Detection of Pseudouridine Synthase Activity with a Fluorescence-Based Yeast Reporter System

ELIJAH GABRIEL QUINN DUECK
Bachelor of Science, **University of Lethbridge, 2017**

A thesis submitted
in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in

BIOCHEMISTRY

Alberta RNA Research and Training Institute
Department of Chemistry and Biochemistry
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

DETECTION OF PSEUDOURIDINE SYNTHASE ACTIVITY WITH A FLUORESCENCE-BASED YEAST REPORTER SYSTEM

ELIJAH GABRIEL QUINN DUECK

Date of Defence: December 10, 2020

Dr. Ute Kothe Thesis Supervisor	Professor	Ph.D.
Dr. Hans-Joachim Wieden Thesis Examination Committee Member	Professor	Ph.D.
Dr. Roy Golsteyn Thesis Examination Committee Member	Associate Professor	Ph.D.
Dr. Michael Gerken Chair, Thesis Examination Committee	Professor	Ph.D.

For my parents Candace and Randy Dueck. Their constant love and support made the creation of this thesis possible.

Abstract

Pseudouridine is an isomer of uridine and a post-transcriptional RNA modification in all domains of life. Its formation is catalyzed by pseudouridine synthases functioning as standalone proteins or protein-RNA complexes called H/ACA small nucleolar ribonucleoproteins (H/ACA snoRNPs). Whereas pseudouridylation was long known to assist in stabilizing noncoding RNAs, it was discovered also in mRNAs in 2014. mRNA pseudouridylation varies in response to stress, and artificial pseudouridylation of stop codons facilitates nonsense suppression. However, the functional significance of mRNA pseudouridylation and the regulation of the pseudouridine synthases responsible are only partially understood. To study mRNA pseudouridylation by specific pseudouridine synthases, I developed a fluorescence-based reporter system in *S. cerevisiae* and used it to assay for standalone Pus7, snR5 and snR81 H/ACA snoRNP activity. This novel system detects snR5 H/ACA snoRNP activity and can be used to further investigate the role of pseudouridine synthases in gene expression.

Acknowledgements

I would like to thank my supervisor Dr. Ute Kothe for her compassion and understanding throughout my undergraduate and graduate studies in her lab. She has been a wonderful mentor and helped me to progress with my studies in the face of immense hardships. I am truly grateful for all that she has done.

I would like to thank Dr. Julia Guegueniat for lending her expertise in yeast genetics to my project. Her insights and assistance were crucial in obtaining the yeast strains which were the basis of my research.

Thanks to Dr. Andy Hudson for sharing his knowledge of sequence and ligation-independent cloning early on in my project, allowing me to establish a cloning pipeline with relative ease.

I would also like to thank Luc Roberts and Justin Vigar for helping me to conduct flow cytometry assays and independent study student Dixon Jensen for completing the cloning of the suppressor tRNA yeast expression vectors.

Finally, I would like to thank Dr. Roy Golsteyn and Dr. HJ Wieden for serving on my supervisory committee and offering feedback on my work.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

4SU	4-Thiouridine
A site	aminoacyl tRNA site
ADE2	Phosphoribosylaminoimidazole carboxylase (selective marker)
ASL	anti-codon stem loop
ATP	Adenosine triphosphate
a.u.	Arbitrary unit
bp	base-pair
CAB-box	Cajal body box
CB	Cajal Body
Cbf5	Centromere binding factor 5
CDC33	Cell Division Cycle protein 33
cDNA	complementary DNA
CDS	Coding sequence
CeU-Seq	N_3-CMC-enriched pseudouridine sequencing
CIP	Calf intestinal phosphatase
CLIP	Cross linking and immuno-precipitation
CMCT	N-cyclohexyl-N′-(2-morpholinoethyl)carbodiimide methyl-*p*-toluenesulfonate
CUP1	Copper metallothionein protein 1
DAVID	Database for Annotation, Visualization and Integration Discovery
dCTP	deoxycytidine triphosphate
ddH_2O	Double-distilled H_2O
DGCR8	DiGeorge syndrome critical region 8
DKC1	Dyskerin: human homolog of Cbf5
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
dsDNA	double-stranded DNA
EDTA	Ethylenediaminetetraacetic acid
EEF1A1	Eukaryotic translation elongation factor 1 alpha 1
EFT1	Elongation factor 2
EF-Tu	Elongation factor thermo unstable
EGFP	enhanced green fluorescent protein
eRF1/3	eukaryotic release factor 1/3
ESI	electrospray ionization
EtBr	Ethidium Bromide
FACS	Fluorescence-activated cell sorting
FITC-A	Fluorescein isothiocyanate-area

FMDV 2A	Foot-and-mouth disease virus 2A element
FSC-A	Forward scatter area
GAL1,10	Bi-directional galactose-inducible promotor
Gar1	Glycine-Arginine rich Protein 1
GFP	Green fluorescent protein
GO	Gene Ontology
GTP	Guanosine triphosphate
HEK293	Human embryonic kidney 293 cells
HeLa	Henrietta Lacks cells
HGMD	Human Gene Mutation Database
HIS3	Imidazoleglycerol-phosphatase dehydratase (selective marker)
hnRNP C	heterogenous nuclear ribonucleoprotein C
HRP	Horseradish peroxidase
HSP104	Heat shock protein 104
HSP60	Heat shock protein 60
IDT	Integrated DNA Technologies
IgG	Immunoglobulin g
KanMX	bacterial aminoglycoside phosphotransferase (geneticin resistance marker)
K_d	Dissociation constant
L7Ae	50S ribosomal protein L7Ae
LB	Lysogeny broth
LEU2	Beta-isopropylmalate dehydrogenase (selective marker)
LYS2	Alpha aminoadipate reductase (selective marker)
m^1A	N1-methyl-adenosine
m^1G	N1-methyl-guanosine
m^5C	5-methylcytosine
m^5U	5-methyluridine
m^6A	N6-methyl-adenosine
m^7G	7-methylguanosine
MALDI	matrix-assisted laser desorption ionization
MDJ1	Mitochondrial DnaJ protein
MeRIP-Seq	methylated RNA immunoprecipitation with next generation sequencing
MET15	O-acetyl homoserine-O-acetyl serine sulfhydrylase (selective marker)
miRNA	micro RNA
mRNA	messenger RNA
mRuby2	monomeric Ruby 2 protein
NAB2	Nuclear Polyadenylated RNA-binding protein 2

NAF1	Nuclear assembly factor 1
NAM7	Nuclear Accommodation of Mitochondria protein 7
ncRNA	noncoding RNA
NEB	New England Biolabs
Nhp2	Nucleolar Protein Family A, Member 2 protein
NMD	nonsense mediated decay pathway
Nop10	Nucleolar protein 10
NOPP140	Nucleolar phosphoprotein 140
nt	nucleotide
ONT	Oxford Nanopore Technology
P2A	Porcine teschovirus-1 2A element
P site	peptidyl tRNA site
PAR-CLIP	Photo-Activatable Ribonucleoside enhanced CLIP
PCR	Polymerase chain reaction
PE-Texas-Red-A	Phycoerythrin Texas red-area
Pho85	Phosphate metabolism protein 85
PMT	Photomultiplier tube
PNK	Polynucleotide kinase
Poly (I:C)	Polyinosinic:polycytidylic acid
Prp5	Pre-mRNA Processing protein 5
PTC	premature termination codon
PUA	Pseudouridine synthase and Archaeosine transglycosylase domain
Pus1-10	Standalone pseudouridine synthases 1 through 10
qPCR	quantitative PCR
RBS-Seq	RNA bisulfite sequencing
RF1	release factor 1
RFP	Red fluorescent protein
RNA	Ribonucleic Acid
RPB10	RNA Polymerase subunit ABC10-beta
RPL11A	Ribosomal 60S subunit protein L11A
rRNA	Ribosomal RNA
RT	reverse transcription
SBA1	Sensitivity to Benzoquinone Ansamycins protein 1
SBP1	Single-stranded nucleic acid-binding protein 1
SCARLET	Site-specific Cleavage And Radioactive-labeling followed by Ligation-assisted Extraction and Thin-layer chromatography
scaRNA	small Cajal body-specific RNA
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SLIC	Sequence and ligation-independent cloning
SMG1	Serine/threonine protein kinase involved in NMD pathway

snoRNP	small nucleolar ribonucleoprotein
snR5	small nucleolar RNA 5
snR81	small nucleolar RNA 81
snRNA	small nuclear RNA
sRNP	small ribonucleoprotein
SSA4	Stress-Seventy subfamily A protein 4
SSB2	Stress-Seventy subfamily B protein 2
SSC-A	Side scatter area
ssDNA	single-stranded DNA
SUP4	UAA stop codon suppressor tRNA charged with tyrosine
SUP53	UAG stop codon suppressor tRNA charged with leucine
T2A	Thosea asigna virus 2A element
Tan1	tRNA Acetylation protein 1
TBS	Tris-buffered saline
TEF1	Translational elongation factor EF-1 alpha
TEF2	Paralog of TEF1
TLC	Thin layer chromatography
TRM4	tRNA:m^5C-methyltransferase
tRNA	transfer RNA
TRP1	Phosphoribosylanthranilate isomerase (selective marker)
U2A′	U2 small nuclear ribonucleoprotein A′
U2B″	U2 small nuclear ribonucleoprotein B″
UPF1	Up-frameshift suppressor 1 homolog
UPF2	Up-frameshift suppressor 2 homolog
UPF3b	Up-frameshift suppressor 3 homolog B
URA3	Orotidine-5′-phosphate decarboxylase (selective marker)
UTR	Untranslated region
UV	Ultra-Violet
WDR79	WD Repeat-containing protein 79
WT	Wild type
yoEGFP	yeast optimized enhanced green fluorescent protein
YPD	Yeast extract peptone dextrose
Ψ	Pseudouridine

Chapter 1–Introduction

1.1–Pseudouridine and its properties

Post-transcriptional RNA modifications constitute a diverse array of over 160 currently identified chemical alterations to the basic A, U, G, and C nucleosides in RNA (*1*). Once introduced, these modifications can influence an RNA molecule's structure and stability, and can also impact the RNA's binding properties with other biomolecules (*2*). Among these modifications, pseudouridine (Ψ) is the most abundant and present across the three domains of life (*3*). First discovered in the 1950s by chromatographic analysis of RNA hydrolysates, pseudouridine has since been identified in a variety of functional RNAs, including ribosomal RNA (rRNA), transfer RNA (tRNA), and small nuclear RNA (snRNA) (*3-5*). As of 2014, pseudouridine has also been found in messenger RNAs (mRNAs) (*6, 7*). Pseudouridine is an isomer of uridine. It is generated from uridine by cleavage of the N1-C1 glycosidic bond between the base and ribose sugar, followed by base reorientation and reattachment via a C5-C1 glycosidic bond (**Figure 1A**). In the resultant pseudouridine molecule, the nitrogen atom (that was previously in the N1-C1 glycosidic bond of uridine) now forms an imino (N1H) group, thus increasing the hydrogen-bonding potential of the modified base (*8, 9*). Pseudouridine's increased hydrogen bonding relative to uridine has been observed in numerous crystal structures, including the *Escherichia coli* (*E. coli*) ribosome and in multiple tRNAs, wherein the imino proton of pseudouridine forms a water bridge with the phosphate backbones of these respective RNA species (**Figure 1B**) (*9-11*).

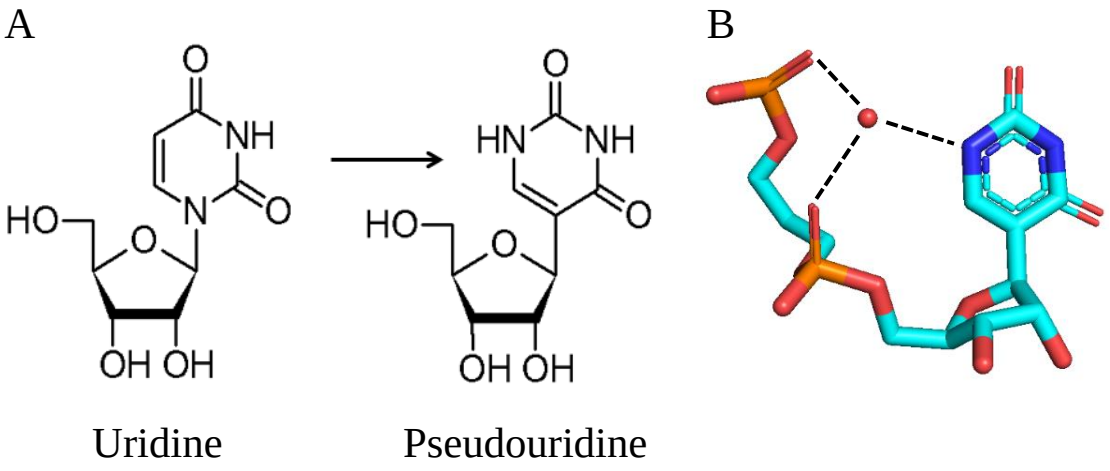


Figure 1. The chemical structure of uridine and pseudouridine. (A) To form pseudouridine from uridine, the N1-C1 glycosidic bond in uridine is first cleaved. Subsequently, the uracil base is repositioned and connected to the ribose sugar via a novel C5-C1 glycosidic bond, thus forming pseudouridine. **(B)** In pseudouridine, the nitrogen atom that was previously in the uridine glycosidic bond now forms an imino group in the nucleobase, which can coordinate a water molecule (red sphere) with the phosphate backbone of both pseudouridine and the residue preceding pseudouridine, thereby stabilizing the local secondary structure in an RNA (model of Ψ55 and the phosphate backbone of the preceding m^5U residue in yeast $tRNA^{Phe}$; PDB ID 1EHZ, (*11*)).

The water bridge interaction mediated by pseudouridine reduces the flexibility of the sugar-phosphate backbone and limits the rotational freedom of the base, locking it in an *anti* conformation. Furthermore, in this state pseudouridine favours a C3′-endo conformation of its ribose moiety, leading to a rigidifying effect on the sugar-phosphate backbone that enhances local base stacking (*8, 12*). Thus, relative to uridine, pseudouridine has a stabilizing effect when present in an RNA duplex, which may explain why this modification is found clustered in functionally important regions in non-coding RNAs (*13*). For instance, 6 pseudouridines are located in the peptidyl transferase center of yeast 25S rRNA, where it has been shown that loss of a conserved pseudouridine in the A site loop drastically impairs translation and that further loss of the remaining 5 pseudouridines in the peptidyl transferase center has a negative synergistic effect (*14*). Another example of

pseudouridine-conferred structural stability is found in the RNA duplex formed by the branch site recognition region of U2 snRNA and the branch point sequence of the intron. Here, a conserved pseudouridine in U2 snRNA stabilizes the extrahelical orientation of the branch point adenosine such that its 2′ OH group is positioned for nucleophilic attack on the 5′ splice site (*15*, *16*).

1.2–Pseudouridine synthases: standalone enzymes and protein-RNA complexes

A group of enzymes called pseudouridine synthases are responsible for catalyzing the isomerization of uridine to pseudouridine. These enzymes can be categorized into 6 families. Five of these families are named after a representative bacterial pseudouridylase: RluA, RsuA, TruA, TruB, and TruD (*17*). The sixth family, Pus10, encompasses enzymes that have only been identified in certain archaea and eukaryotes (*18*). Although the enzymes from these 6 families exhibit minimal sequence similarity, all pseudouridine synthases possess a core common fold and active site cleft containing a conserved aspartate residue necessary for catalysis (*17, 19*). This aspartate residue serves as an acid/base catalyst in a reaction mechanism for pseudouridine formation that occurs through a glycal intermediate (*20*).

Pseudouridylation of an RNA substrate occurs in a site-specific manner. With the exception of the TruB homolog Cbf5, which requires a set of accompanying protein and RNA components to create a modification complex (vide infra), all other pseudouridine synthases function as standalone enzymes that recognize their RNA substrates without any additional factors. Pseudouridine synthases achieve site specificity in modification by recognition of an RNA sequence or structural element. For many of the standalone pseudouridine synthases, substrates are identified through the structural dependent-mode

of recognition. For example, the standalone pseudouridine synthase TruB, responsible for the universally conserved Ψ55 modification in the TΨC arm of virtually all elongator tRNAs, recognizes the three dimensional structure of the T stem-loop (*21*).

Contrasting structure-based substrate recognition, the pseudouridine synthase Cbf5 functions in a protein-RNA complex that modifies RNA in a sequence-specific manner. This complex, called an H/ACA small nucleolar ribonucleoprotein (H/ACA snoRNP), is present in Archaea and Eukarya and exists alongside the C/D box class of snoRNPs responsible for site specific 2′-O-methylation of RNA (*22, 23*). The H/ACA snoRNP complex consists of four core proteins: Cbf5, Nop10, Nhp2 (L7Ae in archaea), and Gar1 that associate with a H/ACA guide RNA (**Figure 2A**) (*23, 24*). In this complex, Cbf5 is the catalytic component (*19*). However, unlike its bacterial counterpart TruB or the other standalone pseudouridine synthases, Cfb5 does not directly recognize an RNA substrate and instead relies upon an H/ACA guide RNA. In eukaryotes, H/ACA guide RNAs consist of a two-hairpin structure with a characteristic H box (sequence ANANNA) found in the hinge region between the 5′ and 3′ hairpins and ACA box located downstream of the 3′ hairpin. Each hairpin contains internal loop regions called pseudouridylation pockets which base pair in a bipartite manner to a target sequence in an RNA substrate, thereby isolating an unpaired uridine at the base of the upper stem in the guide RNA (**Figure 2B**) (*25, 26*). Recognized in this sequence-specific manner, the isolated uridine is subsequently isomerized to pseudouridine by Cbf5. Thus, the sequence-dependent mode of RNA substrate recognition conferred by the H/ACA guide RNA allows the core protein complex to modify a diverse number of uridines at different sites and in different substrates depending on which guide RNA is bound to the complex. For example, the *S. cerevisiae*

snR5 H/ACA guide RNA directs the isomerization of two positions in 25S rRNA: the pseudouridylation pocket in the 5′ hairpin directs isomerization of U1124 and the 3′ hairpin's pseudouridylation pocket targets U1004 (*25*). Altogether, 28 H/ACA guide RNAs direct rRNA pseudouridylation at 44 different positions in baker's yeast (*27*). Intriguingly, in some instances, the two hairpins of an H/ACA guide RNA each target a different class of RNA substrate: the 5′ pocket of snR81 directs Ψ42 formation in U2 snRNA whereas the 3′ pseudouridylation pocket specifies Ψ1052 formation in 25S rRNA (*28*). Finally, in addition to directing pseudouridylation of rRNA and snRNAs, H/ACA guide RNAs likely also target snoRNAs and mRNAs (*7*, *29*).

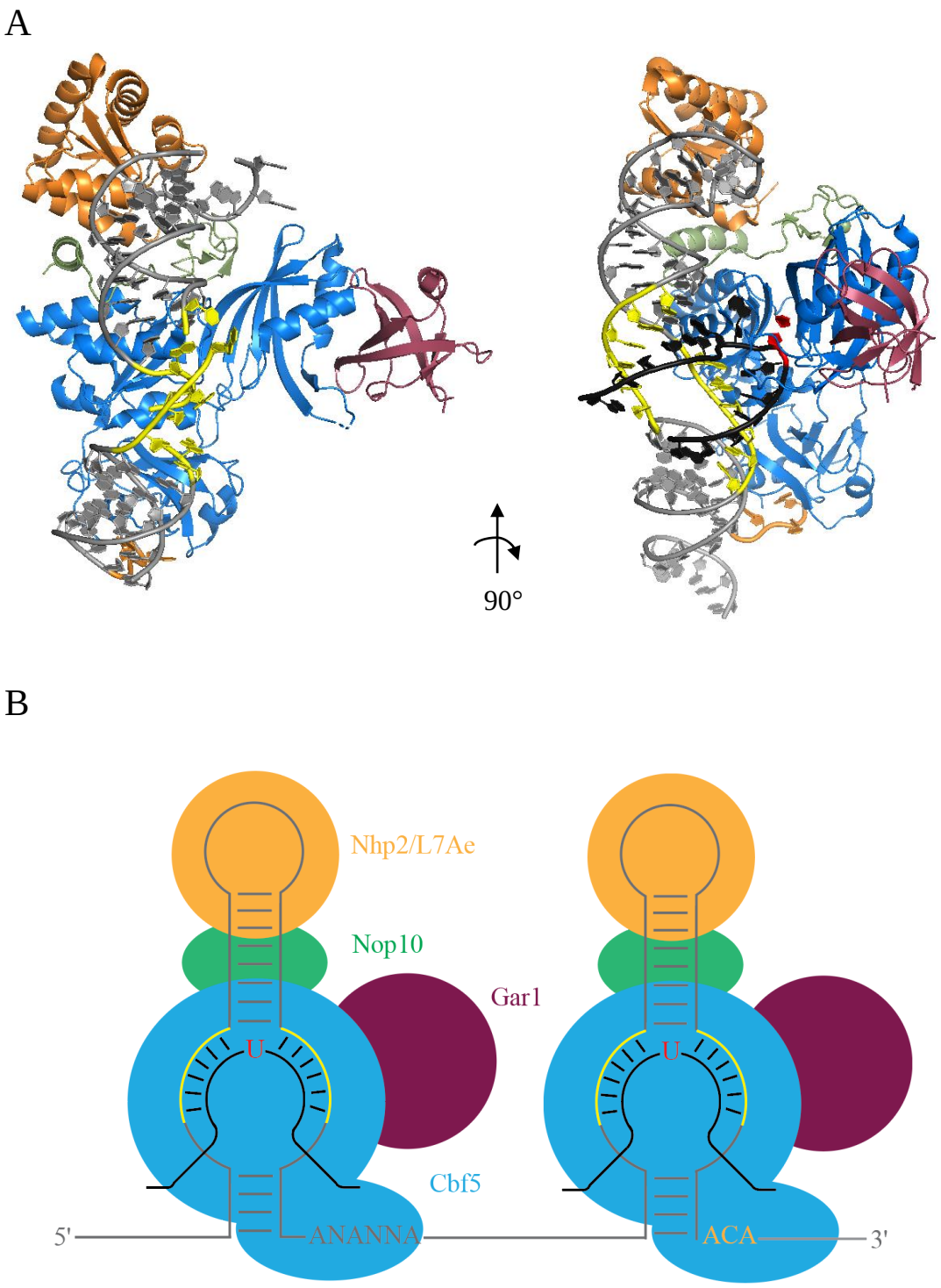


Figure 2. Structure of the H/ACA sRNP. (A) Crystal structures of a *Pyrococcus furiosus* H/ACA small ribonucleoprotein (sRNP) without a substrate (left, PDB ID 2HVY, (*30*)) and with a substrate RNA bound in the H/ACA guide RNA's pseudouridylation pocket (right, PDB ID 3HAY, (*31*)). The H/ACA sRNP structure is composed of the catalytic protein Cbf5 (blue), Gar1 (raspberry), Nop10 (green), and L7Ae (orange) in complex with

a H/ACA guide RNA (grey). Site-specific pseudouridylation is conferred by base pairing interactions between single-stranded regions of the guide RNA, called the pseudouridylation pocket (yellow) with a target sequence in a substrate RNA (black). Upon binding to the complex, the target residue (red; in this structure the highlighted residue is a 5-fluoro-6-hydroxy-pseudouridine product resulting from isomerization of f^5U as a stand-in for uridine) is isolated at the base of the upper stem of the guide RNA and positioned in the active site of Cbf5 for catalysis. A conserved ACA box (orange) binds to the PUA domain of Cbf5 and helps to position the guide RNA's pseudouridylation pocket with the active site. The substrate-bound H/ACA sRNP structure is rotated 90° relative to the unbound structure to better illustrate the base-pairing interactions between guide RNA and substrate and to show the positioning of the target residue within the active site of Cbf5. **(B)** Schematic representation of a two-hairpin H/ACA sRNP. Colouring is the same as in (A). In a eukaryotic H/ACA small nucleolar ribonucleoprotein (snoRNP), Nhp2 takes the place of its archaeal homolog L7Ae. In both Archaea and Eukaryotes, the H/ACA sRNP proteins are thought to associate with each stem loop of the typical double hairpin H/ACA guide RNA structure, wherein the conserved H box (ANANNA) and ACA box elements are necessary for Cbf5 binding (*24, 32*).

Cbf5 contains two domains: the catalytic domain (subdivided into D1 and D2) and the pseudouridine synthase and archaeosine transglycosylase (PUA) domain (*30, 33*). The PUA domain binds the ACA box and lower stem of the guide RNA, and thereby is responsible for correct positioning of the pseudouridylation pocket with the active site of Cbf5 (*26, 30*). Gar1 and Nop10 directly bind Cbf5 to form a core trimeric complex, with Nop10 also making contact with the guide RNA's upper stem (*30*). Thus, Nop10 extends the guide RNA binding face of Cbf5 in addition to stabilizing its active site (*17*). Gar1 exclusively associates with the D2 subdomain of Cbf5's catalytic domain and plays a role in RNA substrate loading and product release by stabilizing the open state of a substrate binding loop found in Cbf5's D2 subdomain (*30*). Nhp2, homologous to L7Ae in Archaea, binds the upper stem loops of the guide RNA hairpins. In Archaea, L7Ae specifically recognizes a kink-turn motif in the upper stem; however, eukaryotic H/ACA guide RNAs lack this motif and Nhp2 exhibits no kink-turn specificity (*30, 34*). Nhp2/L7Ae, in

conjunction with Nop10 anchors the upper stem of the guide RNA to the protein complex, thus facilitating correct positioning of the target uridine in the active site of Cbf5 (*35*).

Finally, it should be noted that the four core proteins of the H/ACA snoRNP complex play a dual functionality in mammals, wherein they also associate with an H/ACA motif present in the RNA template component of the telomerase ribonucleoprotein (*36*). In the telomerase RNP, Cbf5 does not serve a catalytic role. Rather, it has been proposed that Cbf5, Nop10, Nhp2, and Gar1 perform a scaffolding function that stabilizes the telomerase RNA component (*17, 36*). Furthermore, an H/ACA snoRNP complex containing *S. cerevisiae* snR30 H/ACA snoRNA (U17 in vertebrates) is required for the A_0, A_1, and A_2 cleavages of 35S pre-RNA during 18S rRNA processing (*37*).

1.3–Methods of RNA modification detection

The field of RNA modification study had its inception in the 1950s when a peak of unknown origin was observed alongside those of the A, U, G, and C nucleotides on a chromatogram of hydrolyzed calf liver RNA (*38*). Subsequent interrogations identified this peak as a 5′-ribosyl isomer of uridine and designated it pseudouridine (*4*). Since then, a robust repertoire of RNA modification detection methods has been developed, allowing for the identification of over 160 post-transcriptional RNA modifications (*1*). In the following section, various modification detection methodologies will be discussed with an emphasis on those responsible for pseudouridine detection.

1.3.1–Reverse transcription-based modification detection

One of the most prolific modification detection methods is a reverse-transcription (RT) and sequencing based strategy in which RT signatures are identified in complementary DNA (cDNA). When a modified nucleotide in an RNA polymer is

encountered by reverse transcriptase, a disruption in nucleotide incorporation may occur in the cDNA depending on the modification's identity. Upon encountering bulkier modifications or modifications that alter the Watson-Crick face of the nucleobase, the reverse transcriptase tends to abort cDNA synthesis or mis-incorporates a non-complementary dNTP at this position. For instance, N1-methyl-adenosine (m^1A) and N1-methyl-guanosine (m^1G) disrupt the enzyme's ability to incorporate thymine and cytosine, respectively, resulting in apparent mutations in the cDNA and an increased incidence of truncated cDNA products at a position corresponding to the methylated base in the RNA template (*39, 40*). Thus, the detected aberrations in nucleotide incorporation that occur in response to m^1A or m^1G constitute the RT signature of these methylated bases. The signature of m^1G has been used to detect the presence of this modification at position 9 in *Arabidopsis thaliana* $tRNA^{Pro}$ and $tRNA^{Val}$ (*39*). On a larger scale, m^1A's RT signature has been applied to transcriptome wide mapping studies that have identified this modification in mRNAs (*41-43*).

When encountering other modifications, including pseudouridine (Ψ), 5-methyluridine (m^5U), or 5-methylcytosine (m^5C), the reverse transcriptase is capable of faithful dNTP incorporation into the cDNA strand as these modified bases can engage in canonical base pairing interactions (*44*). As such, these modifications are considered RT-silent in the resulting cDNA. However, some RT-silent modifications can be made to elicit a RT signature by chemical pre-treatment of the RNA with a reagent that selectively reacts with the modification of interest. To detect pseudouridine in a reverse-transcription and sequencing assay, the RNA is first reacted with N-cyclohexyl-N′-(2-morpholinoethyl)carbodiimide methyl-*p*-toluenesulfonate (CMCT). This reaction results

in all U, Ψ, and G residues being labelled with the bulky CMCT adduct. Upon a subsequent alkaline hydrolysis treatment, this adduct is readily cleaved from all U and G bases while leaving the Ψ-CMCT adducts intact due to the unique stability of the CMCT bond to N3 of pseudouridine (*45, 46*). Thus, CMCT treatment followed by alkaline hydrolysis allows for the specific labelling of pseudouridines with a bulky adduct capable of blocking reverse transcription one base downstream of the labelled pseudouridine, thereby giving this previously silent modification a discernible RT signature.

CMCT labelling of pseudouridines was first employed by Bakin and Ofengand in 1993 to locate four pseudouridine residues in the peptidyl transferase center of *E. coli* 23S rRNA (*45*). Since then, this method has been leveraged to great effect in a series of seminal 2014 and 2015 investigations that coupled the CMCT/RT approach for pseudouridine detection with next generation sequencing. These studies yielded transcriptome-wide maps of pseudouridine occurrence, establishing the presence of this modification in mRNAs (*6, 7, 47, 48*). As demonstrated by these studies, the greatest strength of this pseudouridine detection method lies in its high-throughput nature, with hundreds of modification sites identified through Illumina sequencing. However, a number of limitations inherent to this method exist, including the labelling efficiency of CMCT; the reaction may not label all pseudouridines in a transcript, thereby leading to false negatives during sequencing. Additionally, the alkaline hydrolysis treatment may leave some G and U-CMCT adducts intact, leading to their false positive identification as pseudouridines upon sequence analysis (*49*). Furthermore, reverse transcription will not always faithfully terminate at a CMCT-labelled pseudouridine, leading to artefacts in the sequencing data that interfere with confident modification assignment to a specific position in the original RNA

molecule. To remedy this, during modification calling (the process whereby a position in an RNA molecule is deemed a good candidate for being modified) the signal-to-noise ratio generated from the sequencing data is addressed by bioinformatic analysis wherein a signal threshold is determined that aims at separating genuine modification events from false positives. Nevertheless, pseudouridylation sites identified through this approach then also run the risk of being overlooked during analysis as false negatives arising from overly stringent modification calling thresholds.

Another limitation of high-throughput pseudouridine sequencing concerns the reproducibility of the pseudouridylation maps generated via this method. A study comparing the transcriptome-wide pseudouridylation data sets found little overlap; of 402 unique coding sequences identified across three independent studies as pseudouridylation targets in log phase yeast, only a single hit corresponding to *RPL11a* mRNA was found in common (*50*). As *RPL11a* is one of the top 30 most expressed genes in yeast, the identity of this one common hit highlights a further limitation of the high throughput sequencing method, namely, that it is biased towards detecting modifications in abundantly expressed transcripts while neglecting those in comparatively rarer RNA species (*50, 51*). A variation of the CMCT/RT method called N_3-CMC-enriched pseudouridine sequencing (CeU-Seq) has been developed to partially address this concern. Specifically, an N_3-CMC derivative is used in the chemical labeling of pseudouridines in an RNA. This derivative is then conjugated to biotin in a subsequent reaction step, allowing for streptavidin pulldown of biotinylated CMCT-Ψ adducts, thus enriching the population of pseudouridylated RNAs prior to sequencing and increasing the number of rarer pseudouridylated RNAs in a data set (*47*). A final limitation to note regarding high-throughput pseudouridine sequencing

following CMCT/RT is that the pseudouridylation extent at a particular site cannot be quantitatively determined, although relative pseudouridine stoichiometries can be (*7*).

A method has been derived from the high-throughput RT-based sequencing detection of pseudouridine that is able to probe the modification status of individual sites in an RNA of interest. Beginning with CMCT treatment of an RNA sample, reverse-transcription of a putatively modified RNA template is coupled to real-time PCR (qPCR). By utilizing Superscript II reverse-transcriptase and reaction conditions containing Mn^{2+}, the incidence of the polymerase bypassing the pseudouridine-CMCT adduct and continuing cDNA synthesis is increased. The result of this readthrough is either a mutation or a deletion in the cDNA. Next, when the mutated cDNA is subjected to qPCR with site specific primers, the mutations/deletions in the cDNA create a characteristic change in the melting curve relative to that of amplified cDNA derived from unmodified RNA, allowing for detection of the pseudouridylated residue (*52*). Designed to investigate the pseudouridylation status of single sites, this method has been employed to validate putative modification targets in *A. thaliana* mRNAs initially identified through high throughput pseudouridine sequencing (*53*).

Alongside CMCT/RT-based pseudouridine identification, another high throughput sequencing-based detection strategy, called RNA bisulfite sequencing (RBS-Seq) has been recently reported. RBS-Seq involves treatment of an RNA with bisulfite reagent prior to reverse transcription and sequencing. Unlike the CMCT/RT method for pseudouridine detection in which CMCT-labelled residues terminate reverse transcription, in RBS-seq, bisulfite treatment ultimately results in pseudouridine residues producing a characteristic 1-2 nucleotide deletion signature in cDNA, thus allowing for detection of the modification

upon sequencing. The reaction mechanism underlying this phenomenon is not completely understood at this time, but initial characterization of the pseudouridine-bisulfite adduct has found that in addition to bisulfite reagent, accompanying 75°C heat treatment and the presence of 20 mM Mg^{2+} are required to manifest the cDNA deletion signature (*54*). A proposed mechanism for bisulfite labelling first involves a bisulfite group being added to the pseudouridine base, followed by heat-induced migration of bisulfite to the ribose sugar, whereupon the base of pseudouridine is re-aromatized and bisulfite renders a ring-opened sugar adduct (*54*, *55*). Once opened, Mg^{2+} is postulated to reorient the ring-opened pseudouridine-bisulfite adduct away from the reverse transcriptase active site or to otherwise alter the local RNA template structure in such a way that induces the reverse transcriptase to bypass the pseudouridine adduct during cDNA synthesis (*54*).

In addition to pseudouridine detection, RBS-Seq is able to simultaneously probe for two other RNA modifications, m^5C and m^1A, in RNA. Bisulfite treatment deaminates unmodified cytidines, thus converting them to uridines in an RNA template whereas m^5C is not susceptible to this deamination. Therefore, upon reverse transcription and sequencing, unmethylated cytidines are detected as thymidine in the cDNA and sites of m^5C methylation are retained as cytidines in the analysis. For m^1A, in the absence of bisulfite treatment this modification tends to cause cDNA mutations during reverse transcription as the methyl group disrupts Watson-Crick base-pairing with thymidine. Upon reacting m^1A with bisulfite, the methyl group is transferred from the N1 to N6 position of adenosine via Dimroth rearrangement, thus generating m^6A in the RNA template, which faithfully leads to thymidine incorporation during reverse transcription (*56*). Therefore, sites of m^1A modification are ascertained by comparing the sequencing

results from a sample reacted with and without bisulfite. RBS-seq, being a sequencing-based method that relies on analyzing RT signatures in cDNA, shares the aforementioned limitations noted for pseudouridine detection via the CMCT/RT approach. However, as this method uses an alternative 1-2 nucleotide deletion signature for pseudouridine detection compared to the CMCT-based reverse transcription termination signature, RBS-seq can be used as an independent high-throughput sequencing method to verify transcriptome-wide modification sites called through CMCT/RT and sequencing analysis. Furthermore, in addition to nanopore sequencing (vide infra), RBS-seq is distinguished as being one of the only high-throughput sequencing methods that can detect multiple types of modifications in a single experiment (pseudouridine, m^5C, and m^1A) and therefore constitutes a notable advancement in the pursuit of comprehensively mapping the post-transcriptional RNA modification landscape (*54*).

Whereas the RT-based high throughput sequencing strategies are able to identify hundreds of *de novo* modification sites in a single assay, most other modification detection methodologies are more limited in scope. However, due to the aforementioned limitations of the high throughput RT-based sequencing strategies, it has been proposed that any modifications identified through these techniques be treated as strictly putative until validated by another detection method (*44*). Therefore, what other detection methods lack in breadth, they make up for in depth when compared with the sheer scope of the data sets generated via RT-based sequencing. Furthermore, the complement of methods that will next be discussed offer dimensions of information not available through high throughput sequencing, including absolute modification stoichiometries and insight into the enzymes responsible for modification.

1.3.2–Radiolabeling of RNA modifications

A versatile method for detecting a single modified position in an RNA involves site-specific radiolabeling. In this approach, the RNA polymer is first cleaved at a position immediately 5′ to the modified site of interest, thus leaving a free 5′ phosphate group attached to the ribose sugar of the modification site. This site-specific cleavage can be accomplished with RNase H, which is directed by a chimeric 2′-O-methyl RNA-DNA oligonucleotide (*57*) or with a DNAzyme (*58*). After purification of the 3′ fragment containing the modification site of interest at its 5′ end, its 5′ terminus is dephosphorylated with calf intestinal phosphatase (CIP) and subsequently rephosphorylated by polynucleotide kinase (PNK) and [ɣ-^{32}P]ATP. Once radiolabeled, the 3′ fragment is completely digested with nuclease P1 and the resultant nucleotides are resolved by thin layer chromatography (TLC). As the modified site of interest is the only labeled position in the final digest, both the unmodified and modified forms of the nucleotide can be distinguished on the TLC plate, thus allowing for quantitative determination of modification extent at that position (*57*). A powerful variation of this method couples site specific radiolabeling of a modification position with a second site-specific selection step, thus increasing the sensitivity of the radiolabeling method. In the SCARLET method (Site-specific Cleavage And Radioactive-labeling followed by Ligation-assisted Extraction and Thin-layer chromatography), a splint ligation reaction takes place after the 3′ fragment (containing the modified site of interest at the 5′ end) has been radiolabeled. Splint ligation entails the use of a splint DNA oligonucleotide to bring the radiolabeled RNA fragment and a long single-stranded DNA oligonucleotide into close proximity, allowing for subsequent ligation such that the radiolabeled RNA is attached to the 3′ end of the ssDNA

oligonucleotide. During subsequent RNase treatment, the radiolabeled position at the 3′ end of the ssDNA probe is protected and subsequently isolated with gel purification. After nuclease P1 treatment, the mononucleosides from the recovered ssDNA oligos are then resolved on a TLC plate. As any detected radioactive signal arises from the ^{32}P-labelled site of interest in the original RNA sample, both the modification status and the modification stoichiometry of that site can be determined (*59*).

In principle, the site-specific radiolabeling detection method can be used to investigate any modification that i) does not disrupt Watson-Crick base pairing with the chimeric 2′-O-methyl RNA-DNA oligo necessary for specific RNase H cleavage (or the base-pairing of the DNAzyme) and ii) migrates at a different rate relative to the unmodified form of the nucleotide on a TLC plate. Since the sequence context of the modification must be known in advance of performing the radiolabeling procedure, this method cannot be used for *de novo* modification apprehension. An additional shortcoming of this method is that a considerable amount of starting RNA material is required for analysis of a specific transcript as the site-specific radiolabeling procedure does not contain an amplification step like in the high-throughput sequencing strategies. Therefore, this method is not well suited to detect modified sites in rare RNA species. However, the single nucleotide resolution at which this method can detect and quantitate modifications in relatively abundant RNAs serves as a powerful form of validation when paired with the high throughput RT-based sequencing strategies. Notably, SCARLET was the first independent method used to corroborate the presence of pseudouridine in mRNA after its initial discovery in the transcriptome-wide pseudouridine sequencing investigations (*47*).

Of final note in regards to radiolabelling is a pseudouridine-specific assay used to investigate pseudouridylation activity *in vitro*: A tritium release assay is conducted by incubating a purified pseudouridine synthase of interest with substrate RNA containing tritiated uridine at the fifth position on the pyrimidine ring. Upon isomerization and formation of pseudouridine's C5-C1 glycosidic bond, the tritium atom is displaced into the solution of the *in vitro* reaction. The amount of radioactivity present in the solution is therefore dependent on the activity of the pseudouridine synthase of interest and can be quantified with a scintillation counter. Subsequent analysis of the tritium level in solution at various time points allows for the kinetic parameters of pseudouridine formation by the enzyme under scrutiny to be determined (*60*). Unlike site-specific radiolabelling with ^{32}P, the tritium release assay is specifically configured to detect the modifying activity of pseudouridine synthases by exploiting knowledge of the pseudouridylation reaction mechanism. The strength of this assay lies in its ability to offer insight into pseudouridine synthase kinetics during pseudouridylation of a specific site in a known RNA substrate reconstituted for *in vitro* observation. As such, the tritium release assay is not designed for large-scale screening of putative modification sites due to the labourious work-up of purified and active pseudouridine synthases and tritiated substrates.

1.3.3–Mass spectrometric detection of pseudouridine and other RNA modifications

Mass spectrometry-based methods have also been developed to detect post-transcriptional RNA modifications. Whereas all other RNA modifications have mass differences relative to their unmodified counterparts, pseudouridine is mass-silent with respect to uridine and therefore presents a unique challenge for mass spectrometric detection. To overcome this challenge, pseudouridine can be labeled prior to analysis via

aforementioned CMCT derivatization, giving pseudouridine a characteristic mass shift of 252 Da (*49*). Other labeling strategies include cyanoethylation by way of acrylonitrile and reaction with methyl vinyl sulfonate (*61, 62*). A common mass spectrometry approach entails performing liquid chromatography on a digested RNA sample prior to electrospray ionization (ESI) or matrix-assisted laser desorption ionization (MALDI). Subsequent mass spectrometry analysis is then used to identify the mass-to-charge ratios of modified nucleotides in the ionized digests (*63, 64*). The chromatographic retention times and mass values obtained from this approach allow for the detection of modifications in the original RNA sample but do not offer any sequence context information. To locate the position of modification, a number of strategies exist, including RNase mapping and the use of biotinylated DNA probes to selectively purify an RNA sample prior to mass spectrometric analysis (*65, 66*). Finally, advancements in the sensitivity of mass spectrometry instruments have made possible a labeling-free method of pseudouridine detection that also provides information on the site of modification. This technique uses collision-induced dissociation to create fragmented ionic species, of which a pseudouridine signature ion ($[C_9H_7N_2O_4]^{1-}$) is searched for that possesses a mass to charge ratio of 207.04 (*67*). Serving as a powerful validation method, mass spectrometric assays can also detect modification sites *de novo*. Suited to investigating the modification status of specific RNA species, mass spectrometry is not well tailored to transcriptome-wide studies; arising from the difficulty it takes to transfer RNA polyanions into a vacuum, the analysis of a total mRNA population would require a prohibitively large sample amount (*44*).

1.3.4–Antibody-based recognition of RNA modifications

Antibodies can be raised to specifically bind a modification of interest, thereby offering an invaluable technique to enrich the population of modified transcripts in a sample prior to analysis with the previously discussed methods. For instance, in a study using MeRIP-Seq (methylated RNA immunoprecipitation with next generation sequencing), a method for m^6A modification detection, anti m^6A antibodies were applied to total mouse brain RNA extracts followed by immunoprecipitation and next generation sequencing to yield a ~70-fold enrichment of modified transcripts (*68*). Another application for anti-modification antibodies has been developed in a protocol analogous to Western blotting, designated immuno-northern blotting, wherein anti-modification antibodies are used to assay for m^1A, m^6A, Ψ, and m^5C (*69*). In principle, the specificity of an antibody is an attractive basis for RNA modification detection. In practice, however, the biggest limitation of antibody-based detection approaches is the fidelity of the raised antibody to its epitope. As a single modified base presents a significantly smaller epitope surface than the 5-12 amino acid epitope recognized by anti-protein antibodies, issues of low affinity and cross reactivity have been reported for antibodies raised against certain modifications, including pseudouridine (*70-72*).

1.3.5–Cross linking and immuno-precipitation of RNA modification enzymes with their RNA substrates

In addition to detecting modifications in an RNA molecule, a complement of methods instead capture the enzyme responsible for modification. Cross linking and Immuno-Precipitation (CLIP) uses UV light exposure to covalently link proteins associated with an RNA *in vivo*. The cells are then lysed and subjected to partial RNase digestion such that any proteins cross-linked to RNA protect their RNA binding sites from complete

degradation and yield cross-linked RNA fragments of ~50 nt or smaller in a typical assay. The cross-linking procedure then allows for stringent purification of a protein-RNA complex of interest. A general purification workflow consists of immunoprecipitation with antibodies raised against an RNA-binding protein of interest or with antibodies raised against an epitope tag that has been transgenically introduced to the protein of interest. The immunoprecipitated protein-RNA complexes are then size-separated using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) before being transferred to a nitrocellulose membrane to remove any non-cross-linked RNA. The RNA fragments resulting from this purification are then isolated from their RNA binding proteins by digestion with proteinase K and subjected to RT-PCR and sequencing analysis, thus revealing the RNA sequence that was bound to the modification enzyme (*73*). Although the exact position of RNA modification cannot typically be identified with this technique, especially if the modifying enzyme does not use a strict sequence motif for substrate recognition and rather binds a structural element, CLIP can elucidate the repertoire of RNAs targeted by a specific modification enzyme while also providing an approximate region in the RNA where modification is likely to occur. A variation of this method, Photo-Activatable Ribonuceloside enhanced CLIP (PAR-CLIP) includes cell culturing in media containing 4-thiouridine (4SU), thereby resulting in 4SU incorporation into the cell's transcripts. The efficiency of subsequent cross-linking is thus enhanced by these 4SU-labeled RNAs.

1.3.6–Nanopore sequencing

A highly promising RNA modification detection method comes in the form of the nanopore device developed by Oxford Nanopore Technology (ONT). A nanopore consists of multiple channel proteins embedded in a synthetic membrane in a flow cell apparatus. Upon applying an electric potential to a nucleic acid sample loaded into the flow cell, RNA molecules are drawn towards the channel proteins and subsequently ratcheted through them at a consistent rate by motor proteins. As individual RNA molecules pass through a channel protein, each nucleotide causes the current across the nanopore to fluctuate and these changes are converted into an RNA sequence by a neural network (*74, 75*). In addition to identifying the four bases in RNA, nanopore current traces have been used to identify RNA modifications. To date, nanopore sequencing has been used to detect m^6A, m^5C, m^7G, pseudouridine, and A-to-I editing events (*74-77*). As nanopore technology allows for direct RNA sequence apprehension, this method avoids the artefacts generated by a RT-based sequencing approach, in which cDNA synthesis and subsequent PCR amplification can introduce bias to a data set. Moreover, whereas most other modification detection assays are configured to detect a single modification type (as in the case of RT-based detection of CMCT-labeled pseudouridine), nanopore sequencing can in principle detect the positions of different RNA modifications within a single RNA molecule. Nevertheless, RNA modifications detected through this direct sequencing approach are similar to those found through the other high throughput sequencing methods in that they should be considered strictly putative until validated by an independent method (*78*).

1.4–Occurrence of pseudouridine in mRNA

Through transcriptome-wide probing reported in 2014, pseudouridine has been found in the mRNAs of all eukaryotes investigated so far, including yeast, mice, various human cell lines, *Arabidopsis thaliana*, and the parasite *Toxoplasma gondii*. With these investigations, characteristics of the mRNA modification landscape have begun to be elucidated such as the total number of pseudouridines present across the mRNA transcriptome, the types of mRNAs modified, the location of pseudouridines in the targeted transcripts, and the pseudouridine synthases responsible for the modifications. Surprisingly, it has been discovered that within the mRNA transcriptomes of the analyzed organisms, distinct pseudouridylation topographies can be induced in response to various forms of stress and environmental stimuli. Moreover, both H/ACA snoRNPs and standalone pseudouridine synthases, which were previously believed to only target ncRNAs, have now been implicated in these mRNA pseudouridylation events. In the following section, the current understanding of mRNA pseudouridylation that has emerged from the various high throughput pseudouridine sequencing data sets will be discussed.

1.4.1–Yeast mRNA pseudouridylation

In yeast, three mapping investigations have been undertaken so far that employed variations of the high throughput CMCT/RT-based sequencing method to probe for pseudouridylation sites across the transcriptome (**Table 1**) (*6, 7, 48*). In the work of Lovejoy et al., their "PSI-seq" methodology was applied on the transcriptome of *S. cerevisiae* grown to log-phase (*48*). In two replicate experiments for pseudouridine detection in log phase yeast, the authors identified 103 unique pseudouridylation sites in 56 different mRNAs (replicate 1) and 335 unique pseudouridylation sites in 150 mRNAs

(replicate 2). Additionally, a heat-shocked sample of log-phase yeast yielded 335 sites across 208 mRNAs (**Table 1**). Consistently identified in all three data sets were two pseudouridylation sites: position 68 in the coding sequence (CDS) of *RPL11a* and position 239 in the CDS of *TEF1* and its paralog *TEF2*. *RPL11a* encodes a 60S ribosomal protein and *TEF1/2* encode translation elongation factors, suggesting that pseudouridylation could be targeting genes with related functional themes. Although gene ontology (GO) analysis performed on the putatively modified transcripts in the log-phase, and heat shock data sets did in fact show an enrichment in the category of cytoplasmic translation, it should be noted that pseudouridine-sequencing strategies are biased towards modification detection in highly abundant transcripts, which may confound any conclusions for mRNA pseudouridylation preference that can be drawn from GO analysis. In this investigation, the authors did not detect a bias for pseudouridine's position within a codon. Moreover, a preference was not observed for the location of pseudouridine within the transcripts, with sites present across the 5′ untranslated regions (UTRs), CDSs, and 3′ UTRs. Having identified mRNA pseudouridylation in *S. cerevisiae*, Lovejoy et al. searched additional yeast species for evolutionary conservation of Ψ68 in *RPL11a* and Ψ239 in *TEF1/2* mRNA. This inquiry revealed that Ψ68 in *RPL11a* is conserved in *S. mikitae* but not *S. pombe*, and that Ψ239 in *TEF1/2* was conserved in both *S. mikitae* and *S. pombe*. As the last common ancestor of *S. cerevisiae* and *S. pombe* is estimated to have lived 600 million years ago, the authors posited that the *TEF1/2* pseudouridylation is an ancient modification granting a hitherto undiscovered fitness advantage, thus alluding to a broader functional importance extending to other sites of mRNA pseudouridylation (*48*).

The findings of S. *cerevisiae* mRNA pseudouridylation by Lovejoy et al. are partially corroborated and expanded upon by the work of Carlile et al., in which they employed a similar pseudouridine sequencing strategy called "Pseudo-seq" to analyze mRNA pseudouridylation in budding yeast during post-diauxic growth (OD_{600} = 12) (**Table 1**) (*6*). The authors of this study reported 260 pseudouridine residues conservatively found across 238 mRNAs. These sites of modification were found in all regions of the transcripts, including the 5′ and 3′ UTRs as well as within the coding sequences. However, unlike in Lovejoy et al., where no preference for pseudouridine occurrence within the transcript was described, Carlile et al. noted an underrepresentation of pseudouridine within the 3′ UTR relative to the other regions of the mRNA. Additionally, GUA valine codons were found to be the most prevalently modified in their data set.

In the third pseudouridine mapping inquiry in yeast conducted by Schwartz et al., the authors first applied their pseudouridine sequencing methodology, called "Ψ-seq", to *S. cerevisiae* grown to mid-log phase (**Table 1**) (*7*). The resulting data set from this initial inquiry yielded 185 sites of pseudouridylation across the mRNA transcriptome, corroborating that pseudouridylation occurs in yeast mRNAs. Unlike the previously mentioned investigations, the authors did not analyze the pseudouridine content of the UTRs. Within the coding sequences of the modified transcripts, a bias in pseudouridine placement was not observed.

Table 1: mRNA pseudouridylation events reported in yeast

Reporting publication and high throuput sequencing methodology utilized	PSI-seq (*48*)	Pseudo-seq (*6*)	Ψ-seq (*7*)
Pseudouridines discovered in mRNA	103Ψs in 56 mRNA species (Log-phase experiment 1)	260Ψs in 238 mRNA species (post-diauxic growth)	185Ψs (30°C log-phase)
	335Ψs in 150 mRNA species (Log-phase experiment 2)		265Ψs (45°C heat-shock induced)
	335Ψs in 208 mRNA species (heat-shock)		
Cell types and growth conditions analyzed	log-phase yeast at 30°C (normal growth) and 45°C heat-shock	log-phase and post-diauxic yeast grown at 30°C	log-phase, growth saturated, 4°C cold-shocked and 45° heat-shocked yeast
Pseudouridine synthases identified as mRNA modifiers	Pus1, Pus4, Pus6	Pus1-4, 6,7, and 9	Pus1, Pus4, Pus7, H/ACA snoRNPs
Method used for pseudouridine validation	pseudouridine synthase gene knockout coupled with CMCT RT-stop assay with gene specific primers, *in vitro* activity assays with yeast cell extracts and purified standalone pseudouridine synthases	pseudouridine synthase gene knockout coupled with resequencing	pseudouridine synthase gene knockout/knockdowns coupled with resequencing

1.4.2–Human mRNA pseudouridylation

The transcriptome-wide study of pseudouridylation has been expanded from initial examination in yeast to inquiries in various human cell lines (**Table 2**). In the Carlile et al. investigation, Pseudo-seq was applied to HeLa cells in normal proliferation conditions and to cells 24-hours post serum starvation (**Table 2**). These experiments yielded 96 putative pseudouridylation sites in 89 human mRNAs, thus establishing the presence of mRNA pseudouridylation in higher eukaryotes. In the HeLa cell line, pseudouridine was found distributed throughout the identified transcripts, although a slight underrepresentation of

modification was found in the 5′ UTR (*6*). Analysis of human mRNA pseudouridylation was also undertaken by Schwartz et al. wherein they used Ψ-seq to analyze HEK293 cells and fibroblast RNAs, finding 353 pseudouridines in the mRNAs of these human cell lines (**Table 2**). In these transcripts, pseudouridine was evenly found throughout the UTRs and CDSs (*7*).

Finally, in a pseudouridine detection study conducted exclusively in mammals, Li et al. employed quantitative mass spectrometry analysis to determine the total content of pseudouridine in the mRNAs of various human cells lines (*47*). They found the Ψ/U ratio of total mRNA samples to range from ~0.20-0.40% depending on the cell line, suggesting mRNA pseudouridylation is more abundant than previously described by Schwartz et al (*7, 47*). Using their CeU-Seq methodology, the authors were able to use streptavidin pulldown to enrich the pseudouridylated mRNA population prior to high throughput sequencing. Consequently, the data set generated by CeU-seq was comparatively larger than those in the previously described pseudouridine mapping studies. Whereas the Ψ-seq inquiry of Schwartz et al. yielded 353 mRNA pseudouridylation sites in HEK293 cells, the pulldown-enrichment of modified targets in CeU-seq ultimately identified 1889 putative pseudouridylation sites in the same cell line (**Table 2**) (*7, 47*). DAVID database analysis of these pseudouridylated transcripts revealed enrichment in gene ontology terms for translation, protein metabolism, and DNA replication, suggesting that functionally related genes are perhaps targets of regulated pseudouridylation (*79*). In agreement with Carlile et al.'s finding that pseudouridine is underrepresented in the 5′ UTRs of HeLa cell mRNAs, here Li et al. note that pseudouridine is also found less abundantly in the 5′ UTRs of HEK293 cell mRNAs relative to the 3′ UTRs and CDSs (*6, 47*). Within the coding

sequences, the phenylalanine codons UUU and UUC were the most commonly modified, with the first two U positions targeted more than the third in the UUU codon. This stands in contrast to Carlile et al.'s observation of GUA valine codons being the most commonly targeted in yeast (*6*, *47*).

Table 2: mRNA pseudouridylation events reported in human cell lines

Reporting publication and high throuput sequencing methodology utilized	Pseudo-seq (*6*)	Ψ-seq (*7*)	CeU-seq (*47*)
Pseudouridines discovered in mRNA	96Ψs in 89 mRNAs	353Ψs	1889Ψs
Cell types and growth conditions analyzed	HeLa cells under normal proliferation conditions and 24 hrs post serum starvation	HEK293 cells and fibroblasts under normal growth conditions	HEK 293 cells under multiple degrees of heat shock, serum starvation, Poly (I:C), HGF, H202, and cyclohexamide treatment
Pseudouridine synthases identified as mRNA modifiers	n.d.	TRUB1 (Pus4 homolog), Pus7, H/ACA snoRNPs	Pus1, TRUB1 (Pus4 homolog), Pus7
Method used for pseudouridine validation	n.d.	recognition motif analysis and dyskerin knockdown coupled with resequencing	pseudouridine synthase knockdown, CRISPR/Cas9 deletion, and pseudouridine synthase overexpression coupled with resequencing; SCARLET for single-site validation; LC/MS/MS for Ψ quantification in total mRNA

1.4.3–*Toxoplasma gondii* and *Arabidopsis thaliana* mRNA pseudouridylation

In addition to the initial 2014 and 2015 mRNA pseudouridylation findings in yeast and humans, subsequent transcriptome-wide investigations have been conducted on the single-celled eukaryotic parasite *Toxoplasma gondii* and in the widely used plant model organism *Arabidopsis thaliana* (**Table 3**). In *T. gondii*, the PSI-seq protocol was adapted from Lovejoy et al. and used to search for pseudouridylation sites in the mRNAs of this parasite. Consequently, 1669 and 394 sites of mRNA pseudouridylation were respectively located in the tachyzoite and bradyzoite forms of *T. gondii*, thus increasing the repertoire of eukaryotic species observed to possess pseudouridines in mRNA (**Table 3**) (*80*). In *T. gondii*, pseudouridines were distributed throughout the 5′ and 3′ UTRs and CDSs, although 3′ UTR pseudouridylation was underrepresented, corresponding to the similar underrepresentation of 3′ UTR pseudouridylation reported by Carlile et al. for yeast mRNAs. Furthermore, within each region of the transcripts, the density of pseudouridylation was shown to be virtually constant and pseudouridines were not found to be clustered around start or stop codons or any other sites within each transcript region. Finally, with respect to codon position, uridine residues at position 1 were statistically determined to be pseudouridylated more frequently than uridines at positions 2 or 3, which supports the finding by Carlile et al. that in the phenylalanine UUU codon the first two positions are more commonly modified than the third. Therefore, in *T. gondii* mRNAs, pseudouridines were found to be distributed nonrandomly in the transcript and corroborate similar findings for mRNA pseudouridylation patterns in yeast.

In *A. thaliana*, the Pseudo-seq method developed by Carlile et al. was employed to detect mRNA pseudouridylation in fully expanded young leaf cells grown under normal

conditions (**Table 3**) (*53*). Using this approach, 451 pseudouridines were identified within 332 transcripts, which accounted for 1.21% of all mRNAs detected in the experiment. Similar to the findings for yeast and *T. gondii*, pseudouridylation was predominantly found in the 5′ UTR and CDS while being less abundant in the 3′ UTR. Furthermore, within *A. thaliana*, UUC, UCU, CUU, and UUU codons were the most frequently modified wherein the first uridine residue out of two consecutive uridine residues was pseudouridylated more often except for UUU, where the second position was the most commonly modified and the third position the least modified.

Table 3: mRNA pseudouridylation events reported in *T. gondii* and *A. thaliana*

Reporting publication and high throuput sequencing methodology utilized	PSI-seq (80)	Pseudo-seq (*53*)
Pseudouridines discovered in mRNA	1669Ψs (tachyzoite form) 394Ψs (bradyzoite form)	451Ψs
Cell types and growth conditions analyzed	parasite strains C32 (W.T. with GFP fused to a bradyzoite-specific *LDH2* promoter) and TBD5 (mutant C32 strain with insertion in *PUS1* gene) transfected into human foreskin fibroblast cells (HFFs)	young leaf samples from wild-type *A. thaliana* (WT, Col-0)
Pseudouridine synthases identified as mRNA modifiers	Pus1	Pus4, H/ACA snoRNPs
Method used for pseudouridine validation	insertional-inactivation of Pus1 coupled with resequencing	recognition motif analysis for pseudouridine synthase assigment; CMCT/RT coupled with qPCR for site-specific pseudouridine validation

1.4.4–Regulated mRNA pseudouridylation

Pseudouridylation of mRNAs has been suggested to be a dynamically regulated process, wherein modification sites are induced in response to different growth states or environmental stimuli (*7, 47*). However, unlike for other RNA modifications, thus far no evidence is available demonstrating the removal/erasing of pseudouridines from RNA besides degrading the entire modified RNA. In the pseudo-seq investigation of Carlile et al, evidence for regulated mRNA pseudouridylation was collected by comparing the modification topographies of log-phase and post-diauxic growth phase yeast. Therein, between the two growth phases, 42% of the modification sites identified in the post-diauxic growth condition were not detected in log-phase yeast. Furthermore, Ψ286 in *CDC33* mRNA (which was also identified as a pseudouridylation target by Lovejoy et al.) was found to be present in both growth conditions, yet more extensively modified during exponential growth. Of 150 mRNA pseudouridylation sites present in both log-phase and post-diauxic growth yeast, 62 sites exhibited a >2-fold change in relative stoichiometry between these two conditions. Overall, these findings support a growth-phase dependent pattern of regulated mRNA pseudouridylation (*6*). Notably, Carlile et al. also established the conservation of regulated mRNA pseudouridylation in mammalian cells using a HeLa cell model system: comparison of pseudo-seq analyses of this cell line grown in either normal proliferation or serum starvation conditions produced partially overlapping pseudouridylation topographies wherein modification sites unique to each condition were observed (*6*).

Regulated mRNA pseudouridylation in yeast is further attested to by the Ψ-seq analysis of Schwartz et al. wherein the authors compared the Ψ-seq profiles of log-phase,

growth saturated, cold shock, and heat shock exposed yeast. Although only a few mRNA pseudouridine modifications were acquired in the stationary phase and cold shock conditions, a wealth of 265 putative mRNA pseudouridylations were induced under 45°C heat-shock when compared to log-phase yeast cultured at 30°C. Some of these modifications occurred in heat-shock induced genes such as *HSP60*, *HSP104*, *MDJ1*, and *SSA4*, yet a greater amount still were induced in transcripts expressed at similar levels at both temperatures, once again suggesting differential pseudouridylation patterns in response to environmental stimulus as observed by Carlile et al. (*6, 7*).

Investigating the effect of various stimuli on human cells, Li et al. found through mass spectrometry analysis that heat shock, poly (I:C) treatment, and H_2O_2 exposure increased the pseudouridine content of the total mRNA population by ~40-50%, whereas other conditions such as starvation lowered the pseudouridine level ~10-15% (*47*). By analyzing the maps of mRNA pseudouridylation from cells either heat-shocked or exposed to oxidative stress in the form of H_2O_2, non-overlapping patterns of transcriptome pseudouridylation were observed. Of the induced pseudouridylation targets under H_2O_2 stress, transcripts encoding proteins involved in telomere maintenance and chromatin remodeling were among those targeted. Conversely, heat-shock exposure induced pseudouridylation in targets encoding proteins for transport and localization pathways (*47*). Of final note is that the pseudouridine profiles of mouse brain and liver were also analyzed by Li et al. using CeU-Seq to see if mRNA pseudouridylation exhibits tissue specificity. Indeed, the authors found that the brain and liver of mice possess tissue-specific transcripts that are modified including targeted genes involved in nervous system development and signal transduction for the brain, and protein transport and macromolecule localization for

the liver. Additionally, many genes expressed in both the brain and liver of mice were found to be selectively modified in only one of the two tissue types. Altogether, Li et al. corroborated and expanded upon the stimulus-specific patterns of mRNA pseudouridylation in human mRNAs and revealed the existence of tissue specificity for mRNA pseudouridylation targets in a mouse model system.

1.4.5–Pseudouridine synthases modifying mRNA

Having determined that mRNA pseudouridylation occurs in a variety of species ranging from parasites, to yeast, to plants, to humans and that distinct patterns of pseudouridylation are induced under certain conditions, the pseudouridine synthases responsible for these modifications have begun to be identified. In yeast, there is experimental evidence for mRNA pseudouridylation by 8 of the 10 pseudouridine synthases present in this organism (Pus1-4,5-7, and 9 and Cbf5) (**Table 1**). In the PSI-seq study of Lovejoy et al, the authors analyzed the sequences flanking two mRNA modification sites that they identified: site 68 in *RPL11a* and site 239 in *TEF1/2 (48)*. By doing so, the enzymes responsible for modification were identified based on the sequence motifs in which these targeted uridines occurred. In *RPL11a*, uridine 68 was found within a UCUGUU motif (where the underlined U residue is pseudouridylated). This recognition sequence is shared with uridine 44 in U2 snRNA, which is targeted by the standalone pseudouridine synthase Pus1p (*81*). The uridine at position 239 in *TEF1/2* was found to occur within a GUUCGA motif (where the underlined U residue is modified). This motif is also present in the TΨC loop of tRNAs, where the eponymous pseudouridine in this region is catalyzed by the standalone enzyme Pus4p (*82*). The assignment of Pus1p and Pus4p to the modification of *RPL11a* and *TEF1/2* mRNAs, respectively, was further

verified *in vivo* through knockout strains lacking genes for these pseudouridylases and *in vitro* through biochemical assays with cell extracts and purified pseudouridine synthases (*48*). Thus, the investigation by Lovejoy et al. suggests that, at least in the case of two target substrates, the pseudouridine synthases responsible for modification are not exclusive to mRNAs, and rather have multi-target specificities that include mRNAs in addition to non-coding RNA targets.

The majority of mRNA pseudouridylation targets found by Carlile et al.'s Pseudo-seq inquiry were assigned to standalone pseudouridine synthases by first generating viable yeast deletion strains each lacking one of eight non-essential pseudouridine synthase genes (Pus1-7 and 9) before performing Pseudo-seq analysis on these strains, thus assigning an enzyme to a specific site if its CMCT-Ψ RT signature present in a wild type yeast control sample was lost in a corresponding deletion strain. Using this methodology, Pus1, known to constitutively modify various tRNA positions and position 44 in U2 snRNA, was found to pseudouridylate the most mRNA substrates, with Pus2-4, 6, 7, and 9 also being found to recognize mRNA targets. Furthermore, some mRNA modification sites were putatively assigned to H/ACA snoRNPs based on perfect matches of their sequence context with known targets of snoRNA-guided pseudouridylation in ncRNAs. Intriguingly, many of the Pus1 sites were found to be modified to a greater extent in post-diauxic growth relative to log-phase growth, suggesting this enzyme's activity on mRNAs is regulated in a growth-state dependent manner. Moreover, through PSI-seq analysis of a Pus1 deficient *T. gondii* strain in the work of Nakamoto et al., this pseudouridine synthase was found responsible for 364 mRNA pseudouridines in the tachyzoite form and 33 mRNA pseudouridines in the bradyzoite form of the parasite (*80*).

In the Ψ-seq study of Schwartz et al, 41 mRNA sites were assigned to specific pseudouridine synthases by applying Ψ-seq to deletion strains of the eight nonessential Pus enzymes (vide supra) in addition to a Cbf5 knockdown strain. Of these 41 sites, 34 were thereby attributed to H/ACA snoRNPs with the remaining 7 assigned to either the Pus1p, Pus4p, or Pus7p standalone pseudouridine synthases. Furthermore, of the 265 heat-shock induced sites identified by Schwartz et al., 159 of them were attributed to Pus7p through knockout analysis of a *ΔPUS7* yeast strain in 30°C and 45°C conditions, suggesting that Pus7 is a critical protein in the heat-shock induced pseudouridylation response. These heat shock induced mRNA sites were found to occur in the established UGUAR recognition motif of Pus7 (where the underlined U is the modified residue and R = G > A) which is also present in U2 snRNA, cytoplasmic tRNAs, pre-tRNA^{Tyr}, and 5S rRNA (**Figure 3**) (*83*). Moreover, the modification extent at these Pus7p-induced sites was comparable to sites in rRNA, suggesting that heat-shock induced pseudouridylation in mRNA is present at high stoichiometries. Of the human mRNA modification sites ascertained from Ψ-seq analysis of HEK293 cells and fibroblasts, TRUB1 (human Pus4 homolog), Pus7, and dyskerin (Cbf5 in yeast) were responsible for the vast majority of them. These pseudouridine synthase assignments are supported in the *A. thaliana* investigation, wherein a variety of mRNA pseudouridines were found to occur within the recognition motifs of Pus4 and specific H/ACA snoRNPs,

Through knockdown and overexpression studies, Li et al. assigned many of the HEK293 mRNA modification sites identified through CeU-Seq to human Pus1, TRUB1, and human Pus7, thus corroborating the finding of Schwartz et al. that TRUB1 and Pus7 play a significant role in human mRNA pseudouridylation (*7, 47*). Indeed, the importance

of TRUB1 and Pus7 in mRNA pseudouridylation has been further attested to by computational analysis of human transcriptome pseudouridine maps wherein these two enzymes have been deemed responsible for ~60% of the high confidence modification sites identified in humans (*84*). In summation, the pseudouridine synthases that have been found responsible for modification of yeast, human, plant, and parasite mRNAs are enzymes that also commonly modify ncRNA species, and thus exhibit multi-substrate specificities. However, it is not fully understood how these enzymes are able to discriminate between constitutively modifying ncRNA substrates while being induced to modify specific mRNAs under certain conditions. For example, Pus1 constitutively modifies tRNAs yet is shown by Carlile et al. to increase the modification extent of mRNAs under post-diauxic growth (*6*). Because mRNA pseudouridylation patterns change in response to growth condition or cell state, the pseudouridine synthases responsible for these modifications are likely subject to a mode of regulation that at this time remains largely unknown. This topic will be addressed in the following section.

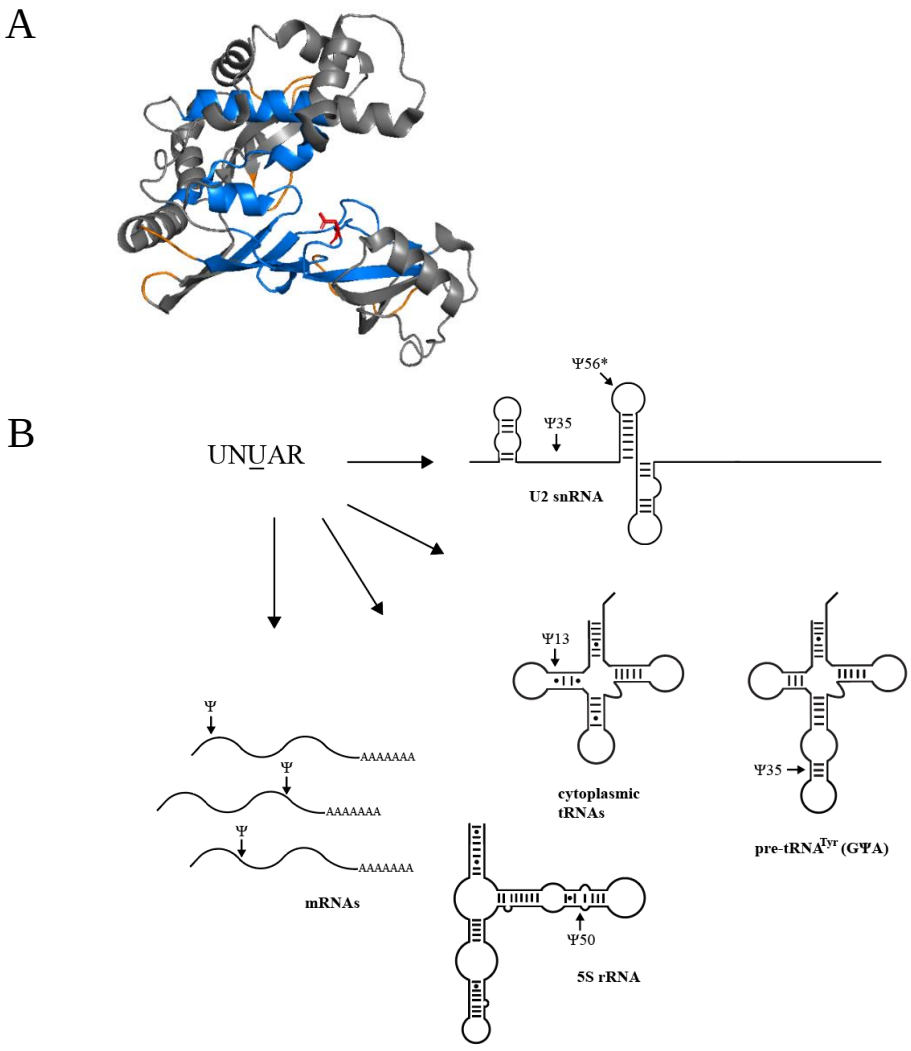


Figure 3. Structure of the Pus7 homolog TruD and the substrates of Pus7. (A) Crystal structure of the standalone pseudouridine synthase *E. coli* TruD which is homologous to Pus7 in yeast (PDB ID 1SB7; (*85*)). The blue-coloured regions of the cartoon representation correspond to sequence segments of TruD/Pus7 that are conserved across all phyla (*86*). The catalytic aspartate residue found in all pseudouridine synthases is highlighted in red and the orange loop regions are where sequence insertions are found in yeast Pus7. TruD/Pus7 is a U-shaped molecule consisting of two domains: the catalytic domain adopts a conserved fold common to other pseudouridine synthases and a second domain unique to TruD/Pus7 hinges over it, possibly functioning to clamp substrate RNAs and thereby position a target uridine within the enzyme's active site. **(B)** Pus7 has a UNUAR (where U is the target residue and R = G > A) consensus motif for substrate recognition and exhibits multi-substrate specificity: Pus7 is responsible for Ψ35 and Ψ56 formation in U2 snRNA, with the latter modification induced by nutrient deprivation and heat shock stress in yeast (denoted by an asterisk next to Ψ56 (*87*)), Ψ13 in cytoplasmic tRNAs, Ψ35 in pre-tRNA^{Tyr} (GΨA), Ψ50 in 5S rRNA in addition to modifying various mRNAs. Moreover, Pus7 mRNA modification has been shown to be upregulated in response to heat shock stress (*7*).

1.5–Emerging roles for pseudouridine in gene expression regulation

Because pseudouridylation of mRNA is generally conserved from yeast to humans and shown to be induced in response to environmental stimulus, it is tempting to postulate a hitherto unknown role for pseudouridine in gene expression regulation. Since pseudouridine confers structural stability to ncRNAs, it has been proposed that pseudouridine in mRNA could modulate the local secondary structure of the transcript, thus effecting interactions with RNA binding proteins and altering gene expression. A precedent for this phenomenon exists; m^6A-switches have been discovered in mRNA wherein this modification alters the accessibility of an RNA binding motif recognized by heterogeneous nuclear ribonucleoprotein C (hnRNP C), thus increasing the affinity of the protein. In turn, hnRNP C affects mRNA abundance and influences alternative splicing patterns, thus demonstrating a direct role for post-transcriptional RNA modification in gene expression (*88*). Furthermore, in yeast U2 snRNA, Ψ42 and Ψ44 have been found to increase the affinity and activity of the ATPase Prp5, an essential spliceosome assembly factor responsible for catalyzing a conformational change in the pre-splicing complex, possibly alluding to a similar factor recruitment role for pseudouridine in mRNA (*89*). Notably, in a recent study that applied high-throughput pseudouridine sequencing to the snRNAs in *Trypanosoma brucei*, pseudouridines were identified in snRNA domains responsible for RNA-RNA and RNA-protein interactions. Through follow up microfluidics analysis of synthesized *T. brucei* U2 snRNA that was either completely pseudouridylated or un-pseudouridylated, the binding affinities of U2 small nuclear ribonucleoprotein B″ (U2B″) and U2 small nuclear ribonucleoprotein A′ (U2A′) with these U2 snRNA variants were determined. Here, it was discovered that a temperature-dependent reduction in the

dissociation constants (K_d) was exhibited by these two proteins for fully pseudouridylated U2 snRNA at 37°C (but not 27°C), suggesting that pseudouridine can increase the binding affinity of these two proteins to U2 snRNA at an elevated temperature and may also have a similar effect on protein binding to mRNAs (*90*).

Pseudouridylation has been proposed to be irreversible, unlike m^6A which can be removed from mRNAs by eraser proteins (*91*, *92*). Therefore, regulated mRNA pseudouridylation is thought to arise mainly from the production or degradation of modified transcripts rather than through a hitherto undiscovered mechanism for reverting pseudouridine to uridine. Alternatively, hyper-modification of pseudouridine (by N1 methylation for instance) has been proposed as a mechanism that could modulate any RNA-RNA or RNA-protein interaction effects conferred by pseudouridine in mRNA, thus serving in place of a conventional eraser mechanism (*93*).

Within the coding sequences of mRNA, it has been suggested that pseudouridine could recode sense codons. This recoding potential has been attested to in an *in vitro* reconstituted *E. coli* translation system wherein pseudouridylated UUU phenylalanine codon variants were shown to direct the incorporation of different amino acids with reasonable efficiencies (*94*). The recoding potential of pseudouridine can also be found in echinoderm mitochondrial tRNA, wherein Ψ35 in the anticodon of mt $tRNA^{Asn}$ (GΨU) is responsible for decoding the AAA codon as asparagine instead of lysine as specified by the universal genetic code (*95*). In addition to recoding, sense codon pseudouridylation has been observed to alter the kinetics of decoding, with pseudouridine lowering the rate of translation elongation and EF-Tu GTPase activation (*94*). There is also evidence that pseudouridine can influence splicing: Pseudo-seq analysis of human pre-mRNAs found an

enrichment of pseudouridines near alternative splice sites and RNA binding protein sites (*96*). Furthermore, knockout of human Pus1 and depletion of the human Pus7 and Rpusd4 pseudouridine synthases results in widespread changes to alternative splicing patterns *in vivo* and single intronic pseudouridylation events have been demonstrated to enhance splicing *in vitro* (*96*). Intriguingly, pseudouridylation of stop codons in engineered systems has been demonstrated to cause translational read-through (vide infra) (*97*). Although there are currently no known stop codons targeted for pseudouridylation in nature, pseudouridine might possibly facilitate the readthrough of stop codons in a few instances that have yet to be discovered. Finally, pseudouridine could potentially alter mRNA stability, thus increasing the longevity of the transcript in the cell. The investigations of Schwartz et al. perhaps attested to this when they quantified the expression levels of genes identified as pseudouridylation targets for Pus7 in WT and *ΔPUS7* yeast. During heat shock, genes known to be targeted by Pus7 were expressed approximately 25% more in WT yeast than in *ΔPUS7* yeast, suggesting a stabilizing role for pseudouridine in mRNA (*7*). Altogether, a variety of theories have been offered to explain the functional significance of pseudouridine in mRNA, often drawing comparisons to the better understood m^6A, but as of this time pseudouridine remains a largely enigmatic mRNA modification.

In addition to the downstream consequences of pseudouridine in mRNA, the upstream regulatory networks that govern the enzymes responsible for these modifications have yet to be comprehensively elucidated (*98*). What is known is that many mRNA pseudouridine modifications are directed by the standalone pseudouridine synthases which were previously understood to only target ncRNAs. In yeast, there are 9 standalone pseudouridine synthases (Pus1-9) belonging to four families. Pus1-3 belong to the TruA

family, Pus4 belongs to the TruB family, and Pus5-9 are found in the RluA family. Of these 9 enzymes, all except for one (Pus8 which has only been demonstrated to target tRNAs) modify mRNAs.

To account for inducible mRNA pseudouridylation in response to growth state or environmental stress, the Pus enzymes are likely subject to a mode of regulation that could consist of altering enzyme expression, localization, interactions with other proteins, or post-translational modification (*98*). In regards to variable expression of enzyme, a yeast stress expression database reveals that Pus1, Pus3, Pus4, and Pus9 are differentially transcribed under different growth conditions (*99*). Therefore, modulation of enzyme abundance could possibly influence the mRNA modification extent. On the localization level, the 9 Pus enzymes are found within distinct organelles in the cell, with Pus1, 3, 4, 7, and 9 present in the nucleus, Pus3, 6, 7, 8, and 9 in the cytoplasm, and Pus2, 4, 5, 6, and 9 localized to mitochondria (*98*). This compartmentalization of enzymes thus offers spatiotemporal control over target RNA modification, which in turn could be used to exert control over gene expression. Notably, when studying heat-shock inducible mRNA pseudouridylation by Pus7 in yeast, it was discovered that the levels of Pus7 transcript and protein, respectively, were found to slightly decrease. However, a potential mechanism for heat-shock induced pseudouridylation was revealed by fluorescence microscopy of yeast harbouring GFP-tagged Pus7. Whereas the majority of Pus7 is localized to the nucleus at 30°C, at 45°C the bulk of this enzyme migrates to the cytoplasm where it presumably modifies its mRNA targets (*7*). In addition to considering the expression level and localization of the Pus enzymes, the interaction of these pseudouridine synthases with other proteins may possibly affect their function. A review of the 9 Pus enzymes has noted that

7 yeast proteins interact with more than one of the enzymes: Pho85, Tan1, SBP1, SBA1, NAB2, SSB2, and NAM7 (*98*). Furthermore, a plethora of other proteins associate with individual Pus enzymes (*100*). Thus, the Pus enzymes could possibly form complexes that alter their modifying activity. Pho85, for instance, is a cyclin-dependent kinase that participates in cell cycle regulation, raising the possibility that phosphorylation of its Pus enzyme binding partners could coordinate pseudouridylation to the cell cycle (*101*). Indeed, post-translational modification of the Pus enzymes is another means through which their activity could be regulated. Currently, Pus1, 2, and 7 are known to be ubiquitinated, Pus1 sumoylated, Pus1, 3, 4, 6, and 9 phosphorylated, and Pus7 succinylated (*98*). Of final note is that pseudouridine synthases may also regulate gene expression through mechanisms that do not depend on their modifying activity. For instance, human Pus10, which pseudouridylates cytoplasmic tRNAs, is also involved in micro RNA (miRNA) processing in the nucleus wherein Pus10 binds to pre-miRNAs and interacts with the DROSHA-DGCR8 microprocessor complex to promote miRNA biogenesis in a process independent of the enzyme's catalytic activity (*102*). Taken together, the mechanisms through which mRNA is inducibly pseudouridylated are largely speculative at this point in time, as are the downstream consequences of pseudouridine in mRNA. However, the exciting possibility exists that mRNA pseudouridylation constitutes a hitherto unknown mode of gene expression regulation occurring on the epi-transcriptomic level.

1.6–Nonsense suppression of stop codons

Protein synthesis is terminated when one of three canonical stop codons in mRNA (UGA, UAG, or UAA) is present in the ribosomal A site. When this occurs, instead of tRNA recruitment, eukaryotic release factors 1 and 3 (eRF1 and eRF3) are recruited to the ribosome in an eRF1·eRF3·GTP·Mg^{2+} complex whereupon peptidyl-tRNA hydrolysis is coupled to GTP hydrolysis, thus releasing the translated peptide (*103*). This termination process is not perfectly efficient, and in some cases, eRF1 and eRF3 must compete for stop codon recognition with a near cognate tRNA in which two of three anticodon positions base pair with the stop signal (*104*). Thus, a low level of natural nonsense suppression occurs during translation. When this happens, an amino acid residue is incorporated into the nascent peptide and translation continues past the stop signal until terminating downstream at the next available in frame stop codon. The resulting protein product therefore possesses a C-terminal extension.

1.6.1–Nonsense suppression of pseudouridylated stop codons

Surprisingly, it was discovered that pseudouridylation of the first U residue in each of the three stop codons significantly enhances nonsense suppression both *in vitro* and *in vivo* (*97*). Specifically, when assessing readthrough with an *in vitro* translation assay, a pseudouridylated UAA stop codon (ΨAA) was found to manifest a ~75% readthrough efficiency compared to an unmodified UAA signal, which had ~0.5% natural readthrough efficiency. Next, when a yeast reporter system was created wherein the pseudouridylation pocket of an H/ACA guide RNA was changed to target a premature stop codon within a reporter gene, *in vivo* stop codon readthrough could be detected by quantifying expression of the full-length reporter gene: ~6.4% of reporter gene expression was restored compared

to the ~0.1% expression level with a UAA stop codon not targeted by a H/ACA guide RNA. Mass spectrometry analysis of readthrough products from the reporter system revealed that ΨAA and ΨAG result in serine and threonine incorporation, whereas ΨGA results in tyrosine and phenylalanine incorporation (*97*). Notably, stop codon pseudouridylation has only been performed in engineered systems; there are currently no established instances of pseudouridylated stop codons in nature.

The structural basis for pseudouridylated stop codon readthrough has been investigated by determining the crystal structure of the anti-codon stem loop (ASL) of $tRNA^{Ser}$ (IGA anticodon) bound to a ΨAG codon in the A site of the bacterial 30S subunit (*105*). In this structure, an expected Ψ-A Watson-Crick base-pair is formed at the first position. At the second and third positions, respectively, unusual base-pairing between the Hoogsteen edges of the codon bases with the Watson-Crick edges of the anticodon bases allow normally forbidden A-G and G-I purine-purine base-pairs (**Figure 4**). Although this unusual base-pairing accounts for decoding of a pseudouridylated stop codon, it is still not understood how exactly pseudouridine facilitates the binding of a non-cognate tRNA, as the additional NH group of pseudouridine is exposed to the solvent and not observed to hydrogen bond with the ribosome or tRNA in the generated crystal structure. One hypothesis is that pseudouridine could stabilize the two noncanonical base pairs in the codon-anticodon helix formed by the pseudouridylated stop codon and the ASL of the non-cognate tRNA, thus facilitating nonsense suppression. However, further biochemical inquiries will be required to test this hypothesis (*105*). Further studies have investigated the mechanistic basis of readthrough, namely, if pseudouridylated stop codons just increase the efficiency of codon recognition by near-cognate tRNAs, or instead achieve readthrough

by also decreasing the efficiency of translation termination. To address this question, a crystal structure has been obtained of *T. thermophilus* 70S ribosome bound to ΨAA in the A site and $tRNA^{fMet}$ in the P site, in complex with *E. coli* release factor 1 (RF1) in the A site. This structure reveals that the conformations of RF1 domains II and III which are critical for stop codon recognition and peptidyl-tRNA bond hydrolysis, respectively, are virtually identical to those of RF1 in a canonical 70S termination complex (*106*). Furthermore, kinetic analysis of peptide release from a termination complex formed on both a pseudouridylated and unmodified stop codon demonstrates comparable rates of termination by RF1, thus showing that RF1 does not discriminate between pseudouridylated and non-pseudouridylated stop codons. Therefore, experimental evidence suggests that pseudouridine does not directly disrupt release factor function, and rather facilitates nonsense suppression via competition with release factors through enhancing stop codon recognition by near cognate tRNAs (*106*).

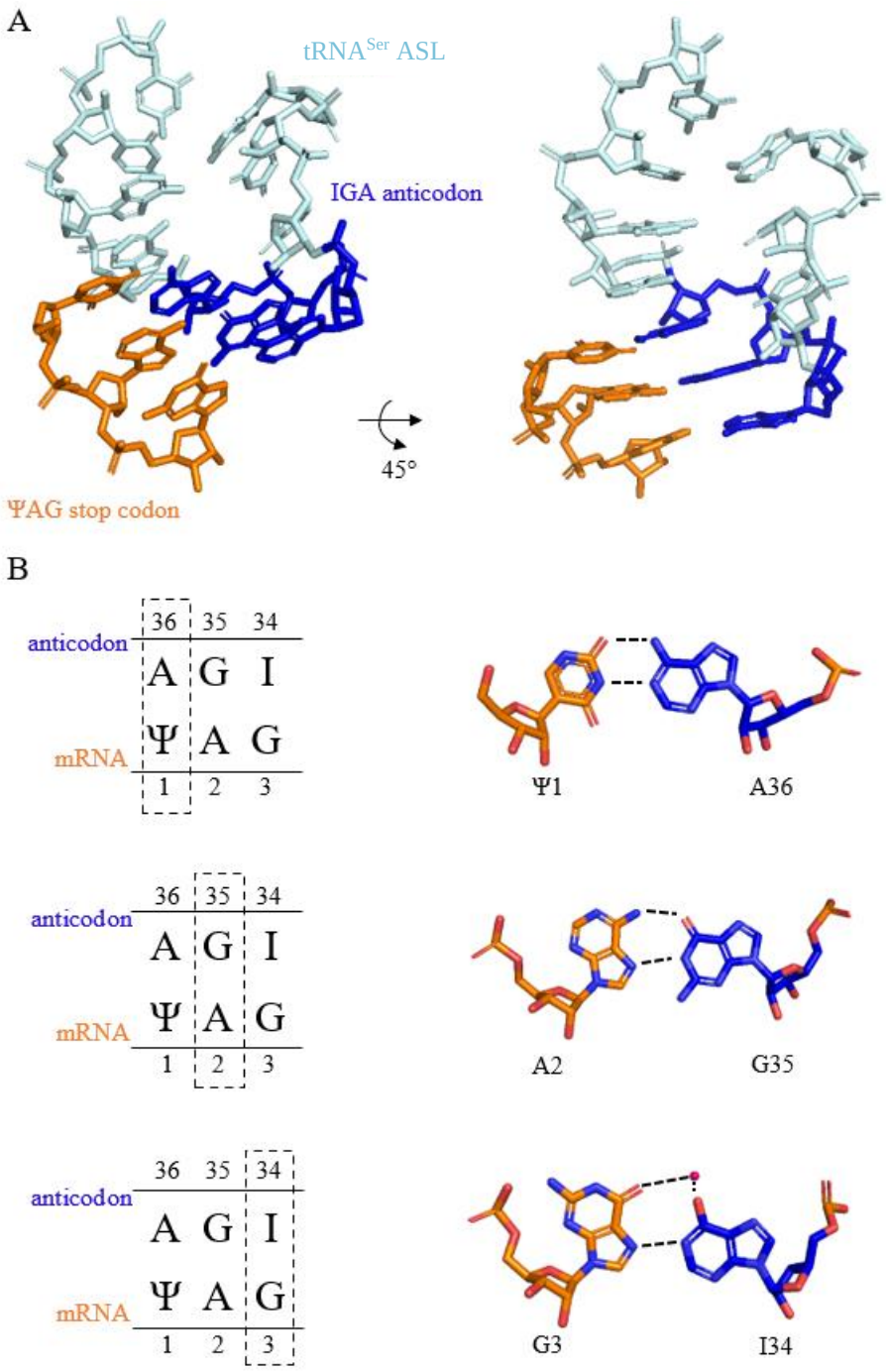


Figure 4. Pseudouridylated stop codon (ΨAG) interaction with the anticodon stem-loop of tRNA^{Ser} (IGA). (A) Stick representation of the anticodon stem loop (pale cyan) of tRNA^{Ser} (IGA anticodon, blue) base-pairing with a ΨAG stop codon (orange) within the decoding site of the *Thermus thermophilus* 30S ribosomal subunit (PDB ID 4JV5; (*105*)). **(B)** Individual Ψ-A, A-G, and G-I base-pairs formed at the first (top), second (middle), and third (bottom) positions of the anticodon-codon interaction shown in (A). The purple sphere participating in the G-I base-pair at the third position is a magnesium ion that alleviates electrostatic repulsion between the O6 atoms of G3 and I34.

During rare translational readthrough events of non-pseudouridylated stop codons (i.e. regular stop codons), the upstream and downstream sequence context of the stop signal has been found to drastically impact the efficiency of termination, with the nucleotides immediately 5′ and 3′ to the stop codon having a particularly important contribution (*107*, *108*). In contrast to these rare stop codon readthrough events, the sequence context of an in-frame pseudouridylated stop codon has been shown to have little to no impact on the efficiency of nonsense suppression (*109*).

1.6.2–Premature termination codons: genetic diseases and therapeutic treatments

Nonsense mutations occur when a single-nucleotide base change converts a sense codon into an in-frame premature termination codon (PTC). A meta-analysis of the Human Gene Mutation Database (HGMD) estimates that approximately 11% of all described gene lesions responsible for inherited human diseases arise from nonsense mutation (*110*). With respect to translation, a PTC in an mRNA can lead to a truncated, non-functional protein product, thus leading to a disease phenotype. Additionally, transcripts harbouring PTCs can be targeted by the nonsense mediated decay pathway (NMD), a eukaryotic quality control mechanism that degrades PTC-containing mRNAs, thus preventing expression of an aberrant gene product, leaving the cell without a functional protein (*111*). To remediate the effects of nonsense mutations causing genetic disease, nonsense suppression strategies have been developed as therapeutic treatments (*112*). For instance, aminoglycoside antibiotics have been shown to suppress pre-mature translation termination by binding to the ribosome decoding center, thus increasing the occurrence of misincorporation events during translation of PTC-containing transcripts (*113*). In patients, short term administration of the aminoglycoside gentamicin could restore a significant level of protein

expression in some cases, but toxicity associated with long term dosages and a tendency to preferentially bind mitochondrial ribosomes hinders the efficacy of this drug (*112*). Efforts are being made to advance nonsense suppression therapy, including synthesizing aminoglycoside derivatives with lower toxicity and conducting high-throughput screens of other compounds with promising nonsense suppression effects. Thus, from a clinical perspective, pseudouridine-based recoding of premature stop codons could constitute a novel therapeutic approach to remediate genetic disease (*104*). Specifically, the sequence of an H/ACA guide RNA's pseudouridylation pocket could be altered to target the uridine of an in-frame premature termination codon, thus facilitating nonsense suppression by reprogramming the cell's endogenous pseudouridylation machinery to recognize a novel substrate (**Figure 5**) (*114*).

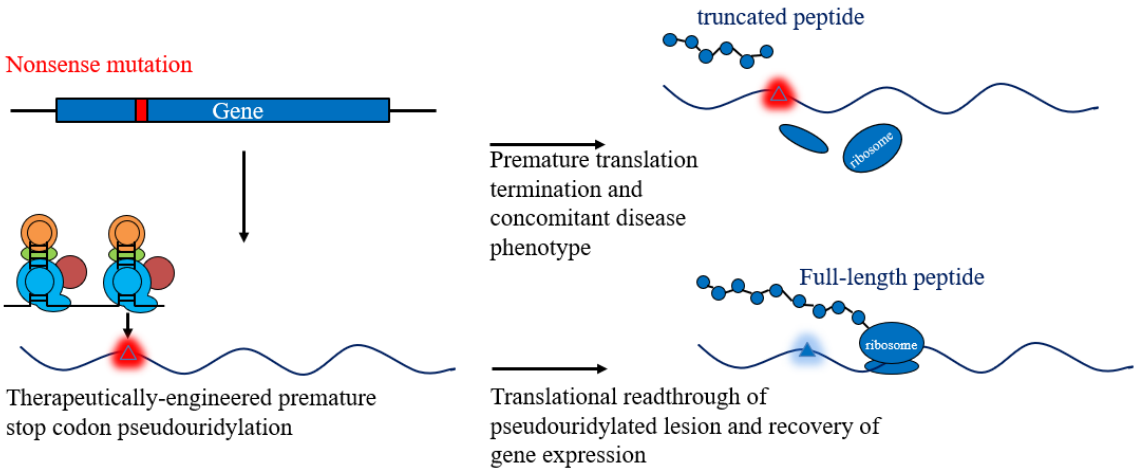


Figure 5. Proposed therapeutic application of targeted stop codon pseudouridylation. In genetic diseases arising from nonsense mutations, a pre-mature termination codon in a critical gene prevents translation of a full-length peptide. Instead, the premature stop signal will either produce a truncated peptide during translation or trigger the NMD pathway resulting in the transcript's degradation. Thus, a disease phenotype results due to the absence of a functional protein product. Administration of a customized H/ACA guide-RNA specifying pseudouridylation of the disease-causing premature stop codon in a target transcript has been proposed as a novel nonsense suppression therapy (*104*). Once pseudouridylated, the modified premature stop codon would facilitate translational readthrough, thus restoring expression of a full-length protein product and remediating the disease phenotype.

As described previously in sub-chapter 1.6.1, a proof of principle yeast reporter system has been developed in which a designer snR81 H/ACA guide RNA variant containing a customized pseudouridylation pocket was used to successfully target a premature stop codon in a reporter gene (*97*). Although artificially guided pseudouridylation is an exciting therapeutic prospect, a number of barriers must be overcome prior to clinical application, including the relative inefficiency of PTC modification observed in the custom guide RNA system (~4% of target transcripts were successfully modified) and the development of a delivery mechanism allowing for administration of a potent custom guide RNA dose in a patient's cells in such a way that does not trigger an innate immune system response (*115*).

Chapter 2–Research objectives and significance

The objective of my thesis is the development of a novel fluorescence-based reporter system in *S. cerevisiae* to allow for *in vivo* detection of pseudouridine synthase activity on mRNA. The motivation behind this research is four-fold. First, in what has been described as a "post mapping" period in post-transcriptional RNA modification research, a multiplicity of modification sites have been putatively identified across a host of transcripts, yet insight into the function of specific mRNA modifications, including pseudouridine, as well as knowledge pertaining to the regulation of the modifying enzymes remains forthcoming (*78*). Consequently, it is of great interest to create an RNA modification detection platform that is able to analyze the modifying activity of specific enzymes and how they behave under various cellular conditions. Second, the transcriptome-wide studies of pseudouridine occurrence using high-throughput sequencing have generated data sets exhibiting little overlap, in part due to stringent bioinformatic cut-offs, issues with sequencing depth, and the possibility of false-positive events (*50*). Therefore, pseudouridine sites found with high throughput sequencing analysis should be regarded with caution—it has been asserted that any sites identified in this manner be treated strictly as putative until verified with an independent method (*44*). With this in mind, the ability to probe specific sites of mRNA modification in a yeast model system would be a highly useful validation method for previously identified pseudouridine sites found through high throughput sequencing methodologies. Third, the mechanisms of substrate recognition by pseudouridine synthases are only partially understood. In nature, the number and continuity of base pairs between H/ACA guide RNAs and their substrates is variable, with internal mismatches between guide and substrate tolerated during

pseudouridylation (*116*, *117*). Therefore, predicting the substrate preferences of an H/ACA snoRNP *a priori* based solely on sequence complementarity between guide and substrate is often challenging. Moreover, the structural elements and sequence motifs used by many of the standalone pseudouridine synthases for substrate recognition are also incompletely defined (*6*). Thus, a reporter system allowing for *in vivo* detection of endogenous pseudouridine synthase activity would be ideally positioned to interrogate the substrate preferences of these enzymes. Fourth, the discovery that pseudouridylated stop codons facilitate nonsense suppression raises the exciting possibility of using directed pseudouridylation as a novel therapy for certain genetic diseases (*97, 104, 112*). Therefore, a reporter system in which fluorescence expression is directly linked to nonsense suppression could serve as an experimental tool to investigate the viability of pseudouridine-based reprogramming of stop codons as a therapeutic approach.

The fluorescence-based yeast reporter system developed in this thesis is adapted from a previous construct used to identify cis-regulatory elements in RNA—it consists of a dual reporter construct containing both red (RFP) and green fluorescent protein (GFP) genes under the control of a galactose inducible promoter (*118*). A premature termination codon (PTC) located upstream of GFP prevents translation of this reporter gene, resulting in a lone red signal upon flow cytometry analysis of the yeast system. Expression of GFP is restored by embedding the premature termination codon within the recognition sequence of a pseudouridine synthase, thus allowing for translational readthrough of this stop signal upon recoding by targeted pseudouridylation. Therefore, this yeast system detects the activity of endogenous pseudouridine synthases based on their ability to recognize and modify a substrate sequence inserted upstream of an EGFP reporter transcript. By changing

the recognition sequence surrounding the PTC, this system can be readily programmed to detect the activity of a wide variety of standalone pseudouridine synthases and H/ACA snoRNPs.

To test the viability of this novel assay to report on *in vivo* pseudouridylation events, prototype yeast reporter systems will be programmed to detect the modifying activities of standalone pseudouridine synthases and H/ACA snoRNP complexes. Furthermore, I plan to assess whether the system can detect differential pseudouridylation activity under different cellular stress conditions following reports of Pus7 being vital for the heat shock inducible mRNA pseudouridylation response (*7, 98*). Finally, in addition to pseudouridine-based reprogramming of stop codons, suppressor tRNAs containing anticodon mutations allowing them to decode stop codons will be used in the fluorescence-based yeast reporter system as a positive control for nonsense suppression (*119, 120*).

Chapter 3–Materials and Methods

3.1–Molecular biology in *E. coli*

Construction of the dual-fluorescing yeast system began with pFA6a-yomRuby2-pGal-EGFP plasmid (henceforth referred to as pRpG), which was cloned by inserting a yomRuby2-pGal gene fragment synthesized by Genewiz into a pFA6a-link-yoEGFP-CaURA3 plasmid (*121*). Cloning of the pRpG plasmid was performed by Genewiz and a map of the entire pRpG plasmid is given in Appendix 1. The pRpG plasmid was transformed via 42°C heat shock into *E. coli* DH5α prior to selection plating on LB agar containing ampicillin. A single colony from this plate was used to inoculate a 5 mL LB-ampicillin culture. After overnight shaking-incubation at 37°C, pRpG was miniprepped from this culture with the EZ-10 spin column plasmid DNA miniprep kit from BIO BASIC INC according to the manufacturer's instructions. From this plasmid template, pFA6a_link_Met15_homology forward and reverse primers described in **Table 4** were used to amplify the yomRuby2-pGal-EGFP-URA3 (RpG) sequence with *MET15* sequence homology arms on the 5′ and 3′ ends of the PCR product, thus allowing for homologous recombination of this reporter construct into the *MET15* locus of *S. cerevisiae* located on chromosome XII (*122*). For this amplification, a 50 µL PCR was formulated with final concentrations: 1× Q5 reaction buffer, 1 mM $MgCl_2$, 200 µM dNTPs, 0.5 µM of each respective primer, and template titrated between 24 and 75 ng. One Unit of Q5 DNA polymerase (NEB) was used for the reaction. Thermocycling consisted of initial 98°C denaturation for 30 seconds followed by 98°C denaturation for 10 seconds, 65.9°C annealing for 30 seconds, and 72°C extension for 2 minutes, 30 seconds. Twenty-five amplification cycles occurred in total before a final extension at 72°C for 2 min.

The presence of a single band between 4000 and 5000 bp (corresponding to an anticipated amplicon of 4336 bp) was checked by electrophoresing 5 µL of the completed PCR on a 1% agarose gel at 100 V for 1 hour prior to ethidium bromide staining and visualization with transillumination. From the remaining 45 µL of reaction mixture, the RpG amplicon was PCR purified using a BIO BASIC INC EZ-10 spin column DNA cleanup miniprep kit following the manufacturer's instructions. This PCR-purified RpG amplicon was ultimately eluted in 30 µL room temperature ddH_2O. The final purity and concentration of the DNA sample was determined via absorbance spectroscopy at 260 nm with a bio-drop instrument (Montreal Biotech Incorporated Lab Equipment).

Sequence and ligation-independent cloning (SLIC) was used to generate pRpG constructs containing sequences upstream of the EGFP coding sequence (*123*). The pRpG plasmid was restricted at its unique XhoI site (10 µg of plasmid digested in a 100 µL volume of 1× buffer R and 15 U XhoI from Fisher Scientific, incubated at 37°C for 2 hours) before DNA purification as previously described. T4 DNA polymerase-mediated 3′ chewback was undertaken by incubating 1 µg XhoI-digested pRpG plasmid with 0.3 U T4 DNA polymerase from NEB in a 20 µL 1× NEBuffer reaction under nucleotide starvation conditions. After 45 minutes of incubation at 37°C, 3′ chew-back from the XhoI cut site was halted by spike-in with 2 µL 10 mM dCTP and transfer of the reaction to an ice bath. The restricted and chewed-back vector was subsequently PCR-purified as previously described.

The oligonucleotides for insertion into pRpG were synthesized by IDT and are described in **Table 4**. Sense oligos were diluted to 100 nM prior to annealing with an identical amount of corresponding antisense oligo in a 10 µL mixture which was heated at

95°C for 5 minutes followed by cooling to room temperature. From the resultant double-stranded insert (with ~30 nt 5′ overhangs complementary to the chewed-back pRpG plasmid), a 10-fold dilution was prepared to yield a 5 nM solution of dsDNA ready for SLIC assembly with restricted and chewed-back pRpG plasmid. Each insert was assembled with 11.25 ng chewed-back pRpG plasmid (4:1 molar insert:plasmid ratio) within a 15 µL total volume of 1× Fermentas ligation buffer. The resultant SLIC assembly reactions were incubated at 37°C for 10 minutes.

Twenty µL high efficiency *E. coli* DH5α (NEB) were subsequently transformed via 42°C heat shock with 1.8 µL of each SLIC reaction, followed by selection plating on LB-Ampicillin media. After miniprep isolation of plasmid from transformed colonies and subsequent restriction screening with XhoI (successful SLIC-insertion is expected to abolish pRpG's unique XhoI cut site), proper insertion of each oligo-pair into a finished pRpG construct was confirmed through Sanger sequencing conducted by GENEWIZ, using the SLIC_RpG_confirm_seq primer specified in **Table 4**. All pRpG plasmid variants generated for this research are detailed in **Table 5.** For yeast transformation, *MET15* homology arm-containing RpG amplicons with sequence-insert upstream of the *EGFP* CDS were prepared in PCR reactions as previously described and checked by 1% agarose gel electrophoresis prior to PCR purification of the remaining reaction mixture.

Two constructs, designated pRS327-Sup53 and pRS327-Sup4, were created to express suppressor tRNAs (**Table 5**). Synthesized by IDT, suppressor tRNA genes *SUP53* and *SUP4* (described in **Table 4**) were subcloned into SmaI-restricted pUC19 via blunt-end ligation. Using M13 forward and reverse primers, *SUP53* and *SUP4* were amplified from pUC19-Sup53 and pUC19-Sup4. After DpnI digestion and subsequent PCR

purification, the *SUP53* and *SUP4* amplicons were EcoRI digested and ligated into 50 ng pRS327 at its unique EcoRI site at a 10:1 mol excess over pRS327 (*124*). These constructs were Sanger-sequence verified using the M13R universal primer.

Table 4. Oligonucleotides used in this study.

Oligo Name	**5′-3′ Sequence**	**Description**
pFA6a_link_Met15_Hom_Fwd	TAA CAT AAT AAC CGA AGT GTC GAA AAG GTG GCA CCT TGT CCA ATT GAA CAC GCT CGA TGA AAT GTA TCA TAC ACA TAC GAT TTA GG	Forward primer used to amplify Ruby-pGal-EGFP from its pFA6a_link plasmid backbone. Underlined region denotes 59 nt of sequence homology to a region upstream of the MET15 locus, situated on chromosome XII in *S. cerevisiae*
pFA6a_link_Met15_Hom_Rev	TTT TAT AAA AGT ATA GTA CTT GTG AGA GAA AGT AGG TTT ATA CAT AAT TTT ACA ACT CAT CGA TGA ATT CGA GCT CG	Reverse primer used to amplify Ruby-pGal-EGFP from its pFA6a_link backbone. Underlined region denotes 59 nt of sequence homology to a region downstream of the MET15 locus situated on chromosome XII in *S. cerevisiae*
Pus7_Sense_Test	TAA GTT TTA ATT ACA AAA TGG GTC TCG AAG AAA AGA TAC TGC TGT AGA AGA ATG ATT CTG	Sense oligo paired with Pus7_Antisense_Test for SLIC assembly
Pus7_Antisense_Test	AAC AAA ACG TTA ATG TGA GAC TTC TCG ACA GAA TCA TTC TTC TAC AGC AGT ATC TTT TCT	Antisense oligo paired with Pus7_Sense_Test for SLIC assembly

Pus7-U2-Test sense	ATC TAA GTT TTA ATT ACA AAA TGG GTC TCG ATC AAG TGT AGT ATC TGT TCT TTT CAG TGT TAC AAC TGA AAT GGC CTC AAC GAG GCT CAT TAC CTT TTA ATT TGT	Sense oligo paired with Pus7-U2-Test antisense for SLIC assembly
Pus7-U2-Test antisense	TCA ACA AAA CGT TAA TGT GAG ACT TCT CGA ACA AAT TAA AAG GTA ATG AGC CTC GTT GAG GCC ATT TCA GTT GTA ACA CTG AAA AGA ACA GAT ACT ACA CTT GA	Antisense oligo paired with Pus7-U2-Test sense for SLIC assembly
PUS7_Sense_ Ser_Positive	TAA GTT TTA ATT ACA AAA TGG GTC TCG AAG AAA AGA TAC TGC TGT CGA AGA ATG ATT CTG	Sense oligo paired with Pus7_ Antisense_Ser_Positive for SLIC assembly
PUS7_Antise nse_Ser_Posit ive	AAC AAA ACG TTA ATG TGA GAC TTC TCG ACA GAA TCA TTC TTC GAC AGC AGT ATC TTT TCT	Antisense oligo paired with Pus7_Sense_Ser_Positive for SLIC assembly
PUS7_Sense_ Negative	TAA GTT TTA ATT ACA AAA TGG GTC TCG AAG AAA AGA TAC TGC TCT AGA AGA ATG ATT CTG	Sense oligo paired with Pus7_Antisense_Negative for SLIC assembly
PUS7_Antise nse_Negative	AAC AAA ACG TTA ATG TGA GAC TTC TCG ACA GAA TCA TTC TTC TAG AGC AGT ATC TTT TCT	Antisense oligo paired with Pus7_Sense_Negative for SLIC assembly
snR5_Sense_ Test	TAA GTT TTA ATT ACA AAA TGG GTC TCG AAG AAA GCG AAT GAT TAG AGG TTC CGG GGT CGG	Sense oligo paired with snR5_Antisense_Testative for SLIC assembly
snR5_Antisen se_Test	AAC AAA ACG TTA ATG TGA GAC TTC TCG ACC GAC CCC GGA ACC TCT AAT CAT TCG CTT TCT	Antisense oligo paired with snR5_Sense_Test for SLIC assembly
snR5_Sense_ Tyr_Positive	TAA GTT TTA ATT ACA AAA TGG GTC TCG AAG AAA GCG AAT ACT TAG AGG TTC CGG GGT CGG	Sense oligo paired with snR5_Antisense_Tyr_Posit ive for SLIC assembly

snR5_Antisense_Tyr_Positive	AAC AAA ACG TTA ATG TGA GAC TTC TCG ACC GAC CCC GGA ACC TCT AAG TAT TCG CTT TCT	Antisense oligo paired with snR5_Sense_Tyr_Positive for SLIC assembly
snR5_Sense_Negative	TAA GTT TTA ATT ACA AAA TGG GTC TCG AAG AAA GTG AGT GAT TGG AGG TTC CGG GGT CGG	Sense oligo paired with snR5_Antisense_Negative for SLIC assembly
snR5_Antisense_Negative	AAC AAA ACG TTA ATG TGA GAC TTC TCG ACC GAC CCC GGA ACC TCC AAT CTC TCA CTT TCT	Antisense oligo paired with snR5_Sense_Negative for SLIC assembly
snR81_Sense_Test	TAA GTT TTA ATT ACA AAA TGG GTC TCG AAG TCA AAC TTT AAA TAT GTA AGA AGT CCT TGG	Sense oligo paired with snR81_Antisense_Test for SLIC assembly
snR81_Antisense_Test	AAC AAA ACG TTA ATG TGA GAC TTC TCG ACC AAG GAC TTC TTA CAT ATT TAA AGT TTG ACT	Antisense oligo paired with snR81_Sense_Test for SLIC assembly
snR81_Sense_Ser_Positive	TAA GTT TTA ATT ACA AAA TGG GTC TCG AAG TCA AAC TTT CAA TAT GTA AGA AGT CCT TGG	Sense oligo paired with snR81_Antisense_Ser_Positive for SLIC assembly
snR81_Antisense_Ser_Postive	AAC AAA ACG TTA ATG TGA GAC TTC TCG ACC AAG GAC TTC TTA CAT ATT GAA AGT TTG ACT	Antisense oligo paired with snR81_Sense_Ser_Positive for SLIC assembly
snR81_Sense_Negative	TAA GTT TTA ATT ACA AAA TGG GTC TCG AAG TCA AGC TCT AAA TCT GCA AGA AGT CCT TGG	Sense oligo paired with snR81_Antisense_Negative for SLIC assembly
snR81_Antisense_Negative	AAC AAA ACG TTA ATG TGA GAC TTC TCG ACC AAG GAC TTC TTG CAG ATT TAG AGC TTG ACT	Antisense oligo paired with snR81_Sense_Negative for SLIC assembly
SLIC_RpG_confirm_seq	GCT TTT ATG GTT ATG AAG AGG	Sequencing primer for validation of all Ruby-pGal-EGFP construct variants created with SLIC

UPF1_KO_LEU2_F	GAC CGA ATA TAC TTT TTA TAT TAC ATC AAT CAT TGT CAT TAT CAA ATG TCT GCC CCT AAG AAG ATC	Forward primer used to amplify the LEU2 marker from its pFL46L backbone. Underlined region denotes 45 nt of sequence homology to a region upstream of the UPF1 locus situated on chromosome XIII in *S. cerevisiae*
UPF1_KO_LEU2_R	ATC ACA AGC CAA GTT TAA CAT TTT ATT TTA ACA GGG TTC ACC GAA TTA AGC AAG GAT TTT CTT AAC TTC TTC GGC	Reverse primer used to amplify the LEU2 marker from its pFL46L backbone. Underlined region denotes 45 nt of sequence homology to a region downstream of the UPF1 locus situated on chromosome XIII in *S. cerevisiae*
UPF1_KO_Confirm_F	CAG GAA AGA AGG AAG GGC	Forward primer designed to anneal upstream of the UPF1 locus situated on chromosome XIII in *S. cerevisiae*
UPF1_KO_Confirm_R	CAT TTC ACG GTT GAG CCG	Reverse primer designed to anneal downstream of the UPF1 locus situated on chromosome XIII is *S. cerevisiae*
SUP53 tL-Caa-C amber suppressor	tgcatgatctacgtgcgtcacatgcagtacGAATT CCAATTGTCCTGTACTTCCTTGTT CATGTGTGTTCAAAAACGTTATAT TTATAGGATAATTATACTCTATTT CTCAACAAGTAATTGGTTGTTTGG CCGAGCGGTCTAAGGCGCCTGAT TCTAGAAATATCTTGACCGCAGTT AACTGTGGGAATACTCAGGTATC GTAAGATGCAAGAGTTCGAATCT CTTAGCAACCATTATTTTTTTCCTC AACATAACGAGAACACAGAATTC cactagctcagattcagtagaccgctgttg	G-Block encoding SUP53 amber suppressor-tRNA. Lower case letters denote adaptor sequences necessary for gene synthesis yet ultimately removed by EcoRI restriction.

SUP4 tY-GUA-J2 ochre suppressor	tgcatgatctacgtgcgtcacatgcagtacGAATT CGTATATGTGTTATGTAGTATACT CTTTCTTCAACAATTAAATACTCT CGGTAGCCAAGTTGGTTTAAGGC GCAAGACTTTAATTTATCACTACG AAATCTTGAGATCGGGCGTTCGA CTCGCCCCCGGGAGATTTTTTTGT TTTTTATGTCTCCATTCACTTCCCA GACTTGCAAGTTGAAATATTTCTT TCAAGCTCTTAACGAAGAGGAAT TCcactagctcagattcagtagaccgctgttg	G-block encoding SUP4 ochre suppressor-tRNA. Lower case letters denote adaptor sequences necessary for gene synthesis yet ultimately removed by EcoRI restriction.

Table 5: Plasmids used in this study. Plasmids created in this study are denoted with an Asterix (*) whereas the source of other plasmids is indicated.

Plasmid	**Description**	**Oligonucleotides used for creation (see Table 4)**
pFA6a_yomRuby2_pGal_yoEGFP (pRpG) *	GFP-Ruby2 reporter, no insert upstream of GFP. URA3 and ampR markers.	-
pRpG-Pus7-Test *	pRpG variant with a stop codon-containing Pus7 recognition sequence upstream of GFP. URA3 and ampR markers.	Pus7_Sense_Test and Pus7 Antisense_Test (for SLIC insertion upstream of GFP)
pRpG-Pus7-Pos *	pRpG variant without a stop codon in the Pus7 recognition sequence upstream of GFP. URA3 and ampR markers.	PUS7_Sense_Ser_Positive and PUS7_Antisense_ Ser_Positive (for SLIC insertion upstream of GFP)

pRpG-Pus7-Neg *	pRpG variant with a mutated Pus7 recognition sequence upstream of GFP. URA3 and ampR markers.	PUS7_Sense_Negative and PUS7_Antisense_Negative (for SLIC insertion upstream of GFP)
pRpG-Pus7-U2-Test *	pRpG variant with a U2 snRNA Pus7 recognition element upstream of GFP. URA3 and ampR markers.	Pus7-U2-Test-sense and Pus7-U2-Test-antisense (for SLIC insertion upstream of GFP)
pRpG-snR5-Test *	pRpG variant with a stop codon-containing snR5 H/ACA snoRNP recognition sequence upstream of GFP. URA3 and ampR markers.	snR5_Sense_Test and snR5_Antisense_Test (for SLIC insertion upstream of GFP)
pRpG-snR5-Pos *	pRpG variant without a stop codon in the snR5 H/ACA snoRNP recognition sequence upstream of GFP. URA3 and ampR markers.	snR5_Sense_Tyr_Positive and snR5_Antisense_Tyr_ positive (for SLIC insertion upstream of GFP)
pRpG-snR5-Neg *	pRpG variant with a mutated snR5 H/ACA snoRNP recognition sequence upstream of GFP. URA3 and ampR markers.	snR5_Sense_Negative and snR5_Antisense_Negative (for SLIC insertion upstream of GFP)
pRpG-snR81-Test *	pRpG variant with a stop codon-containing snR81 H/ACA snoRNP recognition sequence upstream of GFP. URA3 and ampR markers.	snR81_Sense_Test and snR81_Antisense_Test (for SLIC insertion upstream of GFP)
pRpG-snR81-Pos *	pRpG variant without a stop codon in the snR81	snR81_Sense_Ser_Positive and snR81_Antisense_Ser_

	H/ACA snoRNP recognition sequence upstream of GFP. URA3 and ampR markers.	Positive (for SLIC insertion upstream of GFP)
pRpG-snR81-Neg *	pRpG variant with a mutated snR81 H/ACA snoRNP recognition sequence upstream of GFP. URA3 and ampR markers.	snR81_Sense_Negative and snR81_Antisense_ Negative (for SLIC insertion upstream of GFP)
pFL46L	From Bonneaud et al., 1991. Source of LEU2 selective marker. Also contains an ampR marker	UPF1_KO_LEU2_F and UPF1_KO_LEU2_R (for amplification of the LEU2 selective marker out of pFL46L
pRS327-Sup4 *	Sup4 expression construct. LYS2 and ampR markers.	SUP4 tY-GUA-J2 ochre suppressor (synthesized Sup4 gene)
pRS327-Sup53 *	Sup53 expression construct. LYS2 and ampR markers.	SUP53 tL-Caa-C amber suppressor (synthesized Sup53 gene)

3.2–Preparation of competent yeast cells for transformation

Haploid BY4742 *S. cerevisiae* supplied by Dharmacon (subsidiary of Horizon Inspired Cell Solutions) was made competent for transformation by first inoculating a 5 mL YPD culture (2% w/v glucose) with a single colony. After 30°C incubation with shaking overnight, cell concentration of this culture was determined via hemacytometric count using an Olympus CH-2 compound light microscope (CAPSEN Medical & Scientific Co. Ltd.). A volume of overnight culture containing 2.5×10^8 yeast cells was combined with YPD media in a 50 mL final volume (2% w/v glucose). The growth of this culture at 30°C with shaking was monitored via hemacytometer cell-count: at a concentration of 2×10^7 cells/mL, the yeast culture was harvested by centrifugation at 3000

× g for 5 minutes in a 4°C pre-cooled Sorval Legend XTR centrifuge (Thermo Scientific). The YPD supernatant was decanted after initial pelleting. The harvested yeast were washed once in 25 mL sterile ddH_2O prior to another round of centrifugation at 3000 × g for 5 minutes, discarding the wash water after this step. Sorbitol buffer (1 M sorbitol, 100 mM LiOAc, 10 mM Tris-HCl pH 8.0, and 1 mM EDTA pH 8.0) was used to wash the yeast cells once (by cell resuspension in a 10 mL sorbitol buffer volume, centrifugation at 3000 × g for 5 minutes followed by supernatant discard) before final pellet resuspension in 360 µL sorbitol buffer and 40 µL carrier DNA (10 mg salmon-sperm DNA in 1 mL ddH_2O, boiled for 10 min at 100°C, allowed to cool to room temperature before combination with yeast). The resulting, competent BY4742 yeast cells were split into 50 µL aliquots and stored at -80°C without liquid nitrogen shock freezing.

3.3–Creation of a ΔUPF1 yeast strain

Competent BY4742 yeast cells were used to generate a *ΔUPF1* yeast strain via homologous recombination of a *LEU2* marker into the *UPF1* locus. *LEU2* was amplified from pFL46L kindly gifted by L. Minvielle-Sebastia and described in Bonneaud et al., 1991 (**Table 5**) (*125*). Primers flanking the *LEU2* marker in pFL46L (UPF1_KO_LEU2_F and UPF1_KO_LEU2_R described in **Table 4**) were used in a 50 µL PCR reaction consisting of 1× Q5 reaction buffer, 1 mM $MgCl_2$, 0.2 mM dNTPs, 0.5 µM of respective forward and reverse primer, 24, 50, or 75 ng pFL46L plasmid, and 1 U Q5 DNA polymerase with cycling conditions of 98°C initial denaturation for 30 seconds, followed by 98°C denaturation for 10 seconds, 65°C annealing for 30 seconds, and 72°C extension for 45 seconds. Twenty-five amplification cycles occurred in total before 72°C final extension for 2 min. The presence of a single band between 1000 and 1500 bp (corresponding to an

anticipated *LEU2* fragment size of 1094 bp) was checked by electrophoresing a 5 µL sample of the completed PCRs on a 1% agarose gel at 100V for 1 hour prior to ethidium bromide staining and visualization with transillumination. The amplicons from 3 respective 50 µL reactions (containing 24, 50, or 75 ng template) were pooled and PCR purified into a 50 µL ddH_2O final volume.

Following a yeast transformation protocol adapted from Knop et al., 8 µL of PCR-purified *LEU2* marker was mixed into 50 µL competent BY4742 yeast thawed on ice prior to transformation (*126*). In parallel, 8 µL ddH_2O was mixed with two respective 50 µL competent BY4742 yeast aliquots as cell viability controls. Each sample was mixed with a 6-fold volume of PEG solution (40% w/v polyethylene glycol molecular weight 3350, 100 mM LiOAc, 10 mM Tris HCl pH 8.0, and 1 mM EDTA pH 8.0) before vortex mixing and room temperature incubation for 1 hour, 15 minutes. A 1/9 volume of DMSO was added to each sample prior to heat shock. During this heat shock procedure, the *LEU2* transformation sample and one of the ddH_2O-transformated controls were subjected to 42°C heat shock for 15 minutes while the other ddH_2O-transformated control was left at room temperature for the duration of the heat shock treatment as a no heat-shock control. After cell pelleting with an IEC MICROCL 17 centrifuge (Thermo electron corporation), both the *LEU2* transformant and the mock-transformed heat shock controls were resuspended in 200 µL S-broth (synthetic media of yeast nitrogen base lacking amino acids). All 200 µL of the transformation mixture was spread on SD-LEU medium, and the +/- heat shock controls were first diluted 10 000-fold in S-broth before spread plating 200 µL of these 10^{-4} dilutions on solid YPD medium. These plates were incubated for 48 hours at 30°C. Cell viability was calculated from the quotient of colony forming units on the +/-

heat shock control plates. From the SD-LEU transformation plates, 5 mL SD-LEU liquid medium (2% w/v glucose) was inoculated with single colonies and incubated overnight at 30°C with shaking. The resultant cultures were used for *ΔUPF1* yeast glycerol stock preservation.

To validate the putative *ΔUPF1* strain, transformants from the aforementioned SD-LEU plate were re-streaked on new SD-LEU medium and incubated for 48 hours at 30°C to ensure stable integration of the *LEU2* marker into the *UPF1* locus on yeast chromosome XIII. Confirming growth for a second time on the SD-LEU selective medium, single colonies from the re-streak were inoculated into respective 5 mL SD-LEU cultures (2% w/v glucose) and incubated at 30°C for a 48-hour period with shaking. Assessing cell morphology and sample purity via hemacytometer, the cultures were grown to a ~1.5×10^8 cell density. Genomic DNA was extracted from a culture volume containing 2×10^8 yeast cells with a Presto Mini gDNA yeast kit (Geneaid) by following the manufacturer's instructions. Final concentrations and purities of the extracted gDNA from each sample were assessed with biodrop analysis. Using 250 ng of yeast gDNA, primers flanking the *UPF1* locus (UPF1_KO_Confirm_F and UPF1_KO_Confirm_R described in **Table 4**) were used in 50 µL PCR reactions consisting of 1× Q5 reaction buffer, 1 mM $MgCl_2$, 0.2 mM dNTPs, 0.5 µM of respective forward and reverse primer, and 1 U Q5 DNA polymerase. Thermocycling conditions were as follows: 98°C initial denaturation for 30 seconds, 98°C denaturation for 10 seconds, 64°C annealing for 30 seconds, and 72°C extension for 90 seconds with 35 amplification cycles in total before 72°C final extension for 2 min. The presence or absence of *LEU2* marker within the *UPF1* locus on Chromosome XIII of the putative *ΔUPF1* yeast strains was determined by fragment length

polymorphism: 5 µL PCR samples were electrophoresed on a 1% agarose gel at 115 V for 45 minutes prior to EtBr staining and transillumination–a single band present at ~3000 bp was interpreted as intact *UPF1* (expected 3053 bp amplicon) whereas a single band between 1000-1500 bp was interpreted as the strain having the smaller *LEU2* marker present at this locus (expected 1231 bp amplicon), the latter outcome thus validated as having an authentic *ΔUPF1* genotype.

3.4–Establishment of a dual-fluorescing stop codon-readthrough system in *S. cerevisiae*

Using the RpG PCR products described in sub-chapter 3.1, along with competent WT BY4742 and ΔUPF1 yeast cell aliquots prepared with the method in sub-chapter 3.2, dual-fluorescing stop codon-readthrough systems were created in *S. cerevisiae* via the yeast transformation procedure described in sub-chapter 3.3. To assess the effect of nonsense mediated decay on the function of the reporter system, RpG constructs were transformed into both WT BY4742 and ΔUPF1 yeast. As a *URA3* marker was used for RpG selection, and a *LEU2* marker was previously inserted into *UPF1* on Chr. XIII of the ΔUPF1 strain, selection plating on SD-LEU-URA medium was utilized for transformation into the ΔUPF1 strain. Selection plating on SD-URA was utilized for the reporter constructs transformed into WT BY4742 yeast. All yeast strains created for this study are described in **Table 6**. Glycerol stocks of all recombinant yeast strains were made and stored at -80°C. Within the limited scope of this M.Sc. thesis, only single biological replicates of each described yeast strain were prepared.

Using the yeast transformation protocol described previously, the pRS327-Sup53 and pRS327-Sup4 plasmids were respectively transformed into the PUS7-U2-Test-ΔUPF1

and snR81-Test-ΔUPF1 strains created in this thesis, thus creating the Pus7-U2-Test-Sup53 and snR81-Test-Sup4 suppressor strains described in **Table 6**. Additionally, negative control strains were created by transforming pRS327-Sup53 into the snR81-Test-ΔUPF1 strain and pRS327-Sup4 into the Pus7-U2-Test-ΔUPF1 strain, wherein the suppressor tRNAs do not have the correct anticodons to decode the stop signals upstream of *EGFP* (**Table 6**). As the pRS327 plasmids housing the suppressor tRNAs have a *LYS2* selectable marker, suppressor strain transformants were selected by plating on SD-URA-LEU-LYS medium (*124*).

Table 6: Yeast strains created for this study.

Yeast strain	**Ψ-Synthase recruited by reporter-strain**	**Enzyme recognition element upstream of EGFP CDS**	**Yeast reporter-strain genotype**
WT BY4742 (WT *S. cerevisiae)*	No Ψ-Synthase recruited (WT control)	-	*MATα ΔHIS3 ΔLEU2 ΔLYS2 ΔURA3*
RpG	No Ψ-Synthase recruited (positive green and red fluorescence strain)	-	*MATα RpG ΔHIS3 ΔLEU2 ΔLYS2 ΔMET15*
ΔUPF1	No Ψ-synthase recruited (NMD KO strain)	-	*MATα ΔHIS3 ΔLYS2 ΔURA3 ΔUPF1*
Pus7-Test-WT	Pus7	UGUAR-like (R = G) (Ψ138 in *RPB10* mRNA)	*MATα RpG-Pus7-Test ΔHIS3 ΔLEU2 ΔLYS2 ΔMET15*
Pus7-Pos-WT	Pus7	Sense codon positive control	*MATα RpG-Pus7-Pos ΔHIS3 ΔLEU2 ΔLYS2 ΔMET15*
Pus7-Neg-WT	Pus7	Mutated recognition sequence negative control	*MATα RpG-Pus7-Neg ΔHIS3 ΔLEU2 ΔLYS2 ΔMET15*

Pus7-Test-ΔUPF1	Pus7	UGUAR-like (R = G) (Ψ138 in *RPB10* mRNA)	*MATα RpG-Pus7-Test ΔHIS3 ΔLYS2 ΔUPF1 ΔMET15*
Pus7-Pos-ΔUPF1	Pus7	Sense codon positive control	*MATα RpG-Pus7-Pos ΔHIS3 ΔLYS2 ΔUPF1 ΔMET15*
Pus7-Neg-ΔUPF1	Pus7	Mutated recognition sequence negative control	*MATα RpG-Pus7-Neg ΔHIS3 ΔLYS2 ΔUPF1 ΔMET15*
Pus7-U2-Test-ΔUPF1	Pus7	UGUAR-like (R = G) + stem-loop region from U2 snRNA (Ψ35 in U2 snRNA)	*MATα RpG-Pus7-U2-Test ΔHIS3 ΔLYS2 ΔUPF1 ΔMET15*
snR5-Test-WT	snR5-guided H/ACA snoRNP	AGCGAAUGA UUAGA (Ψ1004 in 25S rRNA)	*MATα RpG-snR5-Test ΔHIS3 ΔLEU2 ΔLYS2 ΔMET15*
snR5-Pos-WT	snR5-guided H/ACA snoRNP	Sense codon positive control	*MATα RpG-snR5-Pos ΔHIS3 ΔLEU2 ΔLYS2 ΔMET15*
snR5-Neg-WT	snR5-guided H/ACA snoRNP	Mutated recognition sequence negative control	*MATα RpG-snR5-Neg ΔHIS3 ΔLEU2 ΔLYS2 ΔMET15*
snR5-Test-ΔUPF1	snR5-guided H/ACA snoRNP	AGCGAAUGA UUAGA (Ψ1004 in 25S rRNA)	*MATα RpG-snR5-Test ΔHIS3 ΔLYS2 ΔUPF1 ΔMET15*
snR5-Pos-ΔUPF1	snR5-guided H/ACA snoRNP	Sense codon positive control	*MATα RpG-snR5-Pos ΔHIS3 ΔLYS2 ΔUPF1 ΔMET15*
snR5-Neg-ΔUPF1	snR5-guided H/ACA snoRNP	Mutated recognition sequence negative control	*MATα RpG-snR5-Neg ΔHIS3 ΔLYS2 ΔUPF1 ΔMET15*
snR81-Test-WT	snR81-guided H/ACA snoRNP	UCAAACUUU AAAUAUGUA AGAAGUCCU UG (Ψ1052 in 25S rRNA)	*MATα RpG-snR81-Test ΔHIS3 ΔLEU2 ΔLYS2 ΔMET15*
snR81-Pos-WT	snR81-guided H/ACA snoRNP	Sense codon positive control	*MATα RpG-snR81-Pos ΔHIS3 ΔLEU2 ΔLYS2 ΔMET15*

snR81-Neg-WT	snR81-guided H/ACA snoRNP	Mutated recognition sequence negative control	*MATα RpG-snR81-Neg ΔHIS3 ΔLEU2 ΔLYS2 ΔMET15*
snR81-Test-ΔUPF1	snR81-guided H/ACA snoRNP	UCAAACUUU<u>U</u>AAAUAUGUAAGAAGUCCUUG (Ψ1052 in 25S rRNA)	*MATα RpG-snR81-Test ΔHIS3 ΔLYS2 ΔUPF1 ΔMET15*
snR81-Neg-ΔUPF1	snR81-guided H/ACA snoRNP	Mutated recognition sequence negative control	*MATα RpG-snR81-Neg ΔHIS3 ΔLYS2 ΔUPF1 ΔMET15*
Pus7-U2-Test-Sup53	Sup53 UAG suppressor tRNA targeting UAG stop codon	UGUAR-like (R = G) + stem-loop region from U2 snRNA (Ψ35 in U2 snRNA)	*MATα RpG-Pus7-U2-Test pRS327-Sup53 ΔHIS3 ΔUPF1 ΔMET15*
snR81-Test-Sup4	Sup4 UAA suppressor tRNA targeting UAA stop codon	UCAAACUUU<u>U</u>AAAUAUGUAAGAAGUCCUUG (Ψ1052 in 25S rRNA)	*MATα RpG-snR81-Test pRS327-Sup4 ΔHIS3 ΔUPF1 ΔMET15*
snR81-Test-Sup53	Sup53 UAG suppressor tRNA targeting an incompatible UAA stop codon	UCAAACUUU<u>U</u>AAAUAUGUAAGAAGUCCUUG (Ψ1052 in 25S rRNA)	*MATα RpG-snR81-Test pRS327-Sup53 ΔHIS3 ΔUPF1 ΔMET15*
Pus7-U2-Test-Sup4	Sup4 UAA suppressor tRNA targeting an incompatible UAG stop codon	UGUAR-like (R = G) + stem-loop region from U2 snRNA (Ψ35 in U2 snRNA)	*MATα RpG-Pus7-U2-Test pRS327-Sup4 ΔHIS3 ΔUPF1 ΔMET15*

3.5–Flow cytometry of yeast strains

From SD-URA (for reporter system variants in a WT BY4742 background) or SD-LEU-URA (for reporter system variants in a *ΔUPF1* background) plates, single colonies were used as inoculum for 5 mL YPD cultures (2% w/v glucose) incubated overnight at 30°C. The following day, all cultures were centrifuged at 3000 × g for 5 minutes prior to

cell resuspension in 4.8 mL YPD lacking glucose. OD_{600} measurements of each resuspended cell culture were taken. Subsequently, these resuspended cell samples were re-inoculated into YPD (lacking glucose) to form 5 mL cultures (containing 1% w/v raffinose) with an OD_{600} of 0.3.

The cultures were incubated at 30°C with shaking for 5 hours prior to induction of fluorescent reporter gene expression by adding galactose to a 1.8% w/v concentration. The induced cultures were left to grow overnight at 30°C with shaking-incubation. Alongside the yeast strains housing insert upstream of *EGFP* in the RpG system, a negative-fluorescence strain (either WT BY4742 or *ΔUPF1* yeast) and a positive-fluorescence strain, RpG, were prepared under identical conditions. After overnight incubation, each galactose-induced cell culture was washed three times in cold, phosphate-buffered saline (PBS) solution. Yeast strains to be assayed for heat-shock induced pseudouridylation by way of Pus7 (Pus7-Test, Pos7-Pos, and Pus7-Neg WT and ΔUPF1 strains detailed in **Table 6**) were subjected to 1 hour of 45°C heat-shock with shaking after overnight galactose-induction. After this heat shock treatment, the cells were washed as previously described.

The relative green and red fluorescence of 100,000 yeast cells from each strain was evaluated in flow cytometry assays conducted with a BD FACS aria fusion instrument provided by the SynBridge Synthetic Biology Marker Space at the University of Lethbridge. For these assays, fluorescein isothiocyanate-area (FITC-A) and Phycoerythrin-Texas-Red-area (PE-Texas-Red-A) filters were used to isolate green and red signal from each yeast cell after initial excitation at 488 nm and 561 nm, respectively. The photomultiplier voltages used for signal detection from the forward-scatter-area, side-scatter-area, FITC-A, and PE-Texas-Red-A filters are given in **Table 7**. For the set of test,

positive, and negative control snR5 reporter strains with a ΔUPF1 yeast background, three technical replicate assays were carried out after overnight galactose induction and one follow-up assay was conducted 3 hours post-galactose induction. For the remaining Pus7 and snR81 reporter strains and corresponding controls in wild type and ΔUPF1 yeast, only a single flow cytometry assay was conducted (one technical replicate). Flow cytometry assays for the Sup53 and Sup4-based nonsense suppression control strains were also conducted only a single time.

Red and green fluorescence-histograms and red vs. green dot plot analyses were generated for each yeast strain using FlowJo software (Tree Star). To select for yeast cells that only exhibited a high level of Ruby2 signal expression, a gating step was performed during data analysis. The red fluorescence histogram from either a WT or *ΔUPF1* yeast strain analyzed on the same day as the reporter strains was compared to the red fluorescence histograms from the test, positive, and negative reporter strain variants. From this histogram overlay, a data subset was created that excluded any cells of the test, positive, and negative strains that overlapped with the WT or *ΔUPF1* yeast's red fluorescence distribution. The cells that passed this Ruby2 selection step were then included in the final flow cytometry analysis (Appendix 2).

To quantify the magnitude of the green fluorescence signals emitted by the snR5-Test-ΔUPF1 technical replicates, EGFP fold-change values were calculated wherein the median FITC-A intensity of the snR5-Test-ΔUPF1 strain of each replicate was divided by that of the corresponding snR5-Neg-ΔUPF1 control strain in each replicate. Additionally, the percentage of EGFP fluorescence recovered in the snR5-Test-ΔUPF1 and snR5-Neg-ΔUPF1 reporter systems relative to that of the snR5-Pos-ΔUPF1 positive control strain was

calculated using the median FITC-A values for each strain and averaged across the three technical replicates to give the EGFP signal recovery ± SD for the snR5-Test-ΔUPF1 and snR5-Neg-ΔUPF1 systems.

Table 7. Photomultiplier voltage settings for the flow cytometry assays conducted in this study. The forward-scatter-area (FSC-A), side-scatter-area (SSC-A), fluorescein isothiocyanate-area (FITC-A: green signal detection), and PE-Texas-Red-area (PE-Texas-Red-A: red signal detection) photomultiplier voltages for each flow cytometry assay conducted in this thesis.

Flow cytometry experiment	FSC-A	SSC-A	FITC-A	PE-Texas Red-A
Positive control reporter system (**Figure 8C**)	274	363	495	1000
Pus7 activity at 30°C and 45°C in a WT and ΔUPF1 background (**Figures 10** and **12**)	25	175	380	500
Pus7 activity in Pus7-U2-Test-ΔUPF1 reporter system (**Figure 13**)	25	176	519	560
snR81 H/ACA snoRNP activity in a WT and ΔUPF1 background (**Figure 14**)	25	175	380	500
snR5 H/ACA snoRNP activity in a WT background (**Figure 15**)	25	175	380	500
snR5 H/ACA snoRNP activity in a ΔUPF1 background following overnight induction-replicate 1 (**Figure 16**)	25	175	380	500
snR5 H/ACA snoRNP activity in a ΔUPF1 background following overnight induction-replicate 2 (**Figure 16**)	25	176	519	560
snR5 H/ACA snoRNP activity in a ΔUPF1 background following overnight induction-replicate 3 (**Figure 16**)	20	150	460	600
snR5 H/ACA snoRNP activity in a ΔUPF1 background 3 hours post-induction (**Figure 16**)	25	175	380	857
Stop codon readthrough with Sup4 and Sup53 suppressor tRNAs (**Figure 18**)	20	150	460	600

3.6–Western blot analysis of EGFP expression

The snR5-Test-ΔUPF1, snR5-Pos-ΔUPF1, snR5-Neg-ΔUPF1, and ΔUPF1 yeast strains (**Table 6**) were cultured and induced for fluorescent protein expression as described in sub-chapter 3.5. After overnight shaking-incubation at 30°C following galactose induction, cell pellets were harvested from each 5 mL culture and resuspended in 1 mL 0.2 M NaOH. These mixtures were incubated at room temperature for 45 minutes prior to centrifugation, pellet resuspension in 150 µL 1× SDS running buffer (16.7 mM Tris-HCl (pH 6.8), 3.3 mM 2-Mercaptoethanol, 2.5 mM EDTA, 1% SDS, 0.03% Bromophenol blue, 10% glycercol), and boiling for 3 minutes. The boiled mixtures were pelleted a final time before 30 µL of supernatant from each sample were loaded onto an SDS-PAGE gel (6% stacking layer, 12% resolving layer) alongside a purified GFP standard in amounts of 19.8 and 79.3 pmol (the purified GFP sample was kindly provided by Luc Roberts). Four µL of EZ-Run Prestained *Rec* Protein Ladder (Fisher Scientific) was loaded onto the gel as a size control. SDS-PAGE was subsequently conducted at 100V for 45 minutes and then at 180V until the dye front migrated to the bottom of the gel. The protein contents of the gel were affixed to a nitrocellulose membrane via wet-transfer at 100V for 1 hour in transfer buffer (48 mM Tris, 39 mM Glycine, 0.037% SDS, 20% methanol, pH 9.2, cooled to 4°C). This membrane was subsequently incubated at room temperature for 30 minutes in 10 mL blocking buffer (5% w/v skim milk, 1× TBS, 0.1% Tween 20) before being washed in 2× TBS for 2 min. The membrane was incubated in a 1:1000 dilution of anti-GFP antibody (10 mL blocking buffer, 10 µL anti-GFP antibody ab6556 (Abcam)) overnight at 4°C. After three consecutive washes in 10 mL 1× TBS, 0.1% Tween 20, the membrane was incubated in a 1:1000 dilution of secondary antibody (10 mL blocking buffer, 10 µL goat anti-rabbit

IgG H&L (HRP) antibody ab205718 (Abcam)) at room temperature for 1 hour. Subsequently, the membrane was washed twice in 1× TBS, 0.1% Tween 20, and once in 1× TBS prior to chemiluminescent development. Development consisted of incubating the membrane with activated luminol solution: 110 µL of 91 mM p-coumeric acid (prepared in DMSO) added to 25 mL luminol solution (100 mM Tris-HCl pH 8.5, 11.3 mg luminol dissolved in 0.25 mL DMSO) and then combined with 5 mL H_2O_2 solution (100 mM Tris-HCl pH 8.5, 0.06% v/v H_2O_2). After a 5-minute development time at room temperature, the membrane was visualized by an Amersham Imager 600 instrument (GE Healthcare).

Chapter 4–Results

4.1–Design of a fluorescence-based yeast reporter system to assay for pseudouridine synthase activity *in vivo*

In analogy to the RNA-ID reporter for RNA *cis*-regulatory elements designed by Dean and Grayhack, the pseudouridine reporter system consists of a monomeric Ruby2 gene and a yeast optimized enhanced *GFP* gene under the control of a galactose-inducible, bidirectional promoter *GAL1,10 (118)*. On the genetic level, the fluorescence-based reporter system exists as a single copy integrated into the yeast's genome at the *MET15* locus on chromosome XII. A *URA3* marker is used to select for yeast colonies that have successfully integrated the reporter construct (**Figure 6A**). Situated between the *GAL1,10* promoter and the *EGFP* gene is an intervening sequence consisting of a *TEF1* gene fragment and a 2A peptide. The *TEF1* gene fragment encodes the 5′ UTR and first 10 N-terminal amino acids of translation elongation factor 1 (Appendix 1). In the N-terminal sequence, there is a Lys3 to Leu mutation such that a unique XhoI restriction site is introduced. Immediately following the *TEF1* gene fragment is a sequence encoding the foot-and-mouth disease virus (FMDV) 2A peptide. This cis-acting hydrolase element possesses a DxExNPGP motif that induces a translational stall in which a peptide bond fails to form after the incorporation of glycine into the nascent peptide and before the incorporation of proline (*127*).

The fluorescence-based reporter system is programmed to detect a pseudouridine synthase by insertion of a substrate recognition sequence belonging to the pseudouridine synthase of interest at the XhoI cut site. In the resultant reporter construct, the target uridine within the recognition sequence occurs in an in-frame stop codon context. Once the

reporter system is transformed into yeast and the strain is induced with galactose, transcription of mRuby2 and EGFP mRNA occurs. While there are no obstructions to mRuby2 translation, complete translation of the EGFP transcript is contingent upon the recruitment of a pseudouridine synthase to its recognition element and the subsequent modification of the stop codon-embedded uridine to pseudouridine (**Figure 6B**). If this modification occurs, translational readthrough of the pseudouridylated stop codon will allow for green fluorescent protein to be expressed (*97*). Once translation proceeds past the modified stop codon, the 2A peptide sequence immediately upstream of the EGFP reporter gene results in nascent peptide release before the ribosome continues to the EGFP coding sequence, thus eliminating an N-terminal tag on the EGFP reporter that could interfere with protein folding and subsequent fluorescence. The level of green and red fluorescent output from the reporter system is determined by flow cytometry analysis.

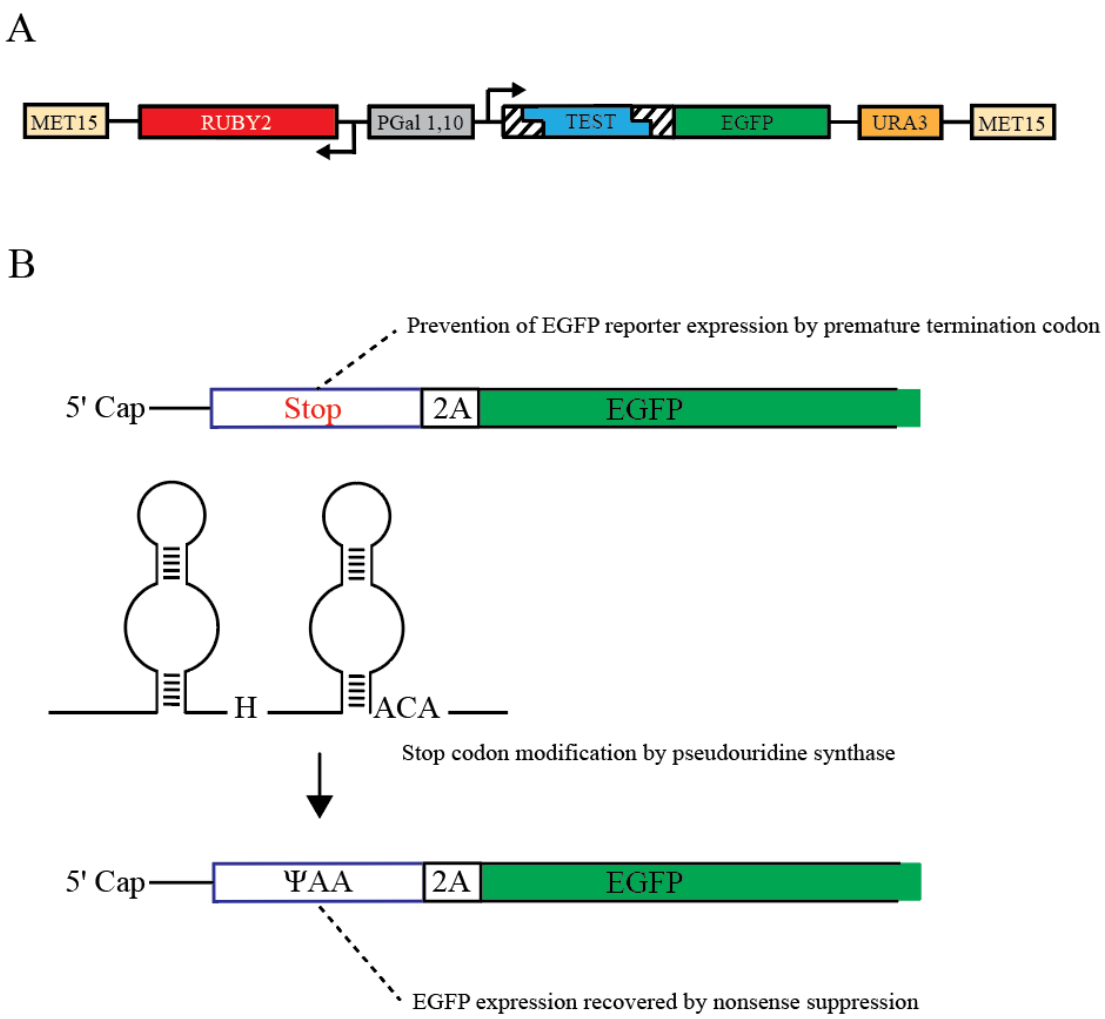


Figure 6. Design of a fluorescence-based yeast reporter system to assay for pseudouridine synthase activity *in vivo*. (A) Schematic representation of the dual-fluorescence reporter construct encoding monomeric Ruby2 and yeast optimized-enhanced green fluorescent protein (EGFP) under control of a bi-directional, galactose-inducible promoter (pGal 1, 10). A short sequence (TEST) upstream of the EGFP reporter contains a premature stop codon embedded within the recognition element of a pseudouridine synthase. **(B)** Upon recruitment of a pseudouridine synthase to its cognate recognition element, modification of the stop codon-embedded uridine will lead to readthrough during subsequent translation. A 2A self-cleaving peptide sequence between the premature stop codon and EGFP coding sequence causes a ribosome skipping event, releasing the nascent peptide resulting from stop codon readthrough before EGFP is translated. EGFP and Ruby2 expression is detected by green and red fluorescent signal emission by flow cytometry.

4.2–Workflow to construct yeast reporter system variants to assay for different pseudouridine synthase activities

Construction of a yeast reporter system variant begins with pFA6a-link-yomRuby2-pGal-yoEGFP (pRpG) plasmid used for sequence and ligation-independent cloning to insert a target sequence. XhoI restriction and subsequent T4 DNA polymerase-mediated 3′ chew back creates linearized plasmid with ~30 nt, single stranded 5′ overhangs for sequence and ligation-independent cloning (*123*). Annealing of sense and antisense oligos encoding a recognition sequence for a natural substrate of a pseudouridine synthase yields a double stranded DNA molecule with ~30 nt 5′ overhangs complementary to the chewed back pRpG plasmid. Incubation of this double-stranded insert with the pRpG vector in a DNA assembly reaction creates an *in vitro* recombination intermediate. Upon transformation into *E. coli*, the cell's repair machinery completes the DNA repair process thus yielding a recombinant pRpG plasmid encoding a recognition sequence for a pseudouridine synthase. Using this methodology, the natural substrates of three pseudouridine synthases were inserted upstream of *EGFP* in the reporter system: a region from 25S rRNA recognized by the snR5 H/ACA snoRNP for Ψ1004 formation, a region from 25S rRNA recognized by the snR81 H/ACA snoRNP for Ψ1052 formation, and a region from *RPB10* mRNA recognized by Pus7 for pseudouridine formation at site 138 (*7, 25, 128*) (**Figure 7**). A complete list of the oligonucleotides used for sequence and ligation-independent cloning is given in **Table 4**.

Using primers with *MET15* sequence homology arms attached to their 5′ ends, the reporter system programmed to a specific pseudouridine synthase is PCR-amplified using the pRpG plasmid variant as template. Transformation of the linear product into yeast

triggers the homologous recombination pathway wherein the reporter system is integrated into the *MET15* locus on chromosome XII of the yeast genome (**Figure 7**).

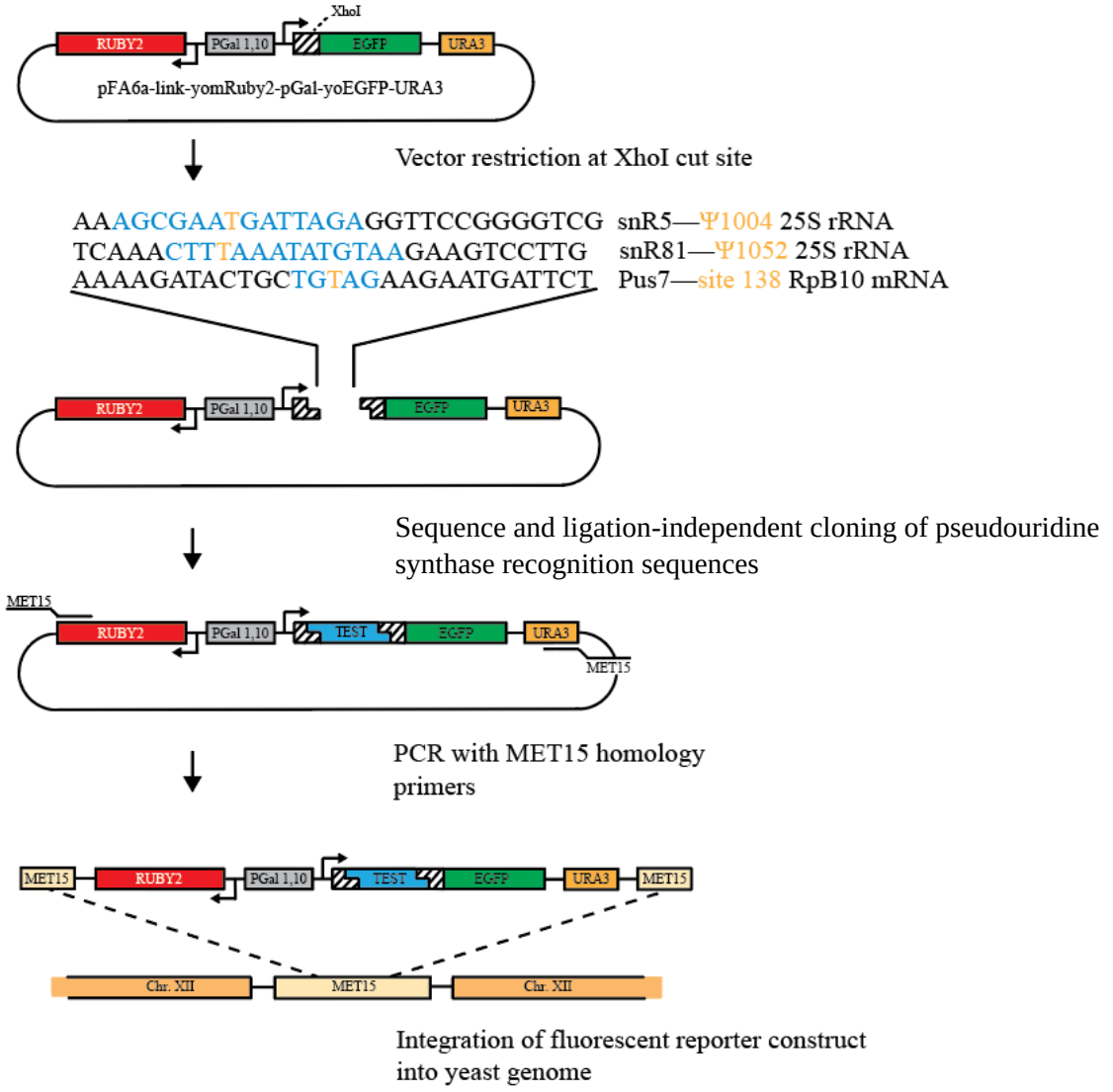


Figure 7. Workflow to construct yeast reporter system variants to detect different pseudouridine synthase activities. Using sequence and ligation-independent cloning, short oligo sequences corresponding to natural substrates of snR5 and snR81 H/ACA snoRNPs as well as standalone pseudouridine synthase Pus7 were inserted upstream of *EGFP* in the pFA6a-link-yomRuby2-pGal-EGFP-URA3 (pRpG) plasmid using a unique XhoI restriction site. Blue letters indicate the pseudouridine synthase recognition sequence and the gold letter indicates the targeted uridine (T in DNA) in a stop codon context. From the resultant plasmids, the dual-fluorescent reporter system is PCR amplified with primers containing *MET15* sequence homology arms appended to their 5′ ends. Subsequently, these PCR products are transformed into yeast where they are integrated into the yeast's genomic DNA by homologous recombination at the *MET15* locus on Chr. XII.

4.3–Red and green fluorescence detection from a positive control yeast reporter system

Prior to creating a yeast reporter strain attuned to the activity of a pseudouridine synthase, a positive control strain lacking insert upstream of *EGFP* was generated to characterize the red and green fluorescent output of the reporter system. Using primers containing *MET15* sequence homology arms on their 5′ ends, amplification of the reporter system lacking insert upstream of *EGFP* from the pRpG plasmid is expected to yield a PCR product of 4336 bp. 1% agarose gel electrophoresis reveals the presence of a single band between 4000 and 5000 bp after thermocycling, corresponding to the amplified reporter construct. This band persists after a PCR purification step to clean up the amplified product (**Figure 8A**). Transformation of the PCR purified reporter construct into yeast results in its homologous recombination into the *MET15* locus on chromosome XII of the yeast's genome (**Figure 8B**). To characterize the fluorescent output of the reporter system, a BY4742 wild type yeast strain, uninduced reporter strain, and galactose-induced reporter strain were cultured and subjected to flow cytometry analysis.

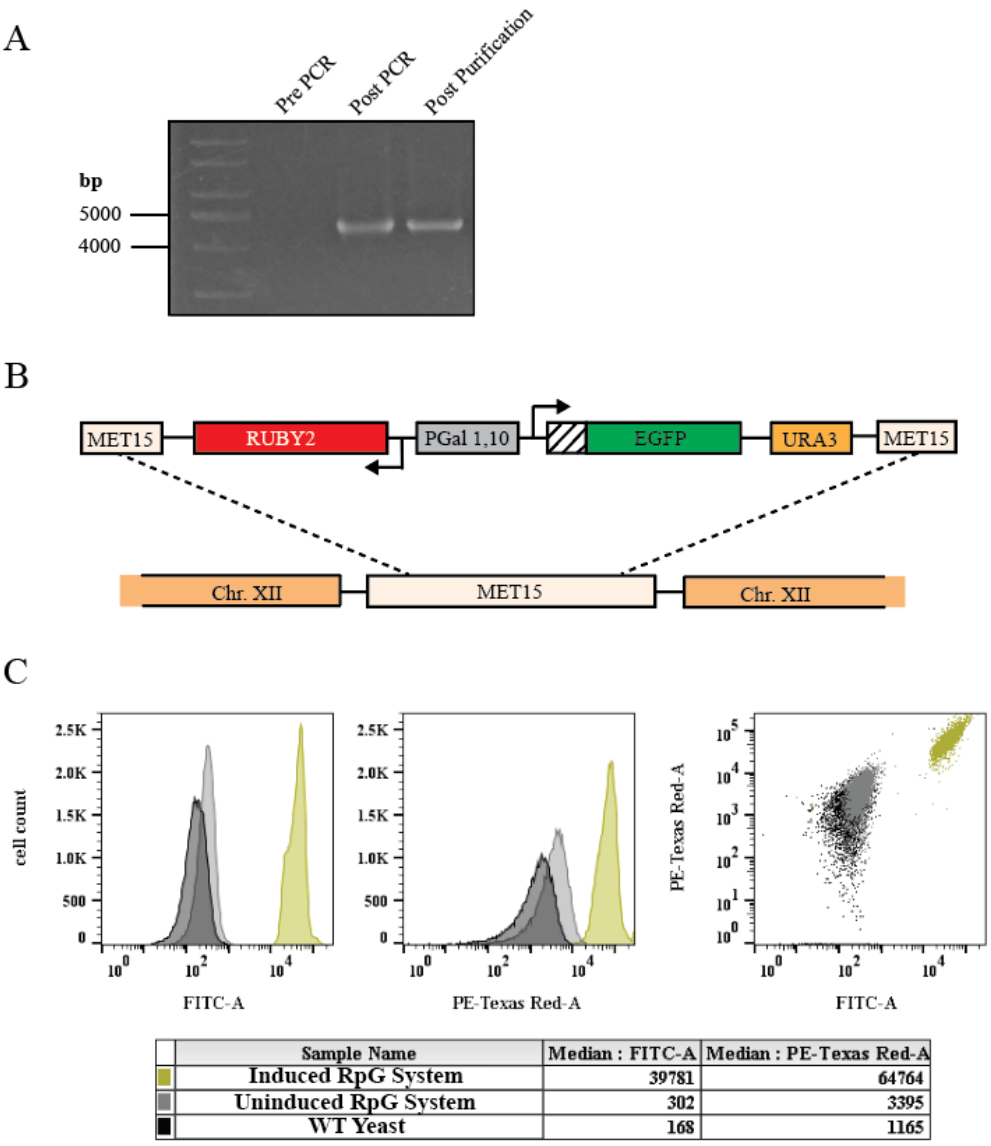
During flow cytometry analysis, yeast cells in suspension are focused through a nozzle whereupon they pass through a laser beam one cell at a time. As a single cell intercepts the laser, it is simultaneously excited by 488 and 561 nm light, corresponding to the excitation maxima of EGFP and mRuby2. A series of channels in the flow cytometer each composed of a filter and photomultiplier tube (PMT) subsequently detect light of specific wavelengths emitted from the cell. The FITC-A channel detects the light emitted by EGFP (emission max. of 507 nm) and the PE-Texas-Red-A channel detects the light emitted by mRuby2 (emission max. of 600 nm). The EGFP and mRuby2 signal intensity

from an analyzed yeast cell population is then visualized in the form of a fluorescence histogram wherein cell count is plotted against the EGFP or mRuby2 signal intensity, given in arbitrary units, of each cell analyzed. Furthermore, the relationship between the mRuby2 and EGFP signals of the yeast population can be observed via dot plot analysis wherein the mRuby2 signal from each cell is plotted against the corresponding EGFP signal from the same cell.

The BY4742 wild type yeast distributions in the EGFP and mRuby2 fluorescence histogram overlays establish base-line values for zero green and red fluorescence, with a median EGFP signal of 168 arbitrary units (a.u.) and a median mRuby2 signal of 1165 a.u., respectively (**Figure 8C**). Relative to the wild type strain, the fluorescence distributions for the uninduced reporter strain are shifted to slightly higher fluorescence intensities in both the EGFP and mRuby2 fluorescence histogram overlays—the median EGFP signal for the uninduced reporter system is 302 a.u. and the median mRuby2 signal is 3395 a.u. Although this strain was not deliberately induced with galactose, a low level of induction likely occurs from a residual amount of galactose present in the undefined YPD media in which the cells were cultured, resulting in the uninduced reporter strain's slight increase of red and green signal above base-line. Compared to the wild type and uninduced reporter strain controls, the induced yeast reporter strain exhibits a distinct fluorescence shift in both the green and red signal channels. As shown in the EGFP fluorescence histogram overlay, the induced reporter strain's cell population has a median signal intensity of 39781 a.u. In the mRuby2 histogram overlay, the induced reporter strain's cell population has a median signal intensity of 64764 a.u. (**Figure 8C**). The dot plot analysis reveals a critical feature of the reporter system. As EGFP and mRuby2 expression are controlled by the same

promoter (*GAL 1, 10*), the green and red signal intensities observed through flow cytometry are correlated—a high level of mRuby2 signal (hence mRuby2 expression) is accompanied by a high level of EGFP signal (hence EGFP expression) in the yeast reporter system (**Figure 8C**).

Of additional note are the magnitudes of the green and red fluorescence shifts from the induced reporter strain relative to the control strains; the median mRuby2 signal from the induced reporter system is one order of magnitude greater than the uninduced reporter system (64764 a.u. vs. 3395 a.u.) and the median EGFP signal from the induced reporter system is two orders of magnitude greater than the uninduced reporter system (39781 a.u. vs. 302 a.u.). Thus, especially in the case of EGFP reporter expression, there exists a large dynamic range between no reporter expression and induced expression suitable for detection of changes in fluorescent signal output over two orders of magnitude.



Sample Name	Median : FITC-A	Median : PE-Texas Red-A
Induced RpG System	39781	64764
Uninduced RpG System	302	3395
WT Yeast	168	1165

Figure 8. Red and green fluorescence detection from a positive control yeast reporter system. (A) Using the pRpG plasmid as template, the bidirectional RpG fluorescence construct was PCR amplified with primers containing *MET15* sequence homology arms appended to their 5′ ends. The PCR product was visualized by agarose gel electrophoresis. The expected amplicon size is 4336 bp. **(B)** Schematic representation of RpG reporter construct integration into the *MET15* locus on Chr. XII by homologous recombination upon transformation of the PCR product into yeast. **(C)** Flow cytometry of the positive control yeast strain expressing both mRuby2 and EGFP. Yeast cells were analyzed for red and green fluorescence in a BD FACS Aria fusion instrument. The left histogram depicts cell count vs. relative green fluorescence detected with a FITC-A filter for wild type, induced, and uninduced reporter strains. The middle histogram shows cell count vs. the relative red fluorescence detected with a PE-Texas Red-A filter. The right overlay presents the red vs. green signal-expression dot plots for the wild type, induced, and uninduced reporter strains.

4.4–Creation of yeast reporter system variants to assay for snR5 and snR81 H/ACA snoRNP activity and standalone Pus7 activity

To attempt to detect the pseudouridylation activity of the two H/ACA snoRNPs, snR5 and snR81, as well as the stand-alone pseudouridine synthase Pus7, a set of reporter constructs were created. The snR5 and snR81 H/ACA snoRNPs were chosen because their targeted uridines in 25S rRNA occur in UGA and UAA stop codon contexts, respectively, allowing for the generation of in-frame stop codons upstream of *EGFP* upon sequence and ligation-independent cloning of these 25S rRNA recognition sequences into the RpG reporter system. Furthermore, the snR81 H/ACA snoRNP is the only pseudouridine synthase that has been manipulated to target stop codon uridines in the literature, so I sought to replicate this instance of engineered pseudouridylated stop codon readthrough and expand this limited research to another H/ACA snoRNP: snR5 (*97, 109*). Pus7 was selected for analysis due to its importance in the heat shock induced mRNA pseudouridylation response, which I hoped to observe with the fluorescence-based reporter system (*7*). Additionally, Pus7 is the only known standalone pseudouridine synthase to possess a consensus sequence for substrate recognition (UGUAR where R = G > A and U is converted to Ψ), meaning this enzyme can be readily tested for activity by inserting a Pus7 substrate sequence upstream of the *EGFP* reporter gene (*98*). Critically, the targeted uridine within the Pus7 recognition sequence occurs in a stop codon context, UAA or UAG, depending on the identity of the purine base two positions downstream of the target U, thus allowing for creation of an in-frame stop codon upon ligation of a Pus7 recognition sequence into the yeast reporter system.

For each pseudouridine synthase, three reporter system variants were generated. The "test" reporter system variant consists of a target uridine in a stop codon context within the recognition sequence of a pseudouridine synthase. In the positive reporter system variant ("Pos") the stop codon within the pseudouridine synthase recognition sequence has been replaced with a sense codon, which enables high expression of EGFP from the induced reporter system. Finally, in the negative reporter system ("Neg"), the pseudouridine synthase recognition sequence surrounding the target stop codon has been abolished by introducing synonymous codon mutations, thus eliminating the possibility of pseudouridylation at this site without changing the encoded amino acid sequence upstream of the *EGFP* reporter (**Figure 9A**).

Sequence and ligation-independent cloning was used to generate reporter system variants containing the recognition sequence of a pseudouridine synthase upstream of the *EGFP* reporter gene (*123*). Successful insertion of a recognition sequence into the pRpG plasmid is expected to destroy the unique XhoI cut site, as revealed by comparing the plasmid with and without XhoI digestion during a restriction screening assay: XhoI digestion of parent pRpG plasmid is expected to yield a 6705 bp band when visualized by 1% agarose gel electrophoresis. In the absence of XhoI, the pRpG plasmid produces the relaxed circular, linear, and supercoiled banding pattern typical of undigested plasmid when electrophoresed (pRpG +/- XhoI, **Figure 9B**). A putative clone of a reporter system variant created through sequence and ligation-independent cloning, pRpG-snR5-Test, fails to linearize when incubated with XhoI and produces a banding pattern identical to the undigested plasmid control, indicating successful insertion of the snR5 H/ACA snoRNP recognition sequence into the pRpG plasmid (pRpG-snR5-Test +/- XhoI, **Figure 9B**). This

XhoI-based restriction screening assay was paired with subsequent Sanger sequencing analysis to confirm the successful creation of nine variants of the pRpG plasmid: a set of test, positive, and negative control constructs designed to assay for snR5 and snR81 H/ACA snoRNP activity and standalone Pus7 activity (**Table 5**).

Using primers containing *MET15* sequence homology arms, PCR amplicons were obtained from the pRpG plasmid variants. Successful amplification is expected to yield a single band of 4387 bp which can be visualized by agarose gel electrophoresis. In an initial round of construct creation, 50 ng of plasmid DNA were used in PCR reactions for all 9 yeast system variants. From this initial round, the snR5-Test, snR5-Pos, snR5-Neg, and snR81-Test PCRs produced a single band of the expected size as visualized by agarose gel electrophoresis, indicating successful amplification (**Figure 9C**). For the remaining constructs that were not successfully amplified, template DNA was titrated in subsequent PCR reactions containing 24, 50, and 75 ng of plasmid until a band of the expected 4387 bp size was observed by gel electrophoresis.

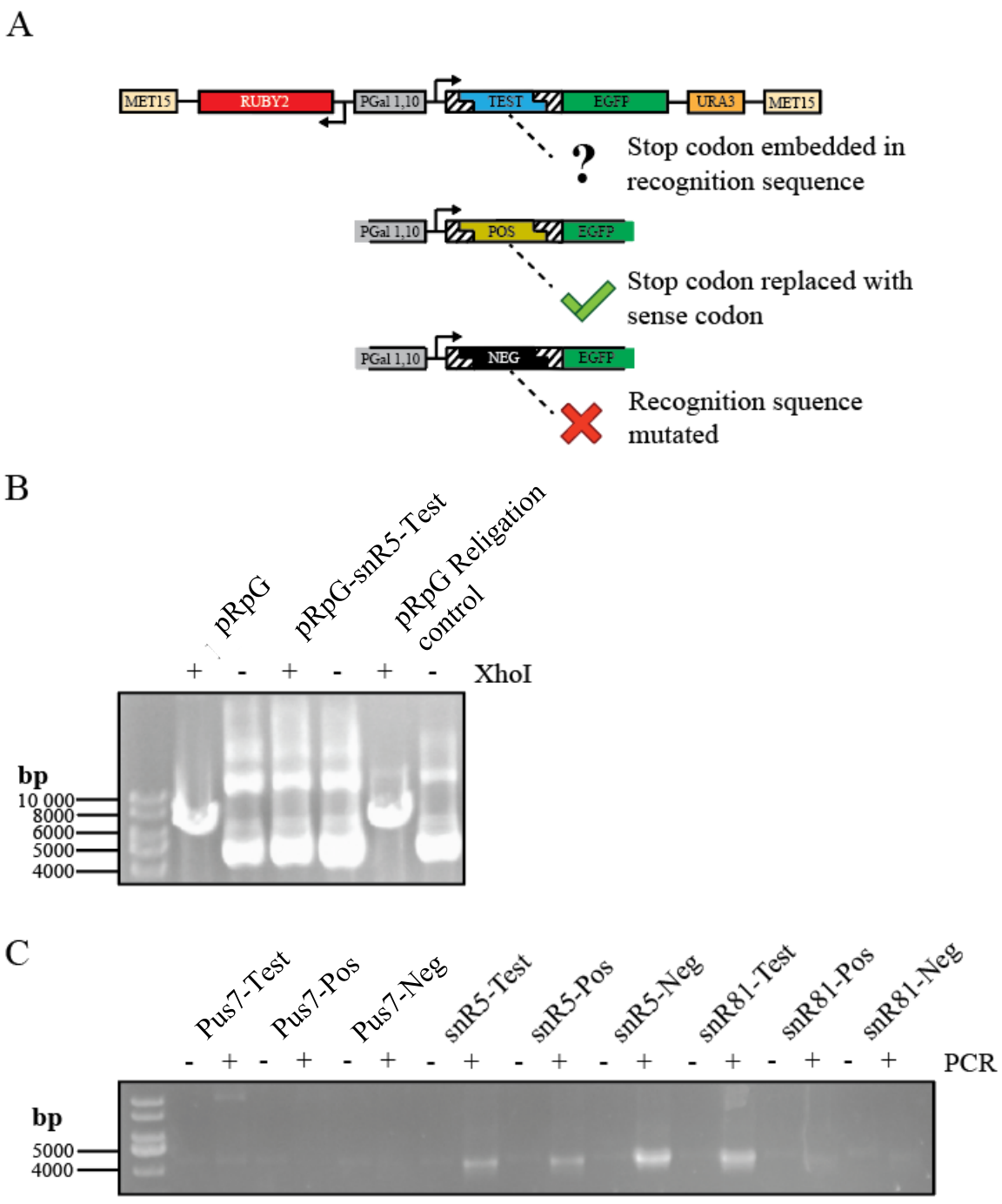


Figure 9. Creation of yeast reporter system variants to assay for snR5 and snR81 H/ACA snoRNP activity and standalone Pus7 activity. (A) The H/ACA snoRNPs snR5 and snR81, as well as the standalone pseudouridine synthase Pus7 were selected for analysis by the yeast reporter system. In addition to a test construct possessing a stop codon embedded in a pseudouridine synthase recognition sequence, corresponding positive and negative control constructs for EGFP expression were created for each pseudouridine synthase analyzed by the yeast reporter system. The positive control constructs contain a sense codon embedded within the pseudouridine synthase recognition sequence and the

negative control constructs have had the recognition sequence surrounding the target stop codon mutated, thus eliminating the possibility of modification at this site. **(B)** Agarose gel electrophoresis of parent pRpG plasmid, pRpG-snR5-Test plasmid after insertion of a short snR5 target sequence by sequence and ligation-independent cloning, and religated pRpG plasmid lacking insert. Successful sequence and ligation-independent cloning is expected to destroy the unique XhoI restriction site as evident by comparing plasmid with and without XhoI digestion. **(C)** Using primers containing *MET15* sequence homology arms, PCR amplicons were generated from the pRpG plasmid variants harbouring snR5, snR81, and Pus7 target sites. Successful amplification is expected to yield a band between 4000 and 5000 bp as visualized by agarose gel electrophoresis.

4.5–Assaying for Pus7 enzyme activity at 30°C

As Pus7 has been identified as a critical pseudouridine synthase in the heat-shock induced mRNA pseudouridylation program, I sought to configure the yeast reporter system to detect the activity of this enzyme at both 30°C and 45°C (*7*). The yeast reporter system was programmed to detect the modifying activity of Pus7 by inserting a 29 nt sequence encompassing positions 124-152 of *RPB10* mRNA upstream of the EGFP reporter gene. *RPB10* is an essential gene encoding a core subunit common to RNA polymerases I, II, and III (*129*). A pseudouridine at site 138 in the mRNA transcript of this gene was previously identified through Ψ-seq transcriptome mapping (*7*). Based on the occurrence of U138 in a Pus7 consensus sequence, in addition to the loss of this modification in a *ΔPus7* yeast strain, the U138 isomerization has been attributed to the standalone pseudouridine synthase Pus7 (*7*). Moreover, modification at this site was shown to be enhanced under heat-shock conditions consisting of culture incubation at 45°C for 1 hour (*7*).

To assay for Pus7 activity at 30°C in a wild type yeast background, yeast cells from the Pus7-Test-WT, Pus7-Pos-WT, and Pus7-Neg-WT strains (**Table 6**) were analyzed by

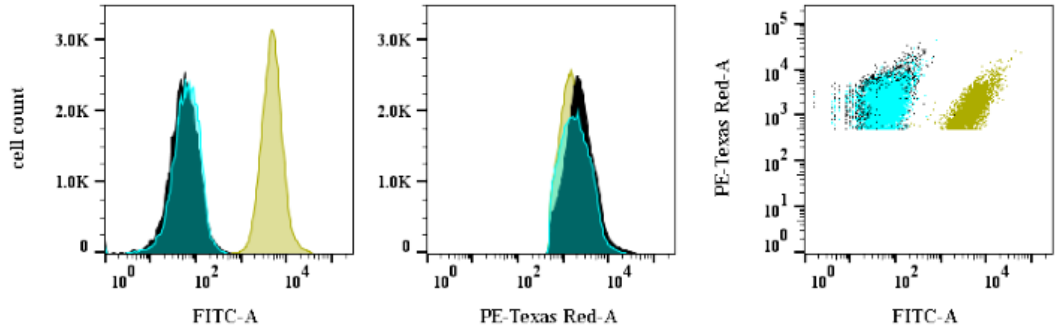
flow cytometry. The green and red fluorescence distributions of the Pus7-Neg-WT yeast strain have median fluorescence values of 55.9 and 2143 a.u., respectively (**Figure 10A**). The EGFP and Ruby2 fluorescence histograms for the Pus7-Test-WT strain are comparable to the negative control strain, with median green and red fluorescence values of 63.9 and 1753 a.u., respectively. The Pus7-Pos-WT strain exhibits a substantial green fluorescence shift, producing an EGFP fluorescence histogram with a median fluorescence of 4589 a.u. The median red signal for the Pus7-Pos-WT strain, 1442 a.u., is comparable to the levels of mRuby2 expression in the test and negative control strains (**Figure 10A**). Overall, the Pus7-Test-WT system grown at 30°C did not exhibit a discernible green fluorescence shift relative to the fluorescence distribution for the Pus7-Neg-WT control strain. This observation could be due to an insufficient level of stop codon modification and/or translational readthrough, yet other molecular reasons may also underlie the lack of green fluorescent signal from this reporter system (vide infra) (**Figure 10A**).

In eukaryotes, a mRNA quality control system called the nonsense-mediated decay (NMD) pathway is responsible for the degradation of transcripts possessing premature termination codons (*130*). As the fluorescence-based reporter system expresses EGFP mRNA transcripts housing a premature stop codon embedded in a pseudouridine synthase recognition sequence, the NMD pathway is likely to interfere with the function of the yeast reporter assay. To eliminate any interference from NMD, the gene for a core factor in the pathway, *UPF1*, was deleted, thus generating a *ΔUPF1* strain into which the Pus7-Test, Positive, and Negative reporter system variants were integrated (**Figure 11**).

To assay for Pus7 activity at 30°C in a *ΔUPF1* background, yeast cells of the Pus7-Test-ΔUPF1, Pus7-Pos-ΔUPF1, and Pus7-Neg-ΔUPF1 strains (**Table 6**) were analyzed by

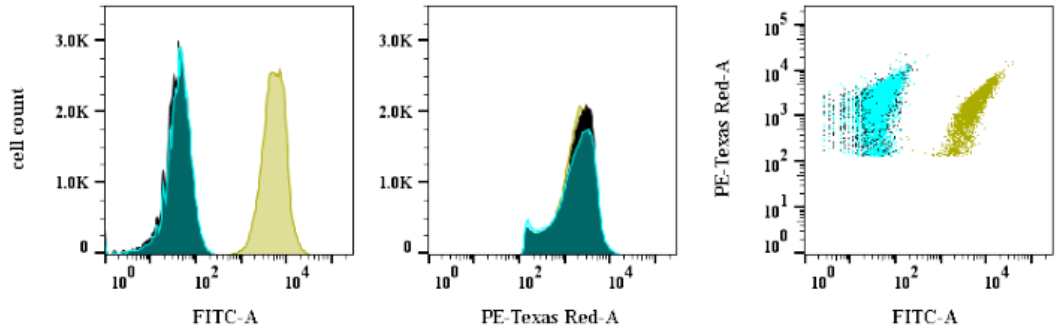
flow cytometry. In a *ΔUPF1* background, the Pus7-Test strain produced EGFP and Ruby2 fluorescence histograms with median fluorescence values of 44.4 and 1926 a.u., respectively. These values were comparable to the median EGFP and mRuby2 fluorescence signals from the Pus7-Neg strain of 41.7 and 2023 a.u., respectively (**Figure 10B**). Thus, the absence of the NMD pathway does not increase the EGFP signal output of the Pus7-Test-ΔUPF1 yeast strain grown at 30°C. Despite stabilization of the premature stop codon-containing transcript due to the *ΔUPF1* genotype, the modification extent of the *EGFP* reporter transcript by Pus7 might still be insufficient to promote a detectable level of EGFP translation at 30°C, possibly because Pus7 pseudouridylation of the *RPB10* mRNA recognition site was shown to be most pronounced under 45°C heat shock stress (*7*). In order to prove that insufficient pseudouridylation limits EGFP expression, it would be required to specifically probe the modification status of the stop codon uridine upstream of the EGFP reporter (see accompanying discussion in Chapter 5.1).

A WT 30°C



	Sample Name	Median : FITC-A	Median : PE-Texas Red-A	Percentage
	Pus7-Test	63.9	1753	83.4
	Pus7-Neg	55.9	2143	87.9
	Pus7-Pos	4589	1442	80.6

B ΔUPF1 30°C



	Sample Name	Median : FITC-A	Median : PE-Texas Red-A	Percentage
	Pus7-Test	44.4	1926	77.1
	Pus7-Neg	41.7	2023	80.9
	Pus7-Pos	5184	1897	81.2

Figure 10. Assaying for Pus7 activity at 30°C. (A) Fluorescent outputs of the Pus7-Test-WT (cyan), Pus7-Pos-WT (gold), and Pus7-Neg-WT (black) yeast strains analyzed with a flow cytometry assay. The left histogram overlay depicts the cell count vs. relative green fluorescence detected with a FITC-A filter for the test, positive, and negative control strains. The middle histogram overlay displays cell count vs. relative red fluorescence detected with a PE-Texas Red-A filter. The right overlay presents red vs. green signal-expression dot plots for the test, positive, and negative control strains. Median red and green fluorescence values for each strain are given in the table alongside the percentage of each cell population that was included in the analysis after gating for cells exhibiting high Ruby2 expression (see Materials and Methods section 3.5 and Appendix 2). **(B)** EGFP and Ruby2 fluorescence analysis of the Pus7-Test-ΔUPF1, Pus7-Pos-ΔUPF1, and Pus7-Neg-ΔUPF1 yeast reporter strains. The analysis of Pus7 activity in a ΔUPF1 background is the same as that for activity in a wild type background described in (A).

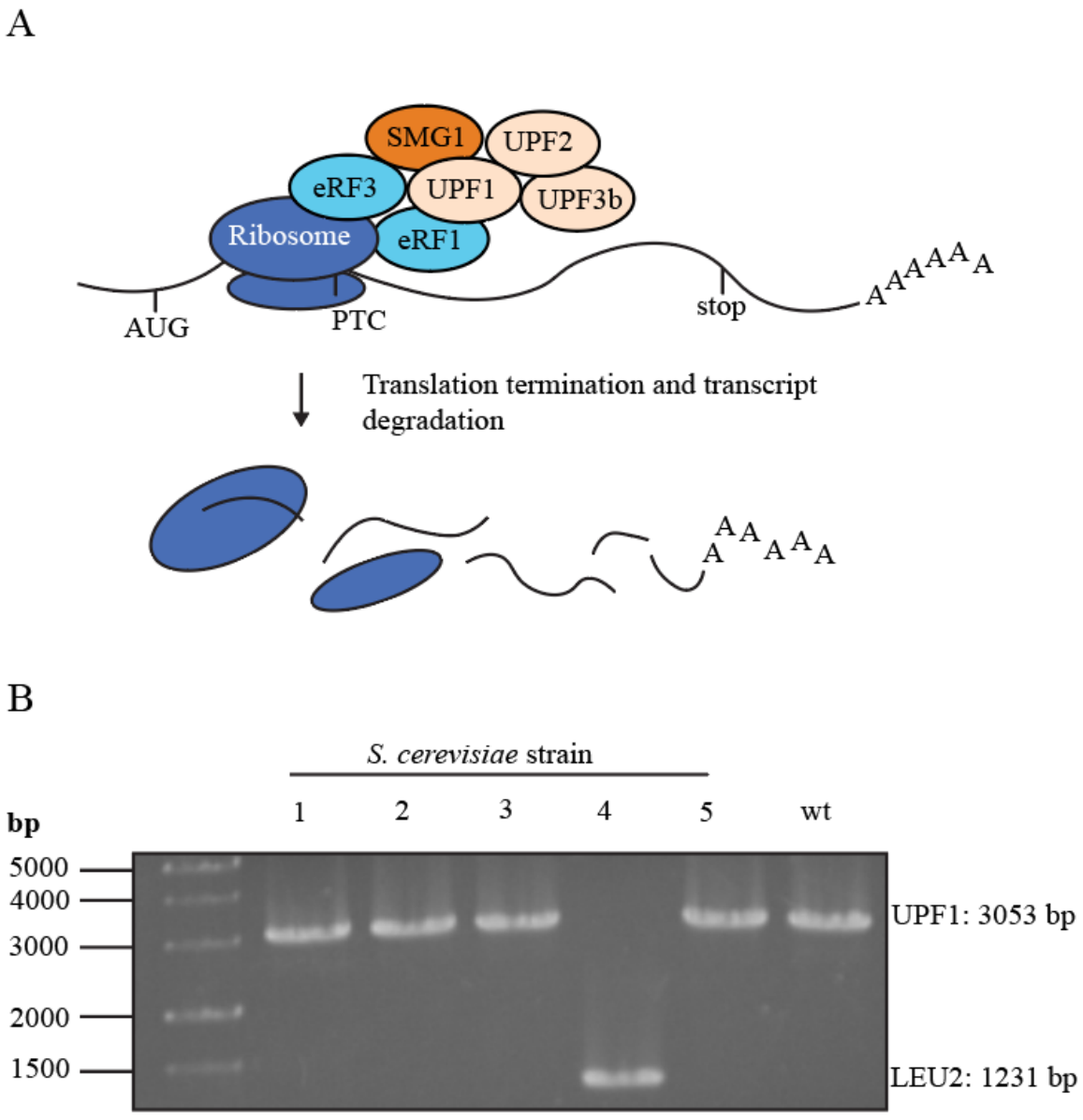


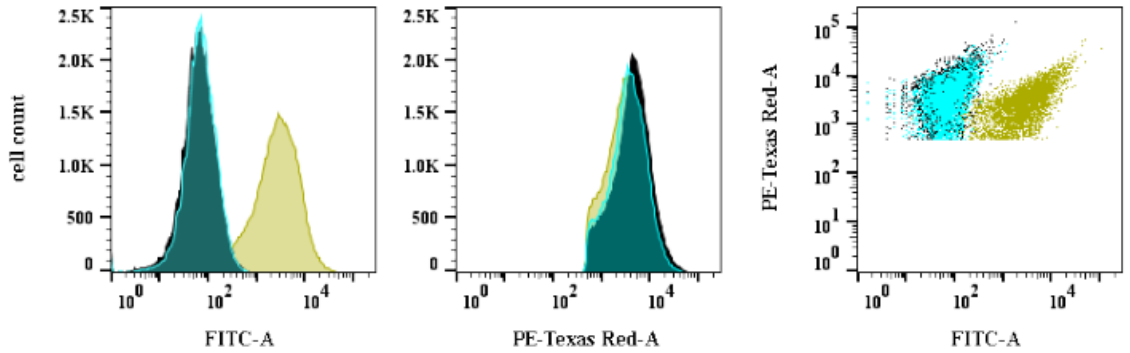
Figure 11. Creation of a ΔUPF1 knockout strain to disable the nonsense-mediated decay pathway. (A) In wild type yeast, the nonsense-mediated decay pathway (NMD) is responsible for destroying mRNAs containing premature termination codons (PTC). During the pioneering round of translation, when a ribosome encounters a PTC, eRF3 and eRF1 recruit a protein complex consisting of SMG proteins and UPF proteins, ultimately resulting in ribosome release and mRNA degradation. **(B)** To disable the NMD pathway in yeast, a *LEU2* marker was used to delete *UPF1* via insertion by homologous recombination. Successful *UPF1* deletion was verified by PCR amplification of the *UPF1* locus. A wild type *UPF1* amplicon is expected to be 3053 bp whereas insertion of the *LEU2* gene results in an amplicon of 1231 bp. 1% Agarose gel electrophoresis was used to screen 5 yeast strains for *UPF1* deletion.

4.6–Assaying for Pus7 enzyme activity at 45°C

Having not observed Pus7 activity at 30°C using the yeast reporter system, I next inquired if heat shock conditions could induce pseudouridylation of the *RPB10* target sequence by Pus7. Recapitulating the heat shock conditions implemented by Schwartz et al., Pus7-Test, positive, and negative control reporter systems in both a wild type and *ΔUPF1* yeast background were subjected to heat shock stress for 1 hour at 45°C prior to flow cytometry analysis (*7*).

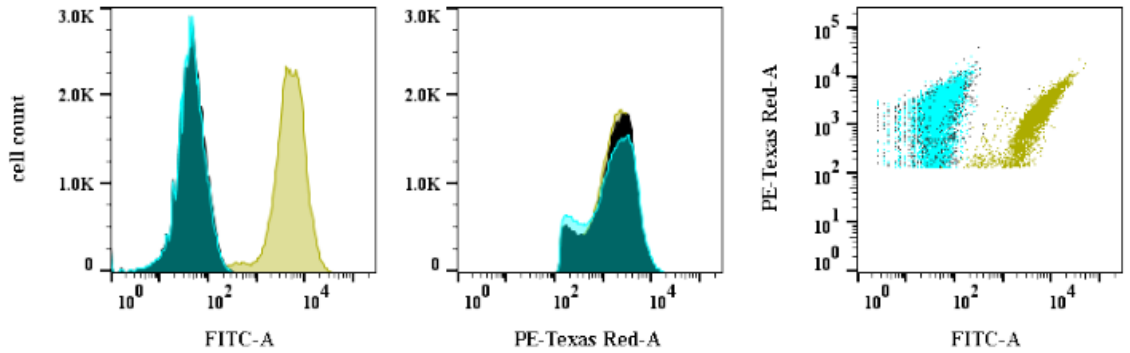
In a wild type yeast background, the Pus7-Test system gave an EGFP fluorescence histogram that was comparable to the Pus7-Neg fluorescence histogram, with respective median EGFP intensities of 70.5 a.u. and 65.5 a.u. (**Figure 12A**). In a *ΔUPF1* yeast background, the EGFP fluorescence histograms for the Pus7-Test and Pus7-Neg reporter systems were also comparable, with respective median EGFP signal intensities of 45.2 and 46.2 a.u. (**Figure 12B**). Overall, the Pus7-Test reporter system did not produce a green fluorescence shift in either a wild type or *ΔUPF1* background at 45°C. This finding suggests that either the Pus7 enzyme activity of pseudouridylating the *RPB10* recognition sequence upstream of the EGFP reporter was insufficient or that another factor limited EGFP synthesis.

A WT 45°C 1 hr



	Sample Name	Median : FITC-A	Median : PE-Texas Red-A	Percentage
	Pus7-Test	70.5	3521	80.6
	Pus7-Neg	65.5	4240	82.4
	Pus7-Pos	2709	2701	74.7

B ΔUPF1 45°C 1 hr



	Sample Name	Median : FITC-A	Median : PE-Texas Red-A	Percentage
	Pus7-Test	45.2	1685	82.3
	Pus7-Neg	46.2	1818	82.8
	Pus7-Pos	4863	1700	82.3

Figure 12. Assaying for Pus7 activity at 45°C. (A) The EGFP and Ruby2 fluorescent outputs of the Pus7-Test-WT (cyan), Pus7-Pos-WT (gold), and Pus7-Neg-WT (black) yeast strains were analyzed after 1 hr of 45°C heat shock treatment following overnight induction with galactose. The left histogram overlay gives the cell count plotted against the relative green fluorescence detected with a FITC-A filter for the test, positive, and negative control strains. The middle histogram overlay depicts the cell count vs. relative red fluorescence detected with a PE-Texas Red-A filter. The right overlay presents the red vs. green signal-expression dot plots for the test, positive, and negative control strains. Median red and green fluorescence values for each strain are given in the table alongside the percentage of each cell population that was included in the analysis after gating selection for cells with a high level of Ruby2 expression (Materials and Methods section 3.5 and

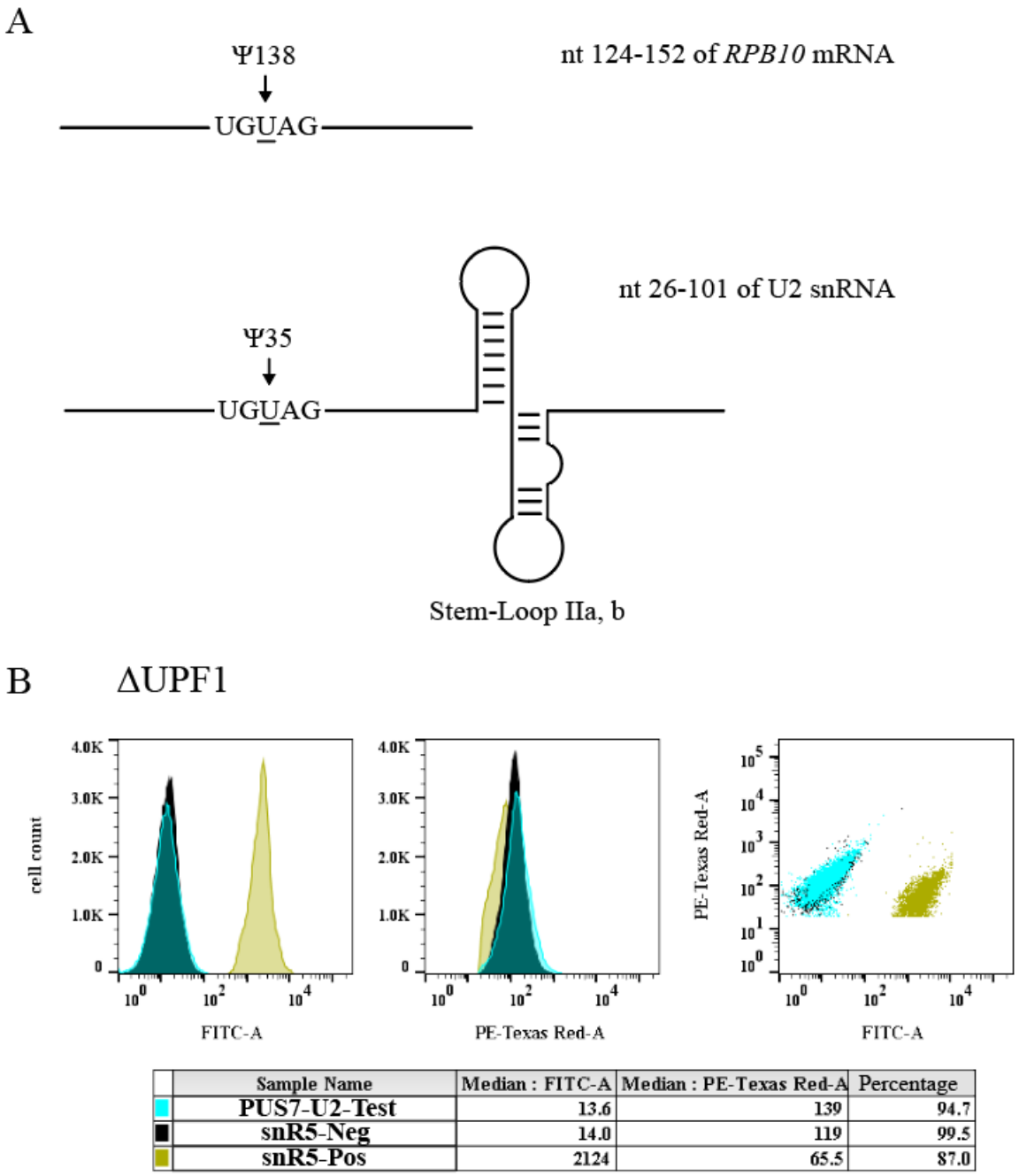
Appendix 2). **(B)** The EGFP and Ruby2 fluorescent outputs of test, positive, and negative control Pus7 reporter systems in a *ΔUPF1* yeast background heat shocked at 45°C for 1 hour prior to flow cytometry analysis. Fluorescence analysis of these strains is the same as in (A).

4.7–Investigation of RNA secondary structure requirements for Pus7 substrate recognition

The standalone pseudouridine synthase Pus7 exhibits multi-substrate specificity. In addition to modifying position 13 in cytoplasmic tRNAs, position 35 in pre-tRNA^{Tyr}, position 35 in U2 snRNA, and position 50 in 5S rRNA, Pus7 also pseudouridylates numerous positions in mRNA (*7, 83, 131*). Therefore, the UGUAR (R = G > A) recognition sequence of Pus7 occurs in a diverse array of sequence contexts across various coding and non-coding RNAs. Studies into the requirements for efficient Pus7 modification suggest that the RNA structure surrounding the target uridine is not a primary determinant for Pus7 specificity, although the presence of a stem loop structure either containing or in close proximity to the target U has been posited as a requirement for efficient Pus7 pseudouridylation (*131*). In the case of uridine 35 in the branch point recognition region of U2 snRNA, *in vitro* modification of this base by Pus7 was shown to rely on the presence of the stem-loop II region downstream of the target uridine (*132*). Therefore, I reasoned that perhaps I was unable to detect a Pus7 signal from the yeast reporter system housing the 29 nt *RPB10* mRNA sequence upstream of EGFP due to the lack of a structural element. In a further attempt to detect Pus7 activity with the yeast reporter system, I investigated the necessity of a stem loop structure for catalysis. A reporter strain was generated in a *ΔUPF1* background, Pus7-U2-Test-ΔUPF1, with sequence insert upstream of EGFP consisting of

positions 26-101 of U2 snRNA, which contains the stem-loop II region downstream of the target U35 residue (**Figure 13A**). Alongside the Pus7-U2-Test-ΔUPF1 strain, flow cytometry analysis was conducted on snR5-Pos-ΔUPF1 and snR5-Neg-ΔUPF1 strains (described in Results section 4.9), which served as positive and negative EGFP expression controls, respectively. These control strains were chosen instead of generating novel Pus7-U2-Neg-ΔUPF1 and Pus7-U2-Pos-ΔUPF1 strains because the snR5-Pos-ΔUPF1 and snR5-Neg-ΔUPF1 strains were already created and shown to reliably set the respective parameters for high and null EGFP expression at 30°C during flow cytometry analysis.

Upon flow cytometry analysis, the EGFP signal emitted by the Pus7-U2-Test-ΔUPF1 strain was comparable to that of the negative control strain: the EGFP fluorescence distributions for Pus7-U2-Test-ΔUPF1 and snR5-Neg-ΔUPF1 had median fluorescence values of 13.6 and 14.0 a.u., respectively. The measured EGFP fluorescence from the snR5-Pos-ΔUPF1 strain was considerably higher than either the test or negative control strains, producing a green fluorescence histogram with a median signal intensity of 2124 a.u. (**Figure 13B**). As demonstrated in the Ruby2 vs EGFP signal dot plot, the red and green fluorescence in each of the three yeast strains is correlated, wherein yeast exhibiting a high level of Ruby2 expression also express EGFP to a greater extent. However, in the dot plot, the Pus7-U2-Test-ΔUPF1 population comprehensively overlaps the snR5-Neg-ΔUPF1 population, thus indicating a similar level of reporter gene expression in both strains with no enhancement in the EGFP channel for the Pus7 test strain (**Figure 13B**). Therefore, a stem-loop region II-containing U2 snRNA sequence inserted upstream of the *EGFP* gene was insufficient to promote EGFP expression.



	Sample Name	Median : FITC-A	Median : PE-Texas Red-A	Percentage
	PUS7-U2-Test	13.6	139	94.7
	snR5-Neg	14.0	119	99.5
	snR5-Pos	2124	65.5	87.0

Figure 13. Assaying for Pus7 activity with a reporter system containing a U2 snRNA recognition sequence. **(A)** In addition to the Pus7 reporter system containing a *RPB10* mRNA recognition sequence, another system was created to test for Pus7 activity with a U2 snRNA recognition region. In this Pus7-U2-Test-ΔUPF1 reporter strain, the Pus7 recognition region upstream of the *EGFP* gene consists of a 5′ fragment of U2 containing the U35 target residue in addition to stem-loop region II downstream of this uridine, which is essential for Pus7 modification of this substrate (*132*). **(B)** Flow cytometry analysis of EGFP and Ruby2 expression from the Pus7-U2-Test-ΔUPF1 strain. Alongside this test

strain (cyan), snR5-Pos-ΔUPF1 (gold) and snR5-Neg-ΔUPF1 (black) strains were used as positive and negative EGFP expression controls, respectively. The left histogram overlay provides the yeast cell count vs. EGFP fluorescence from the test, positive, and negative control strains. The middle histogram overlay gives the cell count vs. Ruby2 fluorescence of each strain and the right overlay presents the red vs. green signal dot plots for the test, positive, and negative control strains. Median red and green fluorescence values for each strain are given in the table alongside the percentage of each cell population included in the analysis after selection for cells exhibiting a high level of Ruby2 expression (Appendix 2).

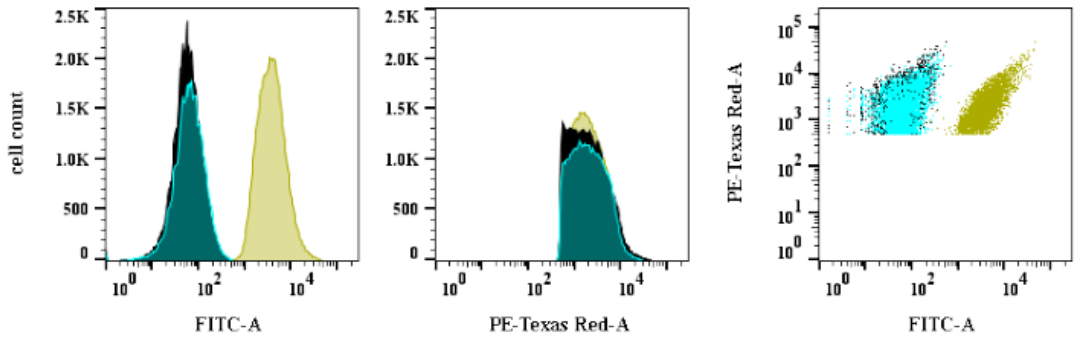
4.8–Assaying for snR81 H/ACA snoRNP activity

The snR81 H/ACA snoRNP complex was the first (and so far only) pseudouridine synthase engineered to target a premature stop codon uridine in order to promote translational readthrough (*97, 109*). This was accomplished by changing the sequence of the snR81 guide RNA's pseudouridylation pocket to recognize a stop codon in a reporter construct. In the work presented here, I sought to investigate the activity of endogenous snR81 H/ACA snoRNP by inserting its rRNA recognition sequence into the fluorescent reporter system—the enzyme's recognition sequence for Ψ1052 formation in 25S rRNA was inserted upstream of the *EGFP* reporter gene, thereby placing the target uridine in an in-frame UAA stop codon context (*128*). Subsequently, snR81 H/ACA snoRNP activity in a wild type and *ΔUPF1* yeast background was investigated by flow cytometry. As H/ACA snoRNP pseudouridylation of rRNA is constitutive, and these protein-RNA complexes have not been implicated in heat-shock induced upregulation of modification (unlike Pus7), the snR81 yeast reporter strains were studied under 30°C growth conditions.

In a wild type background, the snR81-Test system did not exhibit a distinct green fluorescence-shift relative to the snR81-Neg strain; the EGFP fluorescence histograms for

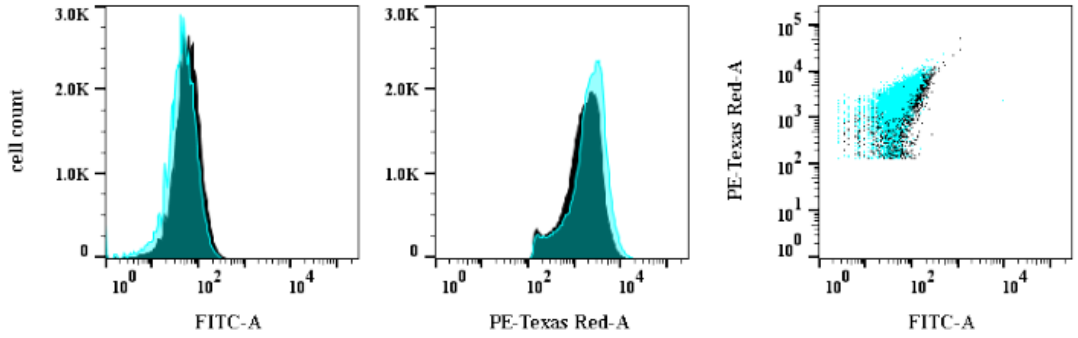
the test and negative control strains overlap with respective median EGFP intensities of 65.5 and 58.4 a.u. (**Figure 14A**). Similarly, in *ΔUPF1* yeast, the snR81-Test and snR81-Neg systems produce overlapping EGFP histograms with respective median EGFP fluorescence intensities of 46.2 and 59.2 a.u. (**Figure 14B**). Thus, significant EGFP expression reflecting snR81 H/ACA snoRNP activity was not observed.

A WT



	Sample Name	Median : FITC-A	Median : PE-Texas Red-A	Percentage
	snR81-Test	65.5	1897	62.4
	snR81-Neg	58.4	1796	72.1
	snR81-Pos	3619	1807	73.0

B ΔUPF1

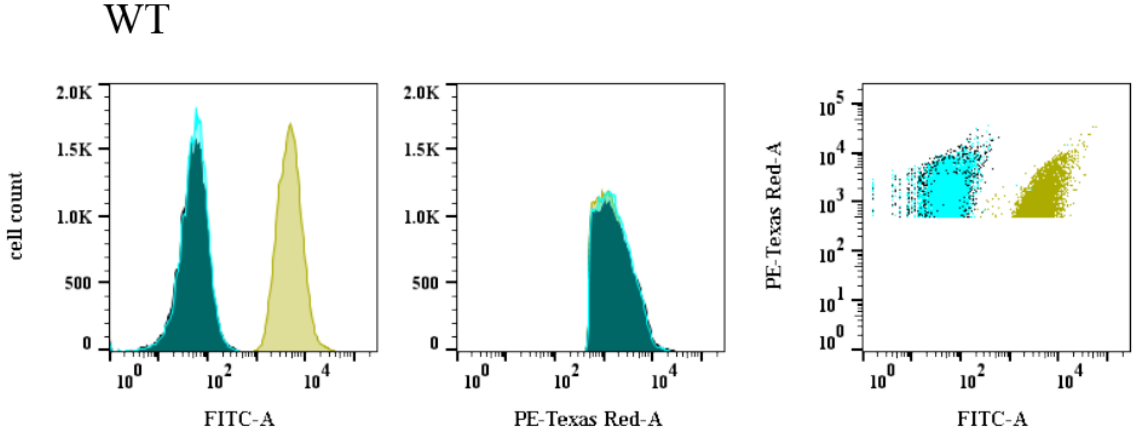


	Sample Name	Median : FITC-A	Median : PE-Texas Red-A	Percentage
	snR81-Test	46.2	2243	87.3
	snR81-Neg	59.2	1664	79.4

Figure 14. Assaying for snR81 H/ACA snoRNP activity. (A) Flow cytometry analysis of EGFP and Ruby2 fluorescent signal outputs from snR81-Test (cyan), snR81-Pos (gold), and snR81-Neg (black) reporter systems expressed in a wild type yeast background. The left histogram overlay shows yeast count vs. EGFP signal intensity for the test, positive, and negative control strains. The middle histogram overlay shows yeast count vs. Ruby2 signal intensity. The right overlay gives the Ruby2 vs. EGFP signal dot plots for the test, positive, and negative control strains. Median Ruby2 and EGFP fluorescence values for each strain are given in the table alongside the percentage of each yeast population that was included in the analysis after selection for yeast cells with a high level of Ruby2 expression (Appendix 2). **(B)** The fluorescent outputs of a snR81-Test and snR81-Neg system in a *ΔUPF1* yeast background were analyzed as in (A). Due to experimental challenges, a snR81-Pos-ΔUPF1 yeast strain was not included in this data set.

4.9–Assaying for snR5 H/ACA snoRNP activity

In addition to snR81, I created yeast reporter strains to detect the activity of another endogenous H/ACA snoRNP: snR5. To program the yeast reporter system to detect the modifying activity of an snR5 H/ACA snoRNP, this enzyme's recognition sequence for Ψ1004 formation in 25S rRNA was inserted upstream of the *EGFP* reporter gene, such that the target uridine occurs in an in-frame UGA stop codon context (*25*). The ability of snR5 to rescue EFGP expression was subsequently tested by flow cytometry. In a wild type yeast background, the snR5-Test reporter system emits an EGFP signal comparable to that of the snR5-Neg reporter system; the EGFP signal intensity histograms of the snR5-Test-WT and snR5-Neg-WT strains have median values of 55.9 and 53.4 a.u., respectively (**Figure 15**).



	Sample Name	Median : FITC-A	Median : PE-Texas Red-A	Percentage
	snR5-Test	55.9	1469	52.8
	snR5-Neg	53.4	1473	50.4
	snR5-Pos	4575	1345	49.5

Figure 15. Assaying for snR5 H/ACA snoRNP activity in a wild type yeast background. EGFP and Ruby2 fluorescence from snR5-Test-WT (cyan), snR5-Pos-WT (gold), and snR5-Neg-WT (black) yeast strains were analyzed by flow cytometry. The left histogram overlay presents the cell count plotted against EGFP fluorescence for the test, positive, and negative control strains. The middle histogram overlay gives the cell count plotted against Ruby2 fluorescence for the yeast strains. The right overlay presents the Ruby2 vs. EGFP signal dot plots for the test, positive, and negative control strains. Median red and green fluorescence values for each strain are given in the table alongside the percentage of each cell population that was included in the analysis after gating for yeast with high Ruby2 gene expression (Appendix 2).

When the snR5-Test, Positive, and Negative control systems were evaluated in a *ΔUPF1* yeast strain, I observed an increased green fluorescent signal from the snR5-Test system relative to the snR5-Neg control system. In an initial flow cytometry assay, the median EGFP signal intensity for the snR5-Test system was 74.9 a.u. compared to the snR5-Neg system's median intensity of 48.8 a.u. thus indicating a median EGFP signal increase of 26.1 a.u. in the snR5-Test strain relative to the snR5-Neg strain. To further quantify the difference in EGFP expression between the Test and Negative control strains, the EGFP fold change was calculated by dividing the median EGFP intensity of the snR5-Test strain by that of the snR5-Neg strain. In this assay, the EGFP signal was 1.5-fold greater in the snR5-Test strain than in the snR5-Neg strain (snR5-ΔUPF1 Replicate 1, **Figure 16** and **Table 8**).

Having observed a green fluorescence-shift from the snR5-Test reporter system, subsequent technical replicate assays were conducted with this reporter strain to ascertain whether the observed green signal increase was reproducible. In a second flow cytometry experiment, the median EGFP intensity of 27 a.u. for the snR5-Test strain was almost two-fold higher than the median EGFP signal for the snR5-Neg strain of 14 a.u., with a EGFP fold change of 1.9 for the test strain (snR5-ΔUPF1 Replicate 2, **Figure 16** and **Table 8**). In a third assay, the median values of the EGFP fluorescence histograms for the snR5-Test and snR5-Neg strains were comparable, although the snR5-Test strain was slightly higher, with respective values of 8.75 and 7.51 a.u., and a modest EGFP fold change of 1.2 (snR5-ΔUPF1 Replicate 3, **Figure 16** and **Table 8**).

To assess the impact of induction time on the fluorescent output of the snR5 reporter system, the snR5-Test-ΔUPF1, snR5-Neg-ΔUPF1, and snR5-Pos-ΔUPF1 strains

were submitted to flow cytometry analysis 3 hours after induction with galactose, whereas the three previous technical replicate assays were conducted after overnight galactose induction. In the 3-hour post-induction experiment, the median EGFP signal of 114 a.u. from the snR5-Test fluorescence histogram was 26 units greater than the median EGFP signal of 88 a.u. from the snR5-Neg fluorescence histogram, with an EGFP fold change of 1.3 (snR5-ΔUPF1 3-hour induction, **Figure 16** and **Table 8**).

Across the three flow cytometry assays of the overnight galactose induction technical replicates, the median EGFP signal from the snR5-Test-*ΔUPF1* strain was 1.20 ± 0.06 % that of the snR5-Pos-*ΔUPF1* strain whereas the snR5-Neg-*ΔUPF1* median EGFP signal was 0.81 ± 0.13% of the snR5-Pos-*ΔUPF1* strain. For the flow cytometry measurements of the snR5 reporter strains 3-hours post-induction, the median EGFP signal from the snR5-Test-ΔUPF1 strain was 1.00% that of the snR5-Pos-ΔUPF1 strain and the median snR5-Neg-ΔUPF1 EGFP signal was 0.77% that of the snR5-Pos-ΔUPF1 strain. Overall, these assays demonstrate that the snR5-Test system in a *ΔUPF1* background exhibits a reproducible green-fluorescence shift relative to the negative control strain. As the main difference compared to the previously described systems is the presence of an snR5 recognition sequence targeting the premature stop codon upstream of EGFP, it is likely that the observed increase in EGFP expression reflects the pseudouridylation activity of the snR5 H/ACA snoRNP.

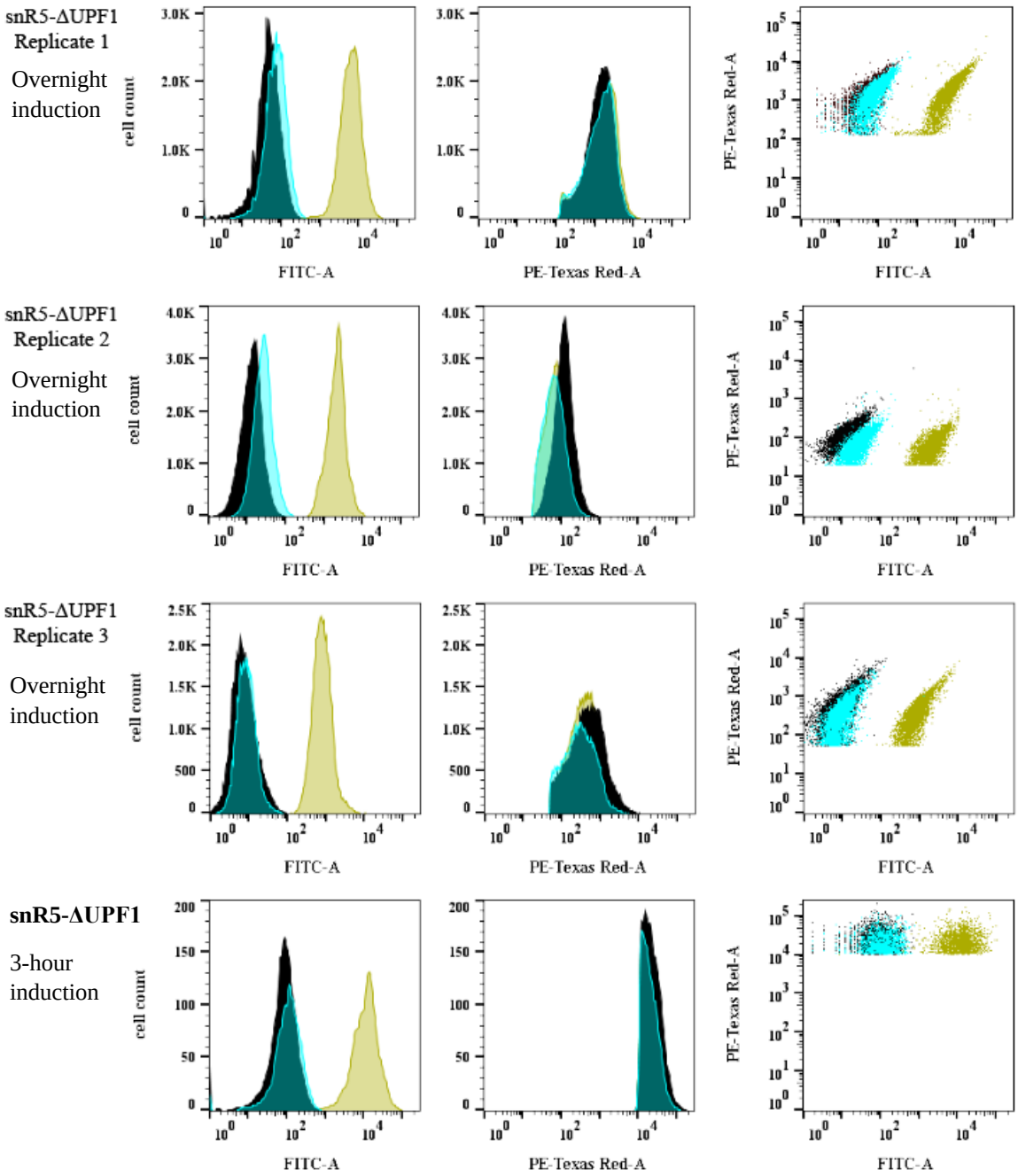


Figure 16. snR5 H/ACA snoRNP activity in ΔUPF1 yeast. Three flow cytometry assay replicates were performed on the snR5-Test (cyan), snR5-Pos (gold), and snR5-Neg (black) constructs in a *ΔUPF1* yeast background following overnight induction with galactose. Additionally, a flow cytometry assay was conducted on the snR5-Test-ΔUPF1, snR5-Pos-ΔUPF1, and snR5-Neg-ΔUPF1 yeast strains three hours post-galactose induction. For each experiment, the left histogram overlay presents the yeast cell count vs. EGFP fluorescence for the test, positive, and negative control strains. The middle histogram overlay gives cell count plotted against the Ruby2 fluorescence of each strain, and the right overlay presents the red vs. green signal dot plots for the test, positive, and negative control strains.

Table 8. Red and green signal intensities from the flow cytometry analysis of a fluorescence-based yeast reporter strain detecting activity of the snR5 H/ACA snoRNP. Median EGFP and Ruby2 signal intensities from three technical replicate flow cytometry assays of the snR5-Test-ΔUPF1, Pos-ΔUPF1, and Neg-ΔUPF1 strains following overnight galactose induction in addition to the median EGFP and Ruby2 signal intensities from the flow cytometry assay conducted on these strains 3-hours post-galactose induction. The EGFP fold change is the quotient of snR5-Test EGFP signal/snR5-Neg EGFP signal. Also given are the percentages of each cell population included in the fluorescence analysis after gating for yeast exhibiting a high level of Ruby2 expression (Appendix 2).

	Flow cytometry experiment	Median EGFP	Median Ruby2	Yeast Percentage Analyzed	EGFP fold change
	snR5-ΔUPF1 Replicate 1 (Overnight induction)				
	snR5-Test	74.9	1473	81.4	1.5
	snR5-Neg	48.8	1420	87.6	
	snR5-Pos	6167	1758	78.7	
	snR5-ΔUPF1 Replicate 2 (Overnight induction)				
	snR5-Test	27.0	65.1	92.4	1.9
	snR5-Neg	14.0	119	99.5	
	snR5-Pos	2124	65.5	87.0	
	snR5-ΔUPF1 Replicate 3 (Overnight induction)				
	snR5-Test	8.75	294	53.4	1.2
	snR5-Neg	7.51	459	70.3	
	snR5-Pos	775	373	67.5	
	snR5-ΔUPF1 3-hour induction				
	snR5-Test	114	17909	3.97	1.3
	snR5-Neg	88.0	20292	5.32	
	snR5-Pos	11375	19092	4.33	

Western blotting against EGFP in the lysates of the snR5-Test-ΔUPF1, Pos, and Neg strains was employed in an attempt to corroborate the slight increase of EGFP expression evinced by flow cytometry analysis of the snR5-Test-ΔUPF1 reporter strain. This blotting assay readily detected a purified GFP control at ~26 kDa (corresponding to an anticipated size of 27 kDa) present in quantities of 79.2 and 19.8 pmol, respectively (**Figure 17**) (*133*). However, at the initial 5 second membrane exposure at which the purified GFP control was visualized, there was little to no signal from the yeast reporter strain lysates. A subsequent 2-minute overexposure of the membrane revealed a ~26 kDa signal in the snR5-Pos-ΔUPF1 lysate, but did not manifest a discernible EGFP signal in the snR5-Test-ΔUPF1 and snR5-Neg-ΔUPF1 lysates (**Figure 17**). Presumably, there is also an absence of EGFP signal from the ΔUPF1 lysate (lacking a fluorescence reporter system), but spillover from the snR5-Pos-ΔUPF1 lane into the adjacent ΔUPF1 lane makes this determination inconclusive (**Figure 17**). Although EGFP protein could not be observed in the snR5-Test-ΔUPF1 lysate upon Western blot analysis, the fold changes in EGFP expression determined by flow cytometry were subtle (1.2-1.9-fold enrichment in snR5-Test-ΔUPF1 EGFP over the negative control). Moreover, across three flow cytometry technical replicates, the median EGFP signal from the snR5-Test-ΔUPF1 strain was merely 1.20 ± 0.06% that of the snR5-Pos-ΔUPF1 strain. Therefore, the level of EGFP produced by the snR5-Test reporter system may be below the detection threshold of the Western blotting assay.

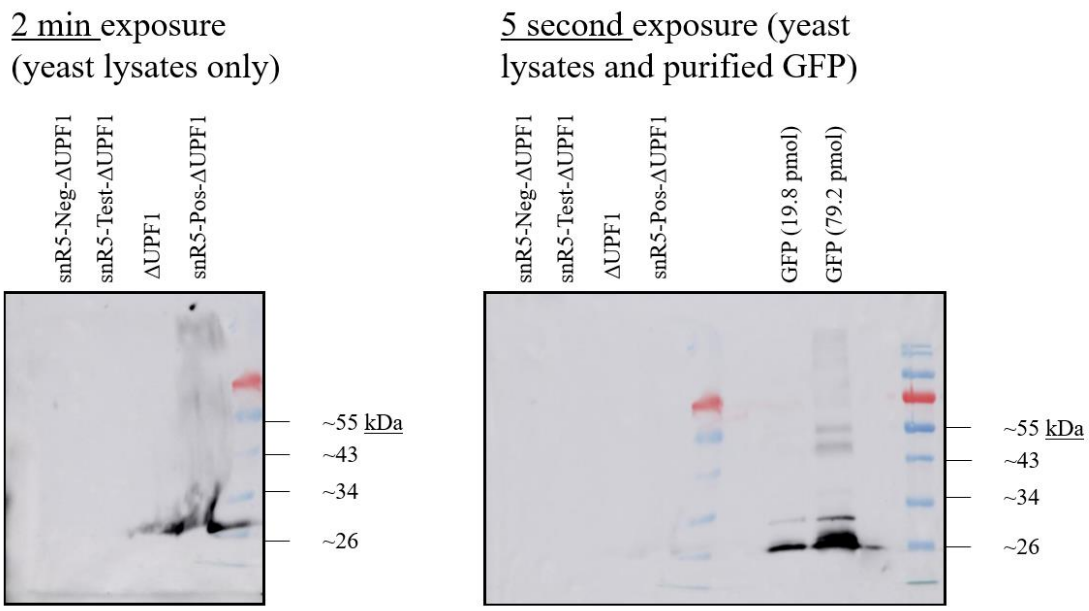


Figure 17. Western blotting against EGFP in the snR5 reporter system. Chemiluminescent detection of EGFP expressed in the snR5-Test-ΔUPF1, snR5-Pos-ΔUPF1, and snR5-Neg-ΔUPF1 reporter strains, in addition to a non-fluorescent *ΔUPF1* strain. Green fluorescent protein has a size of approximately 26 kDa. Blotting was implemented with α-GFP 1° antibody and α-rabbit 2° antibody conjugated to horse radish peroxidase. Purified GFP in quantities of 79.2 and 19.8 pmol was used as a positive blotting control. A 5-second and a 2-minute exposure were implemented to visualize the membrane.

4.10–Nonsense suppression in the yeast reporter system with SUP4 and SUP53 tRNAs

After developing the fluorescence-based yeast reporter system to detect the modifying activity of the snR5 and snR81 H/ACA snoRNPs and standalone Pus7, only the snR5 reporter system manifested a detectable green fluorescent signal upon induction with galactose. To investigate an additional means of achieving stop codon readthrough upstream of *EFGP* in the yeast reporter system, expression constructs encoding suppressor 4 (SUP4) and suppressor 53 (SUP53) tRNAs were transformed into yeast reporter strains which previously did not manifest a green signal. Specifically, the UAA suppressor SUP4 was expressed in the snR81-Test-ΔUPF1 strain (possessing an in-frame UAA stop codon upstream of EGFP) and the UAG suppressor SUP53 was expressed in the Pus7-U2-Test

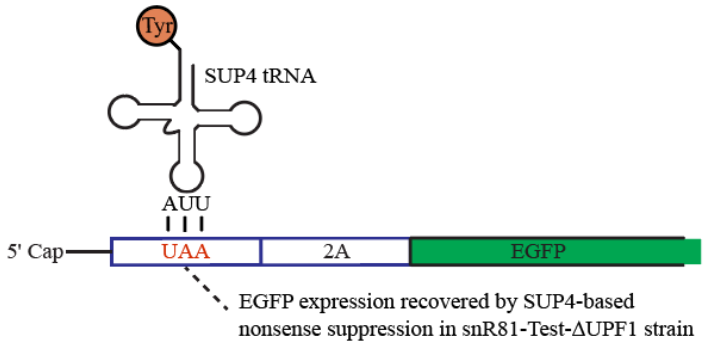
strain, which contains a UAG stop codon upstream of the *EGFP* reporter gene (*119, 120*). In the resultant yeast strains, designated snR81-Test-Sup4 and Pus7-U2-Test-Sup53, respectively (**Table 6**), it was hypothesized that the suppressor tRNAs would be able to recognize the stop codons in the pseudouridine synthase recognition sequences upstream of the *EGFP* reporter gene, thereby granting translational readthrough and rescuing EGFP expression (**Figure 18A** and **C**). As a negative control for SUP4-based UAA stop codon readthrough in the reporter system, the SUP4 tRNA expression construct was transformed into the Pus7-U2-Test strain harbouring an incompatible UAG stop codon upstream of EGFP. This strain was designated Pus7-U2-Test-Sup4 (**Table 6**). Likewise, the SUP53 tRNA expression construct was transformed into the snR81-Test-ΔUPF1 strain (harbouring an incompatible UAA stop signal) to create a negative control strain designated snR81-Test-Sup53 (**Table 6**). As a positive control for EGFP expression in these nonsense suppression assays, the previously examined snR5-Pos-ΔUPF1 strain was utilized.

Flow cytometry analysis of the snR81-Test-Sup4 strain yielded an increased EGFP intensity relative to the Pus7-U2-Test-Sup4 (negative control) strain. This change is evinced in the median EGFP value of 22.8 a.u. for the snR81-Test-Sup4 histogram compared to the median value of 4.53 a.u. for the Pus7-U2-Test-Sup4 histogram (**Figure 18B**). Out of all the manifestations of the fluorescence-based yeast reporter system analyzed throughout this thesis, the snR81-Test-Sup4 strain exhibited the greatest green fluorescence shift relative to the negative control strain, with an EGFP fold-change of 5.0. Thus, the SUP4 expression construct, when transformed into the snR81-Test-ΔUPF1 strain, serves as a useful positive control and benchmark for nonsense suppression in the

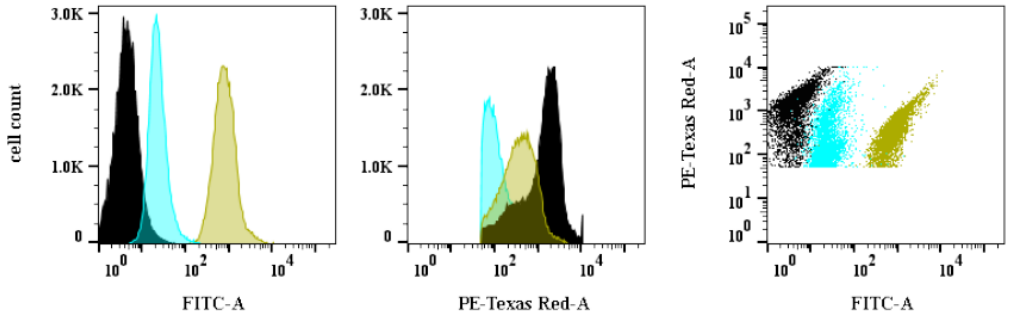
yeast reporter system. The outcome of the SUP53 trials was not as expected—the Pus7-U2-Test-Sup53 yeast strain (with a compatible UAG stop codon-anticodon pairing) exhibited a median EGFP fluorescence of 4.32 a.u. whereas the negative control strain, snR81-Test-Sup53 with an incompatible UAA stop codon-anticodon pairing, yielded a median EGFP intensity of 10.8 a.u. (**Figure 18D**). Therefore, the Pus7-U2-Test-Sup53 strain anticipated to generate an enhanced EGFP signal trace in actuality had 2.5-fold less median EGFP fluorescence than the negative control.

Here, it should be noted that, in addition to this anomalous result, the Ruby2 signal profiles in both the SUP53 and SUP4 flow cytometry experiments exhibited considerable variation between the test and control strains: something not observed in the yeast reporter systems probing for pseudouridine synthase activity (**Figure 18B** and **D**). Since the Ruby2 signal traces serve as a control for fluorescent protein expression from the reporter system, their fluctuation in these suppressor tRNA trials suggests an influence on expression of the reporter system beyond the intended nonsense suppression of the stop signal upstream of the *EGFP* reporter gene. One possible explanation is that the suppressor tRNAs are not just targeting the stop codons situated upstream of *EGFP* but are rather indiscriminately conferring nonsense suppression to other UAA or UAG stop codons present in off-target (i.e. endogenous) mRNAs, thus leading to aberrant gene products with C-terminal extensions that could harm the cell via loss or gain of function. Supporting this hypothesis is the fact, in the lead-up to flow cytometry analysis, the yeast strains transformed with the suppressor tRNA constructs exhibited a slow growth phenotype compared to the snR5-Pos-ΔUPF1 strain.

A

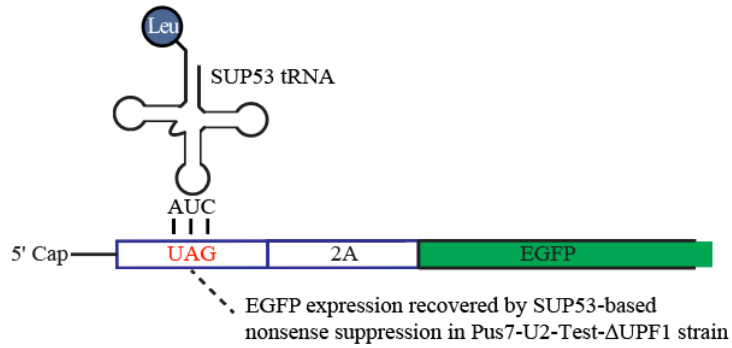


B

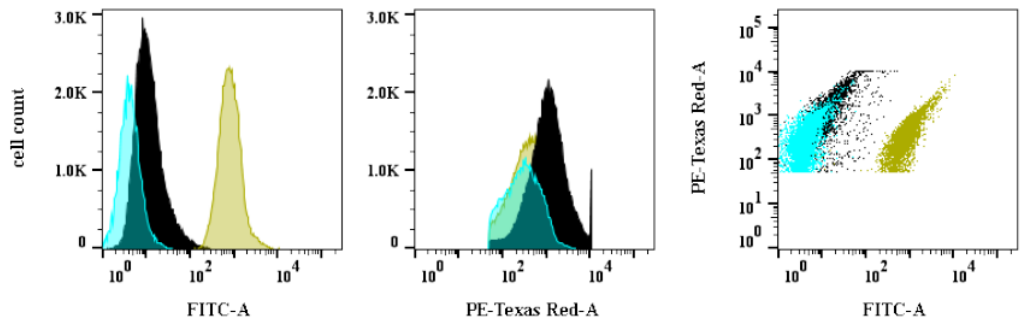


	Sample Name	Median : FITC-A	Median : PE-Texas Red-A	Percentage
	snR81-Test-SUP4	22.8	122	63.3
	Pus7-U2-Test-SUP4	4.53	1469	85.9
	snR5-Pos	775	373	67.5

C



D



	Sample Name	Median : FITC-A	Median : PE-Texas Red-A	Percentage
	Pus7-U2-Test-SUP53	4.32	298	55.7
	snR81-Test-SUP53	10.8	1060	92.2
	snR5-Pos	775	373	67.5

Figure 18. Nonsense suppression in the fluorescence-based yeast reporter system with SUP4 and SUP53 tRNA expression constructs. (A) As an alternative means to achieve stop codon readthrough in the fluorescence-based reporter system, an expression plasmid encoding a suppressor 4 (SUP4) tRNA gene was transformed into the snR81-Test-ΔUPF1 strain, thus creating the snR81-Test-Sup4 strain in which translation termination at the UAA stop codon upstream of EGFP could be suppressed by SUP4. **(B)** Flow cytometry detecting EGFP and mRuby2 expression in yeast strains expressing the SUP4 tRNA. Alongside the snR81-Test-Sup4 strain (cyan), a negative control strain (black) was generated by transforming the SUP4 expression plasmid into the Pus7-U2-Test-ΔUPF1 strain with an incompatible (UAG) codon upstream of *EGFP*. The snR5-Pos-ΔUPF1 strain was used a positive green fluorescence control (gold). The left histogram overlay gives the cell count vs. EGFP fluorescence for each strain tested. The middle histogram overlay gives the cell count vs. Ruby2 fluorescence for each strain, and the right overlay presents the Ruby2 vs. EGFP signal dot plots for the suppressor tRNA strains and the snR5-Pos-ΔUPF1 control. Median red and green fluorescence values for each strain are given in the table alongside the percentage of each cell population that was included in the analysis after gating for cells with a high level of Ruby2 expression (Appendix 2). **(C)** Nonsense suppression of the UAG stop codon upstream of *EGFP* in the Pus7-U2-Test-ΔUPF1 strain was investigated by transforming this strain with a SUP53 amber suppressor tRNA expression construct, thus creating the Pus7-U2-Test-Sup53 reporter strain. **(D)** Flow cytometry analysis of Ruby2 and EGFP fluorescence in the Pus7-U2-Test-Sup53 (cyan), snR81-Test-Sup53 (black; negative control strain with an incompatible stop codon/anticodon pairing), and snR5-Pos-ΔUPF1 (gold) strains. The fluorescence analysis methodology for these strains is the same as stated in (B).

Chapter 5–Discussion

The goal of this research was to construct a fluorescence-based yeast reporter system to detect pseudouridine synthase activity *in vivo*. This reporter system was configured by inserting a substrate recognition sequence for a pseudouridine synthase of interest upstream of an *EGFP* reporter gene wherein a premature termination codon uridine was the target residue in the recognition element. Upon pseudouridylation of this target uridine, the modified stop codon can then facilitate translational readthrough, allowing for EGFP expression. Thus, the emission of green fluorescence from the yeast reporter strain is directly linked to the modifying activity of the pseudouridine synthase of interest. Reporter system variants were constructed to assay for the activities of the standalone pseudouridine synthase Pus7 in addition to the snR5 and snR81 H/ACA snoRNPs. The only pseudouridine synthase reporter system variant to demonstrate a green-fluorescence shift relative to a negative control strain was the snR5 H/ACA snoRNP reporter system, with an EGFP fold change of 1.2-1.9 over the negative control (**Figure 16** and **Table 8**). Therefore, the fluorescence-based yeast reporter system is operational, although a relatively low activity signal is emitted by the system. By contrast, a positive control system for stop codon readthrough using a Sup4 suppressor tRNA to target the UAA stop codon upstream of *EGFP* yielded an EGFP fold change of 5.0 over a negative control strain (**Figure 18A** and **B**). Overall, this novel reporter system is a promising platform for studying mRNA pseudouridylation that now requires further system optimizations to enhance signal output as well as further validations to check that EGFP signal directly reflects pseudouridylation activity.

5.1–Parameters governing fluorescent output of the yeast reporter system

The level of EGFP fluorescence emitted by the yeast reporter system is low—in the only reporter system prototype that elicited a green signal upon induction (snR5-Test-*ΔUPF1*), a modest EGFP fold change of 1.2-1.9 was detected by flow cytometry (**Figure 16 and Table 8**). There are at least three biophysical parameters that can account for this low level of signal output (**Figure 19**). The first is the efficiency of pseudouridine synthase recognition. Once the *EGFP* reporter gene has been transcribed, a pseudouridine synthase, either the snR5 and snR81 H/ACA snoRNPs or Pus7, must bind its recognition sequence upstream of the *EGFP* CDS and isomerize the stop codon uridine in order for EGFP translation to occur (**Figure 19**). Most likely, this recognition and subsequent pseudouridylation is not 100% efficient. In the case of rRNA pseudouridylation, although many positions are modified to a high extent (>85%), some, including Ψ1004 targeted by snR5, are only partially modified (*134*). Moreover, the pseudouridylation stoichiometries of mRNA have been demonstrated as variable; although the high-throughput sequencing methods for pseudouridine detection cannot determine the absolute modification extent at a specific position, the relative stoichiometry can be. Ψ-seq analysis reveals that some mRNA pseudouridylation sites are modified to an extent comparable to those of rRNA, while others exhibit a lower relative stoichiometry (*7*). Additionally, use of the quantitative SCARLET method to validate a putative mRNA Ψ site, U519 of *EEF1A1* mRNA, revealed that it is ~56% modified (*47*). Moreover, the number of EGFP transcripts produced by the reporter system could outnumber the available pseudouridine synthases in the cell nucleus which are already modifying their canonical substrates in addition to the reporter transcripts. Consequently, even undergoing multiple rounds of substrate turnover, the

pseudouridine synthases could be a limiting factor for EGFP transcript pseudouridylation. Therefore, out of the total population of EGFP reporter transcripts in a yeast cell, only a portion of them will be recognized by their cognate pseudouridine synthase and have the stop codon-situated uridine isomerized to pseudouridine, thus limiting the amount of subsequent readthrough and EGFP expression in the yeast reporter strain.

The second system parameter that must be considered is the efficiency of translational readthrough of a pseudouridylated stop codon (**Figure 19**). *In vitro* translation of a reporter transcript harbouring a stop codon uridine that is 100% pseudouridylated yields a readthrough efficiency of ~74% (*97*). Thus, out of the proportion of *EGFP* reporter transcripts that are modified by a pseudouridine synthase, translational readthrough of the modified stop signal in these transcripts will only occur approximately three quarters of the time, resulting in a reduction of *EGFP* reporter gene expression relative to that of a positive control strain that possesses a sense codon instead of a modified stop signal upstream of *EGFP*.

The third parameter of the fluorescence-based yeast reporter system that could affect the final level of EGFP expression is the 2A self-cleaving peptide sequence immediately upstream of the *EGFP* reporter gene (**Figure 6B, Figure 19**). The 2A sequence, also known as a cis-acting hydrolase element, was included in the reporter system design to enhance the green signal output of the yeast by inducing a translational pausing event after the ribosome reads through the pseudouridylated stop codon and before it encounters the *EGFP* CDS. This pause on the 2A element causes the ribosome to release the translated pseudouridine synthase recognition sequence before unpausing and continuing to translate EGFP. This phenomenon has been termed “ribosome skipping”

(*135*). In this manner, an N-terminal fusion tag on the EGFP protein that could interfere with folding and subsequent fluorescence is avoided. However, the "self-cleavage" mechanism of the 2A peptide is not perfectly efficient. Previous characterization of the 2A element used here in the fluorescent reporter system (a 2A element isolated from foot and mouth disease virus; FMDV) found that this sequence in a polycistronic transcript yields a ~2-7 stoichiometric excess of the product N-terminal of the 2A element over the product C-terminal to the 2A element. (*127*). Furthermore, characterization of 2A elements derived from porcine teschovirus-1 (P2A) and thosea asigna virus (F2A) in a bi-cistronic construct reveals a substantial decrease in protein expression at the second position (downstream of the 2A sequence) relative to the first position upstream of 2A owing to the fact that successful ribosome skipping and translation recommencement upon encountering the 2A element was calculated to occur ~30% of the time when using the P2A and F2A variants of the cis-acting hydrolase element. Conversely, ribosome skipping with subsequent translational fall off was determined to occur ~60% of the time with ribosome readthrough of the 2A element occurring during the remaining 10%. (*135*). Thus, in the context of the yeast reporter system using the FMDV 2A element, a significant amount of EGFP expression is potentially being lost during translation of the 2A element. In the case of the positive control strains wherein a sense codon takes the place of the premature stop codon in the test strains, the loss of EGFP from the 2A sequence still results in a high level of fluorescent protein expression as revealed through flow cytometry indicating that in principle the 2A sequence does allow EGFP expression (**Figures 10, 12, 13, 14, 15, and 16**). However, as EGFP expression in the test strain is already contingent upon pseudouridine synthase recognition and translational readthrough of a pseudouridylated

stop signal (two events which are not perfectly efficient), the loss of EGFP expression from ribosome fall off on the 2A element could reduce the signal below the detection limit of the flow cytometry assay. Therefore, inefficient processing of the 2A sequence could be a factor in the apparent absence of enzyme activity in the Pus7 and snR81 H/ACA snoRNP reporter systems and preclude a possibly higher signal output from the snR5 H/ACA snoRNP system. Although substantial progress has been made in characterizing 2A elements derived from various sources, selecting an optimal 2A element for use in the yeast reporter system is complicated by the absence of comprehensive knowledge of cleavage efficiencies by different 2A elements in various cell lines (*135*-*137*). Nevertheless, a 2A element derived from a virus other than FDMV with an experimentally demonstrated higher "cleavage" efficiency could be introduced into the yeast reporter system in an attempt to increase EGFP reporter expression. Alternatively, at the risk of including an N-terminal fusion tag on EGFP that may interfere with folding, maturation, and fluorescence, it may be propitious to create a version of the fluorescence-based yeast reporter system that lacks a 2A peptide element upstream of the *EGFP* reporter gene.

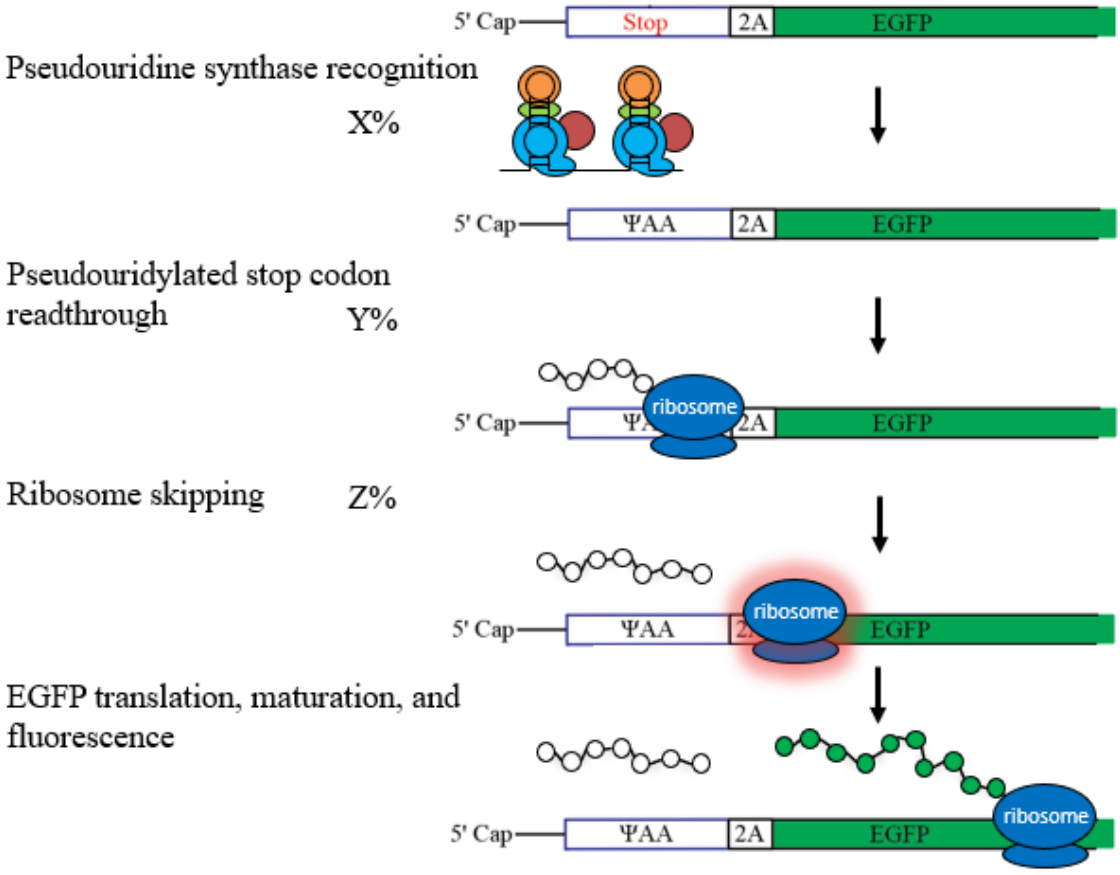


Figure 19. Parameters affecting fluorescent output of yeast reporter system. Following galactose induction and transcription of the *EGFP* reporter gene, three distinct biochemical processes must occur in order for EGFP to be fully translated. First, the stop codon-situated uridine upstream of the EGFP CDS must be targeted and isomerized by the pseudouridine synthase of interest. Second, once pseudouridylated, the premature stop codon must be read-through during translation elongation instead of causing premature termination. Lastly, after bypassing the modified stop signal and before encountering the EGFP CDS, the elongating ribosome must undergo a skipping event upon translation of a 2A cis-acting hydrolase sequence. During this skipping event, the nascent peptide resulting from translation of the pseudouridine synthase recognition sequence is released before translation of EGFP resumes. Once produced, EGFP must undergo folding, chromophore maturation, and eventually fluoresce upon excitation with 488 nm light. The efficiency with which each of these three concerted steps occurs is most likely less than 100% (denoted by X, Y, and Z% for pseudouridine synthase recognition, pseudouridylated stop codon readthrough, and ribosome skipping, respectively), and as such, a certain amount of potential EGFP expression is lost during each step.

In order to determine the respective contributions of stop codon pseudouridylation efficiency, translation readthrough efficiency, and 2A peptide "cleavage" efficiency to the function of the yeast reporter system, and to test the hypothesis that an absence of targeted pseudouridylation is the reason for the lack of EGFP expression in the Pus7 and snR81 reporter systems, three forms of experimental validation must be implemented. First, to determine the extent of stop codon pseudouridylation (if any) of the *EGFP* reporter transcripts produced by the yeast reporter system, the SCARLET method could be applied. After RNA extraction and digestion and specific ^{32}P labelling of the target uridine/pseudouridine, resolution on a TLC plate will then unveil the modification extent at this site (refer to Chapter 1.3.2. for a full summary of this method). The extent of readthrough of the pseudouridylated stop codon could be assessed by combining ribosome profiling and pseudouridine sequencing focussing specifically on the EGFP transcript. Lastly, to ascertain the influence of the FMDV 2A element on EGFP translation, a variant of the reporter system lacking this element can be generated so that its green signal output may be compared to the current 2A element-containing system. Combined, these experiments will allow estimating pseudouridylation efficiency, readthrough efficiency, and 2A element processing efficiency and their contribution to the final EGFP signal output of the reporter system thereby giving further insight into future system optimizations (see Chapter 5.6 for proposed system optimizations).

In addition to the three parameters discussed above, it should be noted here that one obstacle to fluorescence-expression in the yeast reporter system was successfully overcome, namely, the nonsense mediated decay pathway responsible for degradation of premature stop codon-containing mRNAs (*138*). As evidenced by the flow cytometry

experiments in this thesis, a green signal from the snR5-Test reporter system was only observed upon knockout of the NMD pathway gene *UPF1* (**Figures 11, 15 and 16**). Therefore, in a wild type yeast background, the NMD response degrades a large enough proportion of the *EGFP* reporter transcripts before they can be successfully translated that no green signal is detected by flow cytometry; by knocking out the *UPF1* gene critical to this pathway, the NMD response is abolished and the reporter system generates a fluorescent output above the detection limit.

Aside from the genetic composition of the yeast reporter system, a final parameter that could possibly affect fluorescent output is the time of galactose induction prior to flow cytometry analysis. To investigate the impact of induction time on the signal output of the snR5-Test-ΔUPF1, snR5-Neg-ΔUPF, and snR5-Pos-ΔUPF1 reporter strains, flow cytometry measurements of these strains were collected after both an overnight induction with galactose and 3-hours post-galactose induction (**Figure 16**). By doing this, the intent was to determine whether snR5-reporter system cells analyzed in the exponential phase of growth (3-hours post induction) would exhibit a different snR5 activity signal compared to yeast cells captured in the stationary phase (overnight induction). From these trials, it was revealed that the magnitude of the snR5-Test-ΔUPF1 green fluorescence-shift relative to the snR5-Neg-ΔUPF1 control strain was comparable at both sampling times, with the three overnight galactose induction technical replicates evincing EGFP fold-changes of 1.2-1.9, compared to the 3-hour post-induction EGFP fold change of 1.3 for the snR5-Test-ΔUPF1 strain (**Table 8**). However, the median EGFP signal from the 3-hour induction of the snR5-Test-ΔUPF1 strain was 1.00% that of the snR5-Pos-ΔUPF1 strain, which is lower than the 1.20 ± 0.06% value evinced for the snR5-Test-ΔUPF1 strain induced overnight. Further

technical replication of the 3-hour post-induction trial would have to be performed to determine if this subtle signal reduction is statistically significant. Nevertheless, if this slight reduction of signal intensity in the 3-hour post-induction trial is genuine, it is most likely due to the maturation time of EGFP and not indicative of a change in the activity of the snR5 H/ACA snoRNP (vide infra). With respect to mRuby2 signal expression, the 3-hour post-induction time frame reduced the level of mRuby2 fluorescence emitted by the system (**Figure 16: snR5-ΔUPF1 3-hour induction**). Specifically, compared to wild-type yeast, only 3.97, 5.32, and 4.33% of the initial 100 000 yeast cells analyzed for the snR5-Test-ΔUPF1, snR5-Neg-ΔUPF1, and snR5-Pos-ΔUPF1 strains, respectively, had significantly higher levels of red fluorescence and were therefore included in the analysis after gating (**Table 8: snR5-ΔUPF1 3-hour induction**). For comparison, 81.4, 87.6, and 78.7% of the initial 100,000 yeast cells analyzed for the snR5-Test-ΔUPF1, snR5-Neg-ΔUPF1, and snR5-Pos-ΔUPF1 strains, respectively, had a significantly higher mRuby2 level than wild-type after overnight galactose induction (**Table 8: snR5-ΔUPF1 Replicate 1 (Overnight induction); Appendix 2**). These observations are consistent with previous maturation studies of the EGFP and mRuby2-related mRuby3 proteins, wherein the times for 90% fluorophore maturation for these two proteins were calculated to be ~42 minutes, and ~6 hours, respectively, at 37°C (*139*). Because this short induction time of three hours considerably lowers mRuby2 control signal output and may slightly diminish the EGFP signal output of the yeast reporter system, it is therefore advisable to conduct future flow cytometry experiments with longer incubation times (i.e., overnight) so as to allow an optimal level of mRuby2 and EGFP expression, maturation, and subsequent fluorescence.

5.2–snR5 and snR81 H/ACA snoRNP detection with the yeast reporter system

An engineered snR81 H/ACA snoRNP with a customized pseudouridylation pocket is the only pseudouridine synthase so far described in the literature to target premature stop codons and facilitate translational readthrough (*97, 109*). The reporter system described in this thesis was developed to study the activities of endogenous pseudouridine synthases recognizing their canonical substrates in an mRNA reporter construct; the snR81 H/ACA snoRNP was therefore a highly interesting candidate to analyze due to its significance in the nascent field of artificially guided RNA modification. The yeast reporter system was configured to detect the snR81 H/ACA snoRNP by inserting its 25S rRNA recognition sequence for Ψ1052 formation upstream of the *EGFP* reporter gene such that the targeted uridine was placed in an in-frame UAA stop codon context. In both a wild type and *ΔUPF1* background, a green signal enhancement was not observed in the snR81 H/ACA snoRNP reporter system (**Figure 14**). Therefore, the stop codon modifying activity of an engineered snR81 H/ACA snoRNP documented in the literature could not be replicated in this thesis with a wild type equivalent. Because this yeast reporter system was programmed to detect endogenous snR81 H/ACA snoRNP activity, this enzyme could be modifying an insufficient number of EGFP reporter transcripts for a detectable level of EGFP expression as the wild type snR81 H/ACA snoRNPs are already targeting their natural substrates. Conversely, the engineered snR81 H/ACA snoRNP reported in the literature operated orthogonally to the endogenous snR81 pseudouridylation machinery because the yeast system harbouring it was created with a customized snR81 guide RNA expression vector. Thus, a subset of customized snR81 H/ACA snoRNPs was generated in these yeast cells to specifically target the stop codon embedded in a reporter gene and did not have any

endogenous targets, possibly leading to a higher proportion of modified reporter transcripts than in my system analyzing wild type pseudouridine synthases.

To expand the repertoire of H/ACA snoRNP complexes evaluated for their ability to cause nonsense suppression, I also programmed a yeast reporter system with the snR5 H/ACA snoRNP's recognition sequence for Ψ1004 isomerization in 25S rRNA. Consequently, in a *ΔUPF1* yeast background, a 1.2-1.9-fold enrichment in the EGFP signal was detected in the snR5-Test reporter strain relative to a negative control strain with a mutated snR5 recognition sequence surrounding the stop codon (**Figure 16 and Table 8**). This green fluorescence-shift is interpreted as enhanced EGFP expression in the reporter system arising from translational readthrough of the snR5-pseudouridylated stop codon upstream of *EGFP*. The possibility that this green fluorescence increase is not due to stop codon pseudouridylation by snR5 and is rather an artefact from some other process in the cell is ruled out by the lower EGFP signal of the snR5-Neg strain relative to the snR5-Test strain—these yeast strains are genetically equivalent except for the recognition sequence surrounding the *EGFP* premature termination codon wherein the negative control strain has been mutated to abolish snR5 substrate recognition. Therefore, the snR5-Test strain, which emits higher EGFP fluorescence than the negative control strain, is only distinguished from the negative control by having the premature stop codon uridine placed within the recognition sequence of the snR5 guide RNA, which leads to PTC pseudouridylation, subsequent translational readthrough of the stop signal and concomitant EGFP expression. Thus, this reporter system is able to capture the modifying activity of endogenous snR5 H/ACA snoRNPs and subsequent iterations of this system can be used to study the *in vivo* dynamics of pseudouridylation. Furthermore, this enzyme is to-date

the only other pseudouridine synthase alongside engineered snR81 H/ACA snoRNPs that has been experimentally demonstrated to facilitate nonsense suppression of a premature stop codon via targeted pseudouridylation (*97, 109*).

It should be noted that the apparently low level of pseudouridylated stop codon readthrough (hence signal expression) in the snR5 reporter system was similarly observed in the previous work to engineer the snR81 H/ACA snoRNP to target novel substrates. In the system utilized by Karijolich and Yu, a PTC was placed at the second codon position of the *CUP1* gene. The authors engineered the snR81 H/ACA snoRNP to target this PTC by changing the sequence of the snR81 guide RNA's pseudouridylation pocket to recognize the stop codon uridine in the *CUP1* substrate. *CUP1* encodes a copper chelating protein that confers yeast the ability to survive in the presence of Cu^{2+} ions. Therefore, in their *S. cerevisiae* system, nonsense suppression was evaluated by plating yeast on $CuSO_4$ selective media wherein cell survival and proliferation are dependent on translational readthrough of the *CUP1* PTC and expression of the metal chelating gene product (*97*). With this system, the engineered snR81 H/ACA snoRNP was able to restore a partial amount of growth on media containing 0.02 mM $CuSO_4$ relative to a yeast strain with an untargeted *CUP1* PTC. The authors found through subsequent TLC analysis that the *CUP1* PTC was only ~4% pseudouridylated. Furthermore, when they next developed an engineered snR81 H/ACA snoRNP to target a PTC in a *TRM4* reporter gene, they recovered protein expression to ~6% of the level produced from an unmutated *TRM4* control construct (*97*). In comparison, across three technical replicates, the median EGFP signal from the snR5-Test-*ΔUPF1* strain is 1.20 ± 0.06% that of the snR5-Pos-*ΔUPF1* strain whereas the snR5-Neg-*ΔUPF1* median EGFP signal is 0.81 ± 0.13% of the snR5-

Pos-*ΔUPF1* strain (**Table 8**). Therefore, although the readthrough efficiency in the snR5 fluorescence reporter system is lower than in the *TRM4* reporter system, the observed level of reporter gene expression recovery is still in the same order of magnitude as that in the other system. Since the reporter system used by Karijolich and Yu and did not contain a 2A peptide element, it could be that their higher level of reporter expression recovery is owed to the fact that their system does not have to contend with an additional factor that could limit translation of their reporter gene. Nevertheless, both the findings shown here and in Karijolich and Yu's study suggest that targeting a pseudouridine synthase to modify a stop codon in a reporter gene is possible, yet currently inefficient. To increase the stop codon modification efficiency and subsequent signal output of the fluorescence-based yeast reporter system, optimization measures will need to be explored (see further discussion in chapter 5.6).

In both the previous research with an engineered snR81 H/ACA snoRNP and the work described here with snR5, the NMD pathway was shown to affect the respective reporter systems used to assay for stop codon pseudouridylation. In the work of Karijolich and Yu, the *CUP1* mRNA level in a control strain with unmutated *CUP1* was 80% higher than the mRNA level in the strain housing a PTC-containing *CUP1* reporter gene due to NMD-based destruction of the reporter transcripts. Although the level of *EGFP* mRNA in my fluorescence-based reporter system was not directly analyzed, the influence of NMD can be observed by comparing the green fluorescence output of the snR5-Test reporter system in a wild-type yeast background to that of the snR5-Test system in a *ΔUPF1* background (**Figures 15 and 16**). Whereas the snR5-Test system did not elicit a green signal enhancement in wild type yeast, it did when expressed in yeast lacking the NMD

pathway gene *UPF1*. Therefore, in a wild type background, it can be surmised that the NMD response significantly reduces the *EGFP* mRNA level and prevents expression of a detectable level of fluorescent protein. This highlights how kinetics and timing are important features of nonsense suppression by targeted stop codon pseudouridylation in my reporter system. The fate of a target transcript with a premature stop codon depends on what happens first: pseudouridylation before translation or recognition by the NMD response during premature translation termination on an unmodified stop signal. If the PTC in the reporter transcript is not modified before a pioneering round of translation, it will cause premature translation termination which in turn triggers the NMD pathway and ultimately destroys the transcript before it has a subsequent chance to be pseudouridylated (**Figure 11**). Thus, the only way for a PTC-containing *EGFP* transcript to reliably survive is by pseudouridylation prior to a pioneering round of translation, whereupon nonsense suppression of the modified stop signal allows translational readthrough, thereby escaping the NMD response. If the efficiency of reporter transcript pseudouridylation could be increased in my reporter system such that a larger population of *EGFP* transcripts were modified prior to translation, interference by NMD on EGFP expression could be reduced (vide infra). However, it should be noted that since pseudouridylated stop codon readthrough is not perfectly efficient, the NMD pathway will still degrade a portion of modified transcripts in the instances where release factors manage to recognize the pseudouridylation stop codon before a tRNA can be recruited. This has ramifications for the proposed use of pseudouridine-based stop codon reprogramming in the treatment of genetic diseases—even if pseudouridylation can be efficiently directed to a PTC gene lesion (with the intent of restoring a healing dose of functional protein), the NMD pathway

may still degrade the pseudouridylated PTC-harbouring mRNAs and possibly prevent expression of a sufficient dose of vital protein product. With this in mind, future endeavors to develop PTC pseudouridylation as a therapeutic treatment should consider a complementary treatment that can inhibit function of the NMD pathway in addition to ensuring efficient PTC pseudouridylation by a programmed pseudouridine synthase.

5.3–Pus7 mRNA pseudouridylation inquiries with the yeast reporter system

Two iterations of the fluorescent-based yeast reporter system were created in an attempt to detect Pus7 activity. In the first iteration, the reporter system was programmed with a 29 nt sequence consisting of positions 124-152 from the *RPB10* gene, which was previously identified to have a uridine residue at position 138 pseudouridylated by Pus7 (*7*). As the pseudouridylation extent at this site was demonstrated to be enhanced under heat shock stress, this initial reporter system was meant to detect heat shock inducible Pus7 modification with the broader goal of developing a generalized platform to investigate the mechanisms of regulated mRNA pseudouridylation *in vivo*. However, under 30°C and 45°C growth conditions, a green fluorescent signal could not be induced from the Pus7-Test reporter system in both a wild type and *ΔUPF1* background (**Figures 10 and 12**).

As Pus7 has multi-substrate specificity, being known to target cytoplasmic tRNAs, U2 snRNA and 5S rRNA in addition to mRNAs, the second iteration of the Pus7 yeast reporter system was programmed with a 76 nt sequence encompassing positions 26-101 of U2 snRNA with the intent of observing Pus7 modification of U35 in this U2 snRNA sequence (**Figure 13A**) (*7, 83, 131*). Although Pus7 substrate recognition and catalysis requires the UGUAR (R = G > A) consensus sequence in an RNA substrate, some evidence suggests that a hairpin structure in close proximity to the target uridine contributes to target

recognition (*98, 131*). For efficient Ψ35 catalysis in U2 snRNA, the presence of the stem-loop II region downstream of the target uridine has been shown as necessary (*132*). Critically, the U2 snRNA recognition sequence in the second iteration of the Pus7-Test reporter system contains this stem-loop II region; it was hypothesized that its inclusion would lead to efficient EGFP reporter mRNA recognition and catalysis by Pus7 thus leading to enhanced EGFP fluorescence. However, when analyzed by flow cytometry, the Pus7-U2-Test-*ΔUPF1* strain did not emit a greater level of EGFP fluorescence than the corresponding negative control (**Figure 13B**).

Thus, two versions of the Pus7 fluorescence-based yeast reporter system (one encoding an mRNA substrate and the other a snRNA substrate of Pus7) did not result in significant EGFP expression. The current design of the reporter system, namely the 2A peptide feature, may preclude expression of EGFP to a detectable extent even if Pus7 is modifying a small proportion of the reporter transcripts. Furthermore, if specific structural elements are required for Pus7 recognition, perhaps the 76 nucleotides of U2 snRNA sequence inserted upstream of *EGFP* are failing to adopt the precise stem-loop II region that Pus7 relies on when generating Ψ35 in U2 snRNA. Additionally, the reporter transcripts may be competing with natural Pus7 substrates for modification by the available Pus7 enzyme in the yeast cell, thus reducing the efficiency of PTC pseudouridylation to a point where the NMD pathway destroys a large enough proportion of reporter transcripts to lower EGFP expression below the detection threshold of the flow cytometry assay. To capture Pus7's activity, a subsequent yeast reporter system will have to be developed (see further discussion in Chapter 5.6).

5.4–Nonsense suppression in the yeast reporter system using suppressor tRNAs

Because pseudouridine synthase-directed stop codon modification in the yeast reporter system only yielded a minimal level of EGFP expression, I sought to create a positive control for nonsense suppression in the form of a suppressor tRNA-based reporter strain. Two suppressor tRNAs, the UAA suppressor SUP4 and the UAG suppressor SUP53, were evaluated for their ability to promote nonsense suppression in the snR81-Test-*ΔUPF1* (UAA PTC) and Pus7-U2-Test-*ΔUPF1* (UAG PTC) strains, respectively (*119, 120*). Of these two suppressor tRNAs, only SUP4 tRNA expressed from a yeast plasmid in the snR81-Test-*ΔUPF1* strain was able to increase the EGFP output of the reporter system (**Figure 18A and B**). This snR81-Test-SUP4 strain produced the largest EGFP fold change thus far described, with 5.03-fold more EGFP expression over a negative control. This SUP4-based suppression control reading is larger than the snR5-Test-*ΔUPF1* fold change of 1.2-1.9, but still in the same order of magnitude. Although this SUP4 control strain provides a useful benchmark for nonsense suppression-based EGFP expression from my reporter system, it should be noted that previous work with SUP4 tRNA conducted by Guy, et al. has been able to recover a higher degree of nonsense suppression. In a study on the determinants of tRNA function, the authors used a model system quite similar to ours, wherein SUP4 was utilized to suppress a UAA stop codon in a *GFP* reporter gene which was contained in the RNA-ID reporter system such that the level of GFP expression could be compared to a red fluorescent protein control signal (*118, 140*). In this investigation, SUP4 tRNA was able to recover GFP expression to a level virtually identical to that of a control strain with no UAA mutation in *GFP* (*140*). By comparison, SUP4 nonsense suppression in my system only recovers 2.94% EGFP

expression compared to a positive control strain. One possible source of this lower expression level could be the presence of the 2A element upstream of *EGFP* in my reporter system as previously described. However, the disparity in green fluorescence expression recovery between my system and the one in Guy et al., may also owe to the fact that they integrated their SUP4 expression construct as a single copy into the *ADE2* locus of their yeast strain whereas in my work SUP4 was expressed from the multicopy pRS327 plasmid. Thus, in my system there is a higher copy number of SUP4 genes, hence a higher number of suppressor tRNAs (*124*). Consequently, a higher level of UAA stop codon readthrough is most likely occurring in my system compared to the one in Guy et al. Although this means that the PTC upstream of EGFP in my reporter system is being readthrough at a greater frequency, the SUP4 tRNA does not discriminate between the premature UAA stop codon preventing *EGFP* translation and the other UAA stop codons distributed throughout the yeast transcriptome. Therefore, an overabundance of suppressor tRNA expressed from the multicopy vector also enhances nonsense suppression in off target genes, thus generating aberrant protein products with C-terminal extensions that could cause loss or gain of function mutations which harm the cell's metabolism and in turn lower the level of mature EGFP. It may be the case then, that there exists an optimal level of suppressor tRNA expression that facilitates efficient EGFP expression while minimizing the occurrence of inevitable off-target readthrough events. Off-target nonsense suppression may also account for the anomalous outcome of the Pus7-U2-Test-SUP53 strain wherein the presence of the UAG suppressor tRNA *reduced* the EGFP expression level 2.5-fold compared to a negative control strain rather than rescuing EGFP translation by promoting UAG suppression as expected (**Figure 18C and D**).

Overall, the snR81-Test-SUP4 strain manifested a ~5-fold change in EGFP expression over a negative control strain and in so doing serves as a useful positive control for nonsense suppression in the fluorescence-based yeast reporter system developed in this thesis. Going forward, EGFP signal enhancement may be attained by integrating the suppressor tRNA gene as a single copy into the yeast genome to avoid the negative consequences of suppressor tRNA overabundance in the cell.

5.5–Pseudouridylation detection with the fluorescence-based yeast reporter system as compared to other modification detection methods

In the current state of the post-transcriptional RNA modification field, an increasingly large number of modification sites are being identified across a variety of RNAs, in large part due to advances in detection methodologies that include mass spectrometry and next generation sequencing (*78*). Consequently, several modifications including N6-methyladenosine, N6,2′-O-dimethyladenosine, 5-methylcytidine, 5-hydroxylmethylcytidine, inosine, N1-methyladenosine and pseudouridine have been identified as mRNA modifications present throughout the transcriptome of eukaryotic cells (*6, 7, 42, 47, 48, 54, 141-143*). This has led to much excitement in the field, as hypotheses concerning the role of these modifications in mRNA have begun to focus on their potential to be part of a hitherto unknown mode of gene expression regulation. As such, the term "epitranscriptomics" has been proffered to describe the emerging field of study dedicated to elucidating this form of regulation (*144*). However, the current field of epitranscriptomics is at this point largely conceptual and, despite the identification of a multiplicity of modification sites in mRNA, the function of most of these modifications remains to be investigated. Furthermore, of the sites identified through these high

throughput methods, minimal follow up experimentation has validated their presence, thus leading to a large corpus of putative modification sites in numerous transcripts with comparatively little insight into their functional significance and the enzymes responsible for their introduction (*78*).

The fluorescence-based yeast reporter system described in this thesis was designed to address the attendant limitations of the high-throughput sequencing methods for pseudouridine detection and to offer insight into the regulatory mechanisms governing mRNA pseudouridylation. Although the high-throughput pseudouridine sequencing methods are capable of identifying putative modifications sites on the order of hundreds to thousands in a single experiment, issues surrounding stringent bioinformatic cut-offs, sequence coverage and the possibility of false positive and negative events has led to the assertion that all pseudouridylation sites discovered through this manner be considered strictly putative until verified by an independent method (*44*, *50*). Thus, the fluorescence-based yeast reporter system is ideally positioned to verify the putative mRNA modification sites called via high-throughput sequencing. Furthermore, the high-throughput detection methods exhibit a marked bias for pseudouridine identification in highly abundant transcripts, while neglecting putative sites in rarer RNA species. This bias for modification detection is partially addressed in the fluorescence-based yeast system as the putatively modified sequence is placed upstream of the *EGFP* reporter and is under control of a galactose-inducible promoter (**Figure 6**). However, as previously discussed it should be noted that the high transcript expression level may limit efficient pseudouridylation by an endogenous pseudouridine synthase also recognizing its natural substrates.

Sequence and ligation-independent cloning of a short mRNA substrate sequence upstream of *EGFP* prior to integration of the reporter system into yeast removes the difficulties attendant to traditional cloning of recombinant DNA molecules, making it feasible to generate a yeast reporter strain variant offering *in vivo* validation of a putative pseudouridylation site in a relatively short period of time (**Figure 7**) (*123*). Another feature of the fluorescence-based reporter system is its sensitivity; flow cytometry analysis of the yeast strains allows EGFP to be detected in exceedingly small quantities that are below the detection threshold of a Western blotting assay (**Figure 17**). Thus, a low level of nonsense suppression (hence pseudouridylation) that may be missed by other detection methods can be captured using my fluorescence-based reporter approach. Finally, fluorescent readout from the yeast reporter system not only provides information on the status of a putatively modified site, but also reports on the activity of the pseudouridine synthase responsible for catalysis, thus making it possible to study the regulation of the enzyme responsible for the modification of interest, something that no other pseudouridine detection method can provide.

The limitations attendant to pseudouridine detection with the fluorescence-based yeast reporter system include the fact that, for this system to function, the putatively modified site must occur in either a UAA, UAG, or UGA (stop codon) context such that EGFP expression is prevented in the absence of pseudouridylation. In cases where the two positions immediately downstream of the target U do not form a stop codon, the sequence could be mutated to rectify this, but doing so would be to risk destroying the characteristics of the recognition sequence necessary for pseudouridine synthase binding and catalysis. In addition to this limitation, it should also be noted that the fluorescence-based yeast reporter

system does not allow for *de novo* identification of pseudouridylation sites like in the pseudouridine sequencing methods as the target sequence for modification must be known before-hand so that it may be inserted upstream of *EGFP*. Finally, unlike the radiolabeling-based SCARLET method, the fluorescence-based yeast reporter system does not allow for absolute quantitation of pseudouridylation extent at the targeted position and can only offer a relative measure of the modification extent based on the magnitude of EGFP signal expression rescued in the reporter strain (*59*).

5.6–Engineering the fluorescence-based yeast reporter system to increase signal output

As previously noted in chapter 5.1, a modest level of snR5 H/ACA snoRNP activity was detected in the snR5-Test-*ΔUPF1* reporter strain and no discernible activity was detected in the reporter systems configured to analyze the snR81 H/ACA snoRNP and Pus7. There are a number of ways in which the fluorescence-based reporter system can be engineered to increase the EGFP signal output. For the H/ACA snoRNP reporter strains, guide RNA overexpression plasmids encoding the snR5 and snR81 H/ACA snoRNAs, respectively, could be co-transformed into yeast alongside the fluorescence reporter construct, thus increasing the available level of guide RNA in the cell and leading to a greater abundance of mature snoRNP complexes, thereby facilitating an increased level of target uridine modification. Likewise, for the standalone pseudouridine synthase Pus7, a multicopy Pus7 expression plasmid could be introduced into the Pus7 reporter strain, thus increasing enzyme concentration in the cell to a level that would favour *EGFP* reporter transcript pseudouridylation. In regards to the fluorescent reporter construct itself, the 2A peptide element upstream of *EGFP* could be removed in an attempt to increase EGFP

translation by eliminating the occurrence of ribosome fall-off on this sequence (*135*). Furthermore, the *LEU2* selective marker used to disrupt the *UPF1* locus on yeast chromosome XIII may have an unintended impact on the fluorescent reporter system (**Figure 11**). In a systematic characterization of the metabolic effects of yeast plasmid components, it was found that, out of the *HIS3*, *TRP1*, *LEU2*, *URA3*, and *KanMX* selective markers used for yeast genetic manipulation, plasmids harbouring the *LEU2* marker led to yeast exhibiting the slowest growth rate which was 76.6% that of a control strain (*145*). Therefore, the *LEU2* marker in the *ΔUPF1* yeast strain may be limiting the metabolic output of the cell, thereby lowering the final level of EGFP expression from the fluorescence-based reporter system. Consequently, it would be beneficial to engineer a novel *ΔUPF1* strain with a different selective marker inserted into the *UPF1* locus (such as *HIS3*) which does not have the same documented impact on yeast growth, thereby not only increasing the level of EGFP expression in the test strains, but also increasing the maximum level of observed fluorescence in the positive control strains as well.

5.7–H/ACA snoRNP localization and the fluorescence-based reporter system

In addition to the reporter system alterations mentioned in the previous section, there is a final consideration that may impact the enzyme activity measurements from the reporter system, specifically, pseudouridine synthase localization within the cell. The H/ACA snoRNPs are a class of enzymes that are localized to the nucleolus sub compartment where they are known to modify rRNA. Prior to their localization to the nucleolus, H/ACA snoRNPs first undergo maturation in Cajal bodies (CBs): nuclear sub compartments enriched with RNAs and proteins involved in pre-mRNA processing and the site of H/ACA snoRNP and C/D box snoRNP maturation (*146*). CBs are present in various

eukaryotes and are scaffolded by the protein coilin. They contain a high concentration of spliceosomal snRNAs; notably, they also contain H/ACA guide RNA variants called small Cajal body-specific RNAs (scaRNAs) that associate with the WDR79 protein and possess a CAB-box and G·U/U·G wobble stem element that localizes them to CBs (*147*). Thus, CBs have their own complement of pseudouridylation guide RNAs separate from the H/ACA snoRNAs. This has led to a prevailing notion that H/ACA snoRNPs mainly function in the nucleolus to modify rRNA whereas scaRNAs only modify snRNA in CBs. Meanwhile, pre-mRNA transcription and processing is thought to occur in neither nuclear sub compartment and takes place in the nucleoplasm.

Under this prevailing notion then, in the context of the yeast reporter system, the different localizations of the matured snR5 and snR81 H/ACA snoRNPs (nucleoli) relative to the *EGFP* reporter mRNA (nucleoplasm), could serve as an obstacle for union of enzyme and substrate, ultimately preventing translational readthrough via a pseudouridylated stop codon and concomitant EGFP expression. Moreover, once pre-mRNA is fully processed into mRNA, it is subsequently exported to the cytoplasm for translation, at which point the *EGFP* reporter transcript is rendered inaccessible to H/ACA snoRNPs which strictly residue in the nucleus. This concern is not relevant to the Pus7 reporter system, as Pus7 has been established through high-throughput pseudouridine sequencing and knockdown analysis to target mRNA substrates, and is capable of re-localizing to the cytoplasm under heat shock stress (*7*).

It should be noted that, as experimentally demonstrated in this thesis, the snR5 H/ACA snoRNP is capable of directing pseudouridylation to the stop codon upstream of *EGFP* in the reporter transcript, thereby manifesting a green signal enhancement in a snR5-

Test-*ΔUPF1* reporter strain (**Figure 16 and Table 8**). Moreover, nucleolar localization did not eliminate the engineered snR81 H/ACA snoRNP's ability to modify the *CUP1* and *TRM4* reporter mRNAs in the work of Karijolich and Yu, although this factor could perhaps explain why the efficiency of pseudouridylation was low (~4%) (*97*). Therefore, the snR5 H/ACA snoRNP activity captured in this thesis alongside the engineered snR81 H/ACA snoRNP activity previously documented serve as a line of evidence arguing for the ability of predominantly nucleolus-localized H/ACA snoRNPs to pseudouridylate mRNAs. A potential model for H/ACA snoRNP interaction with mRNA comes from the interplay between Cajal bodies and the nucleolus, wherein H/ACA snoRNPs are first matured in CBs before being shuttled to the nucleolus where they modify their canonical rRNA substrates (**Figure 20**). During this shuttling process of a mature H/ACA snoRNP from a CB to the nucleolus, there exists a brief window of time in which the H/ACA snoRNP could encounter an mRNA substrate in the nucleoplasm and modify it provided it contains an appropriate recognition sequence. Furthermore, an underappreciated interplay between the two sub compartments may exist, wherein an H/ACA snoRNP localized to the nucleolus may be able to re-enter the Cajal body to modify certain snRNA substrates: Evidence for this assertion includes the known substrates of the snR81 H/ACA snoRNP. The 3′ pseudouridylation pocket of the snR81 guide RNA directs Ψ1052 formation in 25S rRNA (present in the nucleolus) whereas the 5′ pocket was discovered to direct Ψ42 formation in the CB-localized U2 snRNA (*28*). Moreover, under nutrient deprivation stress in yeast, snR81 inducibly catalyzes Ψ93 formation in U2 snRNA (*87*). Therefore, if H/ACA snoRNPs are actively shuttling between Cajal bodies and the nucleolus, this would increase the number of opportunities the enzymes would have to encounter mRNAs in the

nucleoplasm. Of final note regarding mRNA substrate availability to H/ACA snoRNPs, analysis of the high-throughput pseudouridine sequencing data sets has found pseudouridylated mRNA residues occurring in recognition sequences of known H/ACA snoRNAs, implying they are Cbf5-directed and that the responsible H/ACA snoRNP complexes are colocalized with their mRNA substrates at some point in the life cycle of the cell (*6*).

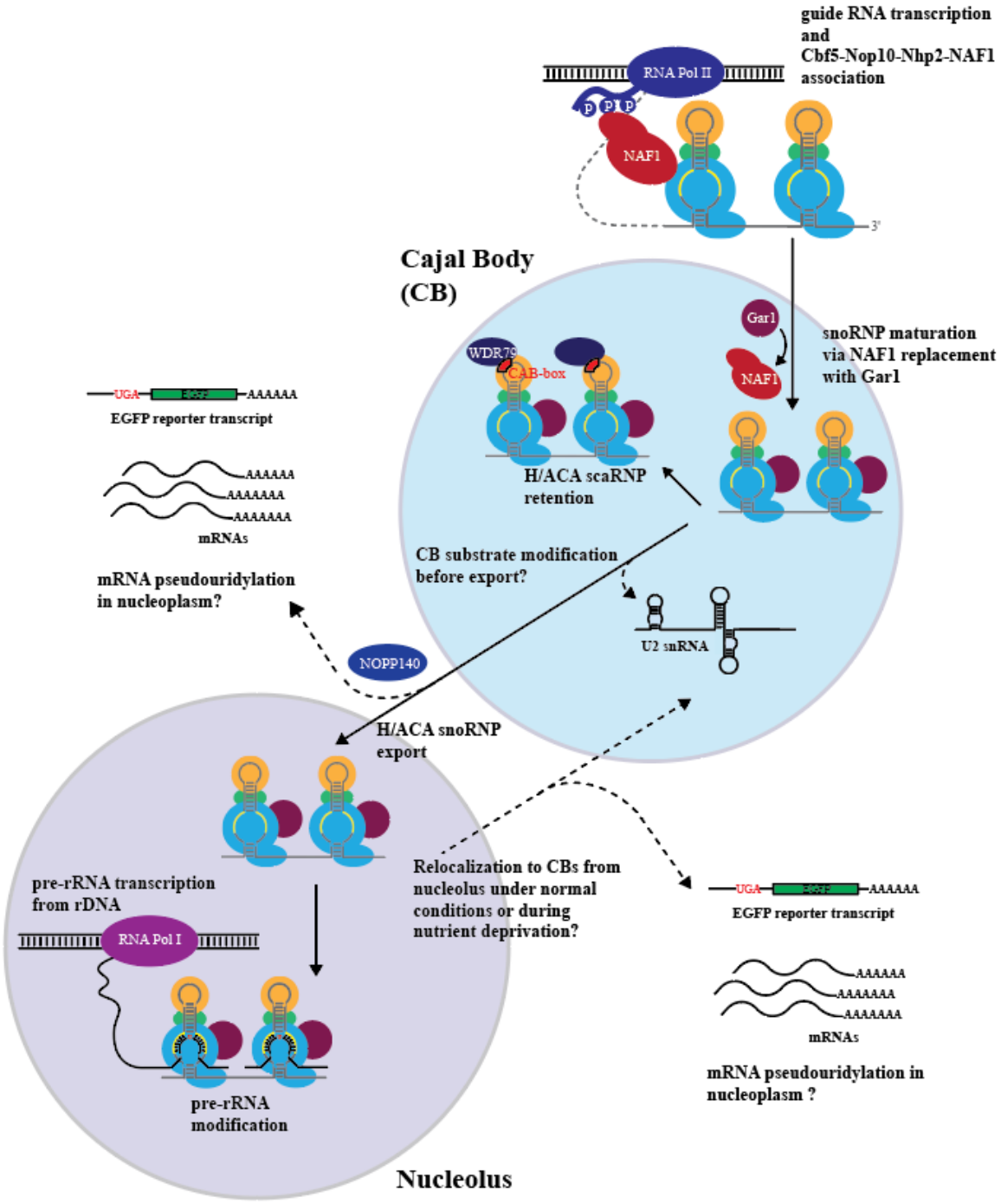


Figure 20. H/ACA snoRNP and scaRNP maturation and localization in the nucleus. A pre-H/ACA sno/scaRNP is formed when a Cbf5-Nop10-Nhp2 complex binds a nascent H/ACA guide RNA during its transcription by RNA polymerase II. The assembly factor NAF1 associates with the CTD of RNA polymerase II and keeps the immature H/ACA RNP inactive. H/ACA RNP maturation is completed in Cajal bodies (CBs), wherein Gar1 replaces the assembly factor NAF1, thus creating an active pseudouridine synthase (*146*).

H/ACA small CB-specific RNPs (H/ACA scaRNPs) are retained in CBs by CAB-box elements in the guide RNA's upper stem-loops, which bind the localization factor WDR79. Conversely, H/ACA small nucleolar RNPs (H/ACA snoRNPs) do not possess CAB-box elements and are exported from CBs to the nucleolus with the assistance of the NOPP140 shuttle protein (*146*). Upon arrival, the H/ACA snoRNPs are able to pseudouridylate their canonical targets in pre-ribosomal RNA, thus participating in ribosome biogenesis. In addition to modifying rRNA, H/ACA snoRNPs are also known to target substrates localized to CBs. For instance, the snR81 H/ACA snoRNP directs pseudouridylation to two positions in U2 snRNA (one under nutrient deprivation conditions (*87*)). Furthermore, the snR5 H/ACA snoRNP can possibly pseudouridylate a target located in the nucleoplasm: a premature stop codon in an *EGFP* reporter mRNA (demonstrated in this thesis), and a series of putative mRNA pseudouridylation targets have also been assigned to a variety of other H/ACA snoRNPs (*6, 7*). The spatiotemporal coordination of mRNA substrate modification by H/ACA snoRNPs could occur when a fully matured H/ACA snoRNP is shuttled from the CB to the nucleolus. Alternatively, if an underappreciated interplay between CBs and the nucleolus exist, H/ACA snoRNPs re-entering CBs from the nucleolus would have additional opportunities to encounter mRNAs in the nucleoplasm.

5.8–Future applications of the fluorescence-based *S. cerevisiae* reporter system

Based on the discovery that pseudouridylated stop codons promote nonsense suppression via noncanonical base-pairing interactions with near cognate tRNAs in the ribosome decoding center, I sought to develop a novel assay for *in vivo* pseudouridine synthase detection by tethering premature stop codon pseudouridylation to readout of a green fluorescent protein gene in yeast (*105*). Upon generating this novel reporter system, I was able to successfully assay for the activity of the snR5 H/ACA snoRNP modifying a stop codon-situated uridine upstream of the *EGFP* reporter, which occurs in snR5's 25S rRNA recognition sequence for Ψ1004 catalysis (*134*). Using this novel reporter system then, the snR5 H/ACA snoRNP is now the only other pseudouridine synthase alongside engineered snR81 H/ACA snoRNP variants that has been experimentally demonstrated to facilitate nonsense suppression of a premature stop codon via targeted pseudouridylation (*97, 109*). As nonsense suppression by pseudouridine recoding has been proposed as a

therapeutic approach for treating diseases caused by premature stop codon mutations, the fluorescence-based reporter system is a useful platform for studying pseudouridine synthase recognition criteria such that novel guide RNAs can be optimized to target gene lesions in specific transcripts (*104*, *112*).

In its current form, the fluorescence-based yeast reporter system is a unique pseudouridine detection method because it allows for *in vivo* apprehension of endogenous pseudouridine synthase activity in addition to merely reporting on the presence or absence of pseudouridine at a putatively modified site. Therefore, in addition to serving as a useful tool for validating the plethora of putative pseudouridylation sites called through high-throughput pseudouridine sequencing studies, the novel reporter system designed in this thesis is positioned to investigate the enzymes responsible for modification, and begin to elucidate the regulatory mechanisms behind their activity. For instance, as different environmental conditions have been shown to impact the transcriptome-wide pattern of pseudouridylation, one could envision using the reporter system to characterize a specific pseudouridine synthase's response to various stress conditions, thereby determining which enzymes are responsible for upregulating or downregulating modification under the applied conditions (*7, 47*). Using the fluorescence-based reporter system in this manner would go a long way in the emerging field of epitranscriptomics, where there is currently a dearth of knowledge surrounding the regulatory mechanisms underlying inducible mRNA pseudouridylation.

Because the fluorescence-based reporter system is easily transformed and integrated into a yeast strain, it could be introduced into yeast strains possessing mutated pseudouridine synthases and used to characterize the activity of these enzymes. In addition

to addressing basic research questions concerning pseudouridine synthase function, doing so is clinically relevant in the case of a series of mutations in DKC1 (human homolog of Cbf5) that are associated with the X-linked Dyskeratosis congenita disorder characterized by nail dystrophy, reticulate skin pigmentation, mucosal leukoplakia and bone marrow failure (*148*). The molecular basis of X-linked Dyskeratosis congenita is not completely understood, however impaired ribosome biogenesis and telomere shortening related to H/ACA snoRNP dysfunction are thought to underly this disease (*149*). By generating Dyskeratosis congenita-like mutations in yeast Cbf5 and characterizing them with the reporter system, the effects of these mutations on pseudouridylase activity could offer critical insights into the molecular mechanisms of this disease.

In summation, the fluorescence-based *S. cerevisiae* reporter system developed in this thesis could be further developed into a robust experimental platform that can be leveraged to not only address basic research questions concerning the mechanisms of regulated mRNA pseudouridylation but also may be utilized to develop nonsense suppression as a viable therapy for premature termination codon-based diseases and characterize the pseudouridine synthase mutations responsible for Dyskeratosis congenita.

References

1. P. Boccaletto *et al.*, MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res* **46**, D303-D307 (2018).

2. C. J. Lewis, T. Pan, A. Kalsotra, RNA modifications and structures cooperate to guide RNA-protein interactions. *Nat Rev Mol Cell Biol* **18**, 202-210 (2017).

3. F. Spenkuch, Y. Motorin, M. Helm, Pseudouridine: still mysterious, but never a fake (uridine)! *RNA Biol* **11**, 1540-1554 (2014).

4. W. E. Cohn, Pseudouridine, a carbon-carbon linked ribonucleoside in ribonucleic acids: isolation, structure, and chemical characteristics. *J Biol Chem* **235**, 1488-1498 (1960).

5. F. F. Davis, F. W. Allen, Ribonucleic acids from yeast which contain a fifth nucleotide. *J Biol Chem* **227**, 907-915 (1957).

6. T. M. Carlile *et al.*, Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **515**, 143-146 (2014).

7. S. Schwartz *et al.*, Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148-162 (2014).

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

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

10. J. Noeske *et al.*, High-resolution structure of the *Escherichia coli* ribosome. *Nat Struct Mol Biol* **22**, 336-341 (2015).

11. H. Shi, P. B. Moore, The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: a classic structure revisited. *RNA* **6**, 1091-1105 (2000).

12. D. R. Davis, Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res* **23**, 5020-5026 (1995).

13. E. Kierzek *et al.*, The contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic Acids Res* **42**, 3492-3501 (2014).

14. T. H. King, B. Liu, R. R. McCully, M. J. Fournier, Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Mol Cell* **11**, 425-435 (2003).

15. M. I. Newby, N. L. Greenbaum, Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. *Nat Struct Biol* **9**, 958-965 (2002).

16. C. Yang, D. S. McPheeters, Y. T. Yu, Ψ35 in the branch site recognition region of U2 small nuclear RNA is important for pre-mRNA splicing in *Saccharomyces cerevisiae*. *J Biol Chem* **280**, 6655-6662 (2005).

17. T. Hamma, A. R. Ferre-D'Amare, Pseudouridine synthases. *Chem Biol* **13**, 1125-1135 (2006).

18. Y. Watanabe, M. W. Gray, Evolutionary appearance of genes encoding proteins associated with box H/ACA snoRNAs: cbf5p in *Euglena gracilis*, an early diverging eukaryote, and candidate Gar1p and Nop10p homologs in archaebacteria. *Nucleic Acids Res* **28**, 2342-2352 (2000).

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

20. G. R. Veerareddygari, S. K. Singh, E. G. Mueller, The Pseudouridine Synthases Proceed through a Glycal Intermediate. *J Am Chem Soc* **138**, 7852-7855 (2016).

21. C. Hoang, A. R. Ferre-D'Amare, Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell* **107**, 929-939 (2001).

22. T. Kiss, E. Fayet-Lebaron, B. E. Jady, Box H/ACA small ribonucleoproteins. *Mol Cell* **37**, 597-606 (2010).

23. T. S. Rozhdestvensky *et al.*, Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res* **31**, 869-877 (2003).

24. D. L. Baker *et al.*, RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. *Genes Dev* **19**, 1238-1248 (2005).

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

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

27. C. Torchet *et al.*, The complete set of H/ACA snoRNAs that guide rRNA pseudouridylations in *Saccharomyces cerevisiae*. *RNA* **11**, 928-938 (2005).

28. X. Ma *et al.*, Pseudouridylation of yeast U2 snRNA is catalyzed by either an RNA-guided or RNA-independent mechanism. *EMBO J* **24**, 2403-2413 (2005).

29. P. Ganot, B. E. Jady, M. L. Bortolin, X. Darzacq, T. Kiss, Nucleolar factors direct the 2'-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA. *Mol Cell Biol* **19**, 6906-6917 (1999).

30. L. Li, K. Ye, Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature* **443**, 302-307 (2006).

31. J. Duan, L. Li, J. Lu, W. Wang, K. Ye, Structural mechanism of substrate RNA recruitment in H/ACA RNA-guided pseudouridine synthase. *Mol Cell* **34**, 427-439 (2009).

32. M. L. Bortolin, P. Ganot, T. Kiss, Elements essential for accumulation and function of small nucleolar RNAs directing site-specific pseudouridylation of ribosomal RNAs. *EMBO J* **18**, 457-469 (1999).

33. I. Perez-Arellano, J. Gallego, J. Cervera, The PUA domain - a structural and functional overview. *FEBS J* **274**, 4972-4984 (2007).

34. A. Henras, C. Dez, J. Noaillac-Depeyre, Y. Henry, M. Caizergues-Ferrer, Accumulation of H/ACA snoRNPs depends on the integrity of the conserved central domain of the RNA-binding protein Nhp2p. *Nucleic Acids Res* **29**, 2733-2746 (2001).

35. E. A. Caton, E. K. Kelly, R. Kamalampeta, U. Kothe, Efficient RNA pseudouridylation by eukaryotic H/ACA ribonucleoproteins requires high affinity binding and correct positioning of guide RNA. *Nucleic Acids Res* **46**, 905-916 (2018).

36. E. D. Egan, K. Collins, Biogenesis of telomerase ribonucleoproteins. *RNA* **18**, 1747-1759 (2012).

37. V. Atzorn, P. Fragapane, T. Kiss, U17/snR30 is a ubiquitous snoRNA with two conserved sequence motifs essential for 18S rRNA production. *Mol Cell Biol* **24**, 1769-1778 (2004).

38. W. E. Cohn, E. Volkin, Nucleoside-5′-Phosphates from Ribonucleic Acid. *Nature* **167**, (1951).

39. H. A. Ebhardt *et al.*, Meta-analysis of small RNA-sequencing errors reveals ubiquitous post-transcriptional RNA modifications. *Nucleic Acids Res* **37**, 2461-2470 (2009).

40. J. E. Jackman, R. K. Montange, H. S. Malik, E. M. Phizicky, Identification of the yeast gene encoding the tRNA m1G methyltransferase responsible for modification at position 9. *RNA* **9**, 574-585 (2003).

41. X. Li *et al.*, Transcriptome-wide mapping reveals reversible and dynamic N(1)-methyladenosine methylome. *Nat Chem Biol* **12**, 311-316 (2016).

42. M. Safra *et al.*, The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature* **551**, 251-255 (2017).

43. D. Dominissini *et al.*, The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **530**, 441-446 (2016).

44. M. Helm, Y. Motorin, Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat Rev Genet* **18**, 275-291 (2017).

45. A. Bakin, J. Ofengand, Four newly located pseudouridylate residues in *Escherichia coli* 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. *Biochemistry* **32**, 9754-9762 (1993).

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

47. X. Li *et al.*, Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat Chem Biol* **11**, 592-597 (2015).

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

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

50. M. Zaringhalam, F. N. Papavasiliou, Pseudouridylation meets next-generation sequencing. *Methods* **107**, 63-72 (2016).

51. V. E. Velculescu *et al.*, Characterization of the yeast transcriptome. *Cell* **88**, 243-251 (1997).

52. Z. Lei, C. Yi, A Radiolabeling-Free, qPCR-Based Method for Locus-Specific Pseudouridine Detection. *Angew Chem Int Ed Engl* **56**, 14878-14882 (2017).

53. L. Sun *et al.*, Transcriptome-wide analysis of pseudouridylation of mRNA and non-coding RNAs in Arabidopsis. *J Exp Bot* **70**, 5089-5600 (2019).

54. V. Khoddami *et al.*, Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution. *Proc Natl Acad Sci U S A* **116**, 6784-6789 (2019).

55. A. M. Fleming *et al.*, Structural Elucidation of Bisulfite Adducts to Pseudouridine That Result in Deletion Signatures during Reverse Transcription of RNA. *J Am Chem Soc* **141**, 16450-16460 (2019).

56. J. B. Macon, R. Wolfenden, 1-Methyladenosine. Dimroth rearrangement and reversible reduction. *Biochemistry* **7**, 3453-3458 (1968).

57. X. Zhao, Y. T. Yu, Detection and quantitation of RNA base modifications. *RNA* **10**, 996-1002 (2004).

58. M. Hengesbach, M. Meusburger, F. Lyko, M. Helm, Use of DNAzymes for site-specific analysis of ribonucleotide modifications. *RNA* **14**, 180-187 (2008).

59. N. Liu, T. Pan, Probing RNA Modification Status at Single-Nucleotide Resolution in Total RNA. *Methods Enzymol* **560**, 149-159 (2015).

60. M. I. Lomax, G. R. Greenberg, A new assay of thymidylate synthetase activity based on the release of tritium from deoxyuridylate-5-3-H. *J Biol Chem* **242**, 109-113 (1967).

61. J. Mengel-Jorgensen, F. Kirpekar, Detection of pseudouridine and other modifications in tRNA by cyanoethylation and MALDI mass spectrometry. *Nucleic Acids Res* **30**, e135 (2002).

62. G. Emmerechts, P. Herdewijn, J. Rozenski, Pseudouridine detection improvement by derivatization with methyl vinyl sulfone and capillary HPLC-mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* **825**, 233-238 (2005).

63. A. Durairaj, P. A. Limbach, Mass spectrometry of the fifth nucleoside: a review of the identification of pseudouridine in nucleic acids. *Anal Chim Acta* **623**, 117-125 (2008).

64. K. W. Gaston, P. A. Limbach, The identification and characterization of non-coding and coding RNAs and their modified nucleosides by mass spectrometry. *RNA Biol* **11**, 1568-1585 (2014).

65. J. A. Kowalak, S. C. Pomerantz, P. F. Crain, J. A. McCloskey, A novel method for the determination of post-transcriptional modification in RNA by mass spectrometry. *Nucleic Acids Res* **21**, 4577-4585 (1993).

66. T. Suzuki, T. Suzuki, Chaplet column chromatography: isolation of a large set of individual RNAs in a single step. *Methods Enzymol* **425**, 231-239 (2007).

67. Y. Yamauchi *et al.*, A mass spectrometry-based method for direct determination of pseudouridine in RNA. *Nucleic Acids Res* **44**, e59 (2016).

68. K. D. Meyer *et al.*, Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**, 1635-1646 (2012).

69. E. Mishima *et al.*, Immuno-Northern Blotting: Detection of RNA Modifications by Using Antibodies against Modified Nucleosides. *PLoS One* **10**, e0143756 (2015).

70. F. Weichmann *et al.*, Validation strategies for antibodies targeting modified ribonucleotides. *RNA*, (2020).

71. M. Helm, F. Lyko, Y. Motorin, Limited antibody specificity compromises epitranscriptomic analyses. *Nat Commun* **10**, 5669 (2019).

72. A. V. Grozhik *et al.*, Antibody cross-reactivity accounts for widespread appearance of m(1)A in 5'UTRs. *Nat Commun* **10**, 5126 (2019).

73. J. C. Darnell, A. Mele, K. Y. S. Hung, R. B. Darnell, Mapping of In Vivo RNA-Binding Sites by Ultraviolet (UV)-Cross-Linking Immunoprecipitation (CLIP). *Cold Spring Harb Protoc* **2018**, (2018).

74. D. R. Garalde *et al.*, Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods* **15**, 201-206 (2018).

75. R. E. Workman *et al.*, Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat Methods* **16**, 1297-1305 (2019).

76. A. M. Smith, M. Jain, L. Mulroney, D. R. Garalde, M. Akeson, Reading canonical and modified nucleobases in 16S ribosomal RNA using nanopore native RNA sequencing. *PLoS One* **14**, e0216709 (2019).

77. A. Viehweger *et al.*, Direct RNA nanopore sequencing of full-length coronavirus genomes provides novel insights into structural variants and enables modification analysis. *Genome Res* **29**, 1545-1554 (2019).

78. M. Schaefer, U. Kapoor, M. F. Jantsch, Understanding RNA modifications: the promises and technological bottlenecks of the 'epitranscriptome'. *Open Biol* **7**, (2017).

79. W. Huang da, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).

80. M. A. Nakamoto, A. F. Lovejoy, A. M. Cygan, J. C. Boothroyd, mRNA pseudouridylation affects RNA metabolism in the parasite Toxoplasma gondii. *RNA* **23**, 1834-1849 (2017).

81. S. Massenet *et al.*, 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 (1999).

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

83. I. Behm-Ansmant *et al.*, The *Saccharomyces cerevisiae* U2 snRNA:pseudouridine-synthase Pus7p is a novel multisite-multisubstrate RNA:Psi-synthase also acting on tRNAs. *RNA* **9**, 1371-1382 (2003).

84. M. Safra, R. Nir, D. Farouq, I. Vainberg Slutskin, S. Schwartz, TRUB1 is the predominant pseudouridine synthase acting on mammalian mRNA via a predictable and conserved code. *Genome Res* **27**, 393-406 (2017).

85. C. Hoang, A. R. Ferre-D'Amare, Crystal structure of the highly divergent pseudouridine synthase TruD reveals a circular permutation of a conserved fold. *RNA* **10**, 1026-1033 (2004).

86. Y. Kaya, J. Ofengand, A novel unanticipated type of pseudouridine synthase with homologs in bacteria, archaea, and eukarya. *RNA* **9**, 711-721 (2003).

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

88. N. Liu *et al.*, N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* **518**, 560-564 (2015).

89. G. Wu *et al.*, Pseudouridines in U2 snRNA stimulate the ATPase activity of Prp5 during spliceosome assembly. *EMBO J* **35**, 654-667 (2016).

90. K. S. Rajan *et al.*, Pseudouridines on *Trypanosoma brucei* spliceosomal small nuclear RNAs and their implication for RNA and protein interactions. *Nucleic Acids Res* **47**, 7633-7647 (2019).

91. G. Jia *et al.*, N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol* **7**, 885-887 (2011).

92. G. Zheng *et al.*, ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol Cell* **49**, 18-29 (2013).

93. W. V. Gilbert, T. A. Bell, C. Schaening, Messenger RNA modifications: Form, distribution, and function. *Science* **352**, 1408-1412 (2016).

94. D. E. Eyler *et al.*, Pseudouridinylation of mRNA coding sequences alters translation. *Proc Natl Acad Sci U S A* **116**, 23068-23074 (2019).

95. K. Tomita, T. Ueda, K. Watanabe, The presence of pseudouridine in the anticodon alters the genetic code: a possible mechanism for assignment of the AAA lysine codon as asparagine in echinoderm mitochondria. *Nucleic Acids Res* **27**, 1683-1689 (1999).

96. N. M. Martinez *et al.*, Pseudouridine synthases modify human pre-mRNA co-transcriptionally and affect splicing. bioRxiv doi: 10.1101/2020.1108.1129.273565 (2020).

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

98. A. C. Rintala-Dempsey, U. Kothe, Eukaryotic stand-alone pseudouridine synthases - RNA modifying enzymes and emerging regulators of gene expression? *RNA Biol* **14**, 1185-1196 (2017).

99. K. Wanichthanarak, I. Nookaew, D. Petranovic, yStreX: yeast stress expression database. *Database (Oxford)* **2014**, (2014).

100. J. M. Cherry *et al.*, Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* **40**, D700-705 (2012).

101. J. Jimenez, N. Ricco, C. Grijota-Martinez, R. Fado, J. Clotet, Redundancy or specificity? The role of the CDK Pho85 in cell cycle control. *Int J Biochem Mol Biol* **4**, 140-149 (2013).

102. J. Song *et al.*, Differential roles of human PUS10 in miRNA processing and tRNA pseudouridylation. *Nat Chem Biol* **16**, 160-169 (2020).

103. V. A. Mitkevich *et al.*, Termination of translation in eukaryotes is mediated by the quaternary eRF1*eRF3*GTP*Mg2+ complex. The biological roles of eRF3 and prokaryotic RF3 are profoundly distinct. *Nucleic Acids Res* **34**, 3947-3954 (2006).

104. J. Karijolich, Y. T. Yu, Therapeutic suppression of premature termination codons: mechanisms and clinical considerations (review). *Int J Mol Med* **34**, 355-362 (2014).

105. I. S. Fernandez *et al.*, Unusual base pairing during the decoding of a stop codon by the ribosome. *Nature* **500**, 107-110 (2013).

106. E. Svidritskiy, R. Madireddy, A. A. Korostelev, Structural Basis for Translation Termination on a Pseudouridylated Stop Codon. *J Mol Biol* **428**, 2228-2236 (2016).

107. B. Bonetti, L. Fu, J. Moon, D. M. Bedwell, The efficiency of translation termination is determined by a synergistic interplay between upstream and downstream sequences in *Saccharomyces cerevisiae*. *J Mol Biol* **251**, 334-345 (1995).

108. H. Yu *et al.*, Stop codons and the +4 nucleotide may influence the efficiency of G418 in rescuing nonsense mutations of the HERG gene. *Int J Mol Med* **44**, 2037-2046 (2019).

109. H. Adachi, Y. T. Yu, Pseudouridine-mediated stop codon read-through in S. cerevisiae is sequence context-independent. *RNA,* (2020).

110. M. Mort, D. Ivanov, D. N. Cooper, N. A. Chuzhanova, A meta-analysis of nonsense mutations causing human genetic disease. *Hum Mutat* **29**, 1037-1047 (2008).

111. T. Kurosaki, M. W. Popp, L. E. Maquat, Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat Rev Mol Cell Biol* **20**, 406-420 (2019).

112. K. M. Keeling, X. Xue, G. Gunn, D. M. Bedwell, Therapeutics based on stop codon readthrough. *Annu Rev Genomics Hum Genet* **15**, 371-394 (2014).

113. M. I. Recht, S. Douthwaite, J. D. Puglisi, Basis for prokaryotic specificity of action of aminoglycoside antibiotics. *EMBO J* **18**, 3133-3138 (1999).

114. C. Huang, J. Karijolich, Y. T. Yu, Post-transcriptional modification of RNAs by artificial Box H/ACA and Box C/D RNPs. *Methods Mol Biol* **718**, 227-244 (2011).

115. R. Barbalat, S. E. Ewald, M. L. Mouchess, G. M. Barton, Nucleic acid recognition by the innate immune system. *Annu Rev Immunol* **29**, 185-214 (2011).

116. E. K. Kelly, D. P. Czekay, U. Kothe, Base-pairing interactions between substrate RNA and H/ACA guide RNA modulate the kinetics of pseudouridylation, but not

the affinity of substrate binding by H/ACA small nucleolar ribonucleoproteins. *RNA* **25**, 1393-1404 (2019).

117. M. D. De Zoysa, G. Wu, R. Katz, Y. T. Yu, Guide-substrate base-pairing requirement for box H/ACA RNA-guided RNA pseudouridylation. *RNA* **24**, 1106-1117 (2018).

118. K. M. Dean, E. J. Grayhack, RNA-ID, a highly sensitive and robust method to identify cis-regulatory sequences using superfolder GFP and a fluorescence-based assay. *RNA* **18**, 2335-2344 (2012).

119. H. M. Goodman, M. V. Olson, B. D. Hall, Nucleotide sequence of a mutant eukaryotic gene: the yeast tyrosine-inserting ochre suppressor SUP4-o. *Proc Natl Acad Sci U S A* **74**, 5453-5457 (1977).

120. A. J. Newman, R. C. Ogden, J. Abelson, tRNA gene transcription in yeast: effects of specified base substitutions in the intragenic promoter. *Cell* **35**, 117-125 (1983).

121. S. Lee, W. A. Lim, K. S. Thorn, Improved blue, green, and red fluorescent protein tagging vectors for *S. cerevisiae*. *PLoS One* **8**, e67902 (2013).

122. W. D. Wright, S. S. Shah, W. D. Heyer, Homologous recombination and the repair of DNA double-strand breaks. *J Biol Chem* **293**, 10524-10535 (2018).

123. M. Z. Li, S. J. Elledge, Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat Methods* **4**, 251-256 (2007).

124. P. Eriksson, L. R. Thomas, A. Thorburn, D. J. Stillman, pRS yeast vectors with a LYS2 marker. *Biotechniques* **36**, 212-213 (2004).

125. O. O.-K. Nathalie Bonneaud, Guoya Li, Michel Labouesse, Lionel Minvielle-Sebastia, Francois Lacroute A Family of Low and High Copy Replicative, Integrative and Single-Stranded *S. cerevisiae*/*E. coli* Shuttle Vectors. *Yeast* **7**, 609-615 (1991).

126. M. Knop *et al.*, Epitope tagging of yeast genes using a PCR-based strategy: more tags and improved practical routines. *Yeast* **15**, 963-972 (1999).

127. M. L. L. Donnelly *et al.*, Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'. *J Gen Virol* **82**, 1013-1025 (2001).

128. P. Schattner *et al.*, Genome-wide searching for pseudouridylation guide snoRNAs: analysis of the Saccharomyces cerevisiae genome. *Nucleic Acids Res* **32**, 4281-4296 (2004).

129. P. Cramer *et al.*, Structure of eukaryotic RNA polymerases. *Annu Rev Biophys* **37**, 337-352 (2008).

130. N. Hug, D. Longman, J. F. Caceres, Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res* **44**, 1483-1495 (2016).

131. A. Urban, I. Behm-Ansmant, C. Branlant, Y. Motorin, RNA sequence and two-dimensional structure features required for efficient substrate modification by the *Saccharomyces cerevisiae* RNA:Ψ-synthase Pus7p. *J Biol Chem* **284**, 5845-5858 (2009).

132. X. Ma, X. Zhao, Y. T. Yu, Pseudouridylation (Ψ) of U2 snRNA in *S. cerevisiae* is catalyzed by an RNA-independent mechanism. *EMBO J* **22**, 1889-1897 (2003).

133. Y. Nishiuchi *et al.*, Chemical synthesis of the precursor molecule of the *Aequorea* green fluorescent protein, subsequent folding, and development of fluorescence. *Proc Natl Acad Sci U S A* **95**, 13549-13554 (1998).

134. K. E. Sloan *et al.*, Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol* **14**, 1138-1152 (2017).

135. Z. Liu *et al.*, Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Sci Rep* **7**, 2193 (2017).

136. M. L. L. Donnelly *et al.*, The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences. *J Gen Virol* **82**, 1027-1041 (2001).

137. J. H. Kim *et al.*, High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS One* **6**, e18556 (2011).

138. S. Brogna, J. Wen, Nonsense-mediated mRNA decay (NMD) mechanisms. *Nat Struct Mol Biol* **16**, 107-113 (2009).

139. E. Balleza, J. M. Kim, P. Cluzel, Systematic characterization of maturation time of fluorescent proteins in living cells. *Nat Methods* **15**, 47-51 (2018).

140. M. P. Guy *et al.*, Identification of the determinants of tRNA function and susceptibility to rapid tRNA decay by high-throughput in vivo analysis. *Genes Dev* **28**, 1721-1732 (2014).

141. X. Li, X. Xiong, C. Yi, Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat Methods* **14**, 23-31 (2016).

142. S. H. Boo, Y. K. Kim, The emerging role of RNA modifications in the regulation of mRNA stability. *Exp Mol Med* **52**, 400-408 (2020).

143. D. Dominissini *et al.*, Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201-206 (2012).

144. E. Peer, G. Rechavi, D. Dominissini, Epitranscriptomics: regulation of mRNA metabolism through modifications. *Curr Opin Chem Biol* **41**, 93-98 (2017).

145. A. S. Karim, K. A. Curran, H. S. Alper, Characterization of plasmid burden and copy number in *Saccharomyces cerevisiae* for optimization of metabolic engineering applications. *FEMS Yeast Res* **13**, 107-116 (2013).

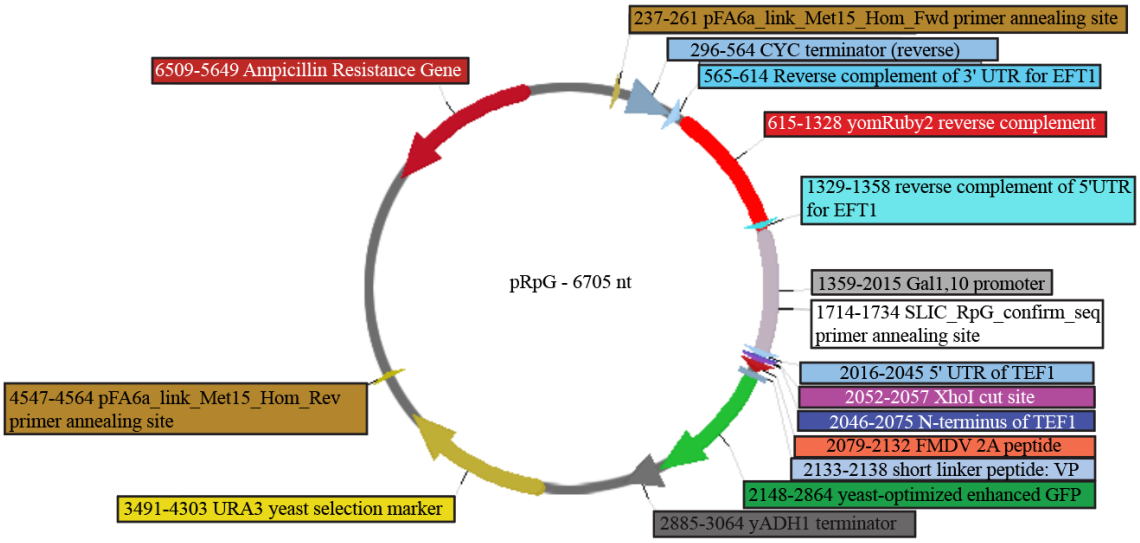
146. S. Massenet, E. Bertrand, C. Verheggen, Assembly and trafficking of box C/D and H/ACA snoRNPs. *RNA Biol* **14**, 680-692 (2017).

147. S. Deryusheva, J. G. Gall, scaRNAs and snoRNAs: Are they limited to specific classes of substrate RNAs? *RNA* **25**, 17-22 (2019).

148. N. S. Heiss *et al.*, X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nat Genet* **19**, 32-38 (1998).

149. D. Ruggero *et al.*, Dyskeratosis congenita and cancer in mice deficient in ribosomal RNA modification. *Science* **299**, 259-262 (2003).

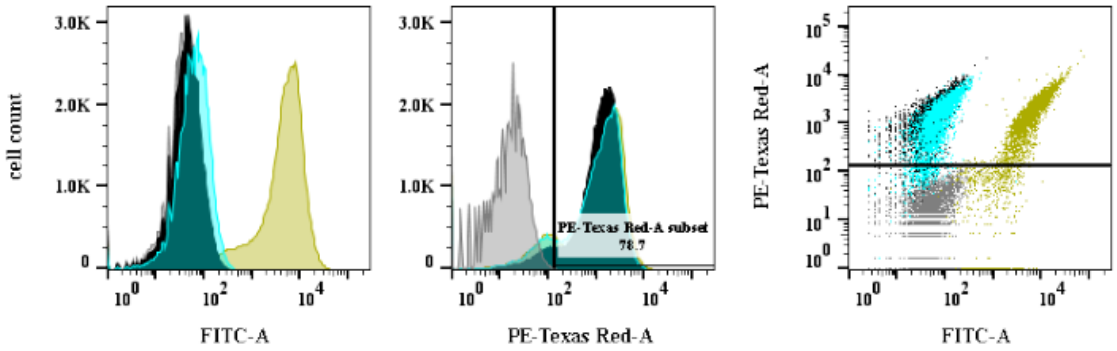
Appendix 1–Map of pFA6a-yomRuby2-pGal1,10-yoEGFP plasmid



S1) Map of pFA6a-yomRuby2-pGal1,10-yoEGFP (pRpG) plasmid.

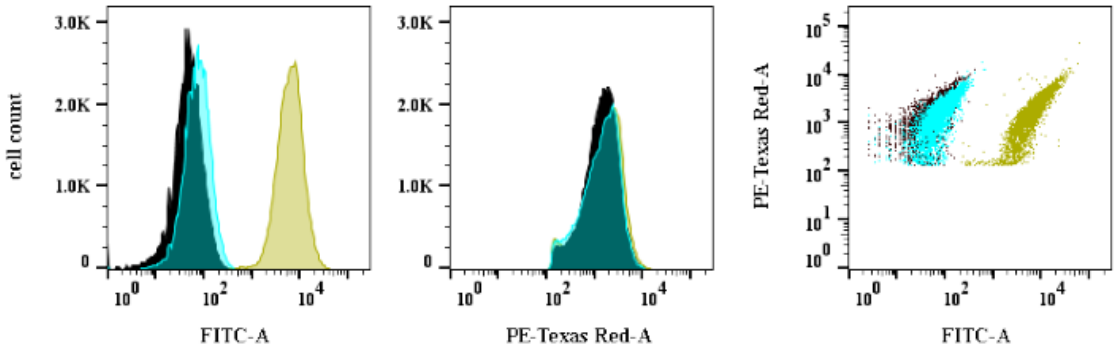
Appendix 2–Gating for yeast with high Ruby2 signal during flow cytometry analysis

A snR5 Assay Pre-Gating



Sample Name	Median : FITC-A	Median : PE-Texas Red-A	Percentage
WT Yeast	40.0	12.0	100
snR5-Test Pre-Gate	64.3	1110	100
snR5-Neg Pre-Gate	45.2	1235	100
snR5-Pos Pre-Gate	4983	1312	100

B snR5 Assay Post-Gating



Sample Name	Subset Name	Median : FITC-A	Median : PE-Texas Red-A	Percentage
snR5-Test Post-Gate	PE-Texas Red-A subset	74.9	1473	81.4
snR5-Neg Post-Gate	PE-Texas Red-A subset	48.8	1420	87.6
snR5-Pos Post-Gate	PE-Texas Red-A subset	6167	1758	78.7

S2) Gating for yeast cells with high Ruby2 expression. (A) For each flow cytometry assay, 100000 yeast cells were initially analyzed. To select for yeast cells that only exhibited a high level of Ruby2 signal expression, a gating step was performed during data analysis. The red fluorescence histogram from a BY4742 wild type or ΔUPF1 yeast strain analyzed in parallel to the fluorescent reporter strains was compared to the histograms from the test, positive, and negative reporter strain variants. From this histogram overlay, a data subset was created that excluded any cells of the test, positive, and negative strains that overlapped with the BY4742 wild type or ΔUPF1 red fluorescence distribution. An

example of this gating process is given for flow cytometry analysis of snR5 H/ACA snoRNP activity from the fluorescent reporter system in a ΔUPF1 background. The left histogram overlay depicts the cell count vs. relative green fluorescence detected with a FITC-A filter for the test, positive, and negative control strains alongside a wild type yeast control. The middle histogram overlay depicts the cell count vs. relative red fluorescence detected with a PE-Texas Red-A filter in which the vertical line denotes the red signal threshold for inclusion in the final data set. The right overlay presents the red vs. green signal-expression dot plots for the test, positive, and negative control strains alongside the wild type yeast control with the red-signal threshold denoted by the horizontal line. Median red and green fluorescence values for each strain are given in the table. **(B)** The cells that passed the Ruby2 selection described in (A) were then included in the final flow cytometry analysis. The percentage of the high Ruby2 expressing cell population from each strain included in the final analysis is given in the table alongside the post-gating median red and green fluorescence intensities of each yeast strain analyzed.