Supplementary Tables

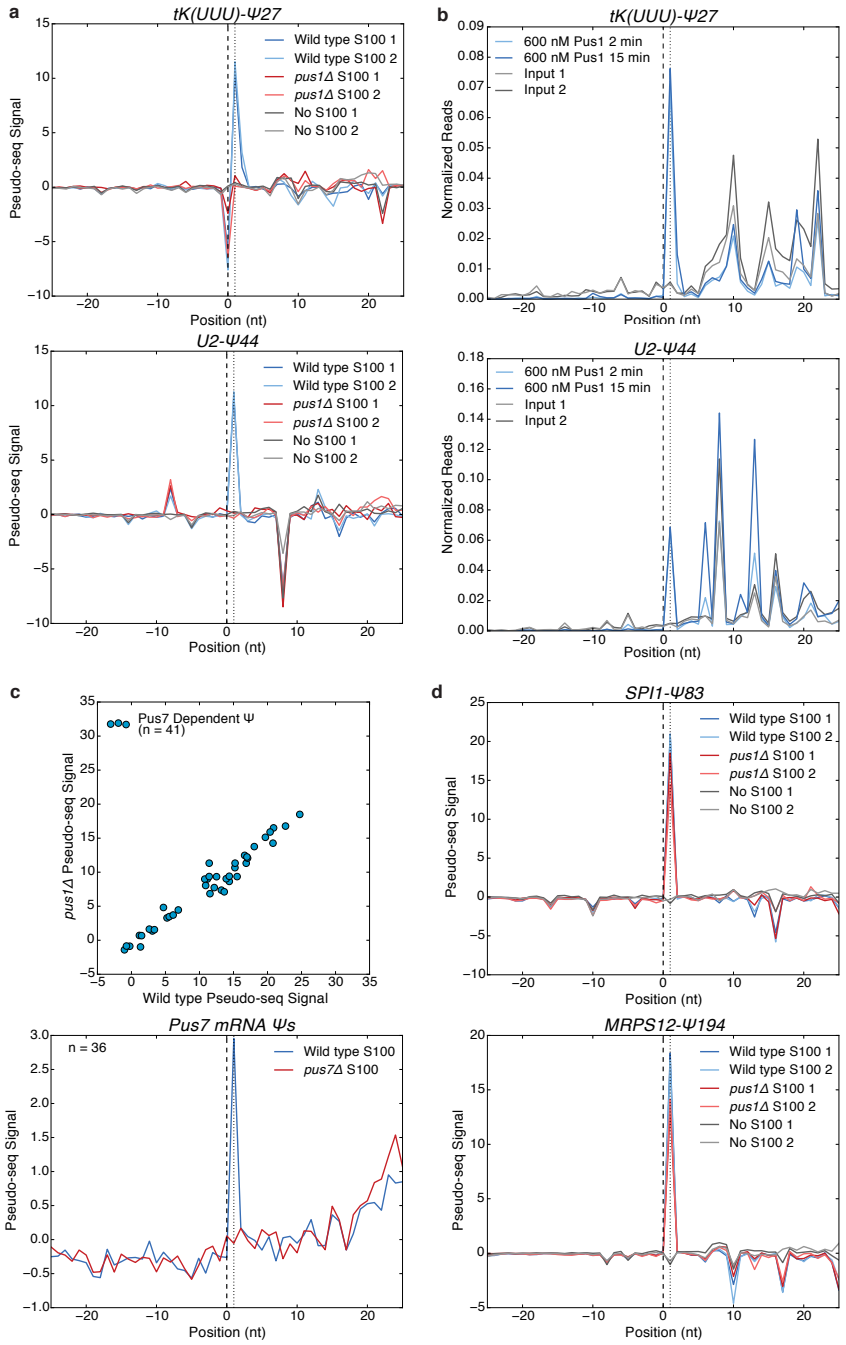
Supplementary Table 1: Pool Sequences

Pool Sequence	Sequence
T7 Promoter	GCTAATACGACTCACTATAGGG
Pool 1 3′ Adapter	GCTGCAAGGTTACGTCTC
Pool 2 3′ Adapter	CACTCGGGCACCAAGGAC

Supplementary Table 2: Primer Sequences

Primer Name	Primer Sequence	Notes
oBZ131	GCTAATACGACTCACTATAGGG	Pool 1,2 PCR F
oTC_pool1_rev	GAGACGTAACCTTGCAGC	Pool 2 PCR R
oTC_pool2_rev	GTCCTTGGTGCCCGAGTG	Pool 1 PCR R
oTC_RT-L2_3′10N	/5Phos/GATCGTCGGACTGTAGAACTCTGAACGTGTAGATC/iSp18/CACTCA/iSp18/CCTTGGCACCCGAGAATTCCANNNNNNNNNNTCCTTGGTGCCCGAGTG	RT Primer, Circ libraries
oTC_RT-L2	/5Phos/GATCGTCGGACTGTAGAACTCTGAACGTGTAGATC/iSp18/CACTCA/iSp18/CCTTGGCACCCGAGAATTCCATCCTTGGTGCCCGAGTG	RT Primer, Circ libraries
ONM_RT-L2	/5Phos/NNNNNNNNNGATCGTCGGACTGTAGAACTCTGAACGTGTAGATC/iSp18/CACTCA/iSp18/CCTTGGCACCCGAGAATTCCAGTCCTTGGTGCCCGAGTG	
RP1	AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA	Library PCR F
BC	CAAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA	Library PCR R, XXXXXX indicates unique barcodes
PFY1-U290_wt	*GCTAATACGACTCACTATAGGG*ACACAAACCGTTATTATTGCTCATTATCCACCAACCGTACAAGCCGGTGAGGCCACCAAGATT*CACTCGGGCACCAAGGAC*	Wild type control for designed Pus1 substrate
PFY1-U290_mutant	*GCTAATACGACTCACTATAGGG*ACACAAACCGTTATTCTGGCTGTGTATCCACAGACAGAACATGCCGGTGAGGCCACCAAGATT*CACTCGGGCACCAAGGAC*	Designed Pus1 substrate

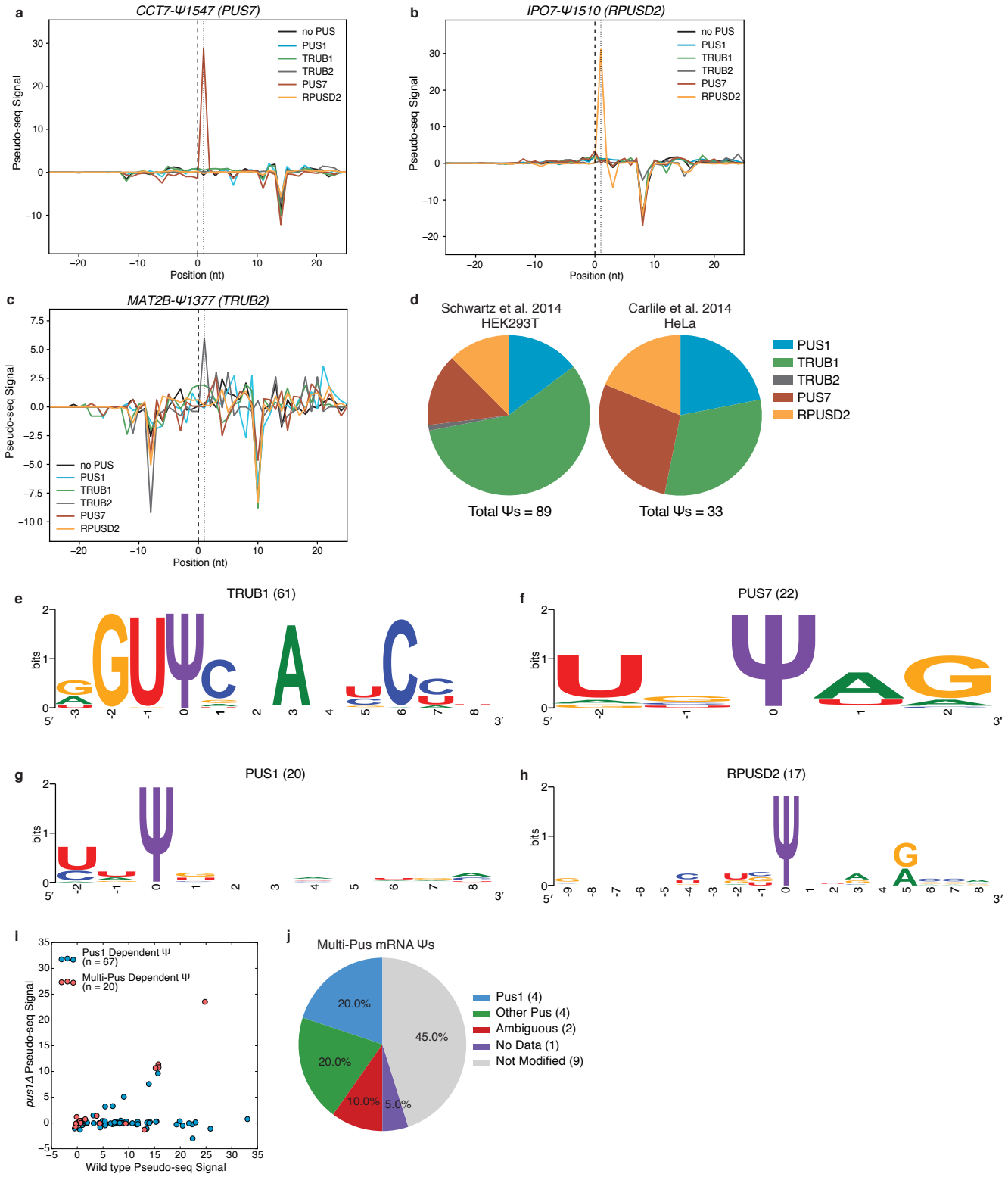
Supplementary Figure 1



Supplementary Figure 1: *In vitro* Pseudouridylation of mRNA targets.

a) Pseudo-seq signal for oligo pool positive control ncRNA Ψs contained in oligo pools incubated with *wild type* S100 (blue), *pus1Δ* S100 (red), or no extract (gray), *tK(UUU)-Ψ27* (upper), *U2-Ψ44* (lower). b) Normalized reads for positive control ncRNA Ψs incubated with rPus1 (blue), or input RNA (gray), *tK(UUU)-Ψ27* (upper), *U2-Ψ44* (lower). c) A scatter plot of Pseudo-seq signal for pool 2 pseudouridylated with *wild type*, or *pus1Δ* extracts (upper, n=41 sequences). A meta plot of Pseudo-seq signal from pool 1 pseudouridylated with *wild type* (blue), or *pus7Δ* (red) S100 extracts (lower, n=36 sequences). Sequences correspond to mRNA Ψs that are genetically dependent on *PUS7*. Values represent an average of n=2 replicates. d) Plots of Pseudo-seq signal for Pus7 mRNA substrates incubated with *wild type* S100 (blue), *pus1Δ* S100 (red), or no extract (gray): *SPI1-Ψ83* (upper), *MRPS12-Ψ194* (lower).

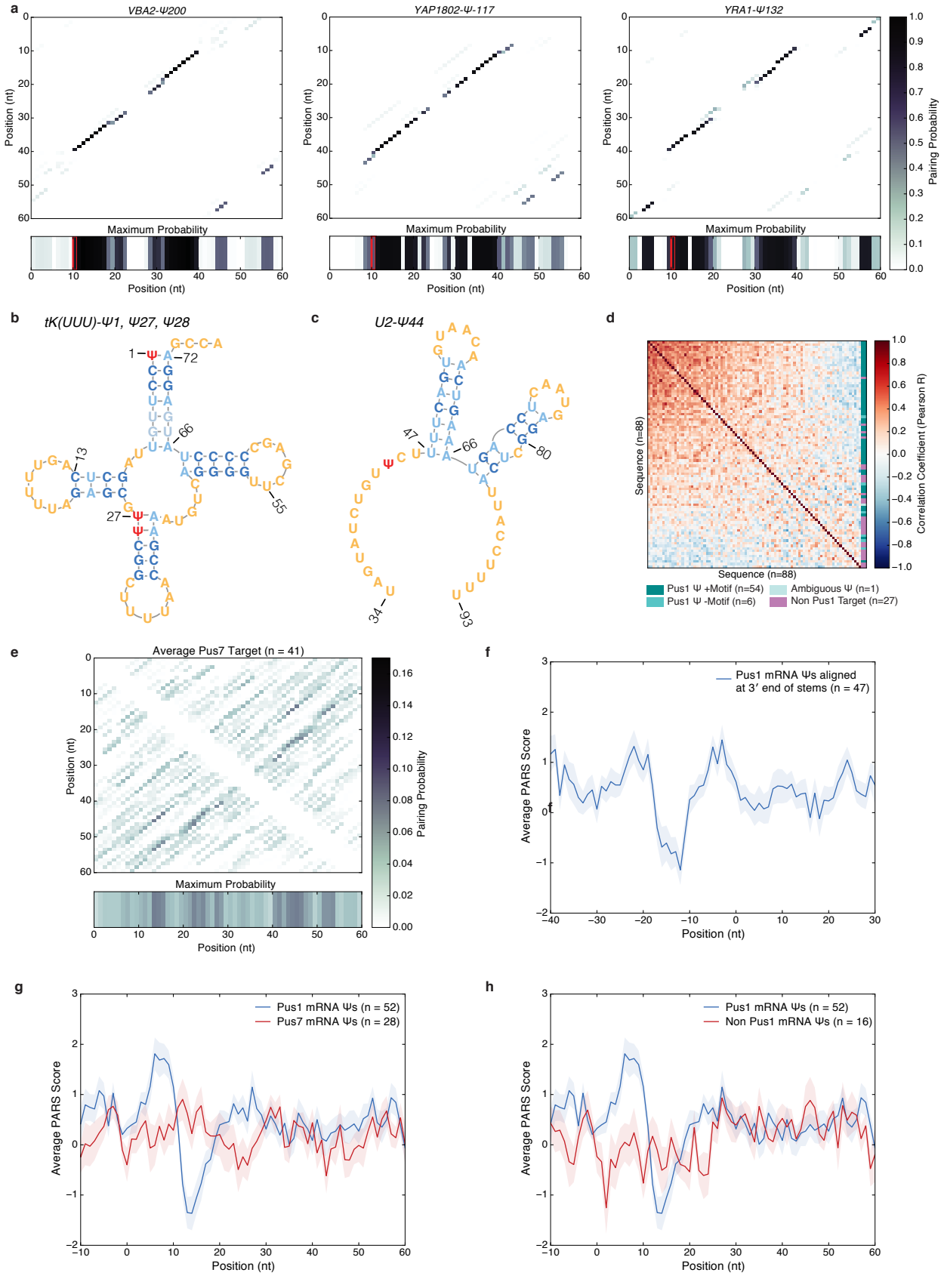
Supplementary Figure 2



Supplementary Figure 2: Identification of Human PUS mRNA Substrates *In vitro*

a-h) *H. sapiens* mRNA sequences were pseudouridylated with recombinant PUS proteins: PUS1 (blue), TRUB1 (green), TRUB2 (gray), PUS7 (red), RPUSD2 (yellow), no PUS (black). a-c) Pseduo-seq signal for a (a) PUS7 mRNA target: *CCT7-Ψ1547*, an (b) RPUSD2 mRNA target: *IPO7-Ψ1510*, and a (c) TRUB2 mRNA target: *MAT2B-Ψ1377*. d) A summary of mRNA Ψs assigned to hPUS proteins in Schwartz et al. 2014 (left), and Carlile et al 2014 (right) data sets. e-h) The sequence motifs surrounding sites modified by (e) TRUB1, (f) PUS7, (g) PUS1, and (h) RPUSD2 generated with WebLogo 3.5. i) A scatter plot of Pseudo-seq signal for yeast pools pseudouridylated with *wild type* or *pus1Δ* S100 extracts. Sequences correspond to mRNA Ψs that are genetically dependent on *PUS1* (blue, n=61 sequences), or are genetically dependent on *PUS1* and another *PUS* (red, n=20 sequences). Values represent an average of n=2 replicates. j) Summary of *in vitro* pseudouridylation of multi-*PUS*-dependent mRNA Ψs.

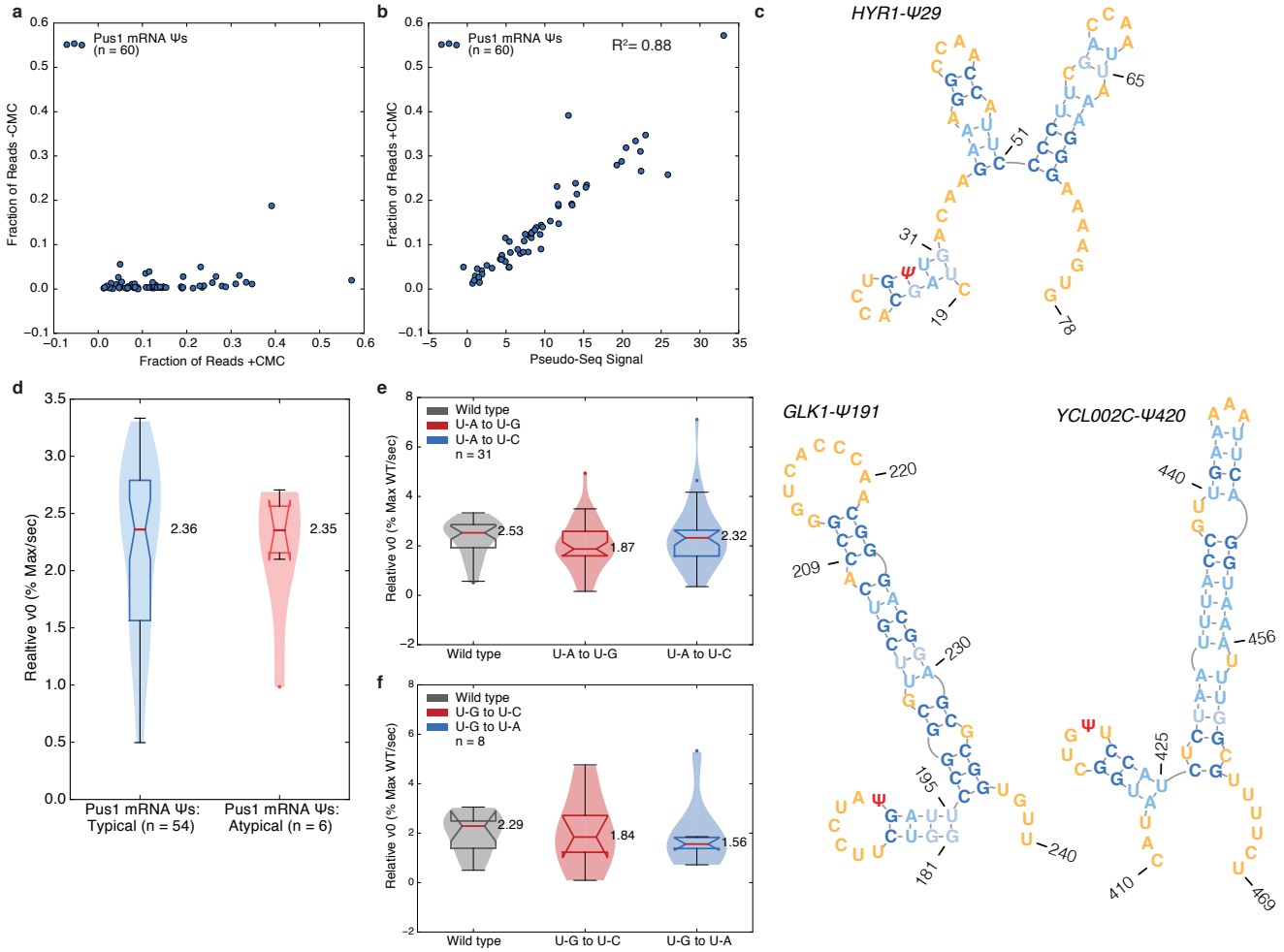
Supplementary Figure 3



Supplementary Figure 3: A structural motif associated with Pus1 mRNA targets.

a) Heatmaps of the RNAfold pairing probability matrices (upper), and maximum pairing probability for each base (lower) for *VBA2-Ψ200* (left), *YAP1802-Ψ-117* (middle), *YRA1-Ψ132* (right). Pseudouridylated positions are indicated by red boxes. b,c) Structures of (b) tK(UUU) and stems (c) IIa and IIb of the U2 snRNA showing positions of Pus1-dependent (b) Ψ1, Ψ27 and Ψ28, and (c) Ψ44. d) A heatmap of pairwise correlation coefficients (Pearson R) between the arrays of the summed pairing probabilities for each mRNA site (n=88 Ψs with genetic evidence for Pus1-dependence *in vivo*). *In vitro* modified mRNA Ψs with a stem-loop motif (dark teal, n=54 sequences), *in vitro* modified without a stem-loop motif (medium teal, n=6 sequences), with ambiguous *in vitro* data (light teal, n=1 sequence), and not modified *in vitro* (purple, n=27 sequences). Rows and columns are ordered by the sum of R values across the row/column. Indicated on the right is the classification of each sequence. e) A heatmap of the average pairing probability matrix from RNAfold for n=41 Pus7 mRNA Ψs. f-h) Average PARS score ±SEM for which there is PARS data. f) High confidence Pus1 mRNA targets containing a structural motif (blue, n=47 targets) aligned at the last base-paired nucleotide in the stem-loop. g) High confidence Pus1 mRNA targets (blue, n=52 targets), and Pus7 mRNA targets (red, n=28 targets). h) High confidence Pus1 mRNA targets (blue, n=52 targets), or mRNA targets not verified as *PUS1*-dependent (red, n=16 targets).

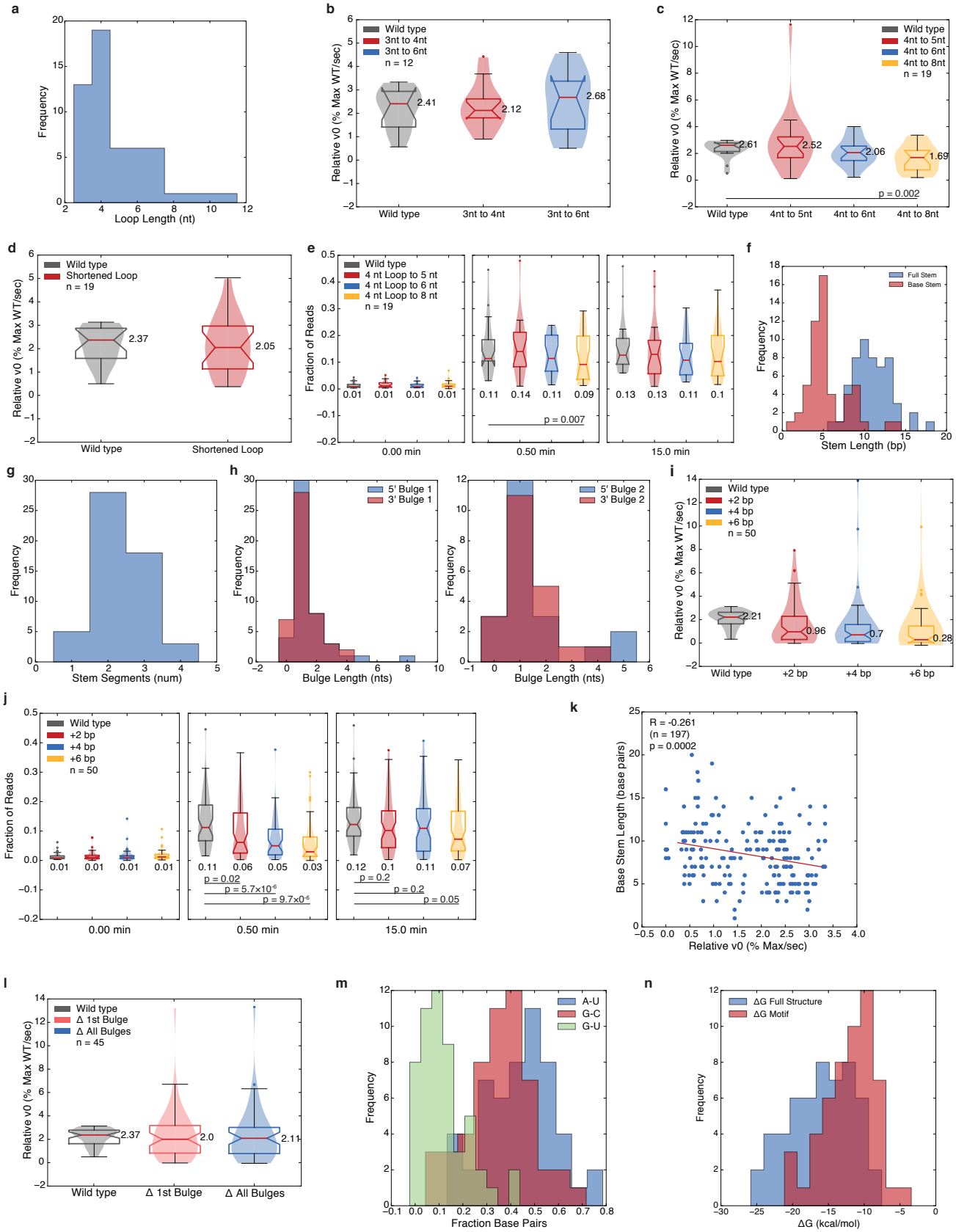
Supplementary Figure 4



Supplementary Figure 4: Kinetic Analysis of Pseudouridylation of Pus1 mRNA Targets

a) A scatter plot comparing the fraction of reads at the expected RT stop position between +CMC and -CMC libraries for n=60 high confidence Pus1 mRNA targets. b) A scatter plot comparing Pseudo-seq signal to the fraction of reads at the expected RT stop position for +CMC libraries n=60 high confidence Pus1 mRNA targets. For a,b, values represent the average for n=2 replicates. c) MFE structures for atypical Pus1 mRNA targets *HYR1-Ψ29* (upper), *GLK1-Ψ191* (lower left) *YCL002C-Ψ420* (lower right). d) Violin plots (center lines, medians; notches, 95% confidence intervals; boxes, 25th to 75th percentiles; whiskers, 1.5X inter-quartile range; dots, values outside of the 1.5X IQR) of v0,rel for typical (blue, n=54 sequences) and atypical (red, n=6 sequences) Pus1 mRNA targets. Medians are indicated. e,f) Violin plots (elements as above) of v0,rel. Medians are indicated. e) Wild type (gray, n=31 sequences), U-A to U-G mutants (red, n=31 sequences), and U-A to U-C mutants (blue, n=31 sequences). f) Wild type (gray, n=8 sequences), U-G to U-C mutants (red, n=8 sequences), and U-G to U-A mutants (blue, n=8 sequences).

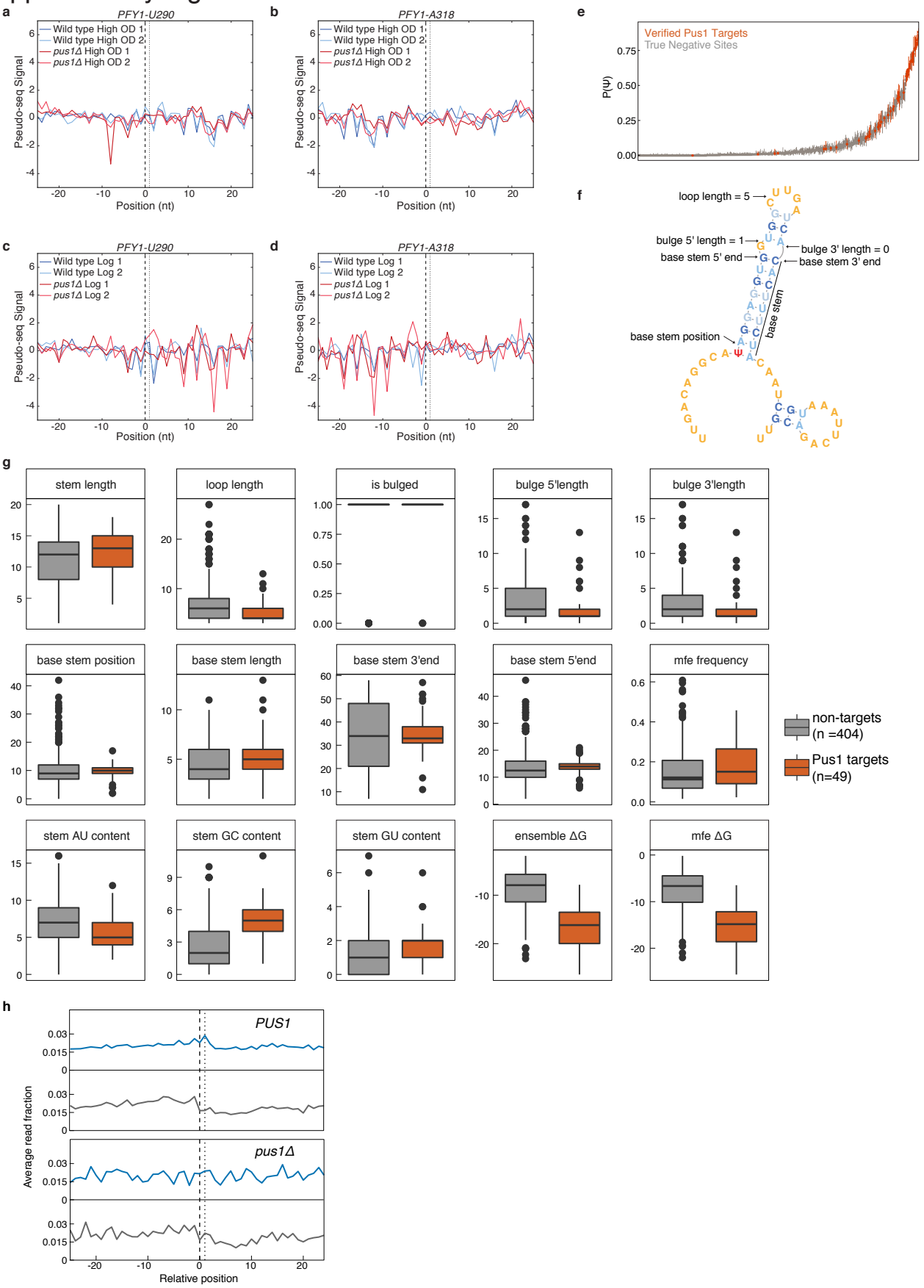
Supplementary Figure 5



Supplementary Figure 5: Sequence and Structure are Pseudouridylation of Targets by Pus1

a) Histogram of Pus1 mRNA substrate loop length. b-e,i,j,l) Violin plots (center lines, medians; notches, 95% confidence intervals; boxes, 25th to 75th percentiles; whiskers, 1.5X inter-quartile range; dots, values outside of the 1.5X IQR) of (b-d, i,l) v0,rel or of (e,j) reads at the expected RT stop positions for wild type and mutant sequences. Medians and p-values (paired t-test, two-tailed) are indicated. b) Wild type (gray, n=12 sequences), 3 to 4 nt (red, n=12 sequences), and 3 to 6 nt (blue, n=12 sequences) loop extensions. c) Wild type (gray, n=19 sequences), 4 to 5nt (red, n=19 sequences), 4 to 6 nt (blue, n=19 sequences), and 4 to 8 nt (yellow, n=19 sequences) loop extensions. d) Wild type (gray, n=19 sequences), and shortened loop mutants (red, n=19 sequences). e) Wild type (n=19 sequences), and mutant (n=19 sequences in each mutant class), labels as in (c). f) Distribution of stem lengths for the full stem (blue) and the stem segment proximal to the first bulge (red). g) Histogram of Pus1 mRNA substrate stem segment number. h) Histograms of Pus1 mRNA substrate bulge dimensions for the first (left), or second (right) bulge (left). 5′ bulge (blue), and 3′ bulge (red) lengths are indicated. i,j) Wild type (gray, n=50 sequences), 2 bp (red, n=50 sequences), 4 bp (blue, n=50 sequences), 6 bp (yellow, n=50 sequences) stem extensions. k) A scatter plot of v0,rel versus base stem length for wild type and stem extension mutants (n=197 sequences). Pearson's R, and p-value (approximation based on Student's t-distribution, two-tailed) is indicated. l) Wild type (gray, n=45 sequences), deletions of the first bulge (red, n=45 sequences), or of all bulged nucleotides (blue, n=45 sequences). m,n) Histograms showing distributions of Pus1 mRNA substrate m) stem base pair compositions (A-U, blue; G-C, red; G-U, green), or n) ΔG values for the full 60 nt (blue) or structural motif only (red).

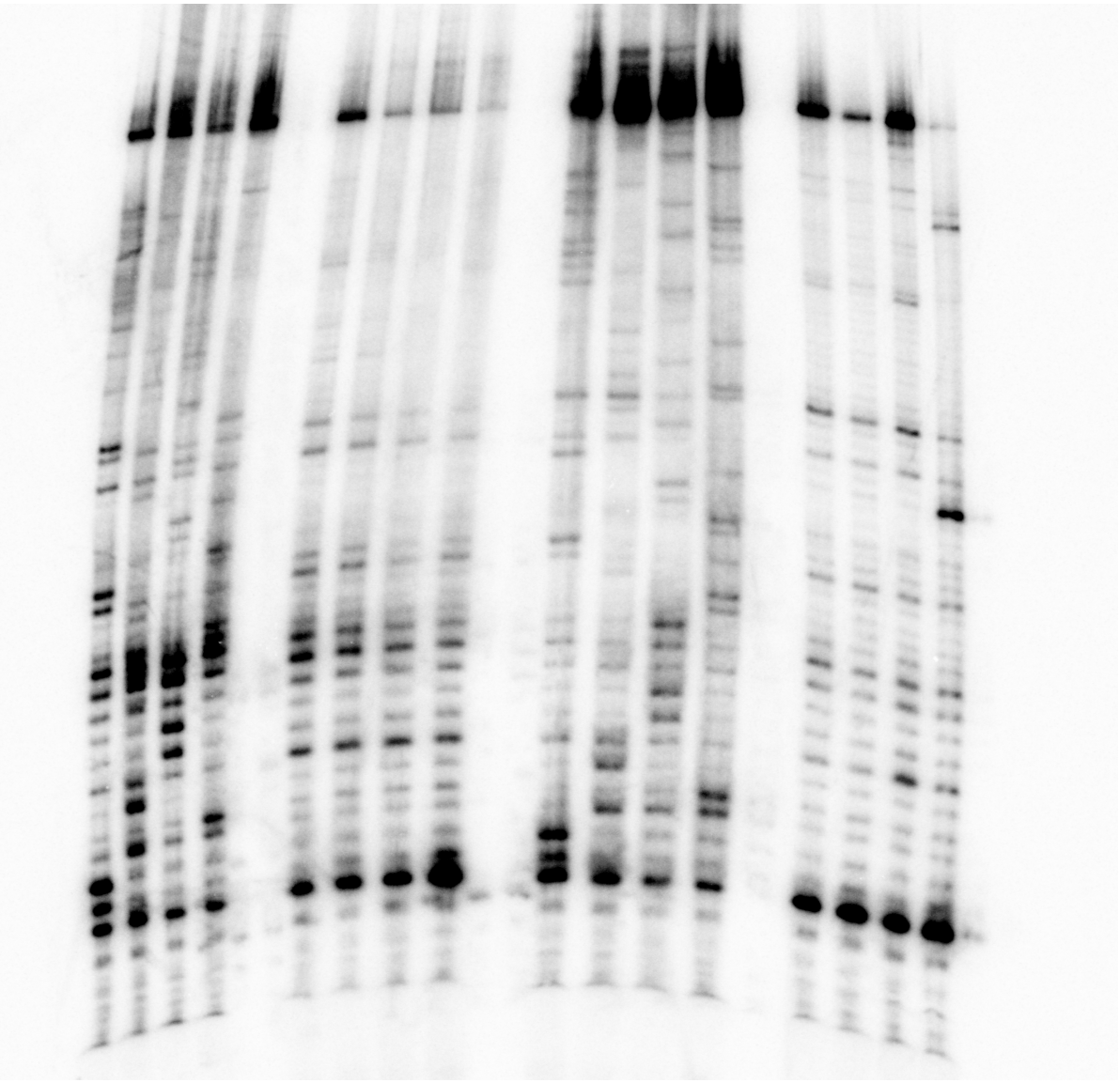
Supplementary Figure 6



Supplementary Figure 6: The Pus1 Structural Motif is Sufficient for Pseudouridylation

a-d) Plots of *in vivo* Pseudo-seq signal for (a,c) *PFY1-U290*, or (b,d) *PFY1-A318* during (a,b) growth to high OD, or (c,d) exponential growth. Wild type (blue), and *pus1Δ* (red) strains are indicated. e) Predicted P(Ψ) for sites in the training set. Sites are sorted from left to right by their median P(Ψ). f) Illustration of structural features used in the classifier, indicated on the structure of a validated Pus1 target. g) Distribution of values (center lines, medians; boxes, 25th to 75th percentiles; whiskers, 1.5X inter-quartile range; dots, values outside of the 1.5X IQR) for each of the classifier features for n=49 positive (Pus1 targets) and n=404 negative (Pus1 non-targets) sites in the training set. h) Metaplot of RT stops for sites predicted to be Pus1 targets with P(Ψ) > 0.8, in +CMC libraries (blue) and -CMC libraries (grey) from log phase *in vivo* Pseudo-seq data. *PUS1* panels show the aggregated reads from knockout libraries for *pus2Δ,3Δ,4Δ,5Δ,6Δ,7Δ, and 9Δ*.

Supplementary Figure 7



Supplementary Figure 7: Uncropped Gel Images Related to Figure 6c,d.

Uncropped primer extension gels of *PFY1-U290* wild type (left lanes, corresponding to Figure 6c), and *PFY1-U290* mutant (right lanes, corresponding to Figure 6d) sequences.

Supplementary Dataset 1: Pseduo-seq Signal for *S.cerevisiae* RNA Pools Pseudouridylated with S100 Extracts

Pseudo-seq signal for Pool2 (a,b) or Pool1 (c) RNAs pseudouridylated with S100 extracts. The sequences correspond to sites genetically dependent on Pus1 (a) or Pus7 (b,c) *in vivo*.

Supplementary Dataset 2: Pseudo-seq Signal for *H. sapiens* RNA Pools Pseudouridylated with Recombinant Human PUS Proteins.

Pseudo-seq signal for hsPool1 and hsPool2 RNAs pseudouridylated with recombinant human PUS proteins, and yeast S100 extract.

Supplementary Dataset 3: Structural Characteristics of *S. cerevisiae* Pus1 Substrates

Characteristics of RNAfold predicted MFE structures of predicted Pus1 substrates validated (upper) or not validated (lower) *in vivo*.

Supplementary Dataset 4: Kinetic Analysis of Pseudouridylation of Wild Type and Mutant Sequences by Recombinant Pus1.

Shown are the fractions of reads mapping to the expected RT stop position for each timepoint, and the calculated relative initial velocity, normalized to the maximum extent of pseudouridylation for wild type sequences. Sequences are wild type and mutant Pus1 substrates.

Supplementary Dataset 5: Kinetic Analysis of Pseudouridylation of Wild Type and Stem Extension Mutant Sequences by Recombinant Pus1.

Shown are the calculated relative initial velocities for wild type and stem extension mutants. Values were calculated by normalizing to the maximum extent of pseudouridylation for each sequence, rather than to the corresponding wild type sequence. Corresponds to Figure 5e.

Supplementary Dataset 6: Random Forest Classifier Predicted Pus1 mRNA Ψs.

Shown are mRNA sites predicted to be pseudouridylated in a *PUS1*-dependent manner by a random forest classifier. Included are the average Pseudo-seq signal from *in vivo* data from exponentially growing and post-diauxic yeast.