Chemical probe-based Nanopore Sequencing to Selectively Assess the RNA modifications

Soundhar Ramasamy[a†]. Vinodh J Sahayasheela[b†], Zutao Yu [a], Takuya Hidaka[b], Li Cai[c], Hiroshi Sugiyama[a,b*] Ganesh N. Pandian[a*]

[a] Institute for Integrated Cell-Material Science (WPI-iCeMS), Kyoto University, Sakyo, Kyoto 6060-8501, Japan

[b] Department of Chemistry, Graduate School of Science, Kyoto University, Sakyo, Kyoto 606-8502, Japan

[c] Cell and Developmental Biology Graduate Program, Rutgers University, Piscataway, New Jersey, United States of America; Department of Biomedical Engineering, Rutgers University, Piscataway, New Jersey, United States of America.

[*] Correspondence: ganesh@kuchem.kyoto-u.ac.jp; hs@kuchem.kyoto-u.ac.jp Yoshida Ushinomiya-cho, Sakyo-Ku, Kyoto 606-8501, Japan, TEL: +818097589979

[†]Both authors contributed equally to the work

ABSTRACT

RNA modifications contribute to RNA and protein diversity in eukaryotes and lead to amino acid substitutions, deletions, and changes in gene expression levels. Several methods have developed to profile RNA modifications, however, a less laborious identification of inosine and pseudouridine modifications in the whole transcriptome is still not available. Herein, we address the first step of the above question by sequencing synthetic RNA constructs with inosine and pseudouridine modification using Oxford Nanopore Technology, which is a direct RNA sequencing platform for rapid detection of RNA modification in a relatively less labor-intensive manner. Our analysis of multiple nanopore parameters reveals mismatch error majorly distinguish unmodified versus modified nucleobase. Moreover, we have shown that acrylonitrile selective reactivity with inosine and pseudouridine generates a differential profile between the modified and treated construct. Our results offer a new methodology to harness selectively reactive chemical probe-based modification along with existing direct RNA sequencing methods to profile multiple RNA modifications on a single RNA.

Keywords: Chemical probe, Direct RNA Sequencing, Oxford Nanopore technology, RNA modifications, Selective reactivity

INTRODUCTION

Transcriptome-wide profiling of RNA modifications have shifted its focus from high abundant non-coding RNAs to mRNAs of minuscules fraction. RNA modifications exert unpreceded regulation over major aspects of mRNA life such as structure [1,2,3,4] ,stability [5,6] decay[7], translation [8,9,10] microRNA binding [11,12] and altering codon potential [10] . To date, 172 modifications (Modomics)[13] are known to exist in biological systems based on mass spectrometry characterizations. Of the above, only N6-methyladenosine (m^6A)[14], inosine (I)[15], pseudouridine(Ψ)[16,17] , N1 -methyladenosine (m^1A) [18,19] and 2′-O-Methylation (Nm)[20,21] are few modifications with transcriptome-wide mapping protocols. Since RNA modifications are silent to the reverse transcription (RT), most of the above protocols employ antibody (or) modification specific chemicals for adduct generation. These adduct induced mutations or truncation profiles are used as a proxy identifier of modifications with single-nucleotide resolution. Major shortcomings of the current methods are 1) multi-step sample preparation results in lesser reproducibility, 2) quantifying stoichiometry of the RNA modifications from these methods are not possible owing to the fragmentation of RNA and 3) simultaneously mapping of co-occurring modifications on the single mRNA molecule is difficult[22].

Recently, oxford nanopore technology (ONT) based direct RNA sequencing provides a solution for the above-described shortfalls. ONT operates by ratcheting DNA/RNA into a protein pore and upon migration triggers a change in the current, which ensues the inference of a nucleic acid sequence. Since the base-calling (current to nucleotide sequence conversion) algorithms are trained on conventional bases such as A, G, C and T/U, any modified bases present in RNA may deviate from the standard model. The resulting difference between modified and unmodified nucleotides could alter nanopore read parameters like base quality, mismatch, deletion, current intensity, and dwell time. Such alterations can be used to detect RNA modifications with single nucleotides resolution[23]. In contrast to second generation NGS like Illumina, ONT direct RNA sequencing does not require reverse transcription or RNA fragmentation. Recently, Liu et al. [24] systematically analysed the m^6A modification behaviour in nanopore by using unmodified and m^6A modified synthetic RNA. The comparative assessment reveals that the presence of m^6A RNA modification could decrease the base quality

and increase the deletion and mismatch frequency with respect to the unmodified adenosine.

In this work, we assessed the behaviour of pseudouridine (Ψ) and inosine RNA modifications using nanopore direct RNA sequencing. Inosine modifications in cells are catalysed by adenosine deaminase (ADAR) by post-transcriptional hydrolytic deamination of adenosine. Inosine base pairs with cytosine and stabilizes or destabilizes RNA structures in a sequence-dependent manner[4]. Inosine modifications are mostly read as guanosine by translational machinery. Hence, it can alter the coding potential of an mRNA, one of the well-studied example is exonic A-I editing of Gria2 gene in the brain, but the majority of ADAR targets are in non-coding regions of the mRNA[10]. A-I editing also prevents cytosolic innate immune response against endogenous double-stranded RNA structures[25,26]. Altered ADAR activities are implicated in complex diseases like cancer[26,27], auto-immune disorder[28], and autism[29]. On the other hand, pseudouridine synthetase catalyse the isomerization of uridine to pseudouridine. Pseudouridine is the most abundant of all modifications enriched in non-coding RNAs and also in mRNA. Pseudouridine is shown to be dynamically regulated in response to stress conditions[28–30]. Mutation in some pseudouridine synthetase are implicated in following diseases X-linked dyskeratosis congenita[31], mitochondrial myopathy and sideroblastic anaemia [32].

For both Inosine and Ψ RNA modification, second generation NGS based transcriptome-wide mapping protocols are available. Acrylonitrile[15] and *N*-cyclohexyl-*N'*-*b*-(4-methylmorpholinium) ethylcarbodiimide (CMC)[16,17] are two chemical probes used for mapping Inosine and pseudouridine modifications, respectively . Sakurai et al first employed acrylonitrile for transcriptome-wide mapping of inosine in mouse and the human brain - inosine chemical erasing sequencing. At pH 8.6 acrylonitrile cyanoethylates inosine and Ψ at N1 position to form N1-cyanoethylinosine (CEI) and N1- cyanoethyl Ψ (CEΨ), with higher reactivity towards inosine. Of the above mentioned adducts, CEI stops the RT and generates truncated short reads, while CEΨ adduct remains silent and undetectable. Hence inosine chemical erasing sequencing only detects inosine modification. Independently, Carlile et al. and Schwartz et al. established the transcriptome profiling protocol for Ψ. Both took advantage of CMC selective reactivity towards Ψ to form N3-CMC-Ψ, which strongly show RT stop signature and aid in mapping Ψ with single-nucleotide resolution.

ONT direct RNA sequencing readily overcomes the shortfalls associated with second generation NGS techniques via its intrinsic ability to detect RNA modifications in full length RNA, and has been deployed successfully for transcriptome wide m^6A detection. In this work, we assessed the nanopore parameters such as base quality, mismatch, deletion, current intensity, or dwell time for inosine and Ψ RNA modifications. We also hypothesized that acrylonitrile adducts CEI and CEΨ can further create a differential profile when compared to the unmodified or Inosine/Ψ RNA. This can be a add stringency for high confident detection of Inosine/Ψ in a comparative manner. Further CEI and CEΨ adduct induced profile can help to differentiate other modifications converging on adenosine and uridine.

RESULTS & DISCUSSION

To understand the changes in nanopore parameters in the context of RNA modifications (Ψ and Inosine) with and without acrylonitrile adduct, we generated a synthetic RNA using *in vitro* transcription (IVT) reaction (**Figure 1**). The synthetic RNA sequences generated from IVT assay are given in the supplementary. In our initial attempt with heavily modified synthetic RNAs (~25% of Ψ and ~25% inosine in the same transcripts), it produces reads that mostly failed base calling, which render it difficult to perform sequence alignment (Data not shown here). In the later attempts, we used synthetic RNA with ~8% of either Inosine (or) Ψ modification, respectively. Acrylonitrile adducts of CEΨ and CEI were generated on the modified RNAs using cyanoethylation reaction at pH 8.6, 70ºC for 30 mins. To confirm the modification, digestion of the synthetic RNA was performed and further analyzed using HPLC. In the acrylonitrile treated modified RNA additional peak was observed thereby confirming the presence of CEΨ adduct formation. (**Figure S2**). Unmodified, modified and cyanoethylated modified RNAs were sequenced using nanopore direct RNA sequencing platform (**Table S1**). Data for inosine modification is under preparation and currently not included in this version.

ONT assign all nucleotide read-out to four letters A, G, C and U (T) during base calling analysis, while the mismatch error indicates high-incidence of unnatural or modified base. Comparison of all three dataset (unmodified, Ψ and CEΨ) reveals that most mismatch errors are enriched at Ψ and CEΨ containing positions, to suggest the fidelity of ONT-based sequencing of base modification (**Figure 2a**). Although few errors in unmodified regions, having comparative datasets including reference sequence, unmodified and CE-treated samples work well in assisting to filter out these unexpected events. The mismatch error in place of Ψ and CEΨ was in the order of C > U > A, but the difference between Ψ and CEΨ mismatches was quantitatively milder (**Figure 2b**). Moreover, base quality analysis shows a substantial decrease between unmodified versus Ψ and CEΨ. There was a slight increase in CEΨ base quality when compared to Ψ alone, which is possibly due to the observed mismatch profile difference between Ψ and CEΨ (**Figure 2c**). Deletion and current intensity parameters of Ψ and CEΨ nucleobases show the difference with respect to unmodified condition, but between Ψ and CEΨ these differences were not substantial (**Figure 2c, d**). The dwell time was not significantly altered across all the three conditions (**Figure 2d**).

Various chemical modifications have been identified in the transcriptome that led to the field of epitranscriptomics. Most of the modifications play a significant role in various biological process, but the lack of generic mapping of transcriptome-wide modifications limits its detailed understanding. In this study, we have reported the identification of Inosine and pseudouridine by direct RNA sequencing as basecalling errors. We observed the base quality and mismatch error are the significant parameters that gets altered due to the presence of Ψ RNA

modification. The mismatch profile for pseudouridine was mostly towards the other two pyrimidine nucleobases- uridine and cytosine. Upon acrylonitrile treatment, direct RNA sequencing with ONT is proved to be feasible for detecting modified bases of pseudouridine and inosine in nucleotide resolution and at single RNA molecule level. Overall, our results show combining selective chemical probes and direct RNA sequencing could be an ideal strategy to identify various RNA modifications. Further studies would be evaluated in studying the chemical probes with different physical and electrochemical properties. This could open up the possibility of using the ONT platform for simultaneous mapping of multiple RNA modification with single nucleotide resolution.

MATERIALS AND METHODS

IVT template design and synthesis

Double-stranded DNA templates with T7 promoter for synthetic RNA were commercially purchased from IDT as blocks, which also has poly-A tail for nanopore adapter binding. IVT reactions were performed using MEGA script™ T7 Transcription Kit (AM1334), for overnight at 37°C followed by DNase treatment and purified using Quick Spin Columns for radiolabelled RNA purification (Roche, 11274015001). For Ψ and inosine modified RNAs synthesize, Pseudouridine-5'-Triphosphate (N-1019, Trilink) and Inosine-5'-Triphosphate (N-1020, Trilink) were used in place of uridine and guanosine, respectively.

Acrylonitrile cyanoethylation reaction

CE solution is prepared with 5 ml of 100% ethanol, 1.530 ml of 7.19 M Triethylamine (34804-85, Nacalai), and the pH was adjusted to 8.6 using acetic acid. The cyanoethylation of modified RNAs used 500ng of RNA, 30 µl of CE solution and 1.6 M of acrylonitrile and incubated at 70°C for 30 mins. Then, the reaction was immediately quenched by adding 160 ul of nuclease-free water on ice. Cyanoethylated RNA was precipitated using 0.1 volume of 3M sodium acetate and three volumes of 100% ethanol.

HPLC validation of cyanoethylation reaction

For nucleoside analysis, the IVT template RNA (0.01– 0.05 A_{260} units) was digested into nucleosides using 10 µg/ml nuclease P1 (New England Biolabs, M0660S), and 0.5 U/ml Bacterial Alkaline phosphatase (Takara 2120A) at 37°C for 1 h in 10 µl of a reaction mixture containing 20 mM HEPES–KOH pH 7.5(Nacalai 15639-84). The nucleosides were separated using a using an HPLC equipped with Chemcobond 5-ODS-H column (4.6 × 150 mm). Analysis was performed with a mobile phase containing solvent A containing 0.1% TFA (Trifluoracetic acid) in water in gradient combination with solvent B acetonitrile. The linear gradient started with 0% solvent B at 0 min to 10% B for 30 min, at a flow rate of 1 mL min^{-1}, monitored at 254 nm.

Direct RNA library preparation and sequencing

500ng of RNA was used for direct RNA seq library (SQK-RNA002) preparation following the ONT protocol version - DRS_9080_v2_revK_14Aug2019. Briefly, 500 ng of unmodified, modified and CE treated IVT RNA were ligated to ONT RT Adapter using concentrated T4 DNA ligase (NEB-M0202T), and was reverse transcribed using SuperScript III Reverse Transcriptase (Thermo Fisher Scientific-18080044). The products were purified using 1.8X Agencourt RNAClean XP beads (Fisher Scientific-NC0068576), washing with 70% freshly prepared ethanol. RNA Adapter (RMX) was ligated onto the RNA:DNA hybrid and the mix was purified using 1X Agencourt RNAClean XP beads, washing with Wash buffer twice. The sample was then eluted in elution buffer and mixed with RNA running buffer (RRB) prior to loading onto a primed R9.4.1 flow cell and ran on a MinION sequencer with MinKNOW acquisition software version v1.14.1.

Base calling and mapping

Fast5 files were basecalled using MinKnow- GUPPY (V 3.4.5) with accurate base calling enabled. Mapping to reference sequence were done using minimpa2 (version- 2.17-r941) with the setting -ax map-ont -L. Mapped reads were further filtered to remove unmapped, secondary, and supplementary and reads mapped on the reverse strand. Reads with low alignment were also removed with cut off MAPQ < 10, using options -bh -F 2324 -q 10. These reads were further sorted and indexed. This mapping workflow is adapted from https://nanocompore.rna.rocks/.

Extraction of nanopore parameters

The nanopore parameters like base quality, mismatch, and deletion were extracted using scripts associated with epinano package (https://github.com/Huanle/EpiNano). The epinano scripts were used to extract the per-site feature of the above described parameters.

- samtools view -h file.bam| java -jar sam2tsv.jar -r ref.fasta > file.bam.tsv
- python2 per_read_var.py file.bam.tsv > file.per_read.var.csv

The dwell time and current intensity of k-mers were extracted using scripts associated with https://nanocompore.rna.rocks/ using Nanopolish (v 0.11.1) and NanopolishComp (v.0.6.11)

- nanopolish index -s {sequencing_summary.txt} -d {raw_fast5_dir} {basecalled_fastq}
- nanopolish eventalign --reads {basecalled_fastq} --bam {aligned_reads_bam} --genome {transcriptome_fasta} --samples --print-read-names --scale-events --samples > {eventalign_reads_tsv}
- NanopolishComp Eventalign_collapse -i {eventalign_reads_tsv} -o {eventalign_collapsed_reads_tsv}

AUTHOR CONTRIBUTIONS

S.R and V.J.S. contributed equally to this work. S.R. conceived the idea, S.R and V.J.S. designed the work. G.N.P. designed the research, S.R and V.J.S. performed research, S.R. analyzed data along with the support of L.C.; T.H., Y.Z. and H.S. gave critical comments to improve the workflow. S.R and G.N.P. wrote the paper. The authors declare no conflict of interest.

This article contains supporting information after the main manuscript.

ACKNOWLEDGEMENTS

This work was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI (JP19H03349 to G.N.P, Grant No. JP16H06356 to H.S and AMED under Grant No. JP18am0301005 (Basic Science and Platform Technology Program for Innovative Biological Medicine), JP20am0101101 (Platform Project for Supporting Drug Discovery and Life Science Research (BINDS)), We also thank JSPS #S19127; NIH/NEI EY031439-01; NJCSCR grant #151RG006 and Rutgers - TechAdvance Fund

REFERENCES

1. Liu, N. *et al.* N6-methyladenosine-dependent RNA structural switches regulate RNA–protein interactions. *Nature* vol. 518 560–564 (2015).
2. Carlile, T. M. *et al.* mRNA structure determines modification by pseudouridine synthase 1. *Nat. Chem. Biol.* **15**, 966–974 (2019).
3. Grohman, J. K. *et al.* A Guanosine-Centric Mechanism for RNA Chaperone Function. *Science* vol. 340 190–195 (2013).
4. Brümmer, A., Yang, Y., Chan, T. W. & Xiao, X. Structure-mediated modulation of mRNA abundance by A-to-I editing. *Nat. Commun.* **8**, 1–13 (2017).
5. Pubmeddev & Stellos K, E. al. Adenosine-to-inosine RNA editing controls cathepsin S expression in atherosclerosis by enabling HuR-mediated post-transcriptional regulation. - PubMed - NCBI. https://www.ncbi.nlm.nih.gov/pubmed/27595325.
6. Karikó, K. *et al.* Incorporation of Pseudouridine Into mRNA Yields Superior Nonimmunogenic Vector With Increased Translational Capacity and Biological Stability. *Mol. Ther.* **16**, 1833 (2008).
7. Yoon, K.-J. *et al.* Temporal Control of Mammalian Cortical Neurogenesis by mA Methylation. *Cell* **171**, 877–889.e17 (2017).
8. Mao, Y. *et al.* m 6 A in mRNA coding regions promotes translation via the RNA helicase-containing YTHDC2. *Nat. Commun.* **10**, 1–11 (2019).
9. Eyler, D. E. *et al.* Pseudouridinylation of mRNA coding sequences alters translation. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 23068–23074 (2019).
10. Licht, K. *et al.* Inosine induces context-dependent recoding and translational stalling. *Nucleic Acids Res.* **47**, 3 (2019).
11. pubmeddev & Meyer KD, E. al. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. - PubMed - NCBI. https://www.ncbi.nlm.nih.gov/pubmed/22608085.
12. pubmeddev & García-López J, E. al. Reprogramming of microRNAs by adenosine-to-inosine editing and the selective elimination of edited microRNA precursors in mouse oocytes and preimpl... - PubMed - NCBI. https://www.ncbi.nlm.nih.gov/pubmed/23571754.
13. pubmeddev & Boccaletto P, E. al. MODOMICS: a database of RNA modification pathways. 2017 update. - PubMed - NCBI. https://www.ncbi.nlm.nih.gov/pubmed/29106616.
14. Hawley, B. R. & Jaffrey, S. R. Transcriptome-Wide Mapping of m A and m Am at Single-Nucleotide Resolution Using miCLIP. *Curr. Protoc. Mol. Biol.* **126**, e88 (2019).
15. Sakurai, M., Yano, T., Kawabata, H., Ueda, H. & Suzuki, T. Inosine cyanoethylation identifies A-to-I RNA editing sites in the human transcriptome. *Nat. Chem. Biol.* **6**, 733–740 (2010).
16. Li, X. *et al.* Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.* **11**, 592–597 (2015).
17. Schwartz, S. *et al.* Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148–162 (2014).
18. Li, X. *et al.* Transcriptome-wide mapping reveals reversible and dynamic N 1 -methyladenosine methylome. *Nat. Chem. Biol.* **12**, 311–316 (2016).
19. Dominissini, D. *et al.* The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **530**, 441–446 (2016).
20. Dai, Q. *et al.* Nm-seq maps 2'-O-methylation sites in human mRNA with base precision. *Nat. Methods* **14**, 695–698 (2017).
21. Incarnato, D. *et al.* High-throughput single-base resolution mapping of RNA 2′-O-methylated residues. *Nucleic Acids Res.* **45**, 1433 (2017).
22. Khoddami, V. *et al.* Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 6784–6789 (2019).
23. Garalde, D. R. *et al.* Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* **15**, 201–206 (2018).
24. Liu, H. *et al.* Accurate detection of m6A RNA modifications in native RNA sequences. *Nature Communications* vol. 10 (2019).
25. Hartner, J. C., Walkley, C. R., Lu, J. & Orkin, S. H. ADAR1 is essential for the maintenance of hematopoiesis and suppression of interferon signaling. *Nat. Immunol.* **10**, 109–115 (2009).
26. Lamers, M. M., van den Hoogen, B. G. & Haagmans, B. L. ADAR1: 'Editor-in-Chief' of Cytoplasmic Innate Immunity. *Front. Immunol.* **10**, (2019).
27. Gannon, H. S. *et al.* Identification of ADAR1 adenosine deaminase dependency in a subset of cancer cells. *Nat. Commun.* **9**, 1–10 (2018).
28. Roth, S. H. *et al.* Increased RNA Editing May Provide a Source for Autoantigens in Systemic Lupus Erythematosus. *Cell Rep.* **23**, 50 (2018).
29. pubmeddev & Tran SS, E. al. Widespread RNA editing dysregulation in brains from autistic individuals. - PubMed - NCBI. https://www.ncbi.nlm.nih.gov/pubmed/30559470.

30. van der Feltz, C., DeHaven, A. C. & Hoskins, A. A. Stress-induced Pseudouridylation Alters the Structural Equilibrium of Yeast U2 snRNA Stem II. *J. Mol. Biol.* **430**, 524 (2018).

31. Knight, S. W. *et al.* X-linked dyskeratosis congenita is predominantly caused by missense mutations in the DKC1 gene. *Am. J. Hum. Genet.* **65**, 50 (1999).

32. Patton, J. R., Bykhovskaya, Y., Mengesha, E., Bertolotto, C. & Fischel-Ghodsian, N. Mitochondrial Myopathy and Sideroblastic Anemia (MLASA). *J. Biol. Chem.* **280**, 19823–19828 (2005).

Figure. 1.

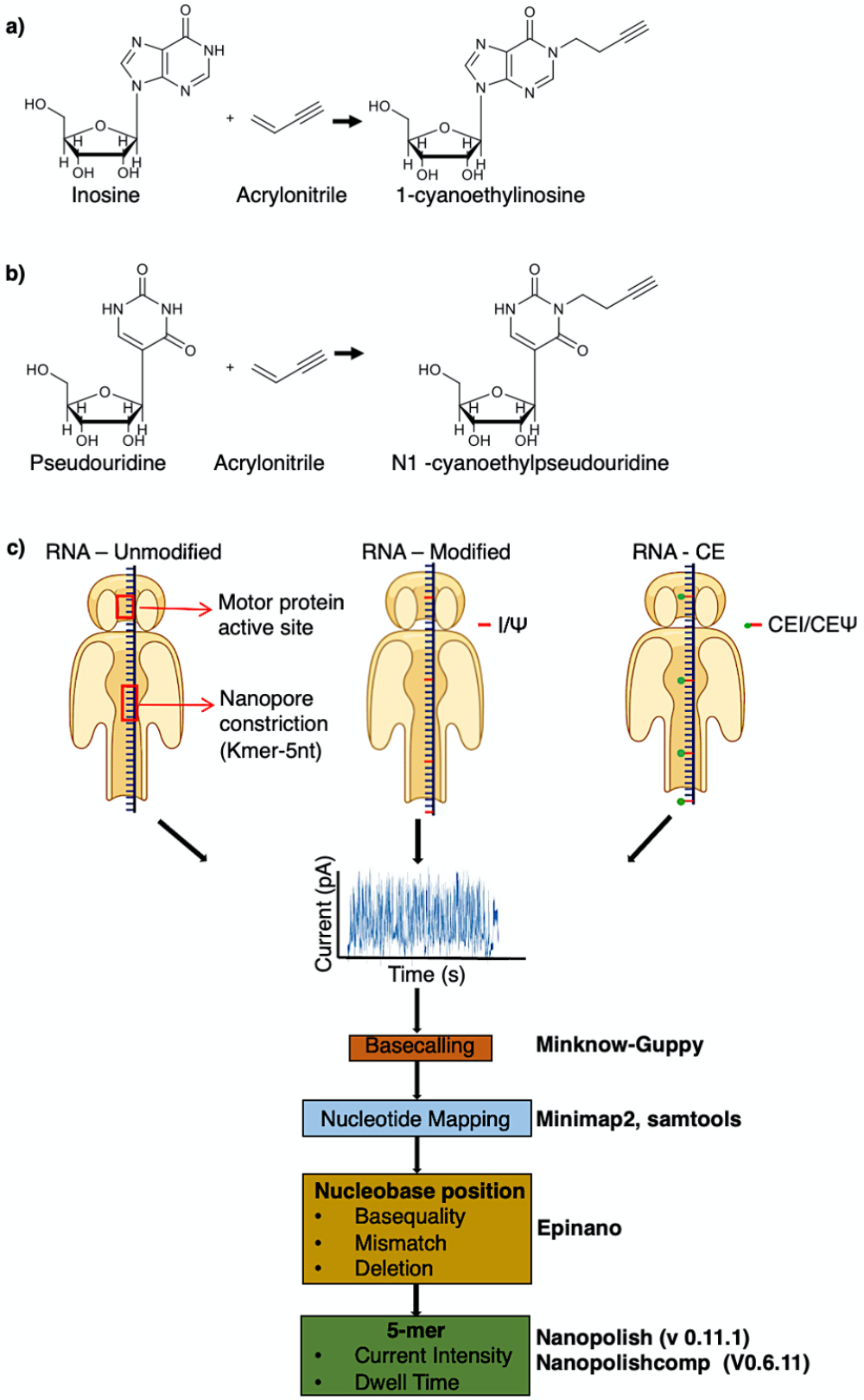

Fig. 1. Schematic illustration. Acrylonitrile cyanoethylation reaction of **a)** inosine, **b)** Ψ pseudouridine and **c)** Schematic workflow of nanopore direct RNA sequencing of unmodified, modified and cyanoethylated RNA and analysis pipeline.

Figure. 2.

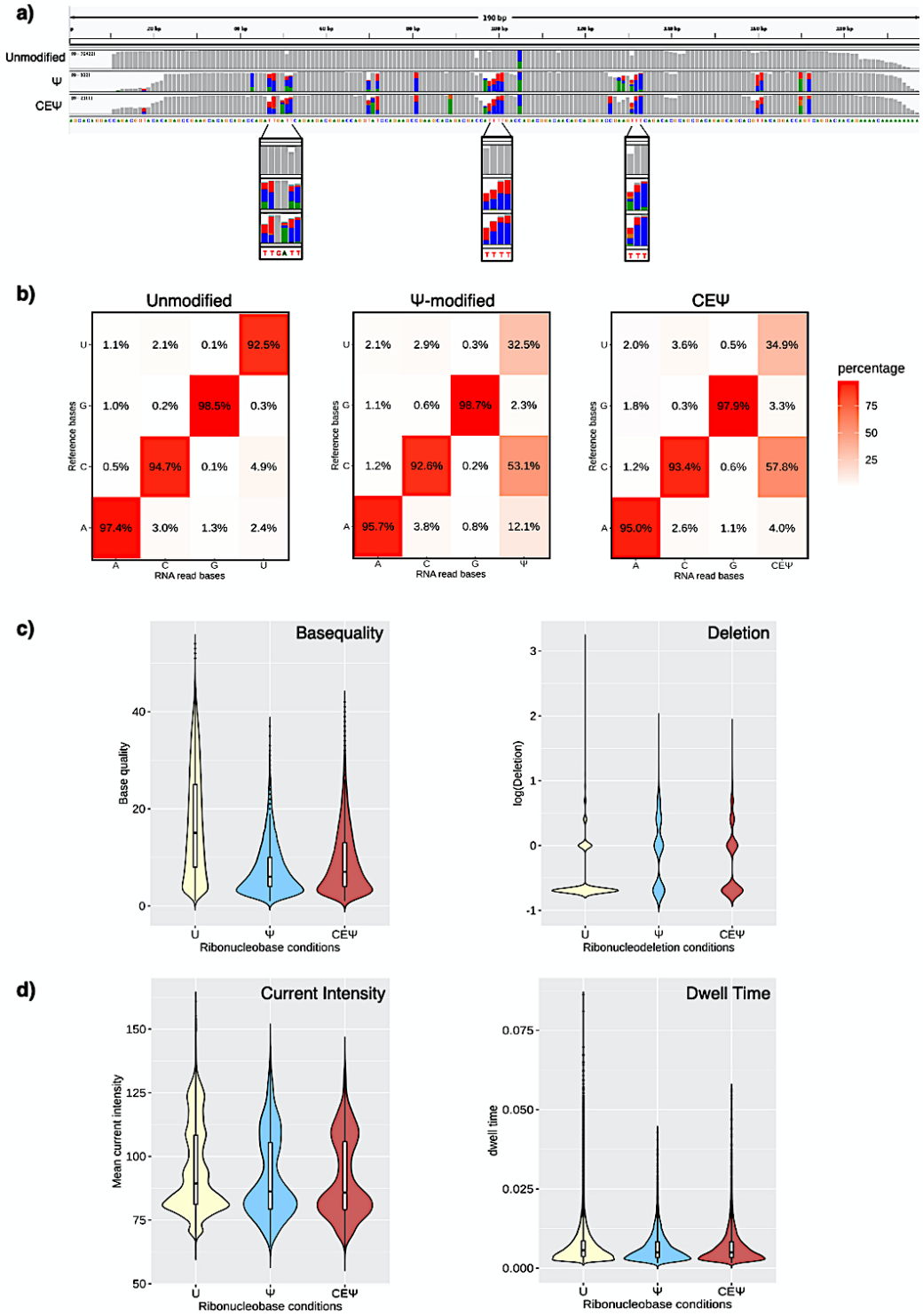

Figure 2. Altered nanopore parameters by Ψ and CEΨ modified nucleobases. a) IGV snapshot of unmodified, Ψ and CEΨ transcripts showing mismatch. Mismatch frequency > 0.2% are represented in colours. Green(adenosine), orange (guan), blue (cytosine) and red (Thymine). **b)** Substitution matrix of unmodified, Ψ and CEΨ transcripts native reads. The x-axis represents the base identity of nanopore reads. The y-axis represents base identity of reference transcript and **c)** Violin plot showing kernel density estimate & inner boxplot showing interquartile range and median **c)** Base quality and deletion of unmodified, Ψ and CEΨ nucleobase. The above parameters are calculated using scripts associated with epinano. **d)** Current intensity and dwell time of identical 5-mers of unmodified, Ψ and CEΨ transcripts. The above parameters are calculated using nanopolish and nanocompolish.

Supplementary Information

Figure. S1. IVT template design

>Inosine
AAGCTAATACGACTCACTATAGGATCCTATACCATACTGTCAACTACTTCAGCATCATACACTACTTACATCATCTACTCCATCAT
GGAGGCTTACCCACATTACCCATATTACTACTACTGAGCGCATACATACATCCATCATACTTACCATTCAGGGGTACCATCATAAC
TCATCAACTACTAGGGCCATCATTACCATTCATCAGGTACACTTACCATTTAGCATCATTACCATCAATACAACAAAAAAAAAA

>pseudouridine
AAGCTAATACGACTCACTATAGGAGCACAGGACCAGACGGTACACAGAGCCGAAGCACAGCAGACCAGATTGATTCAGAAGACG
AGACCAGGTATCCAGAAGCCGAAGCACAGACGACCATTTTGACCAGACGGACAACAGCAGAGACCGAAGTTTCAGACACGCAGC
GACAGAGCAGCACGTTACAGGACCAGTCAGGACAACAGAAAACAAAAAAAAAA

T7 promoter regions are underlined.
Grey and red highlights the reference sequence and modification positions, respectively.

Figure. S2. Total nucleoside analysis of IVT digest by HPLC

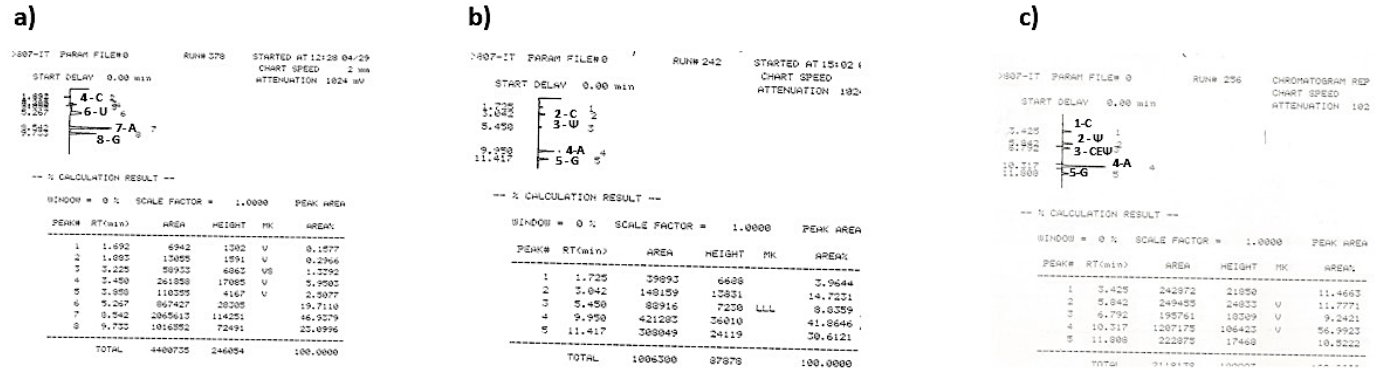

Fig.S2. a) Unmodified IVT showing A, U, G and C peaks, **b)** Modified IVT showing A, Ψ, G and C peaks and **c)** CE treated modified IVT showing A, Ψ, G, C and CE Ψ peaks.

Table. S1. Run statistics

Conditions	Flow cell	Active pores / run time	Reads	N50	Median length	Median PHRED score	Mapped reads
Unmodified (Inosine + Ψ)	New	1442 / 1.43 h	All reads - 9.47e+4 Pass reads - 8.91e+4	185 184	181 181	9.71 9.82	7.68e+4
modified (Inosine + Ψ)	Reuse	900/ 2.59h	All reads Pass reads	190 182	181 180	6.9 7.6	1e+3
ICE-modified (Inosine + Ψ)	Reuse	660/1h	All reads Pass reads	180 179	176 177	7.54 7.89	2.3e+3

All parameters were extracted using pycoQC 2.5.0.19, while mapped reads were extracted using samtools flagstat.