Pseudouridylation meets next-generation sequencing

Maryam Zaringhalam [a,b,*], F. Nina Papavasiliou [a,*]

[a] Laboratory of Lymphocyte Biology, The Rockefeller University, 1230 York Avenue, New York, NY 10065, United States
[b] The Rockefeller Graduate Program, The Rockefeller University, 1230 York Avenue, New York, NY 10065, United States

ARTICLE INFO

ABSTRACT

The isomerization of uridine to pseudouridine (Ψ), known as pseudouridylation, is the most abundant post-transcriptional modification of stable RNAs. Due to technical limitations in pseudouridine detection methods, studies on pseudouridylation have historically focused on ribosomal RNAs, transfer RNAs, and spliceosomal small nuclear RNAs, where Ψs play a critical role in RNA biogenesis and function. Recently, however, a series of deep sequencing methods—Pseudo-seq, Ψ-seq, PSI-seq, and CeU-seq—has been published to map Ψ positions across the entire transcriptome with single nucleotide resolution. These data have greatly expanded the catalogue of pseudouridylated transcripts, which include messenger RNAs and noncoding RNAs. Furthermore, these methods have revealed conditionally-dependent sites of pseudouridylation that appear in response to cellular stress, suggesting that pseudouridylation may play a role in dynamically modulating RNA function. Collectively, these methods have opened the door to further study of the biological relevance of naturally occurring Ψs. However, an in-depth comparison of these techniques and their results has not yet been undertaken despite all four methods relying on the same basic principle: Ψ detection through selective chemical labeling by the carbodiimide known as CMC. In this article, we will outline the currently available high-throughput Ψ-detection methods and present a comparative analysis of their results. We will then discuss the merits and limitations of these approaches, including those inherent in CMC conjugation, and their potential to further elucidate the function of this ubiquitous and dynamic modification.

© 2016 Elsevier Inc. All rights reserved.

Contents

* Address: Laboratory of Lymphocyte Biology, The Rockefeller University, 1230 York Avenue, New York, NY 10065, United States (M. Zaringhalam).
  E-mail addresses: mzaringhal@rockefeller.edu (M. Zaringhalam), papavasiliou@rockefeller.edu (F.N. Papavasiliou).

1. Pseudouridine: an intriguing post-transcriptional modification

The most abundant of over 100 types of post-transcriptional modifications, pseudouridine (Ψ) was the first to be discovered and is often referred to as "the fifth ribonucleoside" [12,14,37]. Ψ is the C5-glycoside isomer of uridine that results when the N1–C1′ bond linking the uracil base to the ribose sugar is broken. The base is then rotated 180° around the N3–C6 axis and a non-canonical C5–C1′ glycosidic bond is formed (Fig. 1a) [11]. Site-specific pseudouridylation is catalyzed by pseudouridine synthases (PUSs) through one of two distinct mechanisms: a protein-only (RNA-independent) mechanism and a box H/ACA snoRNP-catalyzed (RNA-dependent) mechanism [50]. RNA-independent pseudouridylation is catalyzed by a single enzyme that specifically recognizes its particular substrate, either through a consensus motif or secondary structure [6,7,36,49]. On the other hand, RNA-dependent pseudouridylation is mediated by an RNA-protein (RNP) complex, consisting of four core proteins, including the pseudouridine-synthase Cbf5/dyskerin, assembled on a box H/ACA small nucleolar RNA (snoRNA) scaffold. The resulting snoRNP complex is guided to a particular RNA substrate by a unique H/ACA snoRNA, which contains a short guide sequence that recognizes and binds pseudouridylation targets by base complementarity [22,42].

Following isomerization, the Watson-Crick edge of uridine remains unchanged, allowing for Ψ-A base pairing. Nevertheless, the resulting Ψ has unique physiochemical properties, which are largely attributed to the additional hydrogen bond donor at the N1 position. Specifically, Ψ's N1H group adds rigidity to RNA structure through water coordination, increases base stacking with additional hydrogen bonds between the base and its phosphodiester backbone, and has been reported to increase Ψ/A base-pairing stability compared to U/A [1,11,41]. Ψ's additional hydrogen bond donor has also been thought to contribute to novel base-pairing interactions in Ψ-containing RNAs [11,44].

Pseudouridine's distinct structural properties make it unsurprising that Ψs are well known to cluster in evolutionarily conserved and functionally important regions of stable noncoding RNAs. Over the years, appreciation for the significant role pseudouridylation plays in RNA function has grown. Ψ's functional relevance has been well-documented in ribosomal RNAs (rRNAs), where pseudouridylation is required for ribosome biogenesis and translational fidelity, and in small nuclear RNAs (snRNAs), where specific Ψ residues have been identified as necessary for proper pre-mRNA splicing [5,27,33,56,58]. Furthermore, many pseudouridine residues in rRNAs and snRNAs are conserved across species, occurring at identical or near-identical sites [15,55]. More recently, pseudouridylation has been found to be inducible in response to cellular stress and differentiation, suggesting pseudouridylation may provide a dynamic regulatory mechanism for RNA function [4,13,39,54].

Further adding to pseudouridine's intrigue are reports on the effect of pseudouridylated mRNAs on translation. In vitro-transcribed mRNAs in which every U residue is pseudouridylated exhibit enhanced stability and translation efficiency when delivered in vivo [29]. In addition, artificially targeting pseudouridylation to specific U residues within nonsense codons converts them to sense codons in Saccharomyces cerevisiae (budding yeast) [28]. Ψ's recoding potential is strengthened by structural studies that demonstrate the ribosome can accommodate non-canonical codon-anticodon base pairing mediated by a pseudouridylated sense codon [19]. The decoding center's plasticity suggests that Ψ may similarly recode sense codons, generating additional protein diversity. Important to note, however, is that fully pseudouridylated mRNAs synthesized by Karikó et al. were translated into functional proteins (i.e. GFP, lacZ, and luciferase). While the protein products were not sequenced to determine if Ψ facilitated alternate amino acid incorporation, the likelihood that a functional protein would result from multiple codon recoding events is low, suggesting that Ψ content could perhaps play a role in Ψ-mediated recoding. Nevertheless, mRNA pseudouridylation clearly appears to impact the translational machinery

In contrast to the growing body of work pointing to the biological functions of pseudouridylation, further inquiry was limited by the available methods for site-specific Ψ detection. These methods were notably low-throughput, required knowledge of the specific Ψ-containing sequence, and were best suited to highly abundant transcripts. For instance, despite pseudouridine's recoding potential, pseudouridylation of native mRNA transcripts had never been observed. Elucidating the role of pseudouridylation in naturally occurring RNAs would therefore require the development of a high-throughput, unbiased, and sensitive approach to identify Ψs. Fortunately, pseudouridine's potential significance catalyzed the development of four independent RNA sequencing (RNA-seq) techniques for Ψ detection, each taking advantage of the selective labeling of Ψ by the chemical N-cyclohexyl-N′-(2-morpholinoethyl)-carbodiimide metho-p-toluenesulfonate (CMC) (Fig. 2a) [3]. Pseudo-seq, Ψ-seq, PSI-seq, and CeU-seq have all independently confirmed the ubiquity of pseudouridylation in a diverse population of RNA species, including mRNA [9,32,35,47]. Each study also confirmed the existence of a condition-dependent set of additional pseudouridylation events in response to environmental cues in yeast, mice, and human cells. Together, these methods have broadened the scope of Ψ investigation. Still, the existence of four independent methods warrants closer comparative analysis of their
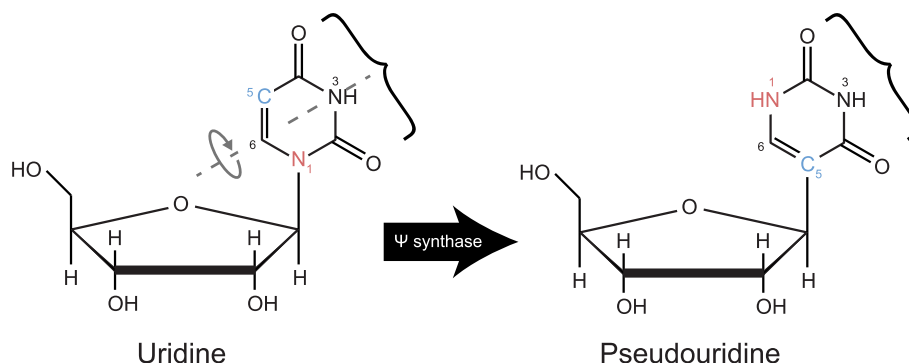

Fig. 1. Schematic of isomerization of uridine to pseudouridine. Pseudouridylation begins with the breakage of the N1–C1′ bond followed by a 180° base rotation around the N3–C6 axis. The resulting Ψ contains an additional hydrogen bond donor (red) and a C5–C1′ base-sugar linkage (blue).
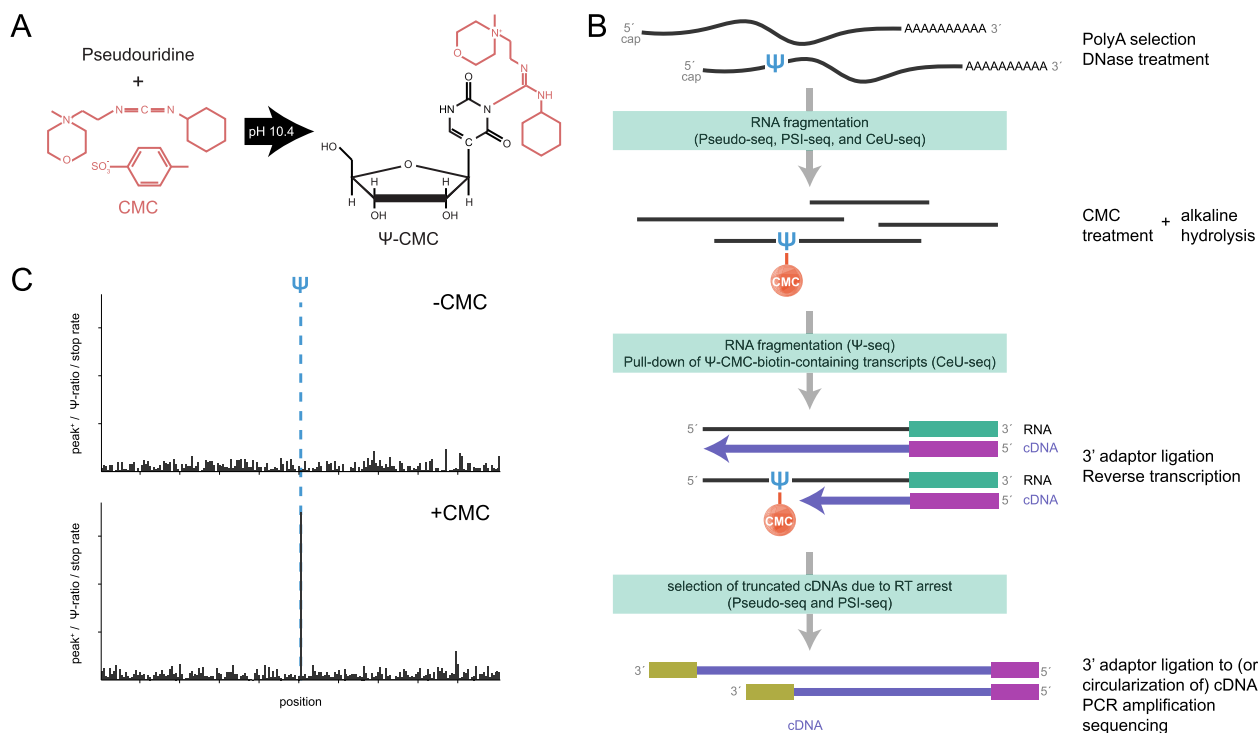
Fig. 2. Transcriptome-wide identification of pseudouridine residues with CMC derivatization. (A) CMC specifically labels pseudouridine. Following alkaline hydrolysis, CMC (red) remains bound to the N3 position of Ψ. (B) Generalized library preparation procedure for Ψ-detection methods. Method-specific details are highlighted in green boxes. PolyA-selected RNA is treated with CMC (or CMC-azide), followed by alkaline hydrolysis, leaving CMC bound specifically to Ψ residues. Adaptors are then ligated to the 3′ ends of RNAs, followed by reverse transcription. The resulting cDNA fragments are then either circularized or subjected to 3′ adaptor ligation. Finally, libraries are PCR amplified and sequenced. (C) Sample output of Pseudo-seq, Ψ-seq, and CeU-seq Ψ-detection metrics. Each method computationally identifies peaks in read starts in CMC-treated (bottom, +CMC) versus mock-treated (top, −CMC) samples as putative sites of pseudouridylation (blue dotted line).

approaches and results. In addition to presenting this analysis, we will here outline the available high-throughput Ψ-detection techniques, discussing the merits and limitations of each approach and their potential to further probe the biological significance of pseudouridylation.

2. The dawn of next-generation sequencing-based pseudouridine detection methods

Pseudouridine was first identified as an unknown ribonucleoside in 1951 by subjecting calf liver RNA isolates to ion-exchange chromatography [12]. Because pseudouridine is mass-silent with respect to its uridine isomer and possesses no readily distinguishable features, rather labor-intensive chromatographic techniques continued to be the prevailing method for Ψ detection. As the field advanced, a combination of RNAse digestion, radiolabeling, and chromatography-based methods produced the first pseudouridine maps in tRNAs and rRNAs [23,24,51]. In 1993, however, a method was developed by Bakin and Ofengand taking advantage of the carbodiimide CMC to label Ψ residues [3]. Under physiological conditions, CMC acylates guanosine (G) at the N1 position and uracil (U) at the N3 position. Notably, isomerization to Ψ creates an additional CMC conjugation site, so CMC acylates Ψ residues at the N1 and N3 positions. CMC adducts are susceptible to alkaline hydrolysis (pH = 10.4), except in Ψ where CMC remains specifically and irreversibly bound at the N3 position. Importantly, Ψ-CMC adducts efficiently block reverse transcription (RT) one base downstream of Ψ, making pseudouridylated residues detectable as a distinct stop (Fig. 2a). CMC/RT detection, however, is not without its own set of shortcomings. Specifically, CMC conjugation is not perfectly efficient; not all Ψs are uniformly labeled, which may lead to false negatives and may preclude abso-

lute quantification of pseudouridylation levels. CMC cleavage from non-Ψ residues (i.e. U- and G-like residues) is likewise incompletely efficient, leading to false positives. Nevertheless, since the CMC/RT approach was introduced, it has become the primary means of Ψ detection, and is the basis of the recently developed transcriptome-wide approaches.

2.1. Scaling up: reinvigorating a popular technique with next-generation sequencing

Four recently published methods, termed Pseudo-seq, Ψ-seq, PSI-seq, and CeU-seq, have all adapted the CMC/RT approach to a high-throughput format for *de novo* identification of pseudouridine residues. The production of truncated reverse transcriptional products due to Ψ-CMC is central to all four methods and poses unique challenges for bioinformatic detection. Each method has therefore developed separate approaches to identify Ψ-CMC generated reverse transcriptional 'stops' to chart the pseudouridine landscape.

Given all four techniques rely on CMC derivatization and subsequent deep sequencing, little technical difference exists between the library preparation protocols for each (Fig. 2b). All began by treating polyA-selected RNA with CMC, followed by alkaline hydrolysis to selectively label Ψ residues. Following treatment, an adaptor was ligated to the 3′ end of RNAs and transcripts were reverse transcribed, with truncated cDNA products resulting from Ψ-CMC-induced RT arrest. Depending on the method used, either a 3′ adaptor was ligated to the resulting cDNAs or RT products were circularized for subsequent PCR amplification and deep sequencing. As a control, libraries were also prepared from mock-treated (i.e. without CMC) samples processed in parallel. Mock-treated libraries ensured that premature RT termination sites were due

specifically to Ψ-CMC and not, for instance, natural stops due to RNA secondary structure.

While all the aforementioned methods follow the general outline detailed above, a few notable exceptions exist, particularly in how each method enriches for Ψ-containing transcripts (Fig. 2b, green boxes). Pseudo-seq and PSI-seq first fragment polyA-selected RNAs to a uniform size range, and select for truncated cDNAs following CMC treatment and reverse transcription [9,35]. While alkaline hydrolysis will fragment RNA species in both treated and mock-treated samples, resulting in nonspecifically truncated cDNAs, treated samples will contain a reproducibly enriched set of fragments due specifically to Ψ-CMC RT arrest. As its full name, N$_3$-CMC-enriched pseudouridine sequencing implies, CeU-seq chemically enriches for Ψ-CMC-containing transcripts. A CMC-azide derivative was utilized for CMC-treatment, which allows for biotin conjugation with click chemistry following derivatization and subsequent hydrolysis. Ψ-CMC-biotin-containing transcripts were then pulled down with streptavidin beads, increasing the method's sensitivity with the benefit of approximately 15–20-fold enrichment of pseudouridylated RNAs [32]. Conversely, Ψ-seq employs no pre-enrichment steps during sample preparation, relying on a purely bioinformatic approach to reduce background [47].

2.2. Seeing a Ψ: bioinformatics approaches to pseudouridine detection

Once again, because all four methods rely on Ψ-CMC-mediated blocks in the reverse transcriptional machinery, all four methods rely on similarly derived bioinformatics approaches to detect sites of pseudouridylation. Reverse transcriptional stops correspond to sequencing read starts. Accordingly, each method computationally identified an increase in CMC-treated reads beginning one position 3′ to a putative Ψ with respect to the mock-treated control. Sites not immediately preceded by a U were filtered out. PSI-seq utilized a regression analysis comparing reads initiating at a given position between treated and mock-treated libraries [35]. Relevant cutoff scores were determined for each individual replicate based on the highest scoring false positive site in rRNA. Pseudo-seq, Ψ-seq, and CeU-seq did not rely on such a statistical approach. Rather, they computationally identified peaks in the number of reads initiating at a particular U-adjacent site (Fig. 2c). Ψ-Seq and CeU-seq calculated the ratio of reads (the 'Ψ-ratio' and the 'stop rate,' respectively) beginning at each mapped position to the total number of reads covering that position (Eq. (1)) [32,47]. The treated and mock-treated ratios were then compared to call putative Ψ sites, requiring the treated ratio, the ratio difference, and the number of reads initiating at that position exceed a particular cutoff. CeU-seq also relied on 'CMC sensitivity'—which was adapted from related work profiling RNA secondary structure using DMS-mediated RT stops—as an additional measure of the difference in stop reads at a particular site [17].

$$\Psi\text{-ratio} = \text{stop rate} = \frac{\text{reads beginning at position}}{\text{total reads at position}} \qquad (1)$$

Pseudo-seq utilized a metric similar to the Ψ-ratio/stop rate, which was calculated with 150-nucleotide windows (Eq. (2), WS) centered on a U site. The number of reads beginning 1 base 3′ of the central U (Eq. (2), URS) and the total number of reads initiating at any other position within the window (Eq. (2), WRS) were determined for treated and mock-treated libraries to calculate the 'peak$^+$' (Eq. (2)) [9]. Peak$^+$ values above a specified cutoff and exceeding a minimal number of supporting reads were used to call putative Ψ sites, requiring reproducibility over a given number of replicates to reduce the possibility of false positives due, for instance, to incomplete cleavage from non-Ψ residues.

$$\text{peak}^+ = \text{WS} \times \frac{\text{URS}^{+\text{CMC}} - \text{URS}^{-\text{CMC}}}{\text{WRS}^{+\text{CMC}} - \text{WRS}^{-\text{CMC}}} \qquad (2)$$

3. The results are in: Pseudo-seq, Ψ-seq, PSI-seq, and CeU-seq in practice

Pseudo-seq, Ψ-seq, PSI-seq, and CeU-seq were all performed on a number of cell types and growth conditions, revealing a tremendous amount of diversity and complexity in the pseudouridylation landscape. Ψ maps have been detailed for budding yeast, human cells (HEK293, HEK293T, HeLa, and fibroblasts), and mouse brain and liver cells. In addition to detecting known sites of pseudouridylation in tRNAs, snRNAs, and rRNAs, pseudouridines were found for the first time in a range of functionally relevant noncoding RNAs and mRNAs [9,32,35,47]. A subset of these newly identified Ψs were attributed to a specific PUS or Ψ-guiding snoRNA through a series of systematic knockdown/knockout experiments. Furthermore, conditionally dependent sites of mRNA pseudouridylation were identified by, for instance, comparing yeast cells grown to different cell densities, human cells grown with or without sera, and mouse cells isolated from different tissues. The main results of each method are summarized in Table 1.

3.1. Establishing Ψ-detecting power with known sites of pseudouridylation in ribosomal RNAs

Positions of pseudouridylation within rRNAs have been well documented experimentally [2,43]. Consequently, known sites of rRNA pseudouridylation were utilized to calibrate each method's respective Ψ-detecting metrics to balance the specificity and sensitivity of each approach. The Ψ-detecting power of each method, outlined in Table 1, was thus determined by the extent to which each was able to predict known Ψs within rRNAs. Notably, CeU-seq's reported sensitivity in Table 1 is likely an underestimate. The method detected 54 sites of pseudouridylation, 47 of which were previously known. Of the remaining seven, three were selected for validation by a CMC-independent method, and were found to be pseudouridylated [32]. The four predicted but as yet unvalidated Ψs may then be true sites of pseudouridylation, and not simply false positives. We thus could not report a false positive rate for CeU-seq.

Each method filters for hits that correspond to a U in the transcriptome. Incomplete hydrolysis of CMC from non-Ψ residues has been observed in previous low-throughput studies and has been a priority in CMC optimization work [16,18]. Computational exclusion of Gs is therefore a potential caveat that might alter the reported false positive rate in the event that CMC hydrolyzed incompletely from G (and U) residues. Nevertheless, according to the Ψ-seq study conducted with log phase yeast, 88 of the predicted 94 Ψ sites (94%) were preceded by a U residue; five of the remaining six sites were adjacent to a called Ψ, and were likely the result of 'stuttered' termination of reverse transcription [2,47]. Additionally, the CeU-seq study demonstrated high specificity of N$_3$-CMC to Ψ, with no cross-reactivity to U or the G-like inosine [32]. The reported Ψ-detecting power is therefore likely unaffected by unhydrolyzed non-Ψ-CMC residues, particularly when coupled with the stringent bioinformatic cutoffs specified by each approach.

Ψ-seq was also able to quantitatively capture the relative level of pseudouridylation by comparing the Ψ-ratios at a particular position across two or more samples. This quantitative power was demonstrated in a synthetic spike-in experiment that mixed different ratios of oligoribonucleotides that either contained a Ψ at a specific site or not [47]. Importantly, however, Ψ-seq was unable to measure absolute levels of pseudouridylation within a

Table 1
Summary of results from Pseudo-seq, Ψ-seq, PSI-seq, and CeU-seq.

Species	Yeast			Human		
Supporting Studies	Pseudo-seq	Ψ-Seq	PSI-seq	Pseudo-seq	Ψ-Seq	CeU-seq
Reported Ψ-detecting Power in rRNA	85% sensitivity 0.14% observed false positive	69% sensitivity 0.07% observed false positive	100% sensitivity (replicate 1) 70% sensitivity (replicate 2)[a]	n.d.	n.d.	>93% sensitivity[b]
Primary Report	ncRNAs 151 mRNAs 260	ncRNAs 107 mRNAs 185	mRNAs 103 (replicate 1) 335 (replicate 2)	ncRNAs 13 mRNAs 96	ncRNAs 43 mRNAs 353	ncRNAs 195 mRNAs 1889
Conditions and Cell Types Surveyed	Logarithmic and post-diauxic growth	Mid-log, logarithmic, and saturated cell growth Heat shock and cold shock	Logarithmic and stationary cell growth Heat shock	Serum starvation of HeLa cells	Fibroblasts and HEK293 cells	Mild and severe heat shock Drug treatment with H_2O_2, cycloheximide, or hepatocyte growth factor Serum starvation All in HEK293T cells

[a] Calculated based on accepting the top 60 sites identified by regression analysis as putative sites of pseudouridylation, and averaging.
[b] 54 Ψ sites were detected, including 47/47 previously known Ψ sites in human rRNA, 3/3 newly confirmed Ψ sites by SCARLET. While four sites were left unvalidated, these Ψ's may well true sites of pseudouridylation.

given sample, perhaps reflecting incomplete CMC derivatization to Ψ residues or Ψ-CMC readthrough events.

3.2. Finding pseudouridines everywhere, methods seek to validate

By applying the Ψ-detecting metrics determined using known sites of rRNA pseudouridylation to whole transcriptome analysis, a great many novel pseudouridines were detected across several transcripts previously thought to lack pseudouridine (Table 1). In yeast, for example, Ψ's were found in the RNAse MRP RNA, which is involved in 5.8S rRNA processing and mitochondrial DNA replication initiation, and a number of H/ACA and C/D snoRNAs, which guide site-specific pseudouridylation and 2′-O-methylation, respectively. In humans, Ψ's were located in disease-related long noncoding RNAs (lncRNAs). Most intriguingly, Pseudo-seq, Ψ-seq, and CeU-seq all identified Ψ5160 and Ψ5590 in the lncRNA metastasis associated lung adenocarcinoma transcript (*MALAT1*), which has been implicated in various types of cancer cells [57]. A great number of pseudouridines—anywhere from 100 to over 1000, depending on the method—were also reported in mRNAs from each profiled cell type. CeU-seq notably identified over five times more putative sites of pseudouridylation in HEK293 cells compared to Ψ-seq, likely owing to the method's pre-enrichment of Ψ-CMC-containing transcripts. CeU-seq's pre-enrichment step quite clearly increased the method's sensitivity while maintaining a high degree of specificity. Gene ontology (GO) analysis was performed on CeU-seq hits from human mRNAs revealing enrichment for mRNAs involved in translation, protein metabolism, and DNA replication. While Ψ's were found to be distributed along the 5′ untranslated region (UTR), coding DNA sequence (CDS), and 3′ UTR, pseudouridylation is underrepresented in the 5′ UTRs of human and mouse mRNAs and in the 3′ UTRs of yeast mRNAs [8,32].

3.2.1. Ψ-Synthase knockdown/knockout experiments indirectly validate putative pseudouridine sites

Pseudouridylation is catalyzed by a number of evolutionarily conserved pseudouridine synthases. To experimentally determine the molecular basis of pseudouridylation, each group systematically knocked down or knocked out activity of a series of non-essential PUSs and Ψ-guiding H/ACA snoRNAs. By mapping a candidate Ψ to a particular PUS, these genetic perturbation experiments served to indirectly validate newly predicted sites of pseudouridylation, strengthening confidence that these sites are indeed true

pseudouridylation targets. Bioinformatic analysis of the sequences flanking putative Ψ's was also able to computationally match Ψ's to particular PUSs that recognize known sequence motifs or complementary H/ACA snoRNA activity. In yeast, pseudouridylation is guided by nine RNA-independent PUSs (eight of which are nonessential) and the essential RNA-guided PUS Cbf5; in humans, 13 proteins with annotated PUS domains have been identified, including the RNA-dependent Cbf5 homolog dyskerin [25]. Of these, Pus4p/TRUB1 and Pus7p respectively recognize the known 'GUUC' and 'UGUA' core sequence motifs [6,7]. Additionally, numerous small RNAs have been computationally predicted to fold into H/ACA snoRNAs that are capable of guiding pseudouridylation via Cbf5/dyskerin activity [26,46]. Carlile et al. were able to computationally identify 157 sites in mRNAs that corresponded to specific H/ACA snoRNA guide sequences; however, only 3 of the 157 sites were actually called by Pseudo-seq, perhaps due to the method's relatively conservative cutoffs [8]. Combined, genetic perturbation experiments and computational analyses linked approximately 20–50% of putative Ψ's to guide RNA or PUS activity, depending on which Ψ-detection method was used and which PUSs and snoRNAs were further investigated.

3.2.2. CeU-seq reveals the first experimentally validated site of mRNA pseudouridylation

While matching predicted sites of pseudouridylation to PUS or guide RNA activity indirectly validated a subset of Ψ candidates, Li et al. went one step further, directly validating four of their hits from CeU-seq. A CMC-independent technique was utilized, termed Site-specific Cleavage And Radioactive-labeling followed by Ligation-assisted Extraction and Thin-layer chromatography (SCARLET), which has the added benefit of quantitatively detecting the extent to which a particular site is modified [34]. SCARLET verified that the aforementioned three previously unknown Ψ sites detected in human rRNA were modified to greater than 90% [32]. Even more intriguingly, SCARLET was applied to demonstrate U519 in *EEF1A1* mRNA was indeed pseudouridylated to approximately 56%, providing the first documented experimental evidence of site-specific mRNA pseudouridylation.

3.3. Evidence for the conditional inducibility of pseudouridylation

Using low-throughput Ψ-detection methods, inducible sites of pseudouridylation have been found in the 28S rRNA of Chinese

hamster ovary cells and in yeast spliceosomal snRNAs in response to changing environmental or developmental conditions [4,13,54]. Having identified pseudouridines in a diverse set of transcripts, each method was applied to cells in a number of environmental conditions to evaluate the extent to which pseudouridylation events are conserved or conditionally dependent. As a result, a subset of tissue-specific, growth state-specific, and stress-specific sites of pseudouridylation have been mapped, further implicating pseudouridine as a highly dynamic modification with important functional implications.

Ψ maps for mRNAs in budding yeast grown to log phase and post-diauxic growth were generated and compared using Pseudo-seq to reveal pseudouridylation events specific to growth state [9]. For instance, 110 of the 260 Ψs found during post-diauxic growth remained undetected in log phase yeast mRNAs. Hundreds of stress-dependent pseudouridylation events were also identified in yeast by Ψ-seq (265 Ψs) and PSI-seq (314 Ψs) analysis of cells following heat shock [35,47]. 60% of Ψ-seq hits perfectly corresponded to the conserved Pus7p recognition motif and became undetectable in the Δpus7 strain, suggesting Pus7p plays a major role in orchestrating heat-shock-specific pseudouridylation. Notably, Pus7p had previously been implicated in the inducible modification of U2 snRNA at Ψ56 following heat shock and nutrient deprivation [54]. Stress-induced Ψs were also found in human cells; CeU-seq profiled sites following heat-shock (464 Ψs) and H_2O_2 treatment (477 Ψs), while Pseudo-seq profiled sites in serum-starved versus serum-fed HeLa cells [9,32]. GO term analysis of the 464 heat-shock-induced pseudouridylated transcripts revealed enrichment in transport- and localization-related pathways, while the 477 transcripts pseudouridylated under H_2O_2 treatment were strongly enriched in telomere- and chromatin-related functions. CeU-seq profiling was additionally performed on mouse cells derived from liver and brain tissue. 1741 and 1543 Ψ sites were identified in brain and liver mRNAs, respectively; however, only 54 of those sites were shared between the two cell types. Remarkably, pseudouridylated transcripts were strongly enriched for tissue-specific function. For instance, Ψ-containing mRNAs from the brain encoded proteins involved in nervous system development and signal transduction.

3.4. Ψ in disease: hypopseudouridylation in X-linked dyskeratosis congenita

Defects in the RNA-dependent pseudouridine synthase DKC1/dyskerin, the mammalian ortholog of Cbf5, are the cause of X-linked dyskeratosis congenita (X-DC), a rare but fatal syndrome resulting in bone marrow failure [38]. In addition to catalyzing pseudouridylation, dyskerin is required for the maturation and stability of snoRNAs and telomerase RNA component (TERC) [31,40]. Telomerase deficiency has been considered the primary cause of X-DC; however, rRNA pseudouridylation deficiency has also been found to contribute to X-DC pathology and severity [40,45]. Given the disease relevance of Cbf5/dyskerin, Ψ-seq was applied to investigate the pseudouridine landscape in the fibroblasts of X-DC patients, confirming that patient rRNA pseudouridylation levels are indeed reduced by an average of 10% at each modified position [47]. By specifically enriching for TERC transcripts in patient and control fibroblasts, Ψ-seq also identified two Ψ sites, both of which had been proposed in a previous study [30]. Both of these sites exhibited reduced levels of modification in patient fibroblasts, suggesting TERC pseudouridylation may contribute to proper telomerase function. That pseudouridylation has similarly been detected in several disease-related transcripts independently by Pseudo-seq, Ψ-seq, and CeU-seq warrants continued investigation into the role Ψ may play in other human diseases.

4. Comparative analysis of Ψ maps; considerations on the robustness of Ψ detection

Given that four independent CMC-based deep sequencing approaches for Ψ detection now exist, we undertook to compare their respective results to determine the robustness of each approach. Because each method was applied to a diverse set of cell types and growth conditions, we were careful to compare Ψ maps provided only for transcripts isolated from the same cell line grown under similar conditions. Consequently, an in-depth analysis was restricted only to yeast cells grown in log phase, though we did also compare human-derived Ψ maps. The resulting comparative analysis revealed a subset of high-confidence Ψ sites, independently detected by multiple methods; however, it also underscored opportunities to improve the available Ψ-detection approaches.

4.1. Comparing pseudouridylation candidates in budding yeast

Pseudo-seq, Ψ-seq, and PSI-seq all profiled pseudouridylation events in yeast undergoing log phase growth ($OD_{600} \approx 1.0$, Pseudo-seq; midlog phase hits were used for Ψ-seq, with log phase defined as $OD_{600} = 2$, though midlog OD_{600} was undefined; $OD_{600} = 0.6–0.8$, PSI-seq), which became the focus of our comparisons. Analysis was further restricted to include only Ψ candidates in coding DNA sequences, as UTRs were not analyzed in Ψ-seq. Because PSI-seq aligned reads to an earlier genome assembly (SacCer2 versus SacCer3), however, site-specific events could only be compared between Pseudo-seq and Ψ-seq. Nevertheless, we were able to interrogate the three methods to uncover a subset of genes with independently called putative Ψs (Fig. 3a, left panel).

In total, pseudouridylation was detected within the CDSs of 402 unique genes. Of those genes, however, only *RPL11a* (a 60S ribosomal subunit protein) was consistently found to contain a CDS-internal Ψ at position 68. On closer inspection, Ψ239 was detected in *TEF1* (a translation elongation factor) by both Pseudo-seq and PSI-seq, and at the same position in *TEF2* by Pseudo-seq and Ψ-seq. Because *TEF1* and *TEF2* are paralogous genes that resulted from gene duplication, it is likely that one or both sequences are pseudouridylated. Importantly, each of these detection techniques is inherently biased towards detecting sites in more abundant transcripts. Indeed, *RPL11a* and *TEF1/TEF2* are both within the top 30 most highly expressed genes in the yeast genome, which may account for their reproducible detection by independent methods [3,53]. Both Pseudo-seq and PSI-seq cite Pus1p dependency for *RPL11a* pseudouridylation, and all three studies cite Pus4p dependency for *TEF1/TEF2*. Furthermore, using the low-throughput CMC-Ψ/RT approach, Lovejoy et al. identified *RPL11a* Ψ68 in the related yeast *Saccharomyces mikitae* and *TEF1* Ψ239 in both *S. mikitae* and *Schizosaccharomyces pombe* [35]. The evident evolutionary conservation of these modifications further points to the potential biological relevance of these particular pseudouridylation events.

Site-specific pseudouridine candidates identified by Pseudo-seq and Ψ-seq were next analyzed. Of the 21 overlapping putatively pseudouridylated CDSs, 10 predicted Ψ positions in 10 genes exactly overlapped (Fig. 3a, right panel, Table 2). The mean distance between the remaining Ψ sites within overlapping CDSs was approximately 740, ruling out the possibility that non-overlapping sites were the result of stuttered CMC-Ψ-mediated RT termination. In both studies, five of the ten Ψ sites were also found to be dependent on activity from the same PUS (either Pus1p or Pus4p). While it would be reasonable to assume that a high Ψ-ratio or peak$^+$ value would increase relative confidence in a given Ψ site, the pseudouridines belonging to this overlapping set did not necessarily have the highest Ψ-detection metrics. In fact, Ψ239 in *TEF1/TEF2* just barely passed the cutoff requirements for
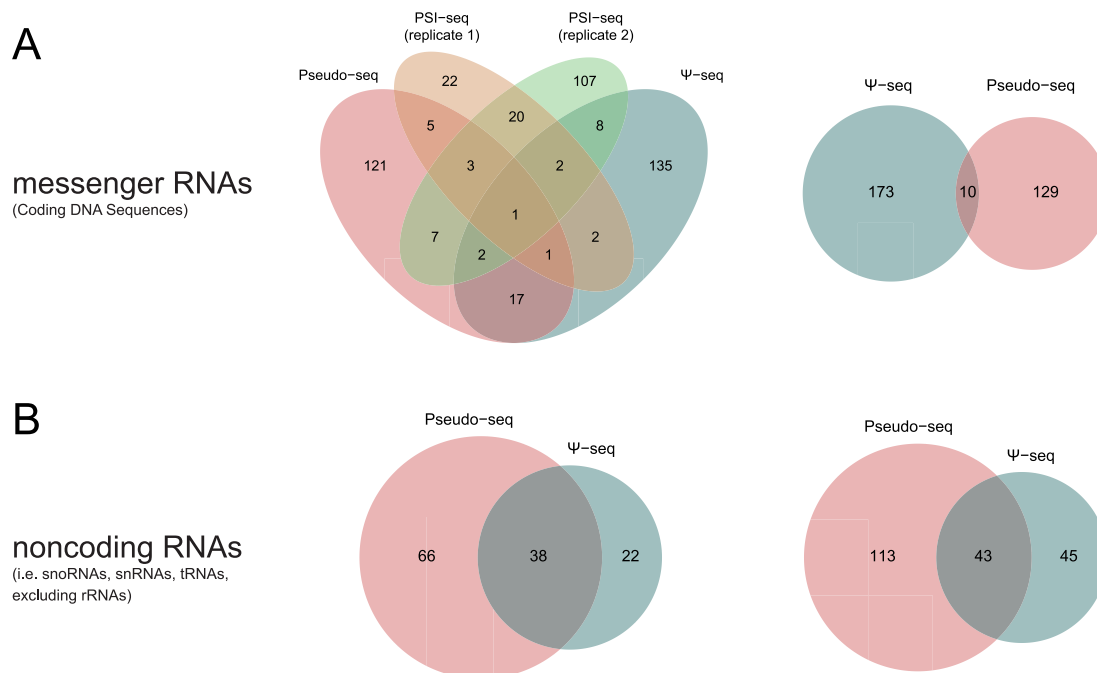
Fig. 3. Comparative analysis of candidate pseudouridylation targets in *S. cerevisiae* during log phase growth. (A) Putatively pseudouridylated coding DNA regions detected by Pseudo-seq, Ψ-seq, and two replicates of PSI-seq (left) and site-specific Ψ sites detected by Pseudo-seq and Ψ-seq (right) were compared to identify overlapping hits. (B) The same analysis was performed for noncoding transcripts (left) and specific ncRNA-internal Ψ sites (right) identified by Pseudo-seq and Ψ-seq.

Table 2
CDS-internal pseudouridine candidates detected by Pseudo-seq and Ψ-seq.

Coordinate	Gene	Position in gene	Ψ-Seq metrics		Pseudo-seq metrics
			Ψ-Ratio	Ψ-fc	Peak+
chr10:383242	KAR2	1916	Not available	Not available	2.82
chr16:126070	YPL225W	65	Not available	Not available	1.15
chr4:331025	BDF2	2	0.66	4.25	13.90
chr3:51028	GLK1	191	0.29	4.31	7.03
chr10:314164	MPM1	709	0.24	4.66	2.02
chr2:477909	TEF2 (TEF1)	239	0.11	3.17	6.25
chr1:32596	GDH3	1030	0.23	5.41	3.52
chr16:731681	RPL11A	68	0.23	3.5	11.97
chr8:499441	RPN10	363	0.17	3.74	5.96
chr7:623051	YGR067C	1736	0.14	3.78	2.68

Ψ-seq (Table 2). Still, given the approximately 2.5 million U residues in yeast coding sequences, an overlap of 10 independently called pseudouridines is highly significant ($P = 1.12 \times 10^{-8}$ by the hypergeometric test), increasing our confidence that this set contains true targets of pseudouridylation. It is worth noting, however, that while all Ψ members of this set were also detected by Pseudo-seq under post-diauxic cell growth, only one (Ψ1916 in *KAR2*) was detected by Ψ-seq following heat shock.

The relatively small percentage of overlapping pseudouridylated CDSs (~0.5%) and specific Ψ positions (~3.2%) does nevertheless highlight the limitations of the high-throughput detection of pseudouridylation events. Specifically, because high coverage at each surveyed position is essential to robust Ψ detection, the output of each method is highly dependent on sequencing depth, which likely varied between each group. All the methods outlined above also favor specificity over sensitivity, which necessitates rather conservative cutoffs for Ψ detection. As a result, the reported Ψs are likely a small sampling of several true pseudouridylation events missed by each method. Additionally, the effi-

ciency of native mRNA pseudouridylation has not been concretely established and may be highly variable [28]. Karijolich et al. noted low isomerization efficiency (~7–10%) when artificially targeting mRNA pseudouridylation, while the one experimentally verified native Ψ target identified by CeU-seq was pseudouridylated to a much higher extent (~56%) [28,32]. High variance in the efficiency of naturally occurring pseudouridylation events coupled with stringent Ψ-detection cutoffs therefore introduces yet another challenge to reproducible Ψ mapping.

A core finding of all four Ψ-detection methods was the conditional inducibility of pseudouridylation, which further complicates Ψ profiling. Changes in the Ψ landscape in response to large environmental perturbations were investigated; however, the robustness of particular pseudouridylation events to smaller environmental fluctuations was not examined. For instance, small differences in CO_2 levels in the incubators of different laboratory spaces may produce different Ψ landscapes. The difference in Ψs identified by these different methods may then be a reflection of biological fluctuations in pseudouridylation in even slightly different environmental contexts. Furthermore, all of the above methods query pseudouridylation events in populations of cells, aggregating cells that likely differ, for instance, in cell cycle stage or microenvironment. These distinct subpopulations may likewise differ with respect to pseudouridylation substrates. Population averaging effects may thus be an additional contributor to variance. We may speculate, then, that the Ψs identified by multiple methods are more frequently pseudouridylated under a broader spectrum of environments, suggesting they play some core role in mRNA structure or function, at least under logarithmic cell growth.

With the above challenges in mind, we turned our attention to analyzing the set of pseudouridines detected in noncoding transcripts in log-phase yeast, excluding rRNAs. Because PSI-seq did not detail Ψs in this subset of transcripts, we compared only the outputs from Pseudo-seq and Ψ-seq. Here, the percentage of pseudouridylated transcripts (~30%) and specific Ψ sites (~20%) independently detected by each method was markedly greater and

highly statistically significant ($P = 6.25 \times 10^{-9}$ by the hypergeometric test) (Fig. 2b). This overlap is well in line with the generally higher expression of ncRNA species with respect to their protein-coding counterparts. Important to note as well, these noncoding RNA transcripts include snRNAs and tRNAs—the long established targets of site-specific pseudouridylation. Moreover, pseudouridines have long been established to be essential for proper structure and function of these classes of RNAs, necessitating constitutive modification of specific uridine residues. Combined, higher expression and functional importance thus facilitate reproducible Ψ detection by multiple methods.

4.2. Comparing pseudouridylation candidates in human cells

Pseudo-seq detailed the Ψ landscape for epithelia-derived HeLa cells, Ψ-seq for a combination of embryonic kidney-derived HEK293 cells and fibroblasts, and CeU-seq for HEK293T cells. While the comparative analysis undertaken above would suggest that these three methods are not directly comparable, we still wondered if we might determine to what extent pseudouridylation was conserved across all transcripts in these different human cell types. Importantly, CeU-seq pre-enriches for Ψ-CMC-containing transcripts by up to 20-fold to increase the method's sensitivity to low-abundance transcripts, which accounts for the large difference in the reported number of hits with respect to Pseudo-seq and Ψ-seq. Pseudo-seq-analyzed HeLa cells shared no putative Ψs with HEK293T cells or the combination of HEK293 cells and fibroblasts, aside from the previously mentioned Ψ5160 and Ψ5590 in the lncRNA *MALAT1*. The lack of commonly predicted Ψs between these cell lines derived from different tissues is in line with the low overlap in Ψ sites detected by CeU-seq in mouse brain and liver cells [32]. On the other hand, HEK293/fibroblast cells and HEK293T cells shared 47 putative Ψs out of the 396 and 2084 called sites in Ψ-seq and CeU-seq, respectively. Rather interestingly, nearly 90% of those overlapping positions were detected in mRNAs, distributed primarily in the 3′ UTR and CDS regions. Once again, the magnitude of each method's respective Ψ-detecting metrics (Ψ-ratio, Ψ-fc, and stop rate difference) does not necessarily correlate with their inclusion in this overlapping set.

4.3. Comparative analyses shed light on opportunities for Ψ-detection improvement

Of the many pseudouridylation events that have been collectively identified by the available Ψ-detection methods, a small subset have been identified by more than one method, further increasing confidence in the Ψ-detecting power of these techniques with the necessary caveats detailed above. Nevertheless, the motivation behind developing such Ψ-detection methods is to elucidate the functional role of this modification. Having further established Ψ's ubiquity by cataloguing a remarkable number of putatively modified sites, it is imperative to next narrow down the list to a set of promising, robustly modified and detectable candidates to interrogate experimentally through, for instance, site-specific Ψ knockout experiments. While comparing the outputs of each respective method has filtered the set of putative Ψs for yeast grown to log phase (and to a lesser extent for HEK293/HEK293T/fibroblast cells), to perform all four methods for every cell type and growth environment of interest is quite obviously impractical. Consequently, each method could benefit from additional parameters that measure the extent to which a given uridine is isomerized, particularly because a high Ψ-detecting metric from any one method does not guarantee detection by an independent technique.

5. Towards quantitative Ψ profiling; a case for molecular barcoding

While Ψ-seq has been demonstrated to quantitatively detect differences in levels of pseudouridylation by comparing Ψ-ratios between different samples, it cannot absolutely quantify the extent to which a particular uridine is pseudouridylated within a given sample [47]. This limitation extends to all available high-throughput Ψ-detection methods. For instance, ribosomal pseudouridines are considered to be constitutively modified at near 100% efficiency. Recent work in *S. pombe* has experimentally demonstrated that the majority of pseudouridylated residues in rRNA are indeed highly modified (>85% isomerization) [52]. We were therefore curious to examine the variation in Ψ-ratios across ribosomal pseudouridines identified by Ψ-seq, as this method has a special focus on quantitative measurement (Fig. 4a). Ψ-Ratios are remarkably reproducible among replicates at a given position, which well supports Ψ-seq's ability to quantitatively compare Ψ levels between samples. However, the relatively uniform level of rRNA pseudouridylation is not reflected in the variable distribution of Ψ-ratios across all rRNA positions. A similar trend can be seen in the peak+ Ψ-detection metric utilized by Pseudo-seq, though peak+ values exhibit a higher degree of variability at each position (Fig. 4b). Variation at a given Ψ residue reflects the variability intrinsic to RNA-seq library preparations using the CMC/RT approach, which requires multiple steps that likewise introduce multiple opportunities for inconsistency in the hands of different operators. Variation across all known rRNA Ψ residues, however, may be the result of chemical limitations inherent in CMC's ability to uniformly derivatize to pseudouridine, which may be due, for instance, to restrictions imposed by RNA secondary structure. To our knowledge, while studies have been undertaken to optimize CMC derivatization efficiency, substrate preferences for CMC derivatization, if any, have not been characterized [18].

An alternative explanation to intrasample variability in Ψ-detection metrics lies in the high sequencing depth required for each method outlined in this review. Increasing sequencing coverage captures more rare cDNA fragments resulting from lowly expressed transcripts; however, increased depth also results in sequencing redundant PCR amplification products more frequently. This trade-off is particularly important given that each of the Ψ-detection techniques identify putative Ψs by *an enrichment in identical reads initiating at the same position*. Importantly, single-end sequencing produces Ψ-CMC-derived reads that are indistinguishable from PCR duplicates (assuming no mismatches). Discarding duplicates therefore interferes with Ψ-detecting power, as multiple Ψ-CMC-initiating reads are collapsed into one (Fig. 4c). Requiring several replicates for confident Ψ detection does mitigate the possibility of false positive Ψ calls due to PCR duplicates. Still, it is difficult to determine the true proportion of reads initiating at a position due to Ψ-CMC, which could more accurately reflect the level of pseudouridylation at that position. Notably, Ψ-seq performed paired-end sequencing, which improves read mapping resolution by sequencing both the 5′ and 3′ ends, to the extent that cDNA fragment length is sufficiently diverse. Redundant reads can therefore be collapsed more easily with reduced loss of sequencing information (Fig. 4c).

The requisite sequencing depth may still be ensured while conserving reads that derive from identical Ψ-CMC-derived cDNA fragments (as opposed to identical PCR duplicates). Coupling molecular barcoding with RNA-seq has been shown to more accurately and reproducibly quantify the absolute number of cDNA fragments in a given sample [21,48]. We therefore propose similarly incorporating short randomized DNA sequences, through end ligation or reverse transcription, prior to PCR amplification

Fig. 4. Quantitative limitations in the current Ψ-detection methods. (A) Ψ-Ratios of known sites of pseudouridylation in 18S and 25S rRNAs detected by Ψ-seq were plotted to assess the variance at each position and across all positions. (B) The same analysis was applied to the peak$^+$ Ψ-detection metric used in Pseudo-seq. (C) Potential improvements with molecular barcoding. Schematics of PCR duplicate removal of Ψ-CMC-derived reads using single-end RNA-seq (purple box), paired-end RNA-seq (purple and green boxes), and paired end RNA-seq coupled with molecular barcoding (purple, green, and multi-colored boxes) are depicted. Depending on the sequencing mode used, more reads may be retained. For instance, while single-end sequencing produces five apparently identical reads that would be filtered out as duplicates and collapsed into one read, coupling barcoding with paired-end sequencing allows for the detection of five uniquely barcoded reads.

to uniquely identify cDNA fragments. The specific length of the barcodes is dictated by the size and complexity of the transcriptome under investigation [10,20]. Replicate clones due to PCR amplification are identified as reads with matching barcodes and sequences that map to the same location. These reads are then collapsed into one, allowing for single-copy resolution. Most importantly, identical reads initiating from the same position due to Ψ-CMC may be distinguished by their unique barcodes, providing absolute quantitation of the number of reads initiating at and covering a given position (Fig. 4c). In other words, the number of unique barcodes, rather than the number of reads, would be used to count and calculate the relevant Ψ-detection metrics. While molecular barcodes cannot completely overcome limitations due to inefficient CMC conjugation, barcoding provides a more quantitative approach that could facilitate a more direct comparison of pseudouridylation levels across putative Ψ sites by comparing absolute proportions of reads. Furthermore, the extent to which given positions are pseudouridylated provides additional information to discern which particular Ψ residues are most promising for further functional investigation.

6. Concluding remarks

Pseudo-seq, Ψ-seq, PSI-seq, and CeU-seq have provided the foundational work for transcriptome-wide investigation of pseu-

douridylation. Collectively, they have confirmed the ubiquity and diversity of pseudouridylation events in a diverse range of RNA species, including mRNA, and have detailed the Ψ landscape in a number of cellular and environmental contexts. Nevertheless, each method comes with limitations likely due, as discussed, to both the technical nature of the approach and the biology of pseudouridylation. We have therefore proposed that some of the more technical limitations may be ameliorated by harnessing the quantitative power of molecular barcoding. Nevertheless, high-throughput methods that circumvent CMC dependency should be explored as the field moves forward. With a growing catalogue of Ψ sites, the quest now turns to further elucidating the functional relevance of pseudouridylation. The field is now left to continue probing Ψ's potential roles in protein coding/recoding, RNA stability, and RNA splicing, and its significance in the pathology of human diseases like X-DC.

References

[1] J.G. Arnez, T.A. Steitz, Crystal structure of unmodified tRNA(Gln) complexed with glutaminyl-tRNA synthetase and ATP suggests a possible role for pseudo-uridines in stabilization of RNA structure, Biochemistry 33 (1994) 7560–7567.

[2] A. Bakin, B.G. Lane, J. Ofengand, Clustering of pseudouridine residues around the peptidyltransferase center of yeast cytoplasmic and mitochondrial ribosomes, Biochemistry 33 (1994) 13475–13483.

[3] A. Bakin, J. Ofengand, 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 (1993) 9754–9762.

[4] A. Basak, C.C. Query, A pseudouridine residue in the spliceosome core is part of the filamentous growth program in yeast, Cell Rep. 8 (2014) 966–973.

[5] A. Baudin-Baillieu, C. Fabret, X.-H. Liang, D. Piekna-Przybylska, M.J. Fournier, J.-P. Rousset, Nucleotide modifications in three functionally important regions of the Saccharomyces cerevisiae ribosome affect translation accuracy, Nucleic Acids Res. 37 (2009) 7665–7677.

[6] H.F. Becker, Y. Motorin, R.J. Planta, H. Grosjean, The yeast gene YNL292w encodes a pseudouridine synthase (Pus4) catalyzing the formation of psi55 in both mitochondrial and cytoplasmic tRNAs, Nucleic Acids Res. 25 (1997) 4493–4499.

[7] I. Behm-Ansmant, A. Urban, X. Ma, Y.-T. Yu, Y. Motorin, C. Branlant, The Saccharomyces cerevisiae U2 snRNA:pseudouridine-synthase Pus7p is a novel multisite-multisubstrate RNA:Psi-synthase also acting on tRNAs, RNA 9 (2003) 1371–1382.

[8] T.M. Carlile, M.F. Rojas-Duran, W.V. Gilbert, Pseudo-Seq: genome-wide detection of pseudouridine modifications in RNA, Methods Enzymol. 560 (2015) 219–245.

[9] T.M. Carlile, M.F. Rojas-Duran, B. Zinshteyn, H. Shin, K.M. Bartoli, W.V. Gilbert, Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells, Nature 515 (2014) 143–146.

[10] J.A. Casbon, R.J. Osborne, S. Brenner, C.P. Lichtenstein, A method for counting PCR template molecules with application to next-generation sequencing, Nucleic Acids Res. 39 (2011) (e81–e81).

[11] M. Charette, M.W. Gray, Pseudouridine in RNA: what, where, how, and why, IUBMB Life 49 (2000) 341–351.

[12] W.E. Cohn, Some results of the applications of ion-exchange chromatography to nucleic acid chemistry, J. Cell. Physiol. Suppl. 38 (1951) 21–40.

[13] F.C. Courtes, C. Gu, N.S.C. Wong, P.C. Dedon, M.G.S. Yap, D.-Y. Lee, 28S rRNA is inducibly pseudouridylated by the mTOR pathway translational control in CHO cell cultures, J. Biotechnol. 174 (2014) 16–21.

[14] F.F. Davis, F.W. Allen, Ribonucleic acids from yeast which contain a fifth nucleotide, J. Biol. Chem. 227 (1957) 907–915.

[15] W.A. Decatur, M.J. Fournier, RRNA modifications and ribosome function, Trends Biochem. Sci. 27 (2002) 344–351.

[16] M. Del Campo, C. Recinos, G. Yanez, S.C. Pomerantz, R. Guymon, P.F. Crain, et al., Number, position, and significance of the pseudouridines in the large subunit ribosomal RNA of Haloarcula marismortui and Deinococcus radiodurans, RNA 11 (2005) 210–219.

[17] Y. Ding, Y. Tang, C.K. Kwok, Y. Zhang, P.C. Bevilacqua, S.M. Assmann, In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features, Nature 505 (2014) 696–700.

[18] A. Durairaj, P.A. Limbach, Improving CMC-derivatization of pseudouridine in RNA for mass spectrometric detection, Anal. Chim. Acta 612 (2008) 173–181.

[19] I.S. Fernández, C.L. Ng, A.C. Kelley, G. Wu, Y.-T. Yu, V. Ramakrishnan, Unusual base pairing during the decoding of a stop codon by the ribosome, Nature 500 (2013) 107–110.

[20] G.K. Fu, J. Hu, P.-H. Wang, S.P.A. Fodor, Counting individual DNA molecules by the stochastic attachment of diverse labels, Proc. Natl. Acad. Sci. U.S.A. 108 (2011) 9026–9031.

[21] G.K. Fu, W. Xu, J. Wilhelmy, M.N. Mindrinos, R.W. Davis, W. Xiao, et al., Molecular indexing enables quantitative targeted RNA sequencing and reveals poor efficiencies in standard library preparations, Proc. Natl. Acad. Sci. U.S.A. 111 (2014) 1891–1896.

[22] P. Ganot, M.L. Bortolin, T. Kiss, Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs, Cell 89 (1997) 799–809.

[23] R.C. Gupta, B.A. Roe, K. Randerath, The nucleotide sequence of human tRNAGly (anticodon GCC), Nucleic Acids Res. 7 (1979) 959–970.

[24] R.W. Holley, G.A. Everett, J.T. Madison, A. Zamir, Nucleotide sequences in the yeast alanine transfer ribonucleic acid, J. Biol. Chem. 240 (1965) 2122–2128.

[25] S. Hunter, P. Jones, A. Mitchell, R. Apweiler, T.K. Attwood, A. Bateman, et al., InterPro in 2011 new developments in the family and domain prediction database, Nucleic Acids Res. 40 (2012) D306–D312.

[26] A. Hüttenhofer, M. Kiefmann, S. Meier-Ewert, J. O'Brien, H. Lehrach, J.P. Bachellerie, et al., RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse, EMBO J. 20 (2001) 2943–2953.

[27] K. Jack, C. Bellodi, D.M. Landry, R.O. Niederer, A. Meskauskas, S. Musalgaonkar, et al., RRNA pseudouridylation defects affect ribosomal ligand binding and translational fidelity from yeast to human cells, Mol. Cell 44 (2011) 660–666.

[28] J. Karijolich, Y.-T. Yu, Converting nonsense codons into sense codons by targeted pseudouridylation, Nature 474 (2011) 395–398.

[29] K. Karikó, H. Muramatsu, F.A. Welsh, J. Ludwig, H. Kato, S. Akira, et al., Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability, Mol. Ther. 16 (2008) 1833–1840.

[30] N.-K. Kim, C.A. Theimer, J.R. Mitchell, K. Collins, J. Feigon, Effect of pseudouridylation on the structure and activity of the catalytically essential P6.1 hairpin in human telomerase RNA, Nucleic Acids Res. 38 (2010) 6746–6756.

[31] D.L. Lafontaine, D. Tollervey, Nop58p is a common component of the box C+D snoRNPs that is required for snoRNA stability, RNA 5 (1999) 455–467.

[32] X. Li, P. Zhu, S. Ma, J. Song, J. Bai, F. Sun, et al., Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome, Nat. Chem. Biol. 11 (2015) 592–597.

[33] X.-H. Liang, Q. Liu, M.J. Fournier, Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing, RNA 15 (2009) 1716–1728.

[34] N. Liu, M. Parisien, Q. Dai, G. Zheng, C. He, T. Pan, Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA, RNA 19 (2013) 1848–1856.

[35] A.F. Lovejoy, D.P. Riordan, P.O. Brown, Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae, PLoS One 9 (2014) e110799.

[36] X. Ma, X. Zhao, Y.-T. Yu, Pseudouridylation (Psi) of U2 snRNA in S. cerevisiae is catalyzed by an RNA-independent mechanism, EMBO J. 22 (2003) 1889–1897.

[37] M.A. Machnicka, K. Milanowska, O. Osman Oglou, E. Purta, M. Kurkowska, A. Olchowik, et al., MODOMICS: a database of RNA modification pathways – 2013 update, Nucleic Acids Res. 41 (2013) D262–D267.

[38] U.T. Meier, Dissecting dyskeratosis, Nat. Genet. 33 (2003) 116–117.

[39] U.T. Meier, Pseudouridylation goes regulatory, EMBO J. 30 (2011) 3–4.

[40] J.R. Mitchell, E. Wood, K. Collins, A telomerase component is defective in the human disease dyskeratosis congenita, Nature 402 (1999) 551–555.

[41] M.I. Newby, N.L. Greenbaum, Investigation of Overhauser effects between pseudouridine and water protons in RNA helices, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 12697–12702.

[42] J. Ni, A.L. Tien, M.J. Fournier, Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA, Cell 89 (1997) 565–573.

[43] J. Ofengand, A. Bakin, Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaebacteria, mitochondria and chloroplasts, J. Mol. Biol. 266 (1997) 246–268.

[44] M. Parisien, C. Yi, T. Pan, Rationalization and prediction of selective decoding of pseudouridine-modified nonsense and sense codons, RNA 18 (2012) 355–367.

[45] D. Ruggero, S. Grisendi, F. Piazza, E. Rego, F. Mari, P.H. Rao, et al., Dyskeratosis congenita and cancer in mice deficient in ribosomal RNA modification, Science 299 (2003) 259–262.

[46] P. Schattner, Genome-wide searching for pseudouridylation guide snoRNAs: analysis of the Saccharomyces cerevisiae genome, Nucleic Acids Res. 32 (2004) 4281–4296.

[47] S. Schwartz, D.A. Bernstein, M.R. Mumbach, M. Jovanovic, R.H. Herbst, B.X. León-Ricardo, et al., Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA, Cell 159 (2014) 148–162.

[48] K. Shiroguchi, T.Z. Jia, P.A. Sims, X.S. Xie, Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes, Proc. Natl. Acad. Sci. U.S.A. 109 (2012) 1347–1352.

[49] B.S. Sibert, J.R. Patton, Pseudouridine synthase 1: a site-specific synthase without strict sequence recognition requirements, Nucleic Acids Res. 40 (2012) 2107–2118.

[50] F. Spenkuch, Y. Motorin, M. Helm, Pseudouridine: still mysterious, but never a fake (uridine)!, RNA Biol 11 (2015) 1540–1554.

[51] Y. Tanaka, T.A. Dyer, G.G. Brownlee, An improved direct RNA sequence method; its application to Vicia faba 5.8S ribosomal RNA, Nucleic Acids Res. 8 (1980) 1259–1272.

[52] M. Taoka, Y. Nobe, M. Hori, A. Takeuchi, S. Masaki, Y. Yamauchi, et al., A mass spectrometry-based method for comprehensive quantitative determination of post-transcriptional RNA modifications: the complete chemical structure of Schizosaccharomyces pombe ribosomal RNAs, Nucleic Acids Res. 43 (2015) (e115–e115).

[53] V.E. Velculescu, L. Zhang, W. Zhou, J. Vogelstein, M.A. Basrai, D.E. Bassett, et al., Characterization of the yeast transcriptome, Cell 88 (1997) 243–251.

[54] G. Wu, M. Xiao, C. Yang, Y.-T. Yu, U2 snRNA is inducibly pseudouridylated at novel sites by Pus7p and snR81 RNP, EMBO J. 30 (2010) 79–89.

[55] G. Wu, A.T. Yu, A. Kantartzis, Y.-T. Yu, Functions and mechanisms of spliceosomal small nuclear RNA pseudouridylation, WIREs RNA 2 (2011) 571–581.

[56] C. Yang, D.S. McPheeters, Y.-T. Yu, Psi35 in the branch site recognition region of U2 small nuclear RNA is important for pre-mRNA splicing in Saccharomyces cerevisiae, J. Biol. Chem. 280 (2005) 6655–6662.

[57] R. Yoshimoto, A. Mayeda, M. Yoshida, S. Nakagawa, MALAT1 long non-coding RNA in cancer, Biochim. Biophys. Acta 2015 (1859) 192–199.

[58] Y.T. Yu, M.D. Shu, J.A. Steitz, Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing, EMBO J. 17 (1998) 5783–5795.