Quantitative profiling of pseudouridylation landscape in the human transcriptome

Meiling Zhang[1,7], Zhe Jiang ●[1,7], Yichen Ma[2,3], Wenqing Liu[4,5], Yuan Zhuang ●[1], Bo Lu[1], Kai Li[2,3], Jinying Peng[1] & Chengqi Yi ●[1,2,6] ✉

Pseudouridine (Ψ) is an abundant post-transcriptional RNA modification in ncRNA and mRNA. However, stoichiometric measurement of individual Ψ sites in human transcriptome remains unaddressed. Here we develop 'PRAISE', via selective chemical labeling of Ψ by bisulfite to induce nucleotide deletion signature during reverse transcription, to realize quantitative assessment of the Ψ landscape in the human transcriptome. Unlike traditional bisulfite treatment, our approach is based on quaternary base mapping and revealed an ~10% median modification level for 2,209 confident Ψ sites in HEK293T cells. By perturbing pseudouridine synthases, we obtained differential mRNA targets of PUS1, PUS7, TRUB1 and DKC1, with TRUB1 targets showing the highest modification stoichiometry. In addition, we quantified known and new Ψ sites in mitochondrial mRNA catalyzed by PUS1. Collectively, we provide a sensitive and convenient method to measure transcriptome-wide Ψ; we envision this quantitative approach would facilitate emerging efforts to elucidate the function and mechanism of mRNA pseudouridylation.

More than 170 different post-transcriptional RNA modifications have been identified to date[1]. Pseudouridine (Ψ), catalyzed by pseudouridine synthase (PUS), is the most abundant modification and occurs in all three domains of life[2,3]. It is widespread in ribosomal RNA (rRNA), small nuclear RNA (snRNA) and transfer RNA (tRNA)[3,4]. In messenger RNA (mRNA) of mammalian cells, it represents the second most abundant[5]. By now, Ψ has been found to impact various biological processes including translation, splicing, RNA stability, protein–RNA interaction and response to cellular environments[6,7]. Recently, its importance is further exemplified by the success of the COVID-19 mRNA vaccine, which is modified by Ψ derivatives[8].

Ψ synthases have been found to affect cellular activities and lead to pathological conditions. PUS7-mediated pseudouridylation can affect cap-dependent protein translation by altering the properties of mTOG tRNA derived fragments[9,10]. In addition, it can inhibit codon-specific translation of tRNAs in glioblastoma[11]. PUS1, PUS7 and RPUSD4 have

been found to modulate alternative pre-mRNA processing[12]. Some functions of PUS are independent of their activity, for instance, PUS10 is reported to promote pri-miRNA processing[13], and TRUB1 regulates the maturation step of miRNA let-7 (ref. 14). Mutations in Ψ synthases are associated with various human diseases, including mitochondrial myopathies and intellectual disability[6].

Our study and understanding of RNA modifications are greatly facilitated by advances in modification detection approaches. A selective chemical labeling strategy utilizing N-cyclohexyl-N′-β-(4-methylmorpholinium) ethylcarbodiimide p-tosylate (CMCT) has been developed to identify Ψ in a primer-extension assay[15]. This chemistry has also been coupled to high-throughput sequencing to enable transcriptome-wide detection of Ψ[5,16–18]. Other than CMCT, additional chemistry has been proposed to realize selective reactions to Ψ[19–21]. More recently, nanopore-based direct RNA sequencing has been developed to profile RNA modifications including Ψ[22,23].

[1]State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing, China. [2]Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China. [3]Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. [4]School of Life Sciences, Tsinghua University, Beijing, China. [5]Tsinghua-Peking Joint Center for Life Sciences, Tsinghua University, Beijing, China. [6]Department of Chemical Biology and Synthetic and Functional Biomolecules Center, College of Chemistry and Molecular Engineering, Peking University, Beijing, China. [7]These authors contributed equally: Meiling Zhang, Zhe Jiang. ✉e-mail: chengqi.yi@pku.edu.cn
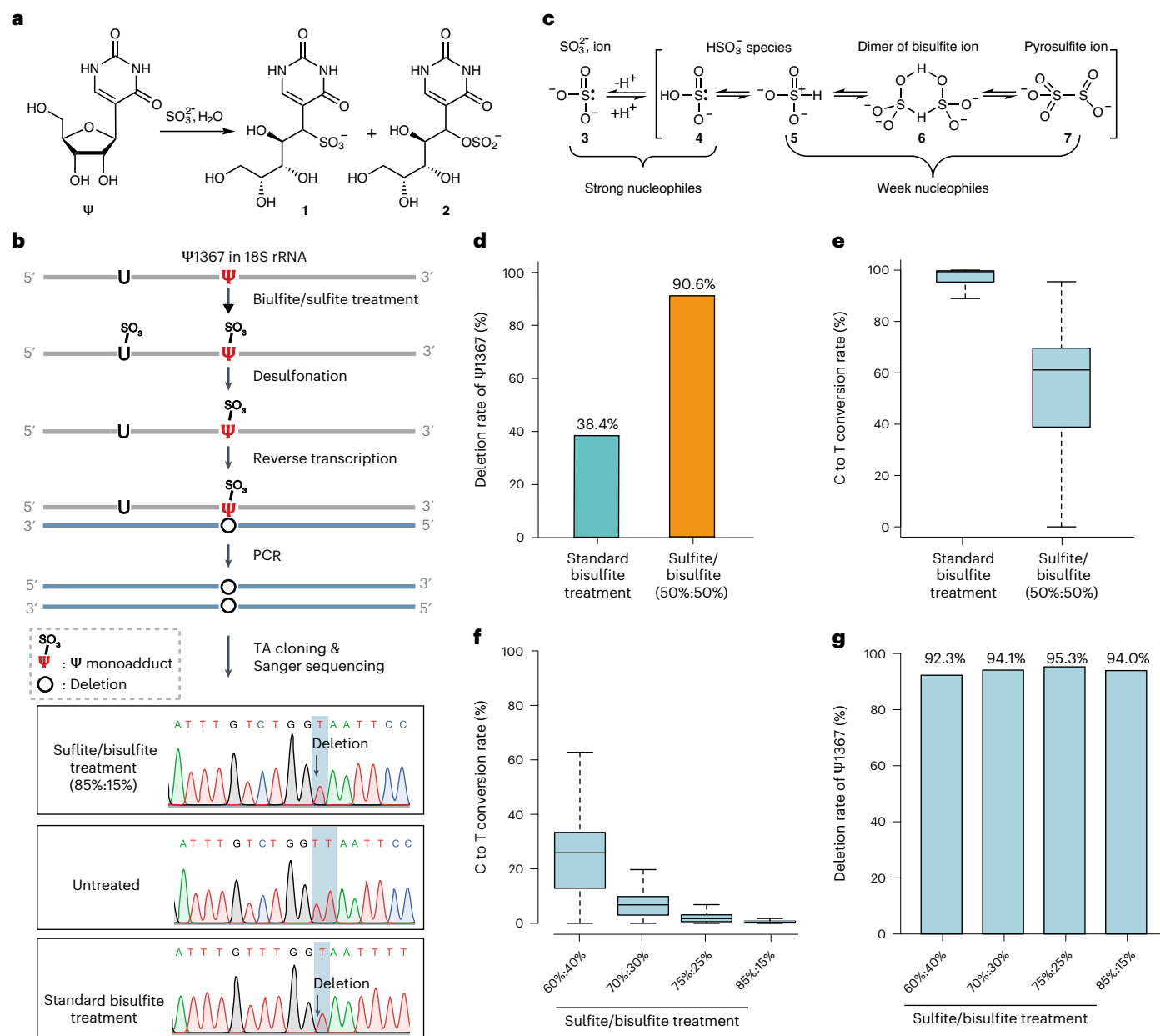
Fig. 1 | Bisulfite-Ψ adduct induces nucleotide deletion during reverse transcription. a, The reaction of bisulfite with Ψ to yield two ring-opened, monobisulfite adducts **1** (S-adduct) and **2** (O-adduct). **b**, Scheme of bisulfite reaction with Ψ. HEK293T total RNA and detection for Ψ1367. The bisulfite reaction contains two sequential steps, which are bisulfite/sulfite treatment and desulfonation. RNA was then reverse transcribed followed by targeted PCR, TA cloning and sanger sequencing of individual clones for Ψ1367 in 18S rRNA. **c**, Structure of sulfite and bisulfite ions and transformations among them. In bisulfite solution (pH = 5.1), sulfite ion **3** (SO_3^{2-}) is in equilibrium with bisulfite species (HSO_3^-) **4** and **5**, which are two tautomers. At high concentration of bisulfite (>10^{-2} M), these two tautomers interact by hydrogen bonding and convert to bisulfite ion dimer **6**, which is in turn in equilibrium with the

pyrosulfite ion **7** ($S_2O_5^{2-}$). **d**, The deletion rate of Ψ1367 site in 18S rRNA under standard bisulfite condition (cyan) and 50% sulfite/50% bisulfite condition (orange). **e**, The conversion rate of C-to-T under standard bisulfite condition ($n = 2{,}205$) and 50% sulfite/50% bisulfite condition ($n = 2{,}205$). **f**, The conversion rate of C-to-T under different proportions of sulfite and bisulfite ($n = 2{,}205$ for all conditions). The ratio of the sulfite reagent ranges from 60% to 85%, and the conversion rate substantially decreases with the increase in the sulfite ratio. In **e** and **f**, boxplot middle lines mark the median and the boundaries of the box indicate the 25th and 75th percentiles; whiskers encompass all data that are not considered outliers. **g**, The deletion rate of Ψ1367 site in 18S rRNA under different sulfite/bisulfite conditions. The ratio of sulfite reagent ranges from 60% to 85%, and the deletion rate remained stable with the increase of sulfite ratio.

Despite the progress, measurement of Ψ stoichiometry in the human transcriptome remains challenging. As a matter of fact, the identification of Ψ sites in the transcriptome is still problematic at the moment. While the four next-generation sequencing (NGS)-based methods first enabled Ψ profiling, the limited overlap between the studies suggests that their accuracy and specificity remain to be

improved[24]. Nanopore sequencing methods were designed to directly detect RNA modifications based on base-calling errors, which are bioinformatically demanding and highly sensitive to the abundance of the transcripts as well as modification levels. Hence, a quantitative, accurate and sensitive method for global Ψ detection is urgently needed.

In this study, we develop a transcriptome-wide method to quantitatively detect Ψ at single-base resolution. This method relies on bisulfite-induced deletion signature during reverse transcription, thus enabling quantitative pseudouridine assessment via bisulfite/sulfite treatment (named 'PRAISE'). In the transcriptome of HEK293T cells, we were able to identify 2,209 confident Ψ sites, showing an ~10% median modification level. We assigned differential Ψ sites to PUS1, PUS7, TRUB1 and DKC1 enzymes. Moreover, we identified known and new Ψ sites in mitochondrial mRNA, which are catalyzed by a mitochondria-localized isoform of PUS1. Collectively, our approach reveals the quantitative landscape of Ψ in the human transcriptome and provides a reliable and sensitive method for the elucidation of biogenesis and function of mRNA pseudouridylation.

Results

Bisulfite-Ψ adduct induces deletion in reverse transcription

We sought to identify a small molecule that could effectively distinguish Ψ from regular U. Interestingly, literature from more than half a century ago reported that pseudouridine can be irreversibly modified by bisulfite to yield two ring-opening monoadducts[25,26] (Fig. 1a). A later study found that ring-opened abasic site structures can block translesional synthesis and lead to deletions during DNA synthesis[27]. We thus hypothesized that Ψ in RNA could be labeled by the bisulfite reaction and subsequently generate nucleotide deletions during reverse transcription.

Because bisulfite treatment has been adopted for m⁵C detection in RNA[28,29], we applied the literature reaction (Methods) to total RNA of HEK293T cells and analyzed the signal of Ψ1367, a highly modified Ψ site in 18S rRNA. We found a 38% deletion rate for the site (Fig. 1b,d). The result is consistent with recent papers for transcriptome-wide m⁵C detection, in which Ψ is simultaneously detected as a side product[28,30]. While the data clearly show the potential of bisulfite in Ψ detection, the chemistry suffers from two fundamental limitations. First, Ψ stoichiometry is severely underestimated; we only observed a 38% deletion signal for a site that is modified to greater than 95% level. Second, the current bisulfite condition converted all cytosines to uridines. Such reduced sequence complexity is expected to lead to low mapping rates and inaccurate Ψ identification, particularly within CU-enriched regions.

To tackle the first challenge and improve the reaction efficiency, we looked into the chemical mechanism of the reaction. While bisulfite (HSO_3^-) is believed to act as the effective nucleophile, literature in the 1970s has clearly documented that within the high concentration of sodium bisulfite solution, the predominant component is bisulfite ion dimer and pyrosulfite ion ($S_2O_5^{2-}$)[31], which is inactive toward Ψ (Fig. 1c). Hence, we speculated that by increasing the proportion of sulfite and decreasing that of bisulfite, the effective nucleophile concentration can be improved. We found that when the molar proportion of sulfite was increased to 50%, the deletion rate of Ψ1367 was increased to 90% (Fig. 1d and Extended Data Fig. 1a). Among the deletion signals, 1-bp deletion accounts for 90%, and 2-bp deletion accounts for 10%. Hence, the efficiency of bisulfite-Ψ addition reaction can be greatly improved by increasing the concentration of the effective nucleophile.

We then set out to solve the C-to-U conversion problem. Under the improved condition for deletion detection, we still observed a high C-to-U conversion rate (~60%; Fig. 1e). Because cytosine deamination prefers acidic conditions while bisulfite-Ψ addition could occur under relatively basic conditions (Extended Data Fig. 1b,c). We thus further increased and tested a range of sulfite proportions. We found that the conversion rate substantially decreased with a higher proportion of sulfite (Fig. 1f). Notably, under the 85% sulfite condition, an extremely low C-to-U conversion rate (<0.5%) was observed, without compromising the deletion signal of Ψ (Fig. 1g).

Lastly, we optimized some of the key steps of the reaction, including the incubation temperature, incubation time, desulfonation time and the choice of RTases (Extended Data Fig. 1e,f). Fragment analysis suggested that the bisulfite reaction caused visible but acceptable RNA degradation (Extended Data Fig. 1g). We further systematically compared the performance of several commercially available RTases. We found that Maxima H minus, which is used in scRNA-seq[32], demonstrated excellent RT efficiency and high deletion frequency at the sites of Ψ-bisulfite adduct (Extended Data Fig. 1h). Collectively, we identified an optimized bisulfite condition that minimizes C-to-U conversion while maximizes Ψ detection based on chemical-induced deletion signature during RT.

A customized bioinformatics pipeline of PRAISE

We then coupled the labeling reaction with NGS to develop the 'PRAISE' technology for global identification and quantification of individual Ψ sites (Fig. 2a). However, there exist several challenges during NGS data analysis. First, the default setting of existing mapping software is to remove reads with gaps, especially those with multiple gaps. Thus, current mapping strategies are anticipated to cause underestimation of Ψ stoichiometry or even false negatives in Ψ identification, particularly for RNA sequences with dense Ψ sites. Another problem is that when a gap is present within the seed region during read mapping, the gap will be soft-clipped, eventually leading to the loss of genuine modification sites. To deal with these challenges, we proposed a bioinformatics pipeline composed of two key procedures. In the first step, we chose hisat2 (ref. 33) and used the 'very sensitive' parameter to prevent overfiltering of sequencing reads. In the second step, we carried out local alignment using a tailored scoring matrix and filter to obtain accurate mapping results (Methods). In case there is a gap in the seed region, we retain the gap and extend the reference sequence to fit results with deletion signals. After incorporating these changes, we were able to achieve more accurate deletion signals and hence Ψ profiling. For instance, there are four highly modified Ψ sites in the region 3665–3720 of 28S rRNA[34]. Using the default alignment parameters, we observed significant underestimation of these sites with ~30% modification level. After adopting our in-house made alignment method, we observed deletion rates of about 90% for the four Ψ sites, while the flanking bases exhibited a very low deletion rate (<1%; Fig. 2b). Thus, we believe that the customized pipeline is suitable for PRAISE-enabled Ψ identification.

Fig. 2 | PRAISE accurately measures Ψ stoichiometry in rRNA and synthetic spike-in oligos. a, Schematic illustration of PRAISE. DNase I treated RNA is first fragmented to ~150 nt, and treated with the optimized bisulfite/sulfite conditions. 'Untreated' sample is set as a negative control. After library construction, Ψ sites are identified as deletion signals during sequencing. An 8 nt UMI is added through the reverse transcription step to remove potential PCR duplications, further improving the confidence of quantitative Ψ identification. **b**, Representative IGV results for dense Ψ sites within region 3665–3720 of 28S rRNA. The NGS data after bisulfite/sulfite treatment is mapped by default (upper panel) or customized strategies (lower panel), respectively. Four highly modified Ψ sites, including Ψ3672–3674, Ψ3694, Ψ3708–3709 and Ψ3713 are shown by the dotted lines. **c**, PRAISE reliably detects Ψs in 18S rRNA. The deletion rates in one replicate are shown. **d**, PRAISE reliably detects Ψs in 28S rRNA. Many Ψ sites tend to densely locate in a short region, for instance, two regions near the 3' end of 28S rRNA. **e**, PRAISE reliably detects the 2 Ψs in 5.8S rRNA. **f**, Deletion rates between the two biological replicates, showing a high quantitative reproducibility of PRAISE. **g**, Venn diagram shows the overlap of rRNA Ψ sites within single U context in PRAISE and MS-based method[34]. **h**, PRAISE quantifies Ψ stoichiometry of five synthetic spike-in RNA. Scatterplot comparing deletion rates measured by PRAISE (y axis) with Ψ stoichiometries as expected (x axis). The Pearson correlation coefficient (R^2) and the trend line equation are denoted.

PRAISE sensitively measures Ψ stoichiometry in rRNA

We then applied the bioinformatics approach to detect Ψ modification in rRNA. In total, PRAISE identified 45, 62 and 2 Ψ sites in 18S, 28S

and 5.8S rRNA, respectively (Fig. 2c–e and Supplementary Dataset 1). PRAISE also demonstrates excellent repeatability for Ψ ratio measurement between two technical replicates ($R^2 = 0.990$; Fig. 2f). We
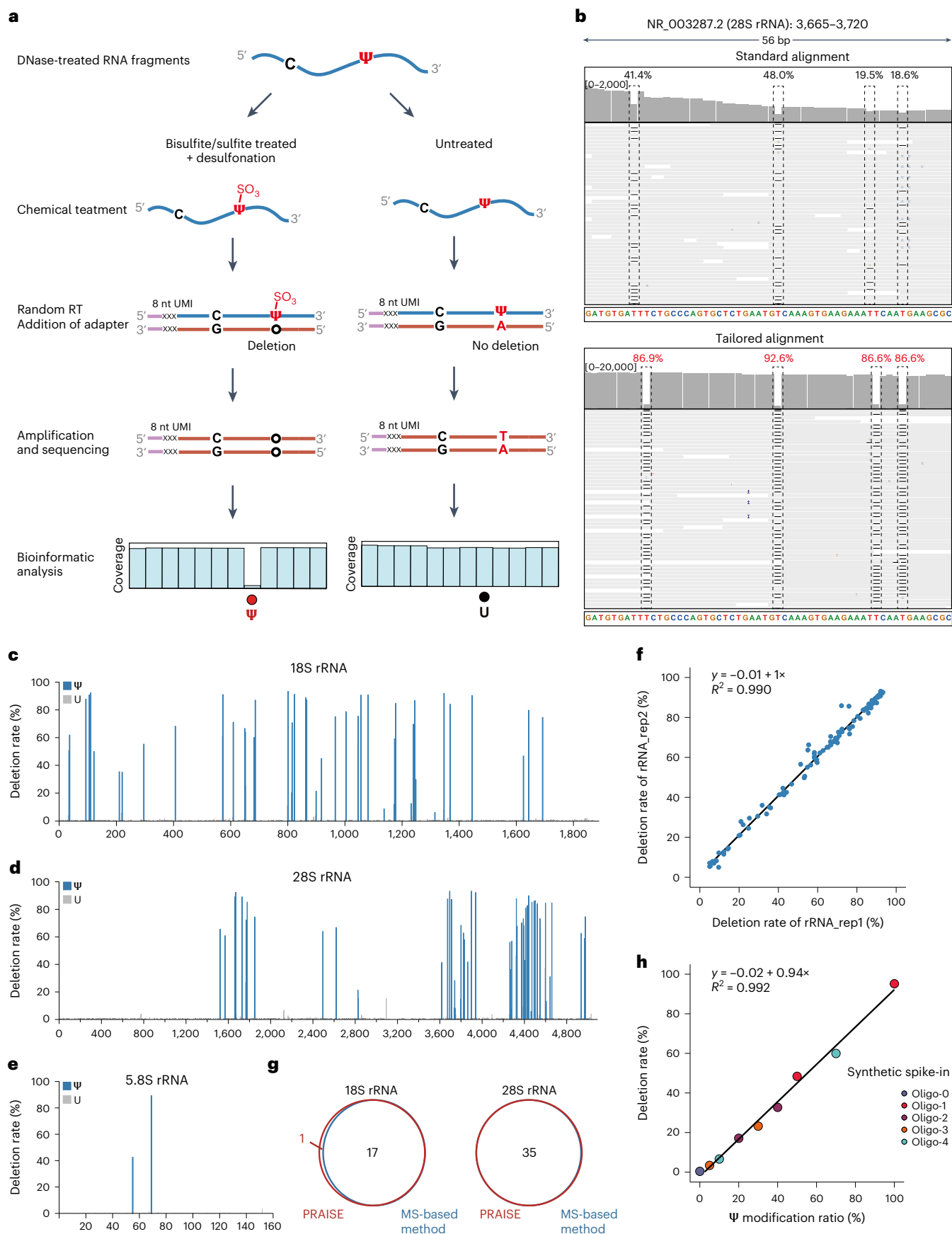
then compared the sequencing results by PRAISE with orthogonal mass spectrometry data[34]. We classified all Ψ sites into two scenarios as follows: Ψ within single U context or within consecutive Us. For both situations, we observed a high consistency between the two methods as follows: for the 17, 35 and 2 Ψ sites residing in single U context in 18S, 28S and 5.8S rRNA, they were all detected unambiguously (Fig. 2g); for the 26 and 26 known Ψ sites residing in consecutive U context in 18S and 28S rRNA, PRAISE is capable to report the modification within the consecutive U sequence. For instance, region 1847–1850 of 28S rRNA has four consecutive Us, the IGV view clearly suggests the presence of two discontinuous Ψ sites (Extended Data Fig. 2a). Besides reliably detecting the known Ψ sites (with reference to the MS and database results[35]), PRAISE also identified three Ψ sites in rRNA that were not reported by the MS-based method. Notably, our previous CeU-seq technology provided orthogonal evidence to the presence of Ψ1315 in 18S rRNA and Ψ4095 in 28S rRNA[5]. To validate the new Ψ1172 site in 18S rRNA, we knocked down the DKC1 enzyme, which is responsible for rRNA pseudouridylation (Extended Data Fig. 2b–d). Indeed, Ψ1172 showed a substantially decreased deletion rate upon DKC1 KD, as well as the other two Ψ sites (Extended Data Fig. 2e). In addition, derivatives of Ψ modification, including $m^1acp^3Ψ$ and Ψm, could also be identified at high confidence (Extended Data Fig. 2f). Moreover, we did not find any of other rRNA modification types showing a deletion rate greater than 2%, demonstrating the high specificity of PRAISE (Extended Data Fig. 2g,h).

To further assess the quantitative capability of PRAISE, we chemically synthesized five spike-in RNA oligos containing a single Ψ site, respectively, mixed the fully modified RNA with the matched, unmodified RNA to ratios ranging from 0% to 100% and measured the deletion signal in PRAISE. We achieved excellent quantitative agreement ($R^2 = 0.992$) between the expected Ψ values and the experimentally determined Ψ levels, even at the 5% modification level (Fig. 2h). Taken together, sequencing results from both rRNA and spike-in oligos showed that PRAISE is capable of quantitatively detecting Ψ sites.

A quantitative Ψ landscape in the human transcriptome

Having validated the unbiased and quantitative ability of PRAISE, we next characterized the transcriptome-wide Ψ modification sites and stoichiometry. We were able to identify 2,960 and 2,971 Ψ sites in the two technical replicates, respectively, with 2,209 shared Ψ sites (~75%), thus demonstrating high reproducibility of PRAISE (Fig. 3a and Supplementary Dataset 2). The marginal RNA degradation by our bisulfite/sulfite condition does not affect the overall quality of the sequencing libraries and Ψ calling (Extended Data Fig. 3). To evaluate the reproducibility of our method, we performed PRAISE to additional biological replicates and/or used additional commercially available library construction kits; the results among these replicates are highly consistent with each other in terms of modification sites and level (Extended Data Fig. 4). To further evaluate the reliability of our method, we employed a recently established in vitro transcribed RNA (IVT RNA) library from HEK293T transcriptome as a negative control[36]. We first confirmed

that ~80% of genes were recovered in the IVT library, and the read coverages across transcripts were similar in IVT and PRAISE libraries (Extended Data Fig. 5a,b). We could only identify 22 and 14 Ψ sites in two technical replicates of IVT RNA, respectively, with five shared Ψ sites (Extended Data Fig. 5c). Moreover, only three Ψ false-positive sites from IVT RNA were found in 2,209 Ψ sites (Fig. 3b). Therefore, the Ψ landscape obtained via PRAISE is highly reliable. As shown in a representative example, ~90% deletion rate was detected in cellular RNA while <1% deletion rate in the IVT replicates (Fig. 3c).

Among the 2,209 shared modification sites, 1,895 and 314 sites are located in mRNA and ncRNA (excluding rRNA) of HEK293T cells, respectively (Fig. 3d). The Gene Ontology (GO) analysis found highly significant enrichment of categories corresponding to metabolic process and the mitochondrial inner membrane. (Fig. 3e and Extended Data Fig. 5d). These results indicate that Ψ-containing transcripts are involved in cellular metabolism and could be important for mitochondria function.

Within mRNA, Ψs are distributed along the 5′ UTR, CDS and 3′ UTR, with an enrichment in CDS and an underrepresentation in 5′ UTR (Fig. 3f and Extended Data Fig. 5e), consistent with our previous finding[5]. We then analyzed the codon preference of Ψ sites (Extended Data Fig. 5f). We found that UUC and GUU codons, encoding phenylalanine and valine, are the most frequently modified (Fig. 3g and Extended Data Fig. 5g). Ψ is mainly present in the second and third positions; interestingly, 2 Ψ sites were detected in the AUG start codons, while 2 Ψ sites were detected in stop codons (Fig. 3d and Extended Data Fig. 5g).

Although transcriptome-wide Ψ distribution has been investigated previously[5,16–18], the absolute level of Ψ modification in mRNA has not been reported. We first analyzed the modification level of the 2,209 Ψ sites, and found that their deletion rates are highly reproducible ($R^2 = 0.946$; Fig. 3h). Globally, ~80% of the Ψ sites were modified at a low level (Ψ level <20%) in HEK293T mRNA and ~10% of the sites were highly modified (Ψ level >40%; Fig. 3i). The median modification level of mRNA Ψ sites was about 10%, which is similar to the level of m^5C in human mRNA[37].

We next analyzed the stoichiometry of Ψ sites in different sequence contexts. We found that 39% Ψ sites possess a modest 'GUG' motif and that 40% 'UU' sites prefer a 'GUUCNA' motif (Fig. 3j). The top 10 Ψ motifs are shown in Fig. 3k−'UGUAG' is the most enriched motif, resembling the preferred sequence context of PUS7, while the 'GUUCNA' motif is reminiscent of the reported targets of TRUB1 (Fig. 3k). Such motifs are consistent with previous studies by us and other labs[5,38]. Ψ sites conforming to the 'GUUCNA' motif exhibit a much higher modification level (median Ψ level > 40%) than those within the 'UGUAG' motif (5–15% median level; Fig. 3l). Thus, such distinct Ψ context and modification level suggests the differential activity of various sequence-specific modification machinery toward human mRNA[5,18].

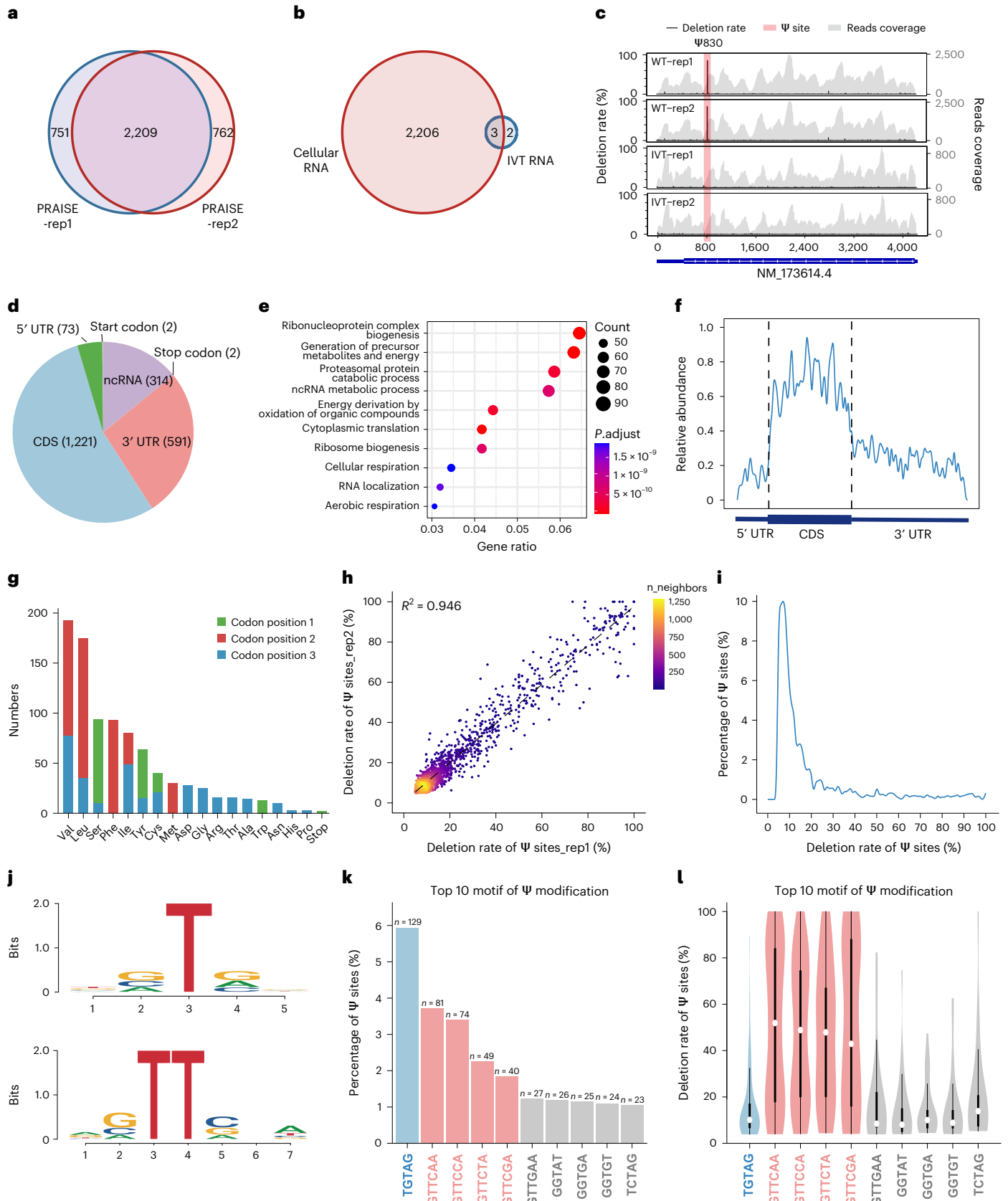## The quantitative PUS-dependent Ψ maps

We next sought to define the molecular basis for targeting the mRNA and ncRNA transcripts for pseudouridylation. Previous efforts to map

Fig. 3 | The quantitative landscape of Ψ in the human transcriptome. a, Venn diagram shows the overlap of Ψ sites in the two biological replicates. **b**, Venn diagram shows the overlap of Ψ sites in the cellular RNA library (2,209 Ψ sites) and the modification-free, IVT RNA library (5 Ψ sites). **c**, Representative views of one Ψ site on the transcript of NM_173614.4 (NOMO2). Deletion rate (left *y* axis) and read counts (right *y* axis) are both shown in PRAISE-rep1, PRAISE-rep2, IVT-rep1 and IVT-rep2 samples. Gray background color denotes reads coverage. Ψ site is indicated with a red background. **d**, Pie chart showing the proportion of Ψ sites in mRNA and ncRNA. **e**, GO enrichment analysis (BP) for the 2,209 transcriptome-wide Ψ sites in **a**. *P* values were adjusted using the Benjamini–Hochberg procedure. **f**, Metagene profiles showing the distribution of Ψ sites in human mRNA. **g**, The counts of Ψ sites in different codons as well as positions

are shown. **h**, Deletion rates of all mRNA Ψ sites between the two biological replicates, showing a high correlation. The color gradient denotes the density of overlapping modification sites. **i**, Curve graph showing the proportion of Ψ sites with different modification levels. **j**, Motif analysis of identified Ψ sites within 'U' (top panel) and 'UU' contexts (bottom panel). **k**, The proportion of top ten sequence contexts containing Ψ sites. The blue and red colors denote the TGTAG and GTTCNA motifs, respectively. The gray color denotes the remaining motifs. **l**, The deletion rate of Ψ sites within the top ten motifs (*n* numbers are shown in **k**). The middle white spot of the violin plot marks the median and the black edges indicate the 25th and 75th percentiles; whiskers encompass all data that are not considered outliers; the outer shape stands for the distribution of all data points.

Ψ sites yielded sites that could be attributed to PUS1/PUS7/TRUB1 (refs. [5],[16],[18],[38]), yet the confidence and stoichiometry information is limited due to the incomplete labeling efficiency of the CMC-based methods. To determine if PRAISE could precisely and quantitatively map PUS-dependent Ψ sites, we applied it to several HEK293T cell lines we generated previously[5],[39], with one candidate PUS enzyme being knocked out at a time. In addition, we generated a new TRUB1 knockout (KO) cell line; using quantitative MS, we found a notable decrease of Ψ
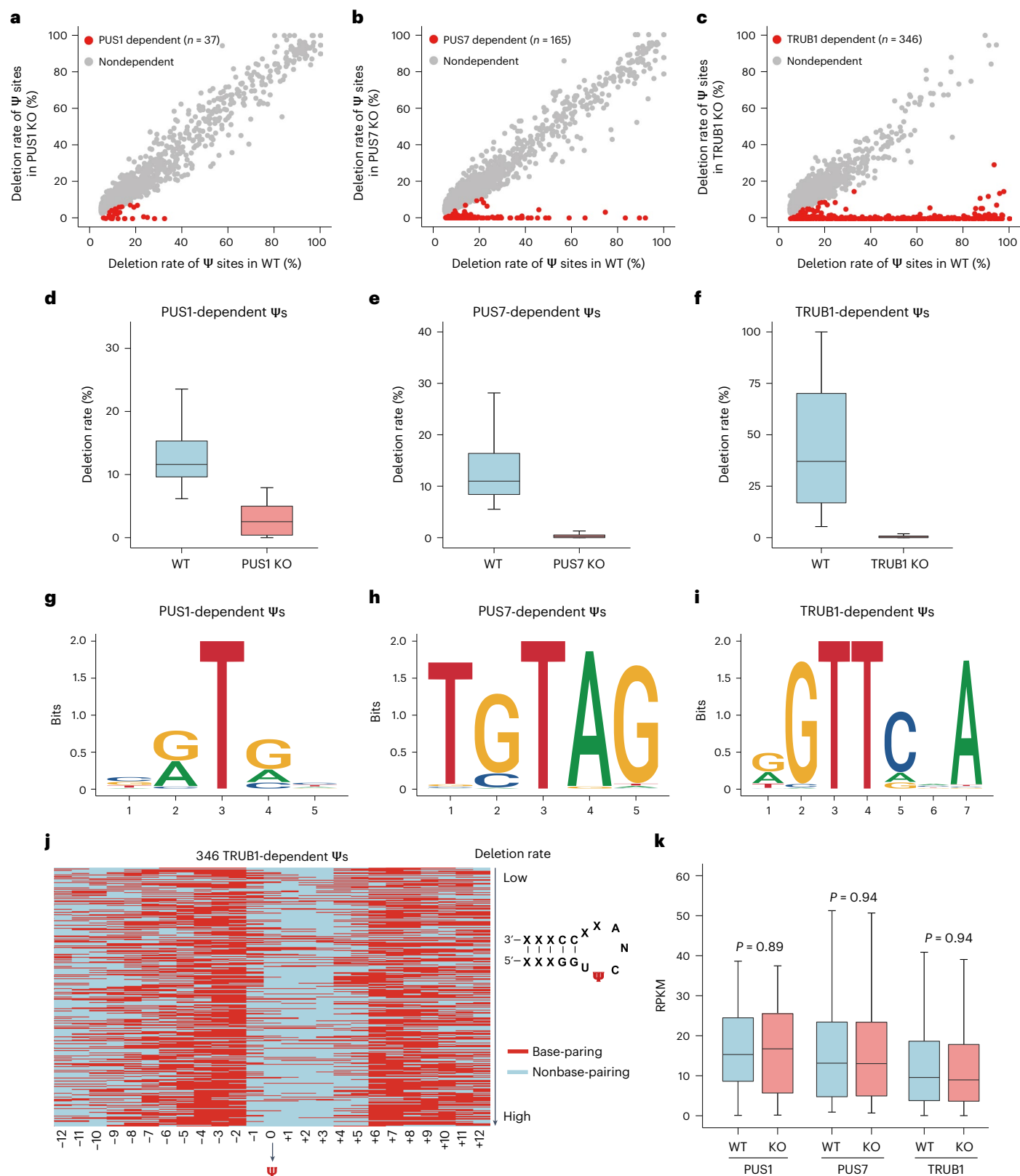
Fig. 4 | The quantitative PUS-dependent Ψ maps. a–c, Scatterplot of Ψ sites of WT versus PUS1/PUS7/TRUB1-KO cells. The red and gray dots denote the dependent Ψ sites and nondependent Ψ sites, respectively. **d–f**, Boxplot showing the deletion rate of the KO-dependent Ψ sites in WT versus PUS1/PUS7/TRUB1 KO-treated samples (*n* = 37 for PUS1-dependent Ψs, *n* = 165 for PUS7-dependent Ψs, *n* = 346 for TRUB1-dependent Ψs). **g–i**, Motif analysis of PUS1/PUS7/TRUB1-dependent Ψ sites identified by PRAISE. **j**, Heatmap depicting the secondary structure of 346 TRUB1-dependent Ψ sites. The red color denotes base pairing while the blue color denotes nonbase-paring. **k**, Boxplot showing the RPKM level of dependent Ψ transcripts in WT versus PUS1/PUS7/TRUB1 KO cells. *n* = 37 for PUS1-dependent Ψs, *n* = 161 for PUS7-dependent Ψs and *n* = 334 for TRUB1-dependent Ψs. *P* values were calculated by a two-sided Student's *t*-test. Boxplot middle lines mark the median and the boundaries of the box indicate the 25th and 75th percentiles; whiskers encompass all data that are not considered outliers.

level for the small RNA (<200 nt) population, consistent with the role of TRUB1 in modifying tRNA and thus the usefulness of the new TRUB1 KO cell line (Extended Data Fig. 6a).

We applied PRAISE to the cellular polyA⁺ RNA of three KO cell lines. Overall, we identified the 37, 165 and 346 targets for PUS1, PUS7 and TRUB1, respectively (Fig. 4a–c and Supplementary Dataset 3). The median modification level of PUS1- and PUS7-dependent Ψs are ~10%, while TRUB1-dependent Ψ sites are ~35% (Fig. 4d–f), indicating a high enzymatic activity of TRUB1 in vivo. The sequence analysis revealed that TRUB1-dependent Ψ sites are highly enriched in a 'GUUCNA' consensus sequence (Fig. 4i), while the majority of PUS7-dependent Ψ sites are highly enriched in a 'UGUAG' motif (Fig. 4h) and PUS1 targets shared a weak 'GUG' sequence motif (Fig. 4g). The secondary structure of the 346 TRUB1 targets was predicted to give rise to a typical hairpin, consisting of a 5-bp stem and a 7-nt loop, with the Ψ site being the second base in the loop (Fig. 4j). Thus, PRAISE data revealed that TRUB1 recognizes a conserved motif and hairpin structure in mRNA and tRNA, confirming a previous study[38]. On the other hand, the structural constraint was modest for the PUS1- and PUS7-dependent Ψs (Extended Data Fig. 6k,l).

Because a large amount of Ψ sites remained unassigned, we further performed DKC1 knockdown and applied PRAISE for Ψ identification. We were able to identify a total of 473 DKC-dependent Ψ sites, accounting for 21.4% (473/2209) of the total mRNA Ψ sites (Extended Data Fig. 6g and Supplementary Dataset 3). Similar to the PUS1 and PUS7 targets, we found that DKC1-dependent Ψs were also lowly modified, with a median of 10% modification ratio (Extended Data Fig. 6h). We did not find a prominent motif for DKC1-dependent Ψ sites, suggesting multiple snoRNAs could be involved in mRNA pseudouridylation (Extended Data Fig. 6i). Collectively, we now have assigned ~50% Ψs (1,021 Ψ sites) to a particular PUS enzyme, and the remaining Ψ sites are expected to be mediated by the other 9 PUS enzymes and need to be further investigated.

We further analyzed the distribution of these PUS-dependent Ψ sites. TRUB1 and DKC1-dependent sites were enriched in the CDS region, with the latter showing a stronger enrichment (Extended Data Fig. 6b,d,f,j). PUS7-dependent sites appear to be more evenly distributed along the transcript (Extended Data Fig. 6c,e). We then examined the potential role of Ψ in gene expression and found no difference in expression level between the wild-type (WT) and KO cells (Fig. 4k and Supplementary Dataset 4). This observation suggested that Ψ does not affect RNA stability.

Comparison of PRAISE and available Ψ sequencing methods

To further evaluate the reliability and performance of PRAISE, we compared our PRAISE results with available CMC-based Ψ sequencing methods[5,16,18]. In general, 108 of 339 (31.8%), 52 of 97 (53.6%) and 567 of 1,440 (39.4%) Ψ sites in Ψ-seq, Pseudo-seq and CeU-seq, respectively, overlap with that of PRAISE (Extended Data Fig. 7a–e). Because CeU-seq is the most sensitive among the CMC-based methods (due to its chemical pull-down feature), we then performed a detailed comparison of Ψ sites between PRAISE and CeU-seq. For the 873 CeU-only Ψ sites, the majority of them have about 2% deletion rates, which are below our 5% cutoff (Extended Data Fig. 7f,g). For the 1,642 PRAISE-only Ψ sites in CeU-seq data, around half of them were not detected because of low coverage while the rest have low stop signal (Extended Data Fig. 7h,i). More recently, the three CMC-based datasets were integrated into high confident Ψ sites[38]. We observed a high overlap ratio, with 62 of the 74 (80.5%) 'highest confidence' sites and 451 of the 993 (45.4%) 'high-confidence' sites detected by PRAISE (Extended Data Fig. 7j–m). Thus, while CMC-based methods could be complicated by partial labeling efficiency and RNA structure, they were able to capture a subset of the genuine Ψ sites in the transcriptome.

During the revision of this work, another method named 'BID-seq' was published, which also uses the bisulfite treatment and achieves quantitative detection of transcriptome-wide Ψ modification[40]. While there are significant differences in both the reaction and RT conditions as well as the bioinformatic approaches, the results of PRAISE and BID-seq agree with each other in general—(1) we found very consistent Ψ sites in rRNA and mRNA stop codon by two studies (Extended Data Fig. 8a–d); (2) both works identified mitochondrial mRNA Ψ sites (see our Mitochondrial Ψ landscape section below; Extended Data Fig. 8e); (3) for transcriptome-wide Ψ sites in HEK293T mRNA, 291 of 543 (53.6%) overlapped with those of PRAISE (Supplementary Fig. 1a), with a highly consistent median deletion rate of over 20% (Supplementary Fig. 1b,c). Of the 252 BID-seq-only Ψ sites, 241 of them have high coverage in PRAISE, but their deletion rate is below our cutoff (Supplementary Fig. 1d–f). As a matter of fact, the lower deletion rates of BID-seq-only Ψ sites in BID-seq hint that they are less confident than the overlapping sites. Of the 1,918 PRAISE-only Ψ sites, 570 Ψ sites show above 5% deletion rate in all three biological repeats, and another 680 Ψ sites show above 5% deletion rate in one or two datasets of BID-seq; thus these sites demonstrate modification signals in BID-seq as well (Supplementary Fig. 1g,h). Moreover, we were able to assign 43.9% of PRAISE-only Ψ sites to one of the four modification enzymes we tested (Supplementary Fig. 1i), further supporting the sensitivity and specificity of PRAISE.

We also picked four mRNA Ψ sites for validation, including two sites (IBA57 and IDI1) also detected in CMC-based datasets and two new sites (NBAS and UNC13B) by PRAISE (Extended Data Fig. 9a,b). We chose an established qPCR-based method[39] that is orthogonal to PRAISE and successfully validated all four Ψ sites (Extended Data Fig. 9c). Moreover, the results showed the melting-curve shift of IBA57 is only sensitive to PUS7 KO, suggesting it is a PUS7-dependent site. Likewise, IDI1, NBAS and UNC13B were validated to be TRUB1-dependent (Extended Data Fig. 9b,c). Thus, the above analysis and data collectively support the reliability of PRAISE.

Mitochondrial Ψ landscape

In addition to the transcriptome-wide Ψ modification in nuclear-coded transcripts, we also investigated Ψ sites in the mitochondrial transcripts. Using our polyA⁺ samples, we identified 2, 7 and 4 Ψ sites from rRNA, tRNA and mRNA in a mitochondrial heavy strand, respectively (Fig. 5a and Supplementary Dataset 5). It is worth mentioning that using a small RNA fraction (<200 nt) of HEK293T cells (Extended Data Fig. 10a), we were able to identify 42 Ψ sites in mt-tRNA, which overlaps well with the reported mt-tRNA sites obtained by MS data[41] (Extended Data Fig. 10b,c and Supplementary Dataset 6).

For mt-mRNA, the Ψ stoichiometry ranges from 10% to 26%. Ψ sites at position 6294 in MT-CO1 mRNA and position 9904–9906 in MT-CO3 mRNA are known[42]; expectedly, they demonstrated the highest modification level among mt-mRNAs, which are 26.0% and 22.1%, respectively (Fig. 5b). A new Ψ sites were identified in MT-ND4 and MT-CYB mRNA, respectively, which are Ψ11154 and Ψ14868. They bear a modification level of about 21.2% and 10.4%, respectively (Fig. 5b).

Mitochondrial Ψ enzymes

We next aimed to explore Ψ synthases responsible for pseudouridylation in mt-RNAs. It is known that one of the two PUS1 isoforms contains a mitochondrial targeting signal in its N-terminal region (Extended Data Fig. 10d) and has been shown to modify Ψ27 and Ψ28 in mt-tRNAs[43]; we thus examined whether PUS1 may have additional targets in the mitochondria. We first performed immunofluorescence staining of the two PUS1 isoforms in Hela cells, and expectedly we found that PUS1 isoform 1 was mainly located in the mitochondria while isoform 2 was abundant in the nucleus (Extended Data Fig. 10e). We then compared the mitochondrial Ψ profile between the wild type and PUS1 KO cells using polyA⁺ samples. We found that modification signals for 1, 5 and 4 Ψ sites in mt-rRNA, mt-tRNA and mt-mRNAs disappeared completely upon PUS1 knockout, suggesting that they are PUS1-dependent
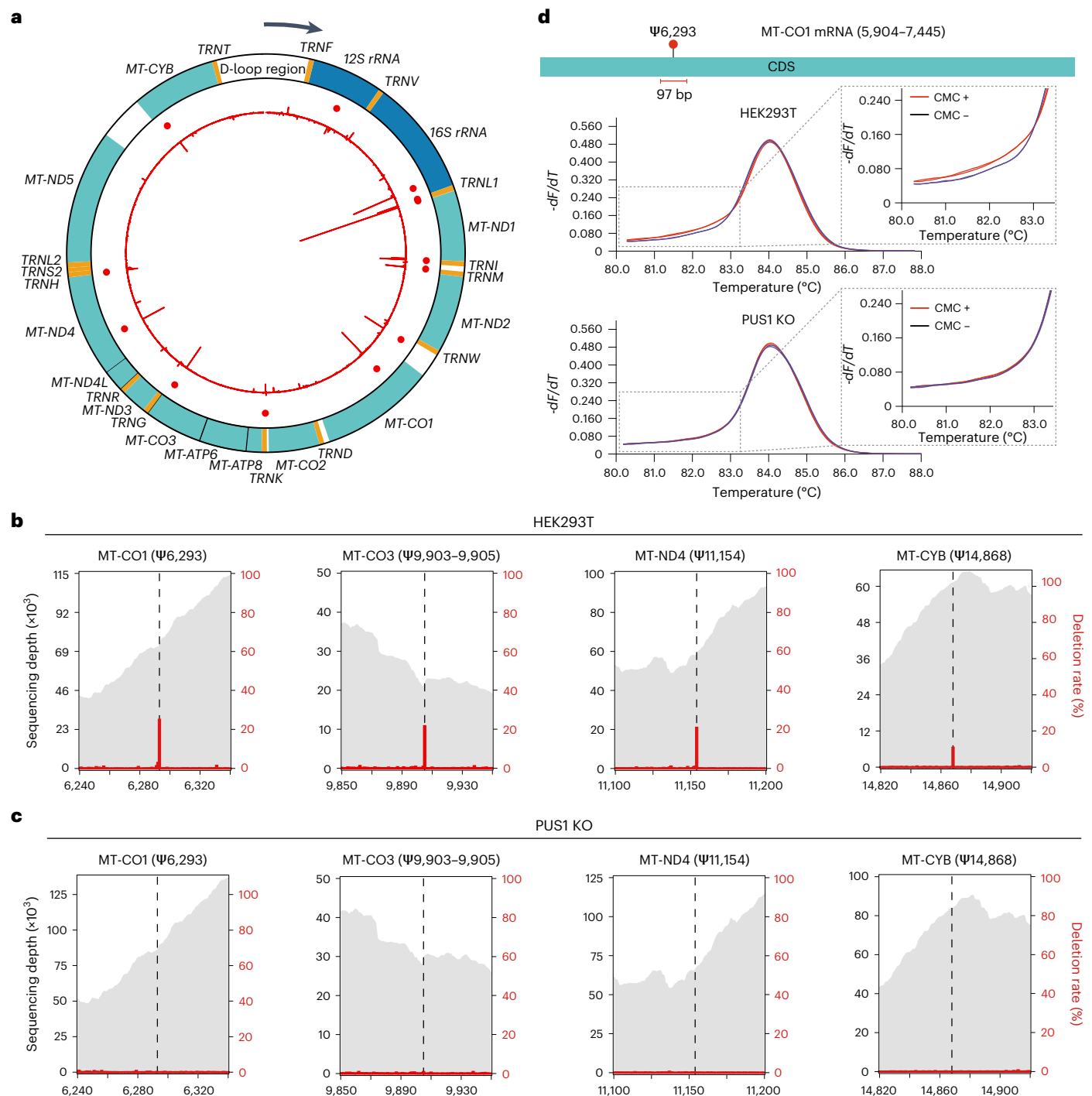
Fig. 5 | Mitochondrial Ψ landscape and Ψ enzymes. a, Ψ sites identified in the heavy strand of mitochondria. Orange, dark blue and light blue colors represent tRNA, rRNA and mRNA, respectively. The red line in the inner circle represents the difference in deletion rate (treated−untreated) of individual nucleotides, while each red dot represents an identified Ψ site. **b**, The sequencing depth of regions surrounding four mt-mRNA Ψs and the corresponding deletion rate in HEK293T (WT) treated samples are plotted. **c**, The sequencing depth of regions surrounding four mt-mRNA Ψs and the corresponding deletion rate in PUS1 KO-treated samples are plotted. **d**, High-resolution melting analysis results for position Ψ6293 in MT-CO1 mRNA, from the HEK293T (WT) and PUS1 KO cell lines.

(Fig. 5c and Extended Data Fig. 10f). We also validated the presence and PUS1-dependency of Ψ6293 in MT-CO1 mRNA[39] (Fig. 5d). Upon PUS1 knockout, the melting-curve alteration disappeared completely, which is in good agreement with our results via PRAISE (Fig. 5d).

We next sought to identify the modification enzyme for the Ψ3286 site in mt-Leu tRNA. Because it shows a 'GUUCAA' motif,

we suspected that it could be catalyzed by TRUB1. Expectedly, the deletion signal of Ψ3286 was almost absent in TRUB1 KO samples (Extended Data Fig. 10g), suggesting that Ψ occurs at position 3286 and is TRUB1-dependent. As a comparison, after the nuclear-localized PUS7 was knocked out, no significant changes in Ψ stoichiometry were observed in any mt-RNAs (Extended Data Fig. 10h).

Discussion

In this study, we developed a new base-resolution technology that enables quantitative Ψ detection. PRAISE uses Ψ-monobisulfite adducts induced deletion signal during cDNA synthesis, which allows for more accurate detection for Ψ compared to CMC-based methods. All reagents used by PRAISE are commercially available, circumventing the multiple-step chemical synthesis required for azide-derivatized CMC compound in our previous method[5]. In addition, PRAISE can be readily adopted for locus-specific and quantitative interrogation of Ψ sites[39,44]. One limitation of PRAISE is that it cannot precisely define Ψ sites within constitutive 'U' contexts. In addition, a high sequencing depth is required for transcriptome-wide quantification of Ψ sites. The detailed mechanism for the generation of deletion signal in RT remains unclear at the moment, but it is clearly influenced by the type of RTases and the condition of RT. Future elucidation of the mechanism may help further optimize the method for more biological contexts.

Epitranscriptomic technologies to detect multiple modifications at the single-molecule resolution are highly desired but remain to be established. PRAISE could be combined to map additional RNA modification types simultaneously[37,45,46]. Potential crosstalk of diverse RNA modifications may be investigated via such approaches. For instance, m6A has been found to modulate A-to-I RNA editing[47]. In addition, the various chemical/biochemical treatments developed for different modifications, including the PRAISE methodology reported herein, can be coupled to third-generation sequencing platforms[22,23], such as nanopore sequencing, to directly identify multiple modifications including Ψ at the single-molecule level.

Ψ in stop codons has been found to suppress translation termination and promote protein readthrough[48]. We found rare but existing cases in which Ψ is present in the stop codon of endogenous mRNA transcripts (two sites) in vivo, which is consistent with BID-seq. In fact, very recently we show that targeted pseudouridylation can mediate readthrough of premature termination codons in cell lines and primary cells[49]. Also, we found two examples in which the start codon is modified by Ψ; whether or not it may alter translation initiation is unclear. In addition, our analysis focused on Ψ distribution and level in mature mRNA; it is known that Ψ could be installed cotranscriptionally and function during mRNA processing, for instance, alternative pre-mRNA splicing[12]. Thus, PRAISE can be readily applied to reveal the modification level of Ψ in nascent transcripts.

Because human contains 13 PUS enzymes, it has been speculated that redundancy may exist in terms of mRNA pseudouridylation[16]. Yet, our data reveal the PUS-dependent Ψ sites do not overlap. Functionally, PUS7- and TRUB1-mediated Ψ has been implicated to stabilize transcripts[18,40]; yet, our results revealed no effect of Ψ on mRNA stability. This could be due to differential biological contexts of the studies, as Ψ affects mRNA expression levels under heat shock but not normal conditions[18], or due to different experimental approaches (transient KD versus stable KO, which could lead to compensation). Alternatively, it is tempting to speculate that Ψ could function in biological processes, for instance, in discriminating 'self' from 'nonself' RNA in nucleic acid sensing, which is an established role for inosine[50]. We hope that the transcriptome-wide, quantitative power of PRAISE could aid new discoveries of the ancient Ψ modification as well as the future development of therapeutics involving RNA modifications.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41589-023-01304-7.

References

1. Boccaletto, P. et al. MODOMICS: a database of RNA modification pathways. 2021 update. *Nucleic Acids Res.* **50**, D231–D235 (2022).
2. Davis, F. F. & Allen, F. W. Ribonucleic acids from yeast which contain a fifth nucleotide. *J. Biol. Chem.* **227**, 907–915 (1957).
3. Li, X., Ma, S. & Yi, C. Pseudouridine: the fifth RNA nucleotide with renewed interests. *Curr. Opin. Chem. Biol.* **33**, 108–116 (2016).
4. Cohn, W. E. & Volkin, E. Nucleoside 5'-phosphates from ribonucleic acid. *Nature* **167**, 483–484 (1951).
5. Li, X. et al. Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.* **11**, 592–597 (2015).
6. Cerneckis, J., Cui, Q., He, C., Yi, C. & Shi, Y. Decoding pseudouridine: an emerging target for therapeutic development. *Trends Pharmacol. Sci.* **43**, 522–535 (2022).
7. 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).
8. Nance, K. D. & Meier, J. L. Modifications in an emergency: the role of N1-methylpseudouridine in COVID-19 vaccines. *ACS Cent. Sci.* **7**, 748–756 (2021).
9. Guzzi, N. et al. Pseudouridylation of tRNA-derived fragments steers translational control in stem cells. *Cell* **173**, 1204–1216 (2018).
10. Guzzi, N. et al. Pseudouridine-modified tRNA fragments repress aberrant protein synthesis and predict leukaemic progression in myelodysplastic syndrome. *Nat. Cell Biol.* **24**, 299–306 (2022).
11. Cui, Q. et al. Targeting PUS7 suppresses tRNA pseudouridylation and glioblastoma tumorigenesis. *Nat. Cancer* **2**, 932–949 (2021).
12. Martinez, N. M. et al. Pseudouridine synthases modify human pre-mRNA co-transcriptionally and affect pre-mRNA processing. *Mol. Cell* **82**, 645–659 (2022).
13. Song, J. et al. Differential roles of human PUS10 in miRNA processing and tRNA pseudouridylation. *Nat. Chem. Biol.* **16**, 160–169 (2020).
14. Kurimoto, R. et al. The tRNA pseudouridine synthase TruB1 regulates the maturation of let-7 miRNA. *EMBO J.* **39**, e104708 (2020).
15. Bakin, A. V. & Ofengand, J. Mapping of pseudouridine residues in RNA to nucleotide resolution. *Methods Mol. Biol.* **77**, 297–309 (1998).
16. Carlile, T. M. et al. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **515**, 143–146 (2014).
17. Lovejoy, A. F., Riordan, D. P. & Brown, P. O. Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in *S. cerevisiae*. *PLoS ONE* **9**, e110799 (2014).
18. Schwartz, S. et al. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148–162 (2014).
19. Emmerechts, G., Herdewijn, P. & Rozenski, J. Pseudouridine detection improvement by derivatization with methyl vinyl sulfone and capillary HPLC-mass spectrometry. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **825**, 233–238 (2005).
20. Mengel-Jorgensen, J. & Kirpekar, F. Detection of pseudouridine and other modifications in tRNA by cyanoethylation and MALDI mass spectrometry. *Nucleic Acids Res.* **30**, e135 (2002).
21. Marchand, V. et al. HydraPsiSeq: a method for systematic and quantitative mapping of pseudouridines in RNA. *Nucleic Acids Res.* **48**, e110 (2020).
22. Huang, S. et al. Interferon inducible pseudouridine modification in human mRNA by quantitative nanopore profiling. *Genome Biol.* **22**, 330 (2021).
23. Begik, O. et al. Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. *Nat. Biotechnol.* **39**, 1278–1291 (2021).

24. Zaringhalam, M. & Papavasiliou, F. N. Pseudouridylation meets next-generation sequencing. *Methods* **107**, 63–72 (2016).

25. Singhal, R. P. Chemical probe of structure and function of transfer ribonucleic acids. *Biochemistry* **13**, 2924–2932 (1974).

26. Everett, D. W. *Part I: Reaction of Pseudouridine with Bisulfite. Part II: Reaction of Glyoxal with Guanine Derivatives: A Spectrophotometric Probe of Molecular Structure* (New York University, 1980).

27. Shibutani, S., Takeshita, M. & Grollman, A. P. Translesional synthesis on DNA templates containing a single abasic site. A mechanistic study "A rule". *J. Biol. Chem.* **272**, 13916–13922 (1997).

28. Yang, X. et al. 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m(5)C reader. *Cell Res.* **27**, 606–625 (2017).

29. Gu, W., Hurto, R. L., Hopper, A. K., Grayhack, E. J. & Phizicky, E. M. Depletion of *Saccharomyces cerevisiae* tRNA(His) guanylyltransferase Thg1p leads to uncharged tRNAHis with additional m(5)C. *Mol. Cell. Biol.* **25**, 8191–8201 (2005).

30. 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).

31. Maylor, R., Gill, J. & Goodall, D. Tetra-n-alkylammonium bisulphites: a new example of the existence of the bisulphite ion in solid compounds. *J. Chem. Soc. Dalton Trans.* **18**, 2001–2003 (1972).

32. Hagemann-Jensen, M. et al. Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nat. Biotechnol.* **38**, 708 (2020).

33. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).

34. Taoka, M. et al. Landscape of the complete RNA chemical modifications in the human 80S ribosome. *Nucleic Acids Res.* **46**, 9289–9298 (2018).

35. Lestrade, L. & Weber, M. J. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res.* **34**, D158–D162 (2006).

36. Zhang, Z. et al. Systematic calibration of epitranscriptomic maps using a synthetic modification-free RNA library. *Nat. Methods* **18**, 1213–1222 (2021).

37. Huang, T., Chen, W., Liu, J., Gu, N. & Zhang, R. Genome-wide identification of mRNA 5-methylcytosine in mammals. *Nat. Struct. Mol. Biol.* **26**, 380–388 (2019).

38. 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).

39. Lei, Z. & Yi, C. A radiolabeling-free, qPCR-based method for locus-specific pseudouridine detection. *Angew. Chem. Int. Ed. Engl.* **56**, 14878–14882 (2017).

40. Dai, Q. et al. Quantitative sequencing using BID-seq uncovers abundant pseudouridines in mammalian mRNA at base resolution. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-022-01505-w (2022).

41. Suzuki, T. et al. Complete chemical structures of human mitochondrial tRNAs. *Nat. Commun.* **11**, 4269 (2020).

42. Antonicka, H. et al. A pseudouridine synthase module is essential for mitochondrial protein synthesis and cell viability. *EMBO Rep.* **18**, 28–38 (2017).

43. Casas, K. A. & Fischel-Ghodsian, N. Mitochondrial myopathy and sideroblastic anemia. *Am. J. Med. Genet. A* **125A**, 201–204 (2004).

44. Xiao, Y. et al. An elongation- and ligation-based qPCR amplification method for the radiolabeling-free detection of locus-specific N(6)-methyladenosine modification. *Angew. Chem. Int. Ed. Engl.* **57**, 15995–16000 (2018).

45. Li, X. et al. Base-resolution mapping reveals distinct m(1)A methylome in nuclear- and mitochondrial-encoded transcripts. *Mol. Cell* **68**, 993–1005 (2017).

46. Suzuki, T., Ueda, H., Okada, S. & Sakurai, M. Transcriptome-wide identification of adenosine-to-inosine editing using the ICE-seq method. *Nat. Protoc.* **10**, 715–732 (2015).

47. Xiang, J. F. et al. N(6)-methyladenosines modulate A-to-I RNA editing. *Mol. Cell* **69**, 126–135 (2018).

48. Karijolich, J. & Yu, Y. T. Converting nonsense codons into sense codons by targeted pseudouridylation. *Nature* **474**, 395–398 (2011).

49. Song, J. et al. CRISPR-free, programmable RNA pseudouridylation to suppress premature termination codons. *Mol. Cell* **83**, 139–155 (2023).

50. Schlee, M. & Hartmann, G. Discriminating self from non-self in nucleic acid sensing. *Nat. Rev. Immunol.* **16**, 566–580 (2016).

Methods

Cell culture

HEK293T cells were cultured in DMEM medium (Corning) supplemented with 10% (vol/vol) FBS (Gibco), 1% GlutaMAX (Gibco) and 0.5% penicillin–streptomycin (Gibco) at 37 °C with 5% CO_2. Cells within passages 3–6 were used for experiments.

Generation of stable knockdown or CRISPR knockout cell lines

The shRNA targeting DKC1 (TRCN0000010325) was cloned into pLKO.1 vector. A scrambled shRNA was used as the mock control. Cell transfection was performed according to Broad Institute. The knockdown efficiency was verified by qPCR and western blot. Sequences of qPCR primers of DKC1 and GAPDH were listed in Supplementary Table 1. The DKC1 and GAPDH proteins were detected using the antibodies anti-DKC1 (Santa Cruz, sc-373956; 1:1,000) and anti-GAPDH (CWBIO, CW0100; 1:2,000).

PUS1 KO HEK293T cells and PUS7 KO HEK293T cells were from the literature[5,39]. TRUB1 KO HEK293T cells were generated by CRISPR–Cas9 technology[51]. sgRNA sequences of TRUB1 were listed as follows: TRUB1-sgRNA−TTCGGATCCGGTCCTGGCCG.

mRNA purification

Total RNA was extracted with TRIzol (Invitrogen, 15596018) followed by isopropanol precipitation, according to the manufacturer's instructions. The resulting total RNA was treated with DNase I (NEB, M0303L) to avoid DNA contamination. For polyA+ RNA isolation, size selection was first performed using MEGAclear transcription clean-up kit (Ambion, AM1908) to deplete small RNA, and RNA was subsequently purified with two sequential rounds of polyA tail purification using oligo(dT)$_{25}$ Dynabeads (Invitrogen, 61005).

Quantification of Ψ level by LC–MS/MS

Isolated mRNA of 200 ng was digested into single nucleosides by 0.5 U nuclease P1 (Sigma-Aldrich, N8630) in 20 µl buffer containing 10 mM ammonium acetate, pH 5.3 at 42 °C for 6 h, then mixed with 2.5 µl 0.5 M MES buffer, pH 6.5 and 0.5 U shrimp alkaline phosphatase (NEB, M0371S), in a final reaction volume of 25 µl adjusted with water, and incubated at 37 °C overnight. The nucleosides were separated by ultraperformance liquid chromatography with a ZORBAX SB-Aq column (Agilent, 827975-914), and then detected by a triple-quadrupole mass spectrometer (SCIEX, QTRAP 6500). A multiple reaction monitoring mode was adopted−m/z 245.0 to 179.1 for Ψ, m/z 245.0 to 113.1 for U. A 5 µl of the solution was injected into LC–MS/MS. Standard curves were generated by running a concentration series of pure commercial nucleosides (Berry & Associates). Concentrations of nucleosides in RNA samples were calibrated by standard curves.

Synthesis of spike-in RNA oligos

Five pairs of 150 nt synthetic RNA oligos containing either Ψ or U were used to examine the deletion signal of Ψ in a quantitative manner. Short Ψ/U-oligos (70 nt) and doner RNA oligos (80 nt) for subsequent ligation were synthesized by Shanghai Primerna Biotechnology (Shanghai, China). The 5′-end of the doner RNA was phosphorylated and 3′-end of the doner RNA was blocked by a ddC (2′,3′-deoxyCytosine) to avoid self-ligation. Splint ligation was performed according to the literature[52]. Briefly, a mixture of short Ψ/U-oligos, doner RNA and the complementary splint DNA strand were annealed at a molar ratio of 1:2:1.5 in T4 DNA ligase buffer (NEB, B0202S) by incubation for 3 min at 65 °C, followed by 5 min at 25 °C. Next, T4 RNA ligase 2 (NEB, M0239L) and RiboLock RNase inhibitor (Thermo Fisher Scientific, EO0384) were added to the annealed mixtures and incubated at 37 °C for 1 h. Then the splint Ψ/U-oligos were gel purified, and selected regions of the gel corresponding to 150 nt were excised and recovered. Mixed Ψ-oligos with paired U-oligos at indicated ratios−100%, 70%, 50%, 40%, 30%, 20%, 10%, 5% and 0%. Four hundred picogram spike-in oligos was mixed with

500 ng total RNA, and the Ψ stoichiometries were detected by PRAISE. Splinted RNA oligos sequences were listed in Supplementary Table 2.

Locus detection of Ψ1367 in 18S rRNA for bisulfite/sulfite conditions screening

RNA preparation. Five hundred nanogram total RNA for one sample was treated with DNase I at 37 °C for 30 min, then RNA was fragmented to ~150 nt by magnesium RNA fragmentation buffer (New England Biolabs, E6150S) for 4 min at 94 °C, followed by chilling on ice. The reaction was stopped by RNA fragmentation stop solution and purified by ethanol precipitation. After ethanol precipitation, RNA was dissolved in 5 µl nuclease-free water.

Bisulfite/sulfite treatment. Standard bisulfite treatment referred to the conditions of literatures[28,29], the bisulfite solution was freshly prepared by dissolving 4.05 g of sodium bisulfite (Sigma-Aldrich, 243973) in 5.5 ml of RNase-free water, adjusting the pH to 5.1 with 10 M sodium hydroxide, and adjusting the volume to 10 ml with water. The 100 mM hydroquinone was prepared freshly by adding 11.01 mg of hydroquinone (Sigma-Aldrich, H9003) to 1 ml of RNase-free water. For standard bisulfite treatment, 5 µl of RNA fragments was dissolved in 50 µl bisulfite solution, which is a 100:1 mixture of bisulfite solution and 100 mM hydroquinone, and subjected to heat incubation at 50 °C for 16 h.

For our improved bisulfite/sulfite treatment conditions, we weighed and mixed potassium sulfite and sodium bisulfite in the proportions. The 100 mM hydroquinone was prepared freshly by adding 11.01 mg of hydroquinone (Sigma-Aldrich, H9003) to 1 ml of RNase-free water. For bisulfite/sulfite treatment, 5 µl of RNA fragments was dissolved in 50 µl bisulfite/sulfite solution, which is a 100:1 mixture of bisulfite/sulfite solution and 100 mM hydroquinone, and subjected to heat incubation at 70 °C for 5 h. The reaction mixture was desalted by passing it through Micro Biospin 6 chromatography columns twice (Bio-Rad, 7326200).

Desulfonation. Desalted RNA was transferred to a new 1.5 ml nuclease-free tube and adjusted to 100 µl with RNase-free water, then incubated with an equal volume of 1 M Tris−HCl (pH 9.0) at 75 °C for 30 min. The reaction was then immediately stopped by chilling on ice and ethanol precipitation.

RT-PCR. Bisulfite/sulfite-treated RNA was reverse transcribed into cDNA using random hexamers (Thermo Fisher Scientific, SO142) with Maxima H minus Reverse Transcriptase (Thermo Fisher Scientific, EP0753) according to the manufacturer's instructions. Briefly, after ethanol precipitation, the RNA was resuspended in 10 µl of RNase-free water, with the addition of 1 µl 100 µM RT primer. RNA-primer mix was denatured at 80 °C for 2 min followed by chilling on ice immediately. Further, 4 µl 5× first-strand buffer, 1 µl 10 mM dNTPs, 2 µl 0.1 M DTT, 1 µl 40 U µl^{-1} RNase Inhibitor and 1 µl Maxima H minus were added into the denatured RNA-primer mix. Reverse transcription was performed by incubating at 25 °C for 5 min, 42 °C for 3 h and heat-inactivated at 70 °C for 15 min. Next, 1 µl of the cDNA was used for PCR reaction (total volume of 50 µl) using specific forward and reverse primer of Ψ1367_18S and Zymotaq DNA polymerase (Zymo Research, E2004) for 34 cycles. Five microliter PCR products for Ψ1367_18S was then assessed on 3% agarose gel. The remaining PCR products were recovered, cloned into TOPO-TA cloning vectors and transformed to *Escherichia coli* competent cells. The deletion rate of Ψ1367_18S and C-to-T conversion rate of PCR products were assessed by Sanger sequencing at least 25 individual clones. PCR primer of Ψ1367_18S is listed in Supplementary Table 1.

Comparison of different reverse transcriptases

Commercially available reverse transcriptases were compared to determine which enzyme caused the highest deletion rate and yielded the

most cDNA. Ten reverse transcriptases including Maxima H minus (Thermo Fisher Scientific, EP0753), SuperScript II (Thermo Fisher Scientific, 18064014), SuperScript III (Thermo Fisher Scientific, 18080044), SuperScript IV (Thermo Fisher Scientific, 18090050), Revert Aid (Thermo Fisher Scientific, EP0441), M-MLV (Thermo Fisher Scientific, 28025013), AMV (New England Biolabs, M0277L), Proto-Script II (New England Biolabs, M0368L), Recombinant HIV (Worthington, LS05003) and HiScript III (Vazyme, R302-01) were tested. All RT enzymes were used according to the manufacturer's protocol. The yield in treated samples of AMV and Revert Aid was too low to be used for sequencing and analysis. For another eight reverse transcriptases, the deletion rate of all rRNA Ψ sites was compared.

Validation of Ψ sites by the locus-specific qPCR-based method

The radiolabeling-free, qPCR-based method was performed as described in ref. 39. Briefly, 6 µg polyA+ RNA was fragmented to ~200 nt at 94 °C for 2.5 min, and purified by ethanol precipitation. Half of the fragmented RNA was treated with 0.2 M CMC (Sigma-Aldrich, C106402) in BEU buffer as CMC (+) samples, while the other half was treated with BEU buffer as CMC (−) samples. Both CMC (+) and CMC (−) RNA samples were then synthesized using SuperScript II reverse transcriptases (Thermo Fisher Scientific, 18064014) with random hexamer primers, and the reactions were carried out at 25 °C for 10 min, 42 °C for 3 h and 70 °C for 15 min. After reverse transcription, 10× diluted cDNAs were analyzed by qPCR using TB Green Premix Ex Taq II (Takara, RR820A) and LightCycler 96 (Roche). The results were analyzed with GraphPad Prism 9 software, and the qPCR primers are listed in Supplementary Table 3.

PRAISE

Fragmentation. First, 500 ng DNase I treated total RNA or mRNA of each technical replicate was fragmented to ~150 nt by magnesium RNA fragmentation buffer (New England Biolabs, E6150S) for 4 min at 94 °C, followed by chilling on ice. The reaction was stopped by RNA fragmentation stop solution and purified by ethanol precipitation. After ethanol precipitation, RNA was resuspended in 6 µl nuclease-free water. One microliter of fragments (~75 ng) is transferred into a new nuclease-free tube and used as 'untreated' and stored at −80 °C. Remained RNA was subjected to bisulfite/sulfite treatment.

Sulfite/bisulfite treatment. The 85% sulfite/15% bisulfite solution was prepared freshly by adding 2.58 g of potassium sulfite (Sigma-Aldrich, 658510) and 0.30 g sodium bisulfite (Sigma-Aldrich, 243973) to 8 ml of RNase-free water and the solution was vortexed to clear (pH value is ~7.5). The 100 mM hydroquinone was prepared freshly by adding 11.01 mg of hydroquinone (Sigma-Aldrich, H9003) to 1 ml of RNase-free water. For sulfite/bisulfite treatment, 5 µl of RNA fragments was dissolved in 50 µl sulfite/bisulfite solution, which is a 100:1 mixture of 85% sulfite/15% bisulfite solution and 100 mM hydroquinone, and subjected to heat incubation at 70 °C for 5 h. The reaction mixture was desalted by passing it through Micro Biospin 6 chromatography columns twice (Bio-Rad, 7326200).

Desulfonation. Desalted RNA was transferred to a new 1.5 ml nuclease-free tube and adjusted to 100 µl with RNase-free water, then incubated with an equal volume of 1 M Tris−HCl (pH 9.0) at 75 °C for 30 min. The reaction was then immediately stopped by chilling on ice and ethanol precipitation. After ethanol precipitation, the RNA was resuspended in 10 µl of RNase-free water, and the concentration of RNA was quantified by Qubit RNA HS Assay Kit (Thermo Fisher Scientific, Q32855). Take 50 ng RNA (~1 µl) as 'treated' sample for subsequent library construction.

Library construction. Fifty nanogram 'treated' RNA sample and 10 ng 'untreated' RNA sample were subjected to library construction using SMARTer Stranded Total RNA-Seq Kit v3−Pico Input Mammalian

(Takara Bio, 634485) according to the manufacturer's protocol with several modifications. Briefly, for first-strand cDNA synthesis, Maxima H minus Reverse Transcriptase (Thermo Fisher Scientific, EP0753) was used instead of SMARTScribe II Reverse Transcriptase. Second, pre-PCR library was purified twice by 0.8× AMPure XP reagent (Beckman Coulter, A63881) and finally eluted with nuclease-free water. Because Ψ-containing rRNA fragments in sulfite/bisulfite treated sample can induce deletion signal, it cannot be digested by R-Probes v3, thereby potentially interfering with quantification of Ψ sites in rRNA. Therefore, we note that the step of ribosomal cDNA depletion was skipped. Eleven cycles were used for untreated and treated RNA in the second round of PCR. Remained procedures were performed with the SMARTer Stranded Total RNA-Seq Kit v3 (Pico Input Mammalian) according to the manufacturer's instructions. Library sequencing was performed on llumina Hiseq ×10 with paired-end 2 × 150 bp read length.

Preprocessing of raw sequencing data of PRAISE

Strand orientation of the original RNA as preserved in the process of library construction and reads R2 yields sequences sense to the original RNA. Thus, only reads R2 was used in our study. Illumina sequencing reads were first treated with cutadapt (version 3.5) for adapter removal and quality trimming. Work commands are as follows: cutadapt -j 28 --times 1 -e 0.1 -O 3 --quality-cutoff 25 -m 55. We then used Seqkit (version 0.13.2) to deduplicate based on 8 bp unique molecular identifier (UMI) at the 5′ end of reads R2, key process parameters are as follows: seqkit rmdup -s. Finally, umi_tools (version 1.0.0) was used to remove the 8 bp UMI in the deduplication read. Six bases after the UMI added during library construction at the 5′ end of inserted sequences and the six bases at the 3′ end of inserted sequences were also removed, using the key parameters: umi_tools extract --extract-method=string --bc-pattern=NNNNNNNNNNNNNN and umi_tools extract --extract-method=string --3prime --bc-pattern=NNNNNN.

PRAISE-tools and code availability

After the reads cleaning steps, the analysis of PRAISE data mainly includes two major parts, which are the alignment of reads and identification of pseudouridine sites with quantified deletion signal level. To make the analysis pipeline easy to implement, we developed PRAISE-tools, a computational pipeline that quantifies the deletion signals with high confidence. PRAISE-tools takes cleaned reads as input and finally reports the deletion rate of T sites as modification level.

Alignment of reads

PRAISE-tools firstly got all possible mapped positions to reference transcriptome (GRCh38) using HISAT2 (version 2.2.1) very sensitive mode. The key parameters are as follows: hisat2 -q --repeat --no-spliced-alignment --very-sensitive. After removing unmapped reads, we used our own tailored realignment based on the customized biopython pairwise alignment to remove low confidence alignments and get accurate deletion sites. The key parameter is python realignment.py −fast -ms 4.8. Next, the multiple-mapping reads with lower mapping scores will be removed and the deletion signal just one bp before reads' terminal will be removed using remove_multi-mapping.py and remove_end_signal.py. After getting the final mapping results in BAM, convert it into mpileup format with the key parameter of samtools mpileup -d 1500000 -BQ0 --ff UNMAP, QCFAIL -aa. Finally, the mpileup file will be counted and transferred into a specific BMAT format file.

Call Ψ signal

To call Ψ signal, we next applied a statistical test to each of the T sites (contingency table test between the treated sample and untreated sample). The calculated significance of P value is required to be <0.0001, and the calculated deletion rate difference between treated and untreated samples is required to be >5%. Because we mapped read

to the transcriptome, the same sites will be counted repeatedly in the different transcripts of one gene. Therefore, we finally removed those repeated signals in one gene by converting sites in the transcript into sites in the genome and kept the longest transcript using site_merge.py to give the Ψ lists of one sample. To give our final Ψ signal list, we overlapped two lists from two samples, only counted overlapped signals into our final Ψ list, and took the average deletion rate of two samples as its Ψ level.

Call Ψ signal in KO samples

To call Ψ signal in KO samples, we took our final Ψ list as a reference to find the deletion rate in KO samples. Then, we calculated the deletion rate difference between KO samples and WT samples as well as the deletion rate decreasing rate. For enzyme-dependent sites, the deletion rate difference is required to be >5%, and the decreasing rate of deletion is required to be >5%. To exclude the situations in which the sites are not covered in the KO samples, the total counts in KO sample is required to be 10. To give our final KO-dependent sites list, we overlapped two lists from two KO repeats, and only counted overlapped signals into our final KO-dependent list.

RNA secondary structure prediction

To analyze RNA secondary structure, we first extract the sequences near the Ψ sites (±12 nt), and sort them by the deletion rate of Ψ to create a fasta file. Then, we used RNAfold to predict RNA secondary structure with the following key parameters—RNAfold−temp=37 -p.

Quantification and statistical analysis

The P values were calculated using two-sided Student's t-test. $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; NS. Error bars represent mean ± s.d.

Reporting summary

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

Data availability

Raw and processed sequencing data have been deposited into the NCBI Gene Expression Omnibus (GEO) under the accession number GSE212210. The public datasets of Ψ-seq, Pseudo-seq, CeU-seq and BID-seq were downloaded from the GEO database with the accession numbers: GSE60047, GSE58200, GSE63655 and GSE179798. The reference transcriptome GRCh38 was downloaded from the following link: https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_000001405.40/. The reference of mitochondrial genome was downloaded from the following link: https://www.ncbi.nlm.nih.gov/nuccore/251831106. Source data are provided with this paper.

Code Availability

PRAISE tools, including a modified biopython package and several python scripts, were deposited on GitHub (https://github.com/Zhe-jiang/PRAISE). Besides, a detailed bioinformatics guide is available to help conduct the whole pipeline of PRAISE.

References

51. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
52. Stark, M. R. & Rader, S. D. Efficient splinted ligation of synthetic RNA using RNA ligase. *Methods Mol. Biol.* **1126**, 137–149 (2014).

Author contributions

M.Z., Z.J. and C.Y. conceived the project and wrote the manuscript with the help of J.P.; M.Z. and Z.J. developed the chemical assay under the guidance of C.Y.; M.Z. designed and performed the experiments with the help of Y.M.; Z.J. designed the tailored alignment method and performed the bioinformatic analysis of NGS data with the help of W.L. and K.L.; Y.Z. performed validation experiments for the mitochondrial site and immunofluorescences; B.L. constructed TRUB1 knockout cells and helped screen RT conditions; C.Y. supervised the project.

Competing interests

All the authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41589-023-01304-7.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41589-023-01304-7.

Correspondence and requests for materials should be addressed to Chengqi Yi.

Peer review information *Nature Chemical Biology* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permissions information** is available at www.nature.com/reprints.
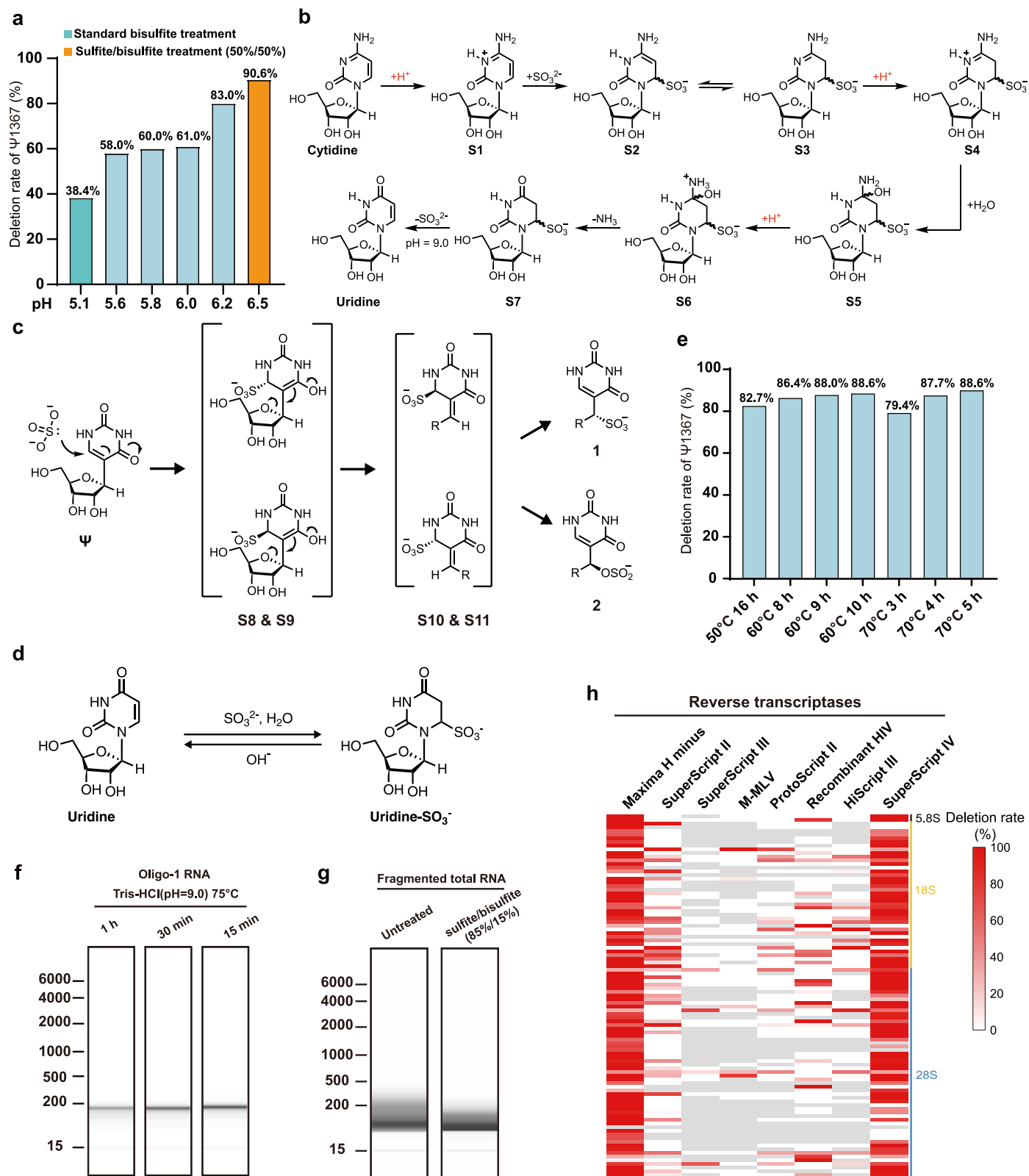
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Performance of different chemical reaction and RTase on capturing Ψ-induced deletion signature under different conditions.
(**a**) The deletion rate of Ψ1367 site in 18S rRNA under different sulfite/bisulfite conditions. The standard bisulfite reagent (pH=5.1) is equivalent to a molar mixture of 95% bisulfite and 5% sulfite (bisulfite:sulfite=19:1). The proportion of sulfite and bisulfite was increased from 95% sulfite/5% bisulfite to 50% sulfite/50% bisulfite, and the pH values are showed in the horizontal axis. 50% sulfite/50% bisulfite reagent is corresponding to pH 6.5. (**b**) Chemical mechanism of conversion from cytidine to uridine, which is both acid-dependent and sulfite-dependent. (**c**) Chemical mechanism of reaction between pseudouridine and sulfite reagent. After Michael addition, Ψ-SO3 generates two isoforms. Then, after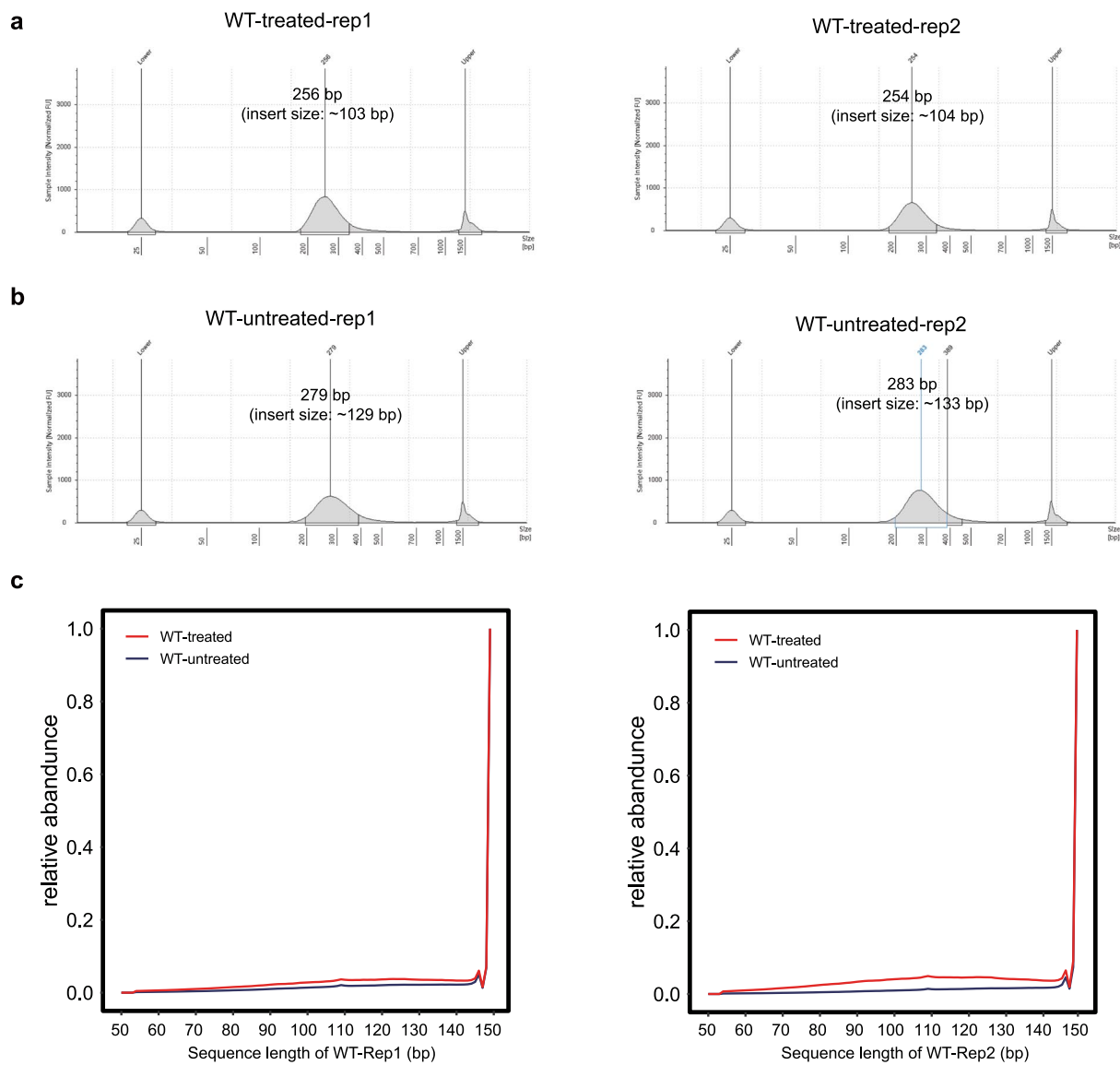 two steps of reverse Michael addition and Michael addition, the final products are two ring-open isoforms. (**d**) Chemical mechanism of transformation from uridine to uridine-bisulfite adduct. After alkaline treatment during the desulfonation step, the bisulfite group can be removed reversibly. (**e**) The deletion rate of Ψ1367 site in 18S rRNA under different time and temperature conditions. The horizontal axis represents time and temperature, and the vertical axis represents the deletion rate. (**f**) Fragments analysis of oligo RNA after Tris-HCl treatment for different time. '1 h', '30 min' and '15 min' represent different treatment time. (**g**) Fragment analysis of fragmented total RNA. 'untreated' represents fragmented total RNA without treatment and 'sulfite/bisulfite (85%/15%)' represents fragmented total RNA with sulfite/bisulfite treated. (**h**) Heatmap of deletion rate of Ψs in rRNA under different reverse transcriptase conditions.
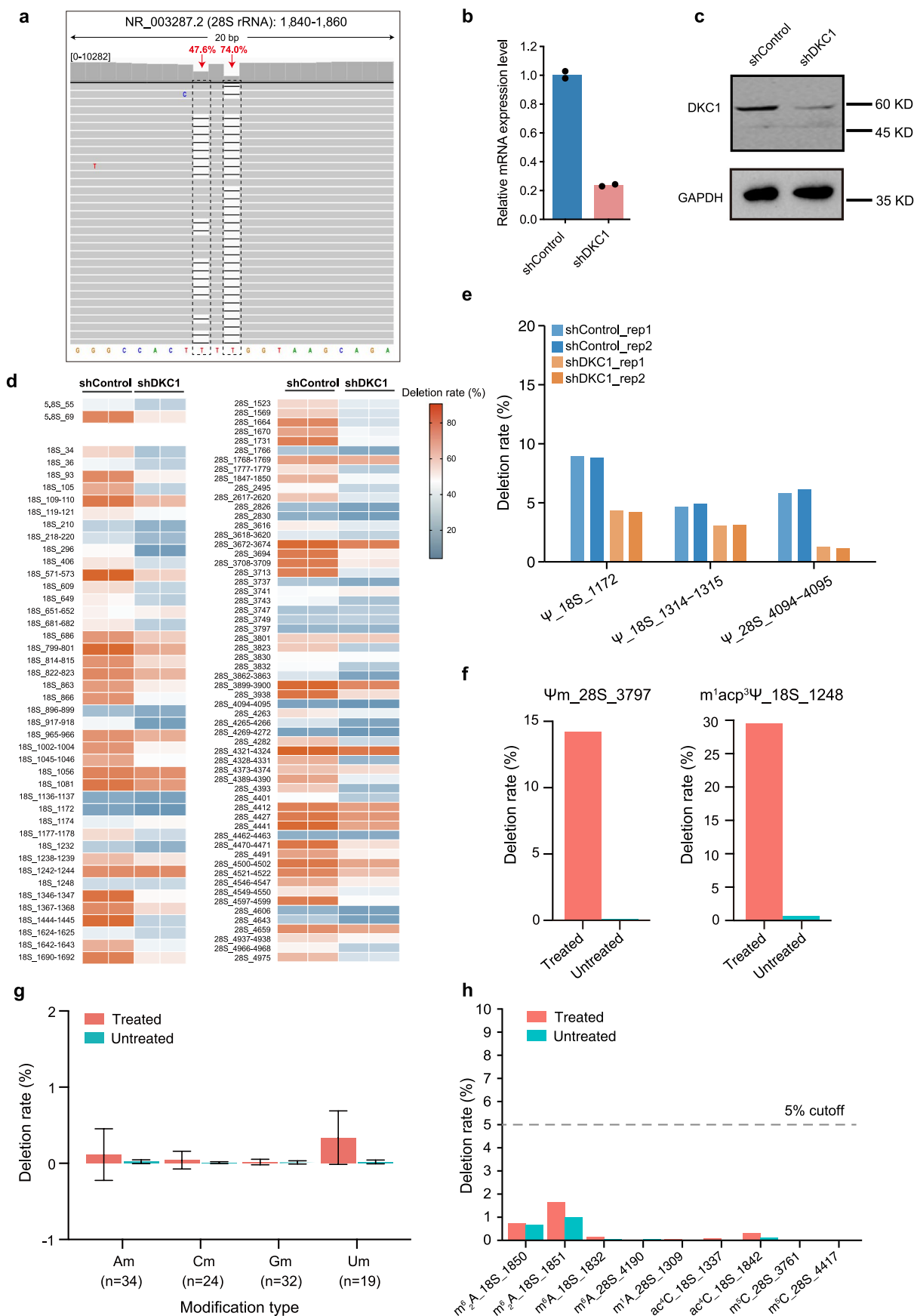
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Identification of known and novel rRNA sites under DKC1 knockdown and evaluation of the deletion rate for other RNA modifications. (**a**) An IGV view of read mapping showing deletion signals in position 1840–1860 of 28S rRNA. Two deletion sites were marked between four consecutive Ts in 1847–1850. (**b**) Boxplot representing the relative mRNA expression level of shControl and shDKC1 samples detected by qPCR. n = 2 biological replicates. (**c**) The protein expression level of DKC1 knock-down cells was determined by Western blot. GAPDH was used as a loading control. (**d**) Heatmap of deletion rates of all Ψ sites in 5.8S rRNA, 18S rRNA and 28 rRNA between shControl and shDKC1 sample. (**e**) Deletion rate of three additional Ψ sites in rRNA in shControl and shDKC1 sample. Each sites have two replicates of shControl and shDKC1 sample. (**f**) Deletion rate of treated and untreated samples on known Ψm and $m^1acp^3Ψ$ sites in rRNA. (**g**) Deletion rate of treated and untreated samples on known Am (n = 34), Cm (n = 24), Gm (n = 32), and Um (n = 19) sites in rRNA. Error bars represent mean ± SD. (**h**) Deletion rate of treated and untreated samples on known m^6_2A, m^6A, m^1A, ac^4C, and m^5C sites in rRNA.

a

WT-treated-rep1

256 bp
(insert size: ~103 bp)

WT-treated-rep2

254 bp
(insert size: ~104 bp)

b

WT-untreated-rep1

279 bp
(insert size: ~129 bp)

WT-untreated-rep2

283 bp
(insert size: ~133 bp)

c

d

Mapping rate of WT and KO libraries in PRAISE

Samples	States	Trimmed reads of Rep1	Mapped reads of Rep1	Mapping rate of Rep1	Trimmed reads of Rep2	Mapped reads of Rep2	Mapping rate of Rep2
WT	treated	158,189,406	100,096,514	63.28%	168,894,766	106,393,583	62.99%
	untreated	204,216,939	138,748,731	67.94%	158,114,211	104,607,124	66.16%
PUS1KO	treated	261,363,836	172,481,777	65.99%	218,626,230	143,835,798	65.79%
	untreated	226,473,382	157,180,129	69.40%	253,080,012	183,810,287	72.63%
PUS7KO	treated	191,237,272	124,020,814	64.85%	158,254,702	99,170,470	62.67%
	untreated	140,211,499	94,815,201	67.62%	152,335,626	105,300,511	69.12%
TRUB1KO	treated	154,507,962	96,779,353	62.64%	165,376,096	102,642,827	62.07%
	untreated	171,647,740	118,969,884	69.31%	178,319,706	122,446,325	68.67%

Extended Data Fig. 3 | Quality and statistics of sequencing libraries.
(**a**) Distribution of fragments size in two treated WT libraries after library construction. (**b**) Distribution of fragments size in two untreated WT libraries after library construction. (**c**) Distribution of sequencing reads length in two pairs of treated and untreated WT libraries. Data from Illumina is analyzed by FastQC. (**d**) Mapping rates of WT and KO libraries from WT, PUS1, PUS7, and TRUB1 knock-out cell lines.
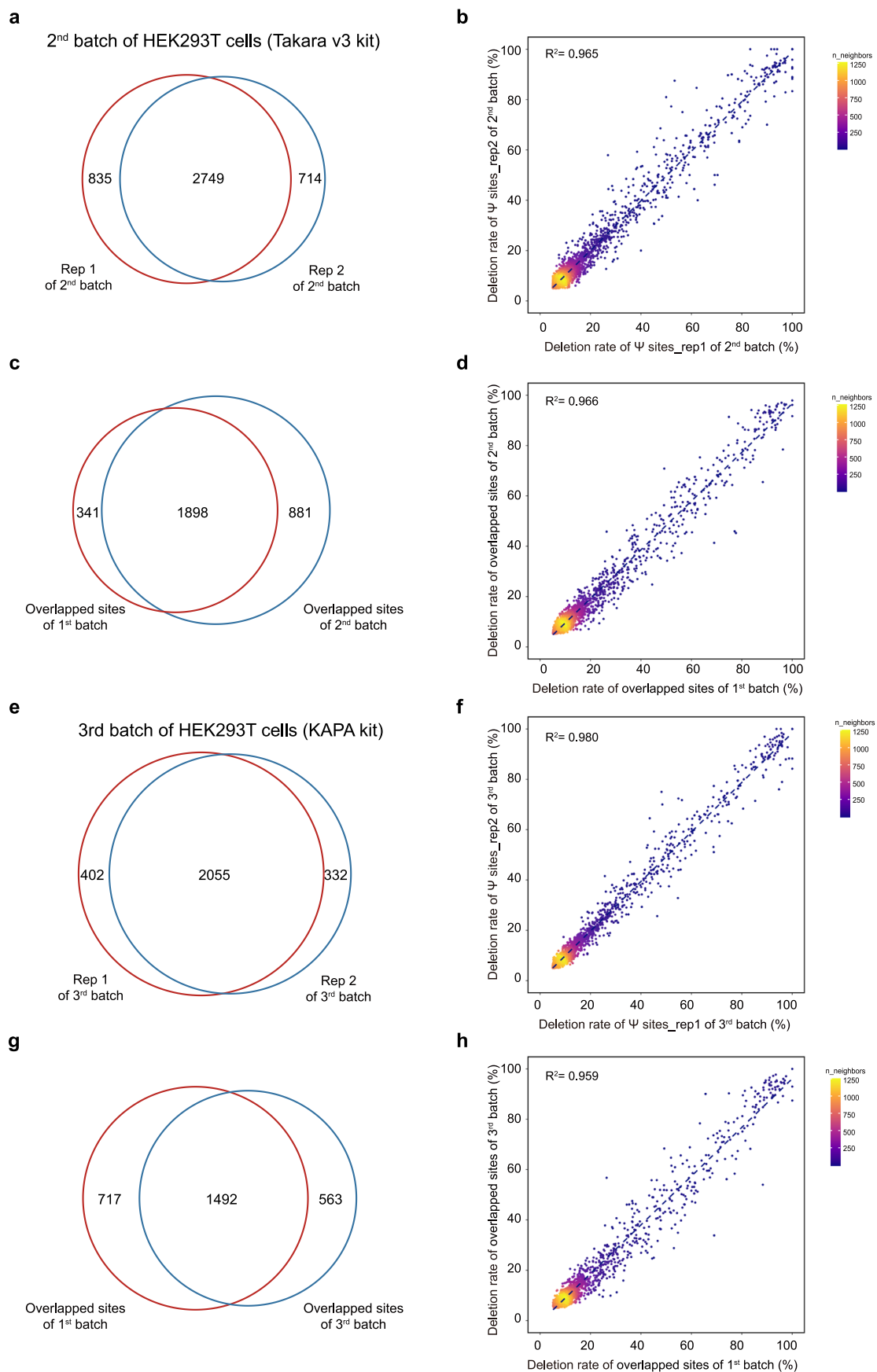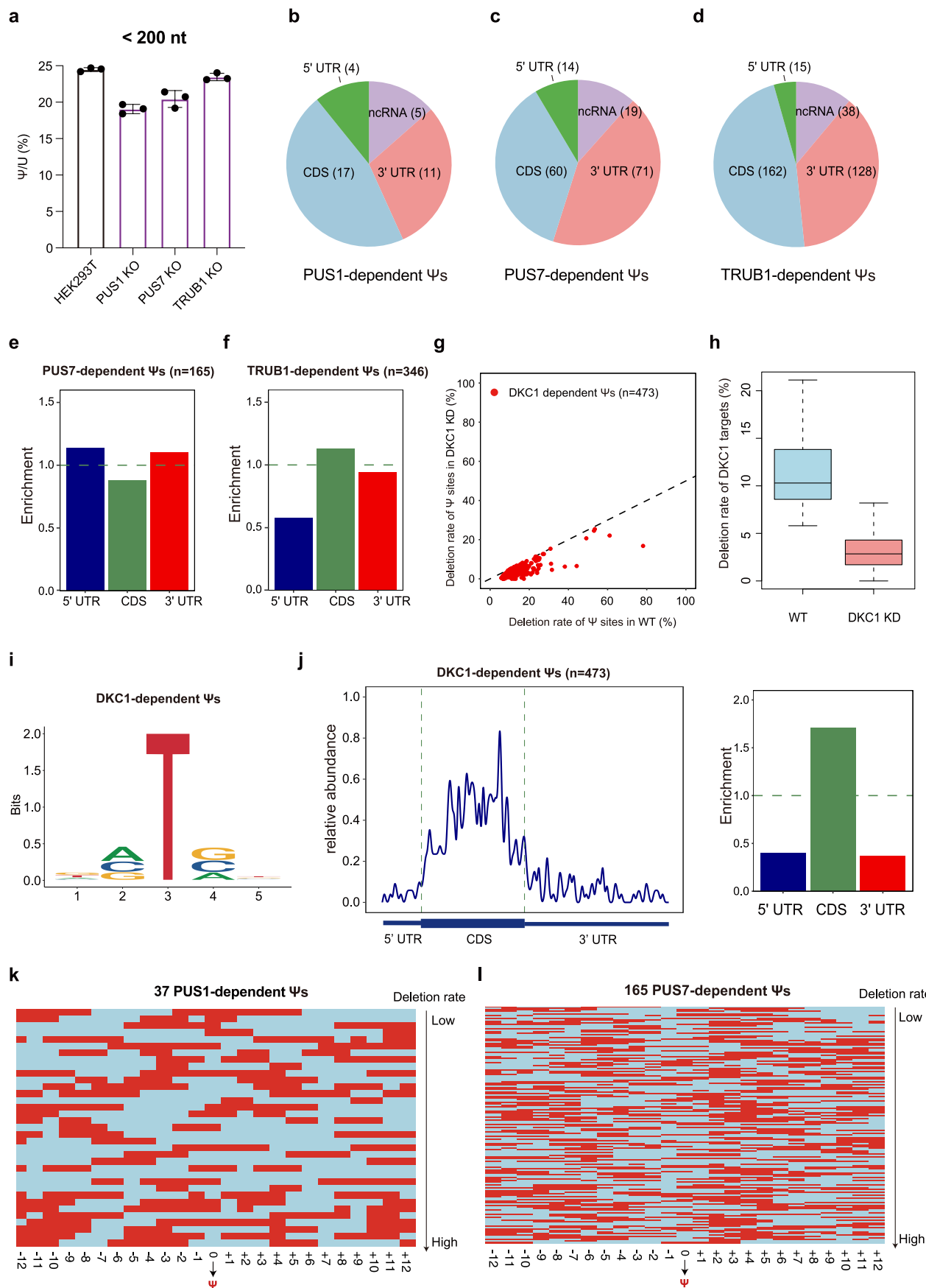
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Performance comparison of PRAISE data from different biological replicates and KAPA RNA-seq kit. (**a**, **b**) The performance of a new batch of HEK293T cells using Takara v3 library construction. a, Venn plot depicting ~78% overlapping between two technical replicates. b, Correlation of deletion rates of overlapping 2,749 Ψ sites detected in both technical replicates (R^2 = 0.965). (**c**) Venn plot depicting the 84.6% overlapping between two biological replicates using Takara v3 library construction. 'Overlapped sites of 1st batch' represents 2,209 Ψ sites originally reported in HEK293T cells, and 'Overlapped sites of 2nd batch' represents 2,749 Ψ sites detected in two technical replicates of 2nd batch cells. (**d**) Correlation of deletion rates of overlapped Ψ sites in (c) of two biological replicates (R^2 = 0.966). (**e**, **f**) The performance of a new batch of HEK293T cells using KAPA library construction kit (KAPA Stranded RNA-Seq Kit, #kk8400). e, Venn plot depicting the ~86% overlapping between two technical replicates. f, Correlation of deletion rates of 2,055 Ψ sites detected in two technical replicates (R^2 = 0.980). (**g**) Venn plot depicting the 67.5% overlapping between two biological replicates using Takara v3 library construction and KAPA kit. 'Overlapped sites of 1st batch' represents originally 2,209 Ψ sites reported in HEK293T cells, and 'Overlapped sites of 3rd batch' represents 2,055 Ψ sites detected in two technical replicates of 3rd batch cells. (**h**) Correlation of deletion rates of overlapped Ψ sites in (g) between two biological replicates (R^2 = 0.959).
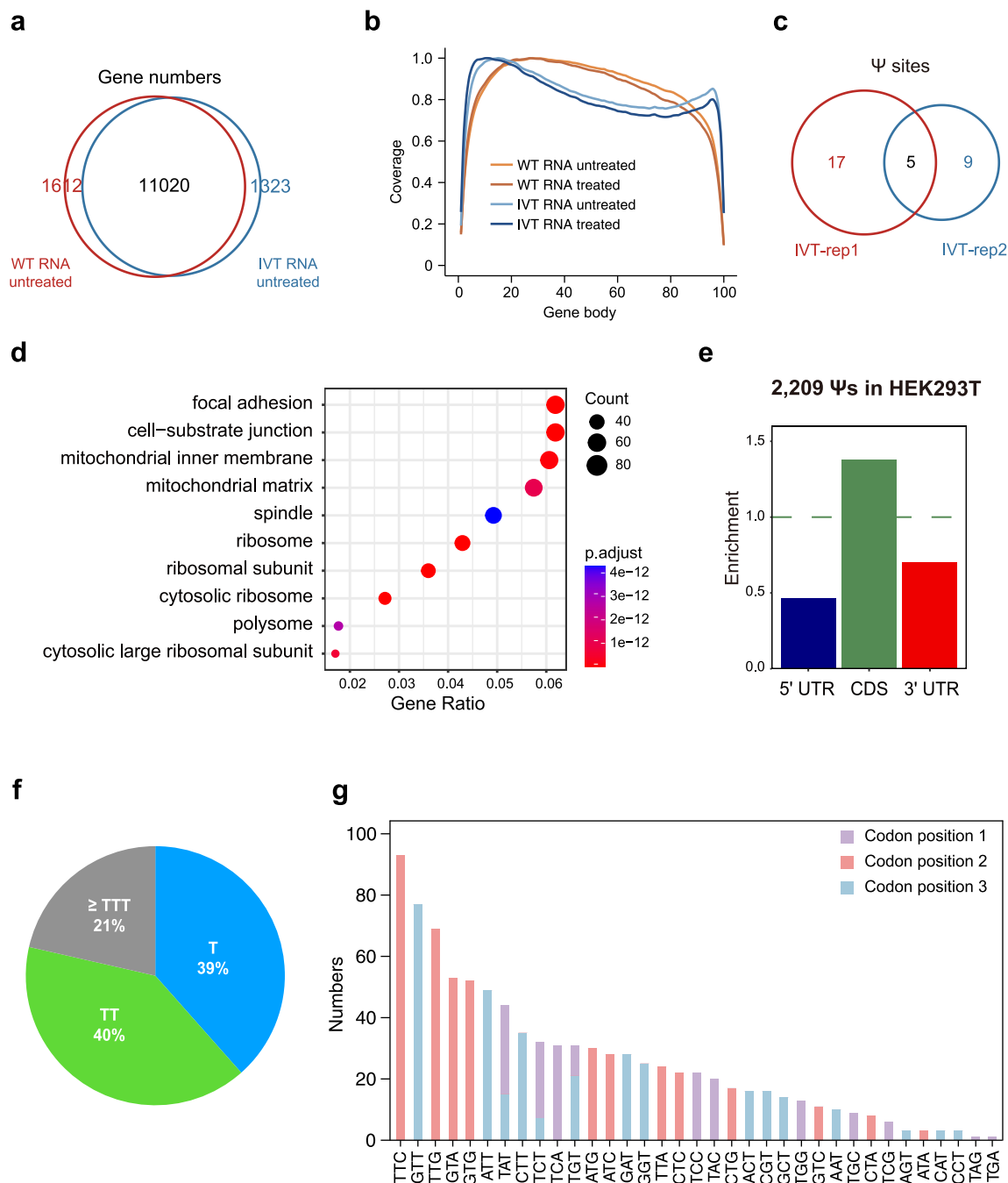
Extended Data Fig. 5 | Quality control of IVT samples and codon usage of cellular Ψ sites. (**a**) Venn diagram depicting gene overlap between wild type cellular RNA sample (red) and IVT sample (blue). IVT, in vitro transcription. (**b**) Read coverage from the 5' end to the 3' end of transcripts in wild type cellular RNA (orange) and IVT RNA (blue) libraries. Each library has two replicates and are combined and scaled to a normalized gene model. (**c**) Venn diagram depicting overlap of Ψ sites between IVT replicate 1 (red) and IVT replicate 2 (blue). (**d**) Genes containing Ψ sites are analyzed using gene ontology of cellular components. P values were adjusted using the Benjamini−Hochberg procedure. (**e**) Bar plot of relative enrichment distribution of 2,209 Ψ sites across transcript segments. CDS, coding sequence. (**f**) Pi chart representing proportions of Ψ sites within 'T', 'TT' and '≥TTT' contexts in human mRNA. (**g**) The counts of Ψ sites in codon types as well as positions are shown. Ψ sites observed at the first (purple), second (pink) and third (blue) codon position are indicated. Note that only Ψ sites present in the single and double U contexts (accounting for approximately 39% and 40% of all Ψ sites) were used for the codon preference analysis.

a < 200 nt

b PUS1-dependent Ψs
5' UTR (4), ncRNA (5), 3' UTR (11), CDS (17)

c PUS7-dependent Ψs
5' UTR (14), ncRNA (19), 3' UTR (71), CDS (60)

d TRUB1-dependent Ψs
5' UTR (15), ncRNA (38), 3' UTR (128), CDS (162)

e PUS7-dependent Ψs (n=165)

f TRUB1-dependent Ψs (n=346)

g DKC1 dependent Ψs (n=473)

h

i DKC1-dependent Ψs

j DKC1-dependent Ψs (n=473)

k 37 PUS1-dependent Ψs

l 165 PUS7-dependent Ψs

Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Distribution and secondary structure of PUS-dependent Ψ sites. (a) Quantification of Ψ modification of small RNA (<200 nt) in wild type and PUS knocked out cell lines. Values and error bars represent mean ± s.d. of n = 3 biological replicates. **(b)** Distribution of PUS1 dependent Ψ sites in human mRNA (32) and ncRNA (5). CDS, coding sequence. **(c)** Distribution of PUS7 dependent Ψ sites in human mRNA (145) and ncRNA (19). **(d)** Distribution of TRUB1 dependent Ψ sites in human mRNA (305) and ncRNA (38). **(e)** Bar plot of relative enrichment distribution of 165 PUS7-dependent Ψ sites in HEK293T mRNA. CDS, coding sequence. **(f)** Bar plot of relative enrichment distribution of 346 TRUB1-dependent Ψ sites in HEK293T mRNA. CDS, coding sequence. **(g)** Deletion rates of Ψ sites in mRNA in WT and DKC1 KD cells. The horizontal axis represents deletion rates of mRNA Ψ sites in WT HEK293T cells, and the vertical axis represents deletion rates of mRNA Ψ sites in DKC1 KD cells. **(h)** Bar plot of deletion rate of Ψ sites (n = 473) in mRNA between wild type and DKC1 KD sample. Box plot middle lines mark the median and the boundaries of the box indicate the 25th and 75th percentiles; whiskers encompass all data that are not considered outliers. **(i)** Motif analysis of DKC1-dependent Ψ sites within the single U context. **(j)** Distribution and relative enrichment across transcript segments of 473 DKC1-dependent Ψ sites in HEK293T mRNA. **(k)** Heatmap depicting secondary structure of 37 PUS1 dependent Ψ sites. A 24 nucleotides interval around Ψ site (at position 0) is selected. Paired bases are marked in red and unpaired bases are marked in blue. **(l)** Heatmap depicting secondary structure of 165 RNA strands containing PUS7 dependent Ψ sites. A 24 nucleotides interval around Ψ site (at position 0) is selected. Paired bases are marked in red and unpaired bases are marked in blue.
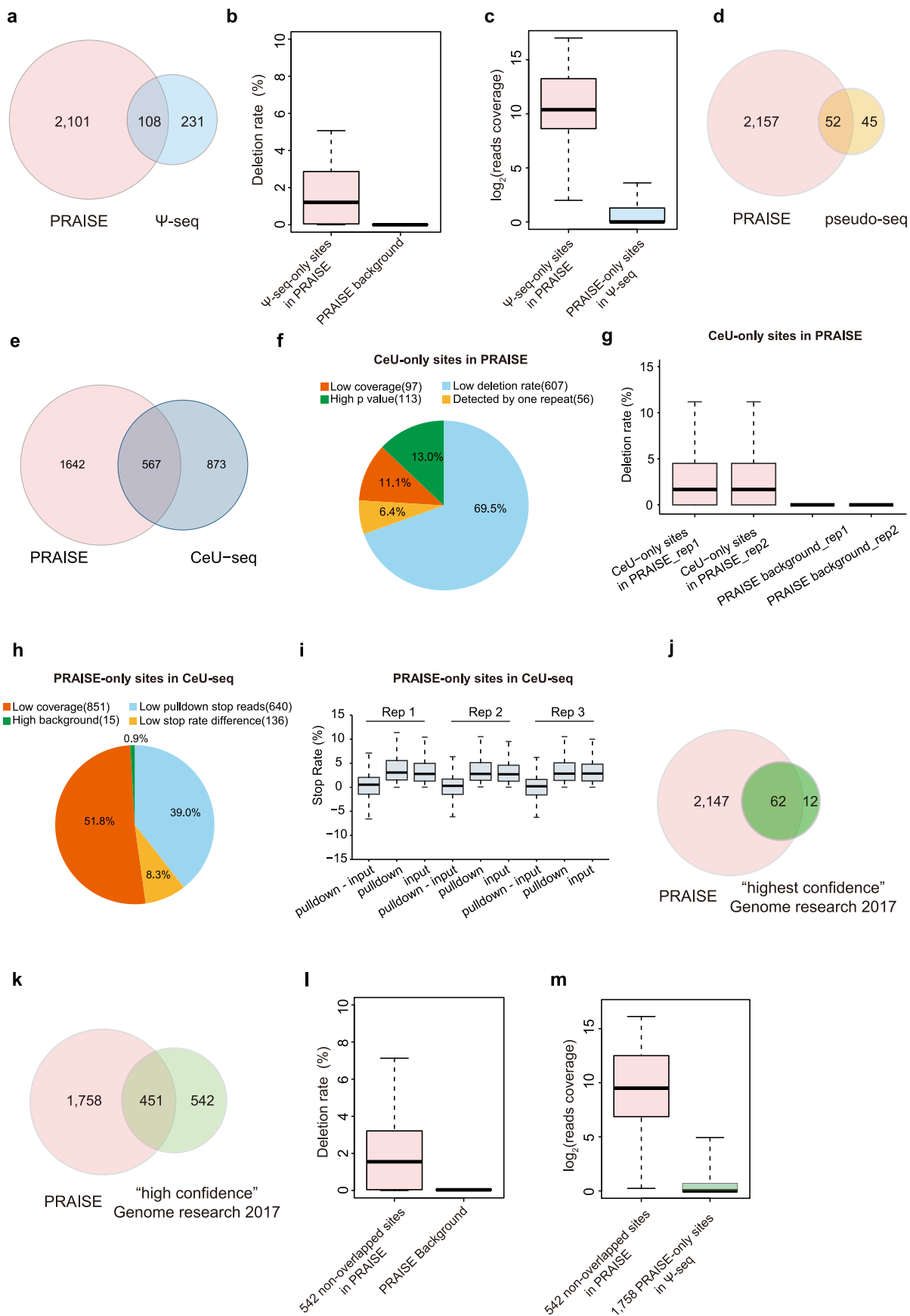
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Comparison of PRAISE and the CMC-based Ψ methods, including Ψ-seq, Pseudo-seq, CeU-seq. (**a**) Venn diagram depicting the overlap of Ψ sites detected in mRNA between PRAISE and Ψ-seq. (**b**) Deletion rates of Ψ-seq-only sites in PRAISE (n = 231) and random 'U' sites in PRAISE (n = 2,000). They had about 2% deletion rates and are higher than the background signal, but were not called because they are below our 5% cutoff. (**c**) Read coverage of Ψ-seq-only sites in PRAISE (n = 231) and PRAISE-only sites in Ψ-seq data (n = 2,101). (**d**) Venn diagram depicting the overlap of Ψ sites detected in mRNA between PRAISE and Pseudo-seq. (**e**) Venn diagram depicting the overlap between PRAISE and CeU-seq Ψ sites. (**f**) Pi chart of CeU-seq-only sites in PRAISE data. 'Low coverage' represents the coverage is lower than 20 reads. 'High p value' represents the p-value is higher than 0.0001 in PRAISE data. (**g**) Deletion rates of CeU-seq-only sites with sufficient reads in PRAISE (n = 776) and random U sites in PRAISE data (n = 2,000). (**h**) Pi chart of 1,642 PRAISE-only sites in CeU-seq data. (**i**) Stop rates of 48.2% of PRAISE-only sites with read coverage in CeU-seq data (n = 791). Data from three replicates are shown. (**j**, **k**) Venn diagram depicting the overlap between PRAISE sites and highest (j) or high (k) confidence sites from Safra et al in Genome research 2017. (**l**) Deletion rate of non-overlapped 'high-confidence' sites in PRAISE (n = 542) and random U sites in PRAISE data (n = 2,000). (**m**) Read coverage of non-overlapped 'high-confidence' sites in PRAISE data (n = 542) and PRAISE-only sites in Ψ-seq data (n = 1,758). Box plot middle lines mark the median and the boundaries of the box indicate the 25th and 75th percentiles; whiskers encompass all data that are not considered outliers.
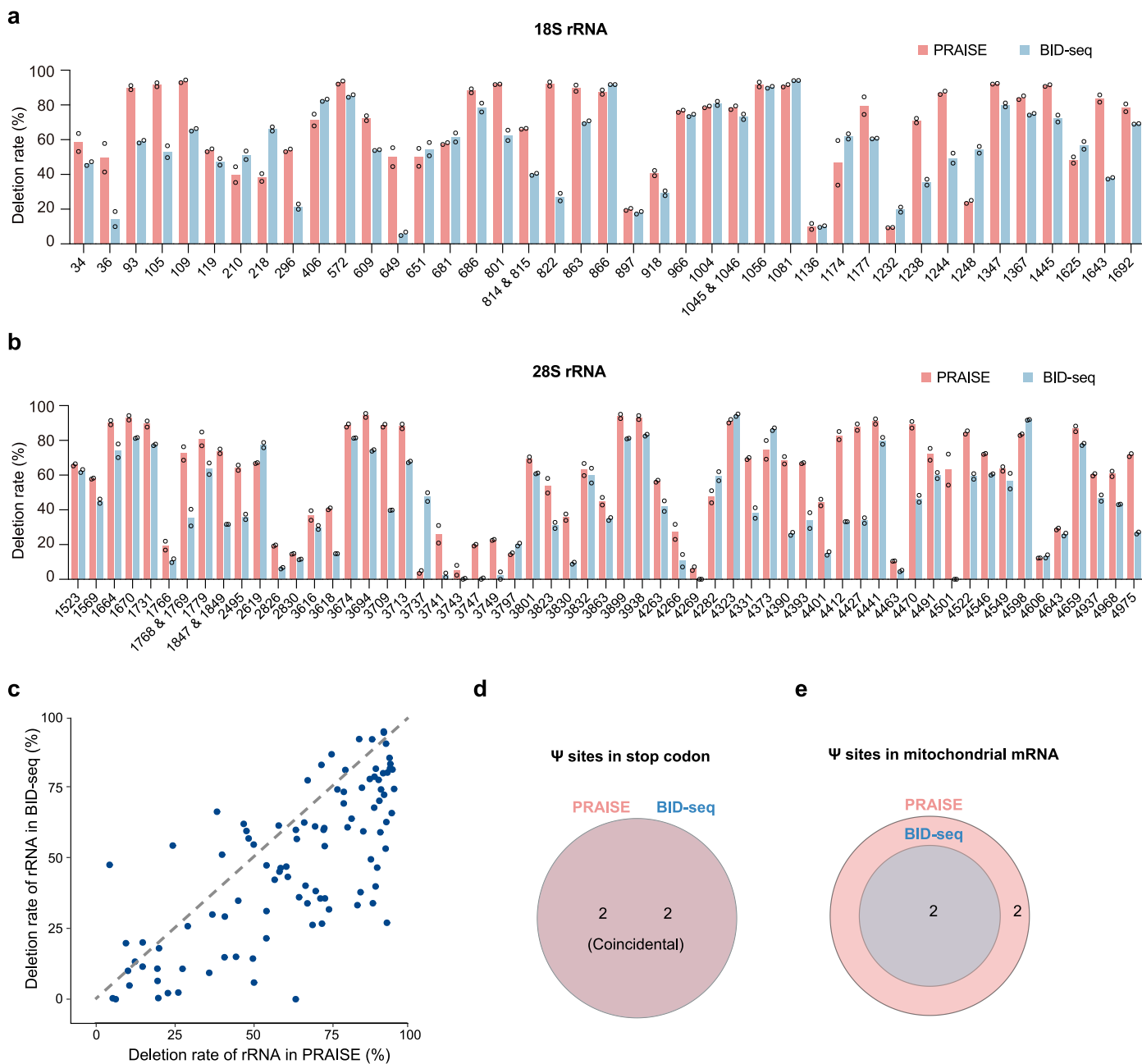
Extended Data Fig. 8 | Comparison of PRAISE and BID-seq data. (**a**, **b**) Deletion rates of Ψ sites in 18 s rRNA(a) and 28 s rRNA(b) detected by PRAISE and BID-seq. (**c**) Deletion rates of all detected Ψ sites in 18 rRNA and 28S rRNA between PRAISE and BID-seq. The horizontal axis represents deletion rates of rRNA Ψ sites in PRAISE, and the vertical axis represents deletion rates of rRNA Ψ sites in BID-seq.

(**d**) Venn diagram depicting two Ψ sites (Ψ in MDK and NDUFS2) detected in stop codon by both PRAISE and BID-seq. (**e**) Venn diagram depicting four Ψ sites detected in mitochondrial mRNA by PRAISE (Ψ in MT-CO1, MT-CO3, MT-ND4, and MT-CYB), two of which are also detected by BID-seq (Ψ in MT-CO1 and MT-CO3).

a

| mRNA	Accession Number	Site	PRAISE	BID-seq	CeU-seq	Ψ-seq	Pseudo-seq	Dependency	Validation
IBA57	NM_001010867.4	4507	√	√	√	-	-	PUS7	Yes
IDI1	NM_001317956.2	696	√	-	√	√	-	TRUB1	Yes
NBAS	NM_015909.4	3879	√	-	-	-	-	TRUB1	Yes
UNC13B	NM_001371189.2	9671	√	-	-	-	-	TRUB1	Yes

b
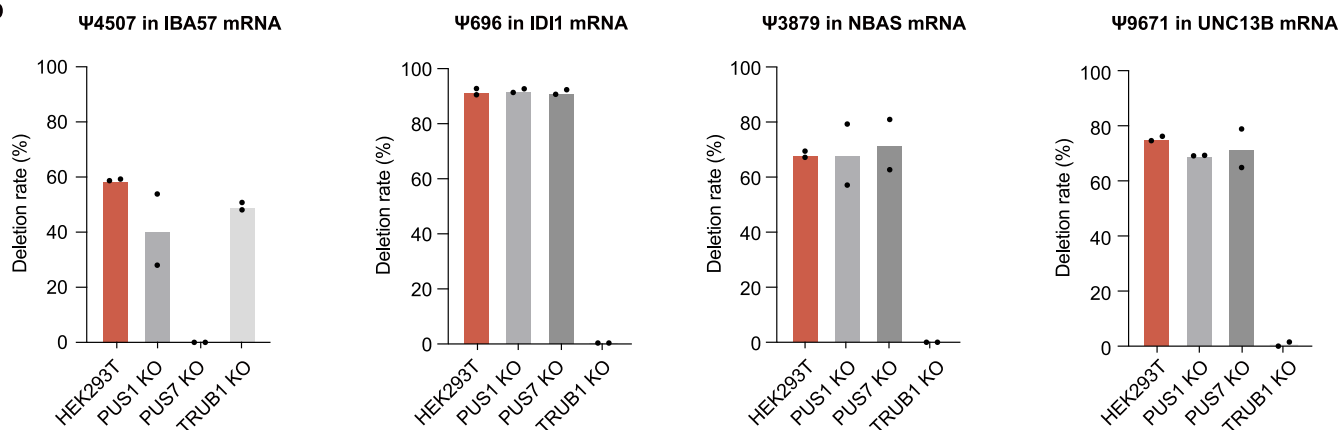
c
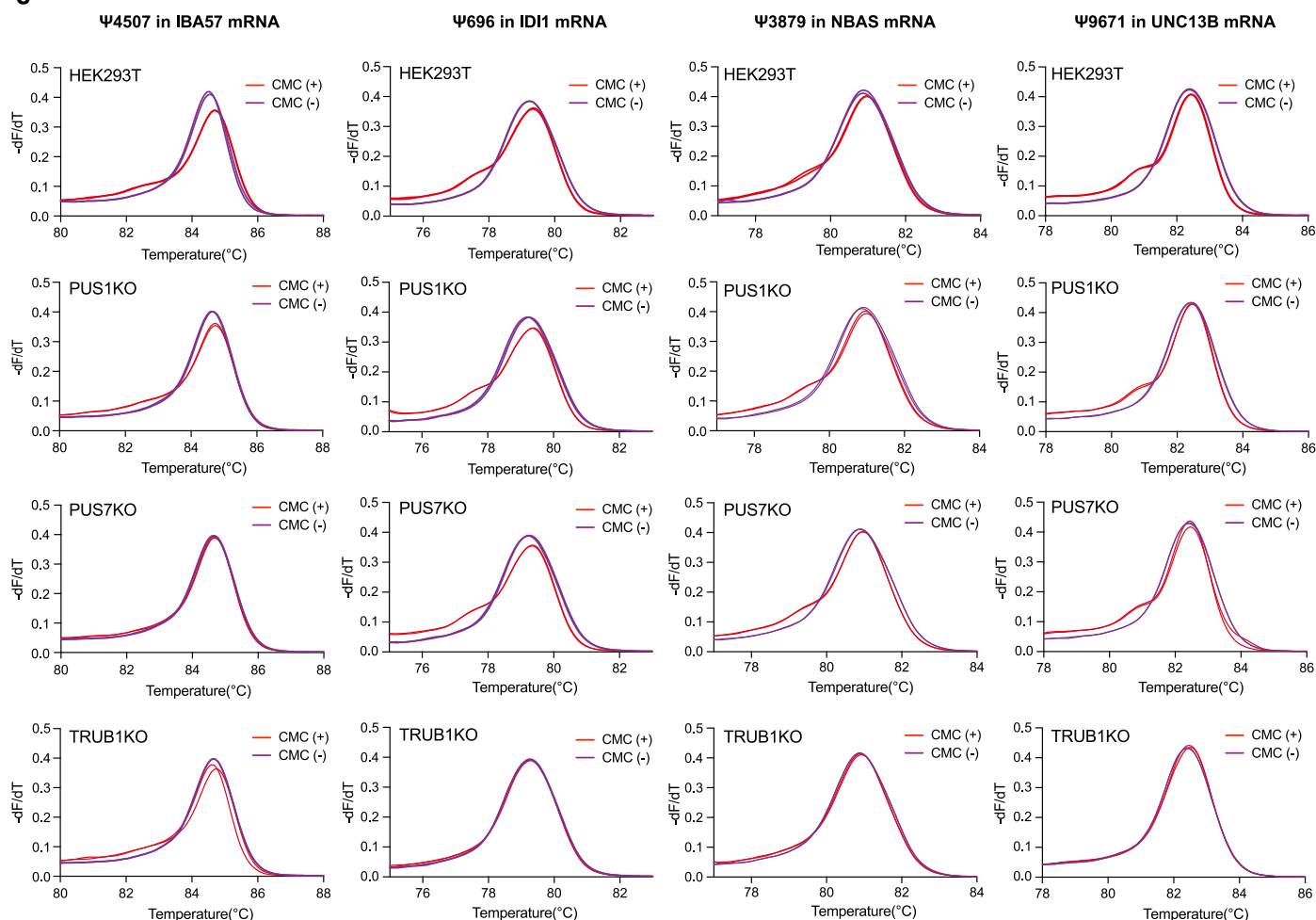
Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Validation of four mRNA Ψ sites in HEK239T cells identified by PRAISE. (**a**) Four representative mRNA Ψ sites identified by PRAISE only or also by CMC-based approaches and BID-seq. Noted that they were all assigned to a particular Ψ writer by PRAISE; Ψ sites in NBAS and UNC13B were only detected by PRAISE. (**b**) Deletion rates of the four mRNA sites in HEK293T wild-type and PUS1, PUS7, TRUB1 knock-out cell lines in PRAISE data. (**c**) Melting curves of qPCR products containing four mRNA sites detected by CMC-assisted, qPCR-based method. The curves were obtained by high-resolution melting analysis.
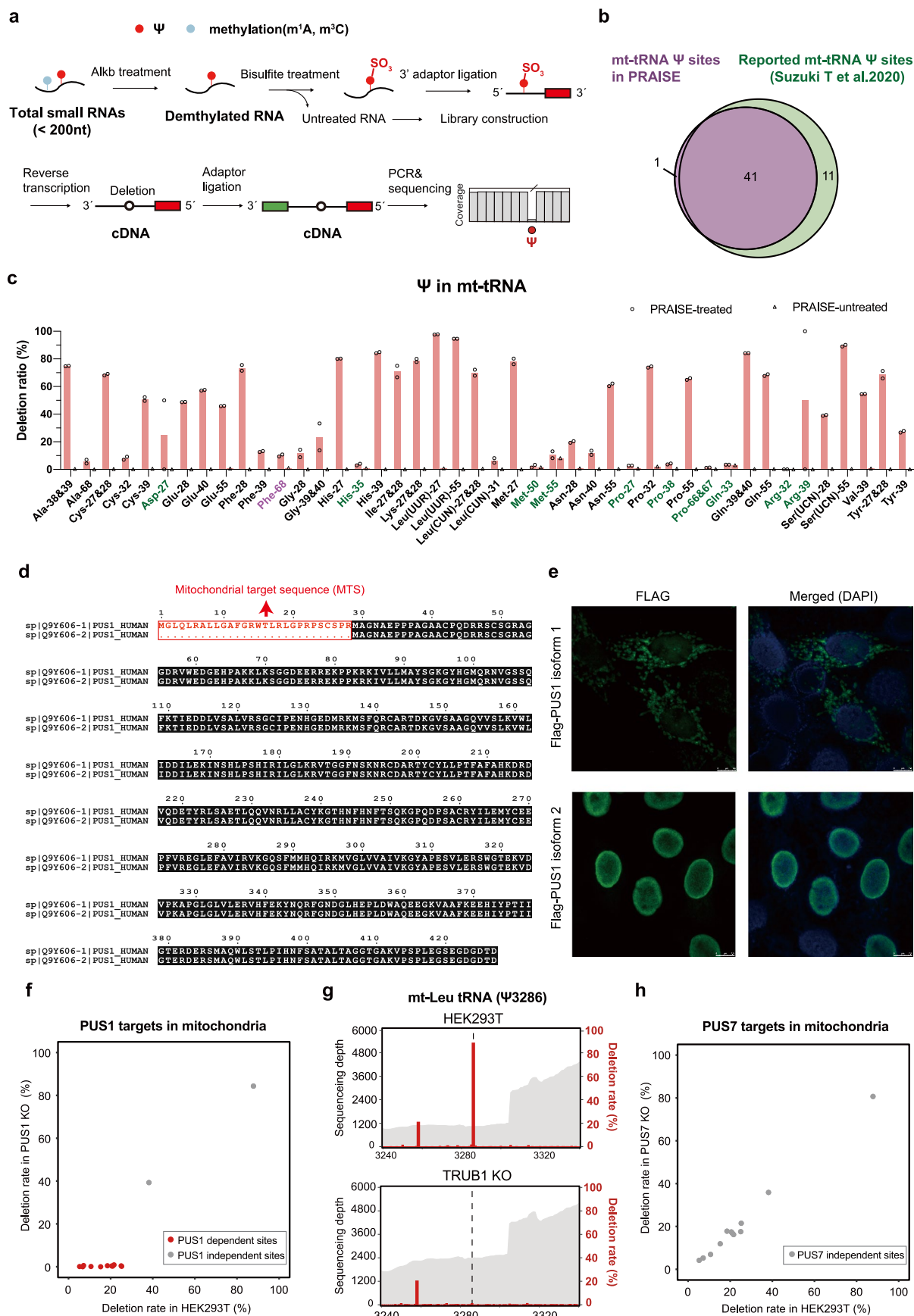
Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Ψ sites in mt-tRNA and mitochondrial targets assigned to PUS1. (**a**) Schematic diagram of PRAISE to identify Ψ sites in mitochondrial tRNAs. Total small RNAs are first treated by Alkb demethylase to demethylase m^1A and m^3C, and then treated with 85% sulfite/15% bisulfite solution. RNA was then followed by 3' adapter ligation, reverse transcription, 5' linker ligation, PCR, and sequencing. The 'Untreated' sample is set as a negative control. (**b**) Venn diagram depicting the overlap of Ψ sites detected in mitochondrial tRNA between PRAISE and previously reported method (Suzuki T et al.[41]). (**c**) Deletion rates of 51 known Ψ sites and 1 novel of Ψ sites in mitochondrial tRNA. For each site, deletion rates of treated and untreated are marked with different symbols. (**d**) Sequence alignment and comparison of PUS1 isoform 1 and isoform 2 in human cell line. Mitochondrial target sequence of PUS1 isoform 1 is marked in red color. (**e**) Double immunofluorescence staining of HeLa cells transfected with expression vectors encoding Flag-tagged PUS1 isoform 1 and 2 (green) and DAPI (blue). The left panels show the expression vectors, and the right panels show the green and blue images merged in Adobe Photoshop. Representative images were obtained under the same exposure condition from n = 3 biological repeats. (**f**) Deletion rates of Ψ sites in mitochondrial RNA (mt-RNA) in PUS1 KO and HEK293T cells. The horizontal axis represents deletion rate of mt-RNA Ψ sites in HEK293T cells, and the vertical axis represents deletion rate of mt-RNA Ψ sites in PUS1 KO cells. (**g**) Sequencing depth of regions surrounding Ψ3286 in mt-Leu tRNA and the corresponding deletion rate in HEK293T (WT) and TRUB1 KO treated samples are plotted. (**h**) Deletion rate of Ψ sites in mt-RNA in PUS7 KO and HEK293T cells. The horizontal axis represents deletion rate of mt-RNA Ψ sites in HEK293T cells, and the vertical axis represents deletion rate of mt-RNA Ψ sites in PUS7 KO cells.

nature portfolio

Corresponding author(s):   Chengqi Yi

Last updated by author(s):   Feb 18, 2023

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	Next-generation sequencing data was observed and demultiplexed by Illumina Hiseq X10 platforms.
Data analysis	Illumina sequencing reads were treated with cutadapt (version 3.5) for adaptor removal and quality trimming. Removed-adapter reads were aligned to reference transcriptome (GRCh38) using HISAT2 (version 2.2.1) very sensitive mode. Seqkit (version 0.13.2) was used to deduplicate PCR of Takara library based on 8 bp UMI. 8bp UMI in the deduplication read was removed using umi_tools (version 1.0.0). Samtools (version 1.10) was used to analyze the mapping results. ViennaRNA (version 2.5.1) was used to predict the RNA secondary structure. Graphpad (version 9.0.0) prism 9 was used to analyze the data. Sanger sequencing data was analyzed using the Snapgene (version 4.3.6). Python (version 3.7) was used to analyze the data. Several packages was used, including modified biopython (version 1.79), pysam (version 0.15.3), multiprocessing (version 2.6.2.1), shutil (version 3.4), random (version 3), string (version 3.7), pandas (version 1.3.4), scipy (version 1.7.3) R (version 4.1.1) was used to analyze and plot the data. Several packages was used, including ggplot2 (version 3.3.6), LSD (version 4.1-0), ggpointdensity (version 0.1.0), clusterProfiler (version 4.2.2), ggseqlogo (version 0.1), vioplot (version 0.3.7), circlize (version 0.4.15), pheatmap (version 1.0.12), ggalt (version 0.4.0) Custom code used for analysis are available at https://github.com/Zhe-jiang/PRAISE.

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

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

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

All the high-throughput sequencing data generated for this study has been deposited in the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE212210. The public datasets of Ψ-seq, Pseudo-seq, CeU-seq, and BID-seq were downloaded from the GEO database with the accession numbers: GSE60047, GSE58200, GSE63655, GSE148603. The reference transcriptome GRCh38 was downloaded from the following link: https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_000001405.40/. The reference of mitochondrial genome was downloaded from the following link: https://www.ncbi.nlm.nih.gov/nuccore/251831106.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research.](#)

Reporting on sex and gender	This study only used cell lines and did not involve relevant content of human research
Population characteristics	see above
Recruitment	see above
Ethics oversight	see above

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

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	For experiments with knock down efficiency analysis, 2 biological replicates were performed to confirm reproducibility. For sequencing data, the sample size was determined based on our prior experience (Li et al., Mol Cell. 2017 (PMID: 31676230); Liu et al., Nat Biotechnol. 2022 (PMID: 36302990)), we performed 3 biological replicates (each biological replicate also corresponds to 2 technical replicates) to confirm the reproducibility because 3 biological replicates are sufficient to support conclusions in this paper. Precise sample sizes are indicated in the figure legends and Methods section.
Data exclusions	No data were excluded from the analyses.
Replication	We performed three independent experiments and successfully of each other.
Randomization	Samples were randomly allocated into experimental groups.
Blinding	Experiments were not blinded in our studies because experimental conditions were evident. All samples were treated identically through standard procedures.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Antibodies

Antibodies used	The DKC1 and GAPDH proteins were detected using the antibodies anti-DKC1 (Santa Cruz, sc-373956, 1:1,000) and anti-GAPDH (CWBIO, CW0100, 1:2,000).
Validation	For western blot, we used control and stable knockdown cell lines (shControl and shDKC1) to validate the specificity of DKC1 antibody. Meanwhile, GAPDH was set as the loading control. 1) Anti-DKC1 antibody (Cited in 22 publications: https://www.scbt.com/p/dyskerin-antibody-h-3; this antibody has been validated by our previous study: https://pubmed.ncbi.nlm.nih.gov/36521489/) 2) Anti-GAPDH antibody (this antibody has been validated by previous studies: https://pubmed.ncbi.nlm.nih.gov/36302990/, https://www.nature.com/articles/s41589-019-0420-5)

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	HEK293T cells were purchased from ATCC (ATCC CRL-3216)
Authentication	No method of cell line authentication was used.
Mycoplasma contamination	All cell lines are tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No such cell lines were used in this study.