

Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome

Xiaoyu Li^{1,5}, Ping Zhu^{2,5}, Shiqing Ma^{1,3,5}, Jinghui Song¹, Jinyi Bai^{1,3}, Fangfang Sun¹ & Chengqi Yi^{1,4*}

Pseudouridine (Ψ) is the most abundant post-transcriptional RNA modification, yet little is known about its prevalence, mechanism and function in mRNA. Here, we performed quantitative MS analysis and show that Ψ is much more prevalent (Ψ/U ratio ~0.2–0.6%) in mammalian mRNA than previously believed. We developed N_3 -CMC-enriched pseudouridine sequencing (CeU-Seq), a selective chemical labeling and pulldown method, to identify 2,084 Ψ sites within 1,929 human transcripts, of which four (in ribosomal RNA and *EEF1A1* mRNA) are biochemically verified. We show that hPUS1, a known Ψ synthase, acts on human mRNA; under stress, CeU-Seq demonstrates inducible and stress-specific mRNA pseudouridylation. Applying CeU-Seq to the mouse transcriptome revealed conserved and tissue-specific pseudouridylation. Collectively, our approaches allow comprehensive analysis of transcriptome-wide pseudouridylation and provide tools for functional studies of Ψ -mediated epigenetic regulation.

More than 100 different types of post-transcriptional modifications to RNA molecules have been characterized so far¹. Among them, Ψ is overall the most abundant^{2,3}. Ψ in ribosomal RNA (rRNA) is required for binding to the internal ribosome entry site and for translational fidelity⁴. In small nuclear RNAs, Ψ can fine-tune branch site interactions and affect mRNA splicing⁵, and Ψ can also make a bifurcate hydrogen bond to stabilize RNA structures⁶. In eukaryotes, pseudouridylation can be catalyzed via two distinct mechanisms: an RNA-guided mechanism that involves the ribonucleoproteins of the box H/ACA class and an RNA-independent mechanism that involves ‘stand-alone’ Ψ synthases^{7,8}. Directly or indirectly, some Ψ synthases (for instance, DKC1 and hPUS1) have been linked to human diseases^{9,10}. Additionally, Ψ modifications in noncoding RNAs (ncRNAs) can also be regulated in response to stress conditions^{11–13}.

In contrast to the extensive literature on pseudouridylation in ncRNAs, our knowledge regarding to the prevalence, mechanism and function of Ψ modifications in mRNA and long noncoding RNA (lncRNA) is limited. Early findings that some box H/ACA small nucleolar RNAs (snoRNAs) are complementary to mRNAs raise the possibility of pseudouridylation in mRNA^{14,15}. In an intriguing study, targeted pseudouridylation was shown to convert nonsense codons into sense codons¹⁶. More recently, it was demonstrated that Ψ can facilitate noncanonical base pairing in the ribosome decoding center so as to promote nonsense suppression¹⁷. However, though the notion that artificial pseudouridylation can generate protein diversity is very tempting, without convincing evidence of naturally existing mRNA pseudouridylation, the physiological significance of mRNA pseudouridylation remains unclear.

Major technical and experimental challenges are present for the detection of Ψ modifications in mRNA. First, compared to rRNA and tRNA, mRNA is of low abundance; hence, a pseudouridylation detection method must be sensitive enough to allow inspection of transcriptome-wide pseudouridylation events. Because Ψ is

mass-silent and indistinguishable from regular U bases during reverse transcription, previous studies relied on a chemical, *N*-cyclohexyl-*N'*-*b*-(4-methylmorpholinium) ethylcarbodiimide (CMC), to specifically label and distinguish Ψ from U^{18,19}. The CMC- Ψ adduct causes reverse transcription to terminate one nucleotide 3' to the CMC- Ψ adduct, thereby enabling Ψ detection at single-base resolution in a primer extension method²⁰. Recently developed CMC-based profiling methods of Ψ , termed pseudo-seq, Ψ -seq and PSI-seq, have allowed identification of Ψ modifications at single-base resolution in yeast and human mRNAs^{21–23}. However, existing profiling methods do not pre-enrich Ψ -containing RNAs. Consequently, pseudouridylation events of low abundance may be missed. Second, the lack of a quantitative method hinders our ability to comprehensively understand the prevalence of pseudouridylation in the transcriptome and to evaluate pseudouridylation dynamics. For instance, recent profiling studies have identified ~50–300 Ψ sites in yeast and ~100–400 Ψ sites in human cells^{21–23}. To what extent the current Ψ profiles represent the genuine landscape of Ψ in the transcriptome has not been addressed. Third, the newly identified Ψ sites in mRNA have not been validated using a CMC-independent method. Taken together, such limitations substantially impede an in-depth, quantitative view of transcriptome-wide pseudouridylation events.

Here, we report a transcriptome-wide profiling method that is enabled via a chemically synthesized CMC derivative that pre-enriches the Ψ -containing RNA through biotin pulldown. We first show by quantitative MS that pseudouridylation is much more prevalent (Ψ/U ratio ~0.2–0.6%) than previously thought; indeed, our chemical pulldown-based profiling method allowed us to identify 2,084 Ψ sites in the transcriptome of HEK293T cells and 1,741 and 1,543 Ψ sites in mouse brain and liver, respectively. Four new Ψ sites (three in rRNA and one in *EEF1A1* mRNA) were experimentally validated. Moreover, we show that known Ψ synthases, including hPUS1, can act on human mRNA. Furthermore, these

¹State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences and Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China. ²Biodynamic Optical Imaging Center, School of Life Sciences, Peking University, Beijing, China. ³Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. ⁴Synthetic and Functional Biomolecules Center, Department of Chemical Biology, College of Chemistry and Molecular Engineering, Peking University, Beijing, China. ⁵These authors contributed equally to this work. *e-mail: chengqi.yi@pku.edu.cn

approaches also allow us to evaluate the dynamics of pseudouridylation in response to stimuli and reveal the stress-specific pattern of inducible pseudouridylation. Collectively, our approaches reveal a dynamically regulated landscape of Ψ modifications in the human and mouse transcriptome and provide robust tools for elucidating the functions of mRNA pseudouridylation.

RESULTS

Ψ is an abundant modification in mammalian mRNA

To determine whether Ψ is present in mRNA *in vivo*, we performed quantitative MS studies. Ψ has been found in tRNA, rRNA and snRNA; in particular, it is present in rRNA at a Ψ/U ratio of ~ 7 – 9% , as detected by radioactivity-based methods^{24,25}. To minimize contamination of such abundant ncRNAs, we first sought to develop a stringent mRNA purification protocol. Starting with total RNA, two successive rounds of poly(A)⁺ selection were first performed; we then carried out one small RNA (<200 nt) removal step to the poly(A)⁺-selected fraction, which was further purified with two successive rounds of rRNA depletion steps.

We systematically measured the Ψ/U ratio for RNAs obtained from every step of such iterative purification procedure by LC/MS/MS (Fig. 1a,b and Supplementary Results, Supplementary Fig. 1a,b). N^6,N^6 -dimethyladenosine (m^6_2A), which is only found in 18S rRNA, was used to indicate rRNA contamination. Our results showed that the Ψ/U ratio remained stationary after the first rRNA depletion step (Fig. 1b and Supplementary Fig. 1b,c). The small RNA removal step and the second rRNA depletion step do not further increase the purity of mRNA, as assessed by the Ψ/U and m^6_2A/U ratios. RNA-seq also showed that rRNA represented only $\sim 0.16\%$ of the mapped reads in such purified RNA samples (Supplementary Fig. 1d). Thus, two successive rounds of poly(A)⁺ selection followed by one rRNA depletion step are necessary and sufficient to quantify Ψ .

We then applied the above purification protocol to various RNA populations and quantified the Ψ/U ratio. The Ψ/U ratio ranges from

approximately 0.20% to 0.40% among different cell lines (Fig. 1c). Remarkably, this is comparable to the ratio of N^6 -methyladenosine (m^6A , the most abundant internal mRNA modification in eukaryotes) to A (m^6A/A), which has been measured by different methods to be 0.1–0.5%^{26–29}. Moreover, we also found high levels of Ψ in mRNAs from various mouse tissues, with particular enrichment in the brain and lung (Fig. 1d). Altogether, these results suggest that pseudouridylation is prevalent in mammalian mRNA and is much more comprehensive than previously believed.

CeU-Seq pre-enriches and identifies thousands of Ψ sites

To comprehensively study pseudouridylation events in the transcriptome, we developed CeU-Seq. For CeU-Seq, we chemically synthesized a CMC derivative, **1** (Supplementary Fig. 2a), which reacts specifically with the Ψ -containing RNA; biotin was subsequently conjugated to the N_3 -CMC- Ψ RNA via click chemistry (Fig. 2a). Notably, no cross-reactivity was detected for uridine or inosine, both well-known internal mRNA modifications (Supplementary Fig. 2b)³⁰. We coupled such ability with high-throughput sequencing methods to map Ψ sites in both mRNAs and noncoding RNAs (Fig. 2b); the biotin pulldown feature of CeU-Seq effectively pre-enriches the Ψ -containing RNA (by ~ 15 - to 20-fold), as measured by spike-in controls (Fig. 2c). The higher percentage of reads ending before U bases in the N_3 -CMC(+) samples also indicates that CeU-Seq can effectively enrich Ψ -containing RNA (Supplementary Fig. 3a).

We first applied CeU-Seq to identify Ψ sites in rRNA. Two metrics were defined to evaluate the CeU-Seq data: ‘stop rate’ and ‘CMC sensitivity’ (Online Methods). Although the current purification protocol negatively selects for rRNA (only $\sim 0.16\%$ of reads are mapped to rRNA in the purified input poly(A)⁺ RNA), the biotin pulldown step of CeU-Seq allowed enrichment of rRNA and mapping of Ψ s. Indeed, the receiver operating characteristic analysis showed that N_3 -CMC(+) samples have a maximum accuracy of $\sim 92.7\%$ in detecting Ψ sites in rRNA (Fig. 2d). Besides reliably detecting known Ψ s in rRNA, to our surprise, CeU-Seq

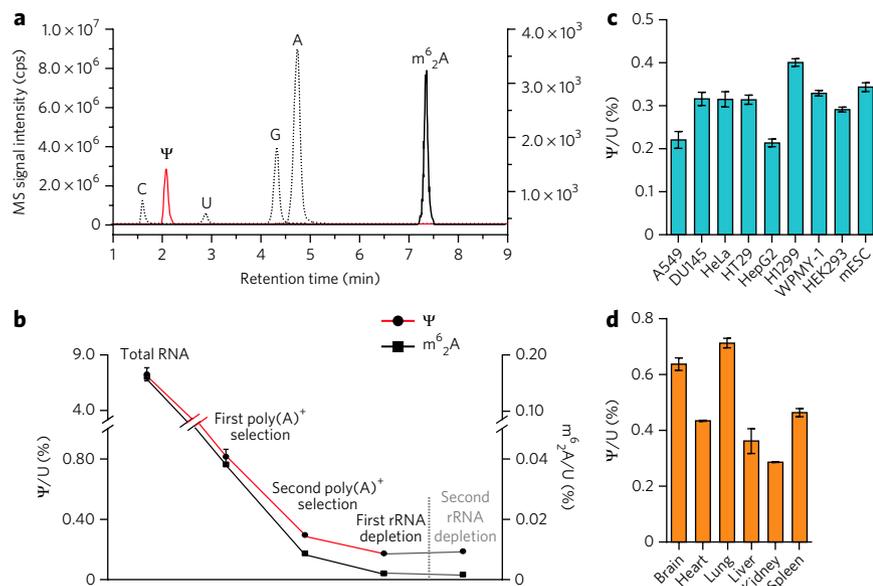


Figure 1 | Quantitative analysis of pseudouridylation in mammalian mRNA. (a) Representative LC/MS/MS chromatograms of purified mRNA samples. The dotted chromatograms of C, U, G and A are scaled to the left y axis, and the solid chromatograms of Ψ (red) and m^6_2A (black) are scaled to the right y axis. cps, counts per second. (b) Quantification of Ψ modification during every step of the RNA purification procedure. The stationary value of Ψ/U ratio after the first rRNA depletion step is indicated by the dotted gray line. (c) Quantification of Ψ modification in different cell lines. (d) Quantification of Ψ modification in various mouse tissues. In b–d, values represent mean \pm s.e.m. ($n = 4$).

repeatedly identified seven sites in rRNA that were previously not known to be pseudouridylated (Supplementary Fig. 3b,c). These sites are not immediately 5' to the known Ψ sites in rRNA, and hence they are not considered the result of ‘stuttering’²⁰. One possibility is that these sites could be previously unidentified new Ψ sites. To investigate the identities of such sites, we chose the ‘SCARLET’ method, which is based on RNase H but is CMC independent³¹. We picked three out of the seven Ψ sites for validation: U1045, U1177 of 18S rRNA and U1781 of 28S rRNA, which are respectively high, medium and low ranking within the list of identified sites (difference of stop rate: ~ 0.59 , ~ 0.40 and ~ 0.34 , respectively, compared to our threshold of 0.30). SCARLET clearly shows that all of these sites are pseudouridylated to $>90\%$ (Fig. 2e). Therefore, CeU-Seq not only detects known rRNA Ψ sites but also is capable of identifying new Ψ sites, thus demonstrating its sensitivity and robustness.

Having validated the unbiased ability of CeU-Seq to detect Ψ in rRNA, we next analyzed pseudouridylation events in the human transcriptome. Using both stop rate and CMC sensitivity, we identified 2,084 Ψ sites in human mRNA (1,889) and ncRNA (195, excluding rRNA) from HEK293T cells (Fig. 3a, Supplementary Figs. 4a,b and 5a,b and

Supplementary Data Set). Two previous reports also profiled pseudouridylation events in human cells: one report identified 96 Ψ sites in 89 human mRNAs from HeLa cells, whereas another report identified 396 Ψ sites combining HEK293 cells and fibroblasts (353 sites in mRNA and 43 sites in noncoding RNAs)^{21,22}. To validate mRNA pseudouridylation events identified by CeU-Seq, we performed SCARLET validation for U519 of *EEF1A1* mRNA³¹, which was identified as Ψ by CeU-Seq but not by previous works. Indeed, U519 is pseudouridylated to a high extent (~56% by thin-layer chromatography; **Fig. 3b**), representing what is to our knowledge the first experimentally validated human mRNA pseudouridylation site. Thus, in comparison with analyses that first reported human mRNA pseudouridylation, our results markedly enrich the understanding of mammalian pseudouridylation events in the transcriptome.

To predict the potential cellular processes that involve Ψ modification, we used the DAVID database to identify the gene ontology terms that are enriched for Ψ -containing transcripts. We found that pseudouridylated mRNAs are involved in various cellular functions, with enrichment in translation ($P = 1.23 \times 10^{-12}$), protein metabolism ($P = 1.09 \times 10^{-6}$), DNA replication ($P = 2.60 \times 10^{-4}$) and others (**Supplementary Fig. 6a**). We also characterized the distribution pattern of Ψ modification in mRNA. Ψ is distributed along the 5' UTR, coding DNA sequence and 3' UTR (**Fig. 3a**) but is underrepresented in the 5' UTR (**Fig. 3c** and **Supplementary Fig. 4a**). Within coding sequences, UUU (U in the first and second position but only in a few cases the third) and UUC codons, both encoding phenylalanine, are the most frequently modified (**Supplementary Fig. 6b,c**), in contrast with observations in yeast that the GUA codon is preferentially modified²¹.

lncRNAs are also frequently pseudouridylated. Out of the 195 Ψ sites identified in ncRNA, 170 Ψ sites are within lncRNAs (**Fig. 3a**). Notably, we identified pseudouridylation events in many disease-related lncRNAs: for example, in addition to the two recently identified Ψ sites (U5160 and U5590) in *MALAT1* (refs. 21,22), we identified U3374 as a new Ψ site; furthermore, U11249 of *XIST* and U64919 of *KCNQ1OT1* are also pseudouridylated (**Fig. 3d**).

CeU-Seq shows that hPUS1 acts on mRNA

To define the molecular basis for pseudouridylation, we searched for candidate Ψ synthases responsible for pseudouridylation in mRNA and ncRNA. In the human genome, 13 proteins with an annotated Ψ synthase domain have been found³², and both RNA-guided and RNA-independent Ψ synthases were actually found to be mRNA bound^{33,34}. To explore whether or not such candidate Ψ synthases can modify mRNA, we performed short hairpin RNA (shRNA) knock-down experiments for all 13 proteins (**Supplementary Fig. 7a,b**). We also overexpressed these proteins individually using a mammalian expression vector and measured mRNA Ψ content (**Supplementary Fig. 7c,d**). For instance, mRNA isolated from HEK293T cells with hPUS1 knockdown or overexpression showed a notable decrease (~10%) and increase (~1.9-fold) of Ψ , respectively

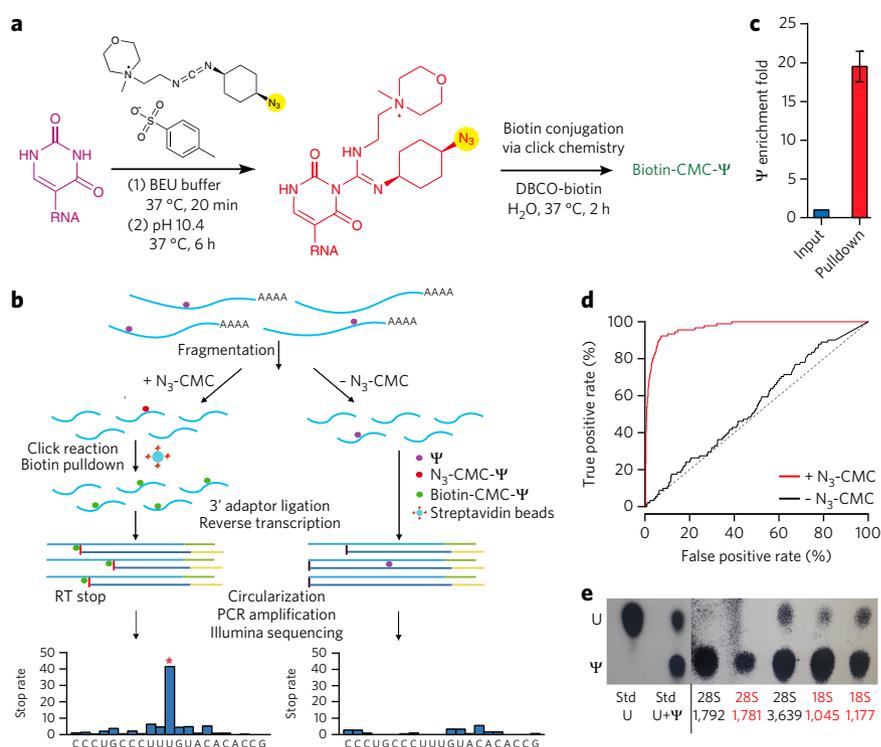


Figure 2 | CeU-Seq pre-enriches Ψ -containing RNA and identifies known and new

pseudouridylation sites in rRNA. (a) Chemically synthesized N₃-CMC specifically labels Ψ . Biotin can then be conjugated to N₃-CMC- Ψ . (b) Procedure of CeU-Seq and scoring Ψ sites. Biotin is conjugated to N₃-CMC- Ψ through click chemistry, and Ψ -containing RNA is enriched by a pull-down step. The biotin-N₃-CMC- Ψ adducts cause reverse transcriptase to stop one base downstream of Ψ modification. The resulting cDNA is then circularized, linearized, PCR amplified and subjected to high-throughput sequencing. 'Stop rate' of position *i* is calculated using the equation $N_{stop}^i / (N_{stop}^i + N_{readthrough}^i)$, where N_{stop}^i is the number of reads with mapping position starting at the *i*+1 base (one nucleotide 3' to position *i*), and $N_{readthrough}^i$ is the number of reads reading through position *i*. (c) N₃-CMC allows -15- to 20-fold enrichment for Ψ -containing RNAs, as measured by qRT-PCR. Values represent mean \pm s.e.m. ($n = 8$). (d) CeU-Seq detects Ψ sites in rRNA. (e) CeU-Seq detects both known and new Ψ sites in rRNA. The two lanes to the left of the black line show the positions of U and Ψ nucleotides on a thin-layer chromatography plate during SCARLET. To the right are five Ψ sites for validation: the two known Ψ sites (U1792 and U3639 from 28S rRNA, black) and the new Ψ sites (U1781 from 28S, and U1045, U1771 from 18S rRNA, red) are all highly Ψ modified. The full TLC images are shown in **Supplementary Figure 11**. Std, standard.

(**Supplementary Fig. 7a-d**). To further confirm the role of hPUS1 in modifying mRNA and ncRNA *in vivo*, we generated *PUS1*^{-/-} cell lines using CRISPR/Cas9 and identified 77 *PUS1*-dependent Ψ sites (a conservative minimum), including two new sites in a lncRNA named hSRA (Ψ 905 and Ψ 1346; **Fig. 3e**, **Supplementary Figs. 7e,f** and **8a** and **Supplementary Data Set**). *PUS1* is known to act on tRNA and hSRA³⁵, and inactivated mutants of hPUS1 are linked to human diseases¹⁰. Our results showing that hPUS1 also works on mRNA raises the possibility that such diseases could be a consequence of misregulation of the mRNA targets of hPUS1.

Unlike *PUS1*, which is structure-specific³⁶, *PUS4* and *PUS7* respectively recognize a 'GUUC' and 'UGUA' core consensus sequence for pseudouridylation^{37,38}. In our Ψ profile, we identified 214 and 125 targets for TRUB1 (*PUS4* homolog in human) and *PUS7* by such motifs, respectively (**Supplementary Fig. 8b**). Indeed, our knock-down and overexpression experiments suggest that they can target mRNA as well (**Supplementary Fig. 7a-d**). In fact, when searching for a potential consensus sequence for Ψ , we only observed a very modest sequence preference around the site of Ψ modification (**Supplementary Fig. 5b**). This could be the result of the combined effects of multiple Ψ synthases with different sequence preferences.

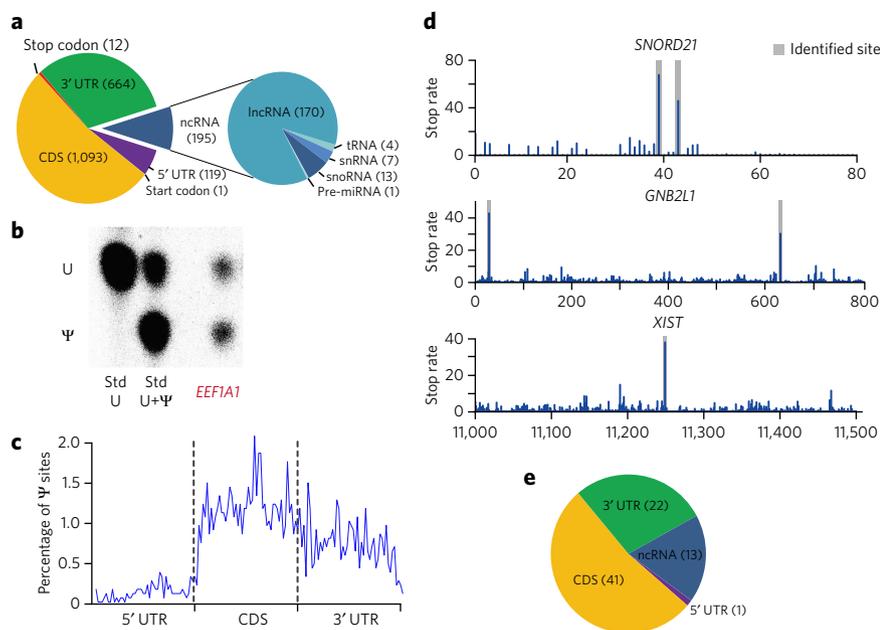


Figure 3 | CeU-Seq identifies thousands of pseudouridylation sites in human mRNA and ncRNA. (a) CeU-Seq identifies 2,084 Ψ sites in the human mRNA (1,889) and ncRNA (195). CDS, coding sequence. (b) U519 of *EEF1A1* mRNA, identified by CeU-Seq, is shown by SCARLET to be ~56% pseudouridylated. The full TLC images are shown in **Supplementary Figure 11**. (c) Ψ distribution in human mRNA. (d) Representative mRNA and ncRNA pseudouridylation events. Stop rate difference across *SNORD21*, *GNB2L1* (Ψ both in 5' UTR and CDS) and *XIST*; identified Ψ sites are marked with gray lines. (e) CeU-Seq identifies 77 PUS1 targets. PUS1 targets were detected by comparing CeU-Seq results of *PUS1*^{-/-} and wild-type HEK293T cells.

Ψ is dynamically regulated and stress specific

Emerging evidence reveals that RNA modifications change in response to stimuli, suggesting that RNA modifications are dynamically regulated^{11–13}. By analogy, we reasoned that the involvement of Ψ in mRNA regulation could be dynamically controlled in response to stimuli.

We tested our hypothesis by comparing the mRNA Ψ content of untreated cells to that of cells exposed to different stress conditions. Notably, heat shock, poly(I:C) and H₂O₂ treatment increased (by ~40–50%) the mRNA Ψ content, whereas starvation

and exposure to hepatocyte growth factor (HGF) mildly decreased (by ~10–15%) the Ψ content (**Fig. 4a**).

To further characterize the inducible pseudouridylation program, we compared the mRNA Ψ profiles of untreated cells to those of cells exposed to heat shock or treated with H₂O₂ (**Supplementary Data Set**). Although the two stimuli increased the mRNA Ψ content to a similarly high extent (**Fig. 4a**), the inducible Ψ sites unexpectedly displayed strong stimuli-specific patterns and were essentially nonoverlapping (**Fig. 4b–d**). Heat shock induced 464 Ψ sites in mRNAs that are highly enriched in transport- and localization-related pathways (nuclear transport ($P = 1.66 \times 10^{-4}$), cellular protein localization ($P = 6.26 \times 10^{-7}$) and others), whereas H₂O₂ treatment induced 477 Ψ sites that are strongly enriched in telomere- and chromatin-related functions (telomere maintenance ($P = 0.00018$), chromatin remodeling ($P = 0.0007$) and others; **Supplementary Fig. 9a,b**). It is worth noting that we ruled out differences in mRNA expression as an explanation for the stress-specific pseudouridylation pattern (**Supplementary Fig. 9c**). Altogether, these data show that mRNA pseudouridylation is dynamically regulated and exhibits strong stimuli-specific patterns.

Ψ profiles in mouse brain and liver

To determine potential evolutionary conservation and functional importance of Ψ modification, we applied CeU-Seq to mouse brain and liver mRNA. We identified 1,741 Ψ sites from the brain and 1,543 Ψ sites from the liver (**Fig. 5a,b** and **Supplementary Data Set**). Although we found 54 sites in mRNA that are both pseudouridylated in brain and liver, the Ψ profiles show many tissue-specific features. For instance, genes encoding Ψ -containing mRNAs in the brain are enriched in nervous system development ($P = 1.3 \times 10^{-6}$; *Ntrk2*, *Dclk1* and so on) and signal transduction ($P = 1.0 \times 10^{-6}$; *Glur2* and other receptors), whereas those in the liver are enriched in protein transport ($P = 1.3 \times 10^{-6}$; *Sec16b* and others) and

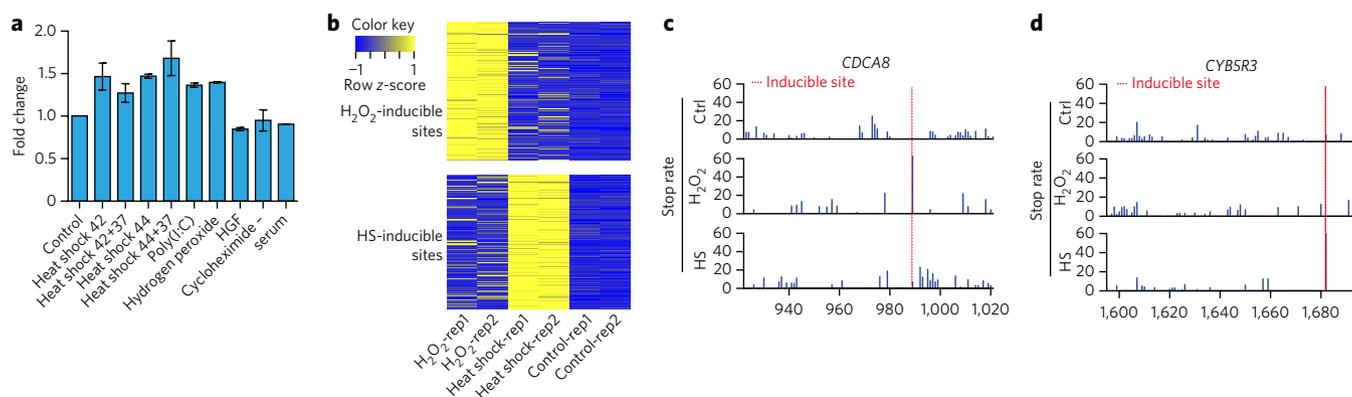


Figure 4 | Inducible pseudouridylation displays stimulus-specific patterns. (a) Ψ modification is highly inducible in response to various stimuli in human mRNA. The Ψ /U ratio was measured by LC/MS/MS, and the fold change (relative to the unstressed control) is plotted on the y axis. Values represent mean \pm s.e.m. ($n = 4$). HGF, hepatocyte growth factor; 'heat shock 42', 8 h incubation at 42 °C; 'heat shock 42+37', 8 h incubation at 42 °C followed by 4 h recovery at 37 °C (see Online Methods). (b) Stimulus-dependent Ψ sites are stimulus specific and dynamically modulated. Inducible sites are categorized into heat shock-inducible and H₂O₂-inducible sites in the heat map, in comparison to the stimulus-independent 'constitutive' sites. Std, standard. 'rep1' and 'rep2' refer to two replicates of the experiment. (c) Examples of H₂O₂-specific regulated Ψ sites in human mRNA. Red lines indicate the inducible sites. Ctrl, control. (d) Examples of heat shock-specific Ψ sites in human mRNA. Red lines indicate the inducible sites. CDS, coding sequence.

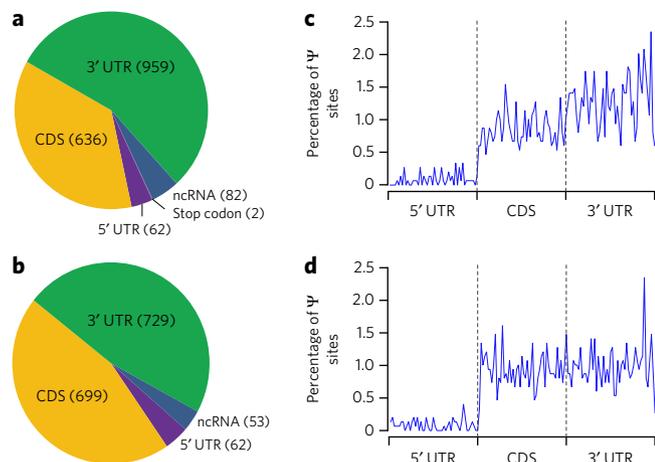


Figure 5 | Human and mouse pseudouridylation exhibits both shared and unique sites across tissues and species. (a) CeU-Seq identifies 1,741 Ψ sites in the mouse brain mRNA (1,659) and ncRNA (82). CDS, coding sequence. (b) CeU-Seq identifies 1,543 Ψ sites in the mouse liver mRNA (1,490) and ncRNA (53). (c) Ψ distribution in mouse brain mRNA. Ctrl, control. (d) Ψ distribution in mouse liver mRNA.

macromolecule localization ($P = 1.0 \times 10^{-6}$). In addition, many genes that are actively expressed both in brain and liver are selectively pseudouridylated only in one of the two tissues (Supplementary Fig. 10a–d). Between human and mouse samples, Ψ s have a similar pattern of distribution along mRNA (Figs. 3c and 5c,d), although the number of conserved Ψ sites are modest (Supplementary Fig. 10e and Supplementary Tables 5 and 6). Yet, Ψ profiles from the same tissue of human and mouse might have to be compared to rule out tissue-specific differences in pseudouridylation. Taken together, pseudouridylation displays tissue-specific features between mouse brain and liver and also exhibits shared sites and pattern between mouse and human.

DISCUSSION

Ψ was among the first discovered post-transcriptional modifications and is the most abundant base modification in RNA. However, the high prevalence of Ψ in mammalian mRNA is striking: the ~0.2–0.4% Ψ/U ratio measured in cell lines is reminiscent of the level of m^6A (m^6A/A ratio of ~0.5%)^{26,27}. Given such m^6A abundance, two recent analyses have identified 8,000–12,000 m^6A sites from 7,000–8,000 human genes^{39,40}. Comparatively, CeU-Seq identified 2,084 Ψ sites from 1,357 genes in HEK293T cells. It is likely that the current number of Ψ sites represents a conserved minimum of bona fide Ψ sites in the human transcriptome. Indeed, by relaxing our Ψ identification criteria (decreasing stop rate difference to >0.20 from >0.30), we readily increased the number of candidate Ψ sites to 4,955, a list that includes three additional known Ψ sites in human U2 (Ψ 43 and Ψ 58) and U5 (Ψ 46) snRNA.

CeU-Seq identified mRNA and ncRNA targets of structure-specific hPUS1 as well as hundreds of targets of PUS7 and TRUB1 by their preferred motifs. However, these Ψ synthases lack an apparent dedication to mRNA; hence, a Ψ synthase that potentially recognizes only the mRNA substrates remains to be identified. In fact, there are 13 proteins in human with a predicted Ψ synthase domain, many of which are still poorly characterized³². Future investigations of such candidates could shed light on this, potentially with the aid of existing Ψ profiling and quantification methods.

Our analysis revealed hundreds of mRNA pseudouridylation sites that are regulated by environmental cues in stress-specific manners. It is tempting to speculate that the dynamically regulated nature of mRNA Ψ modifications would allow them to

manifest as ‘epitranscriptomic marks’ that actively participate in gene regulation^{41,42}. For instance, it has yet to be determined whether or not Ψ could switch the local secondary structure of mRNA to indirectly affect binding of RNA binding proteins⁴³. Alternatively, given recent discoveries that the m^6A modification is reversible^{26,27}, with reader proteins to recognize it for gene regulation⁴⁴, it is also intriguing to infer that there could be readers and erasers for Ψ .

In summary, our study demonstrated that Ψ is a widespread and dynamically regulated mRNA modification. Our quantitative approaches revealed the prevalence of Ψ modification in the mammalian transcriptome and also enabled the evaluation of Ψ dynamics. Our chemically assisted pulldown method allowed the comprehensive profiling of Ψ in the human and mouse transcriptome, providing a reference and tool for future investigations to elucidate the biological roles of Ψ in mRNA and ncRNA.

Received 17 January 2015; accepted 29 April 2015; published online 15 June 2015

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Gene Expression Omnibus. Sequencing data have been deposited under the accession number GSE63655.

References

- Machnicka, M.A. *et al.* MODOMICS: a database of RNA modification pathways—2013 update. *Nucleic Acids Res.* **41**, D262–D267 (2013).
- Charette, M. & Gray, M.W. Pseudouridine in RNA: what, where, how, and why. *IUBMB Life* **49**, 341–351 (2000).
- Ge, J. & Yu, Y.T. RNA pseudouridylation: new insights into an old modification. *Trends Biochem. Sci.* **38**, 210–218 (2013).
- Jack, K. *et al.* rRNA pseudouridylation defects affect ribosomal ligand binding and translational fidelity from yeast to human cells. *Mol. Cell* **44**, 660–666 (2011).
- Yu, A.T., Ge, J. & Yu, Y.T. Pseudouridines in spliceosomal snRNAs. *Protein Cell* **2**, 712–725 (2011).
- Kierzek, E. *et al.* The contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic Acids Res.* **42**, 3492–3501 (2014).
- Kiss, T., Fayet-Lebaron, E. & Jady, B.E. Box H/ACA small ribonucleoproteins. *Mol. Cell* **37**, 597–606 (2010).
- Hamma, T. & Ferre-D’Amare, A.R. Pseudouridine synthases. *Chem. Biol.* **13**, 1125–1135 (2006).
- Heiss, N.S. *et al.* X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nat. Genet.* **19**, 32–38 (1998).
- Bykhovskaya, Y., Casas, K., Mengesha, E., Inbal, A. & Fischel-Ghodsian, N. Missense mutation in pseudouridine synthase 1 (*PUS1*) causes mitochondrial myopathy and sideroblastic anemia (MLASA). *Am. J. Hum. Genet.* **74**, 1303–1308 (2004).
- Wu, G., Xiao, M., Yang, C. & Yu, Y.T. U2 snRNA is inducibly pseudouridylated at novel sites by Pus7p and snR81 RNP. *EMBO J.* **30**, 79–89 (2011).
- Courtes, F.C. *et al.* 28S rRNA is inducibly pseudouridylated by the mTOR pathway translational control in CHO cell cultures. *J. Biotechnol.* **174**, 16–21 (2014).
- Basak, A. & Query, C.C. A pseudouridine residue in the spliceosome core is part of the filamentous growth program in yeast. *Cell Rep.* **8**, 966–973 (2014).
- Cavaillé, J. *et al.* Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Natl. Acad. Sci. USA* **97**, 14311–14316 (2000).
- Hüttenhofer, A., Brosius, J. & Bachellerie, J.P. RNomics: identification and function of small, non-messenger RNAs. *Curr. Opin. Chem. Biol.* **6**, 835–843 (2002).
- Karjolic, J. & Yu, Y.-T. Converting nonsense codons into sense codons by targeted pseudouridylation. *Nature* **474**, 395–398 (2011).
- Fernández, I.S. Unusual base pairing during the decoding of a stop codon by the ribosome. *Nature* **500**, 107–110 (2013).
- Ho, N.W. & Gilham, P.T. The reversible chemical modification of uracil, thymine, and guanine nucleotides and the modification of the action of ribonuclease on ribonucleic acid. *Biochemistry* **6**, 3632–3639 (1967).

19. Bakin, A. & Ofengand, J. Four newly located pseudouridylate residues in *Escherichia coli* 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. *Biochemistry* **32**, 9754–9762 (1993).
20. Bakin, A.V. & Ofengand, J. Mapping of pseudouridine residues in RNA to nucleotide resolution. *Methods Mol. Biol.* **77**, 297–309 (1998).
21. Carlile, T.M. *et al.* Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **515**, 143–146 (2014).
22. Schwartz, S. *et al.* Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148–162 (2014).
23. 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).
24. Hughes, D.G. & Maden, B.E. The pseudouridine contents of the ribosomal ribonucleic acids of three vertebrate species. Numerical correspondence between pseudouridine residues and 2'-O-methyl groups is not always conserved. *Biochem. J.* **171**, 781–786 (1978).
25. Maden, B.E. The numerous modified nucleotides in eukaryotic ribosomal RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **39**, 241–303 (1990).
26. Jia, G. *et al.* N₆-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7**, 885–887 (2011).
27. Zheng, G. *et al.* ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* **49**, 18–29 (2013).
28. Desrosiers, R., Friderici, K. & Rottman, F. Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc. Natl. Acad. Sci. USA* **71**, 3971–3975 (1974).
29. Dubin, D.T. & Taylor, R.H. The methylation state of poly A-containing messenger RNA from cultured hamster cells. *Nucleic Acids Res.* **2**, 1653–1668 (1975).
30. Nishikura, K. Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.* **79**, 321–349 (2010).
31. Liu, N. *et al.* Probing N₆-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA* **19**, 1848–1856 (2013).
32. Hunter, S. *et al.* InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res.* **40**, D306–D312 (2012).
33. Castello, A. *et al.* Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–1406 (2012).
34. Baltz, A.G. *et al.* The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* **46**, 674–690 (2012).
35. Zhao, X. *et al.* Regulation of nuclear receptor activity by a pseudouridine synthase through posttranscriptional modification of steroid receptor RNA activator. *Mol. Cell* **15**, 549–558 (2004).
36. Sibert, B.S. & Patton, J.R. Pseudouridine synthase 1: a site-specific synthase without strict sequence recognition requirements. *Nucleic Acids Res.* **40**, 2107–2118 (2012).
37. Becker, H.F., Motorin, Y., Planta, R.J. & Grosjean, H. The yeast gene *YNL292w* encodes a pseudouridine synthase (Pus4) catalyzing the formation of Ψ55 in both mitochondrial and cytoplasmic tRNAs. *Nucleic Acids Res.* **25**, 4493–4499 (1997).
38. Behm-Ansmant, I. *et al.* The *Saccharomyces cerevisiae* U2 snRNA:pseudouridine-synthase Pus7p is a novel multisite-multisubstrate RNA:Psi-synthase also acting on tRNAs. *RNA* **9**, 1371–1382 (2003).
39. Dominissini, D. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201–206 (2012).
40. Meyer, K.D. *et al.* Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**, 1635–1646 (2012).
41. He, C. Grand challenge commentary: RNA epigenetics? *Nat. Chem. Biol.* **6**, 863–865 (2010).
42. Fu, Y., Dominissini, D., Rechavi, G. & He, C. Gene expression regulation mediated through reversible m₆A RNA methylation. *Nat. Rev. Genet.* **15**, 293–306 (2014).
43. Liu, N. *et al.* N₆-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* **518**, 560–564 (2015).
44. Wang, X. *et al.* N₆-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117–120 (2014).

Acknowledgments

The authors would like to thank R. Meng and S. Huang for technical assistance and N. Liu, X. Wang and X. Liu for discussion. This work was supported by the National Basic Research Foundation of China (no. 2014CB964900) and the National Natural Science Foundation of China (nos. 21472009 and 31270838).

Author contributions

X.L., S.M. and C.Y. conceived the project, designed the experiments and wrote the manuscript. X.L. and S.M. performed the experiments. P.Z. designed and performed the bioinformatics analysis; J.S. contributed in making figures, J.B. synthesized the N₃-CMC derivative, and F.S. contributed in Ψ quantification. All authors commented on and approved the paper.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Supplementary information and chemical compound information is available in the online version of the paper. Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Correspondence and requests for materials should be addressed to C.Y.

ONLINE METHODS

Cell culture and mouse tissues. HEK293T, HEK293, A549, DU145, HeLa, HT29, HepG2, H1299, WPMY-1 cells and mESCs were used in this study. HEK293T, DU145, HeLa, HT29, HepG2 and WPMY-1 cells were maintained in DMEM medium (Life Technologies) supplemented with 10% FBS and 1% penicillin/streptomycin. A549 and H1299 were maintained in RPMI 1640 medium (Life Technologies) supplemented with 10% FBS and 1% penicillin/streptomycin. Mouse ESCs were plated on 10-cm dishes coated with gelatin and grown in GMEM (Life Technologies) supplemented with 10% FBS, nonessential amino acids, glutamine, pyruvate, β -mercaptoethanol and leukemia inhibitory factor⁴⁵. Brains, hearts, lungs, livers and kidneys were isolated from adult C57BL/6 mice.

Antibodies. Polyclonal rabbit anti-PUS1 antibody was affinity purified from rabbits serum immunized with recombinant His₆-tagged human PUS1 protein (dilution 1:2,000 for western blotting). Rabbit anti-HA tag (ab91110; Abcam) was purchased from commercial sources. The secondary antibodies used for western blotting were anti-rabbit-IgG-HRP (CW0103; CWBiotech).

Plasmid construction. Candidate mRNA pseudouridine synthases were cloned from HEK293T cDNA into a modified pCDNA3.1(+) vector (Life Technologies), which has an additional 3' HA tag after the XbaI recognition site. The detailed primers list can be found in **Supplementary Table 2**. For transfection, the recombinated plasmids were extracted using EasyPure Plasmid MiniPrep Kit (Transgen Biotech) followed by ethanol precipitation. Plasmids were transfected into HEK293T by lipofectamine 2000 (Life Technologies) according to the manufacturer's instructions.

Lentivirus-mediated shRNA knockdown. All of the pLKO vectors were obtained from the TRC shRNA library (<http://www.broadinstitute.org/rnai/public/>). Lentiviruses were packaged by cotransfecting HEK293T cells with targeting plasmids together with pCMV-dR8.91 and VSV-G plasmids, following the instructions from Broad Institute. The supernatants from transfected cells were harvested after 2 d and used to infect HEK293T cells. Knockdown cells were selected by puromycin for at least 1 week. Knockdown efficiency was verified by qRT-PCR. The detailed sequences of shRNA and the primers used for qRT-PCR can be found in **Supplementary Tables 3 and 4**.

Procedure of poly(A)⁺ RNA isolation. Total RNA was isolated from cell lines or mouse tissues with TRIzol according to the manufacturer's instructions (Life Technologies). For poly(A)⁺ RNA purification, two successive rounds of poly(A)⁺ selection were performed using oligo(dT)₂₅ dynabeads (Life Technologies), which were followed by one round of rRNA depletion using Mouse/Human RiboMinus kit (Life Technologies). Small RNA depletion before the second round of poly(A)⁺ selection using the mirVana kit (Life Technologies) and a second round of rRNA depletion were optional. The concentration of mRNA was measured using a Nanodrop 2000 (Thermo Fisher Scientific). The quality of mRNA was checked using RNA 6000 Pico Chips on an Agilent 2100 Bioanalyzer (Agilent Technologies).

Quantification of Ψ by LC/MS/MS. 200 ng mRNA was digested by nuclease P1 (1 U, Sigma, N8630) in 20 μ l buffer containing 10 mM ammonium acetate, pH 5.3 at 42 °C for 6 h followed by the addition of 2 μ l 1 M ammonium bicarbonate and alkaline phosphatase (0.5 U, Sigma, P4252). The mixture was incubated at 37 °C for another 6 h and diluted to 50 μ l. 10 μ l of the solution was injected into LC/MS/MS. The nucleosides were separated by ultra-performance LC on a C18 column and then detected by triple-quadrupole MS (AB SCIEX QTRAP 5500) in the positive ion multiple reaction-monitoring (MRM) mode. The mass transitions of m/z 245.0 to 179.1 (Ψ), m/z 296.0 to 164.1 (m^6A) and m/z 245.0 to 113.1 (U) were monitored and recorded. These nucleosides were quantified according to the standard curve running at the same batch of samples. The Ψ/U and m^6A/U ratios were calculated.

Chemically synthesized N₃-CMC specifically labels Ψ . To test the reaction specificity of chemically synthesized N₃-CMC, three synthetic RNA molecules were used in this study: oligo- Ψ , 5'-GCGAGAG Ψ ACGC-3'; oligo-U, 5'-GCGAGAGUACGC-3'; oligo-I, 5'-GCGAGAGIACGC-3'. These three oligos were reacted with 0.2 M N₃-CMC in BEU buffer (7 M urea, 4 mM EDTA, 50 mM Bicine, pH 8.5) at 37 °C for 20 min followed by ethanol precipitation.

The RNA pellet was dissolved in sodium carbonate buffer (50 mM Na₂CO₃, 2 mM EDTA, pH 10.4) and incubated at 37 °C for 6 h to remove N₃-CMC from U and G residues. RNA was purified by ethanol precipitation. Click chemistry was performed by adding 20 mM DBCO-(PEG)₄-biotin (Sigma) to a final concentration of 150 μ M and incubating at 37 °C for 2 h. All of the reaction products of each step were purified by ethanol precipitation and then analyzed on a 20% urea-PAGE.

Spike-in RNA for estimation of pulldown enrichment. With our method, RNA containing Ψ can be labeled with biotin specifically and enriched by streptavidin pulldown. Two synthetic RNA containing a single U or Ψ were described *in vitro* from dsDNA templates to estimate the biotin pulldown enrichment. Ψ -only RNA template: GAAATTAATACGACTCACTATAGGG GAAGAGTACGCCGAAAGAGACACGGACAAGAAGAAAAGAGAGCCA GCACACGAAAACCAAGCAGCAACGAGACAGAACAGGA; U-only RNA template: GAAATTAATACGACTCACTATAGGGGAAGAGTACGCCACAGAA CAACCGAGCGCGAAAGAAAAGAGAGCCAGCACACGAAAACCAAGA CGAACCAGGACCAAGGACC, where Ψ represents Ψ or U. *In vitro* transcription was performed using a MAXIscript T7 kit (Ambion, Life Technologies) followed by a DNase treatment using Turbo DNase I (Ambion, Life Technologies). Ψ -only RNA and U-only RNA were reacted with N₃-CMC followed by click chemistry. The reacted RNAs were used as input and then pulled down by streptavidin beads. Input and pulled-down RNA were reverse transcribed into cDNA using RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) and quantified by qPCR using SYBR GREEN mix (Takara) on an ABI ViiA 7 real-time PCR system. Primer sequences: Ψ -RNA (FWD: CGAAAGAGACACGGACAAGAAG, RVS: GTTCTGTCTCGTTGCTGCTT), U-RNA (FWD: AGAACAACCGAGCGCGAAAG, RVS: GGTCTTGGTC CTGGTTCGT).

CeU-Seq. 10 μ g mRNA was digested into ~100- to 150-nt fragments using magnesium RNA fragmentation buffer (New England Biolabs). Fragmented mRNA was desalted and concentrated by ethanol precipitation. Biotin-CMC labeling of Ψ was performed as follows. The mRNA was heated at 80 °C in 5 mM EDTA for 5 min, followed by chilling on ice. Then, 10 μ l denatured mRNA was added to 100 μ l 0.2 M chemically synthesized N₃-CMC in BEU buffer (7 M urea, 4 mM EDTA, 50 mM Bicine, pH 8.5) followed by incubation at 37 °C for 20 min. The remaining N₃-CMC was removed by ethanol precipitation. The RNA pellet was dissolved in 50 μ l sodium carbonate buffer (50 mM Na₂CO₃, 2 mM EDTA, pH 10.4) and incubated at 37 °C for 6 h to remove N₃-CMC from U and G residues. RNA was purified by ethanol precipitation and dissolved in 40 μ l H₂O. Click chemistry was performed with the addition of 20 mM DBCO-(PEG)₄-biotin (Sigma) to a final concentration of 150 μ M, followed by incubation at 37 °C for 2 h. RNA was purified by ethanol precipitation and dissolved in 20 μ l H₂O. The mock sample was treated in the same way but without addition of chemically synthesized N₃-CMC.

The reacted RNA was pulled down by Streptavidin C1 dynabeads (Life Technologies). 50 μ l C1 beads were washed and resuspended in 200 μ l binding buffer (500 mM NaCl, 1 mM EDTA, 10 mM Tris, pH 7.5). 20 μ l RNA was added into resuspended beads and incubated at room temperature for 30 min with gentle rotation. C1 beads were washed four times with wash buffer (150 mM NaCl, 1 mM EDTA, 10 mM Tris, pH 7.5) and once with 1 \times PNK wash buffer (10 mM MgCl₂, 0.2% Tween-20, 20 mM Tris, pH 7.5).

Library construction. Library construction was performed according to the iCLIP library construction protocol with some modifications⁴⁶. For dephosphorylation of 3' ends, C1 beads were treated with shrimp alkaline phosphatase (New England Biolabs) and incubated at 37 °C for 30 min. The sample was then washed twice with high-salt wash buffer (1 M NaCl, 1 mM EDTA, 0.1% SDS, 0.5% sodium deoxycholate, 1% NP-40, 50 mM Tris-HCl, pH 7.5) and once with PNK wash buffer and RNA ligation buffer. 3' RNA linker (5'-phosphate-UGAGAUCG GAAGAGCGGUUCAG-3'-Puromycin) ligation was performed with T4 RNA ligase 1 (New England Biolabs) at 16 °C overnight. Samples were washed 4 times with wash buffer to remove excess adaptor. Enriched reacted RNA was eluted and purified by ethanol precipitation. Reverse transcription was performed with Superscript III reverse transcriptase (Life Technologies). The sequence of the barcoded RT primer is 5'-phosphate-NNXXXXXXXXNNA GATCGGAAGAGCGTCGTGGATCCTGAACCGCTC-3' (where XXXX represents the barcode). The cDNA was purified with 6% UREA-PAGE to remove

excess RT primers. Purified cDNA was circ-ligated with CircLigase II (Epicentre) followed by ethanol precipitation. Linearization of cDNA was performed by annealing with a cut oligo (5'-GTTCCAGGATCCACGACGCTCTTCAAAA-3') followed by BamHI digestion. The cDNA was amplified by PCR with primers (5'-CAAGCAGAAGACGGCATAACGATCGGTCTCGGCATTCCTGCTG AACCGCTCTCCGATCT-3'; 5'-AATGATACGGCGACCACCGAGATCTA CACTCTTCCCTACACGACGCTCTTCCGATCT-3'). PCR products were purified using an 8% TBE gel and sequenced on Illumina Hiseq 2000 or Hiseq 2500 with single end reads (50 bp).

CeU-Seq data preprocessing and reads mapping. Raw sequencing data were demultiplexed to each sample according to their barcode information. Trim_galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was used to remove or trim adaptor-containing reads and low-quality reads, and the minimum length of trimmed reads was set at 20 bases. Two bases after the barcode added during library construction at the 5' end of inserted sequences were also removed. Remaining reads were mapped to the human transcriptome (hg19, downloaded from University of California–Santa Cruz (UCSC) genome tables) or the mouse transcriptome (mm9, downloaded from UCSC genome tables) using BWA⁴⁷. To increase the accuracy of alignment, mapped reads with more than two mismatches to the reference sequence and reads mapped to transcripts belonging to more than one gene were further altered.

Definition of stop rate and CMC sensitivity. For any given position i on a reference transcript, the stop rate of position i was calculated using the equation $N_{\text{stop}}^i / (N_{\text{stop}}^i + N_{\text{readthrough}}^i)$, where N_{stop}^i (stop reads) is the number of reads with the mapping position starting at base $i+1$ (one nucleotide 3' to position i), and $N_{\text{readthrough}}^i$ (readthrough reads) is the number of reads reading through position i . A stop rate can be calculated for any given position (A, C, G or U) on a transcript.

Our definition of CMC sensitivity was adopted from that in recent work profiling the *in vivo* secondary structure of RNA using DMS⁴⁸. CMC sensitivity for position i is calculated using the equation $P_{\text{norm}}^i - M_{\text{norm}}^i$, where P_{norm}^i and M_{norm}^i are the normalized stop reads for site i in the N_3 -CMC(+) and N_3 -CMC(-) samples, respectively. The difference is then further normalized using the 2–8% normalization method⁴⁸ to give the final CMC sensitivity for position i .

Identification of Ψ . We defined a position i to be Ψ only when all of the following criteria were met: (i) genes of relatively low expression (reads per kb per million mapped reads (RPKM) <0.1) are not used for analysis; (ii) the stop reads of position i (N_{stop}^i) must be no less than 5 in the N_3 -CMC(+) sample; (iii) the stop rate in N_3 -CMC(-) samples must be less than 0.10; (iv) the difference of stop rate for position i between the N_3 -CMC(+) samples and the matched N_3 -CMC(-) samples must be at least 0.30; (v) the difference of stop rate of position i must be at least twofold greater than that of the two flanking bases (position $i-1$ and $i+1$). The Ψ identification criteria can be explained by the following reasons: (i) A minimum stop read of position i (N_{stop}^i) in the N_3 -CMC(+) sample is required to minimize random stopping due to RNA fragmentation process. (ii) A high stop rate in N_3 -CMC(-) samples (probably due to potential RNA secondary structures) will cause inaccurate estimation for the difference of stop rate. Therefore, we excluded all sites with a stop rate >0.10 for more confident identification of Ψ sites. (iii) The difference in stop rate of position i with that of the $i-1$ and $i+1$ positions must be compared to make sure that we exclude sites arising owing to the 'stuttering' of reverse transcriptase when encountering a CMC- Ψ site. (iv) We required that a site had to be identified in at least two independent replicates in order to be included in our final list of Ψ sites. (v) The identified Ψ sites were further evaluated by 'CMC sensitivity'.

Validation of Ψ sites by SCARLET. The validation of Ψ sites by SCARLET was performed as previously described³¹, with a few modifications. The sequences

of all chimera oligos and splint oligos used for this experiment can be found in **Supplementary Table 1**. Briefly, 15 μg poly(A)⁺ mRNA was annealed with 20 pmol corresponding chimera oligo in 2 μl 30 mM Tris-HCL, pH 7.5, by heating at 95 °C for 2 min. The sample was incubated at 44 °C for 1 h in 2 μl 1 \times T4 PNK buffer (New England Biolabs) supplemented with 1 U/ μl RNase H (Epicentre) and 1 U/ μl FastAP (Thermo Fisher Scientific), followed by incubation at 75 °C for 10 min. The RNA was incubated with 1 μl radioactive labeling buffer (1 \times T4 PNK buffer, 6 U/ μl T4 PNK, 150 $\mu\text{Ci}/\mu\text{l}$ [γ -³²P]ATP) at 37 °C for 1 h and then at 75 °C for 15 min. The mixture was then annealed with 10 pmol corresponding splint oligos and 10 pmol 116-mer DNA oligos by heating at 75 °C for 5 min followed by addition of 2.5 μl ligation buffer (1.4 \times PNK buffer, 0.3 mM ATP, 57% DMSO, 5 U/ μl T4 DNA ligase) and incubation for 3.5 h at 37 °C. The ligation was stopped by adding an equal volume of RNA urea loading buffer (8 M urea, 100 mM EDTA) and then was digested by incubated with 1 μl RNase T1/A mixture (Thermo Fisher Scientific) at 37 °C for 1 h. The sample was loaded on 10% denaturing urea-PAGE to isolate the ligation product, which was desalted by ethanol precipitation. The pellet was resuspended in 5 μl sodium acetate/acetic acid buffer, pH 4.8, supplemented with 2 U/ μl nuclease P1 and then was spotted on a TLC plate for separation. The result was visualized by a phosphor imager. The other seven candidate Ψ sites for validation showed either low specificity of hybridization or no detectable radioactive signals during SCARLET.

Generation of $PUS1^{-/-}$ 293T cells with CRISPR/Cas9. The plasmid PX330 containing two expression cassettes, hSpCas9 and the chimeric guide RNA was used in this study. The guide RNA was designed according to the method described before⁴⁹ and the target sequence of *PUS1* is GAGCCGCATGCCCCAGGAC. The plasmid containing the guide RNA sequence was transfected into HEK293T cells using lipofectamine 2000 (Life Technologies) according to the manufacturer's instructions. Cells were diluted and seeded into 96-well dish 48 h after transfection according to the limiting dilution protocol. After 10 d, single colonies were picked and detected by western blotting.

Stress conditions. Different stress conditions were used in this study as follows. (i) Mild heat shock: cells were incubated at 42 °C for 8 h, followed with or without 4 h recovery at 37 °C. (ii) Severe heat shock: cells were incubated at 44 °C for 2 h, followed with or without 4 h recovery at 37 °C. (iii) Drug treatment: cells were incubated with hydrogen peroxide (50 μM) for 30 min or cycloheximide (40 $\mu\text{g}/\text{ml}$) for 4 h or hepatocyte growth factor (10 ng/ml) for 6 h, respectively. (iv) Serum starvation: cells were maintained with serum-free DMEM medium overnight before RNA isolation.

Motif analysis and gene ontology. For motif analysis, the 10-bp sequence surrounding Ψ sites were used and analyzed using WebLogo with default settings (<http://weblogo.berkeley.edu/logo.cgi/>). All Gene Ontology analyses were performed using the DAVID bioinformatics database with default settings (<http://david.abcc.ncifcrf.gov/>).

45. Tremml, G., Singer, M. & Malavarca, R. Culture of mouse embryonic stem cells. *Curr. Protoc. Stem Cell Biol.* **5**, 1C4 (2008).
46. König, J. *et al.* iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* **17**, 909–915 (2010).
47. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
48. Ding, Y. *et al.* *In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505**, 696–700 (2014).
49. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).