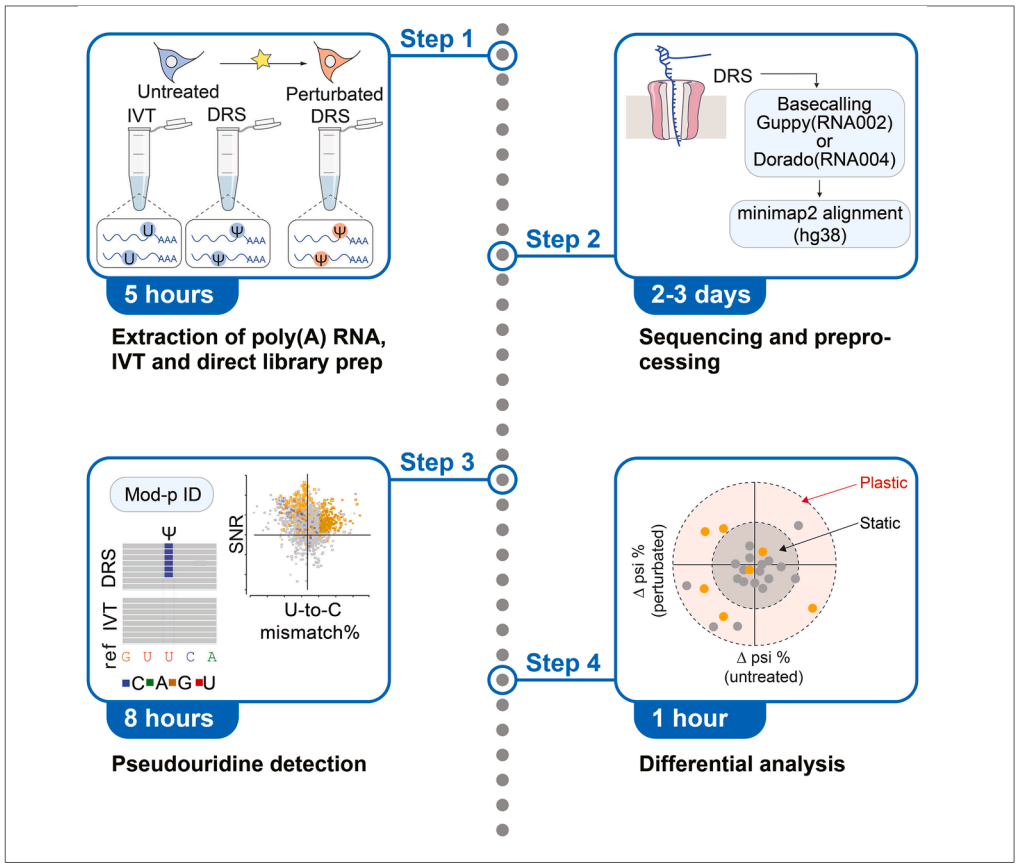
STAR Protocols

CellPress
OPEN ACCESS

Protocol

Protocol for differential analysis of pseudouridine modifications using nanopore DRS and unmodified transcriptome control



In this protocol, we detect pseudouridine using nanopore direct RNA sequencing (DRS 002/004) and analyze dynamic changes to mRNA modifications in response to perturbation. We cover RNA extraction, poly(A) selection, DRS, unmodified transcriptome (*in vitro* transcribed [IVT]) library preparation, and knockdowns of writer proteins as critical controls. We detail preprocessing, psi-site identification, and differential analysis, quantifying the direction and magnitude of changes across cellular types and states.

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

Oleksandra Fanari, Michele Meseonznik, Dylan Bloch, Sara H. Rouhanifard

s.rouhanifard@northeastern.edu

Highlights

Detect pseudouridine in human mRNA using nanopore direct RNA sequencing

Generate and sequence unmodified IVT RNA as a control to improve ψ-site identification

Process DRS and IVT data using the Mod-p ID computational pipeline

Perform differential ψ-site analysis across cellular states

Fanari et al., STAR Protocols 6, 103948
September 19, 2025
https://doi.org/10.1016/j.xpro.2025.103948

STAR Protocols

CellPress
OPEN ACCESS

Protocol

Protocol for differential analysis of pseudouridine modifications using nanopore DRS and unmodified transcriptome control

Oleksandra Fanari,[1,2] Michele Meseonznik,[1,2] Dylan Bloch,[1] and Sara H. Rouhanifard[1,3,4,*]

[1]Department of Bioengineering, Northeastern University, Boston, MA, USA

[2]These authors contributed equally

[3]Technical contact

[4]Lead contact

*Correspondence: s.rouhanifard@northeastern.edu
https://doi.org/10.1016/j.xpro.2025.103948

SUMMARY

In this protocol, we detect pseudouridine using nanopore direct RNA sequencing (DRS 002/004) and analyze dynamic changes to mRNA modifications in response to perturbation. We cover RNA extraction, poly(A) selection, DRS, unmodified transcriptome (*in vitro* transcribed [IVT]) library preparation, and knockdowns of writer proteins as critical controls. We detail preprocessing, psi-site identification, and differential analysis, quantifying the direction and magnitude of changes across cellular types and states.

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

BEFORE YOU BEGIN

This protocol requires various reagent kits and equipment for sample preparation, quality control, and sequencing. Ensure all necessary kits are available and devices, including an Oxford Nanopore Technologies (ONT) sequencing platform (i.e., MinION or PromethION sequencing devices), are correctly installed and tested. Sequencing runs generate large volumes of data, depending on the platform used for sequencing. It is essential to consider platform-specific IT requirements and involve trained experts for the sequencing and preprocessing steps, such as quality control, basecalling, and data analysis.

Considering the length of the protocol, if you plan to prepare an *in vitro* transcribed (IVT), unmodified library, we advise dividing the protocol into two days. We usually perform RNA isolation, poly(A)+ selection, and IVT preparation up to step 26d. We then hold at 4°C and resume the next day, completing the IVT control sample library preparation and your sample library preparation.

Alternatively, IVT libraries that are publicly available may be used (i.e., McCormick et al.[2]), as long as the specific chemistry (i.e., RNA 002) and the specific version of the basecaller match that of the new sequencing library.

> ***Note:*** When you are sequencing the unmodified IVT sample, you still need to follow the RNA sequencing protocol provided by ONT after the IVT steps: the IVT library preparation is a Direct sequencing of an in vitro transcribed RNA sample.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
PBS - phosphate-buffered saline (10×) pH 7.4	Invitrogen	AM9625
TRIzol	Invitrogen	15596026
Chloroform	Thermo Scientific Chemicals	AC423555000
T4 DNA ligase	New England Biolabs	M0202M
Maxima H Minus Reverse Transcriptase (200 U/μL)	Thermo Fisher Scientific	EP0751
SuperScript III Reverse Transcriptase	Thermo Fisher Scientific	18080093
Critical commercial assays		
PureLink RNA Mini Kit	Invitrogen	12183025
Poly(A) mRNA Magnetic Isolation Module	New England Biolabs	E7490L
Qubit RNA High Sensitivity (HS)	Invitrogen	Q32852
Direct RNA Sequencing Kit	Oxford Nanopore Technologies	SQK-RNA002
Agencourt RNAClean XP	Beckman Coulter	A63987
MinION and GridION Flow Cell (R9.4.1)	Oxford Nanopore Technologies	FLO-MIN106D
PromethION Flow Cell (R9.4.1)	Oxford Nanopore Technologies	FLO-PRO004RA
Q5 High-Fidelity DNA Polymerase kit	New England Biolabs	M0491
T4 Polynucleotide Kinase kit	Thermo Scientific	Thermo Scientific
Oligonucleotides		
Nanopore IVT T7 forward primer	IDT (custom order)	5′-TAATACGACTCACTATAGCG AGGCGGTTTTCTGTTGGTGC TGATATTGCT-3′
Nanopore IVT T7 reverse primer	IDT (custom order)	5′-ACTTGCCTGTCGCTCTATCTTC-3′
Deposited data		
SH-SY5Y IVT DRS data	McCormick, C.A., Akeson, S., Tavakoli, S., Bloch, D., Klink, I.N., Jain, M., and Rouhanifard, S.H. (2023). Multicellular, IVT-derived, unmodified human transcriptome for nanopore direct RNA analysis. bioRxiv. 10.1101/2023.04.06.535889.	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA947135
SH-SY5Y DRS data, untreated, differentiated and lead-treated libraries	This work	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1092333
Experimental models: Cell lines		
SH-SY5Y cells	American Type Culture Collection (ATCC)	CRL-2266
Software and algorithms		
Minimap2 v.2.17-r941	Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100. https://doi.org/10.1093/bioinformatics/ bty191.	https://github.com/lh3/minimap2
Rstudio v.4.3.3	RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA	*http://www.rstudio.com/*
SAMtools	N/A	https://www.htslib.org/
Mod-p ID	Tavakoli, S., Nabizadeh, M., Makhamreh, A., Gamper, H., McCormick, C.A., Rezapour, N.K., Hou, Y.-M., Wanunu, M., and Rouhanifard, S.H. (2023). Semi-quantitative detection of pseudouridine modifications and type I/II hypermodifications in human mRNAs using direct long-read sequencing. Nat. Commun. 14, 334.	https://github.com/RouhanifardLab/PsiNanopore
ModQuant	Makhamreh, A. et al. Messenger-RNA Modification Standards and Machine Learning Models Facilitate Absolute Site-Specific Pseudouridine Quantification. Preprint at https://doi.org/10.1101/2022.05.06.490948 (2022).	https://github.com/wanunulab/ModQuant
NeuronalEpitranscriptomePlasticity	This work	https://doi.org/10.5281/zenodo.14706600

MATERIALS AND EQUIPMENT

Strand-switching buffer	
Reagent	Volume
5X RT Buffer	4 μL
Nuclease-Free Water	1 μL
SSP (10 μM)	2 μL
RNaseOUT	1 μL
Total	8 μL

STEP-BY-STEP METHOD DETAILS

RNA isolation from cells

Timing: 2 h

In this section, we describe steps for the total RNA isolation from cultured cells. Total RNA is extracted using TRIzol followed by purification with the PureLink RNA Mini Kit. RNA is collected from cells grown in 100 mm culture dishes at approximately 80–90% confluence. This protocol has been successfully applied to various cell types, including immortalized HeLa, HepG2, NTERA, Jurkat, SHSY5Y, and A549 cells, as well as primary blood T cells.

Note: To ensure sufficient RNA for each library preparation, we recommend starting with six culture dishes. This number may vary depending on cell size and RNA yield—larger or smaller cells may require adjustments. For suspension cells, such as Jurkat, we recommend starting with at least 10^7 cells. If preparing multiple libraries (e.g., one for the IVT control and one for the sample), scale the input accordingly.

1. Prepare all the necessary equipment and reagents during the pre-experiments step.
   a. Pre-cool the centrifuge to 4°C.
   b. Store 1× PBS at 4°C for 10 min (to prevent freezing). Alternatively, hold the tube on wet ice to chill.
   c. Clean around the centrifuge and the chemical fume hood first with ethanol and then with RNaseZap.
   d. Make fresh 70% ethanol with histology grade ethanol (prepare 0.4 ml of 70% ethanol for each 1 ml of TRIzol).
   e. Prepare an ice bucket.
2. Add 3 mL of ice-cold PBS to each culture dish and aspirate. Repeat twice to wash the media from the cells.
3. In the chemical fume hood, add 2 ml TRIzol to each 10 cm dish (or equivalent) and pipette up and down several times around the dish. Incubate for 5 min at 25°C.

   Note: Always handle steps with TRIzol or chloroform in the fume hood.

   Note: When working with suspension cells, add 2 mL of TRIzol per 5×10^6 cells, then proceed with the incubation step. We usually add 12 mL of TRIzol for a 1×10^7 total cell pellet.

   ⚠ CRITICAL: The following steps should be performed quickly to ensure a high-quality, long-read library.

4. Transfer each 1 ml of TRIzol to a 1.5 ml microtube (total of 12 microtubes) and vortex each tube for 30 s.

5. Add 200 μL chloroform to each tube and shake for 15 sec (vortexing not recommended), then incubate at RT for 3 min. Centrifuge at 12000 × *g* for 15 min at 4°C.
6. Transfer 400 μL from the aqueous phase (top layer) to a new tube. Troubleshooting 1.

⚠ CRITICAL: Avoid touching the interphase and organic phase, as this will contaminate the RNA sample. Do this quickly.

7. Add 400 μL of 70% ethanol to the aqueous phase and vortex.
8. Insert 6 PureLink RNA Mini Kit silica columns into the collection tube.
9. Place half of the microtubes (6 microtubes) on ice.
10. From each of the other half of the microtubes, add 700 μL of the solution to one of the 6 PureLink RNA Mini Kit silica columns inserted into the collection tube. Centrifuge at 12000 × *g* for 15 s at 4°C and discard the flow through.

Note: Use one silica column for two tubes containing 700 μL of the RNA+ethanol solution.

11. Take the other half of the microtubes from the ice and repeat step 10, adding 700 μL of the solution to the same 6 silica columns used in step 10.
12. Add 700 μL wash buffer 1 to the column. Then centrifuge at 12000 × *g* for 15 s.
13. Discard the flow through and the collection tube. Use a new collection tube.
14. Add 500 μL wash buffer 2 to the column. Then centrifuge at 12000 × *g* for 15 s. Discard the flow through. Repeat this step for a total of two washes.
15. Centrifuge at 12000 × *g* for 1 min to dry the sample and discard the collection tube.
16. Insert the column into the recovery tube and add 50 μL of nuclease-free (NF) water. Incubate at 25°C for 3 min.
17. Centrifuge at 12000 × *g* for 2 min, then pool the eluates together in one microtube.
18. Measure the concentration of total RNA using Nanodrop. Alternatively, the RNA concentration can be assessed with Qubit.

Note: We need a minimum of 25 μg of pure, high-quality total RNA from this step for each library to perform the Poly(A) RNA selection as the next step. Therefore, when preparing one library, a successful RNA extraction should result in a concentration of at least 100 ng/μL. If you're preparing two libraries, one for the IVT control and one for your sample, you will need 50 μg total, as in the following steps, 25 μg of mRNA will be used for the Poly(A) selection of IVT control, and 25 μg will be used for the sample.

Note: If the concentration is lower than expected after this step, we recommend to repeat the RNA extraction. If this is not possible because the sample is precious, more RNA can be extracted from the aqueous phase left in the tubes from step 6. Transfer the remaining aqueous phase to new tubes being extremely careful to not touch the interphase and proceed with the following steps.

19. Store the total RNA in a 1.5 mL DNA LoBind Eppendorf −80°C freezer or proceed to Poly(A) isolation.

Poly(A) RNA selection

Timing: 1 h 20 min

In this section, we describe steps for Poly(A) RNA selection, which are based on the NEBNext Poly(A) mRNA Magnetic Isolation Module protocol (E7490), available here: https://international.neb.com/-/media/nebus/files/manuals/manuale7490.pdf?rev=8fc35eb8aee94d0b90f0b671aa5a46ac&hash=C8CCC69263E3FEA33B00126133F6A118.

Note: The starting amount of DNA-free total RNA for our protocol is 25 μg. The original protocol provides volumes for one library with 1–5 μg total RNA; therefore, as we use a 5 times higher initial amount, all reagents within the module need to be scaled accordingly (5X).

Alternatives: The thermal cycler steps can be performed using a heat block to reach 65°C and 80°C. The 4°C steps can be performed by holding the sample on ice, while the 25°C steps can be performed by holding the sample at 25°C. We successfully used this alternative in 10+ library preps.

Note: For best results, keep all the reagents used during the Poly(A) isolation (except the NEBNext Oligo d(T)25 beads), on ice when not in use.

20. Prepare a DNA-free total RNA sample containing 25 μg of RNA in a nuclease-free microcentrifuge tube.

Note: The Poly(A)+ selection protocol allows for 1–5 μg of input RNA per reaction, each diluted to 50 μL in nuclease-free water. Dilute the 25 μg of total RNA in nuclease-free water to a final volume of 250 μL (5 × 50 μL).

21. Perform Poly(A)+ RNA isolation by following the steps described in the NEBNext Poly(A) mRNA Magnetic Isolation Module (Standard Protocol): https://international.neb.com/-/media/nebus/files/manuals/manuale7490.pdf?rev=8fc35eb8aee94d0b90f0b671aa5a46ac&hash=C8CCC69263E3FEA33B00126133F6A118.

Note: Make sure to scale all reagents 5X as the starting material is 5 times higher than the standard range. The final elution is performed in 12.5 μL of the NEBNext Tris Buffer.

22. Assess the yield of the poly(A) RNA using RNA HS Qubit assay.

Note: You need 500 ng (RNA002) or 300 ng (RNA004) of Poly(A) selected mRNA for the library preparation of your sample. If the library preparation is going to be performed for the IVT library, the Poly(A) selected RNA needs to be in vitro transcribed before performing the library preparation. The in vitro transcription is described in the IVT preparation protocol, which uses a starting amount of 100 ng. Note that after the IVT preparation is complete, 500 ng (RNA002)/300 ng (RNA004) of in vitro transcribed RNA will be needed for the library preparation of the IVT sample. When the yield at this step is low and we are working with a precious sample for which a new Poly(A) selection is not possible, we usually proceed with the following steps keeping in mind that the final sequencing results will allow us to analyze highly expressed mRNAs only.

23. Freeze the sample in −80°C freezer or proceed with the nanopore library preparation.

Note: If you are proceeding with an *in vitro* transcribed (IVT) library, proceed with the steps described in the IVT preparation section and then follow the RNA sequencing library preparation section. Otherwise, if preparing a library only for your sample, jump to the RNA sequencing library preparation section.

IVT preparation

Timing: 7 h 30 min (total)

Timing: 2 h (for step 25: first-strand RT and strand-switching)

Timing: 3 h 30 min (for step 26: PCR amplification)

Timing: 2 h (for step 28: *in vitro* transcription)

In this section, we describe steps for the unmodified transcriptome (IVT) generation. The reasons for generating an IVT library have been described extensively[2]: the IVT library is an essential control for RNA modification detection performed with DRS, as it provides an unmodified version of the native sample. This allows to detect and filter out problematic k-mers and single nucleotide variations (SNVs), which guarantees that only mismatches due to post-transcriptional modifications will be detected in the computational steps. The following protocol incorporates steps from ONT PCR-cDNA Sequencing Kit (SQK-PCS109), HiScribe T7 High Yield RNA Synthesis Kit (E2040S), and Monarch RNA Cleanup Kit (50 μg) (T2040S). The starting amount of DNA-free RNA is 100 ng. The IVT preparation needs to be followed by the steps described in the RNA sequencing library preparation section before loading the final library into the sequencing device. Therefore, when preparing an IVT library, we perform a Direct RNA preparation of the IVT control.

Note: For best results, all reagents should be kept on ice while not in use.

Briefly, to generate the IVT library, we start with Poly(A)-selected total RNA and synthesize complementary DNA (cDNA) using reverse transcription and strand-switching. These steps are adapted from the ONT PCR-cDNA Sequencing Kit (SQK-PCS109), not for sequencing cDNA but to obtain high-quality double-stranded DNA (dsDNA) as the template for in vitro transcription. The dsDNA is then transcribed into RNA using the HiScribe T7 High Yield RNA Synthesis Kit (E2040S). The resulting unmodified RNA is cleaned up using the Monarch RNA Cleanup Kit (T2040S) and subsequently prepared for nanopore sequencing using the standard ONT Direct RNA library preparation kit.

Note: The ONT PCR-cDNA Sequencing Kit (SQK-PCS109) used in this protocol is a legacy product. The current version of this product is the SQK-PCS114 kit. The steps here described can be applied with the new version of the kit.

24. Before starting the experiment.
    a. Clean the bench with ethanol and RNaseZap.
    b. Thaw VN Primer (VNP) (SQK-PCS109), Strand Switching Primer (SSP) (SQK-PCS109), dNTPs, and 5X RT buffer on ice.
25. Perform the first strand, reverse transcription (RT), and strand-switching.
    a. In a 0.2 mL PCR tube, prepare 100 ng of Poly(A)-selected RNA in 9 μL nuclease-free water.
    b. Add 1 μL VNP (2 μM) and 1 μL dNTPs (10 μM) to the RNA, then gently flick the tube to mix and briefly spin down the sample on a benchtop centrifuge.
    c. Incubate at 65°C for 5 min and snap cool on ice for 5 min.
    d. In a 1.5 mL DNA LoBind Eppendorf tube, prepare the strand-switching buffer by mixing the following reagents:

Note: It is important to use DNA LoBind Eppendorf tubes, as their low-binding surface minimizes RNA adhesion to the tube walls.

Reagent	Volume
5X RT Buffer	4 μL
Nuclease-Free Water	1 μL
SSP (10 μM)	2 μL
RNaseOUT	1 μL
Total	8 μL

e. Add 8 μL of strand-switching buffer to the snap-cooled mRNA, then gently flick the tube to mix and briefly spin down the sample on a benchtop centrifuge.
f. Incubate the reaction at 42°C for 2 min.
g. Add 1 μL of Maxima H Minus Reverse Transcriptase, 1 × 10,000 units (200 U/μL) for a final total volume of 20 μL, then gently flick the tube to mix and briefly spin down the sample on a benchtop centrifuge.

Note: Always verify the enzyme's expiration date before use. Using expired reverse transcriptase can compromise your results and lead to insufficient yield for downstream applications.

h. Incubate the sample using the following thermocycler protocol:

Step	Temperature	Time	Cycles
RT and Strand-Switching	42°C	90 mins	1
Heat Inactivation	85°C	5 mins	1
HOLD	4°C	∞	1

26. Perform PCR amplification.

Note: Each PCR reaction uses 5 μL of reverse-transcribed RNA. Do NOT use all 20 μL of the reverse transcription reaction in a single PCR reaction. Reverse transcriptase is a PCR inhibitor and the RT material must be diluted enough for the PCR to take place.

a. Prepare 2 reactions in separate 0.2 mL PCR tubes at 25°C.

Note: To ensure sufficient yield for IVT, these reactions will be pooled for the following cleanup step.

b. Add 5 μL reverse-transcribed RNA sample to a 0.2 mL PCR tube and add the following reagents:

Note: The primers used in this step were custom-ordered from IDT.

Primer	Sequence
Nanopore IVT T7 Forward Primer	5′-TAATACGACTCACTATAGCGAGG CGGTTTTCTGTTGGTGCTGATATTGCT-3′
Nanopore IVT T7 Reverse Primer	5′-ACTTGCCTGTCGCTCTATCTTC-3′

Reagent	Volume
Nanopore IVT T7 Forward Primer (10 μM)	4 μL
Nanopore IVT T7 Reverse Primer (10 μM)	1 μL
2X LongAmp Taq Master Mix	2 μL
Nuclease-Free Water	1 μL
Total	8 μL

c. Gently flick the tube to mix and briefly spin down the sample on a benchtop centrifuge.

d. Amplify the RT product using the following cycling conditions:

Step	Temperature	Time	Cycles
Initial Denaturation	95°C	30 sec	1
Denaturation	95°C	15 sec	11
Annealing	62°C	15 sec	11
Extension	65°C	15 min	11

Step	Temperature	Time	Cycles
Initial Denaturation	95°C	30 sec	1
Denaturation	95°C	15 sec	11
Final Extension	65°C	15 min	1
HOLD	4°C	∞	1

Note: The 4°C hold can be maintained as needed (up to 24 h) to resume the next steps the following day.

e. Add 1 μL of NEB Exonuclease I directly to each PCR tube, then gently flick the tube to mix and briefly spin down the sample on a benchtop centrifuge.

Note: Ensure the enzyme is not expired for optimal results.

f. Incubate the reaction at 37°C for 15 min, followed by 80°C for 15 min for heat inactivation.

27. Cleanup the PCR product:
    a. Pool the two reactions together in a clean 1.5 mL DNA LoBind Eppendorf tube.
    b. Vortex the Sera-Mag beads and add 250 μL to the PCR product.
    c. Gently mix by pipetting and incubate at ambient temperature for 10 min.
    d. Briefly spin down the mixture, then place the tube in a magnetic rack for 5 min to magnetically separate the beads.
    e. Aspirate the supernatant to waste.
    f. With the tube in the magnetic rack, gently add 1 mL of 70% ethanol. Be careful not to disturb the beads and incubate for one min.
    g. Aspirate supernatant to waste and repeat the previous wash step with 70% ethanol 2X.
    h. Dry the pellet to remove all ethanol residue by quickly spinning it on a benchtop centrifuge, then placing the tube on the magnetic rack for 1 min and carefully pipetting off all residual liquid.
    i. Resuspend the dried magnetic bead pellet in 30 μL Tris-HCl 10 μM, pH 7.0), with gentle pipetting, and heat the sample at 40°C for 15 min to complete DNA extraction from the beads.
    j. Gently flick the beads and briefly spin down before placing the tube on a magnetic rack for 5 min to separate the beads.
    k. Pipet and save the cleared supernatant solution containing the eluted DNA in a fresh 1.5 mL DNA LoBind Eppendorf tube.
    l. Analyze 1 μL of the amplified DNA using Qubit 1X dsDNA HS buffer. You need 1 μg of DNA for the next *in vitro* transcription step.

Note: If DNA concentration is too low (< 1 μg) for the following reaction, repeat the PCR amplification and elute the DNA in 10 μL of Tris-HCl.

28. Perform *in vitro* transcription (IVT) using the HiScribe T7 High Yield RNA Synthesis Kit (E2040S), following the manufacturer's protocol for standard RNA synthesis.

29. Bring the total volume of the *in vitro* transcribed RNA to 50 µL adding 30 µL of nuclease-free water and perform a cleanup using the Monarch RNA Cleanup Kit (50 µg) (T2040S), following the manufacturer's protocol.
30. Use Qubit to measure RNA concentration.

Note: We typically perform a 1:100 dilution and expect the diluted sample to contain at least 10 ng/µL of RNA.

Note: If the concentration is insufficient, we recommend to repeat the in vitro transcription described at step 28.

31. Perform a Poly(A) tail addition.
    a. In a 0.2 mL PCR tube, prepare 12 µg RNA in 29 µL nuclease-free water and add the following reagents in the order specified:

Reagent	Volume
10× E. coli Poly(A) Polymerase Reaction Buffer	4 µL
ATP (10 mM)	4 µL
RNaseOut	1 µL
E. coli Poly(A) polymerase	2 µL
Total	8 µL

Note: Ensure the enzyme is not expired for optimal results.

    b. Incubate the reaction in a thermocycler at 37°C for 10 min.
    c. Stop the reaction by adding 0.8 µL EDTA (0.5 M).
    d. Add 10 µL nuclease-free water.
32. Bring the total volume of the poly-adenylated RNA to 50 µL with nuclease-free water and perform a cleanup using the Monarch RNA Cleanup Kit (50 µg) (T2040S), following the manufacturer's protocol.
33. Quantify the poly(A) RNA using Qubit.

Note: To proceed with library preparation, you'll need 500 ng of RNA for the RNA002 kit and 300 ng for RNA004. Lower input amounts can be used, but this may lead to a reduced library yield.

34. Proceed with RNA sequencing library preparation or store the *in vitro* transcribed Poly(A) RNA in a −80°C freezer.

RNA sequencing library preparation

Timing: 3 h 30 min

In this section, we describe the library preparation steps used for both the IVT control and your sample. 300 ng of starting material is necessary to proceed for RNA004 chemistry and 500 ng for RNA002. If you have insufficient material—either for the IVT control or your Poly(A)-selected sample—you may increase the volume of RNA used in the RTA (Reverse Transcription Adapter) reaction to a maximum of 15 µL by following the table below. In such cases, do not add RCS (RNA Calibration Strand) to the RTA reaction as the RCS is optional according to the manufacturer's protocol. This allows to replace the RCS volume with more RNA.

35. Direct RNA sequencing library preparation is performed on the previously extracted and Poly(A) selected RNA following the ONT Direct RNA Sequencing Kit (SQK-R002) protocol, which can be found here: https://nanoporetech.com/document/direct-rna-sequencing-sqk-rna002.

The updated ONT Direct RNA Sequencing Kit (SQK-R004) protocol for the new chemistry can be found here: https://nanoporetech.com/document/direct-rna-sequencing-sqk-rna004.

Note: We have successfully used 500 ng of Poly(A) selected RNA obtained from the previous Poly(A) RNA selection step as the starting material for RNA002. We have successfully used 300 ng of Poly(A) selected RNA obtained from the previous Poly(A) RNA selection step as the starting material for RNA004 direct RNA sequencing library preparations.

Note: If your RNA yield is low, you can scale the entire RTA reaction of the ONT Direct RNA Sequencing Kit (SQK-R004) protocol up to 2× the standard volume increasing the RNA input volume to a maximum of 15 µL. When scaling, be sure to adjust all reagents proportionally to maintain the correct reaction conditions. Below is an example of how to modify both the RTA reaction mix and the reverse transcription mix accordingly.

Adjust the RTA reaction mix if needed:

Reagent	Original volume	Corrected Volume (2X)
RNA from previous step	8.5 µl	15 µl
NEBNext Quick Ligation Reaction Buffer	3 µl	6 µl
Diluted RNA CS (RCS) or nuclease-free water	0 µl	0 µl
Murine RNase Inhibitor	1 µl	2 µl
RT Adapter (RTA)	1 µl	2 µl
T4 DNA Ligase	1.5 µl	3 µl
Total	15 µl	28 µl

Adjust the reverse transcription mix as follows:

Reagent	Original volume	Corrected volume
Nuclease-free water	13 µl	0 µl
10 mM dNTPs	2 µl	2 µl
5X Induro RT reaction buffer	8 µl	8 µl
Total	23 µl	10 µl

By eliminating nuclease-free water from the reverse transcription mix, you maintain a consistent final reaction volume of 38 µL (28 µL RTA + 10 µL RT mix). This approach allows you to accommodate a higher input RNA volume while preserving reaction efficiency. We successfully used this approach in the past.

Sequencing and preprocessing

Timing: 1.5 days

In this section, we describe steps for analyzing DRS data. Processing of raw ONT sequencing data involves the use of several software tools, including the Guppy basecaller and the Minimap2 aligner. It is strongly recommended to install Conda and perform all analyses within dedicated Conda environments.

To facilitate the use of the protocol by people without an extensive computational background we also provide a quick guide on how to install miniconda3, create a conda environment, and install necessary software in it.

Note: All data processing was performed on Northeastern University Discovery Cluster but this protocol may be adapted for other high-performance computing clusters.

36. Install Miniconda using the following commands:

```
#1. pull miniconda3 from anaconda.com

wget --quiet https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh

#2. check hash key of package

sha256sum Miniconda3-latest-Linux-x86_64.sh

#3. install miniconda3

bash Miniconda3-latest-Linux-x86_64.sh -b -p /work/directory/<YOUR DIR>/SOFTWARE/miniconda3

#4. source miniconda3 (if miniconda3 is active your terminal will show: (base) [<USER>@<NODE>
<CURRENT DIR>])

source /<YOUR DIR>/miniconda3/bin/activate
```

37. Install minimap and samtools in a conda environment using the following commands:

```
#required packages: samtools (v1.3.1) using htslib (v1.3.1)|minimap2 (v2.24)

#1. create secondary environment for alignment

conda create -n align2.24

#2. activate|switch to secondary environment

conda activate align2.24

#3. install required packages

conda install -c bioconda samtools

conda install -c bioconda minimap2

#4. return to (base)

conda deactivate
```

38. All FAST5 files were output from MinKNOW onto the local hard drive and exported to the Northeastern University Discovery Cluster using the following command:

```
scp <filename> <username>@xfer.discovery.neu.edu:/scratch/<username>
```

Note: This command is to be run locally on the machine containing the raw data files

Note: If data are acquired with RNA004 kit, MinKNOW output will be in POD5 format and must be basecalled with Dorado.

39. Basecalling was performed on an NVIDIA A100 GPU using Guppy v6.4.2, and the rna_r9.4.1_70bps_hac_prom model with the following command:

```
/<PATH TO GUPPY FOLDER>/ont-guppy/bin/guppy_basecaller -i /<YOUR FAST5 FOLDER>/ -s /<OUTPUT
FOLDER>/ -c /<PATH TO GUPPY FOLDER>/ont-guppy/data/rna_r9.4.1_70bps_hac_prom.cfg -x cuda:
all:100% -r --u_substitution
```

Note: Basecalling can be performed with any other preferred version of Guppy by using the same command.

Note: If basecalling does not finish within the allocated job time, append the "–resume" tag to the end of the above command and rerun as follows:

```
/<PATH TO GUPPY FOLDER>/ont-guppy/bin/guppy_basecaller -i /<YOUR FAST5 FOLDER>/ -s /<OUTPUT FOLDER>/ -c /<PATH TO GUPPY FOLDER>/ont-guppy/data/rna_r9.4.1_70bps_hac_prom.cfg -x cuda:all:100% -r --u_substitution --resume
```

Note: For RNA004 libraries, basecalling must be performed with the Dorado basecaller on POD5 files. This will output an unaligned bam file that must be converted to FASTQ format before alignment with minimap. The detailed description of available commands is available at the following link: https://github.com/nanoporetech/dorado.

With the following commands, we provide one example of Dorado usage with Modified basecalling enabled for pseudouridine (–modified-bases pseU tag), and how to convert the raw unaligned bam file to FASTQ format retaining the base modification annotations (-TMM,ML,MM,Ml tags).

```
#basecalling

/<PATH TO DORADO FOLDER>/bin/dorado basecaller /<PATH TO DORADO MODELS FOLDER>/rna004_130bps_hac@v5.1.0 /<PATH TO YOUR POD5 FOLDER> --modified-bases pseU --emit-moves > /<OUTPUT FOLDER>/

#unaligned bam to FASTQ conversion

samtools fastq -@ 64 -TMM,ML,MM,Ml /<PATH TO DORADO'S OUTPUT RAW_BAM FOLDER>/ > /<OUTPUT FASTQ>/
```

40. The FASTQ files were then merged into one file using the following command:

```
cat /<YOUR FASTQ FOLDER>/*.fastq > merged.fastq
```

41. Alignment was performed on the merged FASTQ file using minimap 2.24.

Note: We recommend installing minimap2 and samtools in a conda environment as shown in steps 36 and 37.

Note: Alignment protocol is the same for both RNA002 and RNA004 libraries.

```
#1.source your conda environment

source /<YOUR DIR>/miniconda3/bin/activate

#2.activate the environment in which minimap2 and samtools are installed

conda activate align2.24

$REFGENOME=/<PATH TO REFERENGE .fasta GENOME>/

$MERGEDFQ=/PATH TO MERGED .fq/

$SAM_UNF=/<PATH TO OUTPUT UNFILTERED SAM>/
```

```
$SAM=/<PATH TO OUTPUT SAM>/

$BAM=/<PATH TO OUTPUT BAM FILE>/

#3.run alignment steps

minimap2 -t 32 -ax splice -uf -k 14 $REFGENOME $MERGEDFQ | samtools sort -o $SAM_UNF -T
reads.tmp

samtools view -h -F 4 -F 256 -F 2048 $SAM_UNF > $SAM

samtools view -S -b $SAM > $BAM

samtools index $BAM
```

Pseudouridine detection and filtration

Timing: 8 h

In this section, we describe the steps for detecting pseudouridine sites transcriptome-wide. The pseudouridine detection is performed with Mod-*p* ID[3] in R. We advise performing all the steps on RStudio. All the steps required for the pseudouridine detection are described in the INSTRUCTIONS_Pvalue-pipeline.pdf document available at: github.com/RouhanifardLab/PsiNanopore/tree/main/Pvalue-Pipeline.

> *Note:* This tool detects psi sites by comparing the sample library to the IVT control. In the absence of a paired IVT library available for your data and work on human cell lines, you can consider using a paired IVT library from McCormick et al.[2] Alternatively, if your cell line is not in the sequenced list, merge all the bam files in a unique pan-IVT and proceed with the following steps. To generate a panIVT, refer to previous papers from our lab.[1,4]

Mod-p ID is divided in five scripts. The description of key functions of each script is described below.

Step1_PileUp.R script reads aligned and indexed sequencing data from BAM files corresponding to both direct RNA and IVT samples. The script generates distinct pileup summaries for both the sample and IVT, which include counts of nucleotide occurrences at each positions.

Step2_Merge_replicate_pileups.R script reads the pileup files generated from the previous step, each corresponding to a different replicate and IVT, and combines the pileup data across replicates in one single merged file. In this way, each position detected in the sequencing data includes the summaries from all the replicates and the IVT.

Step3_add_kmer.R annotates each candidate psi site with their corresponding k-mer sequence (k=5). As psi sites are modified uridine (U) sites, the code filters the dataset to retain only sites where the reference base in the genome is a T, which corresponds to U in RNA. For each site, the script uses the reference genome to extract the 5-mer sequence centered around the site.

Step4_IVT_Kmer_Analysis.R reads the file containing the U positions and associated k-mers and calculates the frequency of each unique k-mer across the IVT. This allows to establish a baseline k-mer distribution in the IVT sample and find over/underrepresented k-mers in the unmodified RNA. This file is used as a lookup table in the next step, as it provides the expected basecalling error (U-to-C mismatch) for each specific sequence context.

Step5_Add_pvalues.R calculates the p-values for each U position to quantify the likelihood of observed modification being statistically significant. The p-value is calculated by comparing the

U-to-C mismatch of the direct sample to the IVT control. This step produces the final Merged_with_P_vals.csv file, which contains the pileup information, the k-mers and the p-values for each U site found across replicates in the sample and IVT libraries.

42. Move all the .bam and corresponding indexed .bam files (.bai) obtained from the preprocessing step into one folder.

    Note: All the DRS .bam files and the .bam IVT files must be in the same folder. One IVT file is used to process multiple DRS bam files.

43. Clone or download the ZIP folder at https://github.com/RouhanifardLab/PsiNanopore.
44. Unzip the folder and open the PsiNanopore-main/Pvalue-Pipeline folder.
45. Open Step1_PileUp.R script.
    a. Follow the instructions found in the INSTRUCTIONS_Pvalue-pipeline.pdf document to modify the relevant pieces of code:
        i. Write short names for your libraries (if you have replicates, use numbers to distinguish them, i.e., DRS1, DRS2) in the "replicate.name" field of the meta.data.df variable. Troubleshooting 2.
        ii. Provide the names of the bam files moved to the same folder at step 42 following the same order as the "replicate.name" field.
        iii. Provide the path of the reference genome fasta file in hg38.fa variable. The fasta file can be downloaded from https://www.gencodegenes.org/.
        iv. Provide the path of the gff3 gencode file in g variable. The gff3 file can be downloaded from https://www.gencodegenes.org/.
        v. Provide the path of the folder created at step 42 containing all DRS and IVT bam files.
        vi. Provide the output folder, which can be located where desired but named init_gene_pileup.
    b. Create the init_gene_pileup folder in the desired location before running the script.
    c. Open the init_gene_pileup folder and create one folder for each element of "replicate.names" specified in step 45.a.i.

    Note: These folders must have the same names specified in step 45.a.i. Refer to the INSTRUCTIONS_Pvalue-pipeline.pdf document for a practical example. Troubleshooting 3.

    d. Return to RStudio and run the script Step1_PileUp.R script. Wait for it to be completed.
46. Open Step2_Merge_replicate_pileups.R in Rstudio.
    a. Provide the path to the init_gene_pileup folder obtained from the previous script in the data.dir variable, as indicated in the INSTRUCTIONS_Pvalue-pipeline.pdf document.
    b. Modify merged.df using the names specified in step 45.a.i.
    c. Create a folder named Merged under the init_gene_pileup folder. This will store the results of this script.
    d. Return to RStudio and run the script Step2_Merge_replicate_pileups.R and wait for it to be completed.
47. Open Step3_add_kmer.R.
    a. Provide the path to the Merged folder obtained from Step 46 in the data.dir variable as indicated in the INSTRUCTIONS_Pvalue-pipeline.pdf document.
    b. Create the Processed_pileup folder in the init_gene_pileup directory.
        i. Provide the Processed_pileup path in the out.dir variable.
    c. Follow the INSTRUCTIONS_Pvalue-pipeline.pdf to create a raw.pileup.df.2 variable for each nucleotide (A/C/T/G) in each replicate specified in 45.a.i.
    d. Run the Step3_add_kmer.R script.
48. Open Step4_IVT_Kmer_Analysis.R.
    a. Provide the Processed_pileup folder location in the data.dir variable.

b. Create a folder named kmer_analysis inside the init_gene_pileup folder to store the kmer analysis produced in this step.
   i. Provide the path to the kmer_analysis folder in the k.mer.summary.out.dir variable and include the name of the file in which to store the results. We usually name this IVT_kmer_Analysis.csv.
c. Create a folder named finalized_pileup inside the init_gene_pileup folder.
   i. Provide the path of the finalized_pileup folder in the merged.Uonly.Pileup.out.dir variable and include the name of the file in which to store the results. We name this merged_U_only.csv.
d. Modify target.cols as explained in the INSTRUCTIONS_Pvalue-pipeline.pdf.
e. Run the script Step4_IVT_Kmer_Analysis.R.

49. Open Step5_Add_pvalues.R.
    a. Provide the path to the IVT_kmer_Analysis.csv file generated by the previous script in the kmer.summary variable.
    b. Provide the path to the merged_U_only.csv file in the merged.df variable.
    c. Modify the merged.df variable according to your replicate names, as explained in the INSTRUCTIONS_Pvalue-pipeline.pdf.
    d. Add the p.value field to the merged.df variable for all the replicates except IVT following the INSTRUCTIONS_Pvalue-pipeline.pdf.
    e. Add an if statement for each replicate except IVT according to the names provided in 45.a.i.
    f. Modify the final line with the path of the file in which to store the final CSV file containing the detected psi-sites. We usually call this file Merged_with_P_vals.csv and save it under the finalized_pileup folder created at step 48c. Troubleshooting 4.
    g. Run the script.

Differential analysis to investigate psi dynamics in different cellular states

Timing: 1 h

In this section, we describe how to perform differential analysis on the untreated and RA-differentiated libraries described in Fanari et al., where we applied Mod-p ID to untreated and RA-differentiated SH-SY5Y cells. This allowed us to determine the U-to-C mismatch, read coverage and p-values at each U position (T in the genomic reference) using the IVT library as a control to establish the expected background in the unmodified RNA. We then applied a series of filtering steps to define high-confidence modification sites. The filtering steps, described in the original paper and in the next steps of this protocol, took into consideration occupancy levels (U-to-C %mm >10 in the sample and U-to-C %mm <10 in the IVT; Figure 1), coverage (>30 reads per site in the sample and >10 in the IVT), significance (p-value<0.05) and signal-to-noise ratio (SNR>1). After defining the sites, we performed differential analysis to assess whether modification occupancy levels differed between the untreated and treated conditions. To determine the threshold at which the occupancy levels are different between conditions we calculate the confidence interval (CI) of the differential U-to-C mismatch. Although this protocol uses untreated and RA-differentiated cells as an example, it can be used for any perturbation, as shown by the analysis performed on the untreated vs lead-treated SH-SY5Y cells in the same paper.

All the code described in this section can be found at the following link: https://github.com/RouhanifardLab/NeuronalEpitranscriptomePlasticity.

Caution: When working with PromethION datasets, very small differences in modification profiles in DRS and IVT libraries can become significant because of the large read coverage. For this reason, the Mod-*p* ID p-values filtration needs to be coupled with reads and mismatch filtration, and particular caution needs to be taken to ensure a sufficient signal-to-noise ratio (SNR) for each psi position.

Figure 1. IVT filtration criteria to select modified sites in DRS libraries can be visualized with IGV

(A) IGV screenshot shows a valid psi-site with a U-to-C %mm <10 in the IVT library, as shown by the clean (gray) bands.

(B) IGV screenshot shows an example of a candidate psi-site we exclude as the U-to-C %mm >10 in the IVT library, as shown by a colored band.

50. Load the Merged_with_P_vals.csv file in Rstudio.

```
DRS<-read.csv("path_to_Merged_with_P_vals.csv file'')
```

51. Filter the Merged_with_P_vals.csv file.
    a. Filter and keep only significant psi-sites p<0.001. If working with replicates, we usually filter to have at least N-1 out of N significant replicates.

Note: Due to the high read depth in PromethION libraries, even minor differences in basecalling errors between DRS and IVT can become statistically significant. Therefore, after applying the p-value filter in this step, additional filtering based on mismatch rate and coverage in the subsequent steps is necessary to ensure high-quality candidate psi-sites.

```
SH_filter <- SH[which( ((SH$p.value.Diff1 < 0.001 & SH$p.value.Diff2 < 0.001) | (SH$p.value.Diff1 < 0.001 & SH$p.value.Diff3 < 0.001) | (SH$p.value.Diff2 < 0.001 & SH$p.value.Diff3 < 0.001)) ) |((SH$p.value.Undiff1 < 0.001 & SH$p.value.Undiff2 < 0.001) | (SH$p.value.Undiff1 < 0.001 & SH$p.value.Undiff3 < 0.001) |
```

b. Filter and keep only sites with a low U-to-C mismatch in the IVT library (%mm<10). This step filters out potential SNV sites.

```
SH_filter <- SH_filter[which(SH_filter$mm.IVT <10),]
```

c. If working with replicates, calculate the U-to-C mismatch for the merged replicates for each condition.

```
SH_filter$mm.Diff <- 0

SH_filter$mm.Undiff <- 0

for (i in c(1:nrow(SH_filter))) {

   SH_filter$mm.Diff[i] <- (SH_filter$C_Diff[i]/ (SH_filter$C_Diff[i] + SH_filter$T_Diff[i]))*100

   SH_filter$mm.Undiff[i] <- (SH_filter$C_Undiff[i] / (SH_filter$C_Undiff[i] + SH_filter$T_Undiff[i]))*100

   }
```

d. If working with replicates, calculate the reads for the merged replicates for each condition.

```
SH_filter$T_Diff <- 0

SH_filter$C_Diff <- 0

SH_filter$T_Undiff <- 0

SH_filter$C_Undiff <- 0

for (i in c(1:nrow(SH_filter))) {

   SH_filter$T_Diff[i] <- SH_filter$T_Diff1[i] + SH_filter$T_Diff2[i] + SH_filter$T_Diff3[i]

   SH_filter$C_Diff[i] <- SH_filter$C_Diff1[i] + SH_filter$C_Diff2[i] + SH_filter$C_Diff3[i]

   SH_filter$T_Undiff[i] <- SH_filter$T_Undiff1[i] + SH_filter$T_Undiff2[i] + SH_filter$T_Undiff3[i]

   SH_filter$C_Undiff[i] <- SH_filter$C_Undiff1[i] + SH_filter$C_Undiff2[i] + SH_filter$C_Undiff3[i]

  }
```

e. Calculate the total U and C reads found at each position.

Note: An alternative is to calculate the total reads found at each position by summing all the A, C, G, and U occurrences from the Merged_with_P_vals.csv file.

```
SH_filter$total_Diff <- 0

SH_filter$total_Undiff <- 0

SH_filter$total_IVT <- 0

for (i in c(1:nrow(SH_filter))) {

    SH_filter$total_Diff[i] <- SH_filter$T_Diff[i] + SH_filter$C_Diff[i]

    SH_filter$total_Undiff[i] <- SH_filter$T_Undiff[i] + SH_filter$C_Undiff[i]

    SH_filter$total_IVT[i] <- SH_filter$T_IVT[i] + SH_filter$C_IVT[i]

  }
```

f. Filter and keep only sites with many reads in the IVT and DRS libraries. We usually use at least 10 reads.

Note: Steps b–d account for deep coverage in PromethION reads and ensure that significant sites have enough reads.

```
SH_filter <- SH_filter[which(SH_filter$total_Diff > 10 & SH_filter$total_Undiff > 10),]
```

52. Calculate the SNR values. This accounts for deep coverage in PromethION reads and ensures that significant sites have enough reads and occupancy.
    a. Generate files for each condition analyzed. In this case, we provide an example of RA-differentiated and untreated libraries.

```
#write file using the Diff reads

Diff_filter_Amr<-data.frame(SH_filt$Annotation,SH_filt$chr,SH_filt$position,SH_filt
$T_Diff,SH_filt$C_Diff,SH_filt$T_IVT,SH_filt$C_IVT)

colnames(Diff_filter_Amr)<-c
("Annotation","chr","position","DRS_U","DRS_C","IVT_U","IVT_C")

write.csv(Diff_filter_Amr,"path to save Diff_file.csv",row.names = F)

#write file using the Undiff reads

Diff_filter_Amr<-data.frame(SH$Annotation,SH$chr,SH$position,SH$T_Undiff,SH$C_Undiff,
SH$T_IVT,SH$C_IVT)                                          colnames(Diff_filter_Amr)<-c
("Annotation","chr","position","DRS_U","DRS_C","IVT_U","IVT_C")

write.csv(Diff_filter_Amr,"path to save Undiff_file.csv",row.names = F)
```

b. Run the Python script to generate the SNR values for the RA-differentiated library.

```
python3 SNR_marginal_likelihood.py Diff_file.csv SNR_Diff_output.csv
```

c. Run the Python script to generate the SNR values for the untreated library.

```
python3 SNR_marginal_likelihood.py Undiff_file.csv SNR_Undiff_output.csv
```

d. Load the SNR files in RStudio.

```
#load Differentiated SNR file obtained with .py script

snrDiff<-read.csv("path to SNR_Diff.csv file", header=T)

names(snrDiff)[names(snrDiff) == "SNR"] <- "SNR_Diff"

#load Unrteated SNR file obtained with .py script

snrUndiff<-read.csv("path to SNR_Undiff.csv file", header=T)

names(snrUndiff)[names(snrUndiff) == "SNR"] <- "SNR_Undiff
```

e. Filter for sites with an SNR>1 in Rstudio.

```
#define threshold for SNR filtration

sig=1

#merge relevant column from the Differentiated library

selected_columns<-c("Annotation", "position","SNR_Diff")

SH_filter <- merge(SH_filter, snrDiff[selected_columns], by = c("Annotation", "position"),
all.x = TRUE)

#merge relevant column from the Untreated library

selected_columns<-c("Annotation", "position","SNR_Undiff")

SH_filter <- merge(SH_filter, snrUndiff[selected_columns], by = c("Annotation", "position"),
all.x = TRUE)

#calculate the SNR value for each position

SH_filter$N_value<-pmax(SH_filter$SNR_Undiff,SH_filter$SNR_Diff,na.rm = TRUE)

#find all psi-sites with sufficient SNR values

SH_filter_sigma2<-SH_filter[which(SH_filter$N_value>sig),]
```

53. Perform orthogonal validation (optional).

This step enables verification of Mod-p ID-identified psi-sites using independent methods. You may reference previously published techniques, such as those described in Fanari et al., or experimentally validate sites in your specific cell line by knocking down known PUS enzymes and observing a reduction in mismatch errors at modified positions.

Note: When comparing to non-nanopore, chemically-based orthogonal methods, we often observe a position shift of ±2 nucleotides.

Note: Be mindful of the cell line used in the orthogonal method—differences in cell type may explain a lack of site overlap.

54. Perform the differential analysis.
    a. Calculate the U-to-C mismatch difference between the two conditions.

```
SH_filter_sigma2$mm.DiffminusUndiff<-SH_filter_sigma2$mm.Diff-SH_filter_sigma2$mm.Undiff
```

b. Calculate the confidence interval (CI) for the differential U-to-C mismatch. This indicates the differential U-to-C mismatch percentage at which a significant difference occurs between the two conditions.

```
data<-SH_filter_sigma2

mean_data <- (mean(-data$mm.DiffminusUndiff))

std_error <- abs(sd(data$mm.DiffminusUndiff)) /sqrt(nrow(data))

# Confidence level (e.g., 95%)

confidence_level <- 0.95

z_value <- qnorm((1 + confidence_level) / 2)

# Calculate confidence interval

lower_bound <- mean_data - z_value * std_error

upper_bound <- mean_data + z_value * std_error

# View the confidence interval

c(lower_bound, upper_bound)
```

c. Assess sites with a significant difference in U-to-C delta by using the lower bound of the CI. In this work, we use five different intervals to assess groups with different levels of plasticity.

Note: After calculating the CI interval, we decided to use a slightly higher value (the cutoff was set at a 5% mismatch difference) to make our differential analysis more conservative.

```
SH_filter_sigma2$mm.UndiffminusDiff=-SH_filter_sigma2$mm.DiffminusUndiff

int1<-SH_filter_sigma2[which(SH_filter_sigma2$mm.UndiffminusDiff > 20 ),]

int2<-SH_filter_sigma2[which(SH_filter_sigma2$mm.UndiffminusDiff >= 5 & SH_filter_sigma2
$mm.UndiffminusDiff <= 20),]

int3<-SH_filter_sigma2[which(SH_filter_sigma2$mm.UndiffminusDiff > -5 & SH_filter_sigma2
$mm.UndiffminusDiff < 5),]

int4<-SH_filter_sigma2[which(SH_filter_sigma2$mm.UndiffminusDiff <= -5 & SH_filter_sigma2
$mm.UndiffminusDiff >= -20 ),]

int5<-SH_filter_sigma2[which(SH_filter_sigma2$mm.UndiffminusDiff < -20),]
```

EXPECTED OUTCOMES

The expected RNA yield after IVT preparation varies depending on the quality and quantity of input RNA, but typically yields range from 10–20 μg of polyadenylated RNA suitable for RNA sequencing library preparation. For your sample RNA sequencing (DRS) libraries, starting with 500 ng of poly(A)-selected RNA routinely yields sufficient sequencing output for robust downstream analysis.

In terms of computational analysis, successful pipeline execution should result in a table of statistically significant ψ-sites. These are derived by comparing mismatch patterns between DRS and IVT

datasets. The final CSV file used in this pipeline is outputted by Step5_Add_pvalues.R. Although the previous steps produce output CSV tables, these are used by the Mod-p ID pipeline and we don't use them in the next filtration steps.

The CSV file contains: the k-mer centered at each uridine site detected in each direct replicate and the IVT library; the read coverage (number of T and C nucleotides and the total number of reads detected at each position) for each uridine site detected in each direct replicate and the IVT; the U-to-C mismatch percentage calculated for each uridine site detected in each direct replicate and IVT; the p-value calculated for each uridine site detected in each direct replicate and the IVT.

In IGV (Integrative Genomics Viewer) plots, true positive psi-sites are typically characterized by elevated U-to-C mismatches in the DRS sample and negligible mismatches in the IVT sample, which lacks modifications. The IGV traces can be visualized using the same aligned and indexed BAM files used as input by Step1_PileUp.R in Mod-p ID. We use IGV traces to manually inspect positions and ensure that the IVT traces have negligible mismatch to the genomic reference.

To minimize false positives, we apply a stringent filter: we retain only those sites where the U-to-C mismatch in the IVT library is <10%. This threshold helps ensure that retained sites are not sequence variants or artifacts due to problematic k-mers. Importantly, this allows the inclusion of lower-coverage sites, as the low IVT mismatch gives confidence in their validity. A typical output includes several hundred high-confidence psi-sites per library, depending on coverage and cell type.

LIMITATIONS

This protocol, while robust and reproducible, presents several limitations that users should be aware of. Nanopore-based detection of RNA modifications, including pseudouridine, is inherently noisy. The Mod-*p* ID method attempts to overcome this by using an unmodified IVT control, but the approach still results in a high false positive rate. To mitigate this, we apply stringent filtration criteria, which—while effective—also substantially reduce the final number of reported sites (many false negatives). Therefore, we limit our final psi-sites to a rigorous but small dataset.

This method detects the presence/absence and relative occupancy differences of ψ-sites but does not yield precise stoichiometry. For researchers interested in quantifying the fraction of molecules modified at a site, we recommend using synthetic RNA strands and dedicated quantification tools like ModQuant.[5]

Results may be affected by differences in ONT chemistries (e.g., RNA002 vs. RNA004) and platforms (e.g., MinION vs PromethION), especially in terms of read depth and coverage. Users must account for these differences when comparing datasets.

The accuracy of detection relies heavily on having a matched IVT library. If an IVT control for the same cell type is unavailable, results may suffer in accuracy. Public IVT datasets can be used, but users should interpret findings cautiously if there are sequence or expression-level mismatches.

TROUBLESHOOTING

Problem 1

Difficulty isolating only the aqueous phase during RNA extraction, leading to contamination from the interphase.

Potential solution

Consider using Phasemaker Tubes to simplify phase separation and reduce contamination. Detailed usage instructions are available here: https://www.thermofisher.com/order/catalog/product/A33248.

Problem 2

When running a step of Mod-*p* ID, the code stops at the beginning of one of the Steps.

Potential solution

- If an error occurs at the beginning of one of the steps, check if you created all the folders needed to run the script and ensure the spellings are correct and match the names provided in the script.

All the input/output/working folders are:

Step1_PileUp.R script	Step2_Merge_replicate_pileups.R	Step3_add_kmer.R	Step4_IVT_Kmer_Analysis.R
init_gene_pileup folder	Merged folder	Processed_pileup folder	kmer_anlysis folder
One folder for each element of "replicate.names"			

- If the code stops without issues with the folders, check the spelling of bam files and folder names.
- If the code stops without issues with folder creation/bam files, check the reference fasta and gff3 files.
- This type of error could also be generated by misspelling the replicate's names. The replicate's names chosen at step 45.a.i. need to be the same and used as suffixes in all necessary variables with the same spelling specified at 45.a.i. point of the Mod-p ID protocol.

Problem 3

The files created by Mod-p ID contain the same values copy-pasted in multiple columns or only zeroes in a column.

Potential solution

- The names of two different replicates were misspelled in one of the variables. Therefore, the same values were copied in two different columns. Check for misspellings.
- The name of one variable was misspelled, and therefore, the script could not find any non-zero values in the previous files when parsing them. Check for misspellings.
- In general, when this happens, search the oldest file in which the error starts to appear and trace the specific script (Step 1/2/3/4/5) that generated the file with the wrong column. Search for misspellings in the specific Step.

Problem 4

The final step Step5_Add_pvalues.R produced an error at the end, and I don't see a Merged_with_P_vals.csv file in the finalized_pileup folder.

Potential solution

The final line of the Step5_Add_pvalues.R script needs to be modified to include the output Merged_with_P_vals.csv file path. Check for misspellings.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Sara H. Rouhanifard (s.rouhanifard@northeastern.edu).

Technical contact

Technical questions on executing this protocol should be directed to and will be answered by the technical contact, Sara H. Rouhanifard (s.rouhanifard@northeastern.edu).

Materials availability

This study did not generate new materials.

Data and code availability

- All fastq data for Direct Libraries generated in this work have been made publicly available in the NIH NCBI SRA. The BioProject accession number is listed in the key resources table.
- This paper analyzes existing, publicly available data. All IVT Libraries used in this work were sourced from the NIH NCBI SRA. The BioProject accession number is listed in the key resources table.
- All original code has been deposited at github.com/RouhanifardLab/NeuronalEpitranscriptomePlasticity and is publicly available at the date of publication. The DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

S.H.R. acknowledges support from the National Institutes of Health (R01HG011087 and R01HG012856).

AUTHOR CONTRIBUTIONS

Conceptualization, O.F., M.M., and S.H.R.; methodology, O.F. and S.H.R.; software, O.F.; formal analysis, O.F. and M.M.; investigation, O.F. and M.M.; writing – original draft, O.F., D.B., M.M., and S.H.R.; writing – review and editing, S.H.R., O.F., D.B., and M.M.; visualization, O.F.; supervision, S.H.R.; project administration, S.H.R.; funding acquisition, S.H.R.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Fanari, O., Tavakoli, S., Qiu, Y., Makhamreh, A., Nian, K., Akeson, S., Meseonznik, M., McCormick, C.A., Bloch, D., Gamper, H., et al. (2025). Probing enzyme-dependent pseudouridylation using direct RNA sequencing to assess epitranscriptome plasticity in a neuronal cell line. Cell Syst. *16*, 101238.

2. McCormick, C.A., Akeson, S., Tavakoli, S., Bloch, D., Klink, I.N., Jain, M., and Rouhanifard, S.H. (2024). Multicellular, IVT-derived, unmodified human transcriptome for nanopore-direct RNA analysis. GigaByte *2024*, gigabyte129.

3. Tavakoli, S., Nabizadeh, M., Makhamreh, A., Gamper, H., McCormick, C.A., Rezapour, N.K., Hou, Y.-M., Wanunu, M., and Rouhanifard, S.H. (2023). Semi-quantitative detection of pseudouridine modifications and type I/II hypermodifications in human mRNAs using direct long-read sequencing. Nat. Commun. *14*, 334.

4. McCormick, C.A., Meseonznik, M., Qiu, Y., Fanari, O., Thomas, M., Liu, Y., Bloch, D., Klink, I. N., Jain, M., Wanunu, M., and Rouhanifard, S.H. (2024). mRNA psi profiling using nanopore DRS reveals cell type-specific pseudouridylation. Preprint at bioRxiv. https://doi.org/10.1101/2024.05.08.593203.

5. Makhamreh, A., Tavakoli, S., Fallahi, A., Kang, X., Gamper, H., Nabizadehmashhadtoroghi, M., Jain, M., Hou, Y.-M., Rouhanifard, S.H., and Wanunu, M. (2024). Nanopore signal deviations from pseudouridine modifications in RNA are sequence-specific: quantification requires dedicated synthetic controls. Sci. Rep. *14*, 22457.