverse group of RNAs including small interfering RNAs, microRNAs, PIWI-interacting RNAs and RNAs derived from structured RNAs such as transfer RNAs, ribosomal RNAs and others. These sncRNAs have varied termini and RNA modifications, which can interfere with adaptor ligation and reverse transcription during cDNA library construction, hindering detection of many types of sncRNA by standard small RNA sequencing methods. To address this limitation, PANDORA sequencing introduces a refined methodology. The procedure includes sequential enzymatic treatments of size-selected RNAs with T4PNK and AlkB, which effectively circumvent the challenges presented by the ligation-blocking termini and reverse transcription-blocking RNA modifications, followed by tailored small RNA library construction protocols and deep sequencing. The obtained datasets are analyzed with the SPORTS pipeline, which can comprehensively analyze various types of sncRNA beyond the traditionally studied classes, to include those derived from various parental RNAs (for example, from transfer RNA and ribosomal RNA), as well as output the locations on the parental RNA from which these sncRNAs are derived. The entire protocol takes ~7 d, depending on the sample size and sequencing turnaround time. PANDORA sequencing provides a transformative tool to further our understanding of the expanding small RNA universe and to explore the uncharted functions of sncRNAs.

Key points

- PANDORA sequencing employs sequential enzymatic treatment steps to modify the termini of small noncoding RNAs and remove reverse-transcription blocking methylation for optimized in depth genome-wide profiling of these diverse RNA species that extends beyond well-studied classes of small RNAs.
- Complemented by a custom analytical pipeline (SPORTS), the protocol characterizes the features and relative proportions of different sncRNAs and maps them to the parental RNAs they derive from.

Key references

Shi, J. et al. *Nat. Cell Biol.* **23**, 424–436 (2021): https://doi.org/10.1038/s41556-021-00652-7

Shi, J. et al. *Genom. Proteom. Bioinform.* **16**, 144–151 (2018): https://doi.org/10.1016/j.gpb.2018.04.004

Shi, J. et al. *Nat. Cell Biol.* **24**, 415–423 (2022): https://doi.org/10.1038/s41556-022-00880-5

[1]China National Center for Bioinformation and Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China. [2]Clinical and Translational Research Center of Shanghai First Maternity and Infant Hospital, Shanghai Key Laboratory of Signaling and Disease Research, Frontier Science Center for Stem Cell Research, School of Life Sciences and Technology, Tongji University, Shanghai, China. [3]Sycamore Research Institute of Life Sciences, Shanghai, China. [4]Molecular Medicine Program, University of Utah School of Medicine, Salt Lake City, UT, USA. [5]Division of Urology, Department of Surgery, University of Utah School of Medicine, Salt Lake City, UT, USA. [6]Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT, USA. [7]Pudong Medical Center, Institutes of Biomedical Sciences, Shanghai Medical College, Fudan University, Shanghai, China. [8]The Key Laboratory of Cell Proliferation and Regulation Biology, Ministry of Education, Beijing Key Laboratory of Genetic Engineering Drug and Biotechnology, College of Life Sciences, Beijing Normal University, Beijing, China. ✉e-mail: shijc@cncb.ac.cn; zhangyunfang@tongji.edu.cn; yingzhanglab@gmail.com

Introduction

The limitations of traditional small RNA-seq

It is now increasingly recognized that the world of small noncoding RNAs (sncRNAs) extends far beyond the well-studied small interfering RNAs (siRNAs), microRNAs (miRNAs) and PIWI-interacting RNAs (piRNAs), encompassing a diverse universe of previously underexplored sncRNAs derived from various longer structured RNAs such as transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), Y RNAs (yRNAs) and vault RNAs (vtRNAs)[1] (Table 1).

Importantly, these expanding categories of sncRNAs are challenging to discover using traditional small RNA sequencing (RNA-seq) protocols[2,3], which involve terminal adaptor ligation, followed by reverse transcription (RT), to obtain a cDNA library of the sncRNA for high-throughput sequencing and were primarily designed to capture miRNAs. These protocols are based on the premise that sncRNAs such as miRNAs bear 5′-phosphate (5′-P) and 3′-hydroxyl (3′-OH) termini (generated by Dicer) and contain few internal RNA modifications that interfere with reverse transcriptase[1,4]. However, now we understand that the actual sncRNA pool in cells and tissues is much more diverse, with many sncRNAs such as those derived from tRNAs and rRNAs (that is, tRNA-derived small RNAs (tsRNAs) and rRNA-derived small RNAs (rsRNAs)) bearing diverse RNA modifications, including methylation (for example, 1-methyladenosine (m^1A), 3-methylcytidine (m^3C), 1-methylguanosine (m^1G), N^2,N^2-dimethylguanosine (m^2_2G)) that can block the reverse transcriptase widely used in traditional small RNA-seq. In addition, the biogenesis of tsRNAs, rsRNAs and other sncRNAs involves processing by a range of RNases beyond Dicer[5], for example, by RNase T2 or RNase A, resulting in 5′-hydroxyl (5′-OH) and 2′,3′-cyclic phosphate (2′,3′-cP) termini (2′,3′-cP can be further hydrolyzed to a 3′-phosphate (3′-P)), which are distinct from the termini of miRNAs (5′-P and 3′-OH). The sncRNAs bearing 2′,3′-cP or 3′-P and/or 5′-OH termini cannot be efficiently ligated to adaptors using the traditional small RNA-seq protocols (designed for ligating 5′-P and 3′-OH termini); thus, these sncRNAs are excluded from the cDNA library. Owing to the limitations of traditional small RNA-seq, new methodologies that circumvent RT-blocking RNA modifications and the ligation-blocking termini of sncRNAs are essential to unveil a more comprehensive landscape of sncRNAs, and recent methods are burgeoning[1,6–17]. In addition, existing sncRNAs annotation tools primarily focus on canonical or specific types of sncRNA. Furthermore, some annotation tools employ feature-based counting strategies, which group sequencing reads by genomic features (genes or noncoding RNA species) rather than retaining individual sequence details[18]. This loss-of-sequence-level resolution hinders the ability to trace results back to raw data and compromises sequence integrity. Therefore, improved bioinformatics tools are needed to identify and characterize the newly recognized sncRNAs.

Table 1 | SncRNA categories according to their parental origin

Parental RNA	Derivative sncRNA
tRNA	tsRNA
rRNA	rsRNA
yRNA	ysRNA
vtRNA	vtsRNA
snRNA	snsRNA
snoRNA	snosRNA
Long non-coding RNA	lncsRNA
Messenger RNA	msRNA

lncsRNA, long non-coding-RNA-derived small RNA; msRNA, messenger-RNA-derived small RNA; snosRNA, snoRNA-derived small RNA; snsRNA, snRNA-derived small RNA; vtsRNA, vtRNA-derived small RNA; ysRNA, yRNA-derived small RNA.

Development and overview of PANDORA-seq

We developed PANDORA sequencing (PANDORA-seq), which incorporates step-wise enzymatic treatments (T4PNK + AlkB) to eliminate the obstacles presented by ligation-blocking termini and RT-blocking RNA modifications, respectively[6] (Fig. 1a). Subsequent to these treatments, the process includes adaptor ligation, RT and cDNA amplification steps that are tailored to RNA sample quantities to improve the success rate of library construction. PANDORA-seq has revealed the fact that miRNAs, once presumed to be the dominant small RNA species, comprise only 0.1–5% of the total sncRNA population across diverse cell and tissue types. By contrast, tsRNAs and rsRNAs represent the majority of sncRNAs in tissues, including mouse brain, liver, spleen and sperm, and in mouse embryonic fibroblasts and embryonic stem cells, as well as human HeLa cells, induced pluripotent stem cells and embryonic stem cells[6]. The deep sequencing datasets are then processed by our SPORTS1.1 pipeline[19], which can systematically analyze the abundance of various types of sncRNA beyond miRNA and piRNA, including sncRNAs derived from various parental RNAs, such as tRNA, rRNA, yRNA, snRNA, snoRNA and vtRNA, and output the possible location on the parental RNAs from which the sncRNAs are derived (Fig. 1b). The mapping results can facilitate the direct visual comparison of the location and quantity of sncRNAs of different origins between samples (for example, different tissues/cells, or different experimental groups).

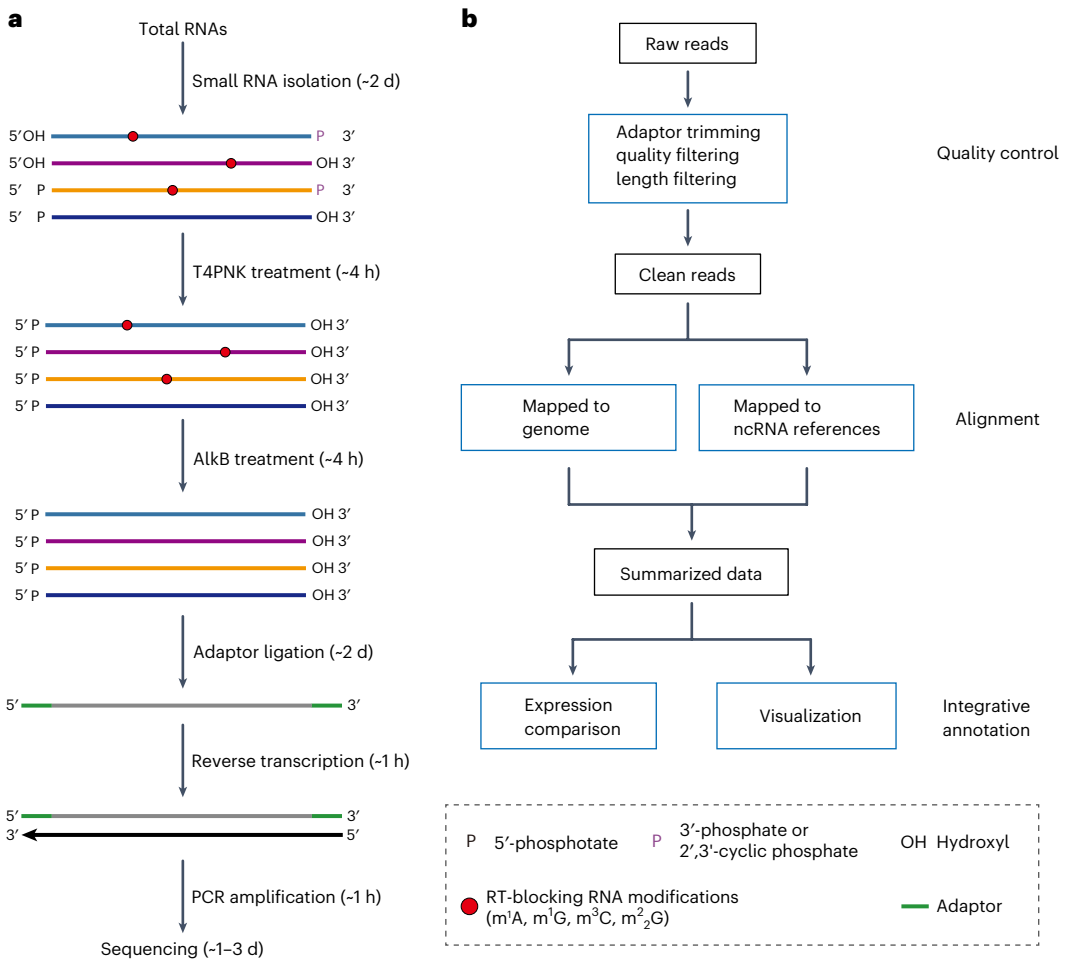


Fig. 1 | Overview of the PANDORA-seq protocol. a, A schematic representation of RNA preparation and library construction steps. Different colors represent RNAs with different RT-blocking modifications and different termini. **b**, A schematic representation of the data analysis pipeline for small RNA analysis using the SPORTS 1.1 pipeline. ncRNA, non-coding RNA.

Applications

Recent findings indicate that tsRNAs and rsRNAs are now emerging as dominant sncRNAs with critical roles in a plethora of biological processes including viral infection[20,21], cancer[22–24], immunological response[25], stem cell differentiation[6,26–28], epigenetic inheritance[29–37] and symbiosis[38,39]. These newly discovered sncRNAs have been increasingly harnessed as new biomarkers for a range of diseases and clinical conditions[40–47]. The systematic discovery of tsRNAs, rsRNAs and other previously less studies sncRNAs by PANDORA-seq under different biological and disease settings would open the next stage of sncRNA exploration. It is also increasingly recognized that many tsRNAs and rsRNAs do not function through the conventional RNA interference mechanisms[48,49]. Instead, they execute their roles through two-dimensional/three-dimensional structural features that go beyond linear complementarity, augmented by RNA modifications. Further explorations along this direction hold the potential to reshape our understanding of sncRNA biology and offer new avenues that could lead to therapeutic interventions based on the structural features of these RNAs.

Advantages, limitations and adaptations

PANDORA-seq provides a new tool in exploring the small RNA universe. Although this approach is not yet refined for single-cell analysis, our current protocol improves the small RNA isolation procedure by eliminating the gel crush step and optimizing the gel elution buffer and procedure, (slightly increasing the isolation efficiency compared to the originally published article[6]). Moreover, it incorporates calculations of primer and adaptor amounts during cDNA library construction, enabling the use of input as low as 1 ng of small RNAs. This allows for the detection of sncRNAs in a wide range of scarce samples. In addition, the sncRNA size selection step in PANDORA-seq prevents the generation of artifact sncRNAs, a problem not addressed in previous studies using AlkB treatment[7,11,12,50]. This issue arises from artificial degradation following the removal of modifications from precursor RNAs in a ferrous iron-dependent manner[51].

Notably, since PANDORA-seq increased the capacity of small RNA-seq in detecting various types of sncRNA, researchers who would like to focus on studying miRNAs or less modified small RNA species will find that the proportion of sequencing reads for these specific RNAs in the result becomes relatively low (as the sequencing reads have been consumed by other sncRNAs such as tsRNAs and rsRNAs). A straightforward solution to this is enhancing sequencing depth to garner more sncRNA reads or to employ traditional small RNA-seq methods[2,3] as an alternative. In addition, northern blot analysis is recommended to validate the observed RNA signatures and confirm potential changes in sncRNAs between samples, as the alterations in small RNA signatures identified solely through traditional small RNA-seq approaches may not accurately represent actual changes in their net expression levels[1].

Experimental design

Preparation and validation of AlkB enzyme activity

The AlkB enzyme, designed to remove several types of RNA methylation that impede the efficiency of RT on full-length sncRNAs during cDNA library preparation, is crucial for PANDORA-seq. The purity and activity of the AlkB enzyme substantially influence its demethylation efficiency and using impure enzyme may lead to RNA degradation during reaction steps, potentially resulting in the failure of library construction or the incorporation of excessively short fragmented sncRNA in the sequencing data. Thus, it is recommended to prepare the AlkB enzyme with expertise in protein expression and purification to ensure high-quality outcomes. In addition, we compared the demethylation efficiency of wild-type AlkB with that of the AlkB mutant D135S, as reported in previous studies[52], and demonstrated that the wild-type AlkB exhibits similar efficiency to the AlkB mixture in removing the corresponding RNA modifications (Extended Data Fig. 1 and Supplementary Table 1).

The AlkB gene used in this protocol features sequence variations at the amino terminus compared with previously documented AlkB sequences[6] (Table 2). These modifications enhance the enzyme's demethylation activity on $m^2{}_2G$, in addition to its established activity on m^1A, m^3C and m^1G. The demethylation activity of the purified AlkB protein must be confirmed through liquid chromatography–tandem mass spectrometry (LC–MS/MS) (Steps 24 to 50).

Table 2 | *AlkB* gene sequence used for protein expression

AlkB gene sequence	5′-CTGGACCTGTTCGCGGATGCGGAGCCGTGGCAGGAACCGCTGGCGGCGGGTGCGGTTATCCTGCGTCGTTTCGCGTTTAACGCGGCGGA GCAACTGATCCGTGACATTAACGATGTGGCGAGCCAGAGCCCGTTTCGTCAAATGGTTACCCCGGGTGGCTACACCATGAGCGTGGCGATGA CCAACTGCGGTCACCTGGGTTGGACCACCCACCGTCAGGGTTACCTGTATAGCCCGATCGACCCGCAAACCAACAAGCCGTGGCCGGCGAT GCCGCAGAGCTTCCACAACCTGTGCCAACGTGCGGCGACCGCGGCGGGTTACCCGGACTTTCAGCCGGATGCGTGCCTGATTAACCGTTAT GCGCCGGGTGCGAAGCTGAGCCTGCACCAAGACAAAGATGAGCCGGATCTGCGTGCGCCGATCGTTAGCGTGAGCCTGGGTCTGCCGGCG ATTTTCCAGTTTGGTGGCCTGAAGCGTAACGACCCGCTGAAACGTCTGCTGCTGGAGCACGGCGATGTGGTTGTGTGGGGTGGCGAAAGCCG TCTGTTCTACCACGGTATCCAGCCGCTGAAAGCGGGCTTTCACCCGCTGACCATTGACTGCCGTTATAACCTGACCTTCCGTCAAGCGGGTAA GAAAGAA-3′

Typically, 500 mL of *Escherichia coli* cell culture yields protein at the milligram level, which is sufficient for processing milligrams of RNA in subsequent experiments. In the demethylation reaction, the enzyme-to-RNA molar ratio is maintained at 1:1. For simplicity, a 1:1 weight ratio is used in practice for easier calculation.

Preparation of RNA samples

PANDORA-seq is adaptable to a wide range of biological samples, and we encourage to follow the ARRIVE guidelines[53] for reporting animal research when obtaining RNAs from animals. RNA degradation can markedly affect the accuracy of small RNA quantification and profiling, making it essential to assess RNA integrity using the RNA integrity number (RIN). Exceptions exist for certain samples, such as mouse sperm[54], sperm head[6] and serum[55], which should not contain 18S and 28S rRNA peaks (presence of these peaks would suggest sample contamination by somatic cells). As a result, the RIN for these samples is typically unmeasurable or very low (<3). We provide an alternative procedure to extract sncRNAs from sperm or sperm head samples (Box 1). To prevent degradation of precursor tRNA/rRNA during enzymatic treatment, it is required to isolate RNAs ranging from 15 to 50 nucleotides. It is also vital to recognize that there is variability in the proportion of small RNAs relative to the total RNAs across different samples. Therefore, the quantification of small RNA concentration (Step 62) is required to determine the appropriate enzyme amount for each sample. Moreover, after evaluating various small RNA isolation protocols, the optimized procedure based on denaturing gel purification is detailed from Step 56 to 62, ensuring the most efficient recovery of small RNAs.

BOX 1

Sample preparation for sperm and sperm head samples

It is crucial to avoid RNA contamination and minimize RNA degradation in biological samples to achieve accurate small RNA profiles. The RIN is commonly used to measure RNA integrity, but it may not be suitable for certain sample types, such as sperm, which typically lack the 18S and 28S rRNA peaks, resulting in an unmeasurable or very low RIN value (<3). For sperm samples, we recommend the following specialized procedure to extract RNAs from sperm and sperm heads.

Procedure

1. Release mouse sperm from the cauda epididymis of one mouse into 5 mL of PBS. Immediately incubate this suspension at 37 °C for 15 min. Filter 4 mL of the upper supernatant though a 40-μm cell strainer. This prepares the sample for subsequent steps, either to isolate sperm or sperm head.
2. For sperm isolation, transfer the filtered flow through into 8 mL of sperm isolation buffer and incubate it on ice for 40 min. Centrifuge the mixture at 600*g* for 5 min at 4 °C and then discard the supernatant.
3. For sperm head isolation, centrifuge the flow through at 3,000*g* for 5 min at 25 °C and discard the supernatant. Add 2 mL of sperm head isolation buffer and incubate at room temperature for 15 min. Follow this with another centrifugation at 3,000*g* for 5 min at 25 °C and then discard the final supernatant.
4. For both sperm and sperm head pellets, wash the pellet by resuspending in 5 mL of PBS, then centrifuge at 600*g* for 5 min at 4 °C. Carefully discard the supernatant. Add at least 500 μL of TRIzol reagent to ensure the adequate extraction of RNA from the samples. Achieve complete lysis by repeatedly passing the suspension through a syringe fitted with a 27G needle until no visible precipitate remains.
5. Purify RNAs from the mixture following Steps 25–32 of the main procedure.

◆ **TROUBLESHOOTING**

Library preparation and controls

We recommend dividing the RNA sample, using half to construct traditional (no modification related enzyme treated) small RNA-seq libraries and the other half for PANDORA-seq library construction followed by sequencing of both libraries. Traditional small RNA-seq data can be used as controls for conducting parallel analyses to distinguish between modified and unmodified small RNA types. To achieve statistically significant results in pairwise comparisons, it is recommended to perform PANDORA-seq with at least three biological replicates. The library preparation requires basic expertise in molecular biology. Caution should be taken to prevent RNase contamination, which can lead to artificial RNA fragmentation during library construction. Incorporating synthetic small RNAs as spike-ins before Step 63 can serve as internal controls to standardize sncRNA expression levels. However, it should be noted that adding spike-in RNA into RNA samples with the same quantity of total RNAs will be problematic if different numbers of cells in the two groups contribute equal quantities of total RNAs[1]. Besides, users can also include additional customized controls based on specific needs and objectives.

Quality control and sequencing considerations

Successful libraries should yield at least 1 nM of purified cDNAs or meet the specific requirements of sequencing center (Steps 66–86). Typically, the peak length of libraries should range from ~130 to ~170 bp, with the average length of a PANDORA-seq library being marginally longer than that of traditional small RNA libraries. Small RNA samples generally require a minimum of 10 million reads; however, sequencing additional reads can improve the accuracy of quantification. During data analysis, an initial assessment can be made by checking the annotation rate of the total clean reads that pass quality control and adaptor removal, which usually exceeds 90%. This serves as a preliminary check for any sample contamination.

Materials

Biological materials

- *E. coli* BL21 (DE3) (New England BioLabs, cat. no. C2527H)
- pET28a-AlkB plasmid (Addgene, cat. no. 228218)
- Mice (The Jackson Laboratory, cat. no. C57BL/6J)
  ▲ **CAUTION** Any experiments involving mice must conform to relevant Institutional and National regulations. All animal experiments and housing conditions were approved by the Animal Care and Use Committee of the University of California, Riverside and Tongji University.

Reagents

- Pure ethanol (Koptec, cat. no. V1001)
  ▲ **CAUTION** Highly flammable liquid and vapor. Avoid open flames and use in a well-ventilated area. Use in a fume hood with gloves, goggles and protective clothing.
- RNaseZap RNase decontamination solution (Thermo Fisher Scientific, cat. no. AM9780)
- Sodium chloride (Sigma-Aldrich, cat. no. S9888)
- UltraPure distilled water (Invitrogen, cat. no. 10977-015)
- Sodium hydroxide (Sigma-Aldrich, cat. no. 415413)
  ▲ **CAUTION** Highly corrosive. Handle with extreme care, using gloves, safety goggles and a face shield. Use in a fume hood with gloves, goggles and protective clothing.
- Glycerol (Sigma-Aldrich, cat. no. G5516)
  ▲ **CAUTION** Highly flammable liquid. Avoid open flames and use in a well-ventilated area.
- Phosphate-buffered saline, 10× (PBS; Gibco, cat. no. 70013032)
- Imidazole (Sigma-Aldrich, cat. no. I202)
- HEPES buffer, 1.0 M solution, pH 8.0 (Alfa Aesar, cat. no. J63578)
- KCl, 3 M solution (Sigma-Aldrich, cat. no. 1048170250)

- Dithiothreitol (DTT; MedChemExpress, cat. no. HY-15917)
  ▲ **CAUTION** Harmful if inhaled or absorbed through the skin. Use in a fume hood with gloves, goggles and protective clothing.
- Tris–HCl, 1 M solution, pH 8.0 (Invitrogen, cat. no. 15568025)
- Tris–HCl, 1 M solution, pH 7.5 (Fisher BioReagents, cat. no. BP1757100)
- Ammonium persulfate (APS; Sigma-Aldrich, cat. no. A3678)
  ▲ **CAUTION** Can irritate the nose and throat. Avoid inhalation and direct contact.
- TBE buffer, 10× (Invitrogen, cat. no. AM9863)
- Ferrous ammonium sulfate (Sigma-Aldrich, cat. no. 215406)
- α-Ketoglutaric acid (Sigma-Aldrich, cat. no. K1128)
- Sodium ascorbate (Sigma-Aldrich, cat. no. PHR1279)
- Bovine serum albumin (BSA; Sigma-Aldrich, cat. no. A7906)
- UltraPure sodium dodecyl sulfate (SDS; Invitrogen, cat. no. 15525017)
  ▲ **CAUTION** Can cause severe skin and eye irritation. Avoid inhaling dust or fumes. Use in a well-ventilated area or fume hood. Wear gloves, goggles and protective clothing.
- Sodium acetate, 3 M, pH 5.5 (Invitrogen, cat. no. AM9740)
- Ethylenediaminetetraacetic acid, 0.5 M, pH 8.0 (EDTA; Thermo Fisher Scientific, cat. no. 15575020)
- Triton X-100 (Sigma-Aldrich, cat. no. X100)
- Proteinase K solution, 20 mg/mL (Thermo Fisher Scientific, cat. no. 25530049)
- Luria-Bertani broth (LB; Sigma-Aldrich, cat. no. L3522)
- LB broth with agar (Sigma-Aldrich, cat. no. L3147)
- Kanamycin (Sigma-Aldrich, cat. no. K0129)
- Yeast extract (Sigma-Aldrich, cat. no. Y1625)
- Tryptone (Millipore, cat. no. T9410)
- Yeast extract tryptone medium powder (YT; Sigma-Aldrich, cat. no. Y2377)
- Isopropyl β-D-1-thiogalactopyranoside (IPTG; MedChemExpress, cat. no. HY-15921)
- DNase I (New England Biolabs, cat. no. M0303)
- Nickel-nitrilotriacetic acid resin (Thermo Fisher Scientific, cat. no. 88221)
- Pre-stained protein marker (Sangon Biotech, cat. no. C610013)
- Coomassie brilliant blue (Sigma-Aldrich, cat. no. 1.15444)
- RNase inhibitor, murine (New England Biolabs, cat. no. M0314)
- TRIzol reagent (Invitrogen, cat. no. 15596026)
  ▲ **CAUTION** Toxic and corrosive. Can cause severe burns and is harmful if inhaled or absorbed through the skin. Use in a fume hood with gloves, goggles and protective clothing.
- Chloroform (Alfa Aesar, cat. no. J67241)
  ▲ **CAUTION** Toxic if inhaled or absorbed through the skin. Use in a fume hood with appropriate protective equipment, including gloves and a respirator if necessary.
- Isopropanol (Fisher BioReagents, cat. no. 17150576)
  ▲ **CAUTION** Highly flammable. Can cause skin and eye irritation. Use in a well-ventilated area away from open flames. Wear gloves, goggles and protective clothing.
- Linear acrylamide (5 mg/mL) (Invitrogen, cat. no. AM9520)
- Ammonium acetate (NH_4OAc; Sigma-Aldrich, cat. no. A1542)
- Nuclease P1 (New England Biolabs, cat. no. M0660S)
- Phosphodiesterase I (Affymetrix/USB, cat. no. J20240EXR)
- Alkaline phosphatase (Sigma-Aldrich, cat. no. P5521)
- Benzonase nuclease (Sigma-Aldrich, cat. no. E8263)
- Formic acid (Thermo Scientific, cat. no. 28905)
- Acetonitrile (Honeywell, cat. no. A9561)
- Urea (Invitrogen, cat. no. AM9902)
- 40% (w/v) Acrylamide–Bis solution, 19:1 (Ambion, cat. no. AM9022)
  ▲ **CAUTION** Toxic. Can cause skin and eye irritation and is harmful if inhaled. Use in a fume hood with gloves, goggles and protective clothing.
- 30% Acrylamide–Bis solution, 37.5:1 (Bio-red, cat. no. 1610158)
  ▲ **CAUTION** Toxic. Can cause skin and eye irritation and is harmful if inhaled. Use in a fume hood with gloves, goggles and protective clothing.

- *N*,*N*,*N*′,*N*′-Tetramethylethylenediamine (TEMED; Sigma, cat. no. T22500)
  ▲ **CAUTION** Highly flammable and harmful if inhaled. Can cause irritation to skin and respiratory tract. Use in a fume hood with gloves, goggles and protective clothing.
- RNA loading dye, 2× (New England Biolabs, cat. no. B0363S)
  ▲ **CAUTION** May cause skin and eye irritation. Avoid inhalation and direct contact. Use gloves, goggles and protective clothing.
- SYBR Gold Nucleic Acid Gel Stain (Thermo Fisher Scientific, cat. no. S11494)
  ▲ **CAUTION** May cause skin and eye irritation. Avoid inhalation and direct contact. Use gloves, goggles and protective clothing.
- Small RNA ladder (Beyotime, cat. no. R0207)
- Low range ssRNA ladder (New England Biolabs, cat. no. N0364S)
- Phenol:chloroform:isoamyl alcohol (25:24:1, v/v) (Thermo Fisher Scientific, cat. no. 15593031)
  ▲ **CAUTION** Toxic and corrosive. Use in a fume hood with gloves, goggles and protective clothing.
- T4 polynucleotide kinase (New England Biolabs, cat. no. M0201)
- ATP solution, 10 mM (Invitrogen, cat. no. AM8110G)
- Qubit miRNA assay kit (Invitrogen, cat. no. Q32880)
- Qubit RNA high sensitivity kit (Invitrogen, cat. no. Q32855)
- Qubit RNA extended range kit (Invitrogen, cat. no. Q33224)
- Agilent RNA 6000 nano kit (Agilent, cat. no. 5067-1511)
- Agilent small RNA kit (Agilent, cat. no. 5067-1548)
- Agilent DNA 1000 Kit (Agilent, cat. no. 5067-1504)
- Agilent high sensitivity DNA Kit (Agilent, cat. no. 5067-4626)
- NEBNext small RNA library prep set for Illumina (New England Biolabs, cat. no. E7560S)
- Monarch PCR and DNA cleanup kit (New England Biolabs, cat. no. T1030)
- Cytidine (C) (Sigma-Aldrich, cat. no. C4654)
- Adenosine (A) (Sigma-Aldrich, cat. no. A9251)
- Guanosine (G) (Sigma-Aldrich, cat. no. G6752)
- m^1A (Sigma-Aldrich, cat. no. SMB00939)
- m^1G (MedChemExpress, cat. no. HY-113136)
- m^3C (MedChemExpress, cat. no. HY-111645)
- $m^2{}_2G$ (MedChemExpress, cat. no. HY-113137)

Equipment

Hardware

- Quintix analytical balances with internal adjustment, 0.1 mg (Sartorius, cat. no. QUINTIX124-1S)
- Pipettes (Eppendorf: 0.1–2.5 µL, cat. no. 3123000012; 0.5–10 µL, cat. no. 3123000020; 2–20 µL, cat. no. 3123000098; 10–100 µL, cat. no. 3123000047; 20–200 µL, cat. no. 3123000055; 100–1000 µL, cat. no. 3123000063; 0.5–5 mL, cat. no. 3123000071)
- 40-µm cell strainer, sterile (Corning, cat. no. 431750)
- Microbiological incubator (Thermo Fisher Scientific, cat. no. 51028130)
- Thermostatic water bath (Thermo Fisher Scientific, cat. no. TSGP2S)
- Shaking incubator (Thermo Fisher Scientific, cat. no. SHKE8000-8CE)
- Refrigerated microcentrifuge (Eppendorf, cat. nos. 5810R and 5425R)
- 1.5-mL microcentrifuge tubes (Corning, cat. no. MCT-150-C-S)
- PCR tubes (Axygen, cat. no. 11376044)
- Electrophoresis chamber (Bio-Rad, cat. no. 1658001FC)
- Affinity chromatography column (Beyotime, cat. no. FCL12)
- Concentration column (10 kDa molecular weight cut-off, PALL, cat. no. MAP010C38)
- Low temperature and ultra-high pressure nano material preparation and dispersion machine (JNBIO, cat. no. JN-Mini Pro)
- Nanosep centrifugal devices with 3K Omega membrane (Cytiva, cat. no. OD003C33)
- ACQUITY UPLC HSS T3 1.8 µm × 2.1 mm × 150 mm column (Waters, cat. no. 186003540)
- QTRAP 6500+ mass spectrometer (Sciex, cat. no. QTRAP6500+)
- ACQUITY UPLC I-Class system (Waters, cat. no. ACQUITY UPLC I-Class)

- 8-mm autosampler inserts (Thermo Scientific, cat. no. C4012-529)
- 9-mm glass screw thread vials (Fisher Scientific, cat. no. 033919)
- 9-mm screw caps (Thermo Scientific, cat. no. 6ASC9RT1F)
- 20-mL syringe (BD, cat. no. 302830)
- 1-mL syringe (BD, cat. no. 309628)
- 27G 12-mm needle (BD, cat. no. 305109)
- ChemiDoc imaging system (Bio-Rad, cat. no. 12003153)
- Roller mixer (Beyotime, cat. no. E0063)
- Thermal cycler (Bio-Rad, cat. no. T100)
- Cathivex-GV filter unit, 0.22-μm filter (Millipore, cat. no. SLGV0250S)
- 2100 bioanalyzer instrument (Agilent, cat. no. G2939A)
- Qubit 4 fluorometer (Invitrogen, cat. no. Q33238)
- Qubit assay tube (Invitrogen cat. no. Q32856)
- Any server, desktop workstation or laptop with a Linux operating system, with enough storge space and random-access memory depending onto accommodate the size of RNA deep sequencing dataset. The example dataset was analyzed on an Ubuntu (version 16.04) Linux platform with ≥8 or more processors and ≥16 GB of random-access memory

Software

- Operating system: Linux
- SPORTS1.1: https://github.com/junchaoshi/sports1.1

Data files

- Required input data: small RNA-seq data; reference database for small RNA annotation (Box 2)
- Example datasets: example mouse liver and mature sperm datasets for running this protocol can be downloaded from the Gene Expression Omnibus under the accession code GSE144666 (ref. 6). The mouse small RNA reference database can be downloaded from https://github.com/junchaoshi/sports1.1.

BOX 2

Software setup for small RNA-seq analysis

sncRNA reference database setup

Standard sncRNA reference database

The most common sncRNA reference databases are precompiled and can be downloaded via GitHub at https://github.com/junchaoshi/sports1.1/blob/master/precompiled_annotation_database.md.

Customized sncRNA reference database

The sncRNA reference database may also be customized. The detailed setup instructions are via GitHub at https://github.com/junchaoshi/sports1.1.

Computational software installation

The small RNA annotation software SPORTS1.1 is designed for use in UNIX or Linux environments. The software can be obtained at GitHub via the following command:

```
$ git clone https://github.com/junchaoshi/sports1.1.git
```

Alternatively, the package can be downloaded as a compressed file by the following command:

```
$ wget https://github.com/junchaoshi/SPORTS1.1/archive/master.zip
```

Additional required software

Install the following programs according to the manufacturer's instructions:

Perl 5 (version: 5.38.2): https://www.perl.org

R (version 4.1.2): https://www.r-project.org

Bowtie (version: 1.3.1): http://bowtie-bio.sourceforge.net/index.shtml

Cutadapt (version: 5.0): http://cutadapt.readthedocs.io/en/stable/index.html

▲ **CRITICAL** The SPORTS1.1 software directory should be added to the 'PATH' environment variable for ease of access:

```
$ echo 'export PATH=$PATH:{your_path_to_SPORTS1.1-master}/source' >> ~/.bashrc
$ chmod 755 {your_path_to_SPORTS1.1-master}/source/sports.pl
```

Alternatively, SPORTS1.1 can be executed directly from its full path:

```
$ perl {your_path_to_SPORTS1.1-master}/source/sports.pl
```

Reagent setup

▲ **CRITICAL** Reagents flagged as RNase-free should be prepared under RNase-free conditions. Clean gloves and the working area with 75% (vol/vol) ethanol and nuclease decontamination solution (for example, RNaseZap). Use RNase-free water as indicated.

YT medium
Dissolve 31 g of 2× YT powder in 1,000 mL ultrapure water. Sterilize by autoclaving for 20 min, then cool the solution to room temperature (25 °C) before use. Store at 4 °C for up to 1 month.

LB medium
Dissolve 1.1 g of LB broth in 48.3 mL of ultrapure water. Sterilize by autoclaving for 20 min, then cool the solution to room temperature before use. Store at 4 °C for up to 1 month.

LB medium (containing agar)
Dissolve 1.68 g of LB broth with agar in 48.3 mL of ultrapure water. Sterilize by autoclaving for 20 min. Allow to cool slightly before adding Kanamycin to achieve a final concentration of 50 µg/mL. Pour into Petri dishes and wait for the agar to solidify. Store at 4 °C for up to 1 month.

50% (vol/vol) glycerol solution
Prepare 200 mL of 50% glycerol solution by adding ultrapure water to 100 mL of glycerol to achieve a total volume of 200 mL. Sterilize by autoclaving for 20 min, then cool the solution to room temperature before use. Prepare fresh before use.

Lysis buffer
Prepare 1 L of lysis buffer by combining 100 mL of 10× PBS, 30 mL of 5 M NaCl and 400 µL of 5 M imidazole solution. Top up with ultrapure water to 1 L total volume. Prepare fresh before use.

Wash buffer
Prepare 1 L of lysis buffer by combining 100 mL of 10× PBS, 30 mL of 5 M NaCl and 1 mL of 5 M imidazole solution. Top up with ultrapure water to 1 L total volume. Prepare fresh before use.

Elution buffer
Prepare 1 L of elution buffer by combining 20 mL of 1 M HEPES buffer (pH 8.0), 33.3 mL of 3M KCl solution, 40 mL of 5 M imidazole solution, 1 mL of 1 M DTT solution and 200 mL of 50% (vol/vol) glycerol solution. Top up with ultrapure water to 1 L total volume. Prepare fresh before use.

Storage buffer
Prepare 50 mL of storage buffer by combining 1 mL of 1 M Tris–HCl (pH 8.0), 25 mL of pure glycerol, 2 mL of 5 M NaCl solution and 100 µL of 1 M DTT solution. Top up with RNase-free water to 50 mL total volume. Store at 4 °C for up to 1 month.
▲ **CRITICAL** RNase free.

50 mM NH_4OAc (pH 5.3)
Dissolve 0.03854 g of NH_4OAc in 5 mL of RNase-free water. Adjust the pH to 5.3 with ammonia or acetic acid. Adjusting the volume to 10 mL with RNase-free water, and store at −20 °C for up to 6 months.
▲ **CRITICAL** RNase free.

10% APS (wt/vol)
Dissolve 1 g of APS in 10 mL of RNase-free water to prepare a 10% (wt/vol) APS solution. The solution can be stored at 4 °C for at least 3 months.
▲ **CRITICAL** RNase free.

1× TBE
Dilute 10× concentrated TBE buffer tenfold with RNase-free water before use.
▲ **CRITICAL** RNase free.

1.5 mM ferrous ammonium sulfate solution
Dissolve 0.1176 g of ferrous ammonium sulfate hexahydrate in 10 mL of RNase-free water to prepare a 30 mM ferrous ammonium sulfate solution and store at −20 °C for up to 6 months. Dilute the solution 20-fold with RNase-free water before use.
▲ **CRITICAL** RNase free.

10 mM α-ketoglutaric acid solution
Dissolve 0.1461 g of α-ketoglutaric acid in 10 mL of RNase-free water to prepare a 100 mM α-ketoglutaric acid solution, and store at −20 °C for up to 6 months. Dilute the solution tenfold with RNase-free water before use.
▲ **CRITICAL** RNase free.

20 mM sodium ascorbate solution
Dissolve 0.3962 g of sodium ascorbate in 10 mL of RNase-free water to prepare a 200 mM sodium ascorbate solution, and store at −20 °C for up to 6 months. Dilute the solution tenfold with RNase-free water before use.
▲ **CRITICAL** RNase free.

0.5 mg/mL BSA solution
Dissolve 0.0500 g of BSA in 10 mL of RNase-free water to prepare a 5 mg/mL BSA solution, and store at −20 °C for up to 6 months. Dilute the solution tenfold with RNase-free water before use.
▲ **CRITICAL** RNase free.

10% (wt/vol) SDS solution
Prepare 40 mL of a 10% (wt/vol) SDS stock solution by dissolving 5 g of SDS in RNase-free water. Adjust the final volume to 50 mL with RNase-free water. Sterilize the solution by filtering through a 0.22-μm filter. Store at room temperature for up to 12 months.
▲ **CRITICAL** RNase free.

Urea–PAGE gel elution buffer
Prepare 50 mL of gel elution buffer by combining 1 mL of a 1 M Tris–HCl (pH 7.5), 4.2 mL of 3 M sodium acetate, 100 μL of 0.5 M EDTA and 1.25 mL of 10% (wt/vol) SDS solution. Adjust the final volume to 50 mL with RNase-free water. Store indefinitely at room temperature. Warm the buffer to 37 °C before use if SDS precipitates.
▲ **CRITICAL** RNase free.

Mixed small RNA ladder
Prepare 4 μL of mixed small RNA ladder by combining 1 μL of small RNA ladder (Beyotime), 1 μL of low range ssRNA ladder (New England Biolabs) and 2 μL 2× RNA loading dye. Store at −20 °C for up to 12 months.
▲ **CRITICAL** RNase free.

5 M imidazole, pH 8.0
Dissolve 17.02 g of imidazole in 40 mL of RNase-free water. Adjust the pH to 8.0 with NaOH. After adjusting the volume to 50 mL with RNase-free water, sterilize through a 0.22-μm filter. Store at room temperature for up to 12 months.
▲ **CRITICAL** RNase-free.

1 M DTT solution
Dissolve 1.543 g of DTT in 8 mL of RNase-free water, top up to 10 mL with additional RNase-free water. The solution can be divided into aliquots of 1 mL and stored at −20 °C for up to 12 months.
▲ **CRITICAL** RNase free.

Sperm isolation buffer
Prepare 10 mL of sperm isolation buffer by combining 150 µL of a 10% (wt/vol) SDS solution and 750 µL of Triton X-100 solution. Top up with ultrapure water to 10 mL total volume, and store at 4°C for up to 1 month.

Sperm-head isolation buffer
Prepare 10 mL of sperm-head isolation buffer by combining 100 µL of 1 M Tris–HCl (pH 8.0), 200 µL of 0.5 M EDTA, 500 µL of 1 M NaCl, 2 mL of 10% (wt/vol) SDS solution and 37.5 µL of proteinase K. Add ultrapure water to reach a total volume of 10 mL, and store at 4°C for up to 1 month.

Procedure

Expression and purification of AlkB protein

● **TIMING ~5 d**

1. Thaw 40 µL of *E. coli* BL21 (DE3) competent cells on ice for 5 min.
   ▲ **CRITICAL STEP** Thaw the competent cells on ice and use them immediately to avoid freeze–thaw cycles.
2. Transfer 1 µL of 500 ng/µL pET28a-AlkB plasmid into 40 µL BL21 (DE3) competent cells. Ensure thorough mixing by gentle agitation. The sequence for the AlkB gene is detailed in Table 2.
3. Incubate the transformed cells on ice for 20 min, followed by a heat shock at 42 °C for 90 s, then immediately cool on ice for 2 min.
4. Incubate the transformed cells in 1 mL of LB medium at 37 °C with shaking at 220 rpm for 30 min. Centrifuge the cells at 1,000*g* for 3 min at 25 °C, discard the medium, then resuspend the cell pellet in 100 µL of LB medium and spread on an LB agar plate containing 50 µg/mL kanamycin.
5. Incubate the plate at 37 °C for 14–16 h or until colonies have grown to the desired size. Isolate a single colony and inoculate it into 3 mL of YT medium supplemented with 50 µg/mL kanamycin. Incubate the culture at 37 °C with orbital shaking at 220 rpm for 12 h.
6. Dilute an appropriate volume of culture at a 1:20 ratio into 3 mL of YT medium supplemented with 50 µg/mL kanamycin. Continue incubation at 37 °C with shaking at 220 rpm. Monitor the absorbance until the optical density at 600 nm (OD_{600}) reaches 0.5–0.6. Either skip to Step 9 or preserve cells for long-term storage at −80°C by creating a glycerol stock.
   ■ **PAUSE POINT** Preserve cells containing the *AlkB* plasmid for long-term storage by combining the cell culture with an equal volume of 50% (vol/vol) glycerol. Store the glycerol stock at −80 °C for up to 1 year, minimizing freeze–thaw cycles.
7. Scrape some of the frozen glycerol stock of cells transformed with the pET28a-AlkB plasmid using a 10 µL pipette tip.
   ▲ **CRITICAL STEP** Do not thaw the glycerol stock.
8. Incubate the cells in 500 mL YT medium containing 50 µg/mL kanamycin at 37 °C with shaking at 220 rpm and monitor the culture's absorbance until OD_{600} reaches 0.5–0.6.
9. Add IPTG to achieve a final concentration of 0.5 mM in the culture and then reduce the temperature to 16 °C. Incubate the cells overnight with shaking at 220 rpm.
10. Centrifuge at 6,000*g* at 4°C for 5 min to collect the cells.
11. Discard the medium and resuspend the cell pellet in 20 mL precooled lysis buffer.
12. Homogenize the cells on ice using a 20 mL syringe, ensuring complete resuspension to prevent clogging in the cell disruptor. (Optional) Add DNase I to a final concentration of 10 U/mL and incubate at 25 °C for 15 min to avoid cell clogging.
    ▲ **CRITICAL STEP** Keep the cell lysate on ice to maintain the enzymatic activity of AlkB protein.
13. Configure the cell disruptor to operate at a temperature of 4 °C and a pressure setting of 1,200 bar. Sequentially rinse the machine interior with 75% (vol/vol) ethanol, ultrapure water and precooled lysis buffer. Pour the cell lysate into the inlet, then collect it at

the outlet. Recycle the cell lysate through the disruptor six to eight times, collecting the lysate on ice throughout the process.

14. Centrifuge the disrupted lysate at 13,000*g* at 4 °C for 40 min. Transfer the supernatant to a new 50 mL centrifuge tube.
15. Add 1 mL nickel-nitrilotriacetic acid resin to a fresh tube. Wash the resin by adding 5 mL of precooled lysis buffer. Centrifuge the mixture at 1,000*g* for 30 s at 4 °C, then discard the supernatant. Repeat this washing procedure two to three times.
16. Add 20 mL of cell lysate to the tube containing the resin and mix thoroughly by pipetting up and down several times. Incubate the mixture on a roller mixer at 4 °C for 2 h.
17. Prepare an affinity chromatography column by washing it with 2–3 column volumes of the precooled lysis buffer.
18. Transfer the mixture from Step 16 to the column. Allow the mixture to flow through the column by gravity at 4 °C.
19. Wash the column with 5–6 column volumes of precooled wash buffer.
20. Elute the bound proteins with precooled elution buffer by adding 1 mL of elution buffer and allowing each 1 mL addition to sit for 3–5 min before collecting the flow through. Repeat the elution process, taking 20 µL of elution from each flow-through-containing tube to stain with Coomassie brilliant blue. Visualize the stain to determine the protein level and continue eluting until no protein is detected.
21. Pool the flow-through fractions containing protein and transfer to a concentration column with a 10 kDa molecular weight cutoff. Centrifuge at 6,000*g* at 4°C for 10 min. Discard the flow through from the collection tube and add an equal volume of storage buffer to the filter membrane. Repeat the centrifugation and buffer replacement steps five to six times to completely substitute the elution buffer with storage buffer. Continue centrifugation at 6,000*g* at 4 °C until no additional fluid collects at the bottom. This process yields several milliliters of protein concentrate, with a protein content in the microgram range.
22. Subject the collected protein to 12% (wt/vol) SDS–polyacrylamide gel electrophoresis (PAGE) electrophoresis followed by Coomassie brilliant blue staining[56] to evaluate the purity and concentration of AlkB protein (Fig. 2). Dilute the protein concentration to 200 ng/µL with storage buffer.

◆ **TROUBLESHOOTING**

23. Aliquot the diluted AlkB protein solution as needed, flash freeze in liquid nitrogen and store at −80 °C for long-term storage. To prevent enzyme inactivation due to repeated freeze–thaw cycles, dilute the protein to a working concentration of 20 ng/µL using storage buffer and store at −20°C for up to 1 month.

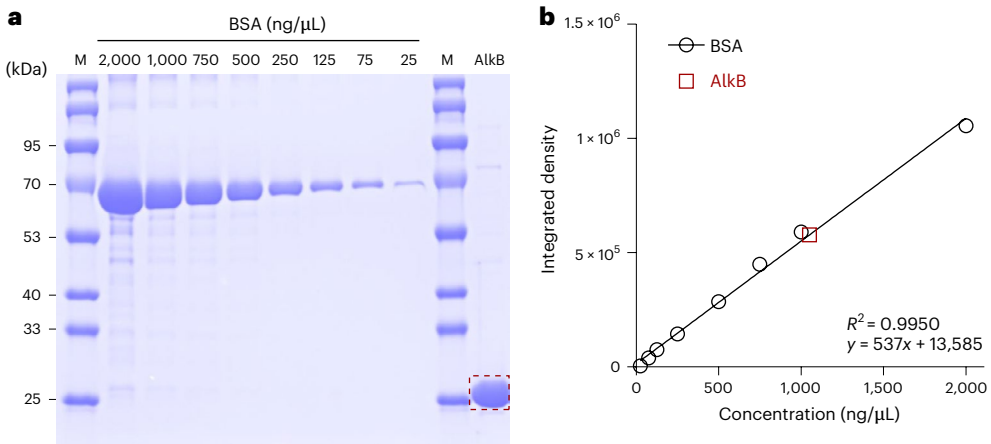


Fig. 2 | SDS–PAGE analysis and quantification of purified AlkB protein. a, SDS-PAGE analysis of AlkB protein (~25 kDa), highlighted by a red dashed square, for molecular weight and purity assessment alongside BSA standards, visualized with Coomassie brilliant blue staining. M, molecular weight markers. **b**, A densitometric analysis of the SDS–PAGE gel to determine the integrated density of the BSA bands and estimate the concentration of AlkB. The linear regression equation and R^2 value indicate the accuracy of the standard curve.

Demethylation activity validation of purified AlkB protein by LC–MS/MS

● **TIMING** 1 d

▲ **CRITICAL** Steps 25–42 should be performed under RNase-free conditions. Clean the gloves and working area with 75% (vol/vol) ethanol and nuclease decontamination solution. Use RNase-free water when indicated.

24. Isolate 50–100 mg of liver sample per 1.5 mL microcentrifuge tube from an euthanized C57BL/6J mouse.
25. Add 1 mL of TRIzol reagent to each tube and homogenize thoroughly. Let the mixture stand at room temperature for 5 min.
    ■ **PAUSE POINT** Samples can be stored at 4 °C overnight or at −20 °C for up to 1 year.
26. Add 0.2 mL of chloroform per 1 mL of TRIzol reagent used. Securely cap the tubes, shake vigorously for 15 s, then let stand at room temperature for 10–15 min.
27. Centrifuge the tubes at 13,000*g* for 15 min at 4 °C.
28. Carefully transfer the upper aqueous phase to a fresh tube. Add 0.5 mL isopropanol per 1 mL of TRIzol reagent used to the aqueous phase and then mix gently by pipetting up and down several times. Optionally, add 1 µL of linear acrylamide per tube to aid in RNA precipitation.
29. Let the mixture stand at 4 °C for 10 min and then centrifuge at 13,000*g* for 10 min at 4 °C.
30. Carefully remove and discard the supernatant. Add 1 mL of 75% (vol/vol) ethanol per 1 mL of TRIzol reagent used.
    ■ **PAUSE POINT** the RNA precipitate can be stored in 75% (vol/vol) ethanol at −80 °C for at least 1 year.
31. Briefly vortex the tube, then centrifuge at 7,500*g* for 5 min at 4 °C. Carefully discard the supernatant. Air dry the RNA pellet for 5–10 min.
32. Resuspend the RNA pellet in 100 µL of RNase-free water by pipetting up and down several times until the RNA pellet is completely dissolved.
    ■ **PAUSE POINT** The RNA solution can be used immediately or stored at −80 °C for up to 1 month.
33. Evaluate and quantify the quality and concentration of RNA using the Qubit fluorometer and the Bioanalyzer 2100 with the RNA nano 6000 kit according to the manufacturer's instructions. Evaluate the integrity of total RNA to identify potential degradation. The acceptable RIN may vary depending on the sample type.
34. Dilute 200 ng of mouse liver total RNA in 10 µL of RNase-free water in preparation for AlkB treatment.
35. Prepare the AlkB enzyme treatment reaction by mixing the specified components. Adjust the volumes for each reaction based on the guidelines provided, including an extra 10% volume to account for pipetting losses. Incubate the reaction mixture at 37 °C for 30 min.

Component	Volume	Final concentration
Input RNA (Step 34)	10 µL	4 ng/µL
1 M HEPES	2.5 µL	50 mM
1.5 mM ferrous ammonium sulfate	2.5 µL	75 µM
10 mM 10× α-ketoglutaric acid	5 µL	1 mM
20 mM sodium ascorbate	5 µL	2 mM
0.5 mg/mL BSA	5 µL	50 ng/µL
RNase inhibitor	2.5 µL	2 units/µL
AlkB (20 ng/µL, Step 23)	10 µL	4 ng/µL
H_2O	7.5 µL	
Total	50 µL	

36. Purify RNA from the mixture following Steps 25–31, then dissolve 200 ng of the purified AlkB-treated liver total RNA in 50 µL of 50 mM NH_4OAc.
37. Dissolve 200 ng of AlkB-untreated liver total RNA from Step 32 in 50 µL of 50 mM NH_4OAc. The AlkB-treated and untreated RNA mixtures are treated in parallel in Steps 38–50.

38. Add 0.2 U of nuclease P1 to the RNA mixture from Step 36 and 37 and incubate at 50 °C for 3 h.
39. Add 0.04 U of phosphodiesterase I to the AlkB-treated and untreated RNA mixture and incubate at 37 °C for 2 h.
40. Add 2 U of alkaline phosphatase to the RNA mixtures and incubate at 37 °C for 2 h. Alternatively, replace Steps 37–40 with a one-step digestion system to save time. Digest 200 ng of AlkB-treated and untreated RNA from mouse liver tissue in 50 µL reaction buffer containing 250 mM Tris–HCl (pH 8.0), 5 mM $MgCl_2$, 500 ng/mL BSA, 1 U benzonase nuclease, 0.2 IU alkaline phosphatase and 0.05 IU phosphodiesterase I at 37 °C for 3 h. However, a shorter digestion time may result in incomplete RNA digestion, which could cause undigested RNA to be removed during the liquid chromatography process, preventing the detection of nucleosides from these RNAs during mass spectrometry.
41. Transfer the mixtures into Nanosep centrifugal devices with a 3K Omega membrane and centrifuge at 10,000*g* at 4 °C for 20 min.
42. Transfer the flow-through containing mononucleosides to a fresh tube and use immediately for LC–MS/MS analysis or store at −80 °C.
    ■ **PAUSE POINT** The mononucleoside sample can be stored at −80 °C for at least 6 months.
43. Use an Acquity UPLC I-class chromatography system with an Acquity UPLC HSS T3 1.8 µm × 2.1 mm × 150 mm column at 40 °C for chromatographic separation of the mononucleoside sample. Set up the flow rate at 0.3 mL/min.
44. Prepare mobile phase A by adding formic acid to distilled water and mobile phase B by adding formic acid to acetonitrile, ensuring a final concentration of 0.1% (vol/vol) in each. Set the mobile phase gradient as follows (A + B = 100%): 0–6 min, 0% B; 6–7.65 min, 0%–1% B; 7.65–9.35 min, 1%–6% B; 9.35–10 min, 6% B; 10–11 min, 6%–10% B; 11–12 min, 10%–20% B; 12–14 min, 20%–50%; 14–17 min, 50%–75% B; 17–17.5 min, 75%–0% B; 17.5–30 min, 0% B.
45. Setup the Sciex QTRAP 6500+ mass spectrometer in a positive ion mode using multiple reaction monitoring scans. Set the parameters: Curtain Gas (CUR): 30.0; Collisionally Activated Dissociation Gas (CAD): 8.0; Ion Spray Voltage (IS): 5500.0; Temperature (TEM): 350.0; Gas 1 (Nebulizer Gas, GS1): 30.0; Gas 2 (Auxiliary Gas, GS2): 30.0.
46. Monitor the following mass transitions for nucleosides: *m/z* 244.1–112.1 for cytidine (C), *m/z* 268.1–136.1 for adenosine (A), *m/z* 284.1–152.1 for guanosine (G), *m/z* 282.1–150.1 for m^1A, *m/z* 298.1–166.1 for m^1G, *m/z* 258.1–126.1 for m^3C and *m/z* 312.1–180.1 for m^2_2G. Optimize the parameters of declustering potential and collision energy for each ion pair to enhance signal quality.
47. Transfer 40 µL of the mononucleoside sample from Step 42 and 40 µL of each of a set of three gradient-diluted nucleoside standard mixtures into separate 100 µL vial insert tubes, each placed inside a 2 mL screw-top vial and carefully cap the vials. The concentration gradient of each nucleoside standard is outlined in Extended Data Table 1.
48. Sequentially inject 10 µL of a set of gradient-diluted nucleoside standards and AlkB-treated and untreated mononucleoside samples into the LC–MS/MS system to quantify the abundance of A, C, G, m^1A, m^3C, m^1G and m^2_2G. Determine the appropriate standard gradient range for each nucleoside to accurately cover the range of sample concentrations.
49. Calculate the molal concentration of each nucleoside according to the standard curve established for the same batch of samples. Ensure each nucleoside's retention time is correctly identified and the mass spectrometry peak is correctly calculated.
50. Normalize the molal concentration of m^1A, m^3C and m^1G and m^2_2G relative to A, C and G, respectively (Fig. 3 and Supplementary Table 2).
    ◆ **TROUBLESHOOTING**

Isolation of specified-size RNA from total RNAs

● **TIMING 2 d**

▲ **CRITICAL** Steps 51–71 should be performed under RNase-free conditions. Clean the gloves and working area with 75% (vol/vol) ethanol and nuclease decontamination aolution. Use RNase-free water when indicated.

51. Purify total RNAs from tissue or cell samples by following Steps 25–33.

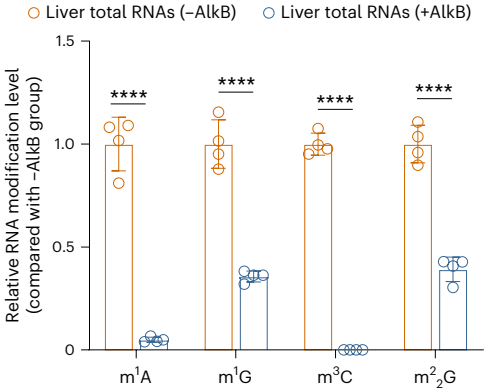


Fig. 3 | Demethylation activity of purified AlkB protein. Relative levels of RNA modifications (m^1A, m^1G, m^3C and m^2_2G) in total RNAs from mouse liver, with or without AlkB treatment, quantified by LC–MS/MS. The data are presented as RNA modification levels relative to the no AlkB treatment group, with $n = 4$ biologically independent samples per group, shown as mean ± standard error of the mean. The statistical significance was determined by two-sided multiple *t*-test. ****$P < 0.0001$.

52. Prepare a 15% (wt/vol) urea–PAGE gel by adding the following components in sequence: urea, acrylamide–Bis solution, 10× TBE, and RNase-free water, without APS or TEMED. This can be done at room temperature. Agitate or shake occasionally, until the urea is completely dissolved. Assemble electrophoresis chamber. Add APS and TEMED to the gel solution, mix thoroughly by vortexing and pour between the gel plates. Allow the gel to polymerize for at least 30 min. Add 1× TBE as running buffer. Prerun the gel for at least 30 min at 200 V (constant voltage).

Component	Volume/mass
40% (wt/vol) acrylamide–Bis solution, 19:1	3.75 mL
Urea	4.2g
10× TBE	1 mL
H_2O	Add to 10 mL in total
10% (wt/vol) APS	100 µL
TEMED	10 µL

53. Add RNA sample from Step 51 (preferably no more than 2 µg in 10 µL RNase-free water) to an equal volume of 2× RNA loading dye. Mix thoroughly by pipetting up and down several times. Heat the samples and mixed small RNA ladders at 75 °C for 5 min to denature them. Flush the urea out of the wells with running buffer using a pipette with a 1 mL tip and then load RNA samples and 4 µL of mixed small RNA ladders into each well. Run the gel for 40 min at 200 V (constant voltage) or until the blue dye reaches the bottom of the gel.
54. Dilute the SYBR Gold Nucleic Acid Gel Stain 10,000-fold with 1× TBE buffer to make a 1× staining solution. Remove the gel from the apparatus and incubate the gel in staining solution for 5 min.
55. Illuminate the stained gel on an ultraviolet transilluminator and excise the 15–50 nt band (Fig. 4). Place the gel slices into a 1.5 mL microcentrifuge tube with 500 µL of gel elution buffer.
56. Freeze the tube containing the gel elution buffer and gel slices on dry ice or −80 °C for 15 min.
57. Incubate the tube overnight at room temperature with continuous rotation. Ensure the tube is properly sealed to prevent contamination and for thorough mixing of the gel components.
58. Centrifuge the tube at 13,000*g* for 10 min at room temperature. Aspirate and save the supernatant in a clean microcentrifuge tube at 4°C.
59. Add 500 µL of phenol:chloroform:isoamyl alcohol to the supernatant and mix thoroughly by vortexing. Centrifuge at 12,000*g* for 10 min at 4 °C. Aspirate and save the supernatant in a clean microcentrifuge tube.
60. Add 500 µL of chloroform to the supernatant and mix thoroughly by vortexing. Centrifuge at 13,000*g* for 10 min at 4 °C. Remove the supernatant.

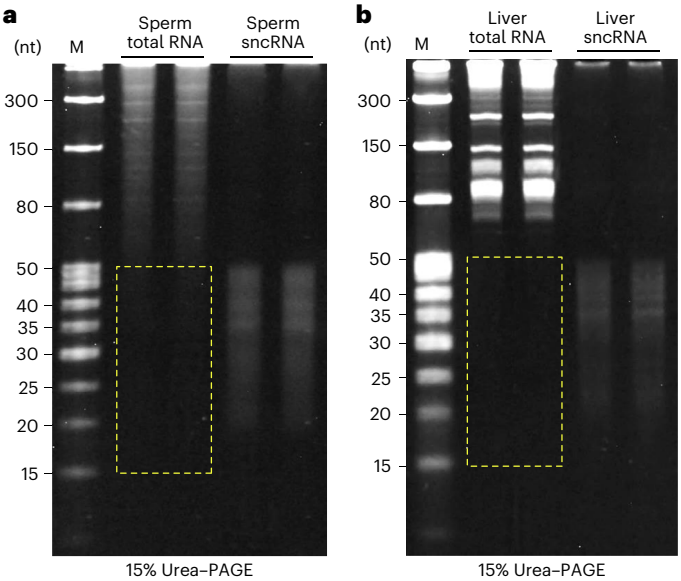


Fig. 4 | RNA size selection from mouse sperm and liver total RNA via PAGE gel electrophoresis. a, Mouse sperm total RNAs and PAGE-isolated 15–50 nt sncRNAs. **b**, Mouse liver total RNAs and PAGE-isolated 15–50 nt sncRNAs. M, RNA markers. The yellow dashed boxes highlight the area excised from the gel for further purification.

61. Wash the pellet in 1 mL of 75% (vol/vol) ethanol by pipetting up and down several times and centrifuge at 13,000*g* for 20 min at 4 °C. Decant the supernatant, removing as much as possible without disturbing the pellet. Dry the RNA pellet for ~10 min.
62. Resuspend the RNA pellet in 12 µL of RNase-free water by pipetting up and down several times until the RNA is completely dissolved. Quantify the concentration of small RNAs using either the Qubit miRNA assay kit or the Bioanalyzer 2100 with the small RNA kit, following the manufacturer's instructions.

PAUSE POINT The RNA should be used immediately or stored at −80 °C for up to 1 month.

◆ **TROUBLESHOOTING**

T4PNK enzyme treatment

● **TIMING ~4 h**

63. Prepare the T4PNK enzyme treatment by mixing the following components. The listed volumes are for one reaction; prepare enough for all reactions plus 10% additional volume. Incubate at 37 °C for 20 min.

Component	Volume	Final concentration
Input RNA (from Step 62)	10 µL	up to 20 ng/µL
T4PNK reaction buffer (10×)	5 µL	1×
ATP (10 mM)	5 µL	1 mM
T4PNK (10 units)	1 µL	0.2 units/µL
H_2O	29 µL	
Total	50 µL	

64. Purify RNA by following Steps 25–32.

AlkB enzyme treatment

● **TIMING ~4 h**

65. Obtain the purified RNAs from Step 64 and proceed with the AlkB treatment outlined in Step 35. Subsequently, purify the RNA according to Steps 25–32. During the AlkB treatment step, adjust the AlkB enzyme amount to match the input RNA amount at a ratio of 1:1, modifying the volume of water as necessary to maintain a total reaction volume of 50 µL.

Optionally, use half of the purified RNA from Step 64 to construct traditional (no-enzyme-treated) sncRNA libraries for comparison.

▲ **CRITICAL STEP** The order of T4PNK enzyme treatment and AlkB enzyme treatment can be interchanged.

Small RNA library construction and sequencing

● **TIMING ~5–7 d**

66. Dilute the 3′ single-read (SR) adaptor (5 μM stock concentration) to match the small RNA concentration proportionally. One possible method to calculate the dilution is as follows: Dilution rate = adaptor concentration (μM) × adaptor volume (μL)/(RNA concentration (ng/μL) × RNA volume (μL) × 10^3/(RNA average length (nt) × RNA average molecular weight (g/mol)). For example, if the adaptor concentration is 5 μM in a volume of 1 μL, the average RNA molecule length is 35 nt and the small RNA concentration is 5 ng/μL in a volume of 6 μL, the dilution rate of the adaptor can be calculated as follows: $5 \times 1/(5 \times 6 \times 10^3/(35 \times 320.5)) \approx 1.9$. This calculation suggests diluting 1 μL of the 3′ SR adaptor with 0.9 μL of RNase-free water. Next, mix the following components and incubate at 70 °C for 2 min.

Component	Volume
Input small RNA (from Step 65)	6 μL
Diluted 3′ SR adaptor	1 μL
Total	7 μL

67. Add the following components to the reaction from Step 66 and mix. Incubate at 16 °C for 18 h.

Component	Volume
3′ Ligation reaction buffer	10 μL
3′ Ligation enzyme	3 μL
Total	20 μL

68. Dilute the RT primer (5 μM stock concentration) to the same rate calculated in Step 66. Add the following components to the reaction from Step 67 and mix.

Component	Volume
Diluted RT primer	1 μL
H_2O	4.5 μL
Total	25.5 μL

69. Place the mixture in a thermal cycler and run the following program:

Temperature	Time
75 °C	5 min
37 °C	15 min
25 °C	15 min
4 °C	∞

70. Incubate the 5′ SR adaptor at 70 °C for 2 min, then immediately place the tube on ice. Use the denatured adaptor within 30 min. Dilute the 5′ SR adaptor (11.25 μM stock concentration) to the same rate calculated in Step 66. Add the following components one by one without premixing to the reaction from Step 69. Incubate at 25 °C for 1 h.

Component	Volume
Diluted 5′ SR adaptor	1 μL
5′ Ligation reaction buffer	1 μL
5′ Ligation enzyme	2.5 μL
Total	30 μL

71. Add the following components to the reaction from the previous step and mix. Incubate at 50 °C for 1 h.

Component	Volume
First strand synthesis reaction buffer	8 µL
RNase inhibitor	1 µL
ProtoScript II reverse transcriptase	1 µL
Total	40 µL

■ **PAUSE POINT** If you do not plan to proceed immediately to PCR amplification, then heat inactivate the RT reaction at 70 °C for 15 min. The samples can be safely stored at −15 to −25 °C for 24 h.

72. Add the following components to the RT reaction and mix.

Component	Volume
Taq DNA polymerase	50 µL
SR primer	2.5 µL
Index primer	2.5 µL
H_2O	5 µL
Total	100 µL

73. Perform PCR with the conditions outlined below.

Cycle step	Temp	Time	Cycles
Initial denaturation	94 °C	30 s	1
Denaturation	94 °C	15 s	Variable
Annealing	62 °C	30 s	
Extension	70 °C	15 s	
Final extension	70 °C	5 min	1
End	4 °C	∞	1

Amplification conditions may vary depending on the RNA input amount, the specific RNA species, and the characteristics of the RNA. Empirically, the number of PCR cycles can be estimated on the basis of the small RNA input (× ng) using the following formula: PCR cycle = round(22 − $\log_2(x)$). For example, if the small RNA input is 100 ng, the PCR cycle number can be calculated as round(22 − $\log_2(100)$) = 15. This number may be adjusted on the basis of the final double-stranded DNA (dsDNA) amount required in Step 86.

■ **PAUSE POINT** It is safe to store the library at −20 °C for 24 h after PCR. Avoid leaving the sample at 4°C overnight if possible.

74. Purify the PCR-amplified cDNA library using a Monarch PCR and DNA kit. Add 700 µL DNA cleanup binding buffer to the 100 µL sample. Mix well by pipetting up and down or by gently flicking the tube. Do not vortex. Insert the cleanup column into a 2 mL collection tube and load the mixture onto the column. Centrifuge at 16,000*g* for 1 min, then discard the flow through.
75. Reinsert the column into the same collection tube. Add 200 µL DNA wash buffer. Centrifuge at 16,000*g* for 1 min, then discard the flow through. Repeat this step.
76. Transfer the column to a clean 1.5 mL microcentrifuge tube.

▲ **CRITICAL STEP** Ensure that the column remains dry and does not come into contact with the flow through.

77. Add 27.5 µL water to the center of the column matrix. Wait for 1 min, then centrifuge at 16,000*g* for 1 min.

■ **PAUSE POINT** It is safe to store the library at −20 °C for 1 week.

78. Load 1 µL of the purified PCR products on the Bioanalyzer using the DNA 1000 kit according to the manufacturer's instructions.

◆ **TROUBLESHOOTING**

79. Prepare a 6% (wt/vol) PAGE ten-well gel by adding the following components in the specified order: acrylamide–Bis solution, 10× TBE and water, without APS or TEMED. This can be done at room temperature. Assemble the electrophoresis chamber. Add APS and TEMED to the gel solution, mix thoroughly by vortexing and pour between the gel plates. Allow the gel to polymerize for at least 30 min. Add 1× TBE as the running buffer.

Component	Volume/mass
30% (wt/vol) acrylamide–Bis solution, 37.5:1	2 mL
10× TBE	1 mL
H_2O	7 mL
10% (wt/vol) APS	100 µL
TEMED	10 µL

80. Add 25 µL cDNA library sample from Step 77 with 5 µL gel loading dye (6×). Mix well by pipetting up and down. Load the samples into two wells and add 5 µL of DNA markers into a separate well. Run the gel for 60 min at 120 V (constant voltage) or until the blue dye reaches the bottom of the gel.
81. Dilute the SYBR Gold Nucleic Acid Gel Stain 10,000-fold with 1× TBE buffer to make a 1× staining solution. Remove the gel from the apparatus and incubate the gel in staining solution for 5 min.
82. Illuminate the stained gel on an ultraviolet transilluminator and excise the band (~130–170 bp). (Fig. 5a).
83. Place the excised gel slice in a 1.5 mL tube. Crush the gel slice with a disposable pellet pestle and then soak the pieces in 250 µL DNA gel elution buffer (1×). Rotate end-to-end for at least 2 h at room temperature. Transfer the eluate and the gel debris to the top of a gel filtration column. Centrifuge the filter for 2 min at 17,000*g* at room temperature.
84. Recover eluate and add 1 µL linear acrylamide, 25 µL 3 M sodium acetate, pH 5.5 and 750 µL of 100% ethanol. Vortex well. Precipitate in a dry ice/methanol bath or at −80 °C for at least 30 min. Spin in a microcentrifuge at 14,000*g* for 30 min at 4 °C.
85. Remove and discard the supernatant, taking care not to disturb the pellet. Wash the pellet with 75% (vol/vol) ethanol by pipetting up and down several times. Spin in a microcentrifuge at 14,000*g* for 30 min at 4 °C. Air dry pellet for up to 10 min at room temperature to remove residual ethanol. Resuspend pellet in 12 µL TE buffer or EB buffer based on the sequencing center requirement.
86. Load 1 µL of the purified library on a 2100 Bioanalyzer using the DNA 1000 or high sensitivity DNA kit according to the manufacturer's instructions (Fig. 5b). Check the size, purity and concentration of the sample. The final concentration should yield at least 1 nM of purified dsDNAs or meet the specific requirements of the sequencing center.
    ◆ **TROUBLESHOOTING**
    ■ **PAUSE POINT** It is safe to store the library at −20°C for 1 week.
87. Pool the dsDNA samples in the desired ratio and then sequence the libraries using an Illumina high-throughput sequencing platform. The sequencing turnaround time is ~1–3 d, but it may vary depending on the type of sequencing instrument used. Each sample typically requires 10 million sequencing reads.

Small RNA-seq data analysis

● **TIMING ~1 d**

▲ **CRITICAL** Replace the parameters within the curly brackets in the remaining steps of the procedure with customized input files and computation resources.

88. Setup the sncRNA annotation software SPORTS1.1 and prepare the reference species annotation database (Box 2).
89. Use SPORTS1.1 (version 1.1.2) with customized settings to trim adaptor and low-quality bases from the raw sequencing data and annotate sequences against the reference database. The following example is provided for mouse sncRNA-seq data. If adaptors

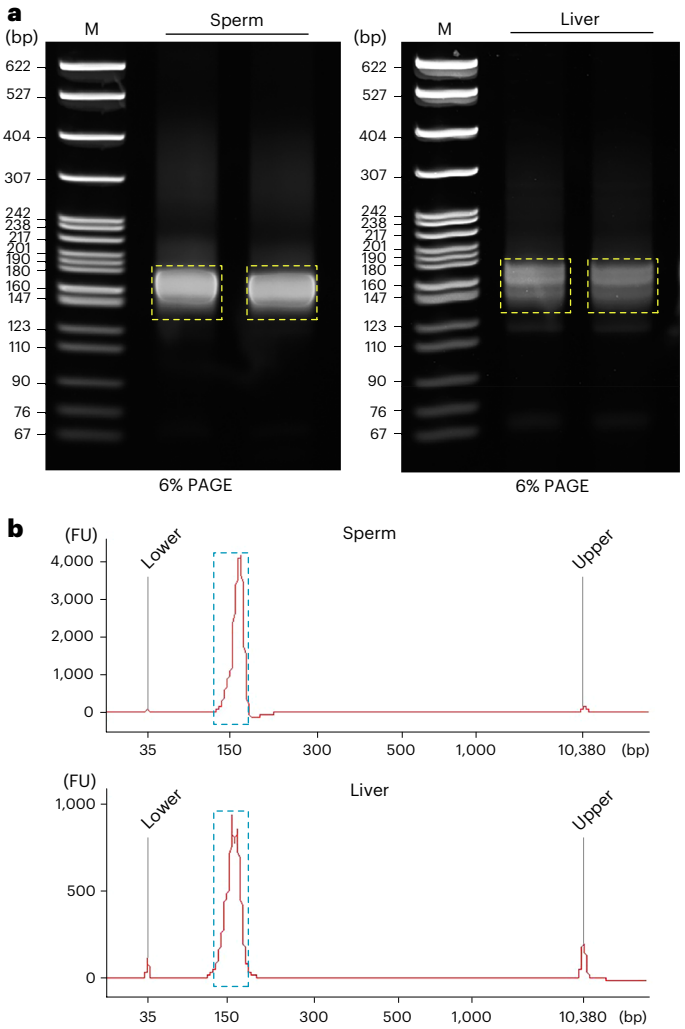


Fig. 5 | Evaluation of mouse sperm and liver PANDORA-seq DNA libraries. **a**, PAGE images of mouse sperm and liver PANDORA-seq DNA libraries before DNA size selection. M, DNA markers. The yellow dashed boxes highlight the area excised from the gel for further purification. **b**, A Bioanalyzer 2100 electropherogram analysis of sperm and liver DNA libraries after size selection. The size of DNA libraries is highlighted with blue dashed boxes. The gray solid lines represent the lower marker and upper marker, respectively. FU, fluorescence units.

have already been removed (as in the provided example data), exclude the parameters *-a -x* GTTCAGAGTTCTACAGTCCGACGATC *-y* AGATCGGAAGAGCACACGTCT. If adaptors are still present in the raw data, use the parameter *-a* to trim adaptors, and optionally use *-x* and *-y* to specify the 5′ and 3′ adaptor sequences, respectively. The default minimum and maximum lengths of the output sequences are set to 15 and 45, respectively, but can be adjusted to align with the user preferences.

```
$ sports.pl -i {input_folder} -p {threads} -M 1 \
-a -x GTTCAGAGTTCTACAGTCCGACGATC -y AGATCGGAAGAGCACACGTCT \
-g {reference_database_folder}/Mus_musculus/genome/mm10/genome \
-m {reference_database_folder}/Mus_musculus/miRBase/21/miRbase_21-mmu \
-r {reference_database_folder}/Mus_musculus/rRNAdb/mouse_rRNA \
-t {reference_database_folder}/Mus_musculus/GtRNAdb/mm10/mm10-tRNAs \
-w {reference_database_folder}/Mus_musculus/piRBase/piR_mouse \
-e {reference_database_folder}/Mus_musculus/Ensembl/Mus_musculus.
GRCm38.ncrna \
-f {reference_database_folder}/Mus_musculus/Rfam/12.3/Rfam-12.3-mouse \
-o {output_folder} > {output.report}
```

Troubleshooting

Troubleshooting advice can be found in Table 3.

Table 3 | Troubleshooting table

Step	Problem	Possible reason	Solution
22	Low purity of AlkB protein	Recombinant protein was degraded	Add protease inhibitors to the lysis buffer
		Insufficient washing of protein during affinity chromatography column purification	Perform additional washes with several column volumes of wash buffer or increase the concentration of imidazole up to 20 mM in the wash buffer as appropriate
	Low concentration of AlkB protein	Low induction efficiency during recombinant AlkB expression	Adjust the IPTG concentration and induction temperature to increase protein expression
		Insufficient disruption of bacterial cells	Increase the pressure of the high-pressure cell disruptor or increase the number of disruptions in Step 13
50	Low enzyme activity of AlkB protein	The protein lost its enzyme activity	Maintain a low temperature throughout the protein preparation process; minimize freeze–thaw cycles during enzyme usage; regularly check enzyme activity
62	Low concentration of isolated small RNAs	The proportion of small RNAs in the particular tissue or cell type is low	Increase the total RNA input amount in Step 51
		Inadequate isolation of small RNAs by gel purification	Repeat Steps 51–61
78 and 86	Low cDNA concentration within the ~130–170 bp region	Insufficient small RNA input	Increase small RNA input amount in Step 65; slightly increase the PCR cycle number in Step 73
Box 1, Step 5	18S and 28S rRNA peaks exist in final RNA samples	Potential somatic cell contamination	Examine the cell purity by microscope during each purification step; slightly increase the isolation buffer incubation time

Timing

Steps 1–23, expression and purification of AlkB protein: ~5 d
Steps 24–50, demethylation activity validation of purified AlkB protein by LC–MS/MS: ~1–2 d
Steps 51–62, isolation of specified-size RNA from total RNAs: ~2 d
Steps 63–64, T4PNK enzyme treatment: ~4 h
Step 65, AlkB enzyme treatment: ~4 h
Steps 66–87, small RNA library construction and sequencing: ~5–7 d
Steps 88–89, small RNA-seq data analysis: ~10 min (10 million reads)
Box 1, sperm and sperm head sample preparation: ~1 h
Box 2, software setup for small RNA-seq analysis: ~1 h

Anticipated results

The PANDORA-seq experimental procedure can be monitored at several steps. At Step 33, the concentration and length of total RNAs must be measured using an Agilent 2100 Bioanalyzer with an RNA Nano 6000 kit to assess RNA integrity. Typically, a RIN of 8 or higher is indicative of high-quality RNA samples. However, in contrast to somatic tissues, high-quality mouse and human sperm and sperm head samples lack the 18S and 28S rRNA peaks, resulting in an unmeasurable or very low RIN value. Following the sncRNA isolation procedure at Step 62, the quantity of small RNAs should also be measured, either using a Qubit miRNA assay kit (range from 0.5 to 150 ng/μL) or the Agilent 2100 Bioanalyzer Small RNA kit (range from 0.05 to 2 ng/μL). The small RNA libraries should achieve a concentration of at least 1 nM or meet the

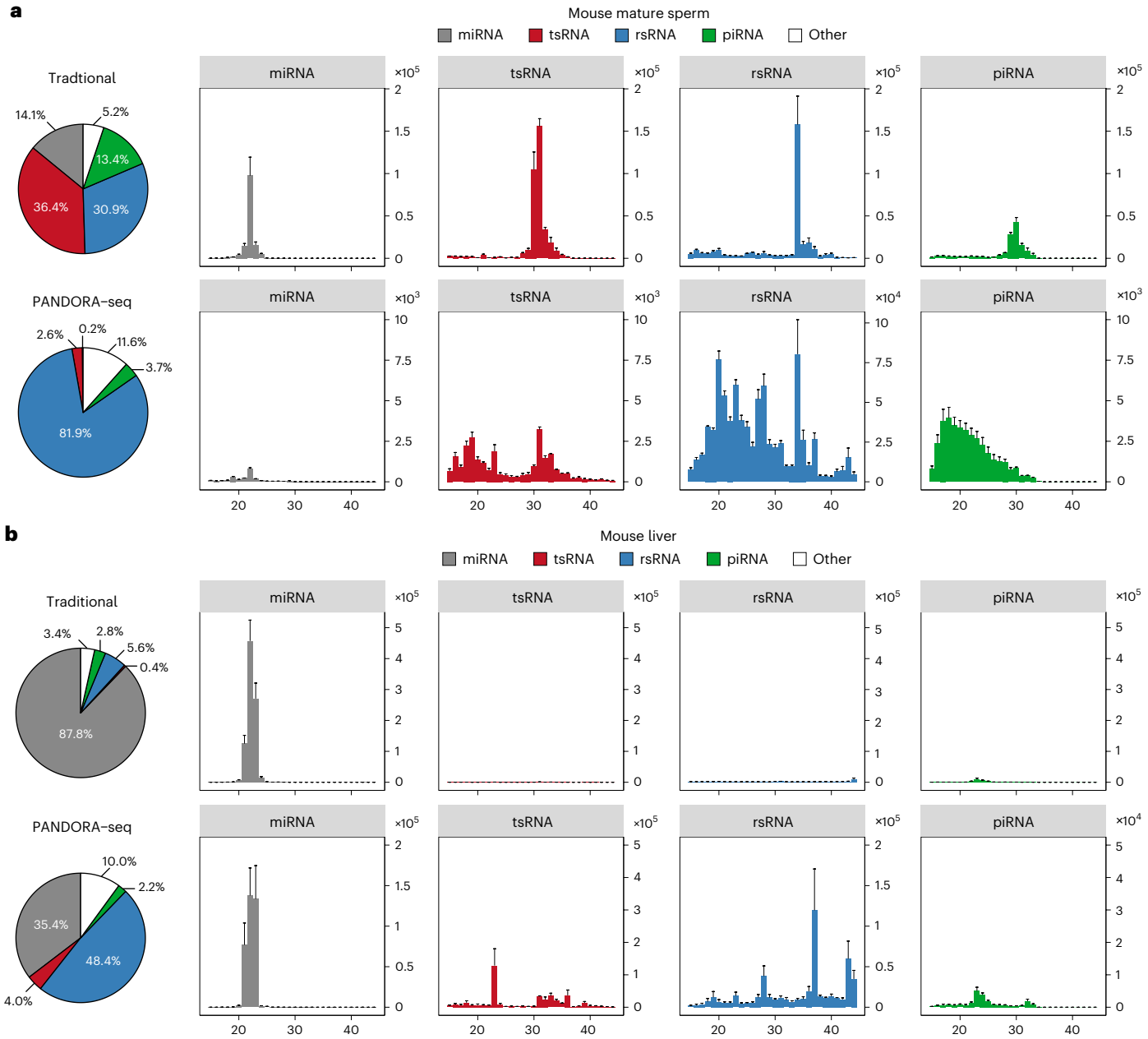


Fig. 6 | Comparison of the sequence length distribution of different types of sncRNA between PANDORA-seq and traditional small RNA-seq. a,**b**, The sequence length distribution of different types of sncRNA of mouse mature sperm (**a**) and mouse liver (**b**). The data are presented as means ± standard error of the mean. Adapted from ref. 6, Springer Nature Limited.

specific requirements of the sequencing center by Step 86. These libraries typically exhibit a peak within the range of ~130–170 bp.

Based on three biological replicates of mouse sperm and liver PANDORA-seq libraries, sequencing results in ~20 million reads on average. Following the SPORTS1.1 annotation process, 91.2% and 92.9% of the reads from sperm (Fig. 6a) and liver (Fig. 6b), respectively, were successfully annotated. For mouse sperm, 81.9% of reads mapped directly to the rRNA sequence, whereas 2.6% of reads were mapped to tRNAs. In contrast, 48.4% of reads from mouse liver mapped to the rRNA sequence, with 4.0% mapping to tRNAs.

Given the wide variation in RNA species across different tissues and cell types, a high sequencing depth is recommended to ensure adequate coverage for lowly expressed sncRNAs. We identified 22,951 types of sncRNA in mouse sperm and 23,995 types in mouse liver, with an expression level of reads per million ≥1.

Moreover, we sequenced a set of three biological replicates of traditional (no-enzyme-treated) small RNA-seq libraries for mouse sperm and liver. From this set, ~17 million reads were obtained. Following a data processing approach similar to that described above, 97.2% and 98.0% of reads from sperm (Fig. 6a) and liver (Fig. 6b), respectively, could be annotated. We identified 19,811 types of sncRNA in mouse sperm and 11,221 in mouse liver, with an expression level of reads per million ≥1. The sncRNAs were detected at significantly higher levels in PANDORA-seq libraries compared with traditional small RNA-seq libraries, suggesting the presence of RNA modifications in these sncRNAs.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The mouse liver and mature sperm small RNA-seq datasets can be accessed through the Gene Expression Omnibus under the accession code GSE144666 (ref. 6). Source data are provided with this paper.

Code availability

The sncRNA annotation pipeline SPORTS1.1 is available via GitHub at https://github.com/junchaoshi/sports1.1.

Received: 12 June 2024; Accepted: 11 February 2025;
Published online: 03 April 2025

References

1. Shi, J., Zhou, T. & Chen, Q. Exploring the expanding universe of small RNAs. *Nat. Cell Biol.* **24**, 415–423 (2022).
2. Chu, A. et al. Large-scale profiling of microRNAs for The Cancer Genome Atlas. *Nucleic Acids Res.* **44**, e3 (2016).
3. Giraldez, M. D. et al. Comprehensive multi-center assessment of small RNA-seq methods for quantitative miRNA profiling. *Nat. Biotechnol.* **36**, 746–757 (2018).
4. Costa, B. et al. Nicked tRNAs are stable reservoirs of tRNA halves in cells and biofluids. *Proc. Natl Acad. Sci. USA* **120**, e2216330120 (2023).
5. Chen, Q., Zhang, X., Shi, J., Yan, M. & Zhou, T. Origins and evolving functionalities of tRNA-derived small RNAs. *Trends Biochem. Sci.* **46**, 790–804 (2021).
6. Shi, J. et al. PANDORA-seq expands the repertoire of regulatory small RNAs by overcoming RNA modifications. *Nat. Cell Biol.* **23**, 424–436 (2021).
7. Cozen, A. E. et al. ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nat. Methods* **12**, 879–884 (2015).
8. Honda, S., Morichika, K. & Kirino, Y. Selective amplification and sequencing of cyclic phosphate-containing RNAs by the cP-RNA-seq method. *Nat. Protocol* **11**, 476–489 (2016).
9. Xu, H., Yao, J., Wu, D. C. & Lambowitz, A. M. Improved TGIRT-seq methods for comprehensive transcriptome profiling with decreased adapter dimer formation and bias correction. *Sci. Rep.* **9**, 7953 (2019).
10. Kugelberg, U., Natt, D., Skog, S., Kutter, C. & Ost, A. 5' XP sRNA-seq: efficient identification of transcripts with and without 5 phosphorylation reveals evolutionary conserved small RNA. *RNA Biol.* **18**, 1588–1599 (2020).
11. Hu, J. F. et al. Quantitative mapping of the cellular small RNA landscape with AQRNA-seq. *Nat. Biotechnol.* **39**, 978–988 (2021).
12. Wang, H. et al. CPA-seq reveals small ncRNAs with methylated nucleosides and diverse termini. *Cell Discov.* **7**, 25 (2021).
13. Upton, H. E. et al. 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 (2021).
14. Tosar, J. P., Castellano, M., Costa, B. & Cayota, A. Small RNA structural biochemistry in a post-sequencing era. *Nat. Protocol* **19**, 595–602 (2024).
15. Kugelberg, U., Natt, D., Skog, S., Kutter, C. & Ost, A. 5 XP sRNA-seq: efficient identification of transcripts with and without 5 phosphorylation reveals evolutionary conserved small RNA. *RNA Biol.* **18**, 1588–1599 (2021).
16. Hernandez-Alias, X. et al. Single-read tRNA-seq analysis reveals coordination of tRNA modification and aminoacylation and fragmentation. *Nucleic Acids Res.* **51**, e17 (2023).
17. Scacchetti, A. et al. A ligation-independent sequencing method reveals tRNA-derived RNAs with blocked 3′ termini. *Mol. Cell* **84**, 3843–3859.e8 (2024).
18. Skog, S., Orkenby, L., Kugelberg, U., Ost, A. & Natt, D. Seqpac: a framework for sRNA-seq analysis in R using sequence-based counts. *Bioinformatics* **39**, btad144 (2023).
19. Shi, J., Ko, E. A., Sanders, K. M., Chen, Q. & Zhou, T. SPORTS1.0: a tool for annotating and profiling non-coding RNAs optimized for rRNA- and tRNA-derived small RNAs. *Genom. Proteom. Bioinform.* **16**, 144–151 (2018).
20. Nunes, A. et al. Emerging roles of tRNAs in RNA virus infections. *Trends Biochem. Sci.* **45**, 794–805 (2020).
21. Wang, Q. et al. Identification and functional characterization of tRNA-derived RNA fragments (tRFs) in respiratory syncytial virus infection. *Mol. Ther.* **21**, 368–379 (2013).
22. Goodarzi, H. et al. Endogenous tRNA-derived fragments suppress breast cancer progression via YBX1 displacement. *Cell* **161**, 790–802 (2015).
23. Liu, X. et al. A pro-metastatic tRNA fragment drives Nucleolin oligomerization and stabilization of its bound metabolic mRNAs. *Mol. Cell* **82**, 2604–2617 e2608 (2022).
24. Honda, S. et al. Sex hormone-dependent tRNA halves enhance cell proliferation in breast and prostate cancers. *Proc. Natl Acad. Sci. USA* **112**, E3816–E3825 (2015).
25. Yue, T. et al. SLFN2 protection of tRNAs from stress-induced cleavage is essential for T cell-mediated immunity. *Science* **372**, eaba4220 (2021).
26. Krishna, S. et al. Dynamic expression of tRNA-derived small RNAs define cellular states. *EMBO Rep.* **20**, e47789 (2019).
27. Blanco, S. et al. Stem cell function and stress response are controlled by protein synthesis. *Nature* **534**, 335–340 (2016).

28. Guzzi, N. et al. Pseudouridylation of tRNA-derived fragments steers translational control in stem cells. *Cell* **173**, 1204–1216 e1226 (2018).
29. Chen, Q. et al. Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science* **351**, 397–400 (2016).
30. Sharma, U. et al. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science* **351**, 391–396 (2016).
31. Zhang, Y. et al. Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs. *Nat. Cell Biol.* **20**, 535–540 (2018).
32. Sarker, G. et al. Maternal overnutrition programs hedonic and metabolic phenotypes across generations through sperm tsRNAs. *Proc. Natl Acad. Sci. USA* **116**, 10547–10556 (2019).
33. Zhang, Y. et al. Angiogenin mediates paternal inflammation-induced metabolic disorders in offspring through sperm tsRNAs. *Nat. Commun.* **12**, 6673 (2021).
34. Alata Jimenez, N. et al. Paternal methotrexate exposure affects sperm small RNA content and causes craniofacial defects in the offspring. *Nat. Commun.* **14**, 1617 (2023).
35. Chen, Q., Yan, W. & Duan, E. Epigenetic inheritance of acquired traits through sperm RNAs and sperm RNA modifications. *Nat. Rev. Genet.* **17**, 733–743 (2016).
36. Tomar, A. et al. Epigenetic inheritance of diet-induced and sperm-borne mitochondrial RNAs. *Nature* **630**, 720–727 (2024).
37. Cai, C. & Chen, Q. Father's diet influences son's metabolic health through sperm RNA. *Nature* **630**, 571–573 (2024).
38. Diallo, I. et al. A tRNA-derived fragment present in *E. coli* OMVs regulates host cell gene expression and proliferation. *PLoS Pathog.* **18**, e1010827 (2022).
39. Ren, B., Wang, X., Duan, J. & Ma, J. Rhizobial tRNA-derived small RNAs are signal molecules regulating plant nodulation. *Science* **365**, 919–922 (2019).
40. Hogg, M. C. et al. Elevation in plasma tRNA fragments precede seizures in human epilepsy. *J. Clin. Invest.* **129**, 2946–2951 (2019).
41. Gu, W. et al. Peripheral blood non-canonical small non-coding RNAs as novel biomarkers in lung cancer. *Mol. Cancer* **19**, 159 (2020).
42. Paris, Z. & Alfonzo, J. D. How the intracellular partitioning of tRNA and tRNA modification enzymes affects mitochondrial function. *IUBMB Life* **70**, 1207–1213 (2018).
43. Balatti, V. et al. tsRNA signatures in cancer. *Proc. Natl Acad. Sci. USA* **114**, 8071–8076 (2017).
44. Zhan, S. et al. Serum mitochondrial tsRNA serves as a novel biomarker for hepatocarcinoma diagnosis. *Front. Med.* **16**, 216–226 (2022).
45. Zhang, X. et al. Small RNA modifications in Alzheimer's disease. *Neurobiol. Dis.* **145**, 105058 (2020).
46. Zhu, L. et al. Exosomal tRNA-derived small RNA as a promising biomarker for cancer diagnosis. *Mol. Cancer* **18**, 74 (2019).
47. Chen, X. et al. Human sperm tsRNA as potential biomarker and therapy target for male fertility. *Reproduction* **161**, 111–122 (2021).
48. Chen, Q. & Zhou, T. Emerging functional principles of tRNA-derived small RNAs and other regulatory small RNAs. *J. Biol. Chem.* **299**, 105225 (2023).
49. Kuhle, B., Chen, Q. & Schimmel, P. tRNA renovatio: rebirth through fragmentation. *Mol. Cell* **83**, 3953–3971 (2023).
50. Zheng, G. et al. Efficient and quantitative high-throughput tRNA sequencing. *Nat. Methods* **12**, 835–837 (2015).
51. Scheepbouwer, C. et al. ALL-tRNAseq enables robust tRNA profiling in tissue samples. *Genes Dev.* **37**, 243–257 (2023).
52. Dai, Q., Zheng, G., Schwartz, M. H., Clark, W. C. & Pan, T. Selective enzymatic demethylation of N^2,N^2-dimethylguanosine in RNA and its application in high-throughput tRNA sequencing. *Angew Chem. Int. Ed. Engl.* **56**, 5017–5020 (2017).
53. Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M. & Altman, D. G. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol.* **8**, e1000412 (2010).
54. Peng, H. et al. A novel class of tRNA-derived small RNAs extremely enriched in mature mouse sperm. *Cell Res.* **22**, 1609–1612 (2012).
55. Zhang, Y. et al. Identification and characterization of an ancient class of small RNAs enriched in serum associating with active infection. *J. Mol. Cell Biol.* **6**, 172–174 (2014).
56. Gallagher, S. & Chakavarti, D. Staining proteins in gels. *J. Vis. Exp.* **8**, 760 (2008).

Acknowledgements

The authors thank T. Zhou for discussions. This work was supported by National Key Research and Development Program of China (grant no. 2019YFA0802600 to Ying Zhang and Yunfang Zhang), Strategic Priority Research Program of the Chinese Academy of Sciences (grant no. XDA0460302 to J.S.), National Natural Science Foundation of China (92357306, 82122027 and 32171110 to Ying Zhang, 32370596 to J.S., 82371727 and 82022029 to Yunfang Zhang), Science and Technology Commission of Shanghai Municipality (23JC1403802 to Yunfang Zhang), Fundamental Research Funds for the Central Universities to Yunfang Zhang.

Author contributions

J.S. and Yunfang Zhang developed and optimized the experimental procedures. L.Z. and Y.L. collected mouse samples under the supervision of J.S. and Yunfang Zhang. Y.L. generated the AlkB enzyme and validated the enzyme activity with help from Yunfang Zhang and M.Y. M.Y. and Y.L. performed the LC–MS/MS RNA modification analyses with the help from L.Z. and Ying Zhang. J.S. developed the analysis pipeline and performed data analysis with the help of L.Z. Y.L., L.Z. and X.Z. contributed to the interpretation and discussion of data. J.S., Yunfang Zhang, Q.C. and Ying Zhang wrote the main manuscript and integrated input from all authors.

Competing interests

The authors declare no competing interests.

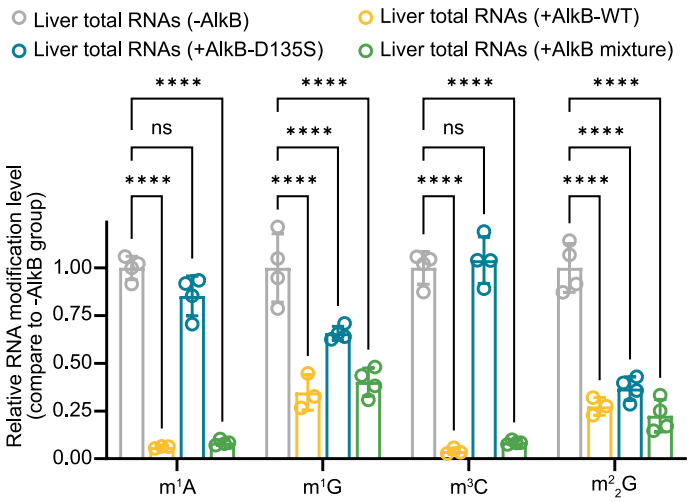
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41596-025-01158-4.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41596-025-01158-4.

Correspondence and requests for materials should be addressed to Junchao Shi, Yunfang Zhang or Ying Zhang.

Peer review information *Nature Protocols* thanks Zhangli Su and Juan Tosar for their contribution to the peer review of this work.



Extended Data Fig. 1 | Demethylation activity of AlkB-WT and AlkB-D135S mutant proteins. Relative RNA modification levels (m^1A, m^3C, m^1G, and m^2_2G) in total RNAs from mouse liver, following treatments with no AlkB enzyme (-AlkB), wild-type AlkB (AlkB-WT), the AlkB-D135S mutant, and a 1:1 combination of AlkB-WT and the AlkB-D135S mutant, as quantified using LC–MS/MS. Data are presented as RNA modification levels relative to the -AlkB group, with $n = 3$ or 4 biologically independent samples per group, shown as mean ± s.e.m. Statistical analysis was performed using Dunnett's multiple comparison test. ****$P < 0.0001$. n.s., not significant.

Extended Data Table 1 | Concentration gradients of nucleoside standards A, C, G, m^1A, m^3C, m^1G, and m^2_2G

Component	Standard#1 (ng/mL)	Standard#2 (ng/mL)	Standard#3 (ng/mL)	Standard#4 (ng/mL)	Standard#5 (ng/mL)	Standard#6 (ng/mL)	Standard#7 (ng/mL)	Standard#8 (ng/mL)	Standard#9 (ng/mL)
A	1000	500	250	100	40	20	10	5	2.5
C	1000	500	250	100	40	20	10	5	2.5
G	1000	500	250	100	40	20	10	5	2.5
m^1A	100	50	25	10	4	2	1	0.5	0.25
m^1G	100	50	25	10	4	2	1	0.5	0.25
m^2_2G	50	25	12.5	5	2	1	0.5	0.25	0.125
m^3C	50	25	12.5	5	2	1	0.5	0.25	0.125

nature portfolio

Corresponding author(s): Junchao Shi, Yunfang Zhang, Ying Zhang

Last updated by author(s): Dec 25, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The mouse liver and mature sperm datasets were obtained from the Gene Expression Omnibus with the accession code GSE144666.
Data analysis	Gray-scale analysis of dyed protein gel samples was performed using ImageJ (version 1.0), bundled with Zulu OpenJDK (version 13.0.6). Protein concentration standard curve was generated using GraphPad Prism (version 10.1.1). The differential analysis of relative RNA modification levels was conducted using GraphPad Prism (version 10.1.1). Small RNA sequences were annotated using the software SPORTS1.1(v1.1.2).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The mouse liver and mature sperm small RNA-seq datasets can be accessed through the Gene Expression Omnibus under the accession code GSE144666.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample-size calculation was performed.
Data exclusions	No data were excluded from the analyses.
Replication	The purification of the AlkB protein was not replicated, as all subsequent experiments were performed using the same batch of purified AlkB to ensure consistency in the experimental results. Four independent experiments were conducted for the LC-MS/MS analysis. Two biological replicates were included in the sncRNA isolation and dsDNA library size selection experiments. Three biological replicates for mouse mature sperm and liver were used for the small RNA-seq data analysis.
Randomization	The mice used for RNA sample collection from mature sperm and liver were randomly selected.
Blinding	RNA sample collection and RNA-seq data analysis were not blinded due to practical limitations.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	9-10 weeks old C57BL/6J mice were given access to food and water ad libitum and were maintained on a 12:12 h light-dark artificial

Laboratory animals	lighting cycle, mice were housed in cages at a temperature of 22-25 °C, with 40-60% humidity.
Wild animals	The study did not involve wild animals.
Reporting on sex	Male mice were used in this study.
Field-collected samples	The study did not involve samples collected from the field.
Ethics oversight	Animal experiments and housing conditions were approved by the Animal Care and Use Committee of the University of California, Riverside and Tongji University.

Note that full information on the approval of the study protocol must also be provided in the manuscript.