Cell

Transcriptome-wide Mapping Reveals Widespread Dynamic-Regulated Pseudouridylation of ncRNA and mRNA

Schraga Schwartz,[1,9] Douglas A. Bernstein,[2,9] Maxwell R. Mumbach,[1,9] Marko Jovanovic,[1] Rebecca H. Herbst,[1,3] Brian X. León-Ricardo,[1,4] Jesse M. Engreitz,[1,5] Mitchell Guttman,[6] Rahul Satija,[1] Eric S. Lander,[1,3,7,10] Gerald Fink,[2,7,10,*] and Aviv Regev[1,7,8,10,*]

[1]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[2]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
[3]Department of Systems Biology, Harvard Medical School, Boston, MA 02114, USA
[4]Department of Biology, University of Puerto Rico, Rio Piedras Campus, San Juan 00931, Puerto Rico
[5]Division of Health Sciences and Technology, MIT, Cambridge, MA 02139, USA
[6]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA
[7]Department of Biology, MIT, Cambridge, MA 02139, USA
[8]Howard Hughes Medical Institute, 4000 Jones Bridge Road, Chevy Chase, MD 20815, USA
[9]Co-first author
[10]Co-senior author
*Correspondence: gfink@wi.mit.edu (G.F.), aregev@broad.mit.edu (A.R.)
http://dx.doi.org/10.1016/j.cell.2014.08.028

SUMMARY

Pseudouridine is the most abundant RNA modification, yet except for a few well-studied cases, little is known about the modified positions and their function(s). Here, we develop Ψ-seq for transcriptome-wide quantitative mapping of pseudouridine. We validate Ψ-seq with spike-ins and de novo identification of previously reported positions and discover hundreds of unique sites in human and yeast mRNAs and snoRNAs. Perturbing pseudouridine synthases (PUS) uncovers which pseudouridine synthase modifies each site and their target sequence features. mRNA pseudouridinylation depends on both site-specific and snoRNA-guided pseudouridine synthases. Upon heat shock in yeast, Pus7p-mediated pseudouridylation is induced at >200 sites, and *PUS7* deletion decreases the levels of otherwise pseudouridylated mRNA, suggesting a role in enhancing transcript stability. rRNA pseudouridine stoichiometries are conserved but reduced in cells from dyskeratosis congenita patients, where the PUS *DKC1* is mutated. Our work identifies an enhanced, transcriptome-wide scope for pseudouridine and methods to dissect its underlying mechanisms and function.

INTRODUCTION

The most abundant RNA modification is pseudouridine (Ψ). In yeast, Ψ is present in many positions in transfer RNAs (tRNAs), in 46 positions across the four ribosomal RNAs (rRNAs) (25S, 18S, 5.8S, and 5S), and in six positions in U1, U2, and U5 snRNA (Charette and Gray, 2000; Ofengand, 2002; Yu et al., 2011). Although Ψ's base pairing is similar to uridine, isomerization allows the potential formation of an extra hydrogen bond and may contribute to structural stability (Durant and Davis, 1997; Kierzek et al., 2014). Pseudouridylation is catalyzed by pseudouridine synthases. In yeast, eight nonessential pseudouridine synthases directly catalyze Ψ formation at specific sites in tRNA, U2 snRNA, and mitochondrial rRNA (Ansmant et al., 2000; Behm-Ansmant et al., 2003; Massenet et al., 1999). In contrast, all positions in rRNA and one position in U2 snRNA are modified by the essential PUS Cbf5, which is guided to its target positions through H/ACA-box small nucleolar RNAs (snoRNAs) based on a 10–12 nt stretch of complementarity to the target sequence (Lafontaine et al., 1998; Watkins et al., 1998). In human, 23 proteins harbor a pseudouridine synthase domain (Hunter et al., 2012), but most have not been studied for function or specificity.

Although the roles of individual Ψ modifications have largely remained elusive, pseudouridylation defects have profound effects. Alteration of bacterial ribosomal RNA (rRNA) pseudouridylation affects antibiotic sensitivity (Toh and Mankin, 2008), ablation of rRNA pseudouridylation by *CBF5* deletion in *S. cerevisae* is lethal (Jiang et al., 1993; Zebarjadian et al., 1999), and defects in *DKC1*/Dyskerin, the mammalian *CBF5* ortholog, cause dyskeratosis congenita (Heiss et al., 1998), a disorder characterized by failure of ribosome biogenesis and an increased risk of cancer (Hoareau-Aveilla et al., 2008). Furthermore, deletion of *S. cerevisiae PUS1* results in growth defects, and mutation of human *PUS1* leads to mitochondrial myopathy and sideroblastic anemia (Fujiwara and Harigae, 2013).

Recent findings that a few Ψ sites in yeast U2 snRNA are induced by nutrient deprivation or heat shock (Wu et al., 2011)

and rRNA sites by rapamycin treatment (Courtes et al., 2014) raise the possibility that the pseudouridine landscape may be more complex and dynamic than previously assumed and that Ψ might also be present in other RNA species, including mRNA. Testing this hypothesis requires a sensitive method for transcriptome-wide mapping of Ψ sites. Such an assay would also permit identification of the enzyme(s) that catalyze the modification at each site and their specificity, as well as an assessment of whether Ψ is a dynamic feature.

Here, we develop Ψ-seq for transcriptome-wide quantitative mapping of Ψ and use it to identify the vast majority of previously reported sites in rRNA, tRNA, and snRNA and dozens of unique sites within snoRNAs and mRNAs. Combining Ψ-seq with genetic perturbations, we associate each site with its cognate PUS and/or snoRNA, finding that mRNA Ψ sites are mediated by at least four different PUSs, conserved from yeast to mammals. Hundreds of sites are induced by heat shock in yeast in a Pus7p-dependent manner, possibly affecting transcript stability. Finally, we find a subtle but significant decrease in rRNA and TERC pseudouridylation in cells from patients with dyskeratosis congenita. Our results show that Ψ is ubiquitous in diverse RNAs and dynamic in mRNA and provide a transcriptome-wide assay for functional studies.

RESULTS

Transcriptome-wide Mapping of Pseudouridine with Ψ-Seq

To map Ψ across the transcriptome, we developed Ψ-seq, relying on the unique stability of N_3-[N-cyclohexyl-N'-β-(4-methylmorpholinium)ethylcarbodiimide-Ψ (N_3-CMC-Ψ) to alkaline hydrolysis, and the ability of N_3-CMC-Ψ to terminate reverse transcription. This was previously used to test individual sites in primer-extension assays (Bakin and Ofengand, 1993, 1995, 1998). Here, we coupled it with strand-specific RNA sequencing (RNA-seq). Specifically (Figure 1A), we (1) treat polyA-selected RNA with CMC, which covalently binds to U, G, and Ψ residues (Ho and Gilham, 1971; Metz and Brown, 1969a, 1969b); (2) incubate CMC-treated RNA at alkaline pH, leading to hydrolysis of the less stable U-CMC and G-CMC adducts; (3) fragment RNA (to 80–150 nt), ligate 3′ adaptor, and reverse transcribe, expecting premature termination at Ψ-CMC sites; and (4) ligate an adaptor to the 3′ end of the cDNA, followed by library preparation. For each experiment, we sequenced both a treated sample and input RNA that was not treated with CMC but otherwise handled identically (Experimental Procedures).

We defined two metrics to analyze Ψ-seq data (Figure 1B). The Ψ-ratio is the proportion of reads supporting reverse transcription termination at a position out of all the reads overlapping it. The Ψ fold change (Ψ-fc) is the $\log_2$ transformed Ψ ratio in the treated samples divided by the Ψ ratio in the nontreated (input) samples. This controls for nonuniform coverage in RNA-seq due to mappability gaps or complex posttranscriptional RNA processing.

We confirmed Ψ-seq's ability to identify Ψ sites and quantified their relative stoichiometry by spiking into a complex sample a 150 nt synthetic RNA oligoribonucleotide with a single Ψ site at position 43 (Experimental Procedures). Examining the Ψ ratio revealed a single strong peak precisely one base downstream of the modified site, as expected (Figure 1C). To assess whether Ψ-seq could quantify Ψ's stoichiometry, we mixed the modified oligoribonucleotide with a nonmodified counterpart at ratios ranging from 0% to 100%. There was excellent quantitative agreement between the known stoichiometry and its Ψ-seq quantification (Pearson $r = 0.98$, Figures 1C and 1D). Thus, Ψ-seq can identify Ψ sites and quantify their relative stoichiometry by the Ψ ratio. However, the Ψ ratio does not capture the absolute stoichiometry—even with a fully pseudouridylated probe, only 43% of the reads terminated at the Ψ site, suggesting partial efficiency of either CMC modification or Ψ-CMC-mediated reverse transcription termination. Finally, downsampling of the spike-in data showed that Ψ ratios depended on sequencing depth; variability of Ψ-ratio estimates decreased with increased depth and with increased Ψ stoichiometry (Figure S1 available online).

Ψ-Seq of Log Phase Yeast Reliably Detects Known Ψ Sites

We applied Ψ-seq to RNA isolated from yeast at log phase and found excellent overlap (Figure 1E) between the experimentally determined Ψ sites and the 45 known sites within the three rRNA subunits (18S, 25S, and 5.8S) (Ofengand, 2002). We had a very high discriminatory power between a gold standard of positive (known) and negative (all remaining) sites in rRNA, thresholding on either the Ψ ratio (area under the receiver-operator curve [AUC] = 0.942), Ψ-fc (AUC = 0.955), or their logistic combination (AUC = 0.967) (Figure 1F). Based on this analysis of known sites, we called a site as pseudouridylated in a transcriptome-wide analysis (Experimental Procedures) using the "pseudouridine detection thresholds" of Ψ-fc > 3 (8-fold enrichment) and Ψ ratio > 0.1. These thresholds favor precision over sensitivity; 31/32 rRNA positions called by these criteria are known sites (precision = 97%), but 14/45 known sites are not called (sensitivity = 69%). Thus, all subsequent analyses may underestimate the number of true Ψ sites. Further supporting call accuracy, 88/94 called sites (94%) in log phase yeast were immediately preceded by a "U" (Figure 1G), and 5/6 remaining sites were one nucleotide away from the 88 called sites, and thus read accumulation at these five sites is likely due to "stuttered" termination of reverse transcription (Bakin and Ofengand, 1998). Downsampling of a deeply sequenced sample shows dependency of site detection on sequencing depth and transcript expression (Figure S2).

Ψ-Seq of Deletion Mutants Associates Ψ Sites and Their Cognate PUS

Applying Ψ-seq to RNA from yeast at midlog, we called both known target sites and ones not previously reported. To validate these putative sites, we reasoned that Ψ-seq in deletion mutants of either PUSs or relevant snoRNAs would allow us to match sites to their cognate enzymes and increase our confidence in their accuracy. Indeed, Ψ-seq of deletion strains for snR34 or snR189 found specific reduction at the two known positions targeted by each in 18S and 25S rRNA, whereas other sites remained correlated, readily identifying the cognate target sites (Figure 2A).

Next, we applied Ψ-seq to deletion strains of each of four additional snoRNAs and eight nonessential PUSs (*PUS1*, *PUS2*, *PUS4*, *PUS5*, *PUS6*, *PUS7*, *PUS9*, and *DEG1*) (Figure 2B).

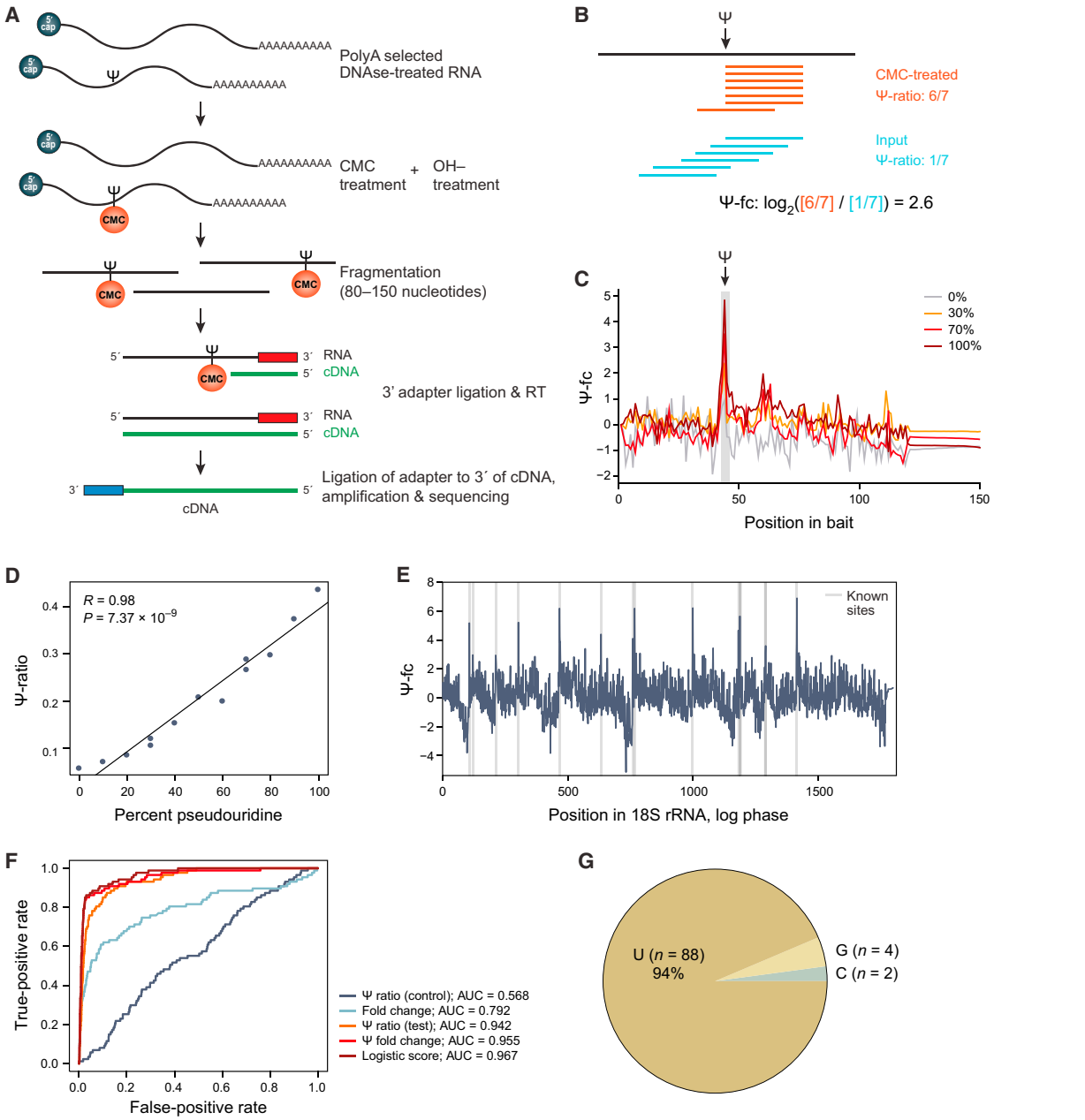


Figure 1. Ψ-Seq Quantitatively Measures Transcriptome-wide Pseudouridylation Profiles

(A) Ψ-seq procedure. Poly(A)-selected RNA is treated with CMC, which covalently binds to U and Ψ residues and, to a limited extent, to G (Ho and Gilham, 1971; Metz and Brown, 1969a, 1969b). This is followed by incubation at alkaline pH, leading to hydrolysis of U-CMC adducts, which are less stable than Ψ-CMC counterparts. RNA is next fragmented to a size range of 80–150 nt, followed by adaptor ligation to the 3′ end. Reverse transcription is primed off of the adaptor, expected to lead to premature termination one base pair immediately downstream of pseudouridylated sites. A second adaptor is ligated to the 3′ end of the cDNA, and libraries are amplified and sequenced in paired end mode.

(B) Scoring modified sites. Ψ ratio between reads terminating at a site and reads overlapping it and Ψ-fc, the fold change in Ψ ratio in the CMC-treated sample (red, top) over nontreated control (blue, bottom).

(C) Validation with a synthetic spike-in. Ψ-fc (y axis) at each position (x axis) across a synthetic spike-in harboring one pseudouridylated site at varying stoichiometries (color legend) at the indicated position (arrow).

(D) Ψ-seq quantifies relative stoichiometry. Scatterplot comparing Ψ ratios (y axis) at the pseudouridylated site in the spike-in with pseudouridine stoichiometries (x axis). The Pearson correlation coefficient (*r*) and associated P value are denoted.

(E) Ψ-seq detects known Ψ sites in rRNA. Ψ-fc values (y axis, dark blue) at each position (x axis) in 18S rRNA, overlaid with all known Ψ sites in this subunit (gray vertical lines).

(legend continued on next page)

Across the 14 experiments, we identified 291 sites with median Ψ-fc > 3 and median Ψ ratio > 0.1, with 55/291 dependent on a single PUS by stringent criteria (Experimental Procedures and Figure 2B). Ψ-seq of a CBF5 conditional knockdown detected 67 Cbf5p-dependent sites (Experimental Procedures and Figure 2C).

Overall, we called 328 unique sites across these experiments ("pan Ψ collection"; Table S1), 108 of which were associated with one or more PUS and/or snoRNA ("PUS-dependent Ψ collection"). These sites spanned various classes of coding and noncoding RNA (Figures 2D and 2E). Although subsequent analysis focuses on the higher-confidence, PUS-dependent collection, many of the additional sites in the pan Ψ collection are likely truly pseudouridylated but did not pass our stringent criteria of being differentially pseudouridylated in one condition. Alternatively, Ψ at some sites may be catalyzed by more than one pseudouridine synthase and thus failed to pass our stringent thresholds.

Unbiased Analysis of Ψ in rRNA, tRNA, and snRNA

The PUS-dependent collection associated the overwhelming majority of known sites with their correct PUS and snRNA dependencies, demonstrating the power of our approach. For rRNA, 24/24 Cbf5p-dependent sites are known targets of pseudouridylation, and 8/9 associations between snoRNAs and rRNA sites are known, as is the identified Pus5p-mediated Ψ site in mitochondrial 21S rRNA. One unknown, snR3-dependent site at position 2140 on 25S rRNA will require further verification.

For tRNA, many known sites and their associated modifiers were identified, including Pus1p-mediated Ψ sites at positions 26–28 in several tRNAs, Pus7p-dependent sites at position 13 of glutamate tRNA, Pus9p-dependent site in mitochondrial aspartate tRNA at position 32, and Deg1p-dependent sites at positions 37–38 in several tRNAs.

For snRNAs, we correctly identified all three previously identified sites in positions 35, 42, and 44 of U2 snRNA and their dependence on Pus7p, Cbf5p, and Pus1p, respectively. We also detected the known site in U5 snRNA and found that it depends on Cbf5p (not previously known).

Cbf5p-Dependent Pseudouridylation in the 5′ Guiding Sequence of C/D Box and H/ACA Box snoRNAs

Although Ψ content within snoRNAs has been reported (Ni et al., 1997), no Ψ site within snoRNA has been localized. The PUS-dependent Ψ collection had six Cbf5p-dependent snoRNA sites, four within C/D box snoRNAs, and two in H/ACA box snoRNAs. All four Cbf5p-dependent sites in C/D box snoRNAs occurred in the 5′ terminus of the snoRNA guiding sequence (Figure 3A, arrows). Pseudouridylation within this stretch is conserved, as 6/9 Ψ sites in C/D box snoRNAs identified in human HEK293 cells (below) are present within the short complementarity stretch of the snoRNA (Figure 3B). The pan Ψ collection had 21 additional sites in H/ACA box snoRNAs: 10/21 in H/ACA box snoRNAs with resolved hairpin (targeting) structures (Piekna-Przybylska et al., 2007) and 4/10 in the 5′ arm of the guiding sequence (none in the 3′ arm) (Figure 3C, arrows). We cannot conclusively determine that these sites are Cbf5p dependent because our detection power is limited by the dramatic decrease in the levels of H/ACA box snoRNAs upon Cbf5p knockdown (Figure 3D), which is consistent with their stabilization via Cbf5p binding (Lafontaine et al., 1998).

PUS-Dependent Ψ Sites in mRNAs at Consensus Sequences of Cognate Pseudouridine Synthases

Whereas no Ψ sites have been previously reported in mRNAs, Ψ-seq uncovered 41 PUS-dependent sites in 41 mRNAs from yeast grown in midlog phase. These include 34 Cbf5p-dependent sites, occurring at a variety of consensus sites, suggesting that they may be mediated by snoRNAs (below). Conversely, the seven sites dependent on one of three nonessential pseudouridine synthases (Pus1p, Pus4p, and Pus7p) were typically associated with distinct consensus motifs. Each of the three Pus4p-dependent sites (in *TEF2*, *MPM1*, and *ICL2*) resided in a "RRUUCNA" motif (underlined U indicates the putative site), perfectly conforming to the known Pus4p "GGUUCRA" target sequence at position 55 of tRNA (Becker et al., 1997a). One Pus7p-dependent site in *SPI1* was at a "UGUAA" sequence, matching a "UNUAR" consensus at Pus7p targets in tRNAs and in snRNAs (Behm-Ansmant et al., 2003; Decatur and Schnare, 2008) (and which we expand to "UGUAR" below).With one exception, all Pus1p-dependent targets (in coding and noncoding genes) were preceded by an A, suggesting partial, but lesser, sequence specificity. We did not observe any positional bias along coding sequences (we only analyzed coding regions due to very short UTRs, limited annotation, and very few Ψ sites in noncoding regions).

Cbf5p-Dependent Ψ Sites in mRNAs and snoRNAs Are Likely snoRNA Guided

As the substrate specificity of Cbf5p is typically mediated through H/ACA box snoRNAs, we hypothesized that the Cbf5p-dependent sites in mRNAs were also snoRNA guided. To test this hypothesis, we compiled all known H/ACA box-targeting arms and used a thermodynamics-based method to assess the folding energy of each snoRNA against each Cbf5p-dependent Ψ site (Experimental Procedures). For each site, we identified the snoRNA with the minimal predicted free energy (strongest cofolding) and calculated a z score for this association with respect to the predicted energies of association with all remaining snoRNAs. Validating our approach, we found that known rRNA Ψ sites had significantly stronger associations with H/ACA box snoRNAs than their shuffled counterparts (Figure 3E).

Overall, Cbf5p-dependent sites in mRNAs and snoRNAs had stronger associations with targeting regions in H/ACA box snoRNAs than shuffled controls. This was partly evident when comparing the two distributions (one-sided Wilcoxon test,

(F) ROC curves for different metrics (color legend) for calling putative Ψ sites. Each classifier was trained based on Ψ-seq data from human rRNA and tested on their performance in yeast rRNA.

(G) Nucleotide distribution across all sites passing our detection thresholds in a sample in midlog yeast.

See also Figure S1.

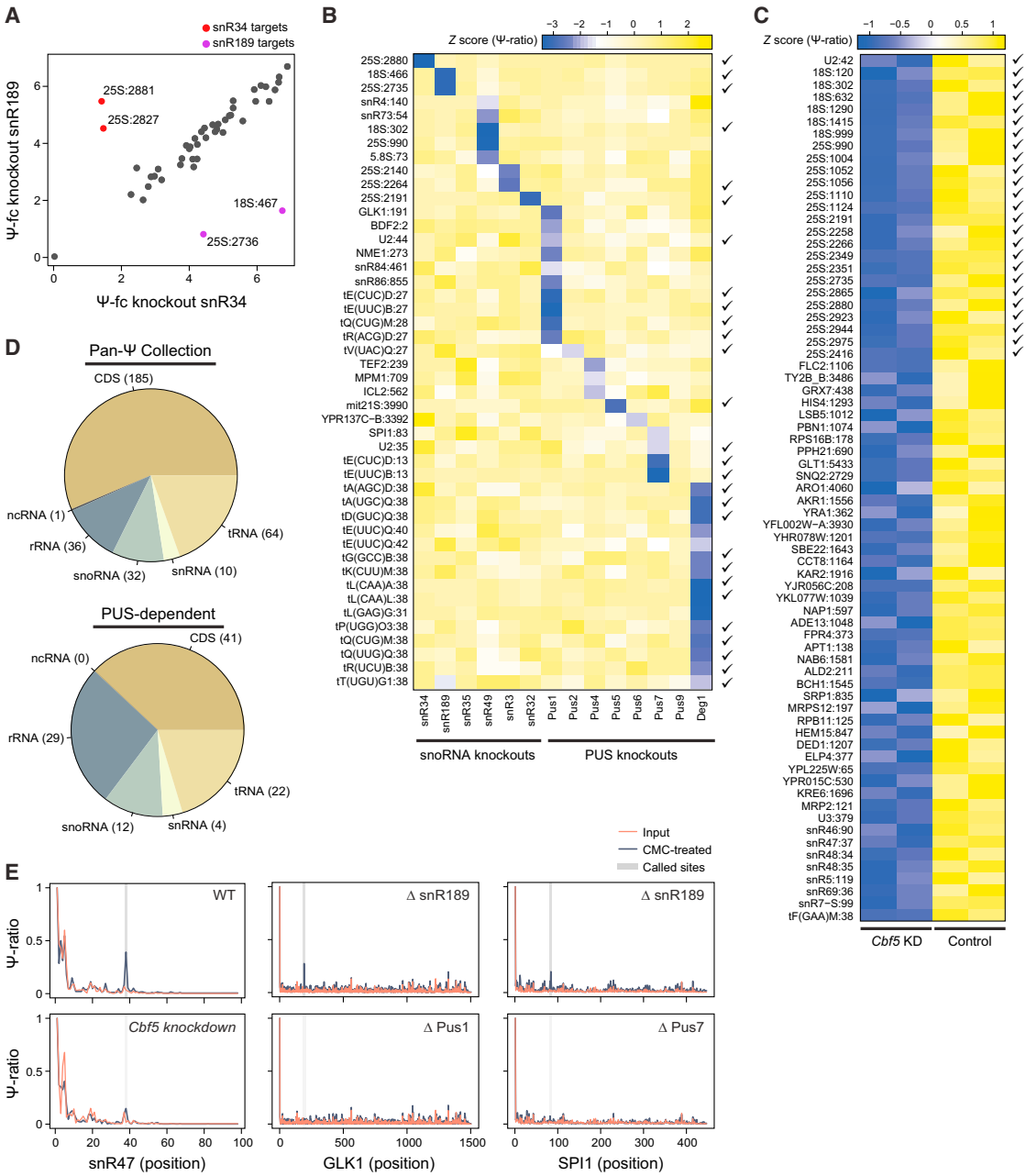


Figure 2. Identification of PUS-Dependent Ψ Sites in Yeast

(A) Ψ-seq correctly distinguishes targets of snoRNAs. A comparison of Ψ-fc scores across all known Ψ sites in three rRNAs (18S, 28S, and 5.8S) measured in yeast strains in which either H/ACA box snR34 (x axis) or snR189 (y axis) were deleted. snR34 is known to target positions 2880 and 2886 in 25S rRNA (red), and snR189 targets positions 2735 in 25S and 466 in 18S rRNA (purple), all detected as outliers in these experiments.

(B) Ψ-seq associates modified bases with their cognate PUS and snoRNAs. Heatmap depicting Ψ ratios (z scores per row, color bar) across all sites (rows, gene, and position labeled on left) in the PUS-dependent collection that are dependent on a snoRNA or a nonessential site-specific PUS (columns). For tRNAs, the labeled position is with respect to a multiple alignment of tRNAs with manual correction at two sites to allow comparison between sites. A check mark indicates that a site was both previously known to be pseudouridylated and that its catalysis was known to be mediated through the enzyme or snoRNA with the minimal z score in the figure.

(C) Ψ-seq identifies Cbf5p-dependent Ψ sites, as in (B) for all sites (rows) dependent on Cbf5p.

(D) Distribution of Ψ sites in different RNA classes, in the Pan-Ψ collection (left), and in the PUS-dependent collection (right). CDS: mRNA coding sequence.

(E) Ψ ratios (y axis) across the snoRNA snR47 (left), the mRNA *GLK1* (middle), and the mRNA *SPI1* (right), which depend on Cbf5p, Pus1p, and Pus7p, respectively. Ψ ratios are presented for CMC-treated (black) and nontreated (red) samples. Called Ψ sites are marked in vertical thick gray lines, and their "absence" position (in the relevant deletion strain) is marked by a more transparent outline.

See also Figure S2 and Table S1.

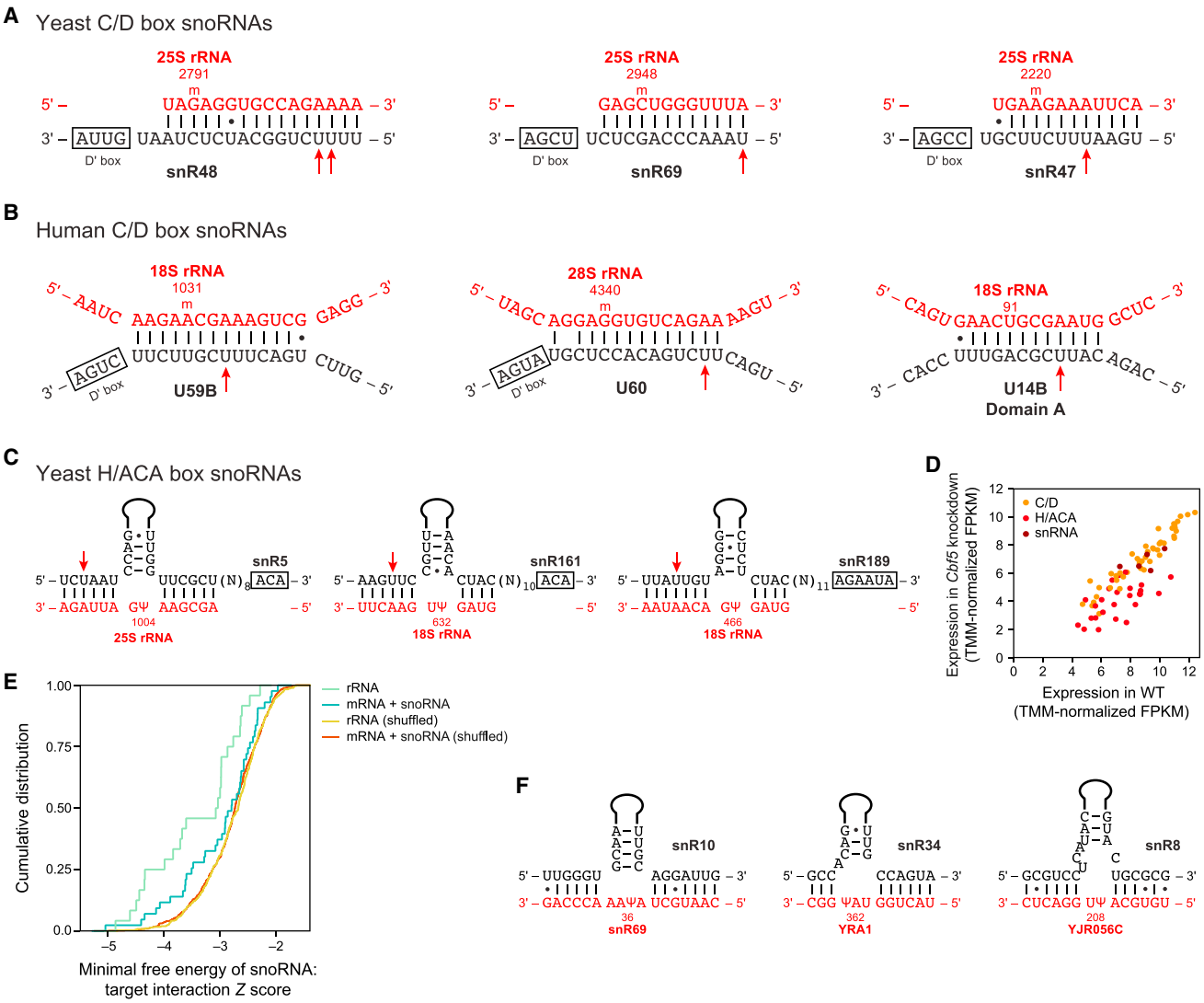


Figure 3. Pseudouridylation in Yeast snoRNAs

(A) Complementarity regions between three C/D box snoRNAs in yeast (black) and target sites in rRNA (red). Red arrows indicate detected Ψ sites.

(B) As in (A), but in human.

(C) As in (A) but for H/ACA box snoRNAs.

(D) snoRNA expression in WT strains (x axis) compared to *CBF5* knockdown strains (y axis), showing decreased expression of H/ACA box snoRNA (red), but not C/D box snoRNAs (orange) or snRNAs (dark red), upon knockdown.

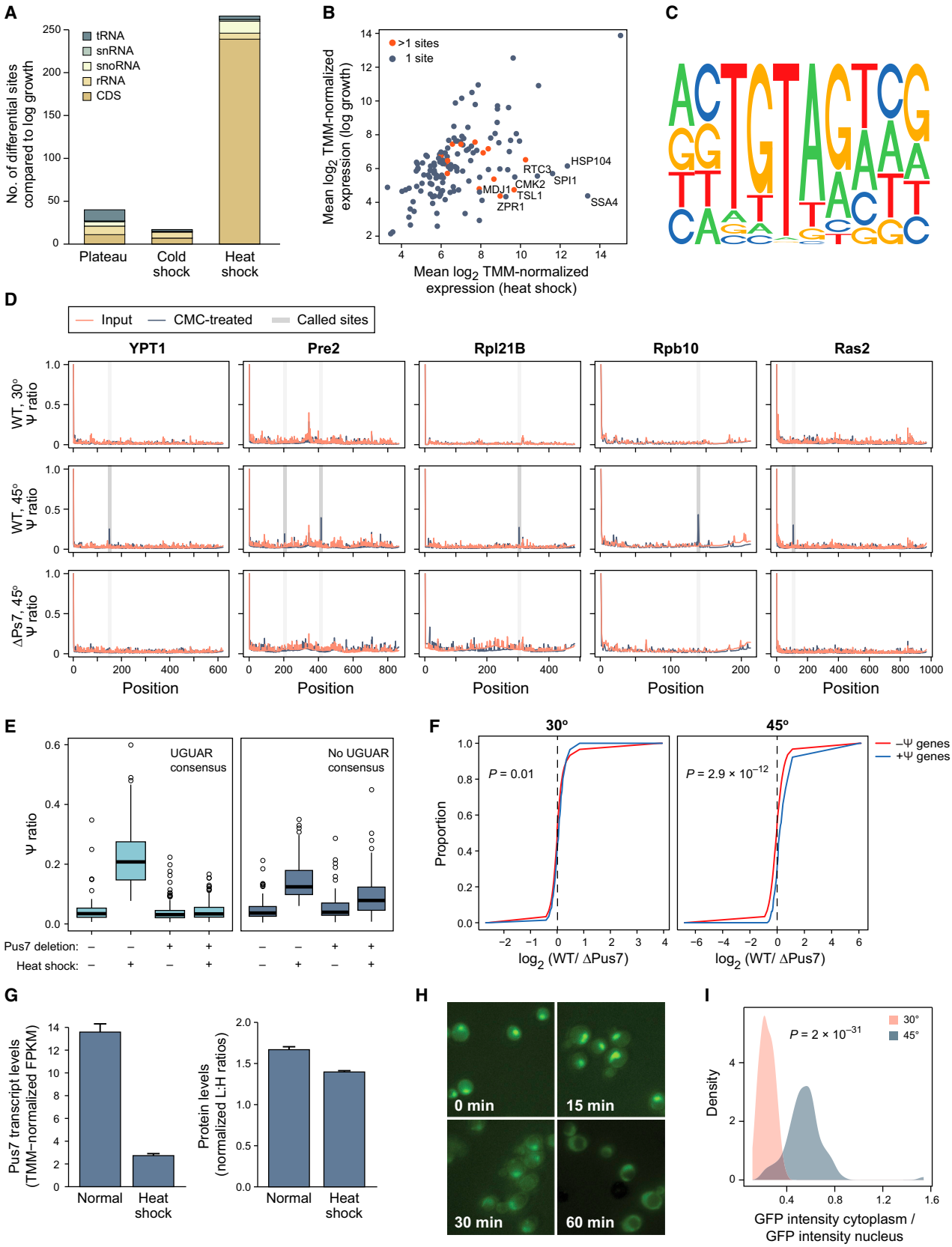
(E) Cumulative distribution frequencies of minimal z score transformed free energies, predicted by cofolding each putative Cbf5p-dependent (or shuffled) site against all known H/ACA box targeting arms. Distributions are plotted separately for sites in the rRNA (green), all remaining sites in mRNA and snoRNAs (blue), and their corresponding shuffled versions (yellow and red, respectively).

(F) Predicted associations among three Cbf5p-dependent sites (red) and top-scoring H/ACA box snoRNAs (black).

p = 0.07; Figure 3E) and was most prominent for outliers (e.g., Figure 3F). For example, 23.2% of Cbf5p-dependent sites in mRNA and snoRNAs were associated with an H/ACA box with a z score < −3.5, as compared to 11% for shuffled controls (χ^2 test, $p = 5.8 \times 10^{-8}$).

Dynamic Pus7p-Mediated Pseudouridylation of mRNAs in Heat Shock May Affect RNA Levels

The PUS-dependent sites in mRNAs raised the possibility that they may be differentially regulated under different conditions. To test this hypothesis, we compared Ψ-seq profiles in log phase to those from yeast in growth saturation, cold shock, and heat shock. Although only a few positions selectively acquired Ψ in stationary phase or cold shock (Figure 4A), 265 putative Ψ sites were induced in heat shock (45°C) versus log growth (30°C) in three independent replicates (Figure 4A and Table S2). A few are in key heat-shock-induced genes (e.g., *HSP60*, *HSP104*, *MDJ1*, and *SSA4*; Figure 4B), but most are in transcripts expressed at similar levels in heat shock and vegetative growth (Figure 4B), suggesting that



(legend on next page)

they are differentially pseudouridylated rather than differentially expressed.

The majority of heat-shock-acquired sites (159/265, 60%) occur in a perfect "UGUAR" Pus7p consensus sequence (R = G > A) (Figure 4C), suggesting that the pseudouridylation program in heat shock is induced through Pus7p. To test this hypothesis, we applied Ψ-seq to a Δ*pus7* strain in normal and heat shock conditions. Ψ sites harboring the Pus7p consensus were reduced to background levels in the Δ*pus7* strain in both conditions (Figures 4D and 4E). Thus, there is a widespread, Pus7p-mediated, pseudouridylation program in heat shock that encompasses the previously identified position in U2 snRNA induced in heat shock (Wu et al., 2011). Importantly, the distribution of Ψ ratios in UGUAR sites in mRNA in heat shock is comparable to that of rRNA sites, suggesting that mRNA sites are pseudouridylated to high stoichiometries (Figure S3A).

To explore the role of pseudouridylation, we compared the expression level (from RNA-seq) between wild-type (WT) and Δ*pus7* yeast for 142 genes that harbor Pus7p-dependent Ψ sites in heat shock versus the remaining, nonpseudouridylated genes (Figure 4F) in either heat shock or normal conditions. In heat shock, pseudouridylated genes were expressed at ~25% higher levels in WT strains than in Δ*pus7* strains, whereas nonpseudouridylated genes were expressed at roughly equal levels in the two strains. In contrast, the two sets were comparably expressed in WT and Δ*pus7* at 30°C. These differences in expression only in pseudouridylated genes and only in heat shock suggest that pseudouridylation may contribute to RNA stability. We did not, however, find a significant correlation between the Ψ ratio and the change in expression. Future studies of RNA half-life with methods compatible with dynamic responses (Miller et al., 2011) will help study this phenomenon.

We explored the basis for Pus7p's increased activity in heat shock. Surprisingly, both Pus7p RNA and protein levels are reduced in heat shock (Figure 4G), but there are significant differences in Pus7p localization; it is predominantly nuclear in 30°C and primarily cytoplasmic in heat shock (Figures 4H and 4I). Thus, Pus7p localization may underlie its enhanced repertoire of substrates in heat shock. Supporting a functional role for Pus7p in heat shock, Δ*pus7* yeast show growth defects after heat shock in comparison to WT (data not shown), which is consistent with a previous report on their increased heat sensitivity (Sinha et al., 2008).

Finally, we explored factors that may govern modification specificity. First, in the most highly expressed genes, ~30% of UGUAG-containing loci are modified, suggesting that the consensus plays a decisive role (Figure S3B). Furthermore, even among the remaining 70%, Ψ ratios were increased in WT compared to Δ*pus7* in heat shock (but not at 30°C), suggesting that many are, in fact, modified (Figure S3C). Nonetheless, UGUAR loci called as modified were more likely to have an A/U at position +3 and a pyrimidine at position −4 (Figure S3D), suggesting that features beyond UGUAR also contribute to specificity.

Pseudouridylation Profiles and Mechanisms Are Conserved in Human

To assess the extent of similarity between yeast and human pseudouridylation, we performed Ψ-seq in HEK293 cells and fibroblasts (below). Applying the same thresholds, we identified 396 putative Ψ sites (Table S3)—353 sites in mRNA and 43 sites in noncoding RNAs (excluding rRNA). We found no functional enrichments in pseudouridylated genes compared to an expression-matched control set, and the sites were relatively uniformly distributed across coding and untranslated regions (data not shown).

To associate pseudouridine synthases to Ψ sites, we examined the sequences surrounding the sites. The most enriched motif (Figure 5A, 70/396 sites) had a GUUC core and was generally characterized by "RGUUCNANYCY," perfectly corresponding to the known Pus4p-modified site in tRNAs and extending the Pus4p consensus in yeast. Another enriched motif (Figure 5B), "UGUAG," present in eight sites, is consistent with a mammalian Pus7p homolog. Thus, homologs of the same two site-specific pseudouridine synthases are likely active on yeast and mammalian mRNA.

The similarity between Ψ sites in yeast and mammals suggested that the other 318 sites, lacking a Pus4p or Pus7p consensus, may be modified by *DKC1*/Dyskerin, the mammalian *CBF5* ortholog. To test this hypothesis, we used siRNAs to knock down *DKC1* to ~8% of WT RNA levels in HEK293 cells (data not

Figure 4. Induction of Pus7p-Dependent Sites in Heat Shock

(A) Bar plots of the number of differentially modified sites (y axis) in each of three conditions (x axis) relative to midlog yeast. Sites are color coded based on RNA class (color legend).

(B) Differential modification in heat shock is not merely due to changes in expression levels. Expression levels in heat shock (x axis) or non-heat-shock (y axis) conditions for all genes harboring heat-shock-induced Ψ sites.

(C) Sequence motif learned across all Ψ sites induced in heat shock is strongly enriched for the UGUAR Pus7p motif.

(D) Ψ ratios (y axis) in CMC-treated (blue) and nontreated (red) samples in five genes (columns) acquiring Ψ sites in heat shock. Ψ ratios are shown in WT in 30° (top) and heat shock (45°, middle) and in heat shock in Δ*pus7* (bottom). Called sites are depicted in vertical gray bars in the samples in which they are called, and in more transparent gray in the remaining samples.

(E) Box plots of Ψ ratios in WT or Δ*pus7* strains, grown in heat shock or 30°, for 159 heat-shock-induced sites harboring a Pus7p consensus (left) and 106 sites without the consensus (right).

(F) Cumulative distribution plots of fold changes of expression levels between WT and Δ*pus7* strains (x axis), comparing genes harboring a Pus7p-induced Ψ site (blue) to genes lacking it (red). At 45° (left), genes harboring Ψ sites are expressed at higher levels in the WT strain than in Δ*pus7*, whereas nonpseudouridylated genes are present at similar levels. At 30° (right), these differences are marginal. p value of the K-S test is depicted.

(G) PUS7 transcript (left; mean TMM-normalized FPKM values from three replicates) and protein (right, data in four replicates from Nagaraj et al. [2012]) levels (y axis) in heat shock and non-heat-shock conditions. Error bars are SEM.

(H and I) Pus7p localization. Fluorescent microscopy images (H) and quantification (I) during a heat shock time course show mostly nuclear localization at 30°C (time 0 in H and red in I), which is significantly reduced after heat shock (time 60 in H, blue in I). p value of a Mann-Whitney test is noted.

See also Figure S3 and Table S2.

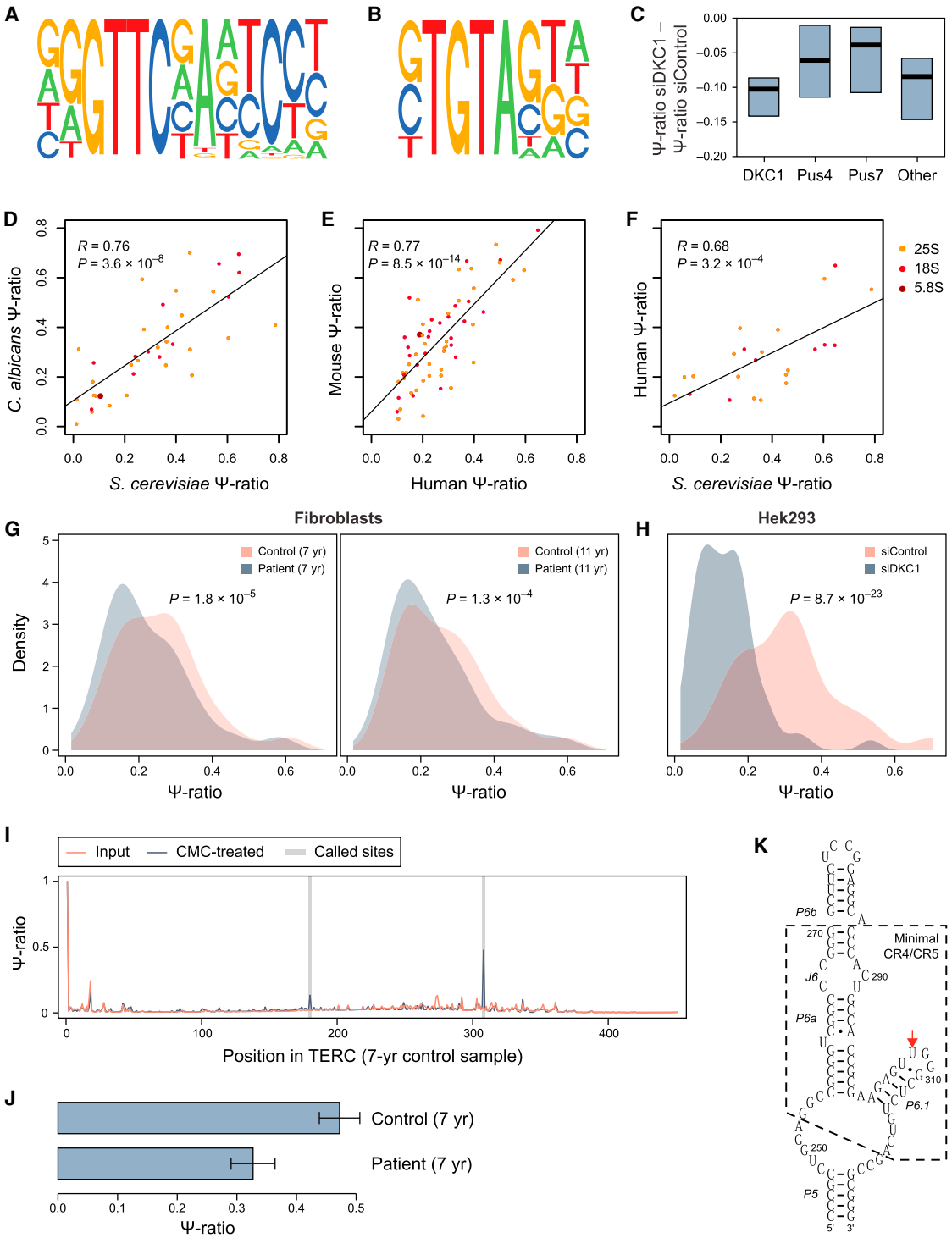


Figure 5. Ψ-Seq of Human RNA Highlights Conserved Features and Disease Relevance

(A and B) Sequence motifs from 70 sites harboring a Pus4-associated GUUC core (A) and 13 sites harboring a Pus7p-associated UGUA core (B) from among 396 modified sites in HEK293 RNA.

(C) Difference in Ψ ratios in *DKC1* knockdown relative to control (y axis) in HEK293 cells. Box plots from left to right: *DKC1*-dependent sites, sites harboring Pus4p (GUUC) or Pus7p (UGUAR) associated motifs, and all remaining sites.

(D–F) Correlation of pseduoridylation stoichiometries in rRNA between homologous positions in *S. cerevisiae* and *C. albicans* rRNA (D), human and mouse (E), and yeast and human (F). Positions are color coded based on rRNA subunit.

(legend continued on next page)

shown) and found significant reduction (Figure 5C and Experimental Procedures) in Ψ levels at 58/396 sites. The 58 *DKC1*-dependent sites had significant complementarity to human snoRNAs ($p < 0.03$, one-sided Kolmogorov-Smirnov (K-S) test versus randomly shuffled sequences using the same approach as in yeast). In contrast, Ψ levels were only negligibly reduced upon knockdown in sites containing either a Pus4p or a Pus7p motif (Figure 5C). Many of the 262 sites that we had not associated with *DKC1* (by knockdown), Pus4p, or Pus7p (by motif) had borderline effects from *DKC1* knockdown and may thus be substrates missed by our stringent criteria.

The number of Ψ sites we report in human is likely a very conservative lower bound due to our stringent thresholds and requirement for very deep coverage, in particular at moderately or lowly expressed transcripts (Figure S2A). Indeed, when slightly relaxing the criteria of calling a Ψ site (minimal Ψ ratio: 0.05; minimal Ψ-fc: 2.5), we identified 176 additional Dyskerin-dependent sites. Only two sites are substantially increased upon *DKC1*/*CBF5* knockdown, suggesting a <1% false detection rate. Thus, as in yeast, mammalian mRNAs have a rich landscape of Ψ sites, mediated by at least three orthologous PUSs.

rRNA Ψ Levels Are Conserved from Yeast to Human

We next examined whether the extent (stoichiometry) of pseudouridylation was conserved across species, focusing on rRNA positions that are largely conserved throughout eukaryotes and substantially covered in our data. To enhance our comparisons, we also collected Ψ-seq data from log phase *Candida albicans* and from mouse bone-marrow-derived dendritic cells. We found that Ψ levels were quantitatively conserved between *S. cerevisiae* and *C. albicans* ($R = 0.76$, Figure 5D), between human and mouse ($R = 0.77$, Figure 5E), and even between yeast and human ($R = 0.68$, Figure 5F), suggesting that both the presence and relative abundance of rRNA Ψ sites are under strong purifying selection.

Reduced rRNA Ψ Levels in Cells from Dyskeratosis Congenita Patients

It has been hypothesized that aberrant rRNA pseudouridylation could underlie X-linked dyskeratosis congenita, caused by mutations in *DKC1*/dyskerin, but the extent to which it is altered remains controversial (Wong and Collins, 2006; Bellodi et al., 2013; Gu et al., 2009; Mochizuki et al., 2004).

Applying Ψ-seq to fibroblasts derived from 7- and 11-year-old patients harboring two distinct mutations (del37L and A386T, respectively) and to cells from age-matched controls, we found subtle—yet significant—decreases in rRNA pseudouridylation among patients (Figure 5G), and Ψ-seq profiles further group in unsupervised analysis by disease and not by age (Figure S4A). The subtle effect, on average a reduction of 10% in Ψ ratio per site, likely explains why these were difficult to observe with less sensitive techniques. As a reference, rRNA pseudouridylation per site is reduced on average a dramatic 55% upon *DKC1* knockdown (Figure 5H).

Decreased Pseudouridylation in Telomerase RNA Component in Dyskeratosis Congenita Cells

Dyskeratosis congenita can also be caused by mutations in the noncoding RNA (ncRNA) telomerase component, telomerase RNA component (TERC) (Hoareau-Aveilla et al., 2008; Trahan and Dragon, 2009). TERC is bound and stabilized by DKC1 but is destabilized in dyskeratosis congenita (Ashbridge et al., 2009). Our findings of Ψ sites within snoRNAs (Figure 3) prompted us to examine whether TERC, which harbors an H/ACA box domain, might be pseudouridylated as well in a DKC1-dependent manner. Although TERC sites were not called in HEK293 cells, the read distribution across TERC revealed a putative Ψ site at position 307 (Figure S4B). This site was not called due to low read counts in the lowly expressed TERC. TERC expression was even lower in the fibroblasts (patients FPKM: 0.15 and 0.28; controls: 2.8 and 0.8), precluding any analysis.

To detect and quantify Ψ sites in TERC, we first enriched it in RNA from fibroblasts of patients and age-matched controls using hybrid capture with RNA antisense purification (RAP) (Engreitz et al., 2014) (Experimental Procedures). This resulted in ~750-fold enrichment in the fraction of reads in the CMC-treated samples aligning to TERC compared to poly(A)-selected RNA. Ψ-seq of the hybrid-captured RNA confirmed substantial pseudouridylation of position 307 (Figure 5I), a highly conserved uridine (Figure 5K) in a region essential for telomerase activity and TERT binding (Chen et al., 2002) and showed that it is modified at significantly higher levels in the control sample (Ψ ratio = 0.47) than in the patient sample (Ψ ratio = 0.33) (Figure 5J, χ^2 test, $p = 0.006$). A similar reduction was observed when comparing samples from the 11-year-old patient to the age-matched control (χ^2 test, $p = 0.01$; Figure S4C). Another putative Ψ site was found at position 179—albeit at substantially lower levels (Ψ ratio = 0.11). This confirms two of the six previously proposed Ψ positions in TERC (we do not find evidence for the others) (Kim et al., 2010), suggests that TERC pseudouridylation may be compromised in dyskeratosis congenita, and provides a general way to quantify Ψ in lowly expressed genes.

DISCUSSION

Ψ-seq profiles Ψ in an unbiased, quantitative, transcriptome-wide manner at single-nucleotide resolution. It discovered pseudouridines in *S. cerevisiae* and human snoRNAs and mRNA,

(G) Distributions of Ψ ratios across rRNA positions in fibroblasts from dsykeratosis patients (gray) and age-matched controls (pink). Left, 7-year-old patient; right, 11-year-old patient.
(H) Ψ ratios in rRNA positions in *DKC1* knockdown (gray) or control (pink) in HEK293 cells.
(I) TERC modification is affected by *DKC1* knockdown. Ψ ratios at each site in the TERC transcript (x axis). Ratios are shown in both CMC-treated (blue) and nontreated (red) samples. The putative Ψ positions are highlighted in a gray vertical bar.
(J) Ψ ratios for position 307 in 7-year-old patient and age-matched control; error bars indicate SE.
(K) RNA secondary structure of the CR4/CR5 region in TERC redrawn based on Zhang et al. (2011). Red arrow indicates putative Ψ site.
See also Figure S4 and Table S3.

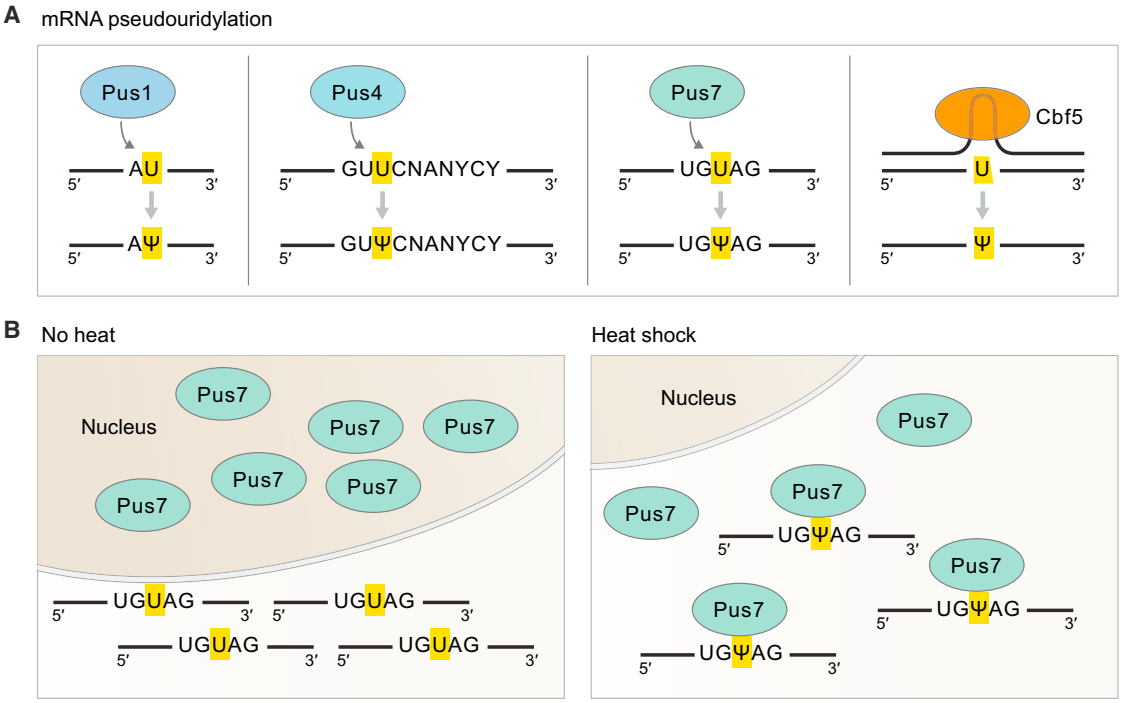


Figure 6. mRNA Pseudouridylation

(A) mRNA pseudouridylation in eukaryotes is mediated by at least four conserved pseudouridine synthases, some site specific (Pus1p, Pus4p, and Pus7p) and others snoRNA mediated (yeast Cbf5p/human DKC1).

(B) Pus7p may orchestrate the yeast mRNA pseudouridylation program in heat shock. At 30°, Pus7p is primarily nuclear, and its localization into the cytoplasm upon heat shock may induce mRNAs pseudouridylation.

catalyzed by both snoRNA-dependent pseudouridine synthases and site-specific pseudouridine synthases guided by consensus sequences. We identified a dynamic program induced in yeast heat shock in which Ψ sites accumulate at hundreds of mRNAs, possibly affecting RNA levels. Ψ-seq allows us to compare Ψ stoichiometries across rRNA positions, finding that these are highly conserved but reduced in cells from dyskeratosis congenita patients. Although we cannot exclude the possibility of some false positives, our stringent controls and excellent correspondence to prior knowledge suggest that their rate is low. The roles of pseudouridine synthases and Ψ are likely more complex than previously assumed, and our genome-wide method can help study this.

Ψ-seq can be further enhanced in several ways. First, to accurately estimate Ψ levels in lowly expressed transcripts, it can be combined with hybrid capture, as demonstrated for TERC, or with new methods to pre-enrich for pseudouridylated RNA. Second, Ψ-seq cannot monitor Ψ levels at 3′ transcript ends, as it relies on reverse transcription termination. To help probe Ψ in short RNA, we could couple Ψ-seq with alternative ligation protocols. Finally, Ψ-seq is theoretically limited in identifying two closely proximal pseudouridines, as reverse transcription is expected to always terminate at the downstream—and never at the upstream—site. In practice, however, this does not appear to be a major limitation; for example, we detect adjacent positions in rRNA (e.g., 25S:1052 and 1056) or in U2 snRNA (positions 42 and 44). This is likely due to incomplete modification stoichiometry, incomplete CMC labeling of Ψ bases, and/or imperfect efficiency of RT stopping at CMC-treated sites, though Ψ quantification at such adjacent sites may nonetheless be impacted.

Pseudouridine Synthase Substrate Specificity

Site-specific, nonessential pseudouridine synthases in *S. cerevisiae* were previously thought to modify only one (or a few) residues by recognizing secondary or tertiary structures in a specific RNA substrate (Becker et al., 1997b). Our results demonstrate that such pseudouridine synthases have a dramatically larger substrate repertoire than previously assumed and suggest that many pseudouridine synthases mediated modifications are guided in part by primary sequence features in their substrates (Figure 6A). This is conserved from yeast to mammals, where the predominant consensus is identical to that in yeast Pus4p, suggesting that Ψ is mediated by the mammalian orthologs TRUB1/TRUB2 (Zucchini et al., 2003). There are at least 23 putative pseudouridine synthases in human (Hunter et al., 2012), and one cannot fully rule out the possibility of additional pseudouridine synthases in yeast; perturbation experiments will help elucidate their targets, as well as resolve possible functional redundancies (Grosshans et al., 2001).

We also find evidence for mRNAs pseudouridylation via snoRNA-dependent PUSs. A more elaborate repertoire of snoRNA substrates, potentially mediated through imperfect complementarities, has previously been hypothesized (Wu

et al., 2011) based on the finding that heat-shock-/nutrient-deprivation-induced Ψ at position 93 in U2 snRNA is guided through imperfect pairing with snR81 (Wu et al., 2011).

Pus7p-Pseudouridylation Program in Heat Shock

The Pus7p-dependent induction of pseudouridylation during heat shock is an attractive system for further studies of Ψ function. The specificity of the modified site to heat shock (despite lower levels of Pus7p than in other conditions) and the similarity of distribution of Ψ ratios to rRNA could be consistent with a regulated system, but further functional studies are warranted. This program may be mediated through Pus7p localization in the cytoplasm in heat shock (Figure 6B). Because RNA levels of pseudouridylated targets are higher in WT compared to Δ*pus7*, it is tempting to speculate that Ψ may serve to stabilize transcript structure in heat shock. Alternatively, Ψ could target specific mRNAs to heat-shock-induced stress granules in a similar manner to m6A destabilization of RNA by directing it to P bodies (Wang et al., 2014). Such responses could help cells to better withstand heat shock, which is consistent with the increased heat sensitivity of the Pus7p deletion strain in *S. cerevisiae* (Sinha et al., 2008). Pseudouridylation of a stop codon was previously shown to suppress translation termination (Karijolich and Yu, 2011), suggesting that Ψ might modify the genetic code. We found only a single putative site in a stop codon in HEK293 cells and none in yeast; greater depth may revise these predictions.

Interestingly, Ψ levels at position 50 in 5S rRNA, catalyzed by Pus7p, also increase in heat shock, suggesting that it is not fully modified during exponential growth and that its pseudouridylation might confer an advantage during heat shock. These data challenge the notion that rRNA modifications are constant across growth conditions and suggest that the stoichiometry of rRNA modification may also change in response to external stimuli.

Pseudouridylation in Dyskeratosis Congenita

Pseudouridine is particularly important in dyskeratosis congenita, which can be caused by mutations in the CBF5 ortholog *DKC1*/dyskerin. Our results contribute two findings. First, we observed a subtle, yet significant, quantitative difference in Ψ levels in patient-derived samples as compared to age-matched controls, which were not previously observed in some studies (Wong and Collins, 2006) but suggested in others (Bellodi et al., 2013), possibly due to less sensitive methods. Second, we detect two putative Ψ sites in TERC, one of which at very high levels and at highly conserved position 307 within the essential P6.1 hairpin (Chen et al., 2002). A Ψ site was previously proposed at that position and was shown to significantly increase stability of the loop structure in vitro (Kim et al., 2010). Our findings may unify the two roles attributed to dyskerin—stabilizing TERC and pseudouridylating RNA—and suggest that dyskerin may both bind to and pseudouridylate TERC and that both processes, perhaps in a mechanistically coupled manner, may be disrupted in the disease.

Pseudouridylation in ncRNAs at Sites of Inter- or Intramolecular Interactions

It has previously been noted that Ψ sites in rRNA and snRNAs tend to occur in "important" regions that directly interact with other molecules. In U1, U2, U4, and U6 snRNAs, for example, Ψ sites occur at or close to a site that base pairs with intronic RNA to facilitate RNA splicing (Charette and Gray, 2000). In rRNA, there is a high density of conserved Ψ sites at the peptidyl-transferase center and at the decoding center of the 25S rRNA subunit, which interacts with mRNA and with the tRNA stem loop (Bakin et al., 1994; Bakin and Ofengand, 1993; Lane et al., 1992). Our study extends these observations to snoRNAs, where we find Ψ sites within both C/D and H/ACA box snoRNAs, at positions involved in base pairing with target sites. The highly conserved Ψ site in TERC is also in a region crucial for telomerase activity, involved in binding to TERT (Chen et al., 2002).

Ψ may also play a role in intramolecular interactions. For example, a Pus1p-dependent Ψ site at position 273 in NME1, the RNA component of yeast RNase MRP, is in a stem structure in the P2 region (Esakova and Krasilnikov, 2010; Esakova et al., 2013; Reilly and Schmitt, 1995-1996), and a site in the helix 2 loop of U3 is involved in preribosomal rRNA cleavage. Mapping Ψ sites in ncRNA molecules may help delineate sites involved in inter- and intramolecular interactions.

Our quantitative and high-resolution methodology coupled with genetic perturbations yielded unbiased, transcriptome-wide maps of Ψ across different RNA species, identified consensus sequences recognized by pseudouridine synthase enzymes, and established pseudouridine as a ubiquitous and dynamic modification of eukaryotic mRNA. Further studies will leverage this method and finding to elucidate Ψ's function in the RNA life cycle, especially of mRNA.

EXPERIMENTAL PROCEDURES

Yeast Growth Conditions

Unless otherwise noted, yeast were grown in YPD supplemented with 0.27 mM uridine and heat shock was performed at 45°C for the designated time points and collected as described in the Extended Experimental Procedures. Deletion strains are listed in the Extended Experimental Procedures.

Ψ-Seq

Ψ-seq was performed on PolyA+ RNA treated with CMC essentially as previously described (Bakin and Ofengand, 1993, 1998) and detailed in the Extended Experimental Procedures. Input controls were handled identically in all steps but without CMC treatment. Following the procedure, RNA-seq libraries were prepared by an RNA ligation method as detailed in the Extended Experimental Procedures and sequenced on Illumina HiSeq 2500 with 30 bp paired end reads.

Detection of Ψ Sites

Reads were mapped to the yeast genome (sacCer3) using Bowtie or to the human (hg19) genome using Tophat, with parameters and scripts as described in the Extended Experimental Procedures. For rRNA and snRNA sites, we aligned reads against specific database of these two classes. Mouse reads were aligned to mouse rRNA from Refseq. *C. albicans* reads were aligned to the *C. albicans* genome (SC5314 version A21). Ψ sites were identified by first calculating a Ψ ratio for each treated or input sample, followed by Ψ-fold change ($\log_2$ fold changes of Ψ ratios in the treated versus nontreated samples). Sites were filtered by parameters listed in the Extended Experimental Procedures. For all unique sites passing the filters in at least one sample, we calculated a Ψ ratio and Ψ-fc in each sample and considered all sites whose median Ψ ratio and median Ψ-fc exceeded 0.1 and 3, respectively (unless otherwise noted). We demanded that each 21 nt sequence surrounding the putative Ψ site be unique and also filtered sites in positions other than

immediately following "U." To compare orthologous rRNA sites across species, we used "needle" (Rice et al., 2000) for pairwise alignments of rRNAs with lift over of known Ψ sites (Machnicka et al., 2013) from human to mouse and from *S. cerevisiae* to *C. albicans*. For further details, see Extended Experimental Procedures.

Synthetic Spike-Ins

We in vitro transcribed an RNA bait containing only a single uridine (Extended Experimental Procedures) and mixed pseudouridylated and nonpseudouridylated RNA at indicated ratios.

PUS-Dependent Sites

We identified differentially pseudouridylated sites across deletion strains of nonessential pseudouridine synthases and snoRNAs, and we tested both the number of reads starting and overlapping at each putatively identified position between a given sample and the background of all samples with a procedure based on a χ^2 test detailed in the Extended Experimental Procedures and similar procedures for heat shock, Pus7p-dependent sites, Cbf5p-dependent sites, and *DKC1*-dependent sites.

Motif Analysis

We used position specific scoring matrices (PSSMs) defined by 21 bp sequences spanning the identified Ψ sites, visualized using SeqLogo in R, and displaying only regions with nonuniform sequence composition.

Prediction of Association between Cbf5p-Dependent Target Sites and snoRNAs

We assessed the hybridization potential between Cbf5p-dependent target sites and H/ACA box snoRNAs by calculating free energies of concatenated upstream and downstream complementary arms (Piekna-Przybylska et al., 2007), separated by a constant hairpin sequence, and folded against a 20 nt region surrounding each CBF5p-dependent site using RNAcofold (Gruber et al., 2008), with parameters and constraints described in the Extended Experimental Procedures. We shuffled each 20 nt region 15 times (maintaining a center U) as controls.

Pus7p Localization

Pus7p localization was measured by microscopy of *S. cerevisiae* encoding GFP-tagged Pus7p grown at 45°C using a Nikon Eclipse Ti with oil immersion at 100× magnification. ImageJ was used to quantitate GFP intensity with DAPI staining to define nuclear boundaries, as detailed in the Extended Experimental Procedures.

Bone-Marrow-Derived Dendritic Cells

Cells were prepared as previously described (Amit et al., 2009). All experiments were performed in compliance with the institutional guidelines and were reviewed and approved by the Institutional Animal Use and Care Committee (IAUCC) of MIT (Protocol 0612-058-15).

DKC1 Knockdown

Human HEK293 cells were plated in 6-well plates at 20% confluence. siRNAs targeting *DKC1* (catalog numbers s4111 and s4112) were transfected using Lipofectamine RNAiMAX (Life Technologies) following the manufacturer's protocols, with boosts at 48 and 96 hr; as negative controls, we used Ambion in vivo negative control 1 siRNA (catalog number 4457287). Cells were harvested at 144 hr.

Samples from Dyskeratosis Congenita Patients

Fibroblasts were obtained from Coriell (GM01774 [del37L, 7 years], AG04646 [A386T, 11 years], GM00409 – [healthy 7-year-old], and GM01864 – [healthy 11-year-old]).

RAP-RNA Enrichment of TERC

We performed RAP on intact isolated total RNA as previously described for RAP-RNA[AMT] (Engreitz et al., 2014).

ACCESSION NUMBERS

Sequencing data have been deposited into the Gene Expression Omnibus (GEO) under the accession number GSE60047.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, four figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2014.08.028.

AUTHOR CONTRIBUTIONS

S.S., D.B., M.R.M., E.S.L., G.F., and A.R. conceived and designed the study. S.S. led the development of experimental and computational methods. D.A.B. developed and executed all yeast experiments. M.R.M. developed and conducted all genomic assays and mammalian experiments. M.J. conducted proteomics experiments. R.H.B., R.S., and B.X.L.-R. analyzed data. J.M.E. and M.G. developed the RNA hybrid selection protocol. E.S.L., G.F., and A.R. guided work. S.S. and A.R. wrote the manuscript with input from all authors.

ACKNOWLEDGMENTS

Work was supported by NHGRI P50HG006193, Pioneer Award, HHMI (A.R.), NHGRI U54 HG003067 (E.S.L.) Broad Institute Funds, NIH GM035010 (G.F.), HFSP fellowship (S.S.), NIH 1F32HD075541-01 (R.S.), and fellowships of the Swiss National Science Foundation and the Marie Curie IOF (M.J.).

Received: July 30, 2014
Revised: August 21, 2014
Accepted: August 22, 2014
Published: September 11, 2014

REFERENCES

Amit, I., Garber, M., Chevrier, N., Leite, A.P., Donner, Y., Eisenhaure, T., Guttman, M., Grenier, J.K., Li, W., Zuk, O., et al. (2009). Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. Science *326*, 257–263.

Ansmant, I., Massenet, S., Grosjean, H., Motorin, Y., and Branlant, C. (2000). Identification of the Saccharomyces cerevisiae RNA:pseudouridine synthase responsible for formation of psi(2819) in 21S mitochondrial ribosomal RNA. Nucleic Acids Res. *28*, 1941–1946.

Ashbridge, B., Orte, A., Yeoman, J.A., Kirwan, M., Vulliamy, T., Dokal, I., Klenerman, D., and Balasubramanian, S. (2009). Single-molecule analysis of the human telomerase RNA.dyskerin interaction and the effect of dyskeratosis congenita mutations. Biochemistry *48*, 10858–10865.

Bakin, A., and Ofengand, J. (1993). 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.

Bakin, A., and Ofengand, J. (1995). Mapping of the 13 pseudouridine residues in Saccharomyces cerevisiae small subunit ribosomal RNA to nucleotide resolution. Nucleic Acids Res. *23*, 3290–3294.

Bakin, A.V., and Ofengand, J. (1998). Mapping of pseudouridine residues in RNA to nucleotide resolution. Methods Mol. Biol. *77*, 297–309.

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

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

Becker, H.F., Motorin, Y., Sissler, M., Florentz, C., and Grosjean, H. (1997b). Major identity determinants for enzymatic formation of ribothymidine and pseudouridine in the T psi-loop of yeast tRNAs. J. Mol. Biol. *274*, 505–518.

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

Bellodi, C., McMahon, M., Contreras, A., Juliano, D., Kopmar, N., Nakamura, T., Maltby, D., Burlingame, A., Savage, S.A., Shimamura, A., et al. (2013). H/ACA small RNA dysfunctions in disease reveal key roles for noncoding RNA modifications in hematopoietic stem cell differentiation. Cell Rep. *3*, 1493–1502.

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

Chen, J.L., Opperman, K.K., and Greider, C.W. (2002). A critical stem-loop structure in the CR4-CR5 domain of mammalian telomerase RNA. Nucleic Acids Res. *30*, 592–597.

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

Decatur, W.A., and Schnare, M.N. (2008). Different mechanisms for pseudouridine formation in yeast 5S and 5.8S rRNAs. Mol. Cell. Biol. *28*, 3089–3100.

Durant, P.C., and Davis, D.R. (1997). The effect of pseudouridine and pH on the structure and dynamics of the anticodon stem-loop of tRNA(Lys,3). Nucleic Acids Symp. Ser. *36*, 56–57.

Engreitz, J.M., Sirokman, K., McDonel, P., Shishkin, A., Surka, C., Russell, P., Grossman, S.R., Chow, A.Y., Guttman, M., and Lander, E.S. (2014). RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent pre-mRNAs and chromatin sites. Cell *159*, this issue, 188–199.

Esakova, O., and Krasilnikov, A.S. (2010). Of proteins and RNA: the RNase P/MRP family. RNA *16*, 1725–1747.

Esakova, O., Perederina, A., Berezin, I., and Krasilnikov, A.S. (2013). Conserved regions of ribonucleoprotein ribonuclease MRP are involved in interactions with its substrate. Nucleic Acids Res. *41*, 7084–7091.

Fujiwara, T., and Harigae, H. (2013). Pathophysiology and genetic mutations in congenital sideroblastic anemia. Pediatr. Int. *55*, 675–679.

Grosshans, H., Lecointe, F., Grosjean, H., Hurt, E., and Simos, G. (2001). Pus1p-dependent tRNA pseudouridinylation becomes essential when tRNA biogenesis is compromised in yeast. J. Biol. Chem. *276*, 46333–46339.

Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R., and Hofacker, I.L. (2008). The Vienna RNA websuite. Nucleic Acids Res. *36* (Web Server issue), W70–W74.

Gu, B.W., Zhao, C., Fan, J.M., Dai, Q., Bessler, M., and Mason, P.J. (2009). Anomalous electrophoretic migration of newly synthesized ribosomal RNAs and their precursors from cells with DKC1 mutations. FEBS Lett. *583*, 3086–3090.

Heiss, N.S., Knight, S.W., Vulliamy, T.J., Klauck, S.M., Wiemann, S., Mason, P.J., Poustka, A., and Dokal, I. (1998). X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. Nat. Genet. *19*, 32–38.

Ho, N.W., and Gilham, P.T. (1971). Reaction of pseudouridine and inosine with N-cyclohexyl-N'-beta-(4-methylmorpholinium)ethylcarbodiimide. Biochemistry *10*, 3651–3657.

Hoareau-Aveilla, C., Henry, Y., and Leblanc, T. (2008). [Dyskeratosis congenita, a disease caused by defective telomere maintenance]. Med. Sci. (Paris) *24*, 390–398.

Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T.K., Bateman, A., Bernard, T., Binns, D., Bork, P., Burge, S., et al. (2012). InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res. *40* (Database issue), D306–D312.

Jiang, W., Middleton, K., Yoon, H.J., Fouquet, C., and Carbon, J. (1993). An essential yeast protein, CBF5p, binds in vitro to centromeres and microtubules. Mol. Cell. Biol. *13*, 4884–4893.

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

Kierzek, E., Malgowska, M., Lisowiec, J., Turner, D.H., Gdaniec, Z., and Kierzek, R. (2014). The contribution of pseudouridine to stabilities and structure of RNAs. Nucleic Acids Res. *42*, 3492–3501.

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

Lafontaine, D.L., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M., and Tollervey, D. (1998). The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. Genes Dev. *12*, 527–537.

Lane, B.G., Ofengand, J., and Gray, M.W. (1992). Pseudouridine in the large-subunit (23 S-like) ribosomal RNA. The site of peptidyl transfer in the ribosome? FEBS Lett. *302*, 1–4.

Machnicka, M.A., Milanowska, K., Osman Oglou, O., Purta, E., Kurkowska, M., Olchowik, A., Januszewski, W., Kalinowski, S., Dunin-Horkawicz, S., Rother, K.M., et al. (2013). MODOMICS: a database of RNA modification pathways—2013 update. Nucleic Acids Res. *41* (Database issue), D262–D267.

Massenet, S., Motorin, Y., Lafontaine, D.L., Hurt, E.C., Grosjean, H., and Branlant, C. (1999). Pseudouridine mapping in the Saccharomyces cerevisiae spliceosomal U small nuclear RNAs (snRNAs) reveals that pseudouridine synthase pus1p exhibits a dual substrate specificity for U2 snRNA and tRNA. Mol. Cell. Biol. *19*, 2142–2154.

Metz, D.H., and Brown, G.L. (1969a). The investigation of nucleic acid secondary structure by means of chemical modification with a carbodiimide reagent. I. The reaction between N-cyclohexyl-N'-beta-(4-methylmorpholinium)ethylcarbodiimide and model nucleotides. Biochemistry *8*, 2312–2328.

Metz, D.H., and Brown, G.L. (1969b). The investigation of nucleic acid secondary structure by means of chemical modification with a carbodiimide reagent. II. The reaction between N-cyclohexyl-N'-beta-(4-methylmorpholinium)ethylcarbodiimide and transfer ribonucleic acid. Biochemistry *8*, 2329–2342.

Miller, C., Schwalb, B., Maier, K., Schulz, D., Dümcke, S., Zacher, B., Mayer, A., Sydow, J., Marcinowski, L., Dölken, L., et al. (2011). Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. Mol. Syst. Biol. *7*, 458.

Mochizuki, Y., He, J., Kulkarni, S., Bessler, M., and Mason, P.J. (2004). Mouse dyskerin mutations affect accumulation of telomerase RNA and small nucleolar RNA, telomerase activity, and ribosomal RNA processing. Proc. Natl. Acad. Sci. USA *101*, 10756–10761.

Nagaraj, N., Kulak, N.A., Cox, J., Neuhauser, N., Mayr, K., Hoerning, O., Vorm, O., and Mann, M. (2012). System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. Mol. Cell. Proteomics *11*, M111.013722.

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

Ofengand, J. (2002). Ribosomal RNA pseudouridines and pseudouridine synthases. FEBS Lett. *514*, 17–25.

Piekna-Przybylska, D., Decatur, W.A., and Fournier, M.J. (2007). New bioinformatic tools for analysis of nucleotide modifications in eukaryotic rRNA. RNA *13*, 305–312.

Reilly, T.H., and Schmitt, M.E. (1995-1996). The yeast, Saccharomyces cerevisiae, RNase P/MRP ribonucleoprotein endoribonuclease family. Mol. Biol. Rep. *22*, 87–93.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. *16*, 276–277.

Sinha, H., David, L., Pascon, R.C., Clauder-Münster, S., Krishnakumar, S., Nguyen, M., Shi, G., Dean, J., Davis, R.W., Oefner, P.J., et al. (2008). Sequential elimination of major-effect contributors identifies additional quantitative trait loci conditioning high-temperature growth in yeast. Genetics *180*, 1661–1670.

Toh, S.M., and Mankin, A.S. (2008). An indigenous posttranscriptional modification in the ribosomal peptidyl transferase center confers resistance to an array of protein synthesis inhibitors. J. Mol. Biol. *380*, 593–597.

Trahan, C., and Dragon, F. (2009). Dyskeratosis congenita mutations in the H/ACA domain of human telomerase RNA affect its assembly into a pre-RNP. RNA. *15*, 235–243.

Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., et al. (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. Nature *505*, 117–120.

Watkins, N.J., Gottschalk, A., Neubauer, G., Kastner, B., Fabrizio, P., Mann, M., and Luhrmann, R. (1998). Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. RNA *4*, 1549–1568.

Wong, J.M., and Collins, K. (2006). Telomerase RNA level limits telomere maintenance in X-linked dyskeratosis congenita. Genes Dev. *20*, 2848–2858.

Wu, G., Xiao, M., Yang, C., and Yu, Y.T. (2011). U2 snRNA is inducibly pseudouridylated at novel sites by Pus7p and snR81 RNP. EMBO J. *30*, 79–89.

Yu, A.T., Ge, J., and Yu, Y.T. (2011). Pseudouridines in spliceosomal snRNAs. Protein Cell *2*, 712–725.

Zebarjadian, Y., King, T., Fournier, M.J., Clarke, L., and Carbon, J. (1999). Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA. Mol. Cell. Biol. *19*, 7461–7472.

Zhang, Q., Kim, N.K., and Feigon, J. (2011). Architecture of human telomerase RNA. Proc. Natl. Acad. Sci. USA *108*, 20325–20332.

Zucchini, C., Strippoli, P., Biolchi, A., Solmi, R., Lenzi, L., D'Addabbo, P., Carinci, P., and Valvassori, L. (2003). The human TruB family of pseudouridine synthase genes, including the Dyskeratosis Congenita 1 gene and the novel member TRUB1. Int. J. Mol. Med. *11*, 697–704.