communications biology

https://doi.org/10.1038/s42003-025-07467-4

Promoted read-through and mutation against pseudouridine-CMC by an evolved reverse transcriptase

Zhiyong He, Weiqi Qiu & Huiqing Zhou

Pseudouridine (Ψ) is an abundant RNA chemical modification that plays critical biological functions. Current Ψ detection methods are limited in identifying Ψs at base-resolution in U-rich sequence contexts, where Ψ occurs frequently. Here we report "Mut-Ψ-seq" that utilizes the classic N-cyclohexyl N′-(2-morpholinoethyl)carbodiimide (CMC) agent and an evolved reverse transcriptase ("RT-1306") for Ψ mapping at base-resolution. CMC selectively labels Ψs in RNA forming the CMC-Ψ adduct and we show that RT-1306 presents promoted read-through and mutation against the CMC-Ψ. We report a high-confidence list of Ψ sites in polyA-enriched RNAs from HEK-293T cells identified by orthogonal chemical treatments (CMC and bisulfite). The mutation signatures resolve the position of Ψ in UU-containing sequences, revealing diverse occurrence of Ψs in such sequences. This work provides methods and datasets for biological research of Ψ, and expands the toolkit for epitranscriptomic studies by combining the reverse transcriptase engineering and selective chemical labeling strategies.

Pseudouridine (Ψ), an isomer of uridine, is the first chemically modified ribonucleotide identified in RNA noted as "the fifth nucleotide" in 1957[1], and an abundant RNA chemical modification occurring in all three domains of life[2]. Endogenous Ψs are known to be installed by stand-alone or RNA-dependent pseudouridine synthases[3,4]. To form a Ψ, a uridine (U) undergoes isomerization starting with the cleavage of the glycosidic bond (N1-C1'), followed by a 180° rotation of the base around the N3-C6 axis, and the reformation of a C5-C1' linkage between the rotated uracil and the ribose (Fig. 1a). Ψ presents the same Watson-Crick-Franklin base-pairing face as U, and contains an extra H-bond donor (N1-H1) and a C–C atypical glycosidic linkage. The chemical structure of Ψ endows unique features in modulating base stacking energetics[5], stability[6], conformation[7,8], and molecular recognition[9,10] of Ψ-modified RNAs. While Ψ was initially known to occur in non-coding RNAs such as rRNAs, tRNAs, and snRNAs, growing evidence over the last decade revealed that Ψ is an abundant modification occurring in mRNAs and long non-coding RNAs that play regulatory functions in gene expression and diseases[11–15]. The occurrence of endogenous Ψs can alter in response to external stress conditions suggesting potential dynamics in Ψ regulation[12,13,16].

Accurate and precise detection of the occurrence of Ψs in the transcriptome is crucial for identifying the functional context of Ψ modification[17]. There have been rapid advances in the transcriptome-wide mapping methods of RNA chemical modifications using RNA-seq based technologies[18,19]. In order to map Ψs, "Pseudo-seq"[11], and "Ψ-seq"[12] coupled the classic "CMC" reaction[20] and RNA-seq to map Ψ at single-base resolution. Briefly, biological RNAs were treated with N-cyclohexyl N′-(2-morpholinoethyl)carbodiimide (CMC), followed by an alkaline treatment step[20]. CMC reacts with deprotonated Ψ-N1, Ψ-N3, U-N3, and G-N1 and forms corresponding CMC-adducts. The adducts Ψ-N1-CMC, U-N3-CMC, and G-N1-CMC get readily reverted back to U-N3-H3, G-N1-H1, and Ψ-N1-H1 under alkaline condition; the Ψ-N3-CMC is resistant to alkaline hydrolysis likely due to the presence of the negative charge at N1 (Fig. 1b)[11,12,20]. The resulted Ψ-N3-CMC presents a bulky and positively charged group at the canonical base-pairing interface and thus promotes stop signatures during reverse transcription (RT); RT stops were subsequently measured by next-generation sequencing and used as signatures to identify Ψ (Fig. 1c)[21]. However, RT stop signatures were subjected to high background noise due to non-random RNA fragmentation, ligation biases, RNA degradation, and stably folded RNA structures. These factors yielded a high false positive rate and low data reproducibility for Ψ identification in low-abundance RNAs, such as mRNAs and long non-coding RNAs. Moreover, RT stop signatures often fail to identify consecutive or clustered Ψs[11,22].

Recently, "RBS-Seq"[23], "BID-seq"[24], and "PRAISE"[25] were developed for Ψ mapping utilizing bisulfite/sulfite treatment to RNA, which produced bisulfite-Ψ adducts and thus RT deletion signatures at bisulfite-Ψs. Bisulfite-based methods demonstrated advanced capabilities in mapping multiple modifications and improved detection sensitivity and reproducibility of Ψ[23–25]. Yet, one caveat for using RT deletion signature is that deletion cannot accurately determine the location of Ψ in any UU sequence context[24,25].

Department of Chemistry, Merkert Chemistry Center, Boston College, Chestnut Hill, MA, USA. e-mail: huiqing.zhou@bc.edu

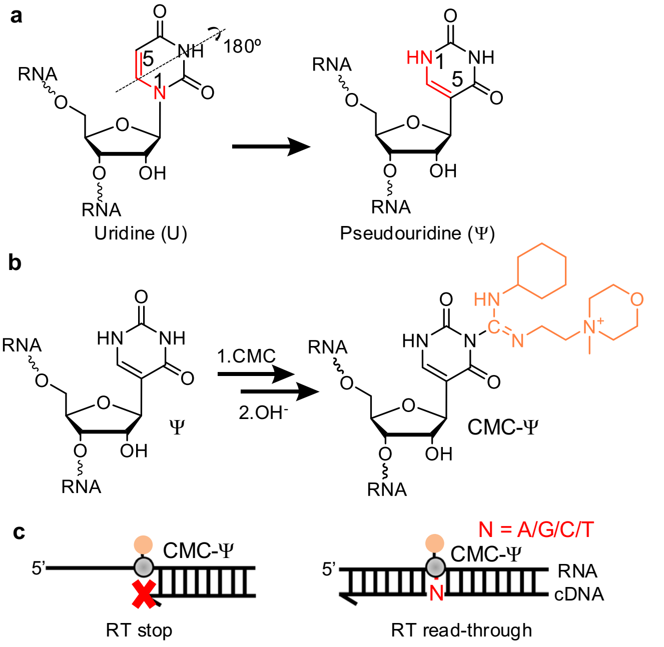


Fig. 1 | Structure and CMC-based detection of Ψ in RNA. a Chemical structures showing the formation of Ψ upon the isomerization of U. **b** Reaction scheme of CMC with Ψ yielding the N3 CMC-Ψ adduct ("CMC-Ψ"). **c** Illustrations of RT stop and read-through events against the CMC-Ψ.

Indeed, Ψ was frequently identified in UU-containing sequences in polyA-enriched RNAs from human cell lines. For instance, 1357 of the total 2209 Ψ sites (61%) reported by PRAISE occurred in UU-containing sequences in polyA-enriched RNAs from HEK-293T cells, and were identified within the UU-containing sequence range rather than at single-base resolution[25]; BID-seq reported the identification of GUUC and poly-U (five Us or more) motifs for Ψ occurrence[24]; and a recently reported Nanopore-based Ψ detection method revealed that the majority of Ψ-containing sequence motifs contained UU sequence contexts[26].

Despite the wide occurrence, the biosynthesis and function for Ψs in UU-containing sequences have yet been established. Several studies showed evidence for the installation of Ψ in the GUΨC motif in mRNAs by the human pseudouridine synthase TRUB1[22,24,25], which was known to recognize and introduce Ψ55 in the GUΨC motif in the T-loop of tRNAs[27–29]. The biosynthesis of Ψs occurring in other UU-containing sequences in polyA-enriched RNAs remain unclear. Knockdown studies show that these Ψs can be partially accounted by the RNA-guided dyskerin pseudouridine synthase 1 (DKC1); however, the guide RNA sequences have yet been reported and many Ψs sites have no identified writers[25]. In addition to biosynthesis, Ψ's occurrence in different codon positions within UU-containing codon sequences altered the translation error rate demonstrated by in vitro translation assay[30]. Resolving the detection challenges of Ψ in U-rich sequence context is critical for understanding the biosynthetic mechanism and regulatory function in gene expression of pseudouridylation in mRNAs[26].

Here we report that a recently evolved reverse transcriptase (RTase) "RT-1306"[31] shows promoted read-through efficiency and mutation rates when processing CMC-Ψ adduct in CMC-treated RNA (Fig. 1c). We developed "Mut-Ψ-seq" utilizing RT-1306 in conjunction with the classic CMC reaction, to map Ψ at single-base resolution in the transcriptome. Mut-Ψ-seq data showed excellent performance in identifying the reported Ψ sites in rRNAs[32] by mutation signatures of RT-1306 via the receiver operating characteristic (ROC) curve analysis. Ψ-identification can be affected by the CMC reaction condition. Sequencing results showed elongated CMC reaction duration (2-h) promoted the RT signatures significantly at Ψ sites compared to the 20-min reaction, which can improve the detection sensitivity for Ψ validation efforts. However, the 2-h reaction duration raised concerns about increased RNA loss and elevated background signatures on unmodified Us, which didn't out-perform the 20-min CMC reaction in terms of Ψ-identification efficiency according to the ROC curve analysis. We then used ROC-guided criteria to identify Ψs in the transcriptome and reported a high-confidence list of 44 Ψ sites in abundant mRNAs and non-coding RNAs identified by orthogonal chemical treatments: CMC for Mut-Ψ-seq and the bisulfite for PRAISE[25]. Seventy seven percentages of these sites occurred in UU-containing sequences, and the RT mutation signature resolved the position of Ψs in most UU sequence contexts. Mut-Ψ-seq has broad applications for transcriptome-wide mapping and locus-specific detection of Ψ in any sequence contexts, and for investigations of biosynthesis mechanisms and regulatory functions of Ψ.

Results and discussion

RT-1306 shows promoted read-through and mutation against CMC-Ψ in RNA oligonucleotides

We recently reported an evolved RTase "RT-1306" from the p66 subunit of human immunodeficiency virus (HIV), with six mutations (D76A R78K W229Y M230L V75F F77A). RT-1306 showed significantly promoted read-through efficiency and robust mutation signature against N^1-methyladenosine (m^1A), which blocks the canonical Watson-Crick-Franklin base-pairing[31]. We hypothesized that this feature may apply to other forms of modifications on the base-pairing interface such as CMC-Ψ. To examine the read-through propensities of RT-1306 against the CMC-Ψ adduct, we first prepared RNA oligonucleotides (Ψ-oligo1 and Ψ-oligo2, sequences shown in Supplementary Data 1) that carry a single Ψ in the sequence into the CMC-Ψ RNA through the reported CMC reaction condition[33]. RNA oligos were treated with excess amount of CMC for 16 h followed with the alkaline treatment step (Fig. 1b and **Methods**). We performed direct electrophoresis analysis of RNA product after CMC reaction, where the RNA product showed as a smeared band, consistent with addition of CMC on the Ψ, Us, and Gs. In contrast, the smeary feature disappeared after alkaline treatment, indicating the removal of the CMC group on Us and Gs (Supplementary Fig. 1)[34].

We then applied Superscript III (SSIII) RT stop assay to examine the products of the RT reaction with fluorescein amidite (FAM)-labeled primers (**Methods**). SSIII was reported to present a near-complete RT stop at the 100% CMC-Ψ site under regular RT condition (with Mg^{2+})[33]; our data showed ~90% RT stop at the CMC-Ψ site quantified by fluorescence intensity of the gel bands, indicating near-complete Ψ into CMC-Ψ conversion in both CMC treated Ψ-oligo1 and Ψ-oligo2 (Fig. 2a and Supplementary Fig. 2). We did not observe additional bands except for the residual primer, RT stop and full-length cDNAs, suggesting that there are no major remaining side products of CMC-U and CMC-G adducts on the RNA oligos after the base-treatment (Fig. 2a and Supplementary Fig. 2).

Next, we assessed the read-through propensity of RT-1306 against the two CMC treated Ψ oligonucleotides (CMC-Ψ-oligo1 and CMC-Ψ-oligo2, sequences shown in Supplementary Data 1), via the RT stop assay with RT-1306. We used the p66 subunit of wild-type HIV-RT (wtRT) as a control. The wtRT showed ~ 94% RT stop at the CMC-Ψ site, comparable to SSIII (Fig. 2a). Interestingly, RT-1306 showed 46–53% read-through efficiencies over CMC-Ψ for both oligos, ~four fold higher than the those of wtRT and SSIII (Fig. 2a and Supplementary Fig. 2). It was reported in the literature that the RT read-through efficiency against CMC-Ψ can be promoted by adding Mn^{2+}[21]; notably, RT-1306 showed improved read-through under the regular Mg^{2+}-based RT conditions, demonstrating unique advances by RTase evolution[31]. Despite that RT-1306 was engineered against m^1A, it showed extended applications onto the CMC-Ψ adduct which carries a charged and more bulky modification.

To assess whether RT-1306 produced any signature that can be deployed to identify Ψ in a read-through cDNA product, we performed colony sequencing assay to characterize the sequence of the read-through RT product by RT-1306 (**Methods** and Supplementary Fig. 3a). Two out of the total 10 sequenced colonies for CMC-Ψ-oligo1 showed T→C

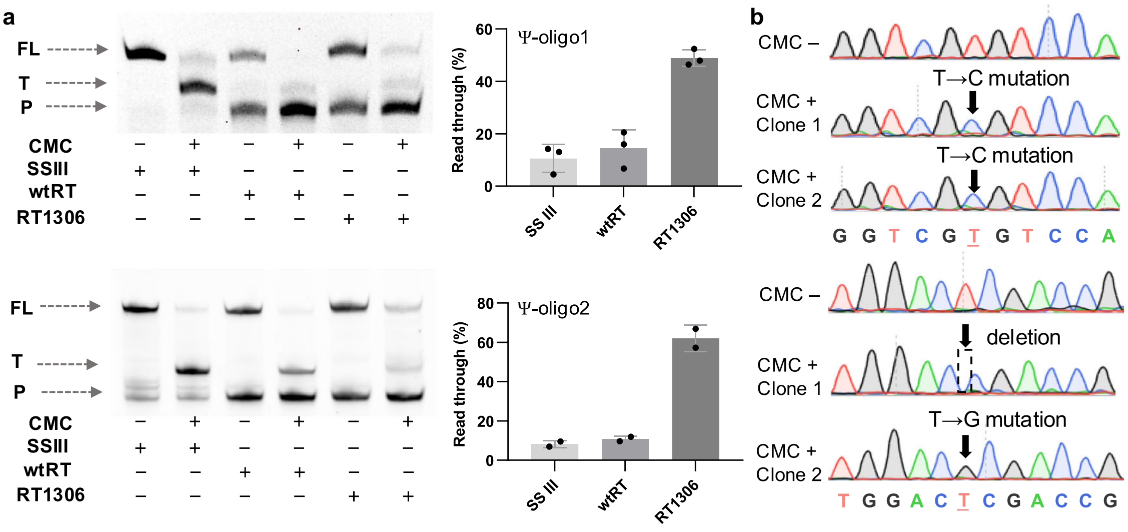


Fig. 2 | Promoted read-through efficiencies and mutations against CMC-Ψ by RT-1306. a Shown on the left are the fluorescence images of RT stop assay gels for SSIII, wtRT, and RT-1306 reading against Ψ-oligo1 and Ψ-oligo2 RNAs with and without CMC treatment. RT products were run and separated on 15% 8 M Urea-PAGE gels and imaged by the fluorescence imaging. Positions of the FAM-labeled RT primer, truncated cDNA at the Ψ site, and the full-length cDNA products are labeled with "P", "T", and "FL", respectively. The same gels were also stained by SYBR-Gold and imaged (Supplementary Fig. 2). Shown on the right are quantified RT read-through based on the RT stop assay. The RT read-through efficiency over CMC-Ψ were quantified by the ratio of the fluorescence intensity of the "T" band over the sum of intensities of the "T" and "FL" bands. Error bars represent the standard deviations of $n = 3$ replicates for Ψ-oligo1, $n = 2$ replicates for Ψ-oligo2; full gel images of 2 replicates are presented in Supplementary Fig. 2. **b** Colony sequencing data of the cDNA products from RT-1306 processing Ψ-oligo1 and Ψ-oligo2 RNAs with ("CMC+") and without ("CMC−") CMC treatments (**Methods**). Shown are the clones that carry mutations at the Ψ sites and data for all sequenced colonies are shown in Supplementary Fig. 3.

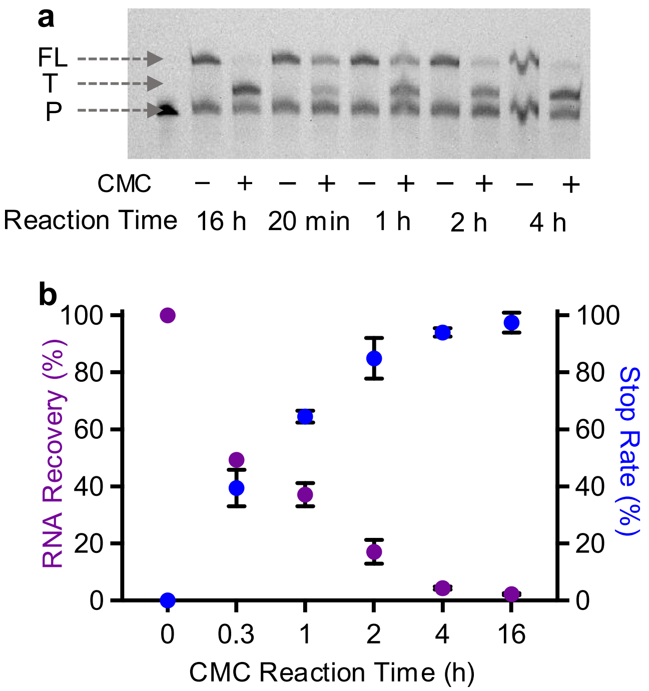


Fig. 3 | Characterization of Ψ into CMC-Ψ conversion and RNA loss by varying the duration of CMC reaction. a Shown is the increasing Ψ into CMC-Ψ conversion efficiency of Ψ-oligo1 with the elongated CMC reaction duration, measured by the RT stop assay. **b** Dependence of the Ψ into CMC-Ψ conversion of Ψ-oligo1 RNA and the RNA recovery after CMC treatment, upon the CMC reaction duration.

mutations. Among the total 7 sequenced colonies for CMC-Ψ-oligo2, one showed a T→G mutation and two showed deletions at the Ψ position (Fig. 2b, Supplementary Fig. 3b and Supplementary Data 1). This indicated that RT-1306 was capable of generating signatures to identify Ψ in RNA, which made it promising to be applied for Ψ mapping. The colony sequencing assay provided a convenient method for site-directed detection of Ψs[21,35,36].

Ψ into CMC-Ψ conversion and RNA loss by CMC treatment

Before applying the RT-1306 directly into Ψ-seq, we noticed that the 16-h CMC treatment condition[33] resulted in significant RNA loss; only 2% of the input RNA oligonucleotides were recovered after the CMC reaction and the alkaline treatment step (Fig. 3). Our data suggested the loss of RNA already occurred at the first CMC reaction step (Supplementary Fig. 4a). Indeed, the reported CMC-based Ψ-sequencing methods applied considerably shorter reaction time for the CMC reaction step: 20–30 min (Supplementary Fig. 4b)[11–13]. However, neither the conversion rate of Ψ into CMC-Ψ, nor the level of RNA loss, were reported under these conditions.

Here we measured the Ψ into CMC-Ψ conversion rates and RNA loss under various CMC reaction conditions by the SSIII RT stop assay. We used the RT stop rate as a proxy for estimating the Ψ into CMC-Ψ conversion (**Methods** and Fig. 3a)[33]. With an increasing duration of the CMC reaction from 20 min to 16 h, we observed increasing Ψ into CMC-Ψ conversion efficiency, accompanied with increasing loss of RNA (Fig. 3b). CMC treatment for 4 or 16 h achieved over 94% Ψ to CMC-Ψ conversion, however, these conditions resulted in over 95% loss of RNA. The loss of RNA can result from RNA degradation mediated by high concentration of CMC, and/or loss during the RNA purification procedure after the CMC reaction step. The 20-min CMC condition (i.e., most frequently used in Ψ-sequencing methods as listed in Supplementary Fig. 4b) recovered 50% of the input RNA oligonucleotide; however, this condition showed only 40% conversion of Ψ to CMC-Ψ measured by the RT stop assay (Fig. 3 and Supplementary Fig. 4c). Interestingly, with 2-h CMC treatment condition, the oligo RNA reaches ~85% Ψ into CMC-Ψ conversion efficiency and was able to recover a decent amount (~14%) of input RNA oligonucleotide after treatment (Fig. 3). We reasoned that although elongated reaction time may reduce RNA recovery, it can potentially promote the detection sensitivity for Ψ due to more favorable Ψ to CMC-Ψ conversion. We decided to investigate how different CMC reaction conditions impact the performance of Ψ-seqs and proceeded testing the 20 min and 2 h CMC treatment conditions by Ψ-seq.

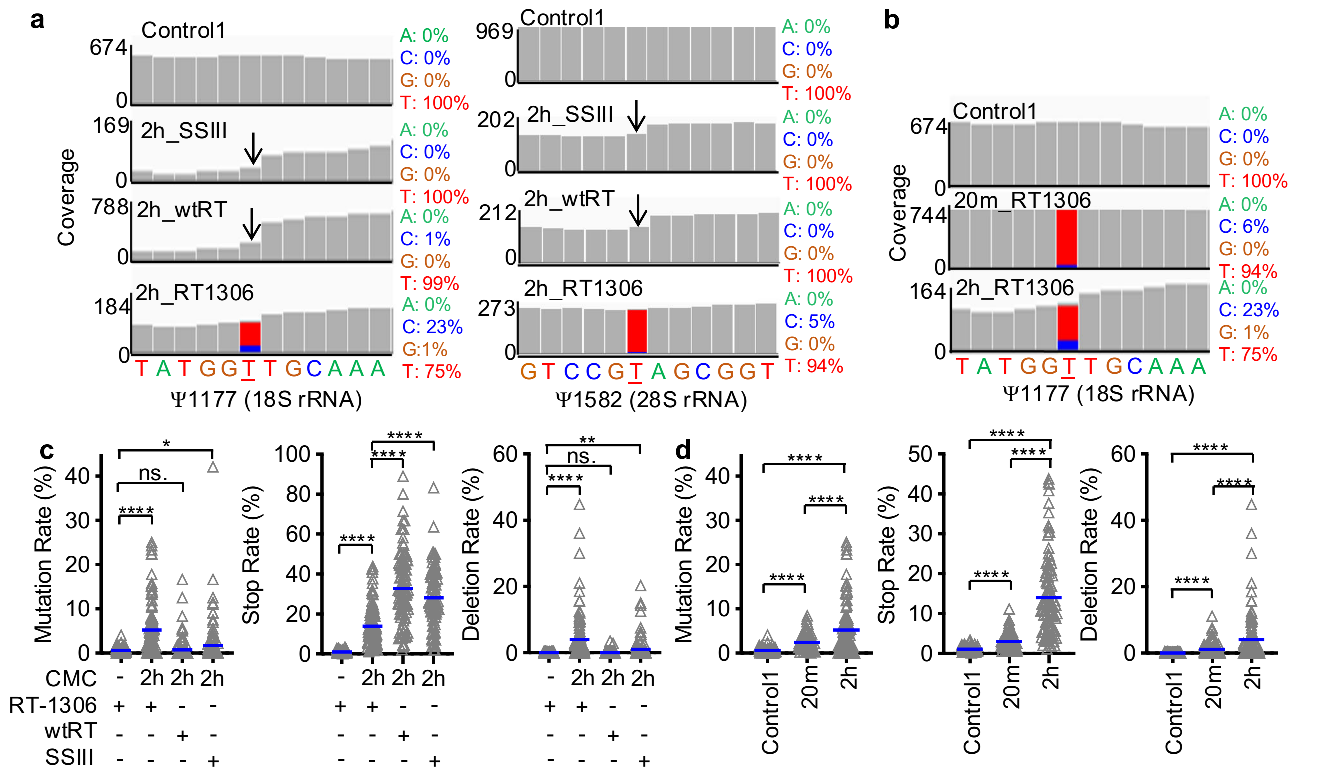


Fig. 4 | Promoted mutation signatures against CMC-Ψ in rRNAs by RT-1306 in "piloting" Ψ-seq libraries. a Representative IGV views showing RT mutations at two example Ψ sites in 18S and 28S rRNAs. **b** Ψ1177 in 18S rRNA with 20-min or 2-h CMC reaction. **c** Profiling of RT mutation, stop and deletion signatures at all reported Ψ sites in rRNAs by RT-1306, with statistical comparisons to wtRT and SSIII. **d** Profiling of RT signatures by RT-1306 at Ψ sites in rRNAs detected with 20-min and 2-h CMC reaction. *P* values were calculated by two-sided Student's *t*-test, with the significance levels noted where ns. not significant, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$.

Profiling of RT signatures via "piloting" Ψ-seq libraries

To examine RT signatures by RT-1306 under various CMC treatment conditions by Ψ-seq, we prepared six "piloting" Ψ-seq libraries from total RNA extracted from HEK-293T cells, utilizing previously reported ligation-based library preparation protocols[11,12,31]. These six "piloting" Ψ-seq libraries include libraries prepared by varying CMC reaction conditions and RTases (Supplementary Fig. 5 and **Methods**), and sequenced with ~50 K reads per library at low cost (~$50 per library). With these libraries, we primarily focused on assessing the RT signatures based on the reported Ψs within human rRNAs[32]. The prepared libraries showed 54–58% alignment to the rRNA reference genome (Supplementary Fig. 6); encouragingly, RT-1306 generated mutation signatures at previously reported Ψs in the 18S and 28S rRNAs upon manual inspection. In contrast, wtRT and SSIII produced stop signatures with no detectable mutations (Fig. 4a). Interestingly, the mutation rate at the Ψ1177 in 18S rRNA increased by ~five fold when the CMC reaction duration increased from 20 min to 2 h (Fig. 4b), which is consistent with the improved conversion of Ψ to CMC-Ψ observed on oligonucleotide RNAs (Fig. 3).

Next, we examined all 105 documented Ψ sites in human 5.8S, 18S, and 28S rRNAs (**Methods**)[32]. Strikingly, RT-1306 generated significantly promoted mutation signatures among Ψ sites with CMC treatment, showing ~nine fold increase of the averaged mutation rate comparing with that of the "Control1" library (only OH^- treatment) (Fig. 4c and Supplementary Fig. 5a). In contrast, wtRT and SSIII showed no or low extent of increase of mutation rates at Ψ sites (Fig. 4c). Additionally, we analyzed all other possible RT signatures beyond mutation rates, including RT stops, deletions, and insertions (**Methods**). The RT signature profiling revealed that RT-1306 generated complex RT signatures against CMC-Ψ adduct in rRNAs, including mutation, stop and deletion, but not insertion. In contrast, both wtRT and SSIII generated predominantly RT stop signature (Fig. 4c and Supplementary Fig. 7a). We observed significantly decreased RT stop rates against CMC-Ψs in rRNAs for RT-1306, comparing to those of wtRT and SSIII (Fig. 4c), which is consistent with the improved read-through efficiency by RT-1306 against the CMC-Ψ RNA oligos (Fig. 2a).

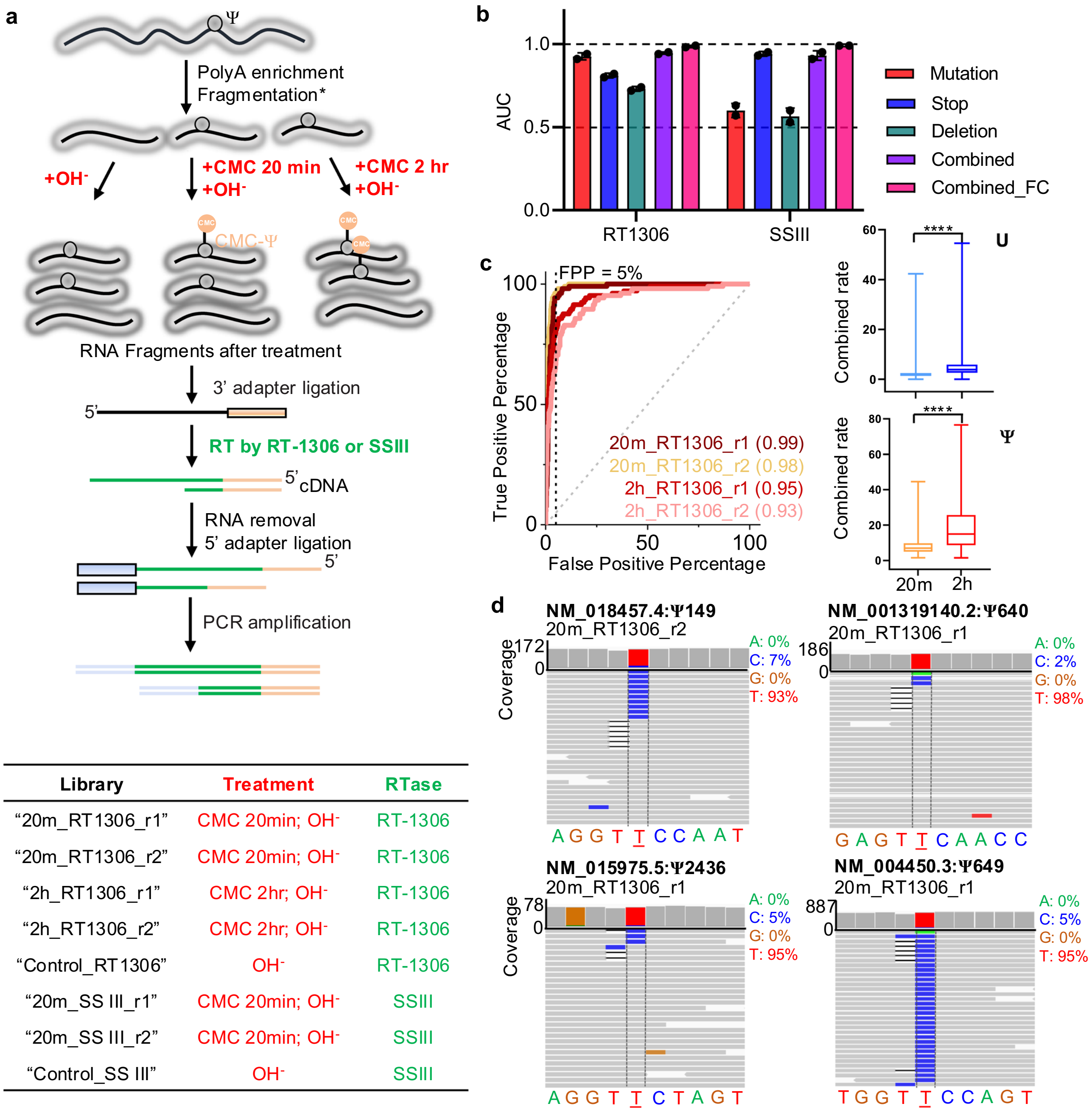
Next, we assessed the background noise level by profiling the RT signatures against 1088 unmodified U sites in 5S, 5.8S, 18S, and 28S rRNAs[37]. Not surprisingly, we found RT stops present much more prominent background noise relative to mutation, deletion and insertion signatures, only considering the non-CMC "Control1" library (Supplementary Fig. 7b). Upon 2-h CMC treatment, all three RTases (SSIII, wtRT, and RT-1306) showed significantly increased RT stops in the CMC treated libraries relative to "Control1"; among the three RTases, the background RT stop is the least significant for RT-1306 (Supplementary fig. 7b).

Interestingly, RT signatures against CMC-Ψ got promoted by 2-5-fold upon increasing the reaction duration of CMC treatment from 20 min to 2 h, suggesting increased conversion ratio from Ψ to CMC-Ψ (Fig. 4d). Importantly, the 2-h CMC treatment did not increase the background mutation and deletion against unmodified Us by RT-1306 compared to the 20-min reaction, though moderate increase was observed for the background stop signature (Supplementary Fig. 7c). In summary, the piloting Ψ-seq provided a low-cost and efficient method for validating library construction method, and profiling RT signatures on the abundant rRNAs with variable chemical treatment and RT conditions. We confirmed that mutation signature of RT-1306 against the CMC-Ψ was promising to map Ψs in biological RNAs and while the elongated reaction duration of CMC treatment can be promising to improve the detection sensitivity of Ψ, the level of background noise must be assessed accordingly.

Development of Mut-Ψ-seq and quantitative assessment of Ψ-identification efficiency by ROC curve analyses

To apply RT-1306 into Ψ identification in coding and non-coding RNAs, we developed "Mut-Ψ-seq". We treated fragmented polyA-enriched RNAs from HEK-293T cells (two biological replicates) with 20 min or 2 h CMC reaction followed by alkaline treatment. The 20 min CMC treatment condition resulted in 84% recovery of the RNA fragments evaluated by the overall mass, whereas the 2 h treatment only recovered around 32.5% of the input RNA (Supplementary Fig. 8a). The degree of RNA loss was slightly reduced for the polyA-enriched RNA fragments compared to the same treatments on short oligonucleotides (Fig. 3b), likely due to the increasing sizes of RNAs. The treated RNA was ligated with the 3′ adapter sequence and reverse transcribed by RT-1306. The cDNA product was ligated with the 5′ adapter followed by PCR amplification into NGS libraries (Fig. 5a, Supplementary Data 1, and **Methods**). To benchmark the ligation-based protocol, we prepared two libraries by 20 min CMC treatment condition and SSIII as previously reported (Fig. 5a)[13]. The resulted library samples (i.e., PCR products) were around 230 base pairs in size, except for the two libraries prepared from RNAs treated with 2-h CMC reaction which showed significantly smaller library size (Supplementary Fig. 8b). All libraries were subjected to deep sequencing with ~40 million reads per library (**Methods**).

We first assessed the Ψ-identification efficiency of the reported 105 Ψ and 1088 non-Ψ U sites in rRNAs[37] by the ROC curve analysis and calculating the area under the curve (AUC). Briefly, AUC = 1 represents the perfect Ψ-identification sensitivity and specificity (i.e., no false positive or false negative discoveries), and AUC = 0.5 represents random selection of Ψs and unmodified Us[11,12,33]. SSIII control libraries showed stop signatures at



Library	Treatment	RTase
"20m_RT1306_r1"	CMC 20min; OH-	RT-1306
"20m_RT1306_r2"	CMC 20min; OH-	RT-1306
"2h_RT1306_r1"	CMC 2hr; OH-	RT-1306
"2h_RT1306_r2"	CMC 2hr; OH-	RT-1306
"Control_RT1306"	OH-	RT-1306
"20m_SS III_r1"	CMC 20min; OH-	SSIII
"20m_SS III_r2"	CMC 20min; OH-	SSIII
"Control_SS III"	OH-	SSIII

Fig. 5 | Development and results of Mut-Ψ-seq. a Workflow and list of prepared libraries for Mut-Ψ-seq. **b** Comparison of Ψ identification efficiency by individual or combined RT signatures generated by RT-1306 or SSIII, through the AUC values derived from ROC curve analyses against reported Ψs and Us in rRNAs (Supplementary Fig. 9 and Supplementary Fig. 10). **c** Shown on the left is the ROC analyses for assessing Ψ identification efficiency of RT-1306 by 20-min or 2-h CMC reaction. Shown on the right are the distributions of observed combined rates generated by RT-1306 against reported 1088 Us (upper panel) and 105 Ψs (low panel) in rRNAs. *P* values were calculated by two-sided Student's *t*-test, with the significance levels noted where *****P* < 0.0001. **d** Representative IGV views of the RT mutation signature suggesting the occurrence Ψ in GUΨC sequence in four mRNAs.

Ψs and the RT stop signatures robustly identified rRNA Ψs with $AUC^{stop}_{SSIII} = 0.93$ or 0.95 for the two biological replicates, which benchmarked the high quality of libraries prepared via the current protocol (Fig. 5b and Supplementary Fig. 9a). Interestingly, under the same 20 min CMC treated condition, the RT mutation signature alone by RT-1306 can identify Ψ sites in rRNAs with $AUC^{mut}_{RT\text{-}1306} = 0.94$ or 0.91 for the two biological replicates, suggesting the mutation signatures by RT-1306 can be used to identify Ψ sites. In contrast, the same ROC analyses yielded $AUC^{mut}_{SSIII} = 0.63$ or 0.57 using the mutation signature generated by SSIII to identify Ψ sites, suggesting near random selection (Fig. 5b and Supplementary Fig. 9b).

Given RT-1306 generated context-dependent RT signatures against CMC-Ψs (Fig. 4c), we systematically assessed the performances of distinct RT signatures for Ψ identification via the ROC analysis. For RT-1306, mutation, stop, and deletion all showed identification power for Ψs with AUC > 0.7; among the three signatures, mutation is the most efficient signature given the highest AUC. In contrast, SSIII produced primarily the stop signature (Fig. 5b). The combined rate (i.e., mutation rate + stop rate + deletion rate) by RT-1306 showed improved identification efficiency rather than single RT signature, where $AUC^{com}_{RT\text{-}1306} = 0.95$ or 0.94 (Fig. 5b and Supplementary Fig. 9c). To rule out background RT signatures (e.g., RT stops due to RNA secondary structures), we calculated the fold change of combined rates between the CMC treated and untreated libraries. The ROC analysis revealed that the fold change gave the most robust identification power for Ψs with $AUC^{com\_fc}_{RT\text{-}1306} = 0.99$ or 0.98 (Fig. 5b and Supplementary Fig. 10). Lastly, the 2-h CMC treatment condition increased the magnitudes of the combined rates compared to the 20 min condition. However, the ROC analysis showed slightly decreased efficiency for Ψ identification, compared to the 20-min treatment, due to elevated noise level on unmodified U sites (Fig. 5c, Supplementary Fig. 9 and Supplementary Fig. 11). Therefore, we continued calling Ψs in mRNAs and non-ribosomal non-coding RNAs (ncRNAs) using the fold change of combined rates between the CMC treated and untreated libraries.

Ψ-identification in UU sequence contexts mRNAs via Mut-Ψ-seq

Mut-Ψ-seq libraries were aligned onto the hg38 RefSeq reference genome including mRNA and ncRNA genes (Supplementary Fig. 12). Us in mRNAs and ncRNAs are initially identified to be potential Ψ sites if they show (i) significant fold change of combined rates for CMC-treated library compared to the untreated library, (ii) significant combined rates above the background level, and (iii) at least 5 read counts for the sum of reads that contain mutation, stop, and deletion in both biological replicates (**Methods** and Supplementary Fig. 12). The thresholds of the fold change of combined rates are determined by the ROC curve analysis at 5% false positive discovery rate for the reported Ψs in rRNAs, where $FC^{Com} = 2.4$ or 7.8 are set as cut-offs for Ψ identification for the 20-min or 2-h CMC treated libraries, respectively (Supplementary Fig. 10 and Supplementary Fig. 12). The thresholds of background of the combined rates were derived based on the observed combined rates for the U sites in rRNAs within the same CMC-treated library; we used 5% or 9.6% as the cut-offs for the combined rates for Ψ calling for the 20-min or 2-h CMC treated libraries, respectively (Supplementary Figs. 11 and 12). Initial Ψ sites were called under such criteria for both the 20-min and 2-h CMC treated libraries (Supplementary Fig. 12 and Supplementary Data 2). We then overlap these initially identified sites with those reported by PRAISE[25] to circumvent limitations caused by side reactions by a single chemical treatment.

We report a list of 294 high-confidence Ψ-containing genes identified by orthogonal chemical treatments (i.e., CMC chemistry in this study and bisulfite treatment used by 'PRAISE'[25]) (Supplementary Fig. 13a). With this list of genes, we performed gene ontology (GO) analysis, which revealed that Ψ-modified mRNA was significantly enriched in translation-related biological processes, especially enriched in multiple ribosomal proteins mRNAs (Supplementary Fig. 13b and Supplementary Data 3). Since the expression of ribosomal proteins tend to be regulated in a concerted manner[38], it is worth pursuing whether pseudouridylation can provide an additional layer of concerted regulation for ribosome biogenesis. While these genes were identified by both Mut-Ψ-seq and PRAISE, the identified Ψs do not always overlap with single-base precision.

When insisting on single-base-level overlap, 44 high-confidence Ψ sites are robustly identified in mRNAs and non-ribosomal ncRNAs by orthogonal chemistry treatments (Table 1). Thirty four out of the 44 sites show the identified Ψ in UU-containing sequence contexts. We observe mutation signatures in 21 out of the 34 sites, which help precisely identify the location of Ψ in the UU sequence contexts (Table 1). For examples, Ψ649 was unambiguously assigned at the second U in the GUUC motif by mutation signature in the 3′-UTR of the *ERH* mRNA (*NM_004450.3*) (Fig. 5d), which was previously assigned to be a substrate site for TRUB1[22,24,25]. PRAISE reported the same location as "648–649" with ambiguity raised by the deletion signature. Similarly, we found three other mRNA sites corresponding to the TRUB1 substrate sequence GUΨC and mutation signatures robustly captured the presence of Ψ in the second U positions (Fig. 5d, Supplementary Fig. 14a and Table 1). In addition, Mut-Ψ-seq identified a previously reported Ψ250 in the stem-loop 3 of the *7SK* RNA written by the writer guided by the H/ACA box small RNA[39]; the mutation signature called the nearby 247 site is also modified by Ψ in the AUΨUG sequence, which was also reported by PRAISE within an ambiguous sequence ("246–248")[25]. Interestingly, the unmodified sequence of stem-loop 3 of *7SK* was reported to show two-state conformational ensemble in solution; U250 occurs as an unpaired internal loop residue in one state, while pairs with A228 adjacent to an asymmetric loop in the other state[40]. It would be interesting to assess whether and how the presence of Ψ at 250 and 247 modulate the structural dynamics of the stem-loop 3 and thus the regulatory function of the *7SK* RNA[40,41].

Notably, the initial assignment of Ψ sites in UU-containing sequences by the combined rates (mutation + stop + deletion) can be inaccurate by any prominent deletion rates (Fig. 5d). The alignment tool in use (Bowtie 2[42]) aligns all deletions to the first U (from the 5′ end) in any sequences containing consecutive Us. This features a major limitation in using deletion signatures to identify Ψ in 'UU' sequences. Here we show the mutation signature against the classic CMC-treated biological RNAs can resolve the most probable Ψ sites in UU-containing sequences.

Mutation signatures revealed the diverse occurrence of Ψ in the UU-containing sequence contexts. In the high-confidence list of Ψ sites, Ψ was found at the second U in all eight GUUV (V = A/C/G) sequences. We found 6 mRNAs that contain one or two Ψ sites in the CUUG sequence context including CΨUG, CUΨG, and CΨΨG. Only a subset of Us were found to be Ψ in short U-tract sequences such as CΨUUG, CΨΨUG, CΨUΨG, GUΨUU and AUΨΨU (Table 1 and Supplementary Fig. 14b). Notably, writers for Ψ in UU-containing sequences have yet been established, except TRUB1 for GUΨC[22,24,25,43]. "PRAISE" reported Ψs in a subset of UU-containing sequences can be regulated by DKC1 by Ψ mapping upon DKC1-knockdown (Table 1)[25]; however, the guide RNAs for the reported sites remain unclear. The precise locations of Ψ in these sequences are critical for the future identification of writers, especially beneficial for predicting guide RNA binding sites for RNA-dependent Ψ writing mechanisms in mRNAs and ncRNAs.

Conclusion

Here we report a pseudouridine sequencing method "Mut-Ψ-seq" with promoted mutation signature generated by the evolved RTase "RT-1306" reading through CMC-Ψ adducts in CMC-treated biological RNA from HEK-293T cells. ROC curve analysis is a powerful method in quantitatively assessing the performance of chemical treatment and RT signatures for identifying modifications. We demonstrated that the RT signatures by RT-1306 robustly identified previously documented Ψ sites in human rRNAs with AUC = 0.98. Our sequencing data showed an increased CMC reaction duration can magnify the detection sensitivity of Ψ; however, it increased the

Table 1 | High-confidence Ψ sites identified by both Mut-Ψ-seq and PRAISE

Refseq	Name	Region	Sequence motif
NM_020992.4:317	PDZ and LIM domain 1	CDS	CUUG
NM_023934.4:234	FUN14 domain containing 2	CDS	CUUUUG**
NM_000983.4:313	Ribosomal protein L22	CDS	GUG
NM_014014.5:62	Small nuclear ribonucleoprotein U5 subunit 200	5′UTR	CUUG
NM_001253384.2:631	Ribosomal protein L15	CDS	GUA
NM_000980.4:247	Ribosomal protein L18a	CDS	GUG
NM_018457.4:148	Proline rich 13	CDS	GUUC*
NM_012423.4:456	Ribosomal protein L13a	CDS	GUUG**
NM_207346.3:305	tRNA splicing endonuclease subunit 54	CDS	GUUG*
NM_213611.3:887	Solute carrier family 25 member 3	CDS	CUUUG**
NM_052844.4:399	Dynein 2 intermediate chain 2	CDS	GUG
NM_003302.3:1434	Thyroid hormone receptor interactor 6	CDS	CUG
NM_033251.2:823	Ribosomal protein L13	CDS	CUUC**
NM_002080.4:241	Glutamic-oxaloacetic transaminase 2	CDS	CUUUA
NM_004077.3:672	Citrate synthase	CDS	CUUUG
NM_017952.6:495	Pentatricopeptide repeat domain 3	CDS	AUUG
NM_001319140.2:639	Adenylate kinase 2	CDS	GUUC*
NM_012111.3:465	Activator of HSP90 ATPase activity 1	CDS	CUUG
NM_001199973.2:242	RPL36A-HNRNPH2 readthrough	CDS	CUUG
NM_015975.5:2436	TATA-box binding protein associated factor 9b	3′UTR	GUUC*
NM_001142285.2:92	Ribosomal protein S24	CDS	CUUC**
NM_006816.3:778	Lectin, mannose binding 2	CDS	CUUC**
NM_001199629.2:594	Myosin light chain 6B	CDS	CUUG**
NM_019059.5:228	Translocase of outer mitochondrial membrane 7	3′UTR	CUUUG**
NM_000990.5:179	Ribosomal protein L27a	CDS	CUUUG**
NM_001679.4:293	ATPase Na+/K+ transporting subunit beta 3	CDS	GUUUUUUA**
NM_003795.6:296	Sorting nexin 3	CDS	AUUUUC**
NM_012268.4:1692	Phospholipase D family member 3	CDS	CUG
NM_001199111.2:699	Malate dehydrogenase 1	CDS	CUUG**
NM_053275.4:748	Ribosomal protein lateral stalk subunit P0	CDS	CUUG**
NM_198552.3:779	Family with sequence similarity 89 member A	3′UTR	CUUUG
NM_004450.3:648	ERH mRNA splicing and mitosis factor	3′UTR	GUUC*
NM_001402.6:519	Eukaryotic translation elongation factor 1 alpha 1	CDS	GUUA
NM_014394.3:166	Growth hormone inducible transmembrane protein	CDS	CUA
NR_104080.1:85	U6 small nuclear 9	–	AUUCG
NR_001445.2:247	7SK nuclear ribonucleoprotein	–	AUUUG
NR_001445.2:250	7SK nuclear ribonucleoprotein	–	GUA
NR_104080.1:40	U6 small nuclear 9	–	AUA
NR_002569.2:69	Small Cajal body-specific RNA 9	–	CUUUC**
NR_003932.2:457	Ribosomal protein L13a pseudogene 20	–	GUUG

Ψs identified by the reported criteria in Mut-Ψ-seq are underlined; when Ψ is detected in UU sequence contexts, the nearby U positions with significant mutation signature which suggests the most probable position of Ψ, are colored in red.

* or ** denotes potential substrates for TruB1 or DKC1, respectively, reported by knockdown studies by PRAISE[25].

level of background noise and RNA loss, and didn't improve overall Ψ identification efficiency in Ψ-mapping efforts. Mut-Ψ-seq demonstrated great potential in identifying Ψ at base-resolution in UU-containing sequence contexts. The promoted mutation signatures generated by RT-1306 against the CMC-Ψ adduct enabled the determination of Ψ occurrence in UU-containing sequences in mRNA and non-ribosomal ncRNAs. We report high-confident lists of Ψ sites and genes identified by orthogonal mapping methods, which provide valuable insights for understanding the biogenesis and function of Ψ.

Methods

DNA and RNA oligonucleotides

DNA primers, primers with FAM labeling and RNA Ψ-oligo1 for in vitro assays and cloning were ordered from Integrated DNA Technologies, Inc. (IDT), with standard desalting. RNA Ψ-oligo2 was ordered from Keck Biotechnology Resource with standard desalting. Ligation adaptors used in piloting Ψ-seq libraries and Mut-Ψ-seq were ordered from IDT with high-performance liquid chromatography purification. Sequences are reported in Supplementary Data 1.

Expression and purification of wtRT and RT-1306

The plasmid of wtRT and RT-1306 were transformed into *Escherichia coli* BL21(DE3) respectively followed by culturing at 37 °C[31]. The expression of RT protein was induced by adding 0.5 mM of isopropyl-β-D-thiogalactoside (IPTG) into 1 L of cell culture with 80 μM kanamycin when the OD600 reached 0.6–0.8. *E. coli* cells were cultured at 37 °C for 2 h and then at 16 °C for 16 h under shaking with 220 r.p.m. Cells were harvested and resuspended in 40 mL lysis buffer (50 mM Na_2HPO_4 and NaH_2PO_4, 0.5 M NaCl, pH 7.8, dissolved half-tablet of the proteinase inhibitor cocktail, Pierce). The cells were then lysed by sonication and centrifuged at 12,000 r.p.m. for 40 min at 4 °C. Solubilized proteins in the supernatant were first purified using His60 Ni Superflow Resins (Takara Bio USA, 635660) and were eluted with 50 mM to 0.5 M gradient imidazole buffer containing 50 mM Na_2HPO_4 and NaH_2PO_4, pH = 6.0, 0.3 M NaCl and 10% glycerol. The eluted protein fractions were run through a desalting column (PD-10, GE Healthcare), the buffer was exchanged into 3 mL ion-exchange Buffer A (50 mM Bis-tris pH 7.0) with an additional 50 mM NaCl. Then, the fractions were subjected to Mono Q ion-exchange chromatography, where the protein was injected onto the column flushing with 97.5% Buffer A and 2.5% Buffer B (50 mM Bis-tris pH 7.0, 1 M NaCl) and the protein was recovered in the flow-through portion. The ion-exchange purification was found to be essential for effectively removing nuclease contamination. All purification steps were carried out at 4 °C or on ice. Fractions containing the expressed protein were combined and concentrated to 2.5 mL with a 30-kDa cut-off centrifugal filter (Millipore), run through the desalting column again, and eluted with the storage buffer (50 mM Tris-HCl, 25 mM NaCl, 1 mM EDTA, 50% glycerol, pH 7.0). Purified proteins were concentrated to 200–300 μL using a 30-kDa cut-off centrifugal filter, aliquoted, flash frozen in liquid nitrogen and stored at −80 °C.

CMC labeling of RNA oligonucleotides and RNA recovery quantification

Synthetic RNA molecules were used in this study: Ψ-oligo1 and Ψ-oligo2 (Supplementary Data 1). 10 μM RNA oligos were reacted with 0.2 M CMC (Chem-Impex) in BEU buffer (7 M urea, 4 mM EDTA, 50 mM Bicine, pH 8.5) at 37 °C for specific time (20 min, 1 h, 2 h, 4 h, or 16 h) followed by Oligo Clean & Concentrator Kits (OCC, Zymo Research, D4061) cleanup, eluted with 20 μL RNase free water. Purified RNA was then mixed with 2 volumes of sodium carbonate buffer (50 mM Na_2CO_3, 2 mM EDTA, pH 10.4) and incubated at 37 °C for 4 h to remove CMC from Us and Gs, and purified with the OCC for subsequent characterizations. The concentration of eluted RNA was quantified through A260 reading measured by Nanodrop. RNA recovery was calculated by dividing the recovered RNA amount after 2-step treatment by the starting RNA amount, molecular weight change during this process is neglected.

RT stop assay

The RT stop assays were performed in 10-μL reaction volumes containing 1× RT buffer, 4 pmol RNA substrates (with or without CMC treatment) and 4 pmol DNA primer with 5′-FAM label (Supplementary Data 1), 1 mM of each dNTP and 2 μM purified RTase. The RNA substrate and DNA primer were added first and incubate in a thermocycler at 65 °C for 4 min, then 55 °C for 2 min, 45 °C for 2 min and 37 °C for 2 min for annealing. The RT reactions were then carried out at 37 °C for 1 h followed by heating at 80 °C for 10 min to inactivate the RT (for SSIII, Invitrogen, 18080093, incubate at 25 °C for 4 min then 42 °C for 10 min, 52 °C for 40 min followed by heating at 70 °C for 10 min to inactivate the RT). Remove the RNA by adding 1 μL 1 M NaOH, and incubating in a thermocycler at 95 °C for 15 min. 5 μL of the reaction was mixed with 5 μL 2× RNA loading dye and heated to 95 °C for 5 min. Seven microliters of that mixture was then immediately loaded onto 15% denaturing 8 M Urea PAGE gel. Gel was pre-run for 30 min at 200 V, and continued running for 1 h after sample loading. The gel was then imaged by fluorescence detection on a Bio-Rad ChemiDoc System and analyzed with Image Lab software. The gel was stained with 1× SYBR Gold (Invitrogen, S11494) in TBE buffer for 30 min, then the stained gel was then imaged by SYBR Gold detection on a Bio-Rad ChemiDoc System and analyzed with Image Lab software.

Colony sequencing assay

The CMC reactions were performed as described above 10 pmol of CMC treated or untreated Ψ-oligo1 and Ψ-oligo2 RNAs were subjected to 20 μL of RT reactions containing 0.5 mM each dNTP, 20 U SUPERase·In RNase Inhibitor (Invitrogen, AM2696), 1 uM RT-1306, 75 mM KCl, 2 mM $MgCl_2$, 50 mM Tris-HCl (pH 8.3); 10 μL of the RT reaction was used as the template for PCR reactions by adding 1 μL 10 mM dNTP Solution Mix, 0.5 μL Q5 High-Fidelity DNA Polymerase (NEB, M0491S), 10 μL 5× Q5 reaction buffer, 23.5 μL water with 2.5 μL 10 μM forward and reverse primers (Supplementary Data 1). PCR reaction was performed according to the manufacturer's manual, and the products were purified by agarose gel purification. The PCR product was cloned into plasmid by Gibson Assembly and individual clone was picked and sent out for Sanger sequencing (Supplementary Fig. 3a). Ten and 7 colonies were picked for CMC treated Ψ-oligo1 and Ψ-oligo2 RNAs, respectively (Fig. 2b and Supplementary Fig. 3b).

Cell culture and total RNA extraction

HEK293T cells were maintained on 100 mm Surface Treated Tissue Culture Dishes (Fisherbrand, FB012924) in DMEM medium (Gibco, 10569010) supplemented with 10% FBS (Gibco) and 1% penicillin/streptomycin (Gibco, 15140148). The cells were maintained at 37 °C under a humidified atmosphere containing 5% CO_2. Total RNA was extracted with TRIzol (Invitrogen, 15596018) followed by isopropanol precipitation, according to the manufacturer's instructions, each plate of cells resulted ~400 μg of total RNA. The resulting total RNA was treated with DNase I (NEB, M0303L) to avoid DNA contamination.

RNA preparation for piloting Ψ-seq libraries

27.4 μg of total RNA was fragmented into ~150–200 nt fragments at 94 °C for 5 min using the magnesium RNA fragmentation buffer (NEB, E6186AVIAL), followed by purification with OCC, and eluted with 30 μL RNase free water (710 ng/μL). Into 3 μL fragmented RNA, added 10 μL water and 10 μL 10 mM EDTA, heated at 80 °C for 5 min, and immediately placed on ice. Twenty microliters denatured RNA was added to 20 μL 0.4 M CMC in BEU buffer as the "CMC+" sample; or added to 20 μL BEU buffer without CMC as the "Control" sample. The CMC reaction was carried out at 37 °C for specific time (20 min, 2 h, Supplementary Fig. 5). This step was then followed by purification with OCC eluted in 15 μL water. RNA recovered from the CMC_20 min, CMC_2h and control samples were next separately dissolved in 30 μL Na_2CO_3 buffer (50 mM sodium carbonate/sodium bicarbonate, pH 10.4, 2 mM EDTA) and incubated at 37 °C for 4 h. An additional purification with OCC was then performed to recover the RNA, eluted with 15 μL RNase free water, and the RNA concentration is around 130 ng/μL.

RNA preparation for Mut-Ψ-seq

Two biological replicates were prepared by harvesting cells from two plates at the same passage. For polyA+ RNA isolation, 75 μg of total RNA were subjected to two sequential rounds of $polyA^+$ selection for each biological replicate using oligo$(dT)_{25}$ Dynabeads (Invitrogen, 61005) according to the manufacturer's instructions. 700 ng polyA+ RNA was fragmented into ~150–200 nt fragments at 94 °C for 3 min using the magnesium RNA fragmentation buffer, followed by purification with OCC. We shorten incubation time due to the possibility that CMC would digest RNA to some degree. Then the fragmented RNA was treated with CMC for 20 min or 2 h (Fig. 5), followed by base treatment. Detailed procedure is as same as described for RNA preparation for piloting Ψ-seq libraries. Finally, the RNA eluted with 18 μL RNase free water.

Library construction

The library was prepared following our previously reported procedure with slight changes[31]. 3′-End repair and 5′-phosphorylation were conducted with T4 polynucleotide kinase (PNK) (NEB, M0201S). Sixteen microliters RNA was mixed with 2 μL 10× T4 PNK reaction buffer and 1 μL T4 PNK, 1 μL SUPERase•In RNase Inhibitor, and incubated at 37 °C for 1 h; followed by RNA Clean and Concentrator (Zymo Research, R1017) purification eluting with 10 μL RNase free water. To perform 3′-adapter ligation, 10 μL 3′-repaired and 5′-phosphorylated RNA fragments were incubated with 2 μL 10 μM RNA 3′ Adapter (5′ App-NNNNNATCACGAGATCGGAAGAGCACACGTCT-3SpC3) at 70 °C for 2 min and placed immediately on ice. Then, 2 μL 10× T4 RNA Ligase Reaction Buffer (NEB), 6 μL PEG8000 (50%), 1 μL SUPERase•In RNase Inhibitor and 1 μL T4 RNA Ligase 2 truncated KQ (NEB, M0373L) were added to the RNA-adapter mixture. The reaction was incubated at 25 °C for 2 h followed by 16 °C for 12 h. One microliter 5′-deadenylase (NEB, M0331S) was added into each ligation mixture by incubation at 30 °C for 1 h followed by adding 1 μL RecJf (NEB, M0264L) for ssDNA digestion at 37 °C for 1 h. Add 1 μL Proteinase K (Invitrogen, EO0491) 37 °C for 15 min. Bring reactions to 50 μL by adding 27 μL RNase free water and perform OCC purification, the 3′-end-ligated RNA was extracted by OCC and eluted with 12 μL RNase free water. 1 μL RT primer was added to RNA and incubated in a thermocycler at 65 °C for 4 min, then 55 °C for 2 min, 45 °C for 2 min, and 37 °C for 2 min for annealing. For RT with HIV reverse transcriptase, to this was added 5× RT buffer, 1 μL 10 mM dNTP Solution Mix, 1 μL SUPERase•In RNase Inhibitor and 2 μL 10 μM RT-1306 or wtRT. The reaction was mixed well and incubated at 37 °C for 1 h, was then heated at 80 °C for 5 min. For RT with SSIII, 4 μL 5× first strand buffer, 1 μL 10 mM dNTP Solution Mix, 1 μL 100 mM dithiothreitol, perform RT at 25 °C for 4 min then 42 °C for 10 min, 52 °C for 40 min followed by heating at 70 °C for 10 min. cDNA clean up following OCC instructions with 7 μL water. Add 0.8 μL 80 μM 5′ adaptor (5′ Phos-NNNNNAGATCGGAAGAGCACACGTCTG-3SpC3) and 1 μL DMSO into 5.5 μL cDNA sample, and mix well. Heat at 70 °C for 2 min and then chill on ice immediately. Then 2 μL 10× RNA ligation buffer (NEB), 0.2 μL 100 mM ATP, 9 μL 50% PEG8000, 1.5 μL Rnl1 (high conc) ligase (NEB, M0437M) were added the ligation mixture was incubated at 25 °C for 12 h. Add 1 μL Protease K to all reactions and incubate at 37 °C for 15 min. Add 29 μL RNase free water to bring reaction volume up to 50 μL and add 350 μL DNA binding buffer in the kit, perform DNA Clean and Concentrator (Zymo Research, D4004) using 1:7 reaction to binding buffer ratio to clean up and elute with 20 μL total volume (10 μL each, two times). Eight microliters eluted cDNA was used for each 13-cycle PCR amplification reaction, which was performed with the NEBNext Universal PCR Primer for Illumina (NEB) and indexed primers (NEB). All libraries were purified on a 3% low melting point agarose gel.

Sequencing data processing

The sequencing data (the R1 reads of the pair-end data) were subjected to deduplication by the BBMap tool "clumpify" (v.38.73)[44] with the option 'dedupe subs = 0' to remove PCR duplicates. Adaptors were then trimmed by the cutadapt tool (v.1.15)[45] while reads were filtered by quality and length with options '-a AGATCGGAAGAGCACACGTCTGAACTCCAGTC -q 20 -m 30'. Processed reads were aligned to the human rRNA genes or RefSeq reference transcriptome (GRCh38) using and Bowtie 2 (v.2.4.0)[42] with parameter --very-sensitive-local (Supplementary Fig. 9a and Supplementary Fig. 12). Read counts for individual base types, deletions, and insertions at each base position were counted by the 'bam-readcount' tool[46] in reference to the script 'bam-readcount.sh'; the output of bam-readcount were further parsed by an in-house python script "parse_R1_with_indel-r.py" or "parse_R1_mut_del.py" to calculate the following rates(Supplementary Data 4). At each U position, mutation, deletion, and insertion rates are calculated as following:

Mutation rate = (A-readcount + C-readcount + G-readcount)/total-readcount;

Deletion rate = deletion-readcount/total-readcount;

Insertion rate = insertion-readcount/total-readcount.

When the RTase stops at the Ψ-CMC adduct, the cDNA terminated significantly at the nucleotide 3'-adjacent to the Ψ[11,12]. To quantify RT stop rates at each base position (e.g., the ***i*** nucleotide position), we used "bedtools genomecov -d -3" to count the number reads of which the 3′ ends aligned at the ***i*** + 1 position (i.e., readcount^{3p}), and "bedtools genomecov -d" to count the total number of reads that aligned to the ***i*** + 1 position (readcount$^{\mathrm{total}}$). The stop rate at the position ***i*** is calculated by "Stop rate [***i***] = readcount^{3p}[***i*** + 1]/readcount$^{\mathrm{total}}$[***i*** + 1]" (Supplementary Fig. 9a, Supplementary Fig. 12a and Supplementary Data 4)[12].

Calling Ψ sites

Combined rates were calculated by the sum of stop rate, deletion rate and mutation rate. Combined rate fold change was calculated using the equation Fold-change$^{\mathrm{Com}}$ = Combined rates$^{\mathrm{CMC+}}$/combined rate$^{\mathrm{Control}}$. We detect a position as Ψ only when the following criteria were met: (i) the sum of reads aligned to the U position containing mutation, stop or deletion must be no less than 5 in the CMC treated libraries; (ii) the combined rate for the position is greater than the maximum value of the combined rates (mean+ standard deviation) of U determined by histogram analysis of combine rates in rRNA; (iii) the combined rate fold change is determined by the ROC. The combined rate fold change should be greater than the threshold value of which the false positive rate is less than 5%.

Gene ontology analysis

The Gene Ontology (GO) analysis was performed using the Metascape[47] bioinformatics database with default settings (https://metascape.org/).

Statistics and reproducibility

The "*t*-test" statistical analysis was performed using GraphPad Prism 9.5.0 (GraphPad Software, Inc.). Asterisks denote statistical significance (ns, not significant; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$). Biological replicates were defined by cells cultured from two independent plates at the same passage number. The number of replicates for in vitro experiments on oligonucleotides are defined by the number of independently performed chemical or biochemical reactions.

Reporting summary

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

Data availability

Raw and processed piloting Ψ-seq and Mut-Ψ-seq data are available at NCBI Gene Expression Omnibus, accession number GSE269406. The plasmid for bacterial expression of RT-1306 is available on Addgene with the ID 131521. The source data for Figs. 2–5 are available at the Figshare repository with doi: 10.6084/m9.figshare.28027715. All other data are available from the corresponding author on reasonable request.

Code availability

Processing scripts for piloting Ψ-seq library and Mut-Ψ-seq and the scripts descriptions are available in the Supplementary Data 4.

Received: 21 August 2024; Accepted: 6 January 2025;
Published online: 11 January 2025

References

1. Davis, F. F. & Allen, F. W. Ribonucleic acids from yeast which contain a fifth nucleotide. *J. Biol. Chem.* **227**, 907–915 (1957).
2. Cappannini, A. et al. MODOMICS: a database of RNA modifications and related information. 2023 update. *Nucleic Acids Res.* **52**, D239–D244 (2024).

3. Rintala-Dempsey, A. C. & Kothe, U. Eukaryotic stand-alone pseudouridine synthases - RNA modifying enzymes and emerging regulators of gene expression? *RNA Biol.* **14**, 1185–1196 (2017).
4. Yu, Y. T. & Meier, U. T. RNA-guided isomerization of uridine to pseudouridine–pseudouridylation. *RNA Biol.* **11**, 1483–1494 (2014).
5. Davis, D. R. Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res.* **23**, 5020–5026 (1995).
6. Kierzek, E. et al. The contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic Acids Res.* **42**, 3492–3501 (2014).
7. Jiang, J., Kharel, D. N. & Chow, C. S. Modulation of conformational changes in helix 69 mutants by pseudouridine modifications. *Biophys. Chem.* **200-201**, 48–55 (2015).
8. Newby, M. I. & Greenbaum, N. L. A conserved pseudouridine modification in eukaryotic U2 snRNA induces a change in branch-site architecture. *RNA* **7**, 833–845 (2001).
9. Davis, D. R., Veltri, C. A. & Nielsen, L. An RNA model system for investigation of pseudouridine stabilization of the codon-anticodon interaction in tRNALys, tRNAHis and tRNATyr. *J. Biomol. Struct. Dyn.* **15**, 1121–1132 (1998).
10. Jiang, J., Seo, H. & Chow, C. S. Post-transcriptional modifications modulate rRNA structure and ligand interactions. *Acc. Chem. Res.* **49**, 893–901 (2016).
11. Carlile, T. M. et al. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **515**, 143–146 (2014).
12. Schwartz, S. et al. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148–162 (2014).
13. Li, X. et al. Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.* **11**, 592–597 (2015).
14. Borchardt, E. K., Martinez, N. M. & Gilbert, W. V. Regulation and function of RNA pseudouridylation in human cells. *Annu. Rev. Genet.* **54**, 309–336 (2020).
15. Cerneckis, J., Cui, Q., He, C., Yi, C. & Shi, Y. Decoding pseudouridine: an emerging target for therapeutic development. *Trends Pharm. Sci.* **43**, 522–535 (2022).
16. Karijolich, J., Yi, C. Q. & Yu, Y. T. Transcriptome-wide dynamics of RNA pseudouridylation. *Nat. Rev. Mol. Cell Biol.* **16**, 581–585 (2015).
17. Rodell, R., Robalin, N. & Martinez, N. M. Why U matters: detection and functions of pseudouridine modifications in mRNAs. *Trends Biochem. Sci.* **49**, 12–27 (2024).
18. Zhang, M., Xiao, Y., Jiang, Z. & Yi, C. Quantifying m(6)A and Psi modifications in the transcriptome via chemical-assisted approaches. *Acc. Chem. Res.* **56**, 2980–2991 (2023).
19. Wang, Y., Zhang, X., Liu, H. & Zhou, X. Chemical methods and advanced sequencing technologies for deciphering mRNA modifications. *Chem. Soc. Rev.* **50**, 13481–13497 (2021).
20. Ho, N. W. & Gilham, P. T. Reaction of pseudouridine and inosine with N-cyclohexyl-N'-beta-(4-methylmorpholinium)ethylcarbodiimide. *Biochemistry* **10**, 3651–3657 (1971).
21. Lei, Z. & Yi, C. A radiolabeling-free, qPCR-based method for locus-specific pseudouridine detection. *Angew. Chem. Int. Ed.* **56**, 14878–14882 (2017).
22. Safra, M., Nir, R., Farouq, D., Vainberg Slutskin, I. & Schwartz, S. TRUB1 is the predominant pseudouridine synthase acting on mammalian mRNA via a predictable and conserved code. *Genome Res.* **27**, 393–406 (2017).
23. Khoddami, V. et al. Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution. *Proc. Natl. Acad. Sci. USA* **116**, 6784–6789 (2019).
24. Dai, Q. et al. Quantitative sequencing using BID-seq uncovers abundant pseudouridines in mammalian mRNA at base resolution. *Nat. Biotechnol.* **41**, 344–354 (2023).
25. Zhang, M. et al. Quantitative profiling of pseudouridylation landscape in the human transcriptome. *Nat. Chem. Biol.* **19**, 1185–1195 (2023).
26. Tavakoli, S. et al. Semi-quantitative detection of pseudouridine modifications and type I/II hypermodifications in human mRNAs using direct long-read sequencing. *Nat. Commun.* **14**, 334 (2023).
27. Zucchini, C. et al. The human TruB family of pseudouridine synthase genes, including the Dyskeratosis Congenita 1 gene and the novel member TRUB1. *Int J. Mol. Med.* **11**, 697–704 (2003).
28. Mukhopadhyay, S., Deogharia, M. & Gupta, R. Mammalian nuclear TRUB1, mitochondrial TRUB2, and cytoplasmic PUS10 produce conserved pseudouridine 55 in different sets of tRNA. *RNA* **27**, 66–79 (2021).
29. Jia, Z. et al. Human TRUB1 is a highly conserved pseudouridine synthase responsible for the formation of Psi55 in mitochondrial tRNAAsn, tRNAGln, tRNAGlu and tRNAPro. *Nucleic Acids Res.* **50**, 9368–9381 (2022).
30. Eyler, D. E. et al. Pseudouridinylation of mRNA coding sequences alters translation. *Proc. Natl. Acad. Sci. USA* **116**, 23068–23074 (2019).
31. Zhou, H. et al. Evolution of a reverse transcriptase to map N(1)-methyladenosine in human messenger RNA. *Nat. Methods* **16**, 1281–1288 (2019).
32. Taoka, M. et al. Landscape of the complete RNA chemical modifications in the human 80S ribosome. *Nucleic Acids Res.* **46**, 9289–9298 (2018).
33. Zhou, K. I. et al. Pseudouridines have context-dependent mutation and stop rates in high-throughput sequencing. *RNA Biol.* **15**, 892–900 (2018).
34. Ho, N. W. & Gilham, P. T. The reversible chemical modification of uracil, thymine, and guanine nucleotides and the modification of the action of ribonuclease on ribonucleic acid. *Biochemistry* **6**, 3632–3639 (1967).
35. Sun, M. et al. Locus-specific detection of pseudouridine with CRISPR-Cas13a. *Chem. Commun.* **60**, 4088–4091 (2024).
36. Fang, X. et al. A bisulfite-assisted and ligation-based qPCR amplification technology for locus-specific pseudouridine detection at base resolution. *Nucleic Acids Res.* **52**, e49 (2024).
37. Yamaki, Y. et al. Direct determination of pseudouridine in RNA by mass spectrometry coupled with stable isotope labeling. *Anal. Chem.* **92**, 11349–11356 (2020).
38. Petibon, C., Malik Ghulam, M., Catala, M. & Abou Elela, S. Regulation of ribosomal protein genes: an ordered anarchy. *Wiley Interdiscip. Rev. RNA* **12**, e1632 (2021).
39. Zhao, Y., Karijolich, J., Glaunsinger, B. & Zhou, Q. Pseudouridylation of 7SK snRNA promotes 7SK snRNP formation to suppress HIV-1 transcription and escape from latency. *EMBO Rep.* **17**, 1441–1451 (2016).
40. Camara, M. B., Lange, B., Yesselman, J. D. & Eichhorn, C. D. Visualizing a two-state conformational ensemble in stem-loop 3 of the transcriptional regulator 7SK RNA. *Nucleic Acids Res.* **52**, 940–952 (2024).
41. Camara, M. B., Sobeh, A. M. & Eichhorn, C. D. Progress in 7SK ribonucleoprotein structural biology. *Front. Mol. Biosci.* **10**, 1154622 (2023).
42. Langmead, B., Wilks, C., Antonescu, V. & Charles, R. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* **35**, 421–432 (2019).
43. Carlile, T. M. et al. mRNA structure determines modification by pseudouridine synthase 1. *Nat. Chem. Biol.* **15**, 966–974 (2019).
44. Bushnell, B. BBMap: a fast, accurate, splice-aware aligner. *Technical report*, Lawrence Berkeley National Lab. LBNL-7065E https://www.osti.gov/biblio/1241166 (2014).
45. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).
46. Khanna, A. et al. Bam-readcount-rapid generation of basepair-resolution sequence metrics. *J. Open Source Softw.* **7**, 3722 (2021).

47. Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).

Acknowledgements

We acknowledge the funding support for this work provided by Boston College (the startup fund to H.Z.) and the National Institute of General Medical Sciences (R35 GM150789 to H.Z.). We thank the Office of the Vice Provost for Research at Boston College for providing the Bio-Rad ChemiDoc System used for the research, and the BC Research Services for providing the Linux cluster in support of the bioinformatic analysis. We thank the laboratory of Prof. Welkin Johnson at Boston College for sharing the HEK293T cell line.

Author contributions

H.Z. and Z.H. conceived the ideas for the project and designed experiments. Z.H. performed all the experiments with RNA oligonucleotides, mammalian cell culturing, RNA extraction and purification, and library preparations of the piloting Ψ-seq and Mut-Ψ-seq. H.Z. wrote the data processing scripts for the piloting Ψ-seq and Mut-Ψ-seq. Z.H. performed sequencing data processing, the ROC analyses, Ψ identification, GO term analyses and statistical analyses for the piloting Ψ-seq and Mut-Ψ-seq. W.Q. expressed and purified the wtRT and RT-1306 used in the study. H.Z. and Z.H. wrote the manuscript with input from W.Q.; Z.H. and H.Z. prepared figures and supplementary materials.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-025-07467-4.

Correspondence and requests for materials should be addressed to Huiqing Zhou.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Michiaki Hamada and Mengtan Xing. [A peer review file is available].

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.