EXTENDED EXPERIMENTAL PROCEDURES

Yeast Growth Conditions

Unless otherwise noted, yeast were grown in YPD supplemented with 0.27 mM uridine and collected by centrifuge at 4°C and flash frozen. For heat stress conditions BY4741 and *pus7* were grown to OD 2.0 600 nm at 30°C and 15 ml cultures was transferred to 45°C or remained at 30°C for 60 min. Cells were harvested by centrifugation at room temperature and flash frozen on dry ice. For cold shock BY4741 was grown to OD 2.0 600nm at 30°C, 15 ml culture was transferred to 4°C for 2 hr, cells were pelleted at 4°C and flash frozen. Controls for these conditions were left at RT for the appropriate length of time. Saturated BY4741 were grown overnight at 30°C. Log BY4741 were harvested from 30°C at OD 2.0. Conditional *CBF5* and control strain were incubated in YPD +.27 mM uridine with 20 μg/ml Dox overnight.

Yeast Strains

snR34, snR189, snR35, snR49, snR3, and snR32 were deleted in BY4741 using NatR. Deletion strains of PUS1, PUS2, PUS4, PUS5, PUS6, PUS7, PUS9, DEG1 in BY4741 were obtained from the Yeast Deletion Collection (Invitrogen). CBF5 conditional knockdown was obtained from the Tet-Promoter Hughes Collection (GE HealthCare). Strains encoding a GFP tagged Pus1, Pus7, Pus4 and Deg1 were obtained from the Yeast GFP Clone Collection (Life Technologies).

Ψ-Seq

Polyadenylated RNA (polyA+ RNA) was enriched from total RNA using one round of enrichment using Oligo(dT) dynabeads (Invitrogen) according to the manufacturer's protocol. CMC-treatment was performed essentially as described in (Bakin and Ofengand, 1993, 1998). Specifically, pelleted polyA+ RNA was resuspended in 30 μl of 0.17 M CMC in BEU buffer (50 mM bicine, pH 8.3, 4 mM EDTA, and 7 M urea) or in 30 μl of BEU buffer (for input samples) at 37°C for 20 min (*i.e.*, input controls were handled identically but without CMC treatment). Reaction was stopped with 100 μl of 0.3 M NaOAc and 0.1 mM EDTA, pH 5.6 (Buffer A), 700 μl ethanol, and 1 μl glycogen. After chilling for 5 min in dry ice, the pellet was recovered, washed with 70% ethanol, dissolved in 100 μl Buffer A, and reprecipitated with 300 μl ethanol and 1 μl glycogen. After washing as above, the pellet was dried, dissolved in 40 μl of 50 mM sodium bicarbonate, pH 10.4, and incubated at 37°C for 3 hr. RNA was precipitated by addition of 100 μl Buffer A, 700 μl ethanol, and 1 μl glycogen. The pellet was washed with 70% ethanol and dissolved with water for library preparation.

Library Preparation

mRNA was chemically fragmented into ~80-150nt fragments using RNA fragmentation reagent and stop solution (Ambion) for 30 s. RNA was subjected to FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific), followed by 3′ ligation of an RNA adaptor using T4 ligase (New England Biolabs). Ligated RNA was reverse transcribed using AffinityScript Multiple Temperature Reverse Transcriptase (Agilent) or SuperScript III (Life Technologies), and cDNA was subjected to 3′ ligation with a second adaptor using T4 ligase. The single-stranded cDNA product was amplified for 9-12 cycles in a PCR reaction. Libraries were sequenced on Illumina HiSeq 2500 with paired end reads (30 bp each).

Read Mapping

Reads were mapped to the yeast genome (sacCer3) using Bowtie (version 0.12.7), with the following parameters: '-k 8–best–strata–sam–maxins 50000'. Reads were mapped to the human (hg19) genome using Tophat (version 1.4.1), with the parameters '–max-multihits 1 –prefilter-multihits' and '–transcriptome-index', assigned to a pre-indexed version of the relevant transcriptomes, based on the UCSC Known Genes annotation (for human) or the Saccharomyces Genome Database (SGD) annotation (for yeast). A script was used to project all reads aligning to the genome to yeast and human transcriptomes. Only read pairs fully matching a transcript structure, as defined by the transcriptomic annotations, were retained. Each read pair was computationally extended in transcriptome space from the beginning of the first read to the end of its mate, whereupon, for each nucleotide, we recorded the number of 'left' reads initiating at each position (corresponding to the last position traversed by reverse transcriptase) and the number of overall reads covering each position.

To examine rRNA and snRNA sites, we used Bowtie to directly align reads against a transcriptomic database of these two classes. These alignments were processed as above to record the number of reads starting and overlapping each position. The mouse sample was directly aligned to mouse rRNA sequences from Refseq. The *C. albicans* sample was aligned against the *C. albicans* genome (SC5314 version A21).

Detection of Putative Ψ Sites

Putative Ψ sites were identified with the following approach: (1) For each treated or non-treated sample, a Ψ-ratio was calculated, corresponding to the number of reads beginning at the position divided by the overall number of reads covering it (Figure 1B). A pseudocount of 1 was added to both the numerator and denominator to stabilize the ratio and avoid division by 0. (2) The Ψ-fold change was calculated as $\log_2$ fold changes of Ψ ratios in the treated versus non-treated samples (Figure 1B). All positions with a Ψ-ratio > 0.1, a Ψ-fc > 3 (8 fold enrichment) and with > 5 reads beginning at the position were considered putative Ψ sites. (3) To compare across conditions, we first merged the positions of all windows passing step (2) from all samples, to define a set of all unique

sites passing the filters in at least one condition. For each such site, we then calculated a Ψ-ratio and Ψ-fc in each sample. We calculated a median Ψ-ratio and median Ψ-fc across all samples, and considered all sites in which these values exceeded 0.1 and 3, respectively (unless explicitly stated). (4) As highly similar paralogs may be present in the genome (e.g., for rRNA and tRNA), we demanded that each sequence of 21 nucleotides surrounding the putative pseudouridylated position in the data set be unique; All redundancies were filtered. We also filtered sites in positions other than immediately following 'U'.

Generation of Synthetic Pseudouridylated Spike-Ins

An RNA bait containing only a single uridine was in vitro transcribed from dsDNA templates. Sequence of the templates (excluding T7 promoter):

GGGAGGCGAGAACACACCACAACGAAAACGAGCAAAACCCGGTACGCAACACAAAAGCGAACAACGCGAAAAAGGACACCGA AGCGGAAGCAAAGACAACCAACAGAAAACAACCGCAAACAAACGGGACCAGACAACGCACCAGCAAAA

In vitro transcription was performed in a volume of 20 μl, using a nucleotide mix containing 150nmol GTP, CTP, ATP, 75nmol Biotin-16 CTP and either 150nmol UTP or Ψ-TP (TriLink BioTechnologies). Pseudouridylated and non-pseudouridylated RNA were mixed at the ratios indicated in Figures 1C and 1D.

Identification of PUS-Dependent Sites

To identify differentially pseudouridylated sites across deletion strains of non-essential PUSs and snoRNAs (Figure 2B), we implemented an approach where for each PUS deletion all remaining strains served as background. We performed two statistical tests based on the number of reads *starting* and *overlapping* at each putatively identified position across all the samples. First, we used a chi-square test to identify whether overall there was a significant change across all samples ("differential P value"). Second, we identified the sample with the lowest Ψ-ratio score ("minimal sample"), and calculated a chi-square P value comparing number of reads starting and overlapping this position with the aggregate number of staring and overlapping reads across all samples ("outlier P value"). A Z score was calculated for the Ψ-ratio in the minimal sample, compared to all other samples ("outlier Z score"). A site was considered as dependent on a PUS if it met the following criteria: both P values were < 0.05, the outlier Z score was < -1.5, and the Ψ-ratio in the minimal sample was < 0.07.

To identify heat-shock dependent sites that are Pus7-dependent (3 replicates / condition), we demanded that (1) the outlier P value be < 0.05, (2) the difference in the mean Ψ-ratio between the WT and the Cbf5 deletion samples be > 0.05, (3) the mean Ψ-ratio in the heat shock samples be > 0.05 and (4) the P value of a Student's t test comparing the WT to the knockdown Ψ-ratios be < 0.1. To identify sites dependent on Cbf5 (2 replicates / condition), we demanded that (1) the outlier P value be < 0.05, (2) the difference in the mean Ψ-ratio between the WT and the Cbf5 deletion samples be > 0.05, (3) the mean Ψ-ratio in the WT samples be > 0.05, (4) the mean Ψ-fc across WT samples be > 2, and (5) the P value of a Student's t test comparing the WT to the knockdown Ψ-ratios be < 0.1. To identify sites dependent on human DKC1, we used identical criteria, and demanded that the mean Ψ-fc across siControl samples be > 2.5.

Motif Analysis

To identify putative motifs we used position specific scoring matrices (PSSMs) for a 21 bp sequence surrounding the identified pseudouridylated sites. PSSMs were visualized using the SeqLogo package in R, and were cropped to display only informative regions (harboring non-uniform sequence composition).

Prediction of Association between Cbf5-Dependent Target Sites and snoRNAs

To assess the hybridization potential between Cbf5-dependent target sites and H/ACA box snoRNAs, we assembled a database of 10 bp of each upstream complementary arm concatenated to the 10 bp downstream complementary arm, separated by a constantly defined hairpin sequence (AGGGCCTTTTGGCCCT). The upstream and downstream targeting arms in yeast were defined based on the Yeast snoRNA Database (Piekna-Przybylska et al., 2007); in human, they were defined based on the snoRNABase database (Lestrade and Weber, 2006). Free energies of these concatenated sequences folded against a 20-nt region surrounding each CBF5-dependent site were calculated using RNAcofold from the Vienna Package suite (Gruber et al., 2008), using the parameters 'RNAcofold -a -d2'. In addition, the folding was given a dual constraint (using the –c parameter) of (1) forcing the hairpin structure, and (2) forcing the putative CBF5-dependent pseudouridylation position to be unpaired. A matrix of free energies across H/ACA boxes paired against all target sites was derived. The matrix was first median-normalized per each H/ACA box (columns). Subsequently, a Z-score was derived across all rows, and the minimal Z score was extracted. As controls, we shuffled each 20-nt region 15 times (maintaining a U in the center), and subjected them to the same cofolding approach against H/ACA box snoRNAs.

Localization of GFP-Tagged Pus7 Proteins after Heat Shock

S. cerevisiae encoding GFP-tagged Pus7 were grown in SC media to OD 2.0 at 30C. Cultures were heat shocked by transfer to 45°C water bath for the time labeled. At the last 5 min of heat shock 100ng/ml DAPI was added to the culture. 500 μl of culture was spun down at 5000 RPM in a desktop centrifuge at room temperature and suspended in 25 μl SC media. Images were collected using a Nikon Eclipse Ti with oil immersion at 100X magnification. Images were captured and visualized with an Andor Camera and

NIS-Elements Imaging Software version 4. ImageJ was used to quantitate GFP intensity from the nucleus and cytoplasm of 100 cells from the 0 and 1 hr time points. DAPI staining was used to define nuclear boundaries.

Bone-Marrow-Derived Dendritic Cells

Bone marrow derived dendritic cells from 6-8 week old female C57BL/6J mice were prepared as previously described (Amit et al., 2009) and profiled at day 9 without any further stimulation. 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 the Massachusetts Institute of Technology (Protocol number: 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 two siRNA boosts at 48 and 96 hr following transfection; As negative controls, we used Ambion In Vivo Negative Control #1 siRNA (catalog number: 4457287). Cells were harvested at 144 hr.

Cross-Species Comparison of rRNA Pseudouridylation

To examine orthologous sites in rRNA positions across different species, we used "needle" (Rice et al., 2000) to generate pairwise alignments of rRNAs and used the alignments to lift over known Ψ sites from human to mouse, and from *S. cerevisiae* to *C. albicans*. Known sites were obtained from the MODOMICS database (Machnicka et al., 2013). We filtered sites that did not contain a "U" in the target position in the projected species.

Samples from Dyskeratosis Congenita Patients

Fibroblasts from two X-linked dyskeratosis congenita patients were obtained from Coriell [(GM01774 (del37L, 7 years), AG04646 (A386T, 11 years), GM00409 - (healthy 7 year old), GM01864 - (healthy 11 year old)].

RAP-RNA Enrichment of TERC

We performed RNA Antisense Purification (RAP) on intact isolated total RNA using the hybridization, capture, and sequencing protocols previously described for RAP-RNA[AMT] (Engreitz et al., 2014) with the following 5′ biotinylated and 3′ blocked DNA oligomers:
gcatgtgtgagccgagtcctgggtgcacgtcccacagctcagggaatcgcgccgcgcgcg,
Cgcggtggcagtgggtgcctccggagaagccccgggccgaccgcggcctccaggcggggt,
Gacatttttgtttgctctagaatgaacggtggaaggcggcaggccgaggcttttccgcc,
Ctcagttagggttagacaaaaaatggccaccacccctcccaggcccaccctccgcaaccc.

SUPPLEMENTAL REFERENCES

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

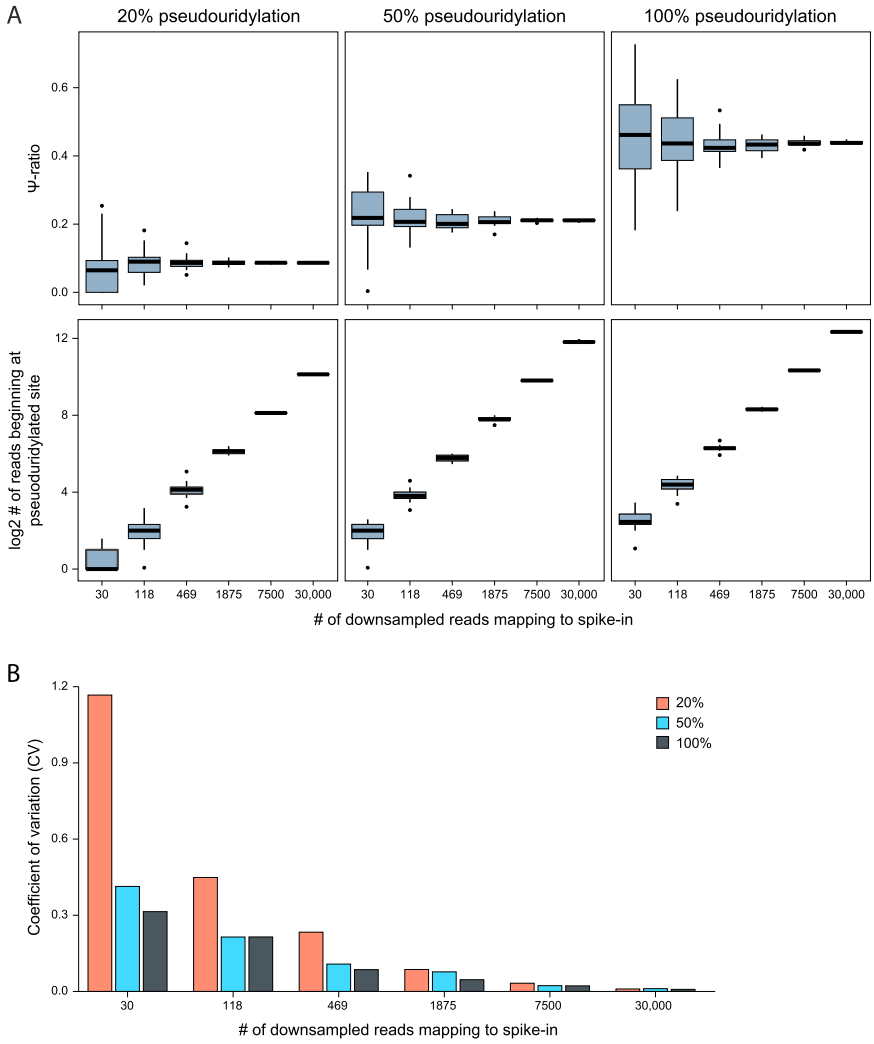

Figure S1. Analysis of Power to Detect Ψ Sites Based on Spike-In, Related to Figure 1

(A) Box plots show the distribution of the # of reads beginning at the pseudouridylated site in the spike in (y axis, bottom plots) or estimated Ψ ratio (y axis, top plots) when data are randomly downsampled such that decreasing number of reads map to the spike-ins (x axis). Data are shown for spike in experiments at (left to right) 20%, 50% and 100% pseudouridylation stoichiometry. The boxes extend from the 25% percentile to the 75% percentile, with the median indicated by a thick line. The whiskers extend 1.5 fold times the interquartile range or to the most extreme value (the smaller of the two), and outliers are indicated as dots.

(B) Coefficient of Variation (CV) of the estimated Ψ ratio at the pseudouridylated site in the spike in (y axis), when data are randomly downsampled such that decreasing number of reads map to the spike-ins (x axis). Data are shown for spike in experiments at 20% (red), 50% (light blue) and 100% (dark blue) pseudouridylation stoichiometry.

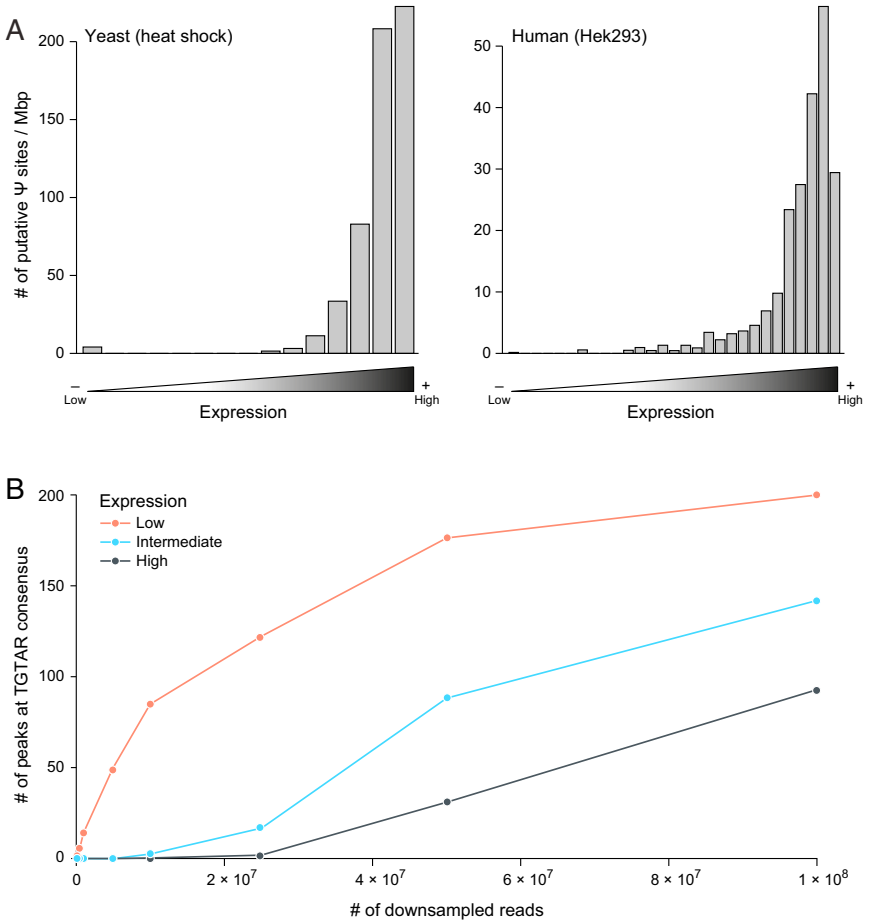

Figure S2. Ψ Site Detection as a Function of Gene Expression, Related to Figure 2

(A) Density of Ψ sites per million bases of annotated transcript are plotted as a function of gene expression levels across 15 equally sized bins for yeast (left, in heat shock, 15 expression bins) and human (right, 29 expression bins) data. Expression data are based on RNA-seq in a matching condition (3 heat shock replicates in yeast, two replicate experiments transfected with siControl in human).

(B) Analysis of power to detect Ψ sites based on downsampling. Show are the number of peaks detected at consensus UGUAR sites (y axis) when reads are downsampled (x axis) for genes at three equally sized expression bins: low (red), intermediate (light blue), and high (dark blue).

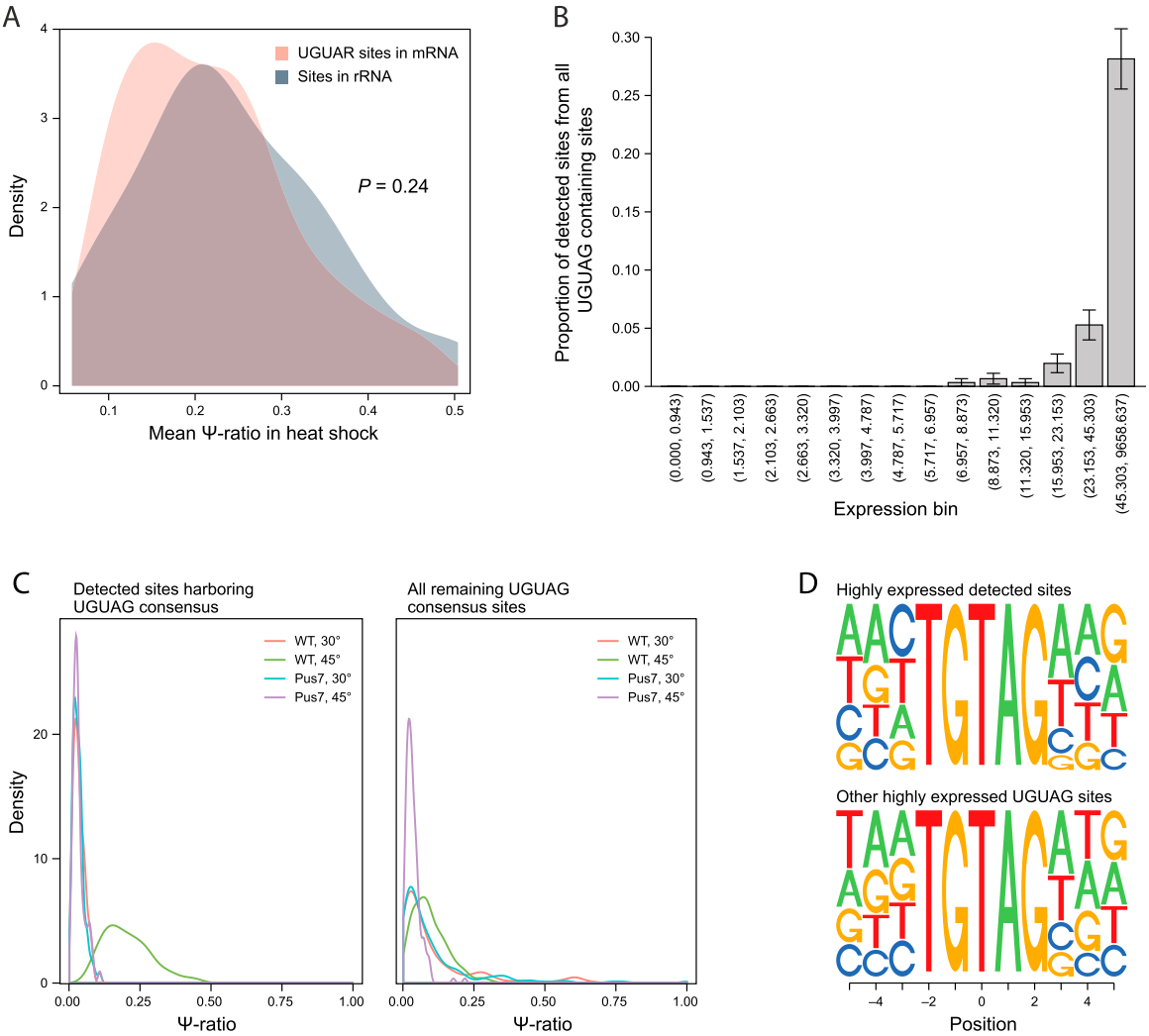

Figure S3. Features of Pus7- and Heat-Shock-Dependent Ψ Sites, Related to Figure 4

(A) Ψ-ratios are comparable between UGUAR consensus sites in mRNA and known Ψ sites in rRNA. Shown are the distributions of Ψ-ratios in heat shock (mean of 3 replicates) in UGUAR consensus sites in mRNA (pink) and known sites in rRNA (gray). P-value of a Wilcoxon test is noted.

(B) Proportion of UGUAG sites called as modified in heat shock (y axis) of all the sites with the consensus among transcripts in each of 15 equally sized expression bin (x axis). Overall 111 of 4,544 sites are detected as modified.

(C) Distributions of Ψ-ratios in UGUAG sites with detected modification (left) and without a detected modification (right), in WT and Δpus7 mutant strains in either 30°C or 45°C (color legend).

(D) Sequence logo of the motif spanning UGUAG sites in highly expressed genes that are either detected as modified (top) or not (bottom).

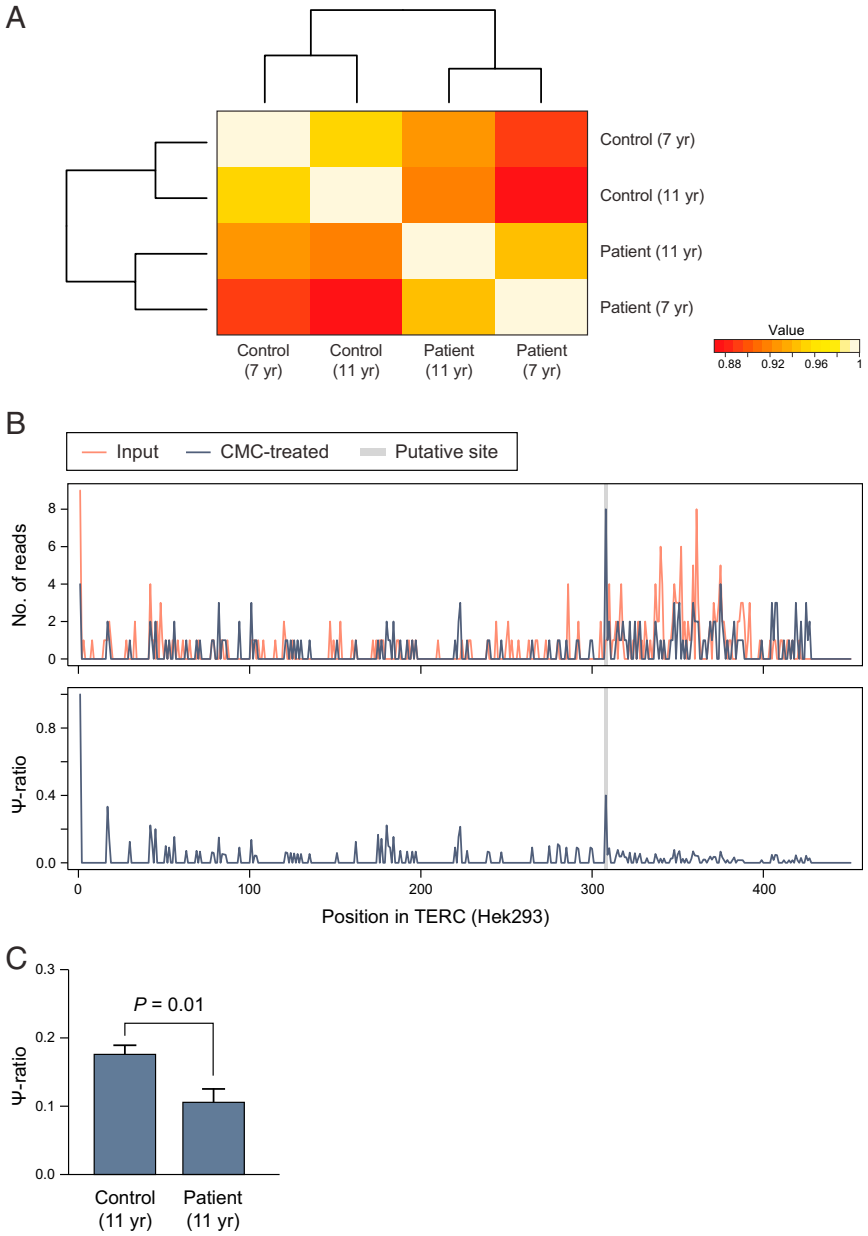

Figure S4. Changes in Pseudouridinylation, Related to Figure 5

(A) Similarity of Ψ profiles from cells of dyskeratosis congenita patients. Heat map shows the Pearson correlation coefficients (white – high; red – low; color bar) between Ψ-ratio profiles of 2 patients (7 and 11 yr) and age-matched controls. Heat map is clustered by hierarchical agglomerative clustering with average linkage. The dendrogram indicates that samples group by disease state rather than age.

(B) Shown are number of reads beginning at each position along TERC for CMC-treated and non-treated samples (top) and Ψ-ratio for the treated sample (bottom) based on aggregated data from two replicates in Hek293 cells (prior to RAP enrichment).

(C) Ψ-ratios for position 307 in 11-yr old patient and age matched control; Error bars: standard error.