

## ABSTRACT

Title of Dissertation: MESSENGER RNA DESTABILIZATION BY -1  
PROGRAMMED RIBOSOMAL FRAMESHIFTING

Ashton Trey Belew, Doctor of Philosophy, 2012

Dissertation Directed by: Professor Jonathan D. Dinman  
Department of Cell Biology and Molecular Genetics

Although first discovered in viruses, previous studies have identified programmed -1 ribosomal frameshifting (-1 PRF) signals in eukaryotic genomic sequences, and suggested a role in mRNA stability. This work improves and extends the computational methods used to search for potential -1 PRF signals. It continues to examine four yeast -1 PRF signals and show that they promote significant mRNA destabilization through the nonsense mediated (NMD) and no-go (NGD) decay pathways. Yeast EST2 mRNA is highly unstable and contains up to five -1 PRF signals. Ablation of the -1 PRF signals or of NMD stabilizes this mRNA. These same computational methods identified an operational programmed -1 ribosomal frameshift (-1 PRF) signal in the human mRNA encoding CCR5. A -1 PRF event on the CCR5 mRNA directs translating ribosomes to a premature termination codon, destabilizing it through the nonsense-mediated mRNA decay (NMD) pathway. CCR5-mediated -1 PRF is stimulated by at least two miRNAs, one of which is shown to directly interact with the CCR5 -1 PRF signal. Structural analyses reveal a complex and dynamic mRNA structure in the -1 PRF signal, suggesting structural plasticity as the underlying biophysical basis for regulation of -1 PRF.

MESSANGER RNA DESTABILIZATION BY -1 PROGRAMMED RIBOSOMAL  
FRAMESHIFTING

By

Ashton Trey Belew

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park, in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
2012

Advisory Committee:

Professor Jonathan D. Dinman, Chair

Associate Professor Jason D. Kahn

Associate Professor Stephen M. Mount

Associate Professor Richard C. Stewart

Associate Professor Michael P. Cummings, Dean's Representative

© Copyright by  
Ashton Trey Belew  
2012

## **Preface**

It would not be called research if we knew what we were doing.



## **Acknowledgements**

I would like to thank Dr. Dinman for his years of critical thought, humor, and support; my committee for their time and assistance in improving my research experience; and the Dinman lab, past and present. Thank you, Sean, for reminding me that it is best to run towards something.

This work was supported in part by a National Institute of Health grant to Dr. Jonathan Dinman.

# Table of Contents

Preface.....	ii
Acknowledgements .....	iii
Table of Contents .....	iv
List of Tables.....	vii
List of Figures.....	viii
List of Abbreviations .....	x
Chapter 1 .....	1
Introduction.....	1
The mRNA Life Cycle .....	4
Translation Subversion .....	5
Subverting Initiation .....	8
Restarting Elongation.....	9
Modifying Termination.....	10
Examples in the Genomic Context .....	11
Models of -1 Programmed Ribosomal Frameshifting .....	13
Nonsense Mediated Decay .....	16
NMD: Increasing Complexity.....	17
NMD: Current Model in <i>Saccharomyces cerevisiae</i> .....	20
The Upf proteins .....	25
Other Factors.....	29
Important Nucleases .....	30
Small RNAs and Genome Regulation .....	32
Computational Searches for mRNA Structure.....	35
Computational Pipelines to Filter Data .....	35
Grammar Searches in Large Datasets (RNAMotif).....	36
Predictive vs. Statistically Informed Searches .....	36
Sequence Randomization Strategies .....	39
Storage, Retrieval, and Visualization.....	40
Summary .....	41
Chapter 2 .....	42
A Database of Computationally Predicted Programmed -1 Ribosomal Frameshift	
Signals .....	42
Introduction.....	42
Database Description.....	43
Process .....	47
Discussion .....	48
Chapter 3 .....	49
Endogenous Ribosomal Frameshift Signals Operate as mRNA Destabilizing Elements	
Through at least Two Molecular Pathways in Yeast.....	49
Introduction.....	49
Results .....	51
The EST2 -1 PRF signal at Nucleotide 1653 is Destabilized by -1 PRF Induced	
NMD .....	54
Ablation of -1 PRF Signals Stabilizes the Yeast <i>EST2</i> mRNA.....	55
Programmed -1 Ribosomal Frameshifting, but not Specific -1 PRF Signals,	
Appears to be Conserved Among Yeasts. ....	59
Discussion .....	60
Materials and Methods.....	66

Strains, Genetic Manipulations, and Media. ....	66
Generation of mRNA Stability Vectors. ....	67
Generation of <i>EST2</i> Open Reading Frame Mutants.....	67
Steady State and Time Course RNA Blot Analyses. ....	68
Quantitative Real Time Reverse Transcription PCR.....	68
Comparative Analyses. ....	69
Chapter 4 .....	70
Ribosomal Frameshifting in CCR5: Regulation by NMD, miRNAs and Conformational Plasticity. ....	70
Introduction.....	70
Results .....	72
A functional -1 PRF signal in the <i>H. sapiens</i> CCR5 receptor mRNA.....	72
The CCR5 -1 PRF Signal Destabilizes a Reporter mRNA through the NMD Pathway in Mammalian Cells. ....	76
Specific stimulation of CCR5-mediated -1 PRF by miR-1224 in HeLa cells.....	79
hsa-miR-1224 directly interacts with the CCR5 -1 PRF signal.....	81
The Native CCR5 mRNA is Affected by NMD and hsa-miR-1224. ....	84
The CCR5 -1 PRF Signal Encodes a Complex and Dynamic set of mRNA Conformers. ....	85
Identification of Additional -1 PRF Signals in Other Human Interleukin Receptor mRNAs. ....	91
Discussion .....	93
-1 PRF and the Immune System. ....	93
Modes of -1 PRF Regulation.....	94
Structural Plasticity. ....	95
Effects on the Viral Context.....	97
Chapter 5 .....	98
Where do we go from here? .....	98
Computational Next Steps .....	99
Correlation Studies .....	100
Possible next steps in <i>Saccharomyces cerevisiae</i> .....	102
Future Mammalian Work.....	107
Chapter 6 .....	111
Experimental Procedures .....	111
Introduction.....	111
Cell Culture.....	111
The Dual Luciferase Assay .....	113
Flow Cytometry, Fluorescence Activated Cell Sorting and Western Blotting .....	114
FACS .....	114
Western Blotting .....	115
Northern Blotting.....	115
Plasmid Construction.....	116
PCR Amplification, Restriction Enzyme Digestions, and Ligation. ....	116
Oligonucleotide Annealing and Ligation. ....	118
Site Directed Mutagenesis and “Mega-oligo” Mutagenesis.....	118
Quantitative Real Time Reverse Transcription PCR. ....	119
Mammalian Cells .....	119
Yeast Cells.....	120
RNA Extraction .....	120
Mammalian Cells .....	120
Yeast Cells.....	121

SHAPE .....	121
Southern Blotting .....	121
Time Course Assays.....	122
Mammalian .....	122
Yeast .....	122
Transfections .....	123
Plasmid.....	123
miRNA .....	123
siRNA .....	124
Transformations.....	124
Bacteria .....	124
Yeast .....	125
Appendices .....	126
Appendix 1: Yeast Strains Used .....	126
Appendix 2: Yeast Plasmids .....	127
Appendix 3: Mammalian Plasmids .....	132
Appendix 4: Cloning Oligonucleotides .....	134
Appendix 5: Site Directed Mutagenesis Oligonucleotide List.....	141
Appendix 6: Oligonucleotides Used for Sequencing .....	145
Appendix 7: Oligonucleotides Used for qPCR and Northern Analyses .....	147
Appendix 8: RNAi Oligonucleotides .....	149
Appendix 9: Oligonucleotides Used for SHAPE / <i>in-vitro</i> Translation .....	150
Appendix 10: Code Summary .....	152
Appendix 11: Partial Summary of the PRFdb.....	154
Appendix 12: Categorized Recoding Signals .....	156
Appendix 13: Observed Viral Origin -1 PRF Signals. ....	157
Appendix 14: -1 PRF Signals are not Conserved across Yeast Species .....	159
SPR6 -1 PRF signals in <i>Saccharomyces</i> species .....	159
EST2 -1 PRF signals in <i>Saccharomyces</i> species.....	159
BUB3 -1 PRF signals in <i>Saccharomyces</i> species .....	160
TBF1 -1 PRF signals in <i>Saccharomyces</i> species.....	160
Appendix 15: Interleukin Receptor Genes Containing -1 PRF Signals .....	162
References.....	163

## List of Tables

Table 1: Classifying intron containing genes in <i>Saccharomyces cerevisiae</i> .....	21
Table 2: Observed RNA abundances for different classes of ORFs.....	21
Table 3: A partial listing of proteins involved in <i>Homo sapiens</i> NMD.....	29
Table 4: Summary of some common RNA prediction algorithms .....	37
Table 5: -1 PRF Candidates Significantly Disregulated in HCC .....	109
Table 6: Yeast strains used in Chapter 3. ....	126
Table 7: Plasmids used with yeast.....	131
Table 8: Mammalian plasmid list.....	133
Table 9: Oligonucleotides used for cloning. ....	140
Table 10: Oligonucleotides used for Site Directed Mutagenesis .....	144
Table 11: Oligonucleotides used for sequencing .....	146
Table 12: Oligonucleotides used as probes for qPCR and Northernns .....	148
Table 13: Oligonucleotides used in RNAi studies .....	149
Table 14: Oligonucleotides used for SHAPE and <i>in-vitro</i> transcription.....	151
Table 15: A partial summary of the PRFdb .....	155
Table 16: A partial bestiary of recoding signals.....	156
Table 17: A catalog of known viral origin -1 PRF signals .....	158
Table 18: -1 PRF signals are not conserved across yeast species .....	161

## List of Figures

Figure 1: The distribution of -1 PRF signal in Homo sapiens. ....	4
Figure 2: The translation cycle.....	6
Figure 3: Peptide bond formation in the ribosome. ....	8
Figure 4: Visualizing circular mRNA by atomic-force microscopy.....	9
Figure 5: The spectrum of IRES factor requirements. ....	10
Figure 6: The organization and results of viral -1 PRF. ....	13
Figure 7: The elements of a -1 PRF signal. ....	14
Figure 8: A Model Mechanism of -1 PRF. ....	15
Figure 9: Kinetic Partitions which lead to -1 PRF. ....	17
Figure 10: Genomic -1 PRF signals do not significantly extend the ORF. ....	18
Figure 11: Translating ribosomes protect mRNPs from NMD. ....	19
Figure 12: The surveillance complex model in yeast.....	22
Figure 13: Toeprints of normal and premature termination. ....	23
Figure 14: The Faux 3' UTR Model of NMD.....	25
Figure 15: The domains and interactions of Upf1p.....	26
Figure 16: Long distance movement of UPF1 CH domain upon UPF2 binding.....	28
Figure 17: Crystal Structure of the Archaeal Exosome .....	31
Figure 18: Methods of mRNA decay.....	31
Figure 19: miRNA Biogenesis.....	33
Figure 20: The PRFdb Search Interface .....	44
Figure 21: A Representative Search Result. ....	44
Figure 22: The distribution of <i>Saccharomyces cerevisiae</i> sequences. ....	45
Figure 23: Details of the <i>Homo sapiens</i> CCR5 -1 PRF signal. ....	46
Figure 24: Schematic of PGK1 reporter vectors.....	52
Figure 25: Genomic -1 PRF signals can function as mRNA destabilization elements in yeast. ....	53
Figure 26: The <i>EST2</i> -1 PRF signal at position 1653 destabilizes mRNA through NMD. ....	56
Figure 27: Four additional high confidence -1 PRF signals in <i>EST2</i> . ....	57
Figure 28: Silent mutations that disrupt slippery sites in <i>EST2</i> gene stabilize its mRNA. ....	58
Figure 29: Map of the full length <i>EST2</i> low-copy plasmids.....	58
Figure 30: MFE 'landscape' of <i>SPR6</i> and predicted PRF signal at position 279. ....	59
Figure 31: -1 PRF signals can destabilize mRNA via NMD and NGD. ....	61
Figure 32: Computationally predicted -1 RF signals in <i>EST1</i> , <i>STN1</i> , <i>CDC13</i> , and <i>ORC5</i> .....	64
Figure 33: Distributions of potential -1 PRF signals from <i>Homo sapiens</i> .....	73
Figure 34: A <i>cis</i> -acting element in the CCR5 mRNA promotes efficient -1 PRF.....	74
Figure 35: The CCR5 -1 PRF signal is active in CHO and Vero cells.....	75
Figure 36: Rabbit reticulocyte lysates confirm CCR5 frameshifting.....	75
Figure 37: The rabbit $\beta$ -globin mRNA stability reporter. ....	76
Figure 38: The CCR5 -1 PRF signal acts as an mRNA destabilizing element. ....	77
Figure 39: Rabbit $\beta$ -globin reporter half-life measurements. ....	78

Figure 40: siRNA knockdown of NMD increases the amount of CCR5 reporter mRNA. ....	78
Figure 41: NMD knockdowns increase other NMD mRNA abundance.....	79
Figure 42: Stimulation of CCR5-mediated -1 PRF by hsa-miR-1224. ....	80
Figure 43: hsa-miR-1224 directly interacts with the CCR5 -1 PRF signal <i>in vitro</i> . ....	82
Figure 44: hsa-miRNA-1224-5p interacts with the CCR5 mRNA <i>in vivo</i> .....	83
Figure 45: Effects of NMD abrogation and miR-1224 on the native CCR5 mRNA.....	84
Figure 46: Prediction and summary of the CCR5 PRF signal. ....	86
Figure 47: Chemical protection analysis of the CCR5 -1 PRF signal. ....	87
Figure 48: Force-extension curves and dE/dx density plots of the CCR5 structure. ....	89
Figure 49: Observed sub-steps when pulling apart the CCR5 -1 PRF signal.....	90
Figure 50: Computationally predicted human cytokine receptor -1 PRF signals.....	91
Figure 51: Efficient -1 PRF is promoted by sequences in additional cytokine receptors. ....	92
Figure 52: Position of 28mer reads with respect to reading frame in ribosome profiling. ....	100
Figure 53: MFE prediction vs. ribosome footprint density vs. PARS score for <i>PDR5</i> . ....	101
Figure 54: Histograms of potential genomic -1 PRF signals.....	103
Figure 55: The <i>EST2</i> mRNA is strongly protected position 1653 -1 PRF signal.....	104
Figure 56: NMD deficient and <i>EST2</i> slippage mutant cells have shorter telomeres. ....	105
Figure 57: Initial attempts to quantify the relative contributions of each <i>EST2</i> PRF signal.....	106
Figure 58: Counting the maximum number of generations of mutant <i>EST2</i> harboring cells.....	106
Figure 59: Dual luciferase of HCC involved -1 PRF signals.....	110
Figure 60: Possible interaction of hsa-miR-101 with the <i>OLA1</i> -1 PRF signal.....	110

## List of Abbreviations

1M7	1-methyl-7-nitroisatoic anhydride
AMPPNP	5'-adenylyl imidodiphosphonate
ATP	Adenosine Triphosphate
DSE	Downstream Element
dsRNA	Double stranded RNA
CCR5	Chemokine receptor 5
CDS	Coding sequence
CHX	Cycloheximide
CrPV	Cricket paralysis virus
DLR	Dual luciferase assay system
DMEM	Dulbecco's modified eagle medium
EDTA	Ethylenediaminetetraacetic acid
GTP	Guanine triphosphate
HIV	Human immunodeficiency virus (usually type 1)
IRES	Internal ribosome entry site
LB	Luria bertani broth
LSU	Large subunit of the ribosome
MCS	Multiple cloning site
MFE	Minimum Free Energy
M <sup>7</sup> G	7-methylguanosine
MOPS	3-[N Morpholino] propane sulfonic acid
mRNA	Messenger RNA
miRNA	MicoRNA
ncRNA	Non-coding RNA
NGD	No-go decay
NMD	Nonsense mediated decay
NSD	Non-stop decay
PABP	Poly-adenosine binding protein
PCR	Polymerase chain reaction
PERL	Practical extraction and report language
PMSF	Phenylmethysulfonyl fluoride
PPCC	Probability plot correlation coefficient
pol	Polymerase
poly A	Poly-adenosine
PRF	Programmed ribosomal frameshifting
PTC	Premature termination codon
qPCR	Quantitative (real time) PCR
RF	Release Factor
RSV	Rous sarcoma virus
RT	Read through (control)
rPTC	Peptidyl transferase center (ribosomal)
SARS	Severe acute respiratory syndrome
SAXS	Small Angle X-ray Scattering
SCFG	Stochastic Context Free Grammar
SHAPE	Selective 2'-hydroxyl acylation analyzed by primer extension
siRNA	Short interfering RNA
ssRNA	Single stranded RNA
SSU	Small subunit of the ribosome



TBE	Tris borate acid, EDTA
TCV	Turnip crinkle virus
T <sub>m</sub>	Melting temperature, in celsius
tmRNA	Transfer-messenger RNA
tRNA	Transfer RNA
UTR	Untranslated region

# Chapter 1

## Introduction

Cells regulate gene expression via diverse mechanisms. From mRNA transcription to protein degradation, many regulatory systems affect the timing, localization, and rate of each reaction. Gene expression is primarily concerned with the abundance and translational activity of mRNA; therefore expression is increased when a message is transcribed more rapidly, stabilized by the cell, or more available to actively translating ribosomes. Increased mRNA degradation, decreased transcription, translational silencing, and the storage of mRNA are the hallmarks of decreased expression. Post-transcriptional regulation of gene expression is a growing field of inquiry; it has primarily concentrated on *cis*-acting elements in the 5' and 3' untranslated regions (UTRs) of mRNAs, and the *trans*-acting factors with which they interact. Protein coding regions have not been as closely examined for effects on post-transcriptional control; however multiple *cis*-acting mRNA elements have been found which cause elongating ribosomes to recode the mRNA sequence [for review see [1], **Appendix 12**]. These include, but are not limited to, sequences responsible for +1 and -1 programmed ribosomal frameshifting, termination suppression, stop-start elements, selenocysteine incorporation in all kingdoms of life, and pyrrolysine incorporation in archaea[2]. More recently, programmed recoding has taken on a whole new dimension, as artificial tRNA synthetases and codon:anticodon pairs provide the means to literally re-encode the genome [introduced in [3] and reviewed in [4]].

At its heart, this work is an attempt to tie together a group of disparate observations; each with its own background, literature, and state of the art. The primary determinants of highly transcribed, functional, and stable mRNAs are well established[5].

In specific instances the opposite is also true; some destabilizing elements have been well characterized. Similarly, the viral determinants of -1 PRF have been described in numerous instances[6]. The mechanisms of -1 PRF are still debated, but center around specific stages during translation elongation[7]. Even the contentious fields of NMD[8] and ncRNA research[9] are coalescing around relatively well defined hypotheses. The questions remain: how do these mechanisms interact? How do we find specific examples which demonstrate the interplay between mRNA structure, translational fidelity, and mRNA stability? Finally, what future avenues of research do these observations suggest? No computational nor laboratory methodologies provide the answers, but existing tools provide glimpses and opportunities to search for and analyze strong candidates. In the following sections, I will provide the background of these post-transcriptional processes and the computational methodologies involved.

This thesis focuses upon -1 programmed ribosomal frameshifting (PRF); which is a recoding mechanism historically associated with viruses[10]<sup>a</sup> and retrotransposons[11,12]<sup>b</sup>. A PRF signal stochastically redirects translating ribosomes into an alternate reading frame. A cursory glance at Appendix 12 shows that, with the notable exception of *Thermus thermophilus*' extraordinary poly-U slippage event, -1 PRF follows a generally consistent pattern of a stimulatory element immediately downstream of a group of weakly pairing bases. On the other hand, +1 PRF, ribosomal shunting, and suppression events intermingle proximal RNA secondary structures, distant RNA structures, mixtures of rare 0 frame and "hungry" +1 frame codons, and even alternate tRNA species. This work therefore limits its scope to canonical -1 PRF as initially described in the Rous sarcoma virus[10]. In this context, a -1 PRF signal leads to

---

<sup>a</sup> Jacks and Varmus introduce the first -1 programmed ribosomal frameshifting signal in this paper.

<sup>b</sup> Craigen et al. describe +1 PRF in the expression of bacterial release factor 2. In the same year, Farabaugh published a speculative paper suggesting +1 PRF in the Ty1 retrotransposon, which was later proved in Clare et al. by sequencing cDNA from the Ty1 mRNA.

expression of the Gag-pol polyprotein from two overlapping ORFs. Thus, a viral -1 PRF signal allows ribosomes to bypass the 0-frame stop codon and continue synthesis of a C-terminal extended fusion protein[13]<sup>c</sup>. Though PRF was first described in viruses, we now have evidence that organisms in all three kingdoms of life employ PRF[14–16]<sup>d</sup>, opening the possibility for a conserved mechanism affecting the expression of actively translated mRNAs[17].

This peculiar post-transcriptional regulation system ties together the mRNA/ribosome interactions which occur in viral -1 PRF (for review, see [18]), with the surveillance mechanisms against aberrant, truncated messages (for reviews, see [8,19,20]). While the requirements of PRF are well understood in viruses and a specific set of transcripts, the potential of PRF outside these contexts has not been fulfilled. For example, while splicing dependent Nonsense Mediated Decay (NMD) substrates have been well characterized[21]<sup>e</sup>, there remain many more transcripts which are known to be up-regulated when NMD is knocked out and cannot be fully explained[22]<sup>f</sup>. One final observation lays the foundation for this work: more than 95% of predicted genomic -1 PRF signals are followed not by a C-terminal extension, but by a -1 frame stop codon[23,24]; as summarized in **Figure 1** for the ORFs of the *Homo sapiens* genome, which is consistent with observations of most other genomes observed (**Figure 54**). Taken together, these disparate observations engender the hypothesis that active translation of -1 PRF signal containing mRNAs negatively regulates these messages via NMD and/or No-Go Decay (NGD). This hypothesis assumes that it is possible to find

---

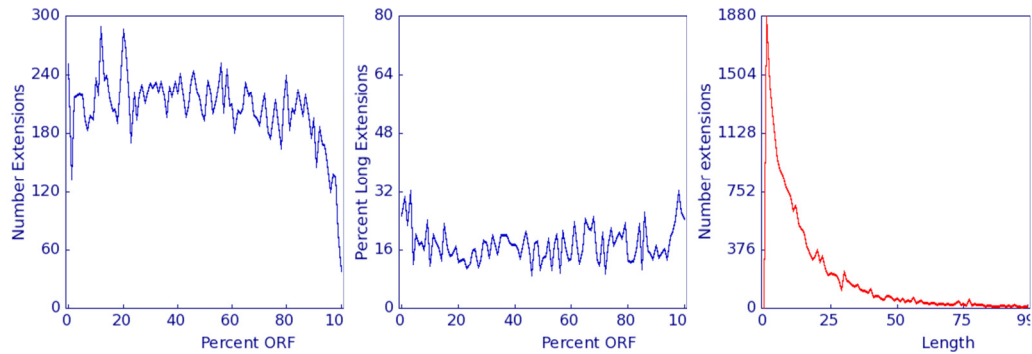
<sup>c</sup> This paper provides some early examples of PRF in viruses, notably in Table 2.

<sup>d</sup> Cobucci-Ponzano observed the frameshift products of the archaeal  $\alpha$ -l fucosidase via Mass Spectrometry (MALDIMS) in their first publications, and later provided more examples in the following review. In Brierley's 2004 paper, they switch the focus from viral PRF to bacterial and eukaryotes.

<sup>e</sup> ~2,000 of 5,693 mRNAs were found with alternative splicing isoforms which lead to PTCs in the coding exons of RefSeq.

<sup>f</sup> Microarray experiments with *Saccharomyces cerevisiae* NMD knockout strains found a set of transcripts which are upregulated without explanation, as explained in the text describing Figure 2.

and evaluate novel decay substrates in sequence databases, and that this form of post-transcriptional regulation is regulated by *trans*-acting factors. Work performed in the mammalian context adds the observation that non-coding RNAs (ncRNA) provide transcript specific regulation of -1 PRF substrates.



**Figure 1: The distribution of -1 PRF signal in *Homo sapiens*.**

**Left:** A histogram of the number of potential -1 PRF signals with respect to relative ORF position in *Homo sapiens*. **Middle:** The percent of the total which is predicted to translate the -1 reading frame by 30 or more codons. Of note is the marked decrease of the total from 90-100% and simultaneous increase in the percentage which extend by 30 or more codons (left vs. middle). **Right:** The distribution of length of -1 PRF signal encoded peptides in *Homo sapiens*. More examples may be found in Figure 54.

## The mRNA Life Cycle

Messenger RNA expression is tightly controlled from transcription until degradation. RNA polymerase II (Pol) recruitment to DNA and activation is controlled by binding; these may recruit or block Pol II as well as stimulate or inhibit transcription elongation[25]. Maturation of nascent transcripts via m<sup>7</sup>G addition, splicing, and polyadenylation protect these new messages from decay, while the UTRs harbor powerful signals which recruit or block mRNA decay. One example which proved useful in this work includes AU rich elements in the 3' UTR[26]; these promote mRNA decay[27]<sup>9</sup>. *Trans*-acting factors are important in other instances; in one particularly powerful example, loss of 3' UTR was shown to activate proto-oncogenes partially

<sup>9</sup> The highlight of this paper is the early time-course assay of  $\beta$ -globin RNA.

because these mRNAs no longer contained binding sites for miRNAs which kept them under control[28]. Indeed, in the case of the yeast HO endonuclease, two separate PUF proteins have been found to simultaneously bind the 3' UTR such that full repression occurs only when **both** are bound[29]. On the other hand, the amino acid starvation response is mediated partially through the relocalization of mRNAs out of processing bodies (P bodies) mediated by the 3' UTR and other miRNA binding sites[30]<sup>h</sup>.

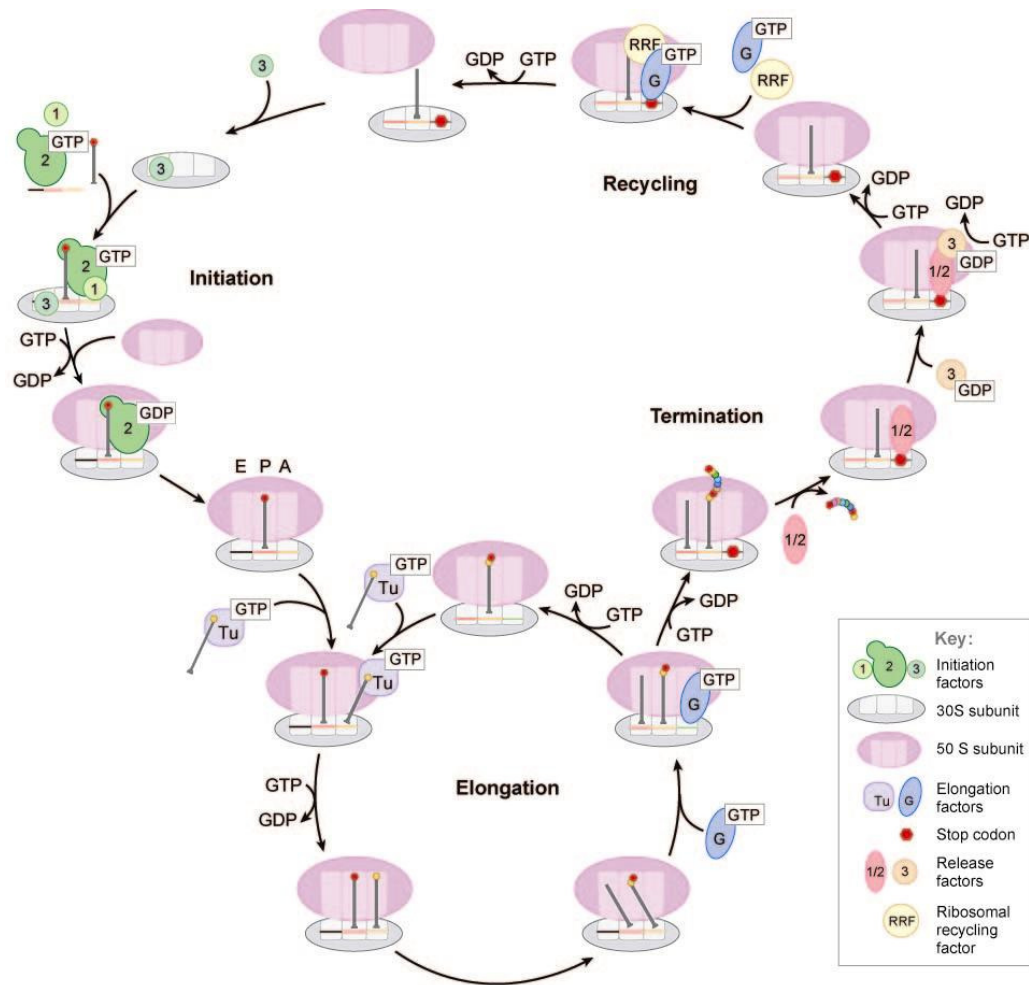
Mature, properly formed, and localized messages remain at risk from attack by ncRNAs; including micro-RNAs (miRNA) (reviewed in [31]) and short interfering RNAs (siRNA) (reviewed in [32]). These quickly recruit the cell's silencing machinery in a sequence specific manner, leading to rapid endonucleolytic cleavage, decapping, deadenylation, and exonucleolytic decay. Messages which maintain the balance of these transcriptional control elements are translated and subject to a separate set of post-transcriptional regulatory mechanisms. In order to address these mechanisms, it is first necessary to address translation as it progresses from initiation, through elongation, termination, and eventually recycling.

### **Translation Subversion**

The actively translating ribosome coordinates an exquisitely complex series of individual activities to complete the cycle of initiation, elongation, and termination (**Figure 2**). Once the translational reading frame is established during initiation, this reading frame must be maintained throughout elongation. During each cycle, the ternary complex (in eukaryotes comprising eEF1A, the tRNA, and GTP) delivers a new elongator tRNA to the decoding center of the small ribosomal subunit. If the amino-acyl tRNA anticodon is complementary to the mRNA codon, they form a helix stabilized by

---

<sup>h</sup> Bhattacharyya et al. used polysome fraction analysis and immunofluorescence microscopy to demonstrate that the CAT-1 3' UTR determines its translational status and localization. It goes on to demonstrate hsa-miR-122 interacting with the CAT-1 3' UTR.



**Figure 2: The translation cycle.**

The current model of the prokaryotic translation cycle includes initiation, elongation, termination, and recycling. The initiation factors bring together the large and small ribosomal subunits (50S and 30S respectively in bacteria) with the initiator tRNA located in the small subunit P site on the mRNA start codon. Elongation factors (bacterial EFG and EF-Tu) continue to bring new tRNAs to the ribosome and promote translocation through the elongation cycle in a GTP dependent manner. When the elongating ribosome reaches a stop codon, release factors are incorporated instead of aminoacyl tRNA, thus freeing the newly synthesized polypeptide. Finally, ribosomal recycling separates the subunits, removes the peptidyl tRNA and mRNA, and leaves the system ready for another round of initiation. This figure is from Marshall et al.[33].

interactions among the small subunit (SSU) rRNA and SSU protein S12[34]<sup>i</sup>. In turn this causes eEF1A to hydrolyze GTP and release the tRNA; it is this process which allows

<sup>i</sup> If you can see stereoscopic images, Figure 3 shows crystal structures of codon:anticodon base pairs.

the aminoacylated 3' end to move from the periphery of the large subunit into the A-site of the peptidyl transfer center (PTC) via the accommodation corridor.

Peptidyl transfer catalysis occurs in the PTC via positioning of the incoming tRNA and an active (requiring energy from GTP hydrolysis) transesterification reaction performed by the ribosome (**Figure 3**, reviewed in [35]). During this reaction, the 3' end of the deacylated tRNA moves into the E-site of the large subunit. Simultaneously, the A-site tRNA, which just received the peptide, moves into the P-site; the anticodon loops of both tRNAs remain in the small subunit P and A sites respectively, creating the hybrid state of the ribosome. The entire ribosome rotates from the “classical” to “rotated” state[36]<sup>j</sup> during this step. Translocation follows and defines the next step in reading frame maintenance. Translocation begins when eEF2 is recruited to the ribosome, leading to GTP hydrolysis, which provides the energy required to separate the tRNAs from the mRNA, and the energy to move the anticodon loops from the P and A sites to the E and P sites respectively[37]<sup>k</sup>. Upon completion, this leaves the A site empty and ready for the next ternary complex. Recent methodologies, including X-ray crystallography, cryo-EM, and single molecule FRET experiments, have elucidated the structural features of the ribosome which ensure that the tRNAs remain correctly positioned, and ensure that translocation is limited to three nucleotides[38]<sup>l</sup>.

Furthermore, actively expressed eukaryotic mRNAs maintain an assembly of factors including: the m<sup>7</sup>G cap and polyA tail circularized by polyA binding protein (PABP); initiation factors 4G and 4E (eIF4G, eIF4E) (**Figure 4**, reviewed in [39]); and the

---

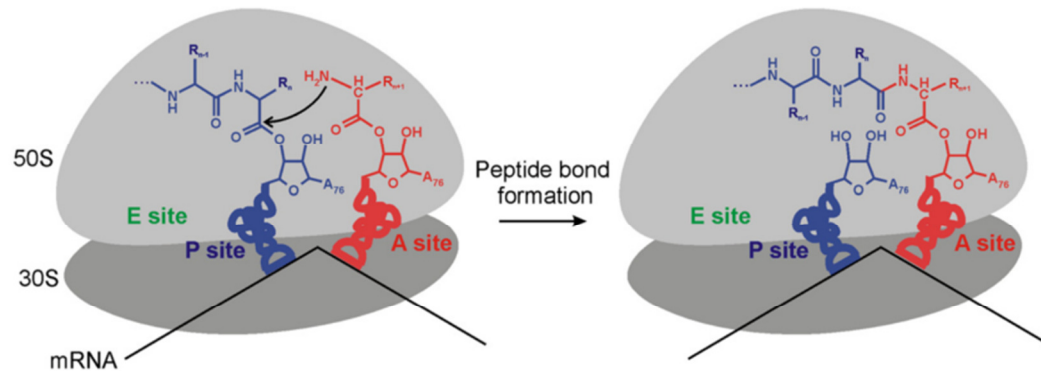
<sup>j</sup> Figure 1 of this paper has a nice way of demonstrating the rotation of the ribosome, while Figure 2 shows the extent of tRNA bending from the P/P to the P/E state.

<sup>k</sup> Filterbinding assays were performed which suggest that tightly coupled ribosomes harbor 3 tRNAs / ribosome (Figure 3C), thus beginning the hypothesis of the E-site and a long-term disagreement with Wintermeyer.

<sup>l</sup> A movie of the cycle of translocation was created from the Cryo-EM data. It is the best part of this paper.



splicing complex (reviewed in [40]). These mRNAs must maintain the processivity of translating ribosomes to stave off the surveillance complex, lest they be rapidly



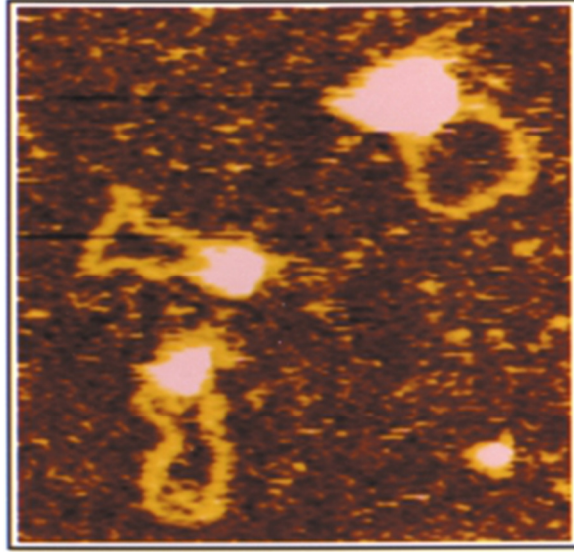
**Figure 3: Peptide bond formation in the ribosome.**

A new peptide bond is formed when the  $\alpha$ -amino group of aminoacyl-tRNA (red) attacks the carbonyl carbon (shown by the black arrow on the left) of the peptidyl-tRNA in the P-site (blue). This leaves a one amino acid longer peptidyl-tRNA in the A site and a deacylated tRNA in the P site. This figure is from Beringer & Rodnina[35].

degraded via the nonsense mediated decay (NMD) [41,42], non-stop decay (NSD)[43,44] or no-go decay (NGD)[45] pathways.

### Subverting Initiation

Given this interplay of individually complex systems, it is no surprise that every phase of translation is subverted by so many viral and cellular systems. Internal Ribosome Entry Sites (IRESes) mimic the initiation complex via many methods, such that the entire spectrum of initiation factors may or may not be required for translation of these RNAs (reviewed in [46]). The CrPV IRES requires none of the endogenous initiation factors, while the polio virus IRES uses all of the endogenous machinery except eIF4e (**Figure 5**). Viral and endogenous elements derail elongating ribosomes to diverse ends. Two methods are primarily employed in this context: disrupting the kinetics of elongation and molecular mimicry. A programmed ribosomal frameshift event is an excellent example of what happens when the normal dynamic of elongation is stalled by either a strong mRNA element, (the kinetic parameters are described



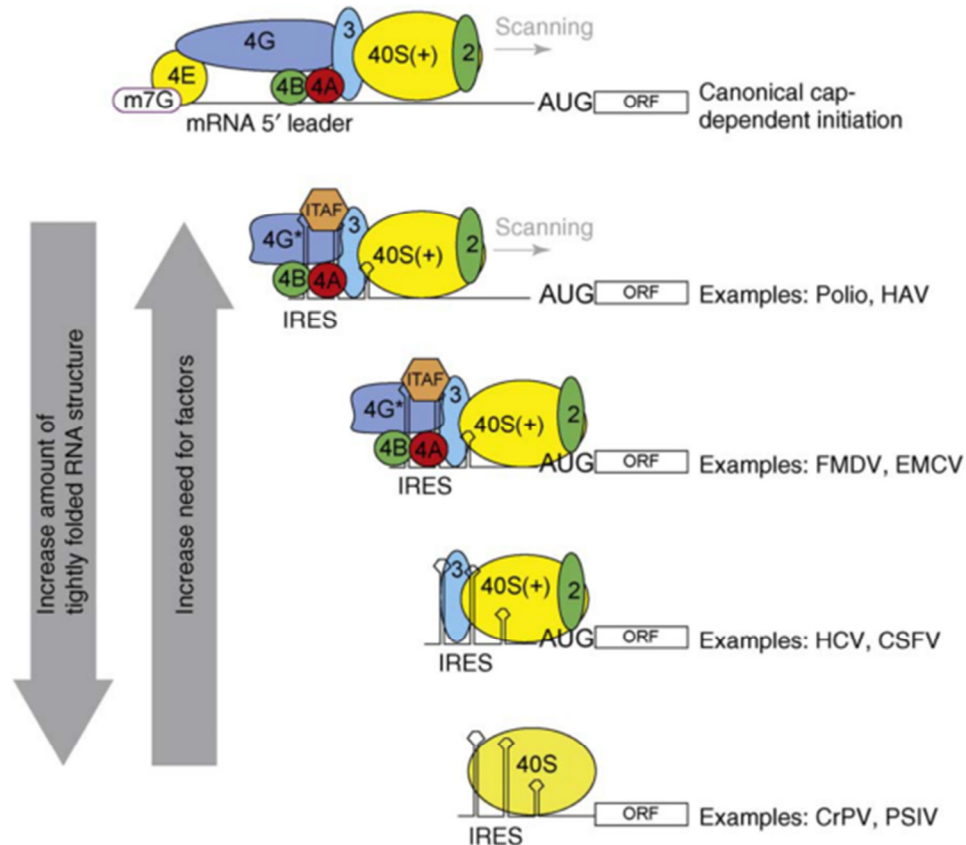
**Figure 4: Visualizing circular mRNA by atomic-force microscopy.**

Microscopy performed by A. Sachs shows complexes formed on capped, polyadenylated mRNAs in the presence of eIF4G, poly(A)-binding protein, and eIF4E. This figure is from Mendez & Richter[39].

in **Figure 9**) or by juxtaposing a very rare codon with a common codon in the +1 reading frame[47].

### **Restarting Elongation**

The bacterial use of transfer-messenger RNA (tmRNA) to free stalled ribosomes[48,49] and the TCV ribosome binding element[50] illustrate two ways in which molecular mimics of the translational players promote elongation. tmRNA performs this function by simply combining the separate elements of an initiator tRNA and short mRNA into a single molecule. This accommodates into stalled ribosomes and causes elongation over its short message and normal termination. Similarly, the TCV binding element mimics a tRNA in order to recruit the ribosomal large subunit as well as competitively bind the RdRp.



**Figure 5: The spectrum of IRES factor requirements.**

These examples show the spectrum of factors required for IRES translation. This study noted the trend that as viral RNA becomes more structured it requires fewer cellular factors for successful translation. This figure is from Filbin 2009[51].

### Modifying Termination

Termination is subverted in multiple ways, leading to suppression and allowing viruses to encode multiple peptides from a single mRNA. The influenza B stop-start signal, though still not fully understood[52]<sup>m</sup>, produces both the M<sub>1</sub> and BM2 proteins while managing to escape NSD. The termination suppression activity demonstrated by the murine leukemia retrovirus is better understood[53]. It uses a strong pseudoknot

<sup>m</sup> This paper has the unusual distinction of using a tri-cistronic reporter containing two fluorescence genes followed by firefly luciferase, each ORF separated by the start-stop element.

RNA structure to force ribosomes to pause, this allows its reverse transcriptase to bind polymerase[54]<sup>n</sup>.

### Examples in the Genomic Context

These few examples illustrate some of the methods employed by viruses to disturb normal translational equilibria. Examples employing these methods are also being found in the genomic context with increasing frequency. The most well documented method described to date is the cellular IRES; first described in the immunoglobulin heavy chain-binding protein mRNA[55], the cellular IRES has since been implicated in many contexts including the transcription factor *c-myc*, which is expressed when cap-dependent translation is otherwise compromised[56]<sup>o</sup>. Overlapping ORFs which lead to translation reinitiation are common in multi-cistronic bacterial messages[57]; but are not generally found in eukaryotes with one glaring exception: upstream open reading frames (uORFs). The most famous example comes from yeast, where the ribosomes of starving cells translate through the uORFs preceeding GCN4, thus translating this powerful activator of many biosynthetic pathways (Reviewed in [58]). While metazoans do not have *GCN4* to turn on our stress pathways, the uORFs before ATF4 demonstrate the same mechanism[59]. uORFs are important not only because they are implicated in an important non-canonical translation system, but also because uORF containing messages are strong substrates for nonsense mediated decay.

This work expands on the hypothesis that there is a linkage between -1 PRF and Nonsense Mediated Decay. Programmed ribosomal frameshifting (PRF) occurs when a *cis*-acting signal in an mRNA directs translating ribosomes into an alternate reading

---

<sup>n</sup> This paper provides a classic example of the mutagenesis strategy: "Mutate the 5' side of a stem, mutate the 3' side of a stem, mutate both."

<sup>o</sup> This is notable not only for the immunodepletion of eIF4GI as a stimulator of IRES driven translation, but also because it introduced the XIAP IRES, which goes on to test for an IRES element in the mammalian dual luciferase reporter system.

frame at a much higher rate than the normal 0.06%<sup>p</sup> or 0.005% frameshift events per message transit<sup>q</sup>. PRF is most commonly observed to shift ribosomes by one base either in the 3' (+1 PRF) or 5' direction (-1 PRF) (comparisons in [62]), but may also include much larger shifts or ribosomal shunting events. Ribosomal shunting is another means to translate sequences which would normally remain quiescent; instead of coding for an overlapping ORF, the shunting target lies far downstream and is often triggered when normal, cap dependent translation is impaired. Adenovirus provides the primary example of this mechanism: its shunting signal uses similarity to 18S rRNA to trigger translation over 220 nucleotides downstream[63].

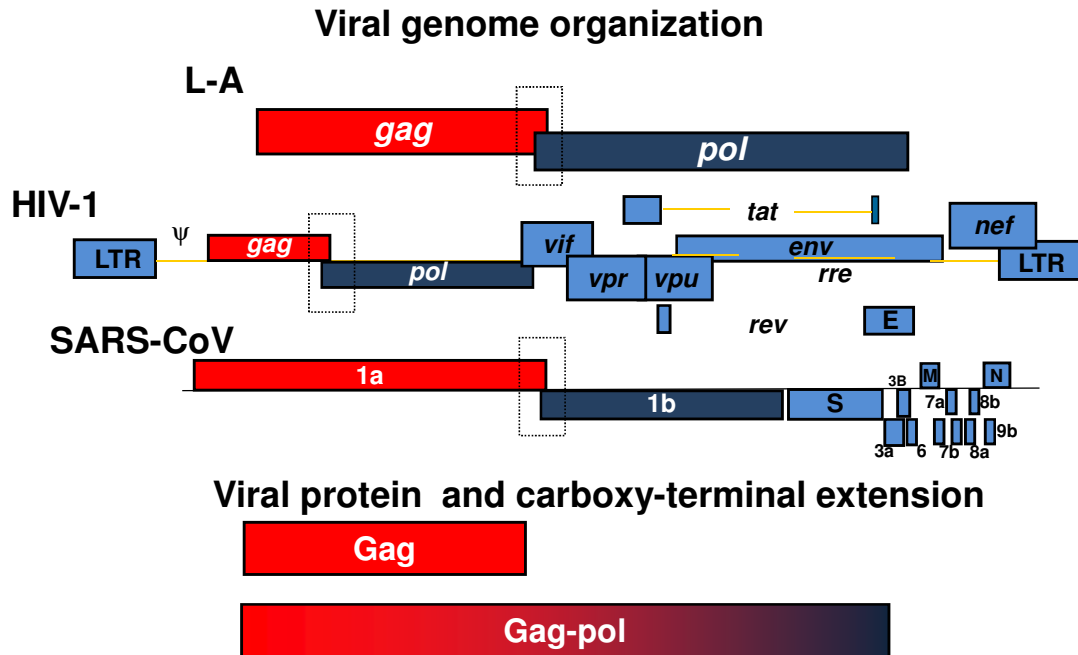
Programmed -1 ribosomal frameshifting was first observed in the *Rous-sarcoma* virus; in this viral context, a single mRNA transcript encodes two peptides and expresses them at a specific ratio via -1 PRF. The majority of the translated product consists of the shorter, unshifted gag structural protein. When -1 PRF occurs, translating ribosomes bypass the zero-frame stop codon, resulting in translation of the Gag-pol fusion product (**Figure 6**). Subsequently, it was found that some plus stranded RNA viruses, dsRNA viruses, and retroviruses also use -1 PRF. For example, the dsRNA L-A yeast Totivirus was shown to produce of its major coat protein and RNA-dependent RNA polymerase using this method[64]. It was further shown that maintenance of the M<sub>1</sub> satellite virus of L-A requires a specific frameshifting rate<sup>r</sup>, demonstrating a distinct phenotype dependent on ribosomal frameshifting. Multiple phylogenetically conserved cellular mechanisms have first been identified in viruses, this may also be true for -1 PRF.

---

<sup>p</sup> This value is presumed to be high because it was measured with  $\beta$ -galactosidase activity.

<sup>q</sup> Though we cite this value often, the logic of this calculation is problematic in the context of programmed -1 ribosomal frameshifting because it is based on observations made with tRNA suppressor mutants.

<sup>r</sup> This marks the first time the killer assay was used as a way to compare the rate of frameshifting.



**Figure 6: The organization and results of viral -1 PRF.**

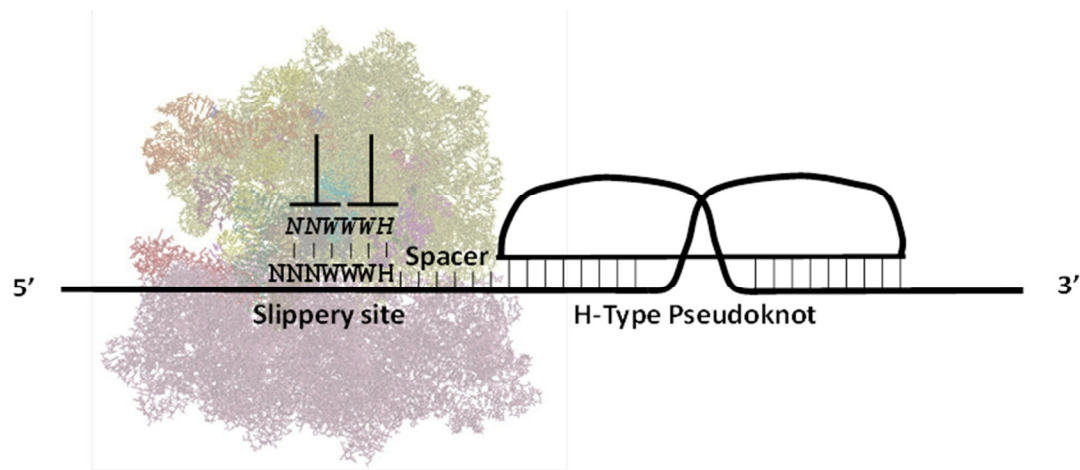
Diagrams of three -1 PRF containing viral genomes show the overlapping open reading frames and the resulting peptides which are synthesized if translation stays in the 0 reading frame (gag) or shifts into the -1 reading frame (Gag-pol fusion).

To date, a few -1 PRF signals of viral origin have serendipitously been found in metazoan genomes, including the mouse *Edr*[66] and the human paraneoplastic *Ma3* genes[67].

### Models of -1 Programmed Ribosomal Frameshifting

The role of -1 PRF outside the viral context was not clear, but the mechanisms have been thoroughly debated. A -1 PRF signal consists of several conserved elements: a “slippery heptamer”, followed by a significantly stable downstream mRNA secondary structure, separated by a spacer (**Figure 7**). The “slippery heptamer” consists of N NNW WWH (spaces delineate reading frame, N, W, H follow IUPAC conventions: N is any three identical bases, W is any three identical weak bases, and H is not guanine). All models of -1 PRF agree that the downstream mRNA structure

causes elongating ribosomes to pause while the tRNAs are positioned over the slippery site. The nature of the slippery site allows the aminoacyl and peptidyl-tRNAs to form a relatively stable mini-helix with the mRNA -1 frame bases[10]. The constituent bases of the slippery site are also important for the rate of -1 PRF, and further have different effects depending on the translational system expressing the sequence[68]. The mRNA secondary structure provides an energetic barrier to translating ribosomes and positions



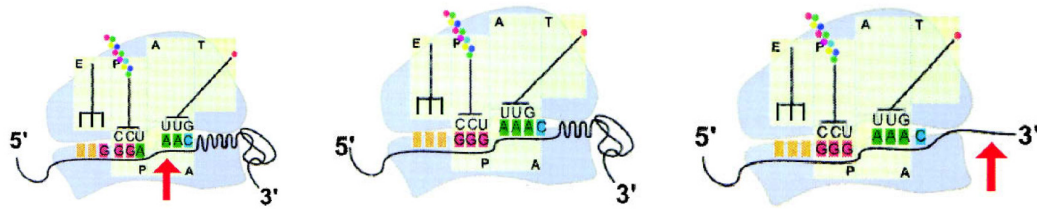
**Figure 7: The elements of a -1 PRF signal.**

A functional -1 PRF signal occurs when an actively translating ribosome is forced to pause at a strong mRNA secondary structure, usually an H-type pseudoknot. The ribosome A and P sites are situated over the NNW and WWH nucleotides of the slippery site due to an appropriately sized (6 bases in this drawing) spacer.

them over the slippery heptamer. H-type mRNA pseudoknots are the most common stimulatory structure, but other structures, including proteins bound to stem-loops[69], variously sized stem-loops[70], and RNA triplexes[71] promote efficient frameshifting.

The 'simultaneous slippage' model[72] of -1 PRF states that translating ribosomes pause on the downstream structure while the peptidyl-tRNA and aminoacyl-tRNA are situated over the 'NNW' and 'WWH' bases of the slippery heptamer (**Figure 8**). In the time it takes the ribosomal helicase to resolve the downstream mRNA, the A and P site codons break their 0-frame hydrogen bonds and subsequently re-form them in the -1 frame, pairing with 'NNN' and 'WWW.' Finally, elongation resumes normally in the -1

frame. This model provides neither details of the timing nor the position of the ribosome at the time of the frameshift event.



**Figure 8: A Model Mechanism of -1 PRF.**

Illustration of the ribosomal A and P sites during a -1 Programmed Ribosomal Frameshifting event. First, the ribosome (light blue) is forced to pause before the mRNA secondary structure (right knot) while its P site tRNA is situated in the classical conformation over a glycine codon and an newly accommodating asparagine tRNA is in the A/P conformation. The physical strain caused on the spacer by the ribosome attempting, but failing to translocate eventually causes the tRNAs to re-pair over the GGG and AAA codons. Finally the ribosomal helicase unwinds the downstream structure and elongation proceeds in the -1 reading frame. The figure is from Plant et al.[73]

Later models have stepped in to fill in these details, in each case the actual slippage event is coupled to GTP hydrolysis during elongation. The “integrated model”[62] hypothesized that the shift occurs after the ternary complex delivers the aa-tRNA to the A-site, but before peptidyl-transfer and therefore translocation by eEF2. This model was later refined in an attempt to explain the role of the downstream secondary structure[73]. In this model, the downstream element resists the movement of the mRNA during translocation, thus causing tension along the mRNA in the entrance tunnel and therefore partially blocking accommodation of the incoming aa-tRNA during eEF1A hydrolysis. This tension is released when the A and P site tRNAs break from the mRNA, allowing the mRNA to shift one base. In this model, the aa-tRNA serves as a lever and the LSU as a fulcrum to shift the mRNA before peptidyl transfer and translocation[74]. A separate model hypothesized that the -1 PRF event occurs during translocation. In this case eEF2 mediated GTP hydrolysis breaks the tRNA/mRNA helix, but the downstream mRNA secondary structure frustrates the movement of the



ribosome, leading to an incomplete translocation of two nucleotides[75]<sup>s</sup>. Interestingly, this model does not specify during which round of translation the incomplete translocation will occur. In one scenario, the NNW WWH nucleotides are in the E and P sites but are forced to slip by the downstream element, leading the incoming tRNA to be in the -1 frame in the A site of the ribosome before peptidyl transfer. The frameshift event may also occur during the following translocation, thus the mRNA shifts by only two nucleotides as the tRNAs transition from classical P/P and A/A to P/E and A/P states.

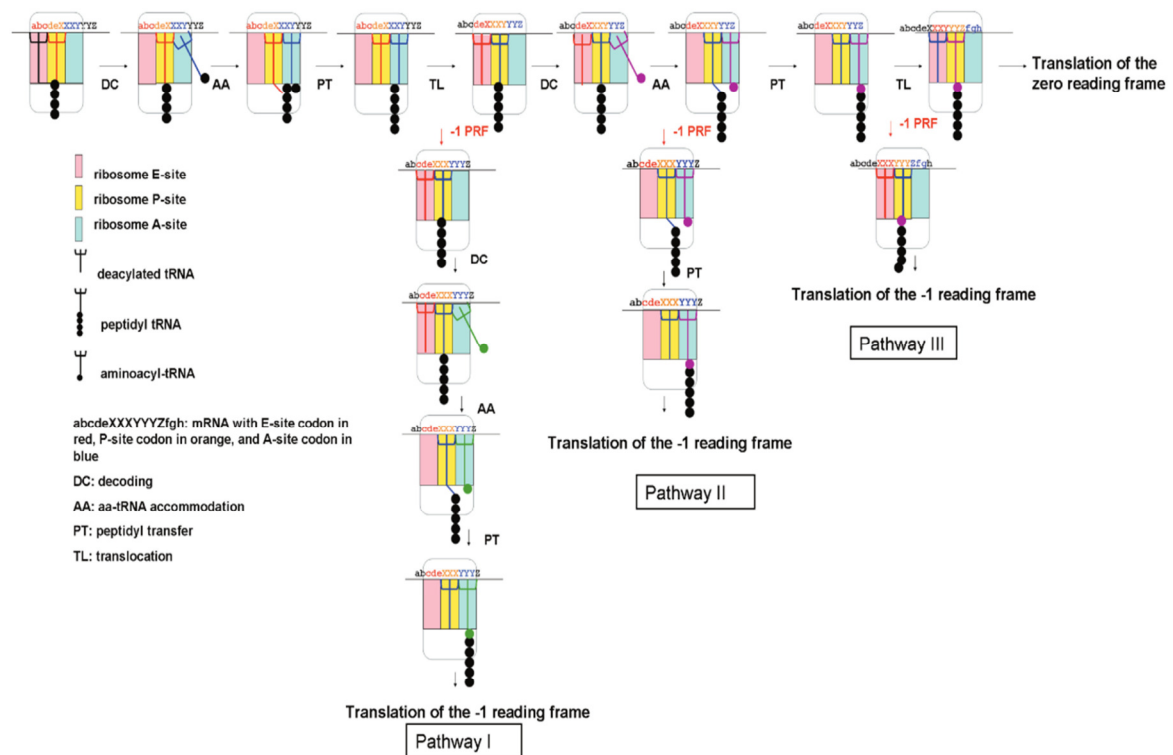
Each of these models has the support of experimental evidence, thus -1 PRF does not follow a single mechanism, but may best be explained as a series of kinetic partitioning events (for examples, see [76,77]) over the course of the elongation cycle such that each type of frameshift event results in a separate off-pathway product. The “kinetic model” of -1 PRF[7] unified these models and illustrated the relative contribution during each step of elongation (**Figure 9**).

### **Nonsense Mediated Decay**

The logical link which makes it possible to hypothesize that -1 PRF affects mRNA abundance comes from comparing the fate of translating ribosomes after translating a viral -1 PRF signal compared to a putative genomic signal. This distinction was suggested in **Figure 1**, but may be more explicitly shown by plotting the -1 frame extension in codons over the length of genomic mRNA (**Figure 10**). While viral -1 PRF signals significantly extend an open reading frame, genomic -1 PRF signals are predicted to truncate the reading frame. From this perspective, a functional -1 PRF signal in the genomic context acts as a stochastic premature termination codon[78],

---

<sup>s</sup> Mutagenesis was performed on the three bases before the slippery heptamer and caused losses of frameshifting by as much as 80%.



**Figure 9: Kinetic Partitions which lead to -1 PRF.**

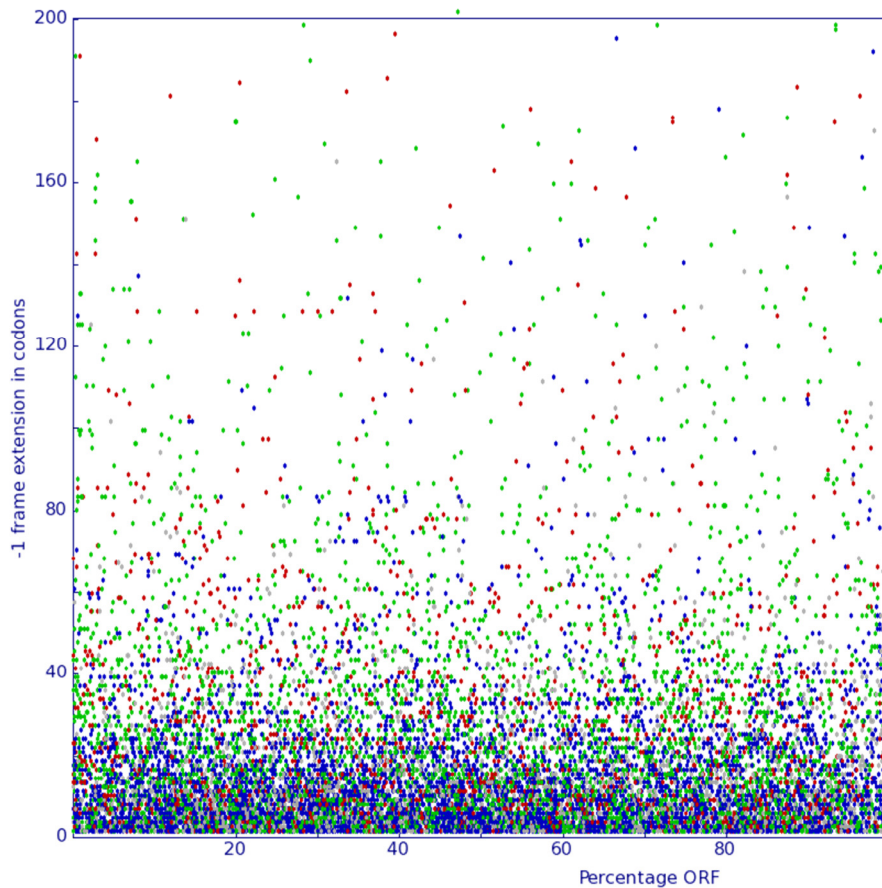
Two elongation cycles are diagrammed showing the on-product 0-frame translation and three separate pathways which result in the ribosome reading in the -1 frame. Thus the aminoacyl-tRNA mechanically wedges into the A-site during the first translocation event (Pathway I); tension of the mRNA caused by the mRNA secondary structure is relieved by slippage during the second accommodation event (Pathway II); or the second elongation cycle is frustrated by the mRNA secondary structure (Pathway III). This figure is from Liao et al[7].

and therefore as a substrate for nonsense mediated decay.

### NMD: Increasing Complexity

Diverse mechanisms exist to ensure the fidelity of gene expression; nonsense mediated decay (NMD) occurs post-transcriptionally and is among the most thoroughly debated. The name harkens to the observation in yeast that nonsense codons in the *ura3* ORF reduce the mRNA's half-life[79]<sup>t</sup> and steady-state levels by as much as 5 fold. Nonsense codons in mRNAs are not unique to eukarya, but bacteria couple transcription and translation and do not remodel messages as extensively as eukaryotes. Thus

<sup>t</sup> Demonstrated with an early example of a pulse chase using [<sup>3</sup>H]RNA.



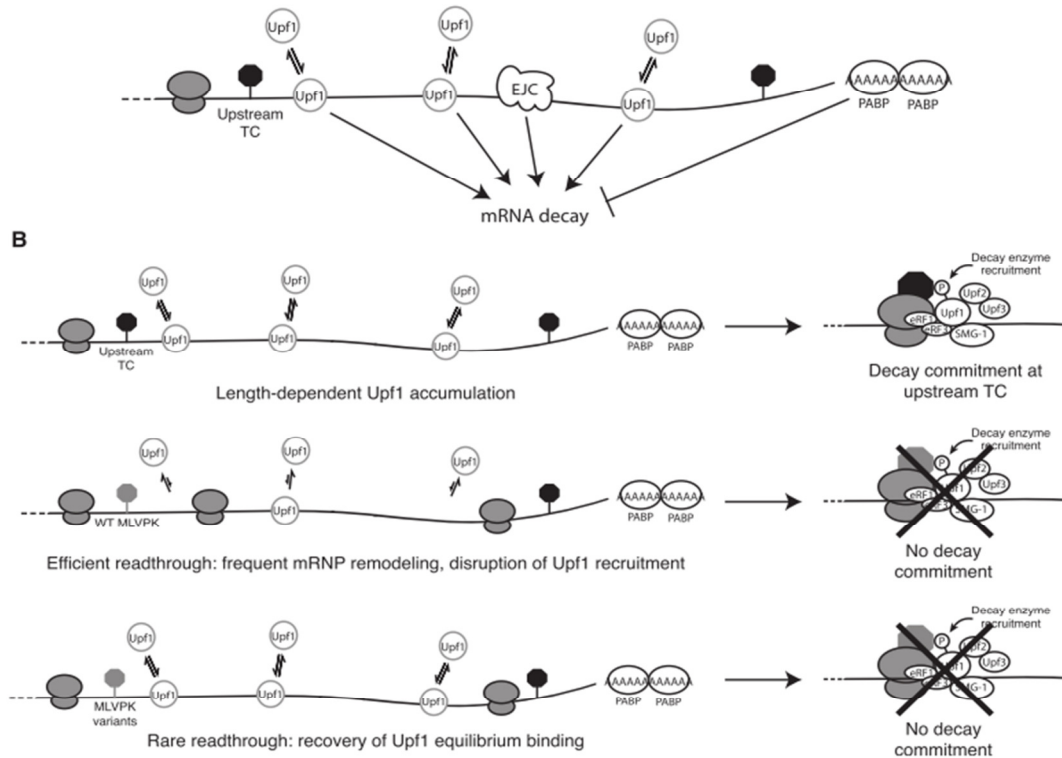
**Figure 10: Genomic -1 PRF signals do not significantly extend the ORF.**

The distribution of length of each -1 frame extension following a slippery heptamer in *Homo sapiens* is plotted with respect to relative ORF position. Each dot is colored according to the identity of the first 3 bases of the slip site (red: AAA, green: UUU, blue: GGG, black: CCC). Approximately 0.07% extend more than 30 codons past the 0 frame stop codon.

bacterial NMD is currently thought to begin when internal cleavage sites are exposed by prematurely terminating ribosomes; these unprotected sites are endonucleolytically cleaved[80] by RNase E in *E. coli*, leaving the mRNA a target for rapid decay.

In eukarya, there are many more proteins involved (**Table 3**) in NMD, and still more models arguing the details; but it is encouraging to note that the general idea that

successfully translating ribosomes protect messenger ribonucleoproteins (mRNPs) from degradation<sup>[81]<sup>u</sup></sup> is shared with the bacterial model (**Figure 11**).



**Figure 11: Translating ribosomes protect mRNPs from NMD.**

This model summarizes the elements which promote and inhibit nonsense mediated decay. Thus if Upf1 is able to mark mRNAs as having long 3' UTRs, or the exon junction complex is not remodeled efficiently, then the mRNP is a strong candidate for decay. Conversely, the poly-A binding protein complex and ribosomal readthrough protect the mRNP from decay. This figure is from Hogg and Goff [81].

The primary sources of disagreement among the eukaryotic models of NMD reside with the importance and function of exon junction complex in NMD; the role of nuclear export in NMD; and the cap binding complex. Therefore, if NMD is to be the

<sup>u</sup> Assayed by inserting Murine Leukemia Virus Pseudoknot mutants into the 3' UTR of a constitutively expressed  $\beta$ -globin construct. Mutants which allow efficient readthrough accumulate more mRNA.

foundation of a model of post-transcriptional regulation, it is imperative to first understand NMD in the context of *Saccharomyces cerevisiae*.

#### **NMD: Current Model in *Saccharomyces cerevisiae***

The complete yeast genome[82] marked a shift in the understanding of the complexity of eukaryotic organisms. As of April 2011<sup>v</sup>, 6,607 open reading frames (ORFs) have been identified across the 16 chromosomes, mitochondria, and the endogenous 2-micron plasmid. The initial report of intron containing genes in yeast identified 228 introns[83], a number which has since grown to 282 (4.3% of the genome). Amazingly though, nearly 25% of the *Saccharomyces cerevisiae* transcriptome is generated from messages containing introns, including 9,168 mRNA molecules per hour per cell (73% of the ribosomal protein mRNAs)[84]. Furthermore, the intron containing ribosomal protein mRNAs comprise 90% of all intron containing transcripts in yeast. In contrast, when microarray experiments were performed using cells deficient in the NMD machinery, 746 transcripts out of 7,839 (9.5%) were upregulated, including 545 out of 6,086 protein coding ORFs (9.0%)[85] assayed. When these transcripts were categorized, it was noted that genes involved in protein synthesis were underrepresented in the population of NMD regulated transcripts (1.7%). From these observations (summarized in **Table 1**), it is unlikely that nonsense mediated decay in *Saccharomyces cerevisiae* is mediated primarily by the exon junction complex. When He et al. examined the chromosomal positions of each transcript, 36% of all up-regulated in NMD deficient cells were positioned within 20 kb of telomere ends; while only 8% of the transcripts farther away were similarly up-regulated. This peculiar spatial arrangement of NMD regulated transcripts led to the observation that strains deficient in

---

<sup>v</sup> The 6,607 ORFs currently assayed at: <http://www.yeastgenome.org/cache/genomeSnapshot.html> were not all available to Ares et al. nor He et al.; therefore these are three separate but comparable datasets.

NMD express genes normally silenced in telomeric regions[86], at least when tested using a telomere localized *URA3* and 5-FOA reporter system.

<b>Class in transcriptome [84]</b>	<b>Number of Genes</b>	<b>Percentage Genes</b>	<b>mRNA/h/cell</b>	<b>Percentage mRNA</b>
Ribosomal proteins with introns	95	1.54%	9,168	24.24%
Other proteins with introns	134	<b>2.17%</b>	1,009	2.67%
Ribosomal proteins without introns	37	0.60%	3,414	9.03%
Other proteins without introns	5,922	95.70%	24,226	64.06%
Total	6,188		37,817	

**Table 1: Classifying intron containing genes in *Saccharomyces cerevisiae***

Ares et al. catalogued the number of mRNAs transcribed by *Saccharomyces cerevisiae* per hour and observed that a majority of the intron containing messages are from ribosomal proteins. Thus 1.5% of the genome is comprised of ribosomal proteins while 24% of the transcriptome is ribosomal protein RNA.

<b>Class in NMDΔ cells [85]</b>	<b>Increased</b>	<b>Percentage of genome Increased</b>	<b>Decreased</b>
Annotated Genomic ORFs	545	8.81%	6
Mitochondrial, plasmid, unannotated ORFS	150	2.42%	5
Small RNA, tRNA, rRNA	51	0.82%	8
Total	746	12.06%	19

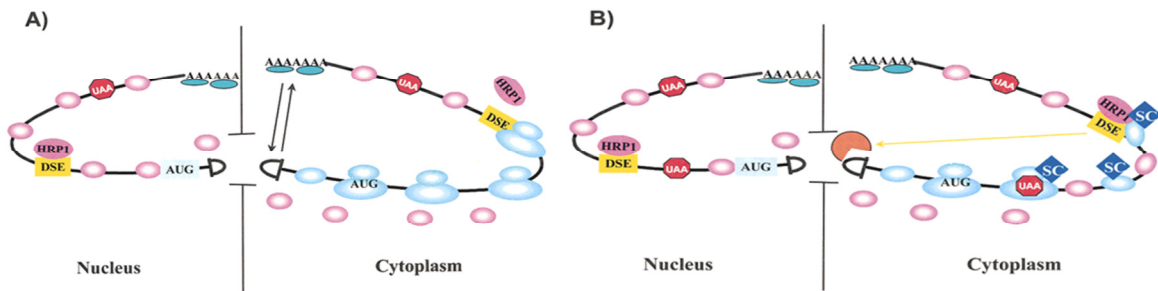
**Table 2: Observed RNA abundances for different classes of ORFs**

He et al. catalogued the ORFs which display increased expression in NMD deficient cells by micro-array. If all NMD is due to splicing, then the number of genes expected to be increased in an NMD deficient strain should be much lower than the observed 8.81% of the entire genome, but closer to 2.17% (bold in Table 1).

This marks one of the first instances in which NMD is cast in a role not only as a protector against deleterious transcripts, but also as a regulator of cellular homeostasis. These roles are in turn defined by the underlying mechanism of NMD. The remaining models of NMD attempt to describe how yeast distinguish normal termination from premature: the “surveillance complex”[87], and “faux-3’ UTR”[88]<sup>w</sup> models. Though the latter is now accepted as predominant mechanism in yeast, the surveillance complex model remains important because it is the direct antecedent of the currently prevalent model of NMD in metazoans, the pioneer round hypothesis[89,90].

<sup>w</sup> This was first demonstrated with some excellent PABP tethering experiments.

The surveillance complex model was posited when it was observed that Hrp1p interacts with Upf1p as well as a substrate for NMD[91]<sup>x</sup>. A poorly defined downstream sequence element (DSE) located 3' of a PTC leads to rapid degradation via NMD; when the DSE was removed these transcripts were no longer degraded. Furthermore, the *hrp1-3* allele increased the half-life of NMD substrates by factors of 5-10. Therefore Hrp1p was identified as a recruiter of the surveillance complex<sup>y</sup>; it binds weakly to the DSE upstream of the stop codon and is easily removed by actively translating ribosomes. When ribosomes terminate before the DSE, then Hrp1p is free to bind the mRNA and recruit Upf1p, leading to rapid degradation (**Figure 12**).



**Figure 12: The surveillance complex model in yeast.**

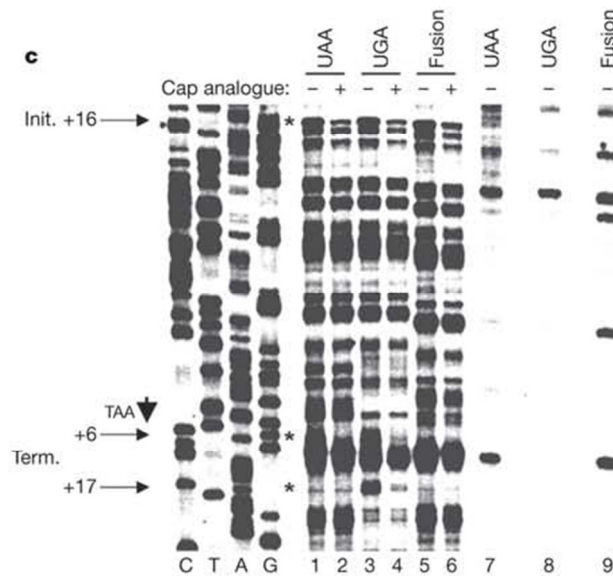
**A.** mRNA binding proteins (including Hrp1p) interact freely with the message inside the nucleus but are removed by translocating ribosomes in the cytoplasm. **B.** If ribosomes terminate prematurely; Hrp1p is free to recruit the surveillance complex, leading to rapid decapping, deadenylation, and decay. This figure is from Czaplinski et al.[8].

The primary strength of this model lies in its implicit coupling of active translation to the degradation of NMD substrate mRNAs. Translating ribosomes prevent the binding of Hrp1p to the DSE and subsequent degradation. Thus the weakness of the model, the poorly defined DSE (TGYYGATGYYYYY), is used to demonstrate that some messages house many separate DSEs[92]; but this model cannot explain rapidly degraded NMD substrates which do not contain this motif. The faux 3' UTR model fills in this logical gap.

<sup>x</sup> This paper serves as a cautionary tale against overinterpreting EMSA data.

<sup>y</sup> Hrp1p is generally identified as part of cleavage factor I. In this role it is responsible for cleaving and polyadenylation of pre-mRNAs.

This model does not focus on active elongation as a protector of mRNPs, but notes the position of terminating ribosomes in the context of a normal stop codon versus premature termination. Toeprint analyses demonstrated that terminating ribosomes do not protect the mRNA 17 bases downstream of a normal stop codon; but do protect the same position when a stop codon is recognized as aberrant (**Figure 13**). Similarly, the +6 position was deprotected in a time dependent manner in the case of premature termination; but adding a m<sup>7</sup>G cap analog attenuated these effects, as did performing the experiment in Upf1p deficient cells.



**Figure 13: Toeprints of normal and premature termination.**

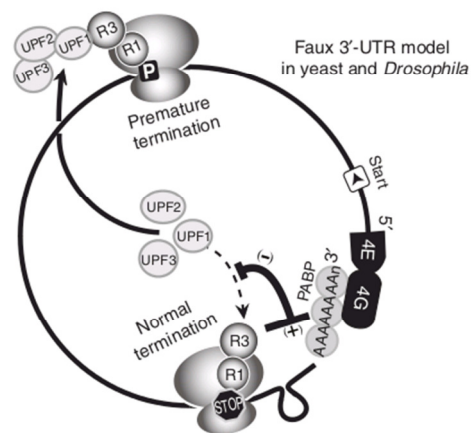
Significant differences were observed when cycloheximide (CHX) was used to stop translation in the context of normal versus premature termination. In this case, the +17 position is protected from transcription elongation by the presence of an abnormally terminating ribosome (lane 7), but not when m<sup>7</sup>G cap analog is supplied (lane 8). This figure is from Amrani et al. [88].

These observations lead to the conclusion that the DSE was not in fact a salient feature of the mRNA which explicitly recruits a surveillance complex, but another sign that the context of termination and the mRNP is aberrant. Under this model, NMD is a function of the interplay between: the m<sup>7</sup>G cap, the initiation factors eIF4G and eIF4E,



poly-A binding protein (PABP), and the poly-A tail of the mRNP; these in concert with the continued successful transition of ribosomes from termination to reinitiation serve to ward off the Upf proteins and rapid degradation. These elements are shared with the components involved in the 3' UTR length surveillance model of NMD[81].

A termination event must be recognized as aberrant in order for the faux 3' UTR model to function (**Figure 14**). Normal termination occurs when the elongating ribosome encounters a UAA/UAG/UGA codon in the appropriate context. Class I Release Factors (RF) (eRF1 is Sup45p in yeast) recognize these stop codons in the A site and stimulate release of the polypeptide chain from the P site tRNA. Class II RFs (eRF3 is Sup35p in yeast) use GTP to stimulate the class I RF; in addition the N-terminal portion of Sup35p interacts with PABP, stimulating mRNA decay[93]. One assumption of the faux 3' UTR model is that premature termination affects this interaction. From the opposite perspective, the faux UTR model suggests that the proteins bound to the 3' UTR are important for the stability of the message, specifically the proximity of PABP to the site of termination. To test this implication of the faux UTR hypothesis, Amrani et al. used a construct containing the MS2 coat protein binding site with a PTC used to bind a portion of Pabp1[88]. This 'tethering' of Pabp1 to the otherwise strong substrate for NMD stabilized the message significantly. Taking one step further, eRF3 was tethered in a similar fashion and also stabilized NMD substrates (PGK1 with a PTC); but when eRF1 was used no effect was observed. These observations lead to the conclusion that termination is normally a process of eRF3 binding near Pabp1, activation by eRF1, and efficient ribosomal decoupling. In contrast, premature termination is inefficient and, without Pabp1, leaves time for the Upf proteins to bind and promote decay. The details of this process are not fully understood, but are suggested by the structures of the Upf proteins, their interactions with the decapping proteins Dcp1p and Dcp2p, the 5' exonucleases Xrn1p and Rat1p, deadenylase, and the exosome.



**Figure 14: The Faux 3' UTR Model of NMD.**

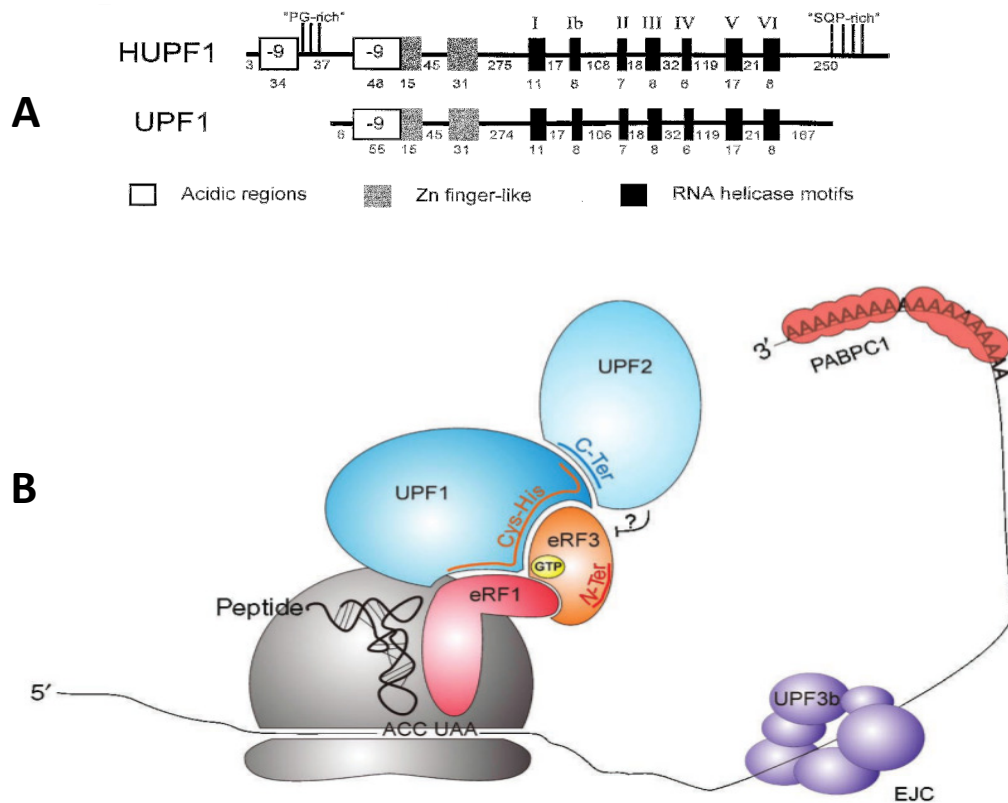
The context of termination is the primary determinant of suitability for NMD in the faux 3' UTR model. This figure is from Brogna et al.[94]

### The Upf proteins

The primary actors of NMD are Upf1p<sup>z</sup>, Upf2p, and Upf3p; they are required for function and partially conserved among eukaryotic species. The complete panoply of NMD associated factors is difficult to define (**Table 1**). Upf1p is the judge of aberrant termination; it subsequently interacts with Upf2p and Upf3p to trigger mRNA degradation. This judicial activity is partially demonstrated by Staufen-Mediated Decay (SMD). In this mechanism staufen recruits Upf1p downstream of a termination codon and leads to rapid mRNA degradation[95]. These activities are mediated by two conserved domains in Upf1p (**Figure 15**); the Upf2p interacting N-terminal domain and a C-terminal helicase region which interacts directly with mRNA. In addition, this region contains 7 superfamily 1 RNA/DNA helicase domains[96].

Crystal structures of the mammalian UPF1 (RENT1) helicase domain bound to ssRNA and AMPPNP[97] provide hints regarding how UPF1 may unwind mRNA

<sup>z</sup> Every effort has been made to follow the yeast[278] and human[279] nomenclature guidelines where appropriate.



**Figure 15: The domains and interactions of UPF1.**

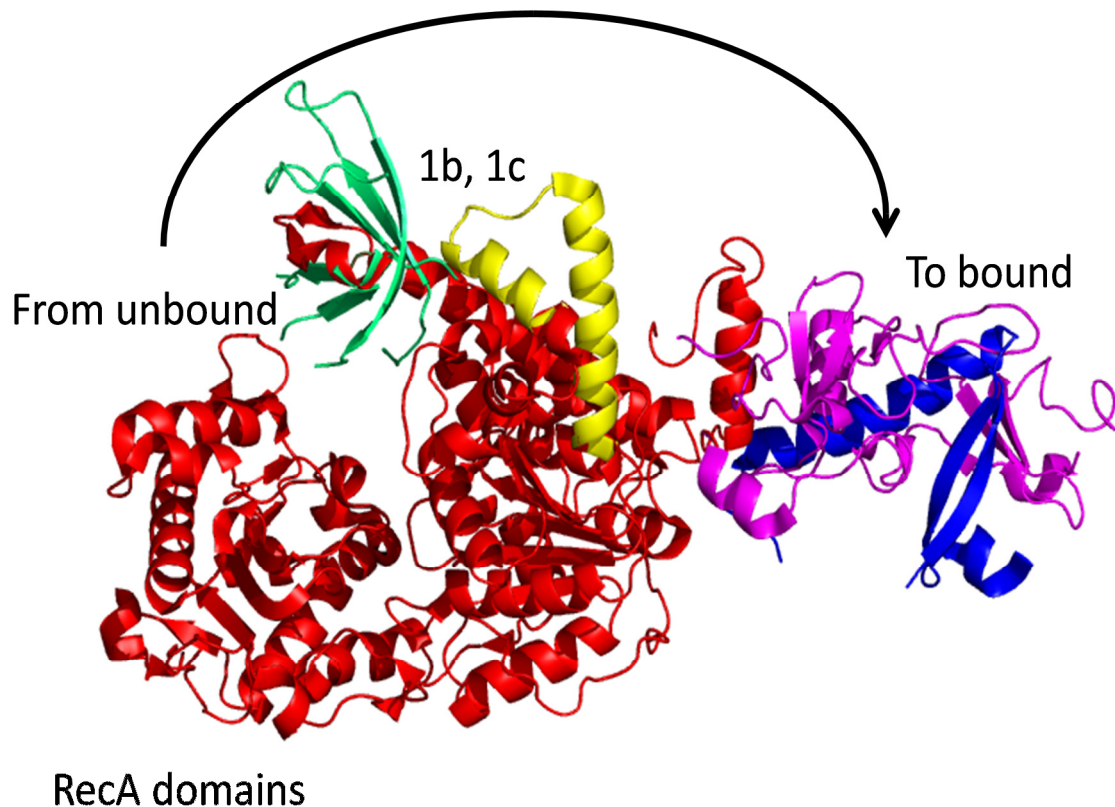
**A.** The order and relative positions of the human and yeast Upf1p domains. This image is from Applequist et al.[96] **B.** Depiction of Upf1p interactions with release factors and Upf2p. This image is from Ivanov et al.[98]

powered by ATP. Cryo-EM structures (along with SAXS and crystal structures) of UPF1 with UPF2[99] provide an image of the C-terminus of UPF2 bound to the UPF1 CH-domain. In mammalian systems this feeds a model of NMD whereby UPF1 and UPF3 transiently join eRF1 and eRF3 (the SURF complex). The EJC in turn meets the SURF complex mediated by UPF2. Once all the components are in place, SMG1 phosphorylates the C-terminal SQ motifs of UPF1. This in turn leads to rapid degradation, at least partially through SMG6's endonuclease activity and partially by SMG7's ability to promote destabilization via DCP2 and XRN1. In addition, phosphorylated UPF1 attracts DCP1, leading to a further stimulation of rapid decapping.

Thus UPF1 stands at the center of a strongly redundant network of interactions which serve to rapidly detect and remove NMD substrates.

UPF2 is usually considered an adaptor protein, bringing either UPF3 or the SURF complex to UPF1. Recent work[100] has illustrated how UPF2 affects the ATPase and mRNA unwinding abilities of UPF1. This model condenses the 7 helicase regions of UPF1 into 2 RecA domains and the N-terminal Zn finger into a CH domain. Crystal structures with various ATP analogues and UPF2 showed that the CH domain inhibits the ATPase activity of UPF1 and binds UPF2. The two RecA domains of UPF1 in these structures each have an additional domain (termed 1C and 1B which are an  $\alpha$ -helix and  $\beta$ -barrel respectively) that affects nucleic acid binding to the RecA domains. When UPF2 was bound to UPF1 in these crystal structures, the CH domain moved to a position almost diametrically opposed to the unbound state (**Figure 16**). This surprising shift of a relatively large 1500 Å<sup>2</sup> surface area was initially thought to be an artifact, but eventually shown to be facilitated by a flexible linker. When the CH domain is bound by UPF2, they drive a shift of the CH domain to the other side of 1B, allowing it to relax and increasing RNA unwinding activity. Therefore, UPF2 is not only an adaptor protein as previously thought, but has a strong role in activating UPF1.

UPF3 is both structurally simpler and evolutionarily more complex than the other two primary members of the NMD system. In yeast it contains 387 amino acids, which are shuttled between the nucleus and cytoplasm by importin- $\alpha$  in order to maintain fully active NMD[101]. However a duplication event during the evolution of higher organisms resulted in two paralogous isoforms, UPF3A and UPF3B. The interactions between UPF3 and the other NMD factors are further complicated because UPF3A and UPF3B compete for binding to UPF2; in addition UPF3A is downregulated by UPF3B in



**Figure 16: Long distance movement of UPF1 CH domain upon UPF2 binding**

Crystal structures of UPF1 (red with lime 1b domain, yellow 1c domain, and purple CH domain) complexed with UPF2 fragments (blue) suggest that the CH domain travels over 1,500 Å<sup>2</sup>. In addition, the RecA domains (bottom left) switch from a contracted to relaxed conformation while the 1B and 1C domains remain relatively stable. This image was generated from PyMol using PDB accessions 2WJV and 2GJK, descriptions from Chakrabarti et al.[100], and the pdb parser from Leshin et al.[102].

a transcription independent manner[103]. The full purpose of UPF3A/B in mammalian NMD is not known, but it has been shown to shuttle between the nucleus and cytoplasm along with SMG5 and SMG7. The SMG (suppressor with morphogenetic effects on genitalia) proteins were identified in *Caenorhabditis elegans* because mutants suppressed specific nonsense containing alleles, some rearrangements, and aberrant 3' UTR containing genes[104]. Later work showed that these are the same players as the yeast Upf proteins with the addition of a phosphorylation control system via SMG1 (kinase activation of UPF1), SMG5-7 (phosphatase inactivation of UPF1).

Gene Name	Role
<b>Primary actors (human)</b>	
RENT1/UPF1	Target of SMG1 phosphorylation. Disrupts translation termination.
UPF2	Interacts with UPF1 and UPF3. Promotes UPF1 phosphorylation.
UPF3X (UPF3B)	Initial interaction with NMD target, attracts UPF2.
UPF3 (UPF3A)	Early interaction with NMD target, attracts UPF2. Immunoprecipitates with Y14, RNPS1, eIF4AIII.
SMG1	Kinase of UPF1.
SMG5	Dephosphorylates UPF1, requires SMG7 to function.
SMG6	Dephosphorylates UPF1, interacts with Y14.
SMG7	Dephosphorylates UPF1, Recruits SMG5 to decay foci for 5' to 3' decay via DCP2 and XRN1.
SMGL1 / hNAG	Unknown, but required for NMD.
SMGL2 / hDHX34	Unknown, but required for NMD.
<b>Epistatic factors (human)</b>	
CBP (CBC80/CBC20)	Binds mRNA m <sup>7</sup> G cap upon nuclear export.
eIF4E / eIF4G	Tethers PABP to m <sup>7</sup> G cap after nuclear export.
PABP	Poly-A binding protein, links m <sup>7</sup> G cap to poly-A tail after nuclear export.
eRF1 / eRF3	Peptide release factors, competing with Upf proteins for ribosomes.
Y14, Magoh	Core of the exon junction complex.
DCP1 / DCP2	The decapping complex.
XRN1 / RAT1	5' to 3' exonuclease, cytoplasmic and nuclear respectively.
PNRC2	Links UPF1 to DCP1.
DHH1 and PAT1	Remove mRNAs from active translation and increase decapping rate.
The Exosome	3' to 5' exonuclease.
<b>Exosome proteins (yeast/archaea)</b>	
Rrp41, Rrp46, Mtr3	Create a portion of the hexameric ring, each has an RNase PH domain.
Rrp42, Rrp43, Rrp45	Also serve to create hexameric ring, their PH domains are inactive.
Rrp4, Rrp40	Create the S1 'pore' structure to guide incoming RNA.
Csl4	Also has a S1 domain.
Rrp44	RNase R domain containing component (RNase II).
Rrp6	Similar to RNase T/D, only found in nuclear exosome.
Rrp47	Putative RNA binding protein, only found in nuclear exosome.
Ski7	GTPase, only found in cytoplasm.

**Table 3: A partial listing of proteins involved in *Homo sapiens* NMD**

The full catalogue of proteins involved in NMD includes these, the components of the exon junction complex, the ribosome, initiation complex, the complexes involved in elongation, RNAi, and factors involved in remodeling the mRNA. With that in mind, these are some of the most immediately involved. Exosome components are transcribed from Houseley et al.[105].

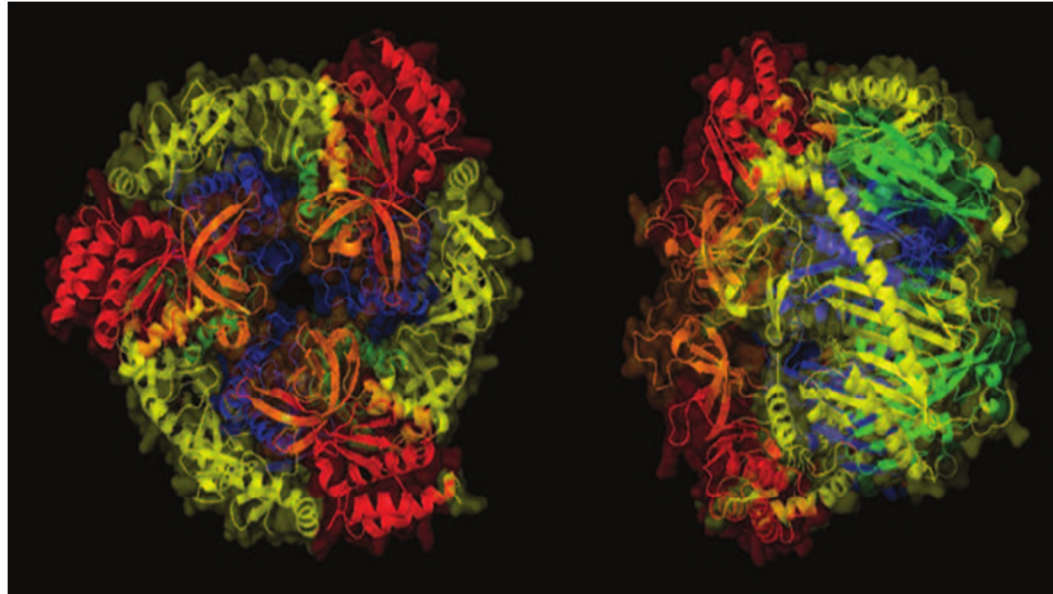
### Other Factors

Y14 and Magoh are cytoplasmic shuttling proteins which bind mRNAs in a position specific manner with respect to the exon junction complex[106]. Together they make an extremely stable heterodimer[107] which is hypothesized to form a clamp

around the mRNA and binding platform for the formation of the rest of the EJC. If this clamp is not removed from an actively translated mRNA, then it becomes a substrate for NMD.

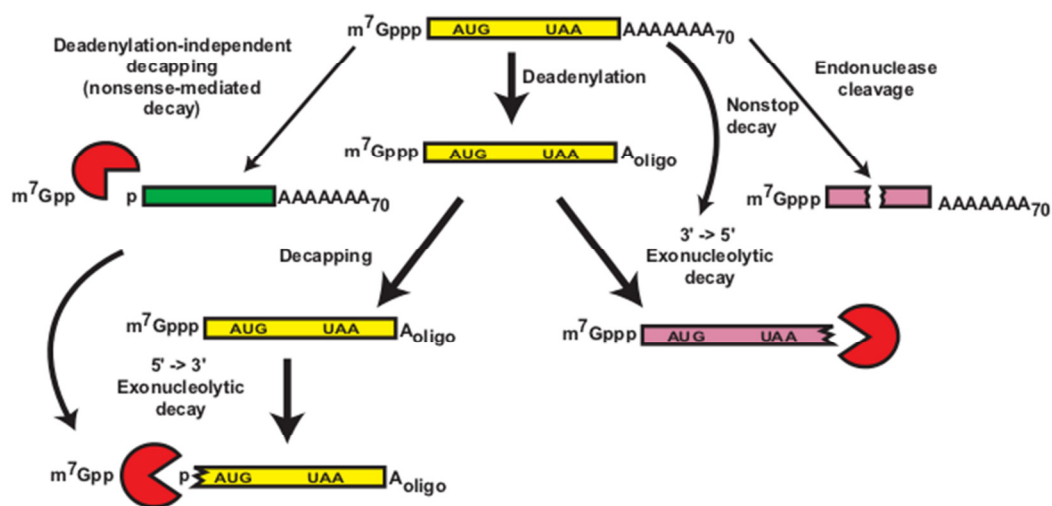
### **Important Nucleases**

The factors which encourage NMD are still being identified and examined, but the proteins which actually perform the degradation of the mRNA are well established. The degradation of a translationally competent mRNA occurs in two stages: the stabilizing 7-methyl-guanosine cap and poly-A tail are removed via the decapping complex and deadenylase respectively. First (in yeast), the poly-A tail is shortened via a complex of Ccr4p and Pop2p. Once this is complete, decapping occurs via Dcp1p and Dcp2p. This process is in turn affected by a series of *trans*-acting factors including: Dhh1p, Pat1p/Mrt1p, the Lsm complex, Vsp16p, Edc1p, and Edc2p.[108] Interestingly however, Dhh1p is not required for NMD in yeast while it does interact with the yeast deadenylase; thus functionally linking deadenylation and decapping. A denuded mRNA is an excellent substrate for 5' to 3' decay via Xrn1p and 3' to 5' decay via the ski complex and exosome (**Figure 17**). These same complexes are required for the decay of mRNAs which have been targeted for RNAi[109] and serve to degrade mRNAs in the same fashion as occur during NGD (**Figure 18**). In these contexts, decapping and deadenylation do not occur, but the mRNA is endonucleolytically cleaved; leaving the 5' fragment available for degradation from the 3' end via the exosome and the 3' fragment available from the 5' end via Xrn1p.



**Figure 17: Crystal Structure of the Archaeal Exosome**

The exosome from *Arcaheoglobus fulgidus*[110]. RNA is guided to the centrally located, catalytically active PH domains by the (orange) S1 and (red) KH domains. The archaeal exosome shows the RNase domains on alternating (blue) Rrp41 and (green) Rrp42 while the RNA binding domains are located on the Rrp4 subunits. Yellow regions are identical to the PNPase of the bacterium *Streptomyces antibioticus*[111]. This figure is from Houseley et al.[105].



**Figure 18: Methods of mRNA decay.**

The various interactions of endo and exo-nucleolytic decay which lead to mRNA degradation. The figure is from Parker and Song.[112]

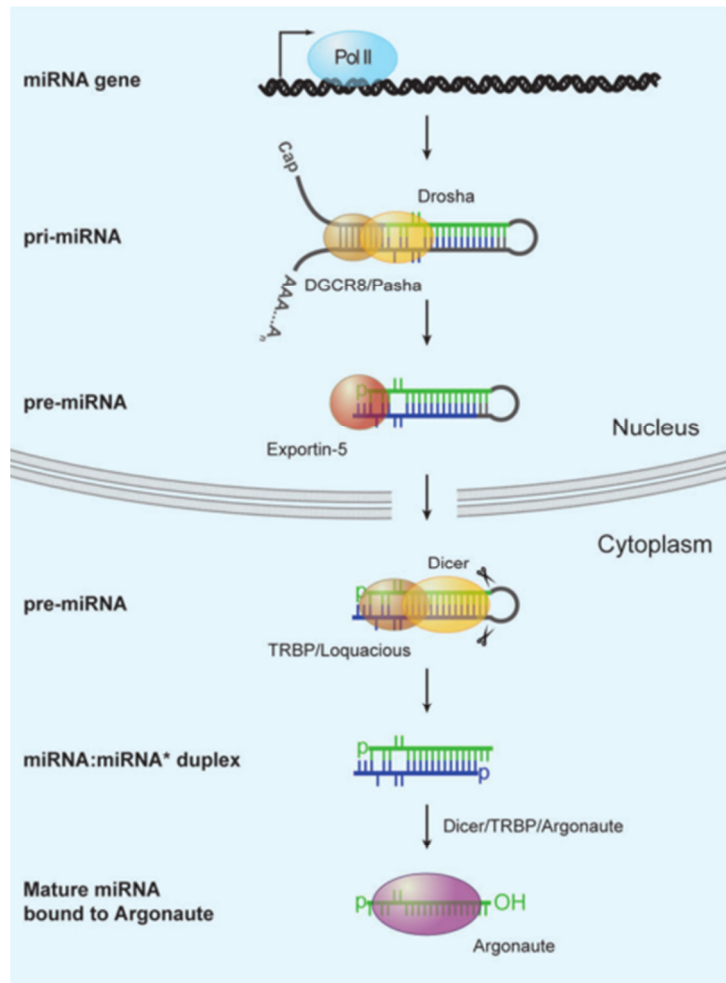


If -1 PRF has a regulatory role in shifting the equilibrium of this enormously complex degradation machinery, then it too must be regulated. Ideally, a regulatory framework for -1 PRF should be sequence specific and able to be rapidly engaged by the cell. The burgeoning field of RNAi provides excellent candidates to fulfill these criteria.

### **Small RNAs and Genome Regulation**

Small non-coding RNAs are assuming an increasingly important role in post-transcriptional regulation. These classes of molecules are predicted to control approximately 1/3 of all protein coding genes in mammals and plants and participate in the regulation of every cellular process thus far examined[113,114]. Most characterized miRNAs control gene expression by modulating translation and mRNA stability in the cytoplasm. In *Arabidopsis*, miRNAs have been shown to directly affect methylation on the chromosomes[115], setting the precedent for other uses of miRNA. Since then, hsa-miR-122 has been shown to directly interact with the Hepatitis C virus (HCV) in liver cells[116], while networks of interactions between cellular miRNAs (specifically hsa-miR-29a) with HIV-1[117] and viral miRNAs with cellular messages[118,119] have been identified in human cells.

Given the multifarious interactions between small RNAs and both endogenous and exogenous mRNAs, it is not a difficult transition to hypothesize that similar interactions may occur between miRNAs and the regions of mRNA which include -1 PRF signals. In order to make these hypotheses, one first must have a general understanding of the lifecycle of miRNAs (**Figure 19**).



**Figure 19: miRNA Biogenesis**

miRNA biogenesis often begins with Pol II transcription of large intron containing sequence, followed by multiple rounds of processing. Drosha and Pasha cleave the stem-loop intronic sequence away, followed by Dicer to remove the loop sequence, leaving a miRNA duplex, and finally an active argonaute bound miRNA. This figure is from Bushati and Cohen.[114]

Most miRNAs are transcribed by RNA polymerase II and are comprised of a stem-loop which contains one or more primary miRNA (pri-miRNA) sequences. These stem-loop sequences range from 100-10,000+ nucleotides and are found primarily (in mammals at least) in intron containing sequences[120]. Processing of these stem-loops occurs in the nucleus by the 'Microprocessor,' which is a multi-protein complex which contains Drosha (an RNase III) and Pasha (which provides dsRNA binding capability). These cleave the end of the stem-loop, leaving a pre-miRNA which has 2-3 bases of

overhang on one side and the RNA loop on the other. This is exported into the cytoplasm where it is bound by the dsRNA binding protein, TRBP, and again cut by Dicer, another RNase III. The remaining duplex contains a miRNA bound to miRNA\* (also 5p vs 3p). Finally, this is split apart and loaded on the RNA-induced silencing complex (RISC). Usually the miRNA strand has less base pairing on the 5' side; thus it is possible to preferentially load the miRNA strand and discard the miRNA\*.

Once the RISC complex is loaded with miRNA, it scans for RNA targets by base-pairing interactions. If (near) perfect hybridization occurs, the mRNA target is cleaved by Argonaute (Ago2p) and rapidly degraded. However, imperfect complementarity still results in translation repression. In animals this relationship is particularly skewed, such that to date only one mRNA has been reported to be directly cleaved via miRNA[121]. Instead, most animal miRNAs pair imperfectly and lead to translational repression and eventual degradation by deadenylation and decapping[122].

The mechanisms through which animal miRNAs repress target mRNAs are not completely understood. Work performed using the dual luciferase reporter system containing multiple copies of the CXCR4 miRNA binding site in multiple contexts showed that miRNA mediated repression requires capped, polyadenylated messages. Thus, when miRNA was applied to cells the CrPV IRES upstream of a luciferase reporter, it had no effect; but when either 5' cap or 3' tail were added back the miRNA had moderate effect on expression[123]. Completely conflicting evidence suggests, however, that mRNA repression occurs during elongation, and that repression requires active translation by polysomes. In this case, adding miRNA to the CrPV IRES was able to completely repress expression, and adding miRNA to actively translated messages led ribosomes to dissociate from the target mRNA[124]. In both cases, the final result is clear: miRNA binding to the message results in strong translational repression.

The scope and process of miRNA repression have been studied primarily by mutating known target sites, looking for evolutionarily conserved miRNA target sites, and by modifying miRNA expression[125]. These analyses have come to the consensus that target identification occurs primarily through a heptameric 5' "seed" region of perfect complementarity. When mismatches occur in the seed, compensatory bases further downstream provide additional specificity. However, when individual miRNAs have been depleted using RNAi or miRNA mutants, the catalog of affected RNAs includes not only the predicted targets but also multiple RNAs which contained unanticipated binding sites. Furthermore, depleting cells of a specific miRNA led not only to an increase in the expected mRNAs, but also decreased expression of other, unexpected messages[126].

## **Computational Searches for mRNA Structure**

### **Computational Pipelines to Filter Data**

The technique of combining previous computational methods into more descriptive tools has been used in many contexts; one of the most eloquent examples is PSI-BLAST[127], a pipeline of BLAST output informing each subsequent database search. Another advantage of implementing a pipeline for performing large-scale searches lies in the potential to optimize each individual step and/or perform steps in parallel. Recent improvements to HMMER3[128] provide not only an excellent example of this concept at work, but also significant improvements over PSI-BLAST in terms of statistical significance and sensitivity. In a similar fashion, excellent work has been performed using RNAMotif[129] in order to implement computational pipelines searching for self-cleaving ribozymes[130,131], generic ribozymes[132], or even localization signals in mRNA[133]. The following catalogue of algorithms and implementations provides a view of some of the choices available for the proposed computational pipeline used in this work and is summarized in **Table 4**.

### **Grammar Searches in Large Datasets (RNAMotif)**

The aspects of RNAMotif which make it such a powerful and widely used tool include its pattern language which is able to distinguish each type of base pair (16 types of canonical and non-canonical) in any arbitrary configuration. In addition, it is able to score aspects of the motif including GC content, number of mismatches, or number of bases paired, among others. Finally, it is able to use experimentally determined thermodynamic parameters[134] in order to approximate the thermodynamic stability of each specific sequence motif it finds. Combined, these attributes make RNAMotif an extensible, multi-purpose tool when searching for potential mRNA secondary structures.

### **Predictive vs. Statistically Informed Searches**

The current implementation of the predicted ribosomal frameshift database (PRFdb) uses RNAMotif to find candidate sequences which have the potential to form H-type pseudoknots. RNAMotif, though powerful, is ill-suited to normalizing its output to find an approximation of the most stable secondary structure for a given sequence window; nor is it intended to take into account evolutionarily conserved mRNA structures. While the venerable mfold[135] suite accomplishes the former, it has no means to search for pseudoknotted sequences. Over time, usage of mfold has migrated towards the Vienna RNA folding applications[136]. This exhaustive suite of tools includes facilities to fold pre-aligned sequences *de-novo* predictions (excluding pseudoknots), RNA duplex and hybridization calculations, and RNA folding kinetics simulation. One missing piece in this suite of tools is the ability to search RNA databases to find similar or homologous sequences; inferNAI[137] nicely fills this niche by implementing stochastic context-free grammars in order to model stem covariance,

Name	Worst case Time	Predict $\psi$	Statistics / Evolutionary Information	Optimal or heuristic	Output sub-optimal predictions
<b>CMFinder</b>	$O(n^3)$	No	Yes	Heuristic	No
<b>HotKnots</b>	$O(n^4)$	Yes	No	Heuristic	Yes
<b>ilm</b>	$O(n^4)$	Yes	No	Heuristic	Yes
<b>infeRNAI</b>	$O(n^7)$	No	Yes	Optimal	Yes
<b>mFold</b>	$O(n^3)$	No	No	Optimal	Yes
<b>NUPACK</b>	$O(n^5)$	Yes	No	Heuristic	No
<b>pknots</b>	$O(n^7)$	Yes	No	Optimal	No
<b>pknots-RG</b>	$O(n^4)$	Yes	No	Heuristic	No
<b>PFinder</b>	$O(n^3)$	No	Yes	Heuristic	Yes
<b>rnamotif</b>	$O(n^2)$	Yes	No	Heuristic	Yes
<b>TT2NE</b>	$O(n^6)$	Yes	No	Heuristic	No
<b>Vienna</b>	$O(n^3)$	No	No	Optimal	No
<b>RNA</b>					

**Table 4: Summary of some common RNA prediction algorithms**

The most important aspects of a program to be used in a predictive pipeline for sequences containing strong secondary structures include: running time, ability to predict pseudoknots, what type of heuristic it uses (if it uses one), whether or not it provides sub-optimal structures, or uses statistical or evolutionary information, and its input/output requirements. Though it is not specifically a predictor of structure, the scoring functions of rnamotif make it a useful addition to this repertoire.

thus dissimilar sequences which fold into similar structures may be used to identify and align other sequences from a provided database. Like RNAMotif, these techniques depend initially upon a context-free grammar which describes the base pairing possibilities at each position, but they then add likelihood estimates for each observed base pair (for examples, see pages 233-269 of [138]). While extremely sensitive, these techniques are complex and can be computationally intensive (thus Tornado runs  $> O(n^3)$  with respect to time, even when using the simplest model[139]). Finally, though stochastic context free grammars (SCFGs) can be used to distinguish ambiguous characters, the nested structure of an RNA pseudoknot is unavailable to current implementations, including Pfold[140] and CMFinder[141].

As the primary purpose of the PRFdb is to identify mRNA sequences which are able to fold into pseudoknots, the limitation to only search for stem-loops is unacceptable. This leaves the simpler minimum free energy (MFE) strategies as the most likely tool. Searching in this manner is NP-complete[142], imposing a limit on searchable length of sequence. The classic dynamic programming algorithm solution to this problem was implemented in  $O(n^{6.8})$  time by pknots[143]; primarily because this implementation iterates through the entire search space for stem-loops and simple pseudoknots. At the time of its implementation, thermodynamic information was unavailable for some aspects (dangling bases in multi-loop structures, for example) of the nested stems found in pseudoknots and so these were estimated or filled in with contemporary values[144]. It is worth noting that the other programs discussed here use more current and complete values[134]. The iterative loop matching algorithm[145] and NUPACK[146] seek to improve the computational complexity of this problem by implementing heuristics to decrease the search space of the dynamic programming matrix. In the first case this is accomplished by initially collecting short optimal secondary structures, subsequently building them up until they attempt to include the entire sequence of interest; and then using the unpaired bases to perform another round of the same process, finally repeating as required for each class of secondary structures (pseudoknot, kissing loop, or stem-loop for examples). Therefore the worst case scenario for this algorithm is  $\geq O(n^4)$  (the  $n^3$  of a normal secondary structure prediction multiplied by another approximate  $n$  for the iterative aspect). HotKnots employs loop matching to achieve a similar simplification, but then applies another heuristic to determine if each matching loop (hotspot) is promising. This causes HotKnots to examine a slightly smaller subset of the available search space and perform at  $O(n^4)$  time. On the other hand, HotKnots uses a more complete free-energy model which may also be easily replaced with other models (including the covariance models used by

Tornado above) in order to add a statistical or phylogenetic aspect to the scoring function. TT2NE is a more recent addition to this list of heuristics[147], performing a depth-first search of the available space. It then applies a branch-and-bound procedure (as had been suggested in the publications of every preceding MFE minimization algorithm as well as Durbin & Eddy) to skip segments of the search space which cannot terminate in a MFE less than the current minimum. This brings it to  $O(n^3)$  or  $O(n^4)$  depending on the sequence, with the pathological case running in  $O(n^6)$  time. Each of these programs uses approximately  $O(n^3)$  memory.

Given these competing criteria for successful mRNA structure prediction, the PRFdb currently is able to use pknots, NUPACK, HotKnots, Vienna RNAFold, and mFold. TT2NE support is intended but has not been implemented while ilm is functional but has not yet had its output passed to a thermodynamic parameter evaluation (Vienna RNAeval is an excellent candidate but requires file-format conversion, indeed it should be used to re-evaluate the pknots MFE calculations) to provide a better predicted MFE.

### **Sequence Randomization Strategies**

Once MFE minimization strategies have been performed and potential mRNA secondary structures predicted, the question shifts to one of scoring. How does one decide among 190,000+ predictions (for *Homo sapiens*) which are the most significant? Previous work has compared 6 quantitation methods[148] for scoring mRNA structure predictions: MFE/base pair, *Z*-score, *p*-value, Shannon entropy, average base-pair distance, and the valley index. This work suggests that MFE/base, *Z*-score, and *p*-value are sensitive to the strength of the given secondary structure while the Shannon entropy, average base-pair distance, and valley index indicate the uniqueness of the given structure. Of the three measurements which are sensitive to the predicted structure's strength, the *Z*-score was shown to be the most useful; however it depends on the



specific sequence randomization strategy employed. Debate continues regarding the optimal sequence randomization strategy; some work suggests that simple Fisher-Yates shuffling is sufficient, while others debate the relative merits of maintaining dinucleotide frequencies[149,150], mononucleotide frequencies[151], amino-acid frequencies, etc. The differences in sensitivity between the Z-scores for these various methodologies are relatively small. The current PRFdb implementation is able to perform them all, but defaults to a simple shuffle.

### **Storage, Retrieval, and Visualization**

It is important not to lose sight of the primary goal of this work: implement a simple to use method to search genomic sequence for potential -1 PRF signals. The methodologies for searching through large amounts of sequence data for strong mRNA structures have been established. Finding potential pseudoknots is NP-complete, but possible for limited cases via heuristics; and methods exist which provide some measurement of significance. The final step is to put these pieces together. Bioperl[152], the “LAMP” (Linux, Apache, MySQL, Perl) software stack, and HTML::Mason<sup>aa</sup> provide a simple solution to this problem. These tools make developing a medium-scale database of potential -1 PRF signals (currently containing 525,000+ ORFs, easily scalable to the 5,000,000+ of Genbank) simple. The only remaining puzzle lies in how to visualize the results from the various MFE minimization algorithms. Bioperl provides methods to convert most of the various output formats, but one might want to see a dot-plot of the pseudoknotted sequence, linear Feynman diagram[153], or planar drawing[154]. The source code for jViz and accompanying thesis made implementing similar visualization strategies possible in Perl.

---

<sup>aa</sup> <http://www.masonhq.com>

## **Summary**

Each of the preceeding fields informs one or more aspects of this thesis. An understanding of programmed ribosomal frameshifting depends upon normal elongation; so too does creating a pipeline to search for strong mRNA secondary structures depend upon an understanding on the dynamic programming algorithm. Similarly, one cannot search for changes in mRNA stability without some idea of the players involved from transcription to decay. The most attractive aspect of the central hypothesis of this work lies in its interdependence on so many fields. This is a two-edged sword: while it provides excellent opportunities to explore hypotheses across the spectrum of computational and molecular biology, it falls prey to the competing hypotheses of the fields and fractious nature of some emergent fields. Thus changing observations in the field of mammalian nonsense mediated decay provide simultaneously worrisome changes in how this work will be interpreted as well as opportunities to fine-tune our hypotheses and experiments. Similarly, the competing hypotheses in the field of microRNA make it difficult to foresee the best method to direct our hypotheses regarding PRF regulation; but provide wonderful new insights into the relationships between translational repression versus mRNA decay. In this environment of astonishingly fast paced change, I hope this document provides a useful foundation to formulate new hypotheses and interpret future observations.

## Chapter 2

# A Database of Computationally Predicted Programmed -1 Ribosomal Frameshift Signals

### Introduction

Canonical decoding of the genetic code requires translating ribosomes to convert triplets of bases (codons) into amino acid sequences. Although this algorithm is employed for translation of the vast majority of mRNA sequences, in some special cases cis-acting mRNA elements direct ribosomes into alternative reading frames, dynamically “recoding” their sequence information (reviewed in [155]). Programmed -1 Ribosomal Frameshifting (-1 PRF) was first discovered in RNA viruses where it enables viral genomes to encode multiple peptides from a single mRNA [10]. An individual -1 PRF signal consists of a heptameric 'slippery site' usually followed by an mRNA pseudoknot secondary structure separated by a suitable spacer region (reviewed in [13,156]). Unlike their viral counterparts, eukaryotic genome-encoded -1 PRF signals are predicted to direct elongating ribosomes into premature termination codons [23]. Such events have been shown to initiate rapid mRNA degradation in yeast through the Nonsense Mediated Decay (NMD) pathway[78]. As such, -1 PRF is hypothesized to add a novel modality for regulation of gene expression at the post-transcriptional level.

There are currently three databases serving the translational recoding community.

RECODE (<http://recode.genetics.utah.edu/>) is a browsable collection of all the published translational recoding signals[6,157,158] RECODE's strength is as central repository of all empirically proven translational recoding signals. FSDB (<http://wilab.inha.ac.kr/fsdb/>) contains a compilation of a handful of known and predicted viral, prokaryotic and eukaryotic -1 and +1 PRF signals, and also allows users to input their own sequences to

search for frameshift signals using a program called FSFinder[159]. This site provides tools not available through RECODE, in particular it integration of PseudoViewer, a powerful graphics tool for that simplifies visualization of H-type pseudoknots[160]. MLOGD (<http://guinevere.otago.ac.nz/aef/MLOGD/index.html>) is a suite of software that allows detection of new protein-coding sequences by identifying overlapping open reading frames[161]. While all three of these sites have their strengths, a common weakness is that they do not provide well catalogued, searchable databases of all potential recoding signals of any one kind. To fill this gap, we have created PRFdb as a database of predicted -1 PRF signals in eukaryotic genomes. The methods used to search for predicted -1 PRF signals have been previously described, and importantly, we have empirically demonstrated that a significant number of -1 PRF signals so identified actually promote significant levels of frameshifting[23]. The strength of the PRFdb is that it provides a tool for researchers outside of the translational recoding field to use to quickly search for and identify potential -1 PRF signals in genes in which they are interested.

### Database Description

In the PRFdb, the predicted -1 PRF signals are represented by: 1) the genes in which they reside; 2) the identity and location of their slippery sites; 3) graphical representations of their predicted secondary structures; 4) computationally identified minimum free energies (MFE); and 5) the thermodynamic significance of these mRNA structures as compared to randomized variants. Currently completed genomes in the PRFdb include: *Saccharomyces cerevisiae*, *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Bos Taurus*, *Danio rerio*, and *Xenopus tropicalis*. A listing of examined genomes with more than 400 ORFs may be found in **Appendix 11**.

Researchers can access data in the PRFdb through four means: (i) Search (Figure 20) provides a way to query a gene of interest using the specific gene name,

**Figure 20: The PRFdb Search Interface**

The PRFdb search interface provides searches against the indexed description text, specific accessions, specific HGNC[162] gene names, or a BLAST interface which allows one to search for genes of interest with nucleotide or protein sequence.

HGNC id or description. The search interface also provides a means to use BLAST to search for genes in the PRFdb similar to a query sequence. The initial result from a search provides further information about the ORF, other search tools, and a summary of the potential -1 PRF signals detected for the given search (Figure 21).

Gene Name: YLR318W EST2

Accession: SGDID:S0004310

Species: *Saccharomyces cerevisiae*

Comments: YLR318W EST2

SGD Link

Run Local Blast or Remote Blast using this gene.

Download: Nucleotide or Amino acid sequence.

Position	Slippery Site	#Folds
1	Start Codon	
72 (71, 2.7%)	AAAUUUA	12
75 (74, 2.8%)	UUUAAAA	13
591 (590, 22.2%)	UUUUUUU	9
1137 (1136, 42.8%)	UUUUUUU	10
1215 (1214, 45.7%)	CCCUUUU	10
1326 (1325, 49.9%)	AAAAAAA	10
1653 (1652, 62.2%)	AAAAAAU	18
1995 (1994, 75.1%)	UUUAAAA	15
2313 (2312, 87.1%)	AAAUUUA	9
2316 (2315, 87.2%)	UUUAAAU	9
2655	Stop Codon	
All		

Green: 0 Frame Start Codon

Red: 0 Frame Stop Codon

Blue: Slippery Site

Orange: First -1 PTC

Maroon: SNP

>SGDID:S0004310 | YLR318W EST2

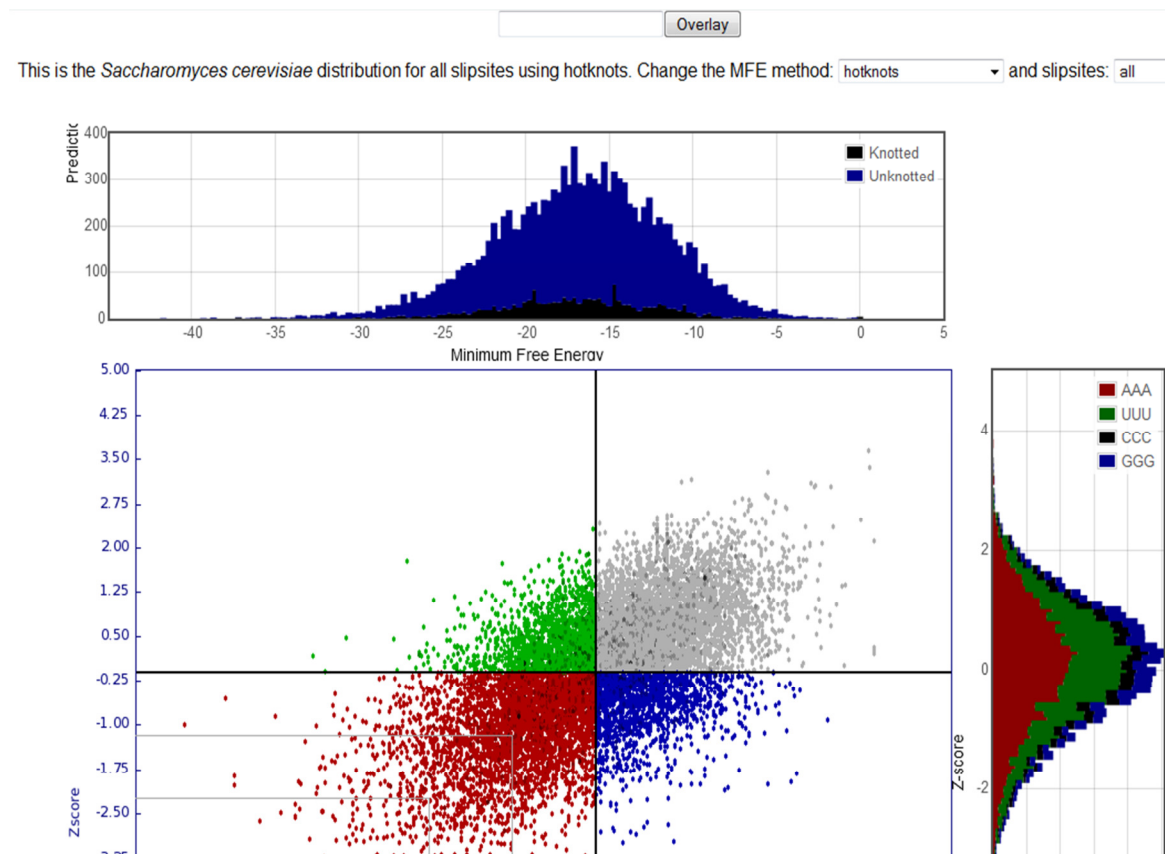
1	AUG	AAA	AUC	UUA	UUC	GAG	UUC	AUU	CAA	GAC	AAG	CUU	GAC	AUU	GAU	CUA	CAG	54	
55	ACC	AAC	AGU	ACU	UAC	AAA	GAA	AAU	UUA	AAA	UGU	GGU	CAC	UUC	AAU	GGC	CUC	GAU	108
109	GAA	AUU	CUA	ACU	ACG	UGU	UUC	GCA	CUA	CCA	AAU	UCA	AGA	AAA	AUA	GCA	UUA	CCA	162

**Figure 21: A Representative Search Result.**

The top of this page provides information pertaining to a specific gene (*S. cerevisiae* EST2), its genome database (SGD) entry, a link to perform BLAST searches for similar genes, MFE minima graph, and a link to download its sequence. Following this information is a list showing the locations of the translational start site, potential slippery sites and the number of secondary structure solutions that have been computed for them, and the 0-frame termination codon. At the bottom is a display of the gene where the ATG start site is displayed in green, slippery sites are shown in blue, and -1 frame termination codons are shown in orange. The specific entry for each potential frameshift signal may be viewed by clicking on the slippery site's position or its link in the

sequence. In addition, locations of human single nucleotide polymorphisms catalogued in the NCBI Single Nucleotide Polymorphism Database rendered in maroon. Clicking these will open links to the database <http://www.ncbi.nlm.nih.gov/projects/SNP/>.

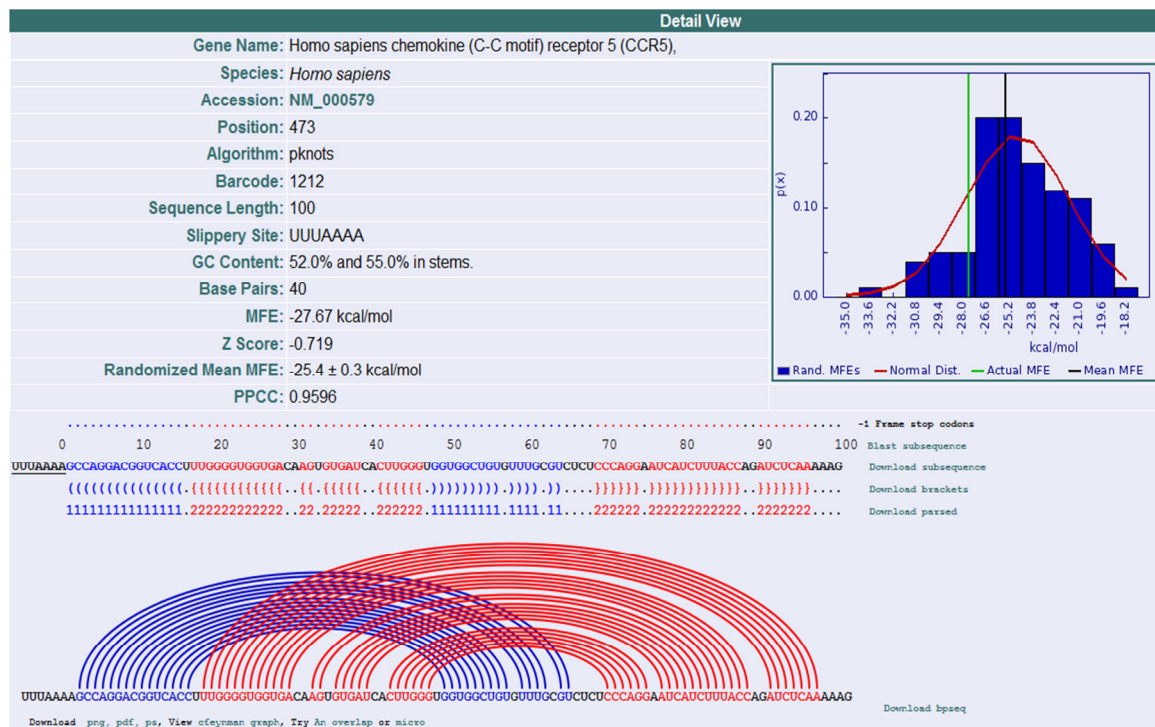
*ii) Distribution (Figure 22)* enables browsing for sequences containing statistically significant putative -1 PRF signals through a graphical representation of computed minimum free energies with respect to randomized Z scores for all sequence windows. It is also possible to limit this distribution to sequences that are preceded by a specific slippery site. *iii) Filter* prints sequences from a given genome that meet specific criteria including: species, pseudoknotted sequence, sequences with a specific number of base



**Figure 22: The distribution of *Saccharomyces cerevisiae* sequences.**

Computed minimum free energy is on the x-axis, Z-score is on the y-axis. Black lines denote the mean values and gray lines define sequence windows that are one and two standard deviations less than mean. Clicking on any region links to the closest -1 PRF signals with respect to MFE and Z score. Distributions of the numbers of sequences at each MFE and Z sorted by pseudoknot status and slippery site identity are along each axis, respectively. It is possible to overlay the values an accession of interest or view the distribution using different MFE minimization algorithms using the fields at the top.

pairs and/or MFE. *iv)* *Download* provides a format suitable for parsing all sequences of a given genome/sequence dataset. The search, distribution, and filter interfaces lead to a detailed description (**Figure 23**) of individual putative PRF signals. This provides a summary of all data gathered for a given sequence including: background information on the gene and location of the -1 PRF signal, information regarding the program used to perform the MFE prediction, multiple methods to view the secondary structure, and a comparison of the distribution of randomized sequences to the MFE of the folded sequence.



**Figure 23: Details of the *Homo sapiens* CCR5 -1 PRF signal.**

This demonstrates that pknots was used to compute an MFE of -27.7 kcal/mol for the 100 bases following the UUUAAAA slippery site at position 473 of the CCR5 ORF. When randomized 100 times using Fisher-Yates shuffling, a mean MFE of -25.4 kcal/mol was computed for a normal distribution of correlation coefficient 0.96. The MFE distribution of the randomized sequences is on the right; with the idealized normal distribution in red. The black vertical line marks the mean MFE of the randomized sequences, and the green vertical line marks the MFE of the native sequence. This secondary structure is more stable than random (z score = -0.72). The predicted mRNA secondary structure of this sequence is shown below using both bracket notation, and using a Feynman diagram. Links below provide download links for png/ps/pdf images, the bpseq format for this structure, a circular Feynman diagram, an overlap of all MFE predictions, and a microRNA prediction using miRanda.

If a sequence of interest is not currently in the PRFdb, it can be imported via its NCBI accession number. Sequences added in this manner will be filtered within hours of import. Sequences imported into the PRFdb are also folded using sequential windows across the entire sequence in order to create a graphical minimum free energy 'landscape.' This enables users to submit longer or shorter sequence strings for computational folding, a particularly useful feature e.g. for eliminating extraneous sequence that may not be involved in actual RNA folding. For example, the computational analysis of the 100 nucleotide sequence downstream of the slippery site of the mouse Ma3 -1 PRF signal provided by the PRFdb predicts tandem stem loop structures. However, when only 55 nt of downstream sequence are provided, PRFdB predicts the empirically documented pseudoknot structure[66].

## **Process**

Sequences to be analyzed by the PRFdb are imported into the database, filtered using *RNAMotif*[129], folded with secondary structure prediction algorithms, randomized using one or more randomization methods, and refolded. To avoid complications of untranslated intronic sequences, the PRFdb contains only mature mRNA sequences (cDNA sequences primarily). Sequences are imported into the PRFdb from a web interface using Genbank accession numbers, yeast genome accessions, or raw sequences. Each new sequence is first passed through a simple text filter that searches for slippery sites following the International Union of Pure and Applied Chemistry (IUPAC) pattern 'N NNW WWH' (or X XXY YYZ), where N (X) denotes any three identical bases, W (Y) denotes AAA or UUU, H (Z)  $\neq$  G, and spaces indicate the incoming (zero) reading frame. Since the distance between the end of the slippery site and the downstream stimulatory sequence is important, a spacer of 1 to 8 nucleotides was incorporated into the search. Remaining sequences are passed to *RNAMotif* with a



descriptor looking for the potential to form an mRNA pseudoknot. Sequence windows passing this minimal test are passed to multiple pseudoknot predicting mRNA secondary structure prediction algorithms, including *Pknots*[143], *Nupack*[146,163], *Hotknots*[164], *Mfold*[165], and the Vienna RNA package[136]. After folding, every sequence is randomized using one or more algorithms including: Fisher-Yates shuffling to maintain dinucleotide frequencies, codon frequencies, or nucleotide frequencies. The resulting random sequence windows are then refolded without searching for pseudoknots. This process is repeated a fixed number of times (100 by default) to create a distribution of sequence specific randomized MFEs. These resulting distribution of randomized MFEs is then compared the MFE of the original sequence window. These values are used to compute a z score, thus providing a measurement of the significance of the native sequence.

## **Discussion**

The BLAST interface to the PRFdb is currently being used to discover Genbank sequences similar to the most statistically significant sequences in the database, thus providing a means to expand the PRFdb in a depth first manner. As more similar sequences are completed, comparative genomics studies using sequence and/or mRNA structure alignments will be incorporated to enable identification of conserved -1 PRF signals across species and/or genes. As time progresses, additional computational and empirical information will allow for improved scoring, helping to increase the statistical relevance of the predicted secondary mRNA structures. These improvements will continue to make the PRFdb more useful and accessible to the research community, providing a resource allowing individual users to identify -1 PRF signals in genes of interest, and as a metasource of information for cross referencing with other databases, e.g. genomes and DNA microarray databases.

## Chapter 3

# Endogenous Ribosomal Frameshift Signals Operate as mRNA Destabilizing Elements Through at least Two Molecular Pathways in Yeast.

### Introduction

Programmed ribosomal frameshifting (PRF) is has been historically associated with viruses. PRF signals stochastically redirect ribosomes into new reading frames, and viral PRF promotes synthesis of C-terminally extended fusion proteins. The most well defined PRF signals direct ribosomes to slip by one nucleotide in the 5' (-1) direction. -1 PRF signals typically contain three elements: a “slippery site” composed of seven nucleotides (N NNW WWH, incoming 0-frame indicated by spaces) where shifting occurs; a short spacer sequence; and a downstream stimulatory structure, typically an mRNA pseudoknot[1,13]. Current models posit that the pseudoknot directs ribosomes to pause with their aminoacyl- (aa-) and peptidyl-tRNAs positioned over the slippery sequence, where re-pairing of the non-wobble bases of both tRNAs with the -1 frame codons occurs[7,166,167]

It is now clear that PRF is employed by organisms representing every branch in the tree of life, suggesting an ancient and possibly universal mechanism for controlling the expression of actively translated mRNAs[17]. The past few years have witnessed several reports describing *in silico* identification of recoding signals using a variety of computational approaches[16,23,24,168–171]. While the methodologies of each study covered a broad range of bioinformatics techniques, the general goal was to first find out-of-frame ORFs followed by the identification of PRF signals in the overlapping region

between them. While this can identify new classes of PRF signals, it is based on the assumption that PRF outcomes should mimic those observed in viral genomes and thus cannot identify new functional outcomes of frameshifting.

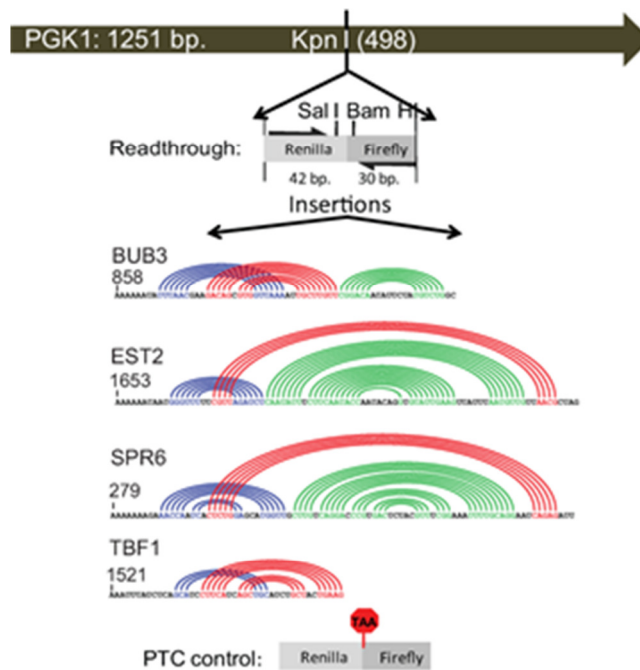
While “outcome-neutral” approaches using mRNA motifs known to promote efficient PRF cannot identify new classes of frameshift signals, they enable an expansion of our understanding of functional uses for PRF. The seminal study in this field searched the yeast genome for -1 PRF promoting motifs resembling well characterized examples of viral -1 PRF signals, identifying ~260 putative such elements[24]. This work was limited by incomplete annotation of the yeast genome and insufficient computational resources available at the time. New bioinformatics tools were subsequently developed and applied using faster and more robust computational platforms. The results showed that: pattern matching approaches coupled with a predictive method for folding RNA sequences provided a dramatic improvement in the results; -1 PRF motifs are widespread in the genome of *S. cerevisiae*; and many have predicted secondary structures with statistically significant measures of free energy[23]. This analysis showed that ~11% of yeast genes contain at least one high probability -1 PRF signal. Furthermore, we demonstrated that 9 putative -1 PRF signals selected from a variety of *S. cerevisiae* genes promoted efficient recoding *in vivo*. More recently, this bioinformatics protocol has been applied to more genomes. Currently, more than 25 genomes have been analyzed, and it appears that 8-10% of genes contain at least one potential -1 PRF signal (See the PRFdb at: <http://prfdb.umd.edu/>)[172].

A key finding was that the outcome and function of -1 PRF differs significantly between the viral and ‘genomic’ contexts. In viruses, PRF controls the stoichiometries of structural versus enzymatic proteins[65]. In contrast, ‘genomic’ PRF events redirect elongating ribosomes to premature termination codons, suggesting that -1 PRF is used to control cellular mRNA abundance and stability through the nonsense-mediated mRNA

decay (NMD) pathway. A proof-of-principle experiment demonstrated that a viral -1 PRF signal can function as an mRNA destabilizing element and that mRNA destabilization required NMD[78]. Here, rapid degradation of a reporter mRNA by -1 PRF through NMD is demonstrated for four genomic yeast -1 PRF signals. Further, the presence of the PRF-stimulating pseudoknot can promote mRNA destabilization through no-go decay(NGD)[45]. The *EST2* gene, encoding the catalytic subunit of telomerase[173], was used to delve deeper into the relationships between -1 PRF and mRNA stability. The *EST2* mRNA is destabilized by -1 PRF primarily via NMD, and ablation of its five -1 PRF signals resulted in stabilization of the *EST2* mRNA.

## Results

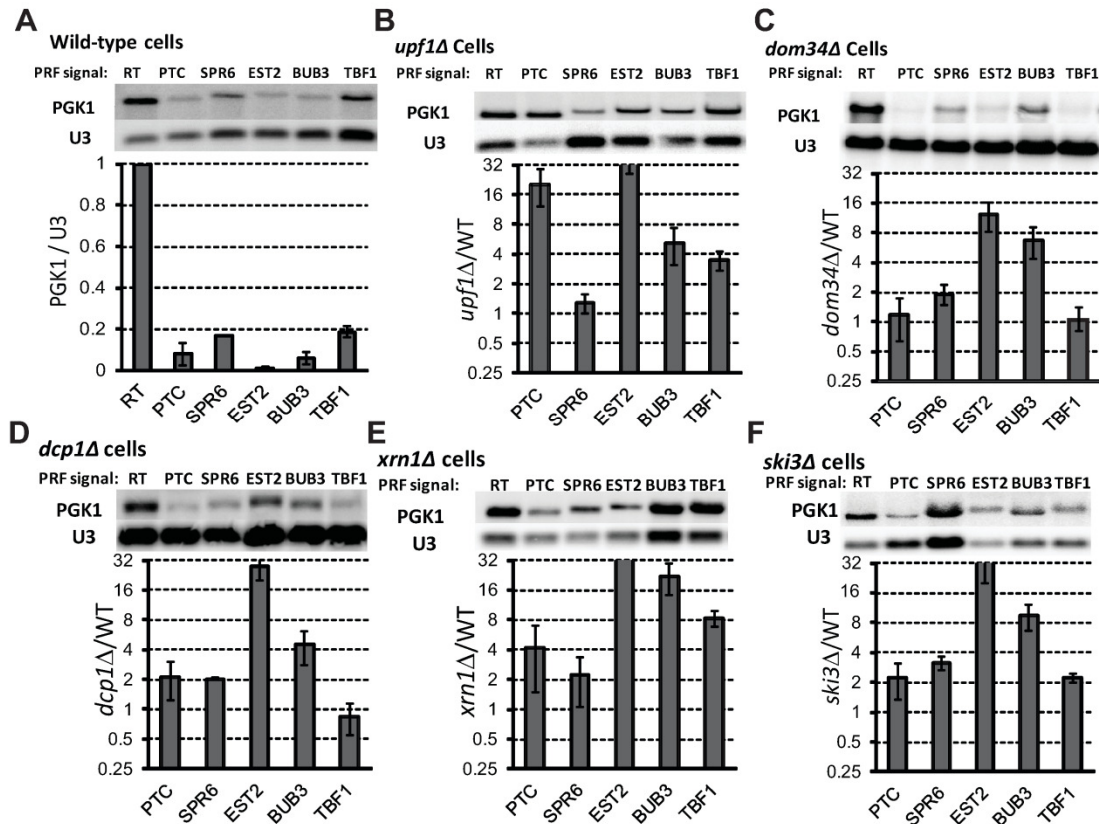
Genomic -1 PRF signals are mRNA destabilizing elements. Four functional yeast genomic -1 PRF signals derived from the *BUB3*, *EST2*, *SPR6*, and *TBF1* genes were employed to test the hypothesis that -1 PRF function as mRNA destabilization elements. The slippery heptamers for these -1 PRF signals begin at nucleotides 858, 1653, 279, and 1521 of their respective ORFs. These were cloned into a yeast *PGK1* reporter gene so that frameshifted ribosomes are directed to PTCs. All inserts were flanked by sequences derived from *Renilla* and firefly luciferase genes, providing unique exogenous sequences for specific detection of the reporter mRNAs. Two additional *PGK1* reporters without -1 PRF signals, were used as controls: a readthrough reporter encoded a continuous ORF, while a PTC control contained an in-frame UAA termination codon (**Figure 24**).



**Figure 24: Schematic of PGK1 reporter vectors.**

The indicated Renilla and firefly luciferase derived sequences from pJD375 were cloned into the unique Kpn I restriction site in a high copy PGK1 expression vector to create the readthrough control (pJD753). The indicated -1 PRF signals derived from BUB3, EST2, SPR6, and TBF1 were cloned into Sal I/Bam HI digested pJD753. Colored arcs depict computationally predicted base-paired stems[129]. The premature termination control (PTC) was constructed by mutagenizing pJD753 to create an in-frame TAA codon.

Reporters were introduced into wild-type yeast cells; their steady state mRNA abundances were determined by RNA blot analysis and normalized to U3 snoRNA controls. A minimum of three independent blots were performed for all experiments. In wild-type cells, all four of the genomic -1 PRF signals and the in-frame PTC containing control destabilized the PGK1 reporter mRNA (**Figure 25**).



**Figure 25: Genomic -1 PRF signals can function as mRNA destabilization elements in yeast.**

The -1 PRF signals from *SPR6*, *EST2*, *BUB3*, and *TBF1* were cloned into a *PGK1* reporter such that frameshift events would cause elongating ribosomes to encounter premature termination codons (PTC). Readthrough (RT) and in-frame PTC containing reporters are included as controls. Northern blots of total mRNAs extracted from logarithmically growing cells were probed with a reporter-specific oligonucleotide (PGK1), stripped and re-hybridized with a U3 snoRNA-specific probe for normalization. All blots were repeated at least two times. **A**: steady-state abundance of reporter mRNAs in wild-type cells. Each graph shows abundances of test mRNAs relative to the readthrough control. **B**: Same as panel A, but in *upf1Δ* cells. Graph plots abundance of specific test mRNAs in *upf1Δ* versus wild-type cells. **C – F** are similar to panel B, except that samples were extracted from *dom34Δ*, *dcp1Δ*, *xrn1Δ*, and *ski3Δ* cells respectively.

The extent of mRNA destabilization varied from ~0.01 fold of the readthrough control (*EST2*) to ~0.19 fold of wild-type (*TBF1*). Experiments were also performed in *upf1Δ* and *dom34Δ* strains, and the U3-normalized signal intensities were compared among the same signals between wild-type and mutant strains to determine the relative contributions of NMD and NGD to the ability of the -1 PRF signals to act as mRNA destabilization elements. The PTC containing mRNA was only destabilized through the

NMD pathway: 28-fold stabilization in *upf1Δ* cells relative to wild-type cells, but no stabilization in *dom34Δ* cells. The TBF1 -1 PRF signal similarly destabilized the reporter signal only through NMD (~4-fold). In contrast, the EST2 and BUB3 -1 PRF signals functioned through both pathways: the EST2 signal stabilized the reporter ~35-fold in *upf1Δ* cells and ~14-fold in *dom34Δ* cells, while the values for the BUB3 signal were ~7-fold and ~8-fold respectively. The mRNA destabilization activity of the SPR6 -1 PRF signal was primarily through NGD (~6.0-fold stabilization in *dom34Δ* cells). Deletion of DCP1, XRN1 and SKI3, all of which are epistatic to UPF1 and DOM34, also generally stabilized the reporter mRNAs. We note however that, in the case of the *dcp1Δ* cells, the continued presence of Dcp2p likely provided residual decapping activity. These results establish that endogenous genomic -1PRF signals can function as mRNA destabilizing elements in yeast through at least two mRNA degradation pathways.

#### **The EST2 -1 PRF signal at Nucleotide 1653 is Destabilized by -1 PRF Induced NMD**

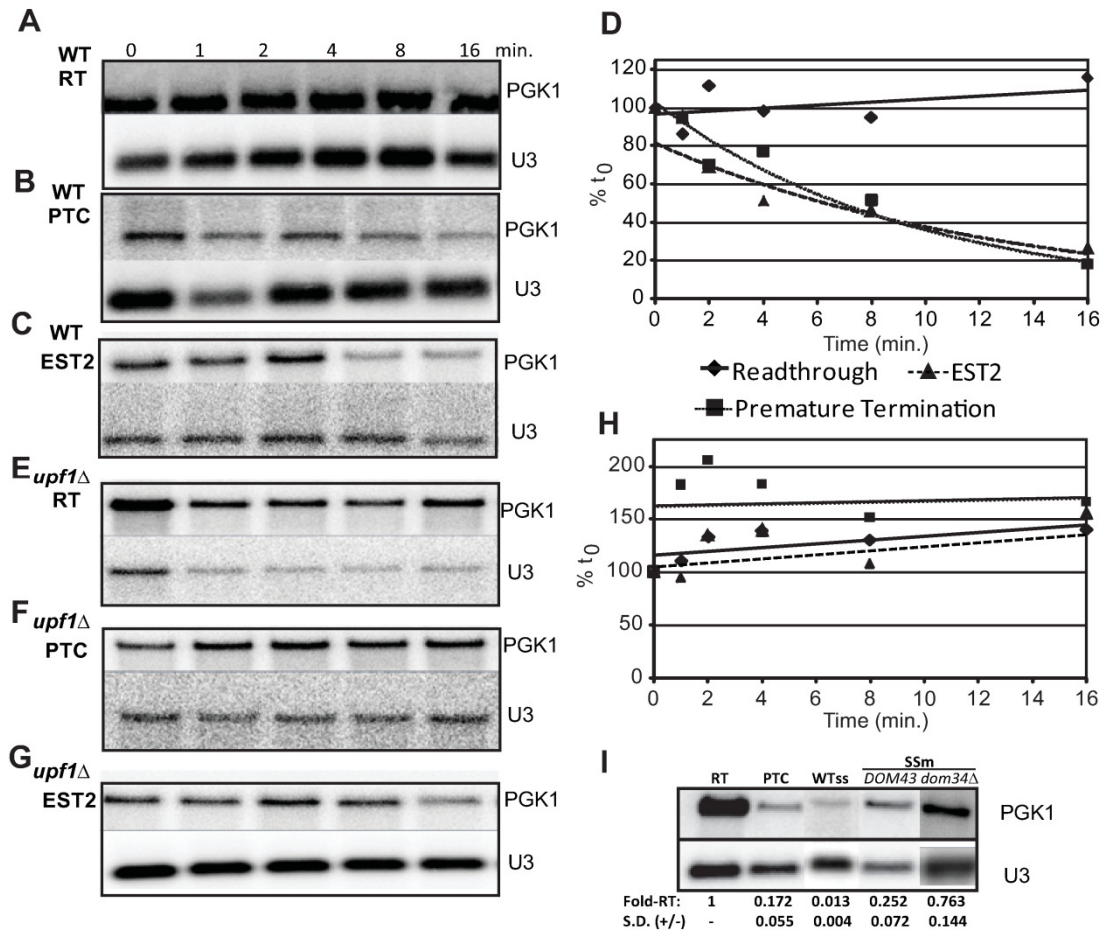
**Figure 26** suggests that -1 PRF induced NMD is the major cause of mRNA destabilization by the EST2 -1 PRF signal beginning at nucleotide 1653. To confirm this, a series of time course mRNA decay assays were performed employing the PGK1-EST2 -1 PRF reporter, the readthrough control, and the PTC containing construct in cells harboring the temperature sensitive *rpb1-1* allele of RNA polymerase II. At the zero timepoint, cells were shifted to the nonpermissive temperature (42°) to arrest transcription of mRNAs, total cellular mRNAs were extracted at 0, 1, 2, 4, 8, and 16 minutes following the temperature shift, and RNA blots were hybridized with the firefly luciferase and *U3* snoRNA probes. While the readthrough control was stable in wild-type cells (**Figure 26 A, D**) both the PTC containing control and the reporter containing the EST2 -1 PRF signal promoted rapid exponential decay of the reporter mRNA (**Figure 26 B, C, D**). In a parallel experiment using *rpb1-1 upf1Δ* cells, all of the reporter

mRNAs remained stable (**Figure 26 E-H**). The rapid decay kinetic profile of the *EST2* -1 PRF containing reporter, and its stabilization in NMD-deficient cells are consistent with NMD being the major decay pathway triggered by this element[78]. To independently test this, the A AAA AAT slippery site was partially inactivated by mutating it to G AAG AAC. This silent mutation stabilized the reporter mRNA ~19-fold compared to the wild-type slippery site (**Figure 26 I**). Interestingly, this is less than the 35-fold stabilization in *upf1Δ* cells. One would expect that, since this destabilization is dependent on -1 PRF, then inactivation of -1 PRF should be quantitatively the same as inactivation of NMD. However, this particular slippery site mutant is predicted to have some residual frameshifting activity due to the ability of the uracils in the A- and P-site tRNA anticodons that base pair with the A residues at the first position of the two 0-frame codons to re-pair with the G bases in the -1 frame. Thus, we suspect that some low levels of -1 PRF may still contribute to destabilization of this reporter.

#### **Ablation of -1 PRF Signals Stabilizes the Yeast *EST2* mRNA.**

The *EST* family of yeast genes is named after their “Ever Shortening Telomere” phenotype[174]. *EST2* encodes the catalytic subunit of telomerase and the other three *EST* genes either encode protein subunits of telomerase (*EST1* and *EST3*) or a telomere-associated regulator of telomerase (*CDC13/EST4*)[175]. Telomere elongation occurs in late S phase, although Est2p is associated to varying extents with telomeric chromatin throughout the cell cycle, and telomerase defects result in chromosome instability and rapid senescence[176]. The very low abundance *EST2* mRNA is stabilized in NMD-deficient cells[177,178].

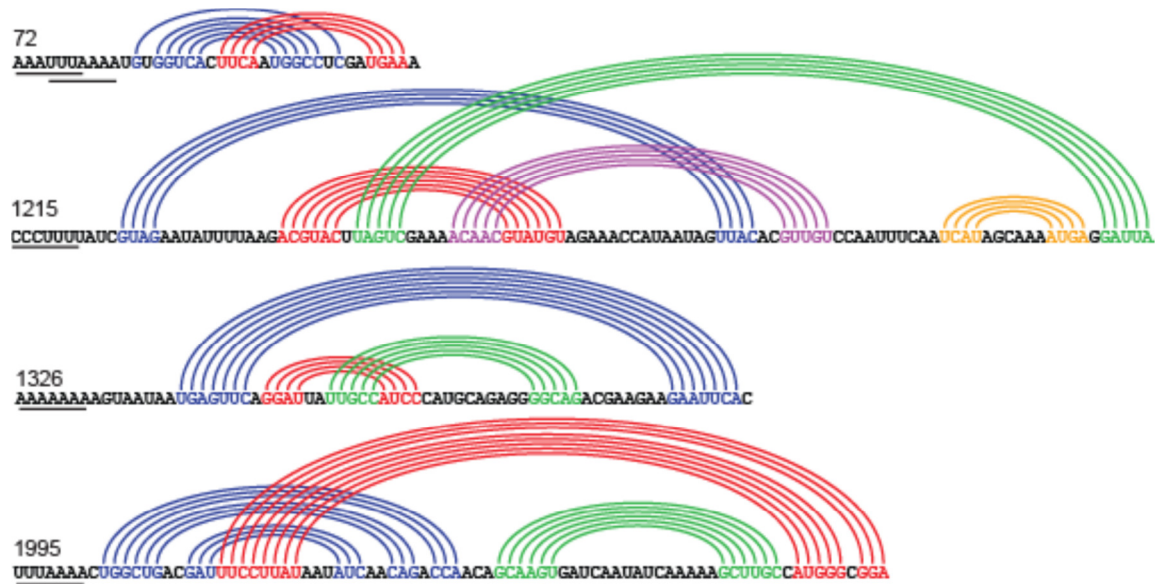




**Figure 26: The *EST2* -1 PRF signal at position 1653 destabilizes mRNA through NMD.**

**Panels A – H:** the readthrough control, in-frame PTC control, and *EST2* -1 PRF containing PGK1 reporters were introduced into either wild-type (**A – D**) or *upf1Δ* (**E – H**) cells harboring the temperature-sensitive *rpb1-1* allele of RNA polymerase II. Total mRNAs were harvested from cells after temperature shift at the indicated timepoints, and Northern blots were probed using the PGK1 reporter-specific and U3 snoRNA specific probes. Graph in panel **D** plots normalized PGK1 reporter mRNA abundances in wild-type cells, and graph in panel **H** plots these data in *upf1Δ* cells. **I:** the wild-type A AAA AAT slippery site of the *EST2* -1 PRF signal in the PGK1 reporter was changed to G AAG AAC, and steady state northern blot analyses were performed using mRNAs extracted from cells expressing the readthrough control, the in-frame PTC containing control, and cells expressing either the wild-type or mutant slippery sites. Fold-RT denotes fold readthrough control. S.D. (+/-) denotes standard deviation.

Computational analyses revealed that *EST2* contains four additional high confidence -1 PRF signals beginning at positions 72, 1215, 1326, and 1995 (**Figure 27**). The positions of the five predicted -1 PRF signals in the *EST2* ORF are shown in **Figure 28 A**. Silent protein coding changes were introduced into the slippery sites of all 5 of the



**Figure 27: Four additional high confidence -1 PRF signals in *EST2*.**

The predicted -1 PRF signal beginning at nucleotide 72 contains two overlapping slippery sites (AAU UUA and U UUA AAA). Slippery sites beginning at nucleotides 72, 1215, 1326, and 1995 are underlined. Colored arcs depict base pairing of stems.

-1 PRF signals in a full-length *EST2* clone expressed from a low copy vector

pEST2wt, pEST2ssmut, and derivative plasmids share the backbone plasmid: pJD641.

(pEST2ssΔ, **Figure 29**). Clones expressing either wild-type *EST2* (pEST2wt) or

pEST2ssΔ were introduced into isogenic *est2Δ* or *est2Δ upf1Δ* cells, and qRT-PCR

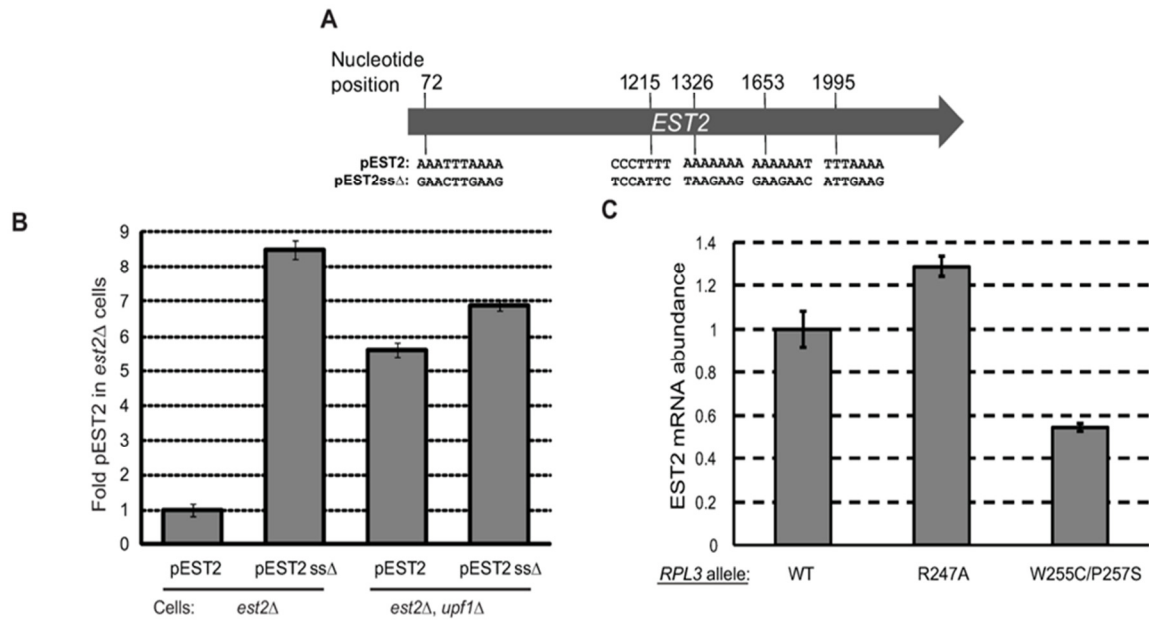
analyses were performed. These silent mutations resulted ~8.5-fold stabilization of the

full-length *EST2ssΔ* mRNA relative to wild-type *EST2* mRNA (**Figure 28 B**). Similarly,

abrogation of NMD stabilized the wild-type *EST2* and *EST2ssΔ* mRNAs ~5.8-fold and

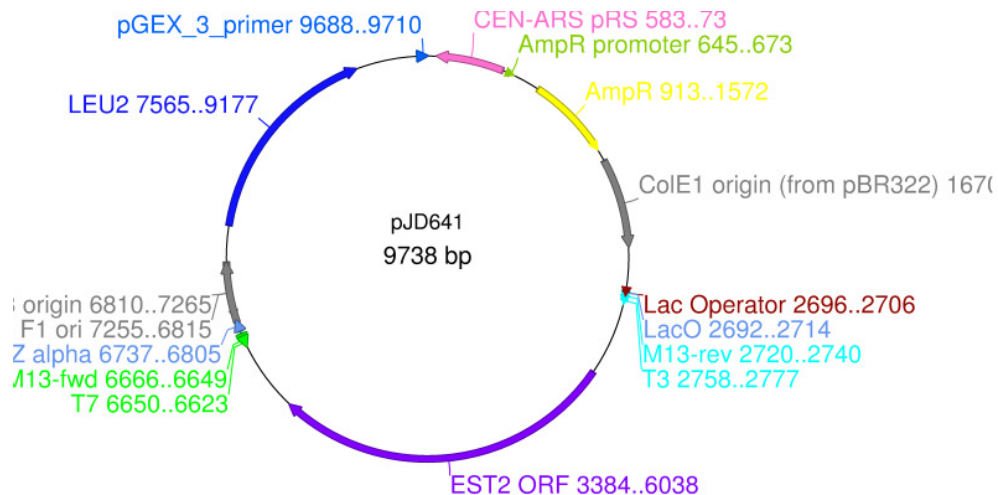
~7.0 fold respectively, thus demonstrating that -1 PRF induced NMD plays a significant

role in destabilizing the *EST2* mRNA.



**Figure 28: Silent mutations that disrupt slippery sites in *EST2* gene stabilize its mRNA.**

**A**, schematic of the *EST2* coding sequence. Positions of the slippery sites of 5 predicted -1 PRF signals and their sequences are indicated. The full-length gene including native 5' and 3' UTR sequences were cloned into a low-copy yeast vector to create pEST2. Silent coding mutations that are predicted to inactivate -1 PRF were introduced to produce pEST2ssΔ. **B**, pEST2 or pEST2ssΔ were introduced into *est2Δ* or *est2Δ upf1Δ* cells and *EST2* mRNA steady state abundances were determined by quantitative real-time PCR.

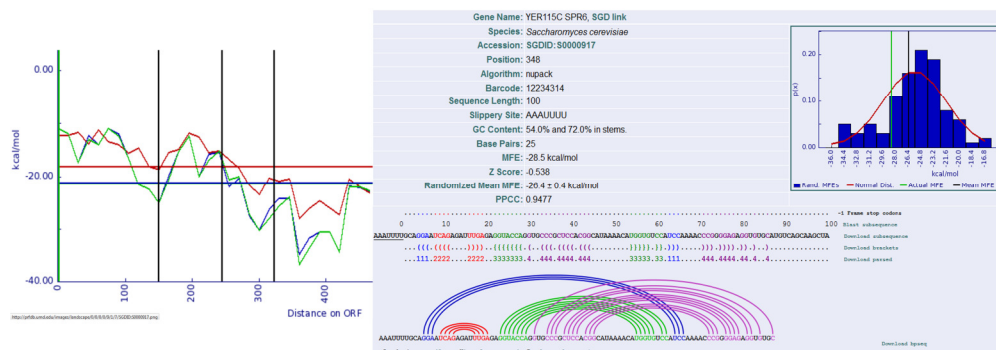


**Figure 29: Map of the full length *EST2* low-copy plasmids.**

Of particular note is the Centromeric yeast origin of replication (12 o'clock) and the sequence flanking the *EST2* ORF, this includes the endogenous UTRs and M13, T7, and T3 regions.

### Programmed -1 Ribosomal Frameshifting, but not Specific -1 PRF Signals, Appears to be Conserved Among Yeasts.

If regulation of gene expression through -1 PRF is biologically significant, then -1 PRF signals should be present in homologous mRNAs from other budding yeast species. To address this, the *BUB3*, *EST2*, *SPR6* and *TBF1* homologs were identified in *S. paradoxus*, *S. mikatae*, *S. bayanus*, *S. castellii*, *S. kudriavzevii* and *S. kluyveri*, and analyzed for potentially significant -1 PRF signals as previously described[23]. At first glance, these analyses reveal that no single -1 PRF signal is completely conserved among the budding yeasts (**Appendix 14**).



**Figure 30: MFE ‘landscape’ of *SPR6* and predicted PRF signal at position 279.**

The minimum free energy of the *SPR6* ORF decreases from approximately position 200 to 400. Though the structure at position 279 was studied, position 348 also appears significant.

However, closer analysis shows that strong candidate -1 PRF signals can be identified in the homologs of all of these genes, although not in every species. For example, as noted above, the *S. cerevisiae* *EST2* mRNA contains 5 potential -1 PRF signals. Similarly, the *S. paradoxus* homolog also contains 5 potential -1 PRF signals, although none share elements identical to *S. cerevisiae*. *S. mikatae* *EST2* appears to harbor two potential -1 PRF signals, *S. bayanus* has three, and *S. castellii* contains two. However, none were identified in the *S. kudriavzevii* *EST2* homolog, and we were not able to identify an Est2p homolog in *S. kluyveri*. Turning to *SPR6*, the *S. cerevisiae*

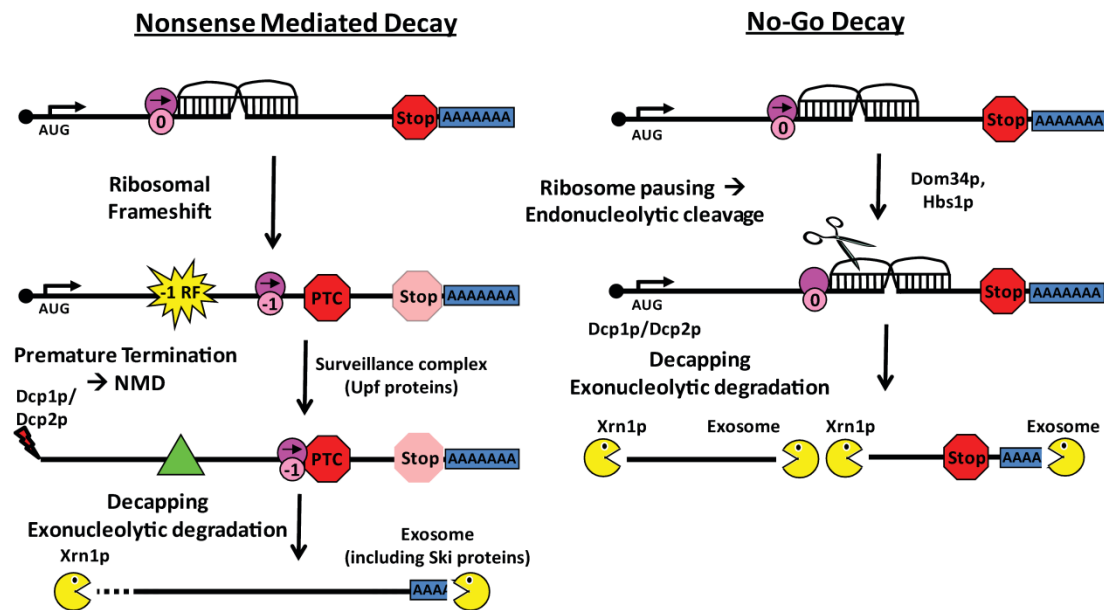
mRNA contains a second potential -1 PRF signal beginning at nucleotide 348 (**Figure 30**) in addition to that identified beginning at nucleotide 279.

Both the *S. paradoxus* and *S. kudriavzevii* SPR6 homologs contains three potential -1 PRF signals, but none were identified in the *S. mikatae* or *S. bayanus* homologs. In addition, BLAST analyses failed to identify SPR6 homologs in *S. castelli* or *S. kluyveri*. *S. cerevisiae* BUB3 contains the functional -1 PRF signal at nucleotide 858, plus potential -1 PRF signals beginning at nucleotides 27 and 732. The homologous mRNAs in *S. paradoxus*, *S. bayanus*, *S. castelli*, and *S. kudriavzevii* each appear to have one potential -1 PRF signal, but the none were identified in *S. mikatae* or *S. kluyveri*. Lastly, the *S. cerevisiae* TBF1 mRNA has only the single confirmed -1 PRF signal. The *S. mikatae* and *S. kluyveri* homologs appear to also have one, *S. kudriavzevii* contains two, but no -1 PRF signals were identified in either *S. paradoxus* or *S. bayanus*, while no Tbf1p homolog was identified in the *S. castelli* genome.

## Discussion

In a prior proof-of-principle experiment, we utilized the well characterized -1 PRF signal from the yeast L-A dsRNA virus to demonstrate that these elements can generally function as mRNA destabilizing elements through the NMD pathway[78]. Subsequently, a bioinformatics approach was used to determine that potential -1 PRF signals are widely found in all genomes examined, and that the great majority of these are predicted to direct elongating ribosomes to premature termination codons[23,172]. Here, we show that these chromosomally encoded, endogenous -1 PRF signals can also function as cis-acting mRNA destabilizing elements, both in the context of a reporter mRNA, and also in one case in a natural context. Further, we demonstrated that -1 PRF signals can differentially destabilize mRNAs through at least two pathways; NMD and NGD.

These are modeled in **Figure 31**. In the case of a ribosome shifting reading frame into a PTC, the surveillance complex lead by the Upf proteins signals rapid decapping by Dcp1p/Dcp2p, followed by deadenylation and exonucleolytic decay via Xrn1p and the exosome. In parallel, the NGD pathway can be activated by ribosomes that are stalled at strong secondary structures. are freed from mRNAs by Dom34p/Hbs1p, promoting exonucleolytic cleavage at unpaired nucleotides near the pause, thus resulting in two mRNA fragments which become substrates for decapping and exonucleolytic decay (reviewed in [179]). The findings presented here suggest that cells are not only well equipped to deal with abhorrent messages which contain premature termination codons and to clear stalled ribosomes from mRNAs, but have also evolved to capitalize upon these functions to post-transcriptionally regulate gene expression.



**Figure 31: -1 PRF signals can destabilize mRNA via NMD and NGD.**

**Left panel:** A -1 PRF event directs an elongating ribosome to encounter a premature termination codon (PTC). This leads to recruitment of the surveillance complex (Upf proteins), leading to mRNA decapping and 5' → 3' degradation by Xrn1p and deadenylation and 3' → 5' degradation by the degradasome. **Right panel:** the mRNA pseudoknot in a -1 PRF signals causes elongating ribosomes to pause, recruiting the Dom34p/Hbs1p complex, thus initiating No-Go Decay.

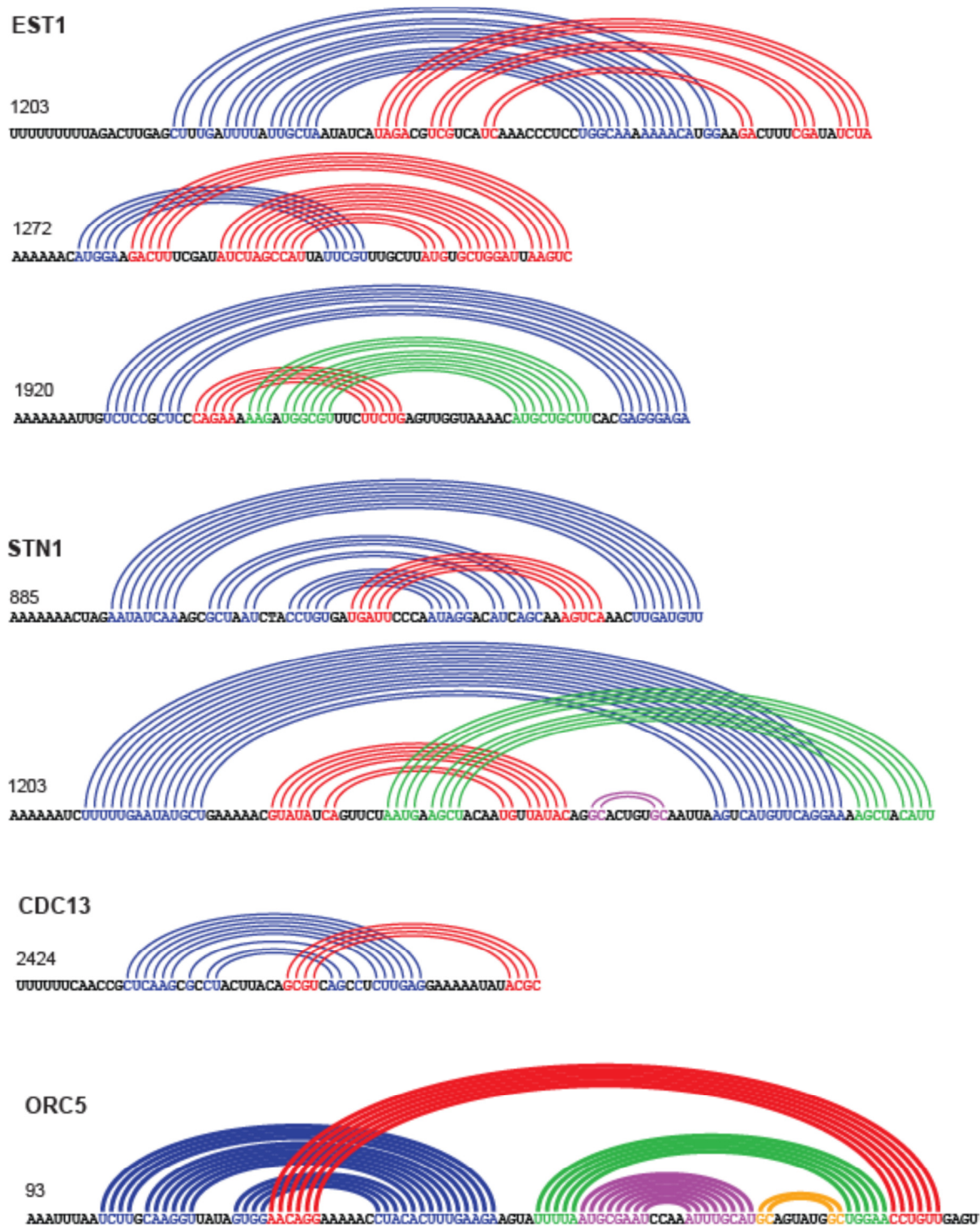
The strength of these signals to function as mRNA destabilizing elements should be equal to a combination of 1) their strengths as -1 PRF signals, and 2) their abilities to block ribosome progression, i.e. their thermodynamic stability. The *EST2* signal is both highly efficient at promoting -1 PRF (~55%, see [23], and is predicted to be quite stable (approximately -27 to -24 kcal/mol depending on the particular folding solution). It is important to note however that the software used to predict mRNA pseudoknots can neither identify base triples, which make major contributions to frameshifting [71,180–183], nor calculate their contributions to thermodynamic stability. Regardless, this combination of high frameshifting and thermodynamic stability results in very strong destabilization via both NMD and NGD (**Figure 25 C**). Interestingly, examination of **Figure 26 D** shows that the single exponential analysis of the *EST2* data yielded a *y*-axis intercept at ~80%. From **Figure 25 B** and **C**, it can be calculated that the contribution of NGD to total destabilization of the PGK1 reporter mRNA was approximately 30%. Taken together, these data suggest that the NGD component in the degradation of this message is very rapid, likely during the pioneer round of translation. In contrast, as discussed previously[78], the exponential decay profile shows that NMD can occur beyond the pioneer round.

In contrast to *EST2*, the *TBF1* signal promoted ~5% frameshifting, but is not predicted to be highly stable (-9.5 kcal/mol). Thus, all of its mRNA destabilization activity was through NMD (compare **Figure 25 B** with **C**). The thermodynamic stability of the *BUB3* signal is predicted to have an intermediate value to *EST2* and *TBF1* (approximately -12 kcal/mol), and hence the contribution of NGD to the stability of its reporter was significant. Interestingly, this signal only promoted ~1% frameshifting, yet the contribution of NMD to its destabilization was greater than observed for *TBF1*. One possible explanation for this apparent discrepancy may stem from the fact that, in order to measure frameshifting, one base had to be deleted from the spacer region between

the slippery site and the stimulatory pseudoknot. Changes in the length and composition of this spacer are known to affect rates of -1 PRF[184], and thus the -1 PRF values so determined cannot be taken as absolute. In contrast, the reporters used to monitor mRNA stability contained the native sequences. In light of this, it is likely that the native *BUB3* -1 PRF signal promotes more frameshifting than the *TBF1* signal. Lastly, the *SPR6* -1 PRF signal is predicted to be quite stable (approximately -20 kcal/mol), yet promoted very low levels of frameshifting (~0.5%). Accordingly, destabilization via NMD was negligible for this element, while NGD was the major contributor.

Beyond the *pro forma* demonstration that -1 PRF signals can destabilize cellular mRNAs, it is important to begin to understand the biological function of this phenomenon. As a first step in this direction, we showed that silently mutating the slippery sites in 5 predicted -1 PRF signals within a full-length clone of *EST2* significantly stabilized its encoded mRNA (Figure 28). Similarly, abrogation of NMD stabilized this message. Est2p is the reverse transcriptase subunit of the telomerase holoenzyme[173]. Interestingly, prior studies have demonstrated that this mRNA, along with other mRNAs encoding proteins having telomere-associated functions, are stabilized in NMD<sup>-</sup> yeast cells [85,177]. Analysis of the Programmed Ribosomal Frameshift Database (<http://prfdb.umd.edu/>) reveals that, along with the other 4 putative -1 PRF signals in the *EST2* mRNA, the mRNAs encoding Est1p, Stn1p, Cdc13p, and Orc5p, all components or regulators of telomerase that are stabilized in NMD<sup>-</sup> cells, also contain high confidence -1 PRF signals (**Figure 32**). In addition, the *EST3* mRNA contains a +1 PRF





**Figure 32: Computationally predicted -1 RF signals in *EST1*, *STN1*, *CDC13*, and *ORC5*.**

*EST1* contains three predicted -1 RF signals beginning at nucleotides 1203/1206 (overlapping UUU UUU UUA), 1272, and 1920. *STN1* contains two beginning at nucleotides 885 and 1203. *CDC13* contains one beginning at nucleotide 2424. *ORC5* has one beginning at nucleotide 93. Colored arcs depict base pairing of stems.

signal[185]. Intriguingly, telomerase is limiting in cells: while a yeast cell contains 64 chromosome ends, there are only ~29 telomerase molecules per cell and that telomerase is preferentially recruited to short telomeres[186]. Additionally, Tbf1p is a telobox containing general regulatory factor that binds to TTAGGG repeats within subtelomeric anti-silencing regions[187]. Intriguingly, ablation of NMD [188] or overexpression of single components of telomerase-associated proteins, i.e. the *TEL1* RNA, Est2p, Stn1p, or Cdc13p results in changes in telomere length [186,189,190]. We hypothesize that yeast cells use -1 PRF to limit the expression of these proteins in order to maintain the correct stoichiometric balance among telomere associated components. Corollary to this, mutations that alter -1 PRF and/or NMD should affect telomere function, and should thus show phenotypic defects similar to those observed in telomerase mutants, e.g. cell cycle progression defects. Indeed, we have isolated numerous such mutants(reviewed in [191]), and have reported that the *mof2-1* and *mof5-1* mutants, which affect both NMD and -1 PRF tend to accumulate large mother-daughter cells, and/or multiply budded cells, typical of G2/M cell cycle defects[192]. Similarly, *upf1Δ* cells have abnormally elongated buds, and decreased telomere lengths [193,194]. Intriguingly, *mof6-1* mutants, which only affect -1 PRF, arrest as large, unbudded cells, typical of M-phase exit defects[192]. These observations suggest that stabilization of the mRNAs encoding multiple telomere-associated proteins may have dominant negative effects on telomere homeostasis, and that NMD and -1 PRF may regulate different aspects of the cell cycle. Additionally, the central role of Bub3p at the mitotic cell cycle spindle assembly checkpoint and the progeroid phenotypes caused by Bub3p deficiency suggest a more general role for -1 PRF in control of cell growth and division. Finally, the expression of Spr6p during sporulation[195] suggests a role for -1 PRF in this developmental process as well. Future studies will dissect the roles of the -1 PRF signals in these mRNAs.

Finally, if -1 PRF is widely used to regulate gene expression, then it should be well conserved. The major problem associated with attempting a phylogenetic analysis of -1 PRF signals is the inherent limitations of the software used to predict them. In short, it is not well enough developed to automatically identify matching motifs. In an attempt to begin to address this issue, BLAST alignments were used to identify the homologous *BUB3*, *EST2*, *SPR6* and *TBF1* ORFs in six closely related yeast species, their nucleotide sequences extracted, and analyzed for the presence of potential -1 PRF signals. These analyses revealed that while specific -1 PRF signals do not appear to be evolutionarily conserved, -1 PRF itself may be relatively well-enough conserved as a mechanism to post-transcriptionally regulate the expression of these genes across many but not all species examined (See **Appendix 11** and **Appendix 14**). However, given the large degree of divergence among the budding yeast sequences so analyzed, if -1 PRF is so conserved, its usage in specific mRNAs would appear to be rapidly evolving.

## **Materials and Methods**

### **Strains, Genetic Manipulations, and Media.**

*Escherichia coli* DH5 $\alpha$  was used to amplify plasmid DNA. Transformations of *E. coli* were performed as described previously using the calcium chloride method[196]. Yeast cells were transformed using the alkali cation method[197]. Yeast strains used in this study are shown in **Appendix 1**. Yeast were grown on YPAD and synthetic complete media (H-)[198]. yRP2056, yRP2077 were kind gifts from R. Parker. YJB2659 (generously provided by Judith Berman) was sporulated and strains JD1276, JD1281, JD1287 and JD1288 were obtained by tetrad dissection.

### **Generation of mRNA Stability Vectors.**

Dual luciferase and mRNA stability plasmids have been previously described[23]. Oligonucleotide primers were purchased from IDT (Coralville, IA) and are shown in **Appendix 4**, the plasmids created are in **Appendix 2**. Computationally identified putative -1 PRF signals were amplified from yeast genomic DNA using PCR using Oligonucleotide primers which terminated in a *Sa*I restriction site at the 5' and *Bam* HI at the 3'. The zero-frame dual-luciferase reporter plasmid (pJD375) along with the PRF signal containing dsDNA fragments were digested using these restriction enzymes and ligated together to generate endogenous PRF signal containing dual-luciferase vectors. Oligonucleotide primers were chosen to terminate in *Kpn* I restriction sites and amplify 41 and 30 bases of *Renilla* and firefly luciferase derived sequences respectively. The resulting amplicons were cloned into the *Kpn* I site 492 bases into the *PGK1* open reading frame of the unmodified *PGK1* containing vector (pJD741). A premature termination codon vector (pJD828) was generated by cutting the readthrough (pJD753) with *Bam* HI and backfilling with Klenow fragment.

### **Generation of *EST2* Open Reading Frame Mutants.**

Full length *EST2* in a centromeric plasmid and the diploid *S. cerevisiae* *EST2* deletion strain were generously provided by the Berman lab and have been previously described[86]. Individual mutant strains were obtained by tetrad dissection. Five potentially significant -1 PRF signals were identified in the *EST2* open reading frame using the Predicted Ribosomal Frameshift Database[172]. The wobble bases of 5 slippery heptamers were synonymously mutagenized by oligonucleotide site-directed mutagenesis using the QuickChange II XL Site-Directed Mutagenesis Kit (Stratagene). Oligonucleotide design and reaction conditions were performed as recommended by the manufacturer with minor modifications. All mutations were confirmed by sequencing, the oligonucleotides for sequencing are in **Appendix 6**.

### **Steady State and Time Course RNA Blot Analyses.**

mRNA stability vectors were transformed into wild-type yeast (JD1158), *upf1*Δ or *upf2* Δ (JD1181 or JD1367), *xrn1*Δ (JD1170), *dcp1*Δ (JD1122), *ski2*Δ (JD1345), *ski3*Δ (JD19), and *dom34*Δ (JD1363) cells. The EST2 mRNA stability vector (pJD754) was transformed into *rpb1-1* (JD977) and *rpb1-1/Upf*<sup>-</sup> (JD978) cells and time courses were performed as described previously[199]. Total RNA was extracted with acid phenol/chloroform (pH = 4.5) from mid-logarithmic cell cultures [200], or with Trizole<sup>®</sup> Reagent following the manufacturer's directions (Invitrogen, Carlsbad, CA). RNA (northern) blotting was performed as previously described[78]. Equal amounts of RNA (1 μg, 2 μg, or 4μg) were separated through 1% agarose-formaldehyde gels. RNA samples were transferred and UV cross linked to Hybond-N-membranes (Amersham). Blots were hybridized with γ[<sup>32</sup>P] 5'-end-labeled oligonucleotides specific for *U3* snoRNA (loading control) and the exogenous *Renilla* fragment (experimental); these oligonucleotides are displayed in **Appendix 7**. Messenger RNAs were identified using a GeneStorm phosphoimager (Bio-Rad) and quantified using QuantifyOne (Bio-Rad). Blots were repeated three or more times and averaged to generate graphs.

### **Quantitative Real Time Reverse Transcription PCR.**

Full length EST2 expression vectors (pJD641), EST2 mutant vectors (pJD796), and null plasmids (pJD315) were transformed into WT (JD1281), EST2 deletion (JD1287), UPF2 (JD1288) and EST2/UPF2 (JD1276) deletion strains. Total RNA was extracted with acid phenol/chloroform (pH=4.5) from mid-logarithmic cell cultures. In parallel, total RNAs were extracted from isogenic *rpl3*Δ strains expressing wild-type RPL3 (JD1228), the down-frameshifting *rpl3*-R247A allele (AM-L3R247A), or the up-frameshifting *rpl3*-W255C/ P257S allele (JD1229). To prevent amplification from contaminating cellular DNA, RNA was treated with DNase I before reverse transcription using Turbo DNase (Ambion). cDNA was generated using the Bio-Rad iScript cDNA

synthesis kit and used in the LightCycler real-time PCR system. PCR reactions were performed with 2 µl of cDNA in 20 µl reactions containing ~10 nM each sense and antisense primer, and 1x LightCycler 480 SYBR Green I Master Mix (Roche). PCR cycles were run as follows: 1 cycle of 95° for 10 min; 40 cycles of 95° for 10 s, 54° for 20 s and 72° for 20 s. U3 snoRNA was chosen as a reference gene.

#### **Comparative Analyses.**

The Spr6p, Est2p, Bub3p, and TBF1p peptide sequences for were extracted from the *S. cerevisiae* genome and local BLAST alignments were performed to identify the homologous peptides in the genomes of *S. paradoxus*, *S. mikatae*, *S. bayanus*, *S. castellii*, *S. kudriavzevii*, and *S. kluyveri*. Homologs were identified in all cases except for Spr6p in the *S. castellii* and *S. kluyveri* genomes, for Est2p in the the *S. kluyveri* genome, or for Tbf1p in the *S. castellii* genome. For the remaining cases, the nucleotide sequences were analyzed for the presence of potential -1 PRF signals as previously described[23,172].

## Chapter 4

### **Ribosomal Frameshifting in CCR5: Regulation by NMD, miRNAs and Conformational Plasticity.**

#### **Introduction**

Gene expression is a metabolic process: rates of mRNA transcription delineate the anabolic phase, while rates of mRNA degradation define the catabolic parameters. While much attention has been paid to transcription, the impact of mRNA stability on this equation is tremendously important. Many mRNAs are post-transcriptionally regulated, and this field has mainly concentrated on identifying *cis*-acting elements in the 5' and 3' untranslated regions (UTRs) of mRNAs, and the *trans*-acting factors with which they interact. Less emphasis has been placed on identifying those regulatory elements located within protein coding regions. A variety of *cis*-acting mRNA elements have been identified that cause elongating ribosomes to “recode” the primary information contained in mRNAs[201].

Programmed ribosomal frameshifting is one such translational recoding mechanism historically associated with viruses and retrotransposons. A PRF signal stochastically redirects translating ribosomes into a new reading frame (*i.e.* by +1 or -1 nucleotide). In the viral context, a PRF signal allows ribosomes to bypass the 0-frame encoded stop codon and continue synthesis of a C-terminally extended fusion protein. PRF is likely employed by organisms representing every branch in the tree of life, suggesting an ancient and possibly universal mechanism for controlling the expression of actively translated mRNAs [17]. Computational analyses revealed that approximately

10-15% of genes in the 20+ eukaryotic genomes analyzed thus far contain at least one potential -1 PRF signal [172]. A key observation was that the outcome and function of -1 PRF differs significantly between viruses and eukaryotic cellular mRNAs. More than 95% of 'cellular' -1 PRF signals are predicted to direct elongating ribosomes to encounter premature termination codons (PTC), suggesting that -1 PRF may be used by cells to control mRNA abundance and stability through the nonsense-mediated mRNA decay (NMD) pathway. While this hypothesis has been demonstrated in yeast using -1 PRF signals of both viral and cellular origin[78,202], it has not yet been tested in higher eukaryotes. Furthermore, if -1 PRF is used to control expression of cellular genes, it should be subject to regulation: the issue of how sequence-specific regulation of -1 PRF may be achieved has been a central unanswered question in this field.

This study addresses these questions using a functional -1 PRF signal identified in the CCR5 mRNA, which encodes the co-receptor for HIV-1[203]. We demonstrate that it is able promote efficient levels of -1 PRF, and that it can function as an mRNA destabilizing element which is dependent on both NMD and efficient -1 PRF. CCR5-mediated -1 PRF is specifically stimulated by at least two human microRNAs (miRNA), hsa-miR-1224-3p (MI0003764) and hsa-miR-141(MI0000457). miR-1224 specifically interacts with the CCR5 -1 PRF signal *in vitro* and *in vivo*, suggesting a mechanism through which sequence-specific regulation of -1 PRF may be effected. Chemical protection, native polyacrylamide gel electrophoresis (PAGE), and single molecule optical trap experiments suggest that the -1 PRF mRNA stimulatory element can assume multiple dynamic structures, suggesting inherent structural plasticity as the biophysical basis for regulation of -1 PRF. Importantly, hsa-miR-1224 does not change the chemical protection pattern, nor does it result in the presence of new species resolvable by PAGE. Rather, the PAGE analyses show that hsa-miR-1224 alters the



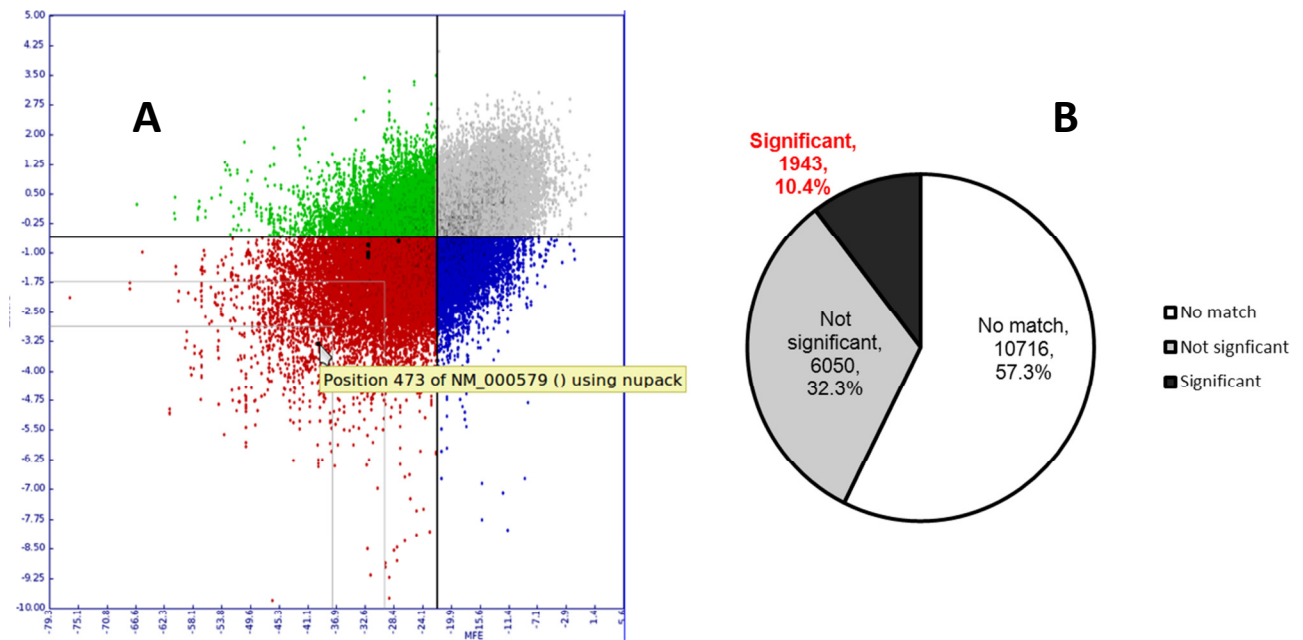
distribution of the existing CCR5 -1 PRF signal containing RNA conformers, consistent with the optical trap experiments.

Operational -1 PRF signals were also identified in the mRNAs encoding the interleukin 7  $\alpha$ -chain receptor subunits in both humans and in mice, in the mRNAs encoding the human IL8 receptor  $\alpha$ - and  $\beta$ -chains, and IL27 receptor  $\alpha$ -chain. These findings increase our understanding of cytokine receptor expression, how the immune response may be controlled at the local level (i.e. by individual cells), and presents a potentially new approach to the control of HIV/AIDS and the immune response.

## Results

### A functional -1 PRF signal in the *H. sapiens* CCR5 receptor mRNA.

A computational analysis of 18,709 human coding sequences (CDS) revealed that 1943 (~10.4%) contain at least one high confidence -1 PRF signal (**Figure 33 B**). The distribution of *Homo sapiens* signals is shown in **Figure 33 A**. High confidence -1 PRF signals are defined as those having acceptable slippery sites followed closely by predicted mRNA pseudoknot structures in which the computed minimum free energies (MFE) with respect to randomized sequences result in probabilistic z-scores greater than one standard deviation from the population of all sequences. These criteria revealed a strong candidate -1 PRF signal beginning at nucleotide 473 in the human CCR5 mRNA (NM\_000579) that begins with the classic U UUA AAA slippery site. This sequence is 100% conserved in the *Pan troglodytes* CCR5 mRNA beginning at nucleotide 408. Computational analyses predicted two nearly equivalent, highly stable potential



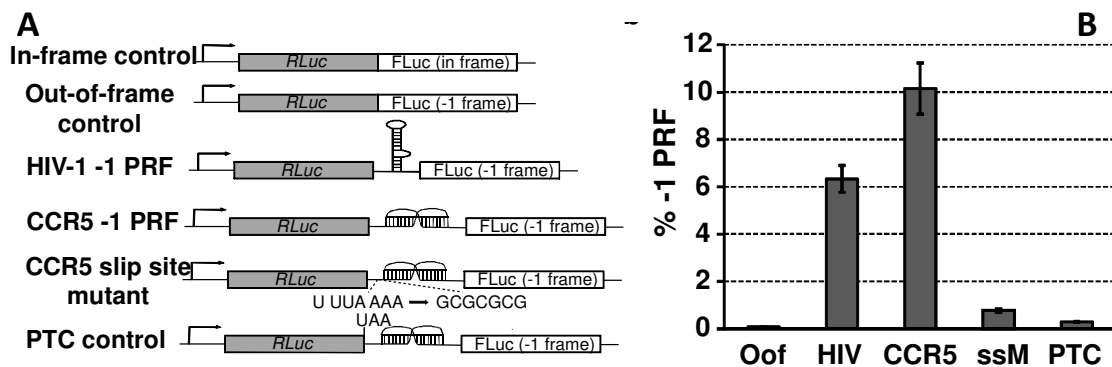
**Figure 33: Distributions of potential -1 PRF signals from *Homo sapiens*.**

**A.** All potential -1 PRF signals from *Homo sapiens* are plotted as the z-score with respect to minimum free energy. The z-score is defined as the number of standard deviations between the predicted minimum free energy and the mean predicted free energy of the same sequence shuffled 100 times. Using these criteria, a significant PRF signal is one which is predicted to form a pseudoknot, has a minimum free energy less than one standard deviation from the mean of the population of *Homo sapiens* predictions, and a z-score less than one standard deviation from the mean of the population of *Homo sapiens* z-scores. **B.** Pie chart showing number and fraction of human genes without predicted -1 PRF signals (no match), with potential, but non-significant -1 PRF signals, and with predicted significant -1 PRF signals.

downstream mRNA pseudoknot structures, two less stable pseudoknots, and a stable stem-loop. A dual-luciferase reporter system[204] was used to test the ability of this sequence to promote -1 PRF in cultured mammalian cells (**Figure 34**).

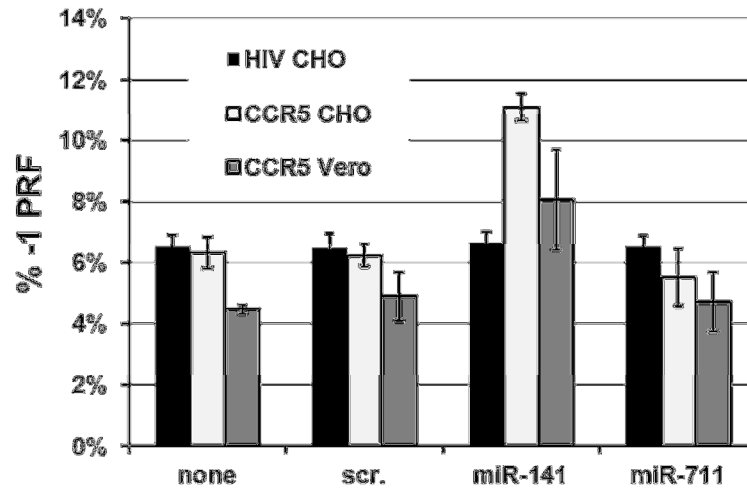
The putative human CCR5 -1 PRF sequence promoted 8-10% -1 PRF in HeLa (**Figure 34 B**) using an *in vivo* assay, 4-7% in CHO or Vero cells (**Figure 35** see "none"), and 20-25% using an *in vitro* assay (**Figure 36**). Mutagenesis of the slippery site from TTT AAA to G CGC GCG (ssM) reduced -1 PRF to 1-4%. This degree of -1 PRF in the mutant was unexpectedly high, and suggested the possibility that an mRNA splicing donor site or internal ribosome entry signal (IRES) may have been introduced into the reporter with the CCR5 sequence. To test this, an in-frame termination codon was

introduced into the reporter immediately 3' of the *Renilla* luciferase open reading frame, but 5' of the CCR5-derived sequence (PTC control). This mutation reduced apparent -1 PRF levels by more than two orders of magnitude. Similarly, placing the firefly luciferase reporter out of frame (Oof) with respect to *Renilla* without any intervening sequences resulted in even lower levels of frameshifting (**Figure 34B**, Oof). A rabbit reticulocyte based *in vitro* translation assay independently observed that the CCR5 derived sequence promoted synthesis of a peptide consistent with a -1 PRF event, and at levels comparable to that promoted by the HIV-1 PRF signal(**Figure 36**). While they do not directly demonstrate -1 PRF, these two independent sets of experiments are consistent with the hypothesis that the computationally identified sequence in the CCR5 mRNA promotes efficient -1 PRF.



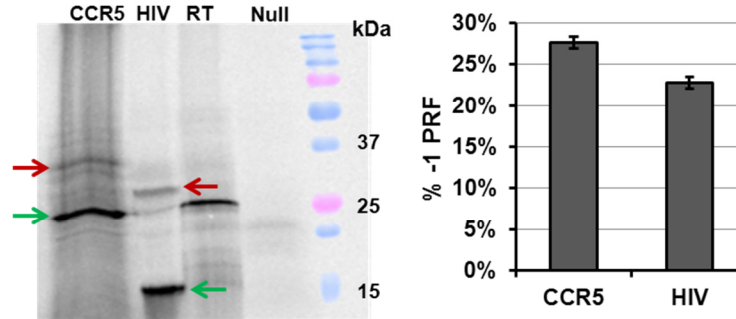
**Figure 34: A *cis*-acting element in the CCR5 mRNA promotes efficient -1 PRF.**

**A.** Schematic of dual luciferase constructs used to test the CCR5 frameshifting signal. Transcription is driven from the SV40 early enhancer/promoter and transcription termination and polyadenylation utilizes the SV40 late poly(A) signal. The in-frame control is p2luci[204], encoding a firefly/*Renilla* luciferase fusion protein. In the out of frame reporter (Oof), firefly luciferase lies in the -1 reading frame with respect to the *Renilla* open reading frame. In the HIV -1 PRF reporter, the -1 PRF signal of HIV-1 was cloned in between the two luciferase reporters, and the firefly ORF is in the -1 frame with respect to *Renilla*. The CCR5 -1 PRF reporter is the same, except that it contains the CCR5 -1 PRF signal. In the CCR5 slip site mutant (ssM), the UUUAAAA slippery heptamer of the CCR5 -1 PRF signal was mutated to GCGCGCG. The PTC reporter is based on the CCR5 -1 PRF reporter in which a premature termination codon was inserted following the *Renilla* open reading frame. **B.** Measurements of -1 PRF in HeLa cells using these constructs. Error bars approximate standard error[205].



**Figure 35: The CCR5 -1 PRF signal is active in CHO and Vero cells.**

Efficient -1 PRF is also promoted by the CCR5 sequence in CHO and Vero cells. Furthermore, -1 PRF is stimulated by hsa-miR-141 in these cell types.

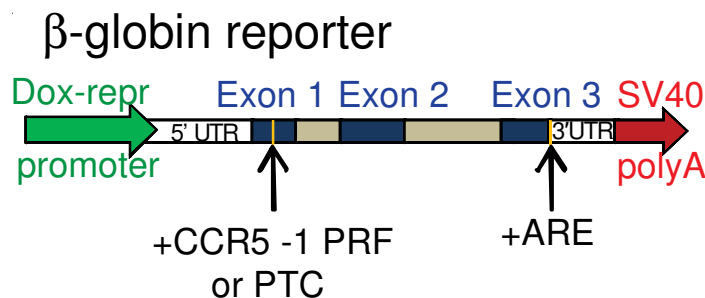


**Figure 36: Rabbit reticulocyte lysates confirm CCR5 frameshifting.**

Efficient -1 PRF promoted by the CCR5 sequence *in vitro*. **Left:** autoradiogram of *in vitro* translation reaction using synthetic mRNAs harboring CCR5 or HIV-1 derived -1 PRF signals. Green arrows denote 0-frame encoded products. Red arrows denote -1 PRF encoded peptides. RT indicates the readthrough control. **Right:** Percent -1 PRF promoted by CCR5 and HIV-1 frameshift signals *in vitro*. This experiment was designed and performed by Sharmishtha Musalgaonkar.

### The CCR5 -1 PRF Signal Destabilizes a Reporter mRNA through the NMD Pathway in Mammalian Cells.

A -1 PRF event at the slippery site is predicted to direct elongating ribosomes to a premature termination codon (PTC) 45 codons after the beginning of the slippery site, suggesting that this may function as an mRNA destabilizing element through the NMD pathway as described in *Saccharomyces cerevisiae*[202]. To test this hypothesis, the human CCR5 -1 PRF signal was cloned into exon 1 of a rabbit  $\beta$ -globin reporter construct (**Figure 37**), and its effects on mRNA stability were assayed in HeLa cells. A reporter containing a TNF $\alpha$ -derived AU-rich element (ARE) cloned into the 3' UTR was employed to independently monitor mRNA destabilization by AUF1[206], this construct was mutagenized to contain both the CCR5 -1 PRF signal and ARE, and a fourth contained an in-frame PTC at the same site as the CCR5 -1 PRF signal insertion. All

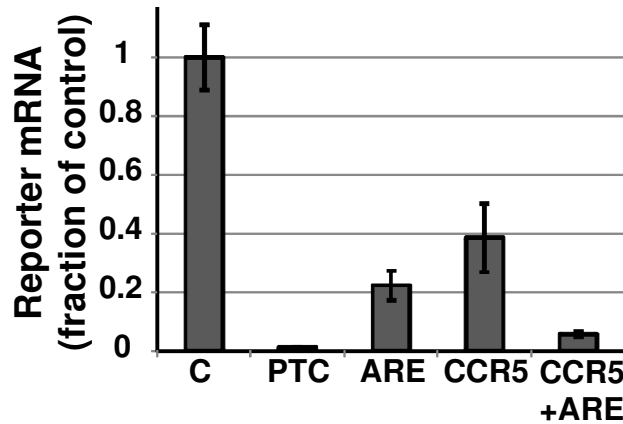


**Figure 37: The rabbit  $\beta$ -globin mRNA stability reporter.**

A rabbit  $\beta$ -globin reporter containing a doxycycline repressible promoter and SV40 derived polyA signal is shown. The native CCR5 -1 PRF signal was cloned into exon 1. Controls included insertion of a PTC at this position, or insertion of the 27 nucleotide TNF $\alpha$  derived A-U rich element (ARE) immediately following the rabbit  $\beta$ -globin open reading frame[207].

cells were co-transfected with the dual-luciferase readthrough control plasmid to control for transfection efficiency variation. The steady-state mRNA abundance of the CCR5 -1 PRF containing reporter was ~38% of the control  $\beta$ -globin reporter sample lacking the

frameshift signal (**Figure 38**). Similarly, the TNF $\alpha$ -derived ARE reduced the steady-state abundance of the reporter to ~22% of the readthrough control. In combination, the

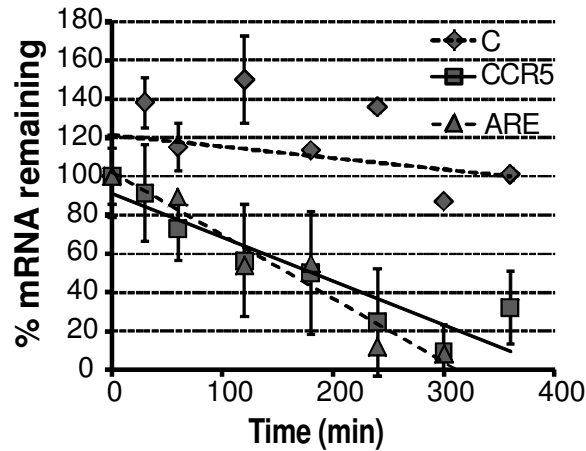


**Figure 38: The CCR5 -1 PRF signal acts as an mRNA destabilizing element.**

Quantitative reverse-transcriptase PCR (qRT-PCR) analysis of rabbit  $\beta$ -globin steady-state abundance in HeLa cells co-transfected with reporters shown in Figure 37, and with the readthrough dual-luciferase reporter; reported as fold of the control, native  $\beta$ -globin reporter (denoted as C). The readthrough dual-luciferase reporter mRNA was used to control for differences in transfection efficiencies. Error bars denote standard error.

CCR5 -1 PRF signal and ARE decreased  $\beta$ -globin reporter mRNA steady state abundance to ~6% of the control, suggesting that the two elements promote mRNA destabilization by two independent pathways. As expected, the presence of an in-frame PTC strongly decreased reporter mRNA abundance (~1% of readthrough).

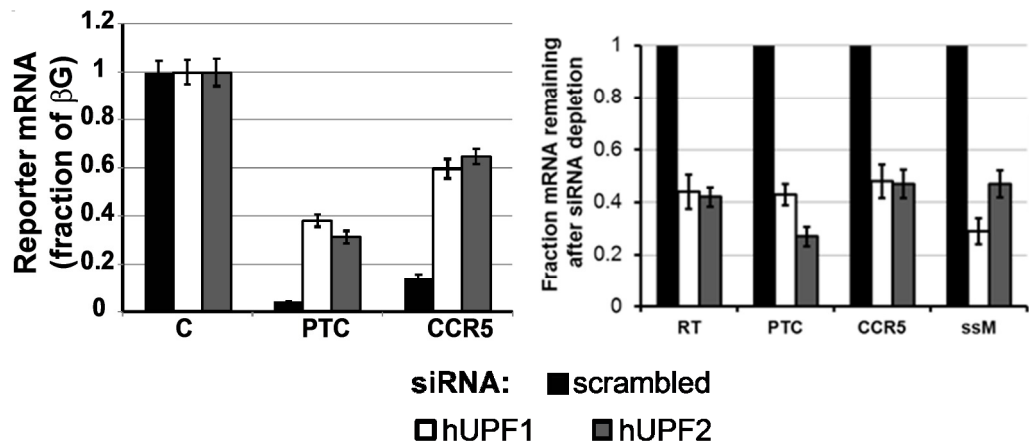
A transcriptional arrest time course experiment employing a Tet-repressible system was performed[206,207] to determine whether introduction of the CCR5 -1 PRF signal rendered the  $\beta$ -globin reporter a direct substrate for NMD. This experiment revealed that while the native  $\beta$ -globin mRNA was very stable, the CCR5 -1 PRF signal lowered the half-life of the reporter mRNA to 80-160 minutes (**Figure 39**). The ARE containing reporter also promoted rapid decay of its mRNA ( $t_{1/2}$  120-160 minutes). Time-course assays were attempted for the PTC containing reporter, but the data were



**Figure 39: Rabbit  $\beta$ -globin reporter half-life measurements.**

Time course measurements of rabbit  $\beta$ -globin reporter abundances transcriptionally arrested with doxycycline. Measurements of the PTC control were undetermined because its abundance was too low to determine after the first time point.

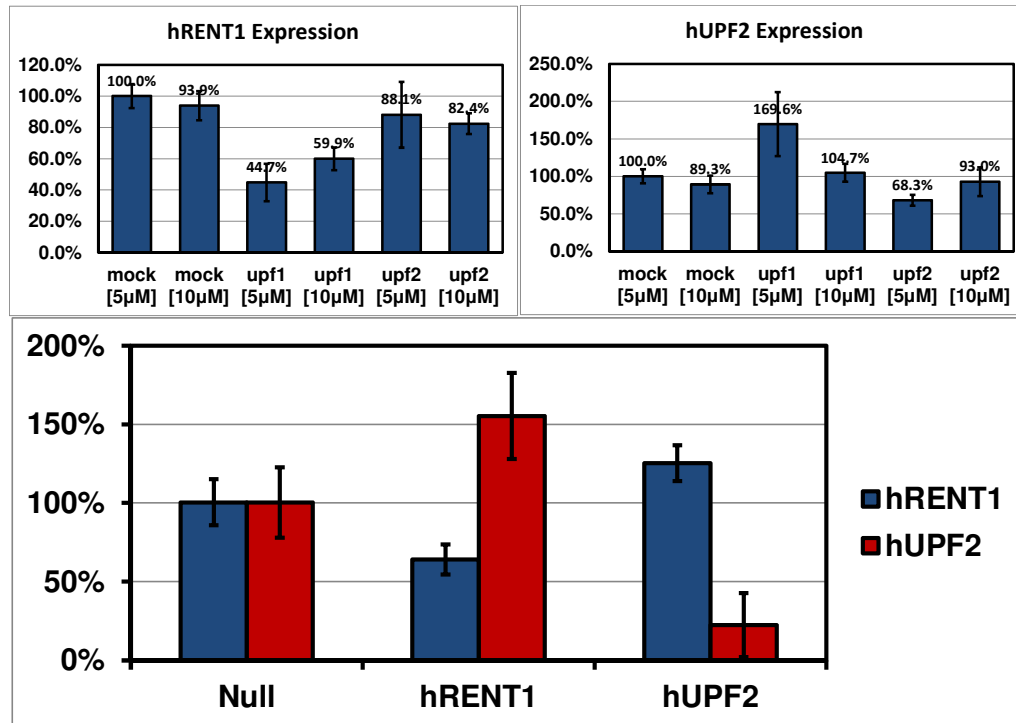
uninterpretable due to consistently low levels of expression. **Figure 40** shows that siRNA depletion of either hUPF1 (hRENT1) or hUPF2 resulted in ~4.4-fold increase in the abundance of the CCR5 -1 PRF signal containing reporter mRNA as compared to control cells transfected with a scrambled siRNA. In contrast, the in-frame PTC



**Figure 40: siRNA knockdown of NMD increases the amount of CCR5 reporter mRNA.**

siRNA transfection experiments for hRent1 (Upf1) and hUpf2. Rabbit  $\beta$ -globin abundance in cells transfected with scrambled, hRent1, or hUpf2 siRNAs. The differences between un-normalized RT abundances compared to one another in the different siRNA-treatments were insignificant. Error bars denote standard error. Knockdown efficiency of hRENT1 and hUPF2 mRNA are shown to the right.

containing reporter was increased ~8.5 fold. Figure 40B shows the siRNA depletion of hUPF1 and hUPF2 ranged from 30% - 50% in these experiments; interestingly, knockdown of either NMD factor significantly increased expression of the other (**Figure 41**), as has been recently described[208].



**Figure 41: NMD knockdowns increase other NMD mRNA abundance.**

Successful hUPF2 knockdowns were observed at [5 μM] siRNA but increase the amount of hRENT1 mRNA. Interestingly, hUPF2 knockdown did not significantly affect hRENT1 in this experiment, but in other trials increased expression by as much as 180% (lower panel). Other trials display a strong inverse relationship between hRENT1 and hUPF1 after knockdown and provide a useful benchmark to assay success.

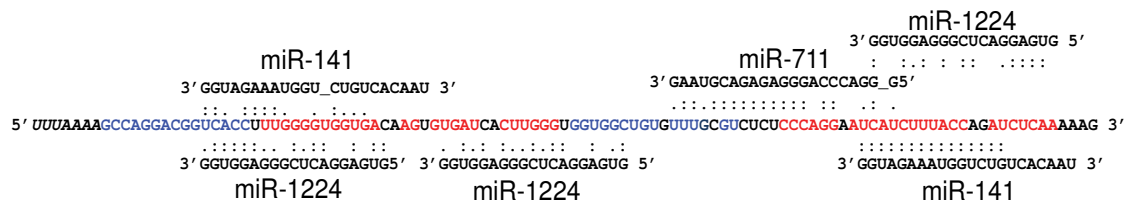
#### Specific stimulation of CCR5-mediated -1 PRF by miR-1224 in HeLa cells.

If -1 PRF is used by cells to post-transcriptionally control gene expression, then it is reasonable to hypothesize that this process may be regulated in a sequence-specific manner. One mode includes small ncRNAs that are capable of interacting with the frameshift promoting mRNA pseudoknot through base-pairing. This hypothesis dovetails observations that antisense oligonucleotides induce +1 programmed ribosomal frameshifting of eukaryotic antizyme genes[209], and observations that antisense

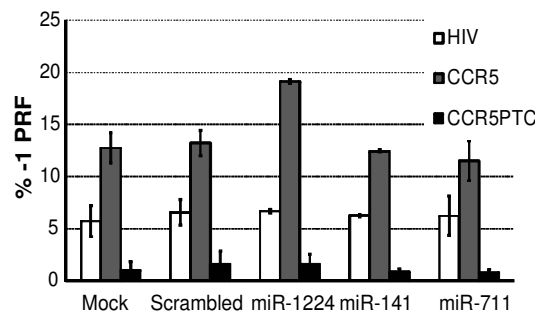


oligonucleotides increase rates of -1 frameshifting [210–212]. A search of the NCBI miRNA database revealed that hsa-miR-1224, hsa-miR-711, and hsa-miR-141 are potentially able to interact with the CCR5 -1 PRF signal in multiple locations (**Figure 42 A**). To test this, HeLa cells were transfected with constructs expressing each of these three miRNA precursors [5nM], or with a construct containing scrambled sequences. A

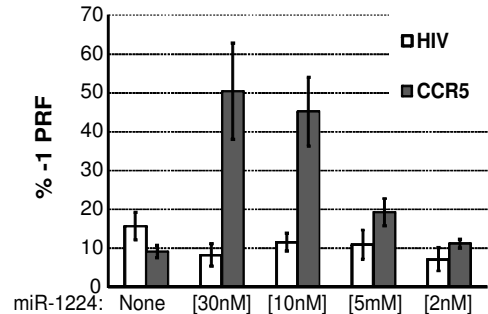
**a**



**b**



**c**



**Figure 42: Stimulation of CCR5-mediated -1 PRF by hsa-miR-1224.**

**a.** Sequence of the CRR5 -1 PRF signal is shown. The UUUAAAAA slippery site is italicized, stems 1 and 2 of the mRNA pseudoknot are colored blue and red respectively, and unpaired bases are black. Sequences of hsa-miR-1224-5p, hsa-miR-711, and hsa-miR-1413, and their predicted hybridization patterns with CCR5 sequence are indicated. **b.** HeLa cells were transfected with [5nM] of the indicated miRNA expressing constructs, or mock transfected. After 24 hours, cells were transfected with the indicated -1 PRF dual-luciferase reporters, and frameshift assays were performed 24-36 hours later. **c.** HeLa cells were transfected with indicated concentrations of hsa-miR-1224 miRNA expressing constructs, transfected 24 hours later with either HIV-1 or CCR5 -1 PRF dual-luciferase reporters, frameshift assays were performed 24-36 hours later. Error bars approximate standard error.

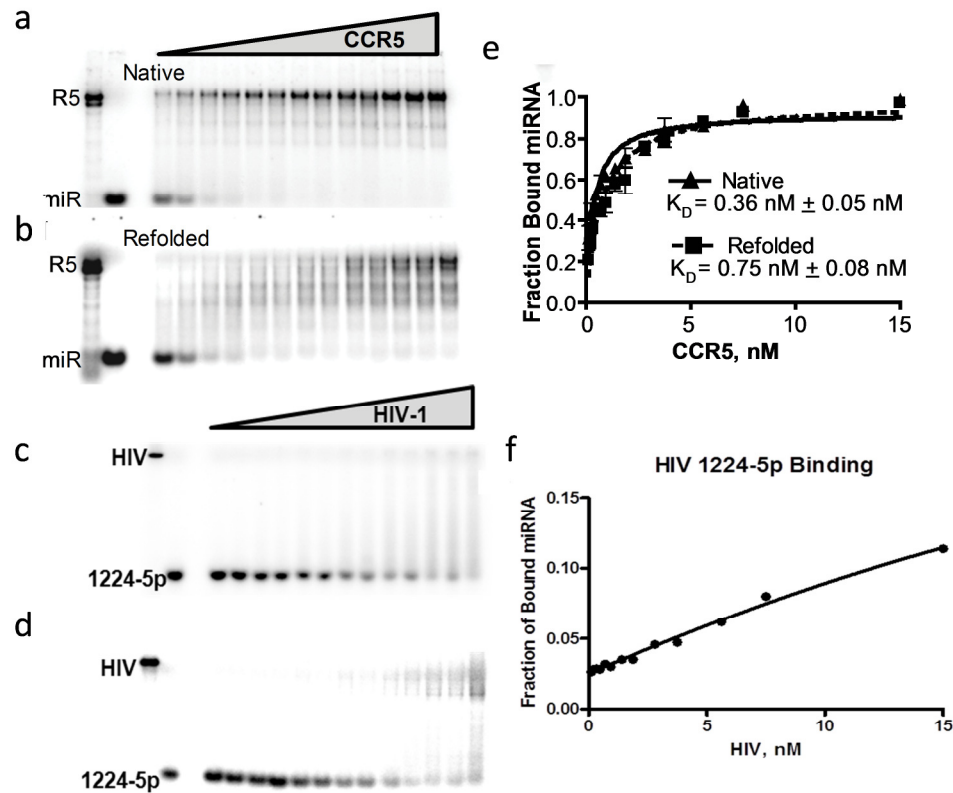
control set of cells were also mock transfected. Either immediately, or 24 hours post-miRNA transfection, cells were transfected with -1 PRF reporters containing the native CCR5 sequence, CCR5PTC, or HIV-1 frameshift signal, and frameshifting assays were performed after an additional 24 hour incubation. **Figure 42 B** shows that hsa-miR-

1224 promoted ~1.5-fold enhancement CCR5-mediated -1 PRF (~19.1% PRF compared to ~12.8% in mock transfected controls). HIV-1 frameshifting was affected by neither hsa-miR-1224, nor by the scrambled miRNA control. To determine whether this effect was dose-dependent, cells were first transfected with a hsa-miR-1224 precursor at [30nM], [10nM], [5nM], and [2nM], and then transfected with either the HIV-1 or CCR5 reporters. In this experiment, [30nM] and [10nM] of hsa-miR-1224 greatly enhanced CCR5-mediated -1 PRF while [2nM] promoted little to no effect (**Figure 42 C**). hsa-miR-1224 did not affect HIV-1 mediated -1 PRF. We note however that the higher doses of hsa-miR-1224 were toxic to cells, resulting in much lower baseline levels of the reporter proteins. This added a significant amount of noise into the system as reflected by the larger error bars. While neither hsa-miR-141 nor hsa-miR-711 affected -1 PRF in the HeLa cell based assay, the presence of endogenous hsa-miR-711, and/or hsa-miR-141 in these human derived cells may have masked the ability of these transfected miRNAs to affect -1 PRF. Transfection of these miRNAs at [5nM] into CHO and Vero cells revealed that hsa-miR-141 was also able to specifically stimulate CCR5-mediated -1 PRF (**Figure 35**).

#### **hsa-miR-1224 directly interacts with the CCR5 -1 PRF signal.**

Two different assay systems were employed to determine whether the effects of hsa-miR-1224 on CCR5-mediated -1 PRF were direct or indirect. In vitro gel shift experiments were performed using increasing concentrations of a 247 nucleotide T3-RNA polymerase generated transcript containing the CCR5 -1 PRF signal and a constant amount of  $\gamma$ [<sup>32</sup>P]-labeled synthetic hsa-miR-1224-5p. In one experiment, the two RNAs were mixed, incubated at 37° for 30 min, and then resolved through native PAGE (“native”). In parallel, the RNAs were mixed, heated to 90° and then allowed to slowly re-fold (“refolded”). Both experiments show that hsa-miR-1224-5p interacted with

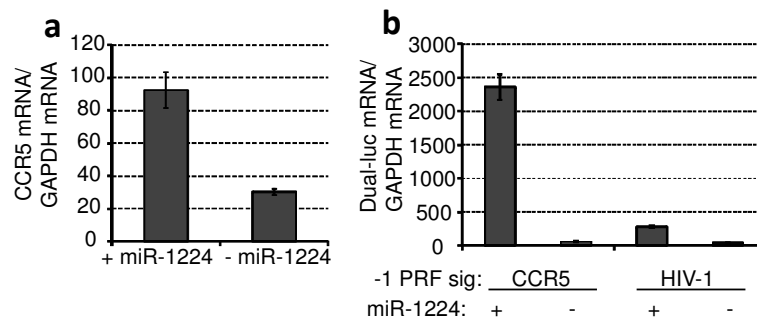
the CCR5 -1 PRF signal containing RNA with subnanomolar  $K_d$ 's (**Figure 43 a-c**). In contrast, hsa-miR-1224 did not interact with a 315 nt. transcript containing the HIV-1 PRF signal (**Figure 43 d-f**). The CCR5/miR-1224 gelshift experiments generated some additional interesting observations. First, hsa-miR-1224 interacted with the CCR5-derived sequence ~ 2-fold more strongly in the “native” as opposed to the “refolded” context. Second, addition of miR-1224 enhanced the appearance of three pre-existing conformers, particularly in the “refolded” context. The significance of these observations is discussed below.



**Figure 43: hsa-miR-1224 directly interacts with the CCR5 -1 PRF signal *in vitro*.**

Two-fold dilutions of a CCR5 -1 PRF signal containing transcript were mixed with equal volumes of 1.0 nM [ $^{32}$ P]-labeled synthetic hsa-miR-1224 RNA, and incubated for 30°C for 30 min (**a,c**, Native), or incubated at 90° for 5 sec, cooled quickly to 60° and then slowly to 37° (**b,d**, Refolded). Samples were separated through 10% native PAGE, dried, and radioactive activities were determined using a phosphorimager. **e,f**. Single site binding isotherms generated for CCR5 and HIV-1.  $K_D$  values and standard deviations are indicated. These experiments were performed by Dr. Meskauskas.

To ascertain whether hsa-miR-1224 interacts with the CCR5 -1 PRF signal *in vivo*, a biotin-labeled hsa-miR-1224 precursor was transfected into HeLa Tzm-BL cells expressing CCR5[213], cell lysates were passed through a streptavidin-agarose slurry, washed extensively, bound RNAs were eluted, and qPCR was used to quantify the amount of CCR5 mRNA relative to a GAPDH loading control[214]. These miRNA-mediated mRNA pulldown experiments demonstrated ~3-fold enrichment for CCR5 mRNA (**Figure 44 a**). Repetition of the experiment using lysates from HeLa cells transfected with the CCR5 PRF signal containing dual luciferase reporter plasmid revealed a >2000-fold enrichment for this mRNA as compared to no-miRNA controls (**Figure 44 b**). In contrast, enrichment of a dual-luciferase reporter harboring the HIV-1 PRF signal was approximately 10-fold above the no-miR-1224 control. These findings



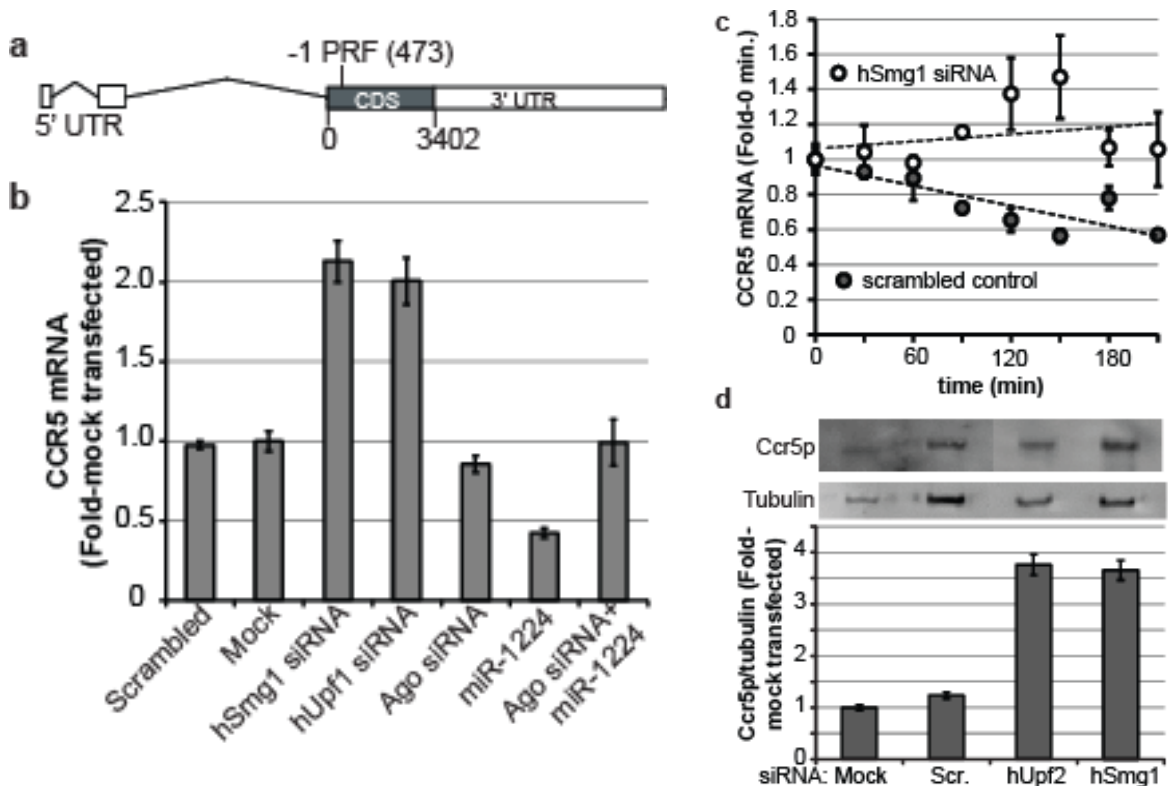
**Figure 44: hsa-miRNA-1224-5p interacts with the CCR5 mRNA *in vivo*.**

**a.** *In vivo* pulldown of native CCR5 mRNA in live cells. Biotinylated hsa-miR-1224 precursor (+ miR-1224) or a scrambled control (- miR-1224) were transfected into HeLa TZM BL cells expressing CCR5. Cell lysates were passed through a streptavidin slurry and bound mRNAs were eluted after washing. qPCR using CCR5 or GAPDH specific primer sets were used to quantitate the enrichment of CCR5 mRNA relative to GAPDH in the samples. **b.** HeLa cells were co-transfected with dual luciferase plasmids containing either the CCR5 or HIV-1 -1 PRF signal sequences, and with biotinylated miR-1224 (+) or scrambled precursor RNAs (-). Cell lysates were processed and analyzed as in panel a. Error bars approximate standard error.

directly demonstrate that hsa-miR-1224 specifically interacts with the CCR5 PRF signal in live cells, and not with other sequence elements within the CCR5 mRNA.

### The Native CCR5 mRNA is Affected by NMD and hsa-miR-1224.

The 5' UTR of the CCR5 gene contains two introns, while the CDS and 3' UTR are encoded by a single large exon (**Figure 45 a**). The CDS is 3402 base-pairs in length, and the -1 PRF signal begins at nucleotide 473. HeLa Tzm-BL cells expressing CCR5 were transfected with siRNAs targeting hUpf1, hSmg1, containing scrambled sequences ([5nM] each), or mock transfected. qPCR analyses of total RNA revealed



**Figure 45: Effects of NMD abrogation and miR-1224 on the native CCR5 mRNA.**

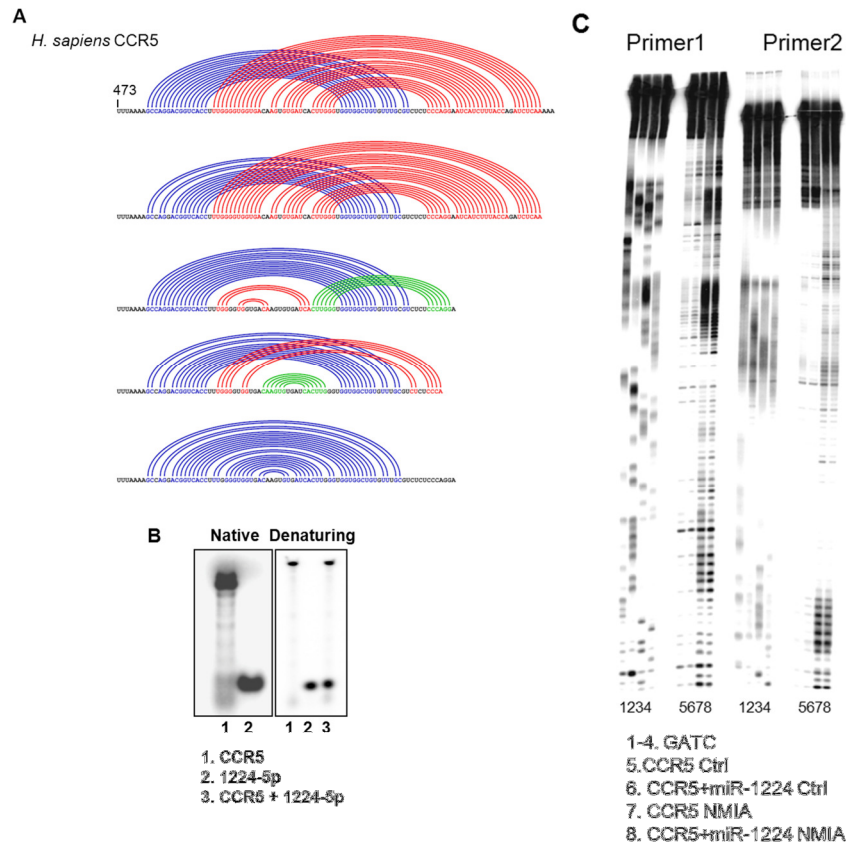
**a.** The full length CCR5 locus is diagrammed, showing two introns in the 5' UTR, the relative position of the -1 PRF signal at position 437, the termination codon at nt 3402, and the long 3' UTR. **b.** CCR5 mRNA steady-state abundance in Tzm-BL cells was monitored by qPCR. hSmg1, hUpf1, and argonaute (Ago) were partially knocked down by siRNA. miR-1224 indicates transfection with precursor expressing this miRNA. Scrambled denotes transfection with a scrambled siRNA precursor, and mock denotes mock transfected cells. **c.** Time course measurements of native CCR5 mRNA reporter abundances transcriptionally arrested with actinomycin D. Cells were transfected with Smg1 siRNA or scrambled siRNA control. **d.** Immunoblot analysis: HeLa Tzm-BL cells were either mock transfected, transfected with siRNAs targeting hUpf2, hSmg1, or scrambled RNA controls. Top panel: Immunoblots of total lysates were probed with anti-CCR5 or anti-tubulin monoclonal antibodies. Lower panel: Relative abundance of Ccr5p in each sample normalized to tubulin loading controls. Error bars approximate standard error.

that partial knockdown of hUpf1 (64% knockdown) or hSmg1 (48% knockdown) increased the steady-state abundance of the CCR5 mRNA ~2-fold each relative to controls (**Figure 45 b**), consistent with the previous reporter construct experiments and the hypothesis that the native CCR5 mRNA is a substrate for NMD. Transfection of cells with hsa-miR-1224 decreased CCR5 mRNA steady-state abundance to ~40% of controls, while co-transfection of cells with hsa-miR-1224 and siRNA directed against Argonaute (Ago) reversed this effect. Transfection with the Ago siRNA alone had no effect on CCR5 steady state abundance. These results are consistent with the notion that argonaute-mediated processing of hsa-miR-1224 is required for stimulation of -1 PRF, and that this stimulation increases the proportion of ribosomes directed to the -1 frame PTC. To determine whether the CCR5 mRNA is a direct substrate for NMD, cells were transfected with hSmg1 or scrambled siRNAs, transcription was arrested with actinomycin D, and mRNAs were harvested at 30 min timepoints. While the CCR5 mRNA was decreased to ~50% of initial levels after 3.5 hours in cells transfected with the scrambled siRNA control, hSmg1 siRNA transfection stabilized this mRNA. Immunoblot analysis showed that abrogation of NMD by transfection of Tzm-BL cells with hSmg1 or hUpf2 siRNAs resulted in a nearly 4-fold increase in Ccr5p compared to mock or scrambled siRNA controls (**Figure 45 d**). These experiments demonstrate that the native CCR5 mRNA is directly regulated by NMD, and that the extent of mRNA degradation is in turn inversely proportional to frameshifting efficiency.

**The CCR5 -1 PRF Signal Encodes a Complex and Dynamic set of mRNA Conformers.**

As previously noted, multiple folding solutions can be predicted for the sequence downstream of the slippery site. While the Stem 2 structures of the two most stable predicted pseudoknots are identical, the slippery-site proximal ends of the Stem 1 structures differ in their base-pairing solutions (**Figure 46 A**). Three additional mRNA

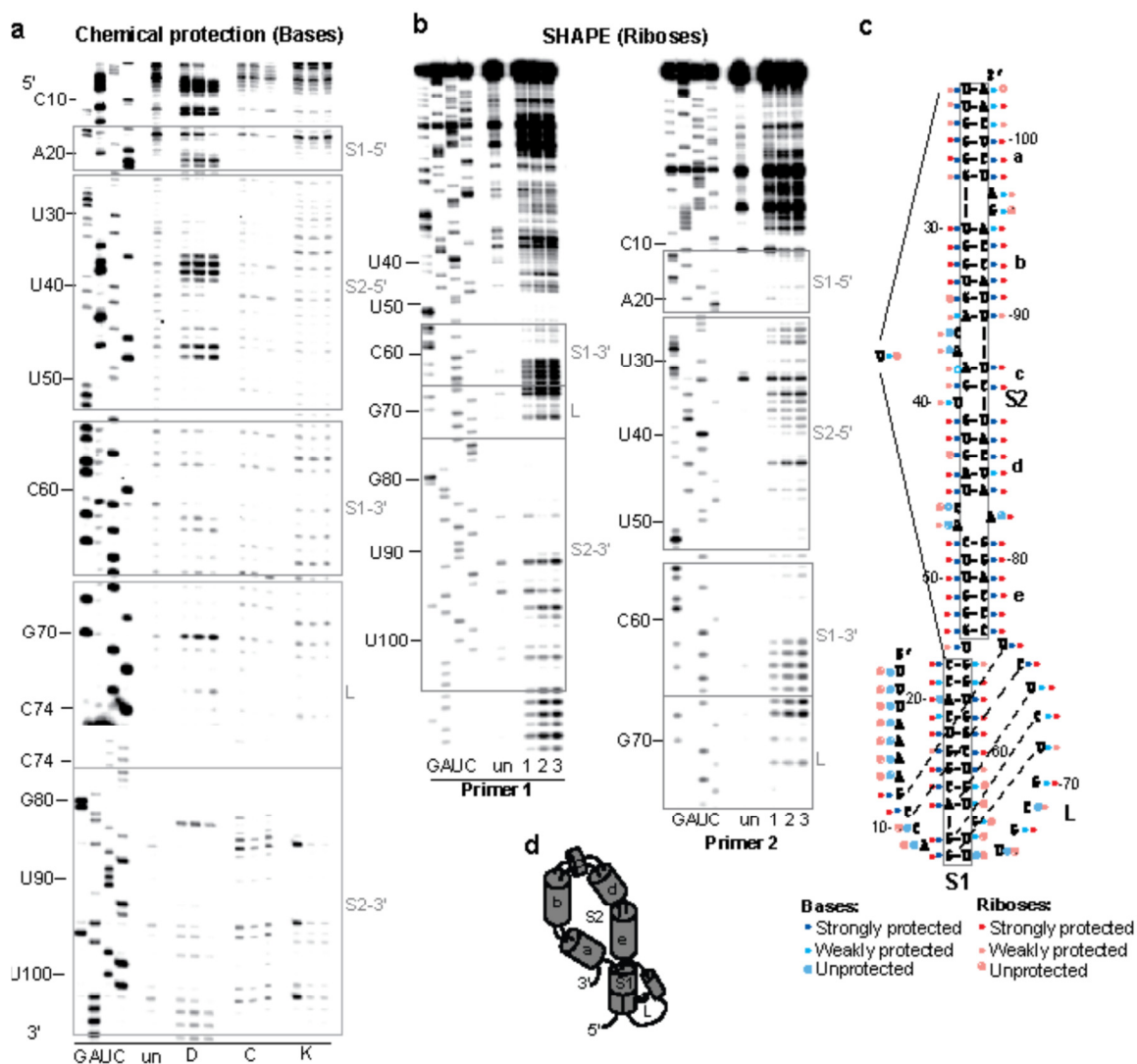
folding solutions for this sequence are also shown. While denaturing polyacrylamide gel electrophoresis (PAGE) of a  $\gamma[^{32}\text{P}]$ -3' end labeled transcript containing this element showed that it contained a single transcript of uniform size, native PAGE of the same transcription reaction revealed the presence of two major, and multiple minor bands, demonstrating that this sequence is able to fold into multiple conformers (**Figure 46 B**).



**Figure 46: Prediction and summary of the CCR5 PRF signal.**

**A.** Five computationally predicted folding solutions for the CCR5 -1 PRF signal as annotated in the PRFdb. The predicted stems 1, 2 and 3 are displayed as red, blue, and green respectively. **B.** Native and denaturing PAGE of a CCR5 -1 PRF signal containing RNA and synthetic miR-1224. **C.** miR-1224 does not change the SHAPE pattern of the CCR5 -1 PRF signal. These experiments were performed by Dr. Meskauskas.

To further investigate its structural aspects, a runoff transcript containing the CCR5 -1 PRF signal was subjected to chemical protection methods and the products were analyzed by primer extension. Dimethylsulfate, kethoxyl and CMCT were used to probe the solvent accessibility of individual bases (**Figure 47 a**), while NMIA was



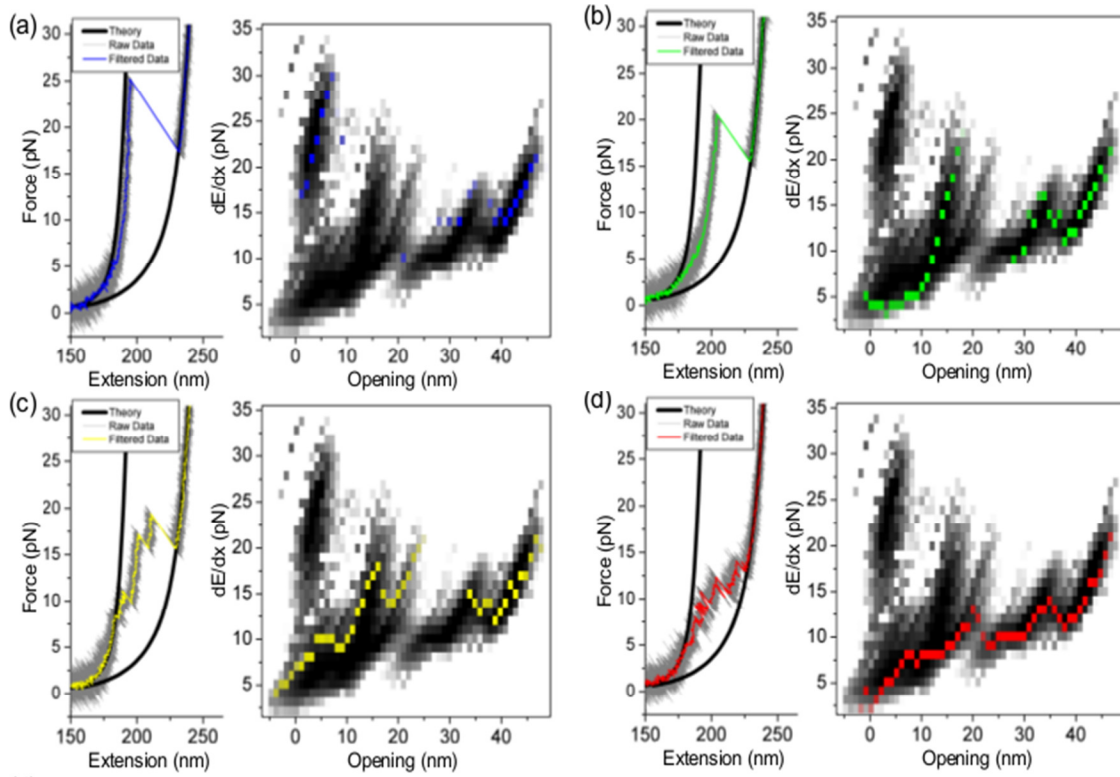
**Figure 47: Chemical protection analysis of the CCR5 -1 PRF signal.**

**a, b.** Autoradiograms of reverse transcriptase primer extensions performed on T7 transcribed RNA amplified from the CCR5 containing dual luciferase plasmid. Bands correspond to strong RT stops 1 nucleotide 5' of bases modified by chemical reagents. **a.** The CCR5 mRNA was either left unmodified (un), or modified with 3 increasing concentrations of dimethyl sulfide (DMS, reacts with A and C), 1-cyclohexyl-(2-morpholinoethyl)carbodiimide metho-p-toluene sulfonate (CMCT, reacts with U), or 1,1-Dihydroxy-3-ethoxy-2-butanone (Kethoxal, reacts with G) respectively. **b.** Primer extension reactions were performed on unmodified samples and samples incubated with 30, 65, and 110 nM NMIA (N-methylisatoic anhydride). These are labeled 1, 2 and 3 respectively beneath each sample, and un denotes untreated RNA. **c.** Data from panels a and b mapped onto a flat representation of the CCR5 -1 PRF signal. Stems 1 and 2 (S1 and S2) are boxed, and the loop is indicated as L. Alternative base-pairing schemes for the base of Stem 1 are indicated by dashed lines. The five different segments of Stem 1 are labeled a – e. Sugars protected from 2'OH attack by NMIA and bases strongly protected from chemical modification by DMS, CMCT, and kethoxal are noted as dark red or blue filled circles respectively. Weakly protected sugars and bases are denoted by light pink or light blue filled circles. Strongly modified (unprotected) sugars and bases are represented as light blue or pink open circles. **d.** Cartoon representation of the CCR5 -1 PRF signal. These experiments were performed by Dr. Meskauskas.



employed to probe ribose 2'-OH groups (**Figure 47 b**)[215]. Analyses of these data revealed a single solution for the structure of Stem 2 that roughly consists of 4 semi-helical segments (labeled a, b, d, e in **Figure 47 d**) and a small segment in the middle (c), each of which is separated by unpaired bases. This stem 2 closely conforms to the most stable computationally predicted structure (see the top two examples in **Figure 46 A**). Importantly, these unpaired bases should allow the entire structure to bend, enabling U23 to bridge the gap between C22 and U24 (**Figure 47 d**). Analysis of the Stem 1 forming region was consistent with the computational predictions: that while the distal region of Stem 1 is relatively stable, the proximal region is conformationally dynamic. Specifically: G8 and C9 were protected; the 3' half of Stem 1 (G62 – U66) was more deprotected than predicted; and four bases in loop 2 that are predicted to be single stranded (G70, and C72 → U75) were protected from chemical modification. These data suggest that the 5' half of the base of Stem 1 can interact with either the 3' half of the base of Stem 1, or with the 3' bases in the loop (L), thus accounting for the two main conformers observed in the native gels (**Figure 46 b**). Similarly, when this mRNA was stretched many times in an optical trap, two or three main conformers were consistently observed (**Figure 48**). These alternative interactions are shown as dashed lines in **Figure 47 c** and half cylinders in **Figure 47 d**. These experiments probed the structural ensemble as it exists in steady-state equilibrium; thus the two alternate conformations denoted with an arrow in **Figure 47 d** are likely represented by the two major, slowly migrating species shown in the gelshift experiments. In contrast, the optical trap experiments monitor pre-steady state refolding. In those experiments, the blue species (**Figure 48** and **Figure 49**) correlates very well with the pseudoknot, while the forms designated by red, green and yellow may correspond to the minor species observed in native gels (**Figure 43 b**). Together, these data reveal the presence of a complex and

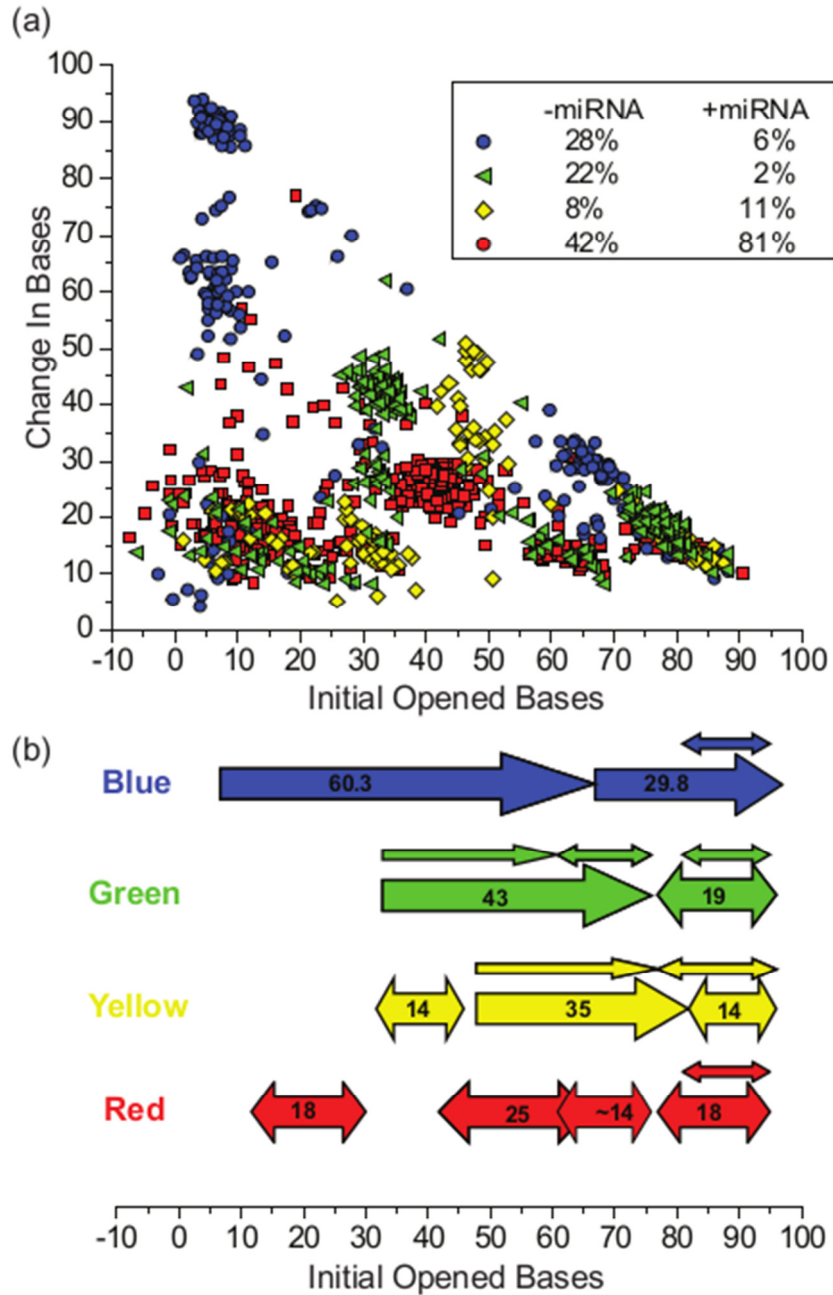
dynamic ensemble of mRNA pseudoknot structures that can be formed by the CCR5 -1 PRF stimulating sequence. Importantly, SHAPE analysis did not reveal any



**Figure 48: Force-extension curves and  $dE/dx$  density plots of the CCR5 structure.**

Black lines: Worm-like chain theory[216] for hybrid DNA/RNA handles with and without 97 additional bases of extended single-stranded RNA. Change in energy per opening distance ( $dE/dx$ ) calculated as described previously[217]. Gray scale density plot: The accumulated result from all scans. Colored squares: The result from the single scan shown in the force-extension curve. This work was performed by Michel deMessieres in the LaPorta laboratory.

differences in protection patterns of the CCR5 -1 PRF signal containing RNA in the presence of miR-1224 (**Figure 46 c**). This suggests that that hsa-miR-1224 does not function to alter the topology of this sequence per se, but that it may help to drive the equilibrium toward structures that promotes elevated rates of -1 PRF. This is further supported by changes between **Figure 43 a/b** as well as the distribution changes seen in **Figure 49 a**.

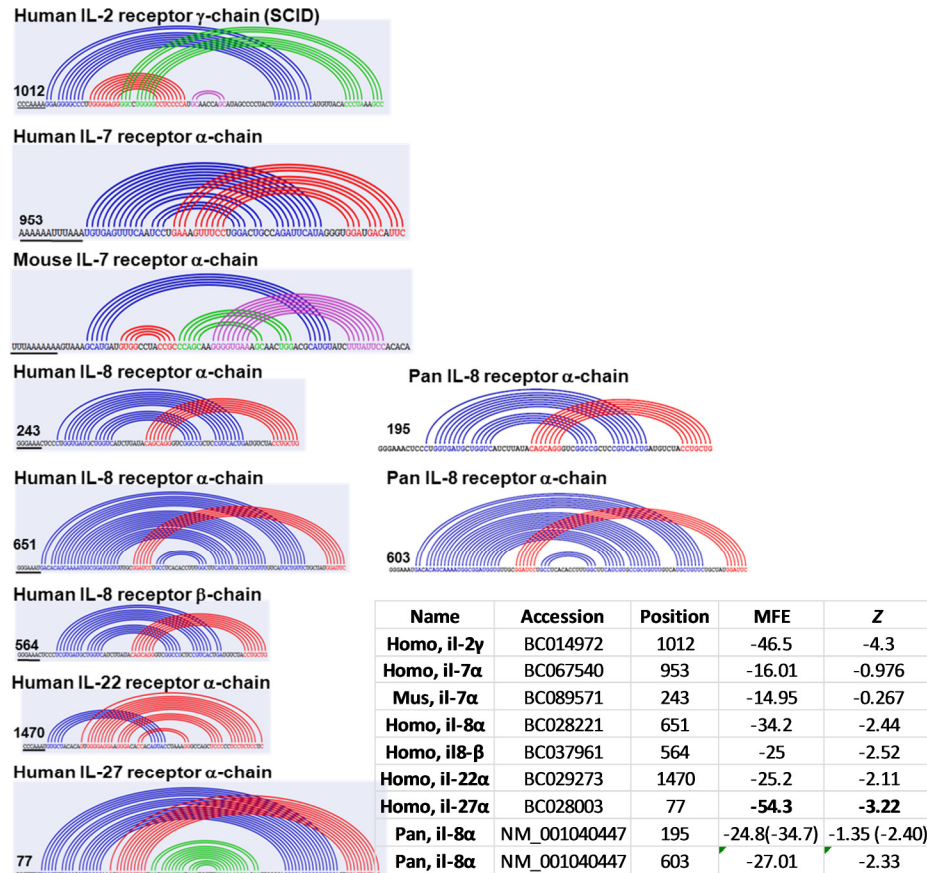


**Figure 49: Observed sub-steps when pulling apart the CCR5 -1 PRF signal.**

**a.** An alternate view of the  $dE/dx$  data in Figure 48 a. Each disruption path is composed of one or more substeps, plotted where  $x$  is the initial opened bases and  $y$  is the change in bases for that substep. **b.** Arrows indicate the mean behavior of the disruptions plotted in a. Numbers indicate how many bases were released for the given substep on average. Double-ended arrows indicate substeps which were observed to be reversible. Thinner arrows indicate substeps observed with less frequency. This work was performed by Michel deMessieres in the LaPorta laboratory.

## Identification of Additional -1 PRF Signals in Other Human Interleukin Receptor mRNAs.

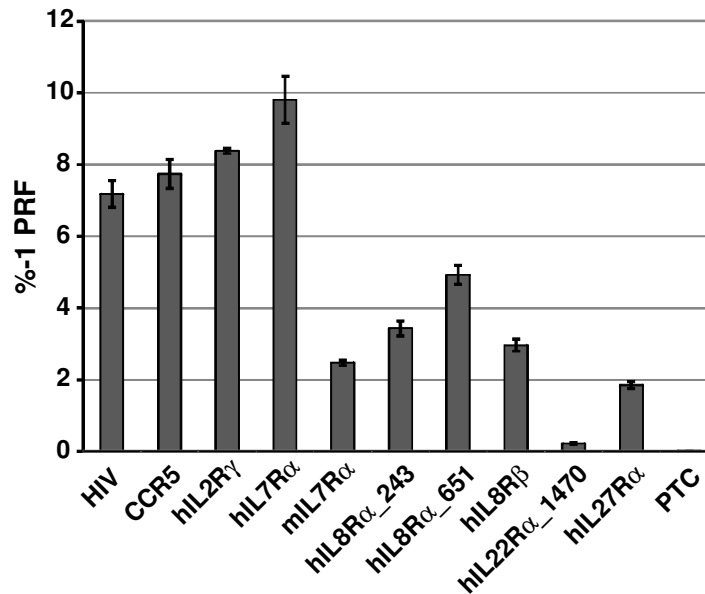
A search of the PRFdb revealed the presence of putative -1 PRF signals in mRNAs encoding seven additional interleukin receptor subunits (**Figure 50**). These



**Figure 50: Computationally predicted human cytokine receptor -1 PRF signals.**

One predicted solution structure for each predicted -1 PRF signal is shown as a linear Feynman diagram. The name, accession, position, predicted MFE and Z score against randomized sequence is provided in the inset table. Two conserved sequences in the *P. troglodytes* IL-8 receptor α-chains are also shown.

were cloned into dual-luciferase reporters and *in vivo* frameshifting assays were performed in HeLa cells (**Figure 51**). Efficient -1 PRF, as defined by the ability of a sequence to promote >1% -1 PRF, was effected by six of these sequences. Sequence



**Figure 51: Efficient -1 PRF is promoted by sequences in additional cytokine receptors.**

Computationally identified putative -1 PRF signals cloned into dual luciferase reporters were assayed in HeLa cells. Numbers in hIL8R $\alpha$  and hIL22R $\alpha$  denote the nucleotide positions of the slippery sites in the native mRNAs. Error bars denote standard error. See **Appendix 15** for more information regarding these sequence elements. This work was performed by Sharmishtha Musalgaonkar, Vivek Advani, and ATB.

beginning at nucleotide 1012 of the human IL-2 receptor  $\gamma$ -chain mRNA was a strong promoter of -1 PRF (8.4%). The human (9.8%) and mouse (2.5%) IL-7 receptor  $\alpha$  chain mRNAs contain each contain one functional -1 PRF signal; this mRNA was recently identified as an NMD substrate[208]. The human IL-8 receptor  $\alpha$  chain mRNA harbors two functional -1 PRF signals (3.4% and 4.9%), and like CCR5, these are conserved in the *P. troglodytes* IL8RA mRNA (**Figure 50**). The human IL-8 receptor  $\beta$  chain and the human IL-27 receptor  $\alpha$  chain mRNAs harbor one functional -1 PRF signal each (3.0% and 1.9% -1 PRF respectively). Potential -1 PRF signals identified in the mRNAs encoding the human IL-2 receptor  $\gamma$  chain and the human IL-22 receptor  $\alpha$  chain were not initially functional as defined in this assay, although they did promote frameshifting rates approximately 1 order of magnitude above the PTC containing control. When the

IL-2 receptor  $\gamma$  was modified to lengthen rather than shorten the spacer, its rate of frameshifting increased from ~1% to >8%.

Some informative trends may be gleaned from these data. While strong stem 1 structures appear to be important stimulators of -1 PRF, large internal loops may hinder this activity. However, the presence of overlapping slippery sites, e.g. A AAA AAU UUA AAU (human IL7RA) and U UUA AAA AAA (mouse IL7RA) may compensate for weaker stem 1 structures.

## **Discussion**

Prior to the current study, only three -1 PRF signals were known to exist in mammalian genomes, all of which are thought to be remnants of ancient retroviral insertion events[66,67,218]. The discovery of operational -1 PRF signals in the mRNAs encoding 5 different human and one mouse cytokine receptor mRNAs represents the first such examples in mammalian genes of non-retroviral origin, and as demonstrated with CCR5, the first in which -1 PRF may be used to control gene expression through mRNA stability.

### **-1 PRF and the Immune System.**

Each of the cytokine receptors identified in the current study have significant roles in human health as described in greater detail in **Appendix 15**. The discovery of operational -1 PRF signals in the mRNAs encoding five cytokine receptors, a subset of which appear to be evolutionarily conserved, has a potentially profound impact on our understanding of immune homeostasis. While a robust immune response is critical for limiting and controlling infection, left uncontrolled, it can rapidly result in pathology and death. Although there is a large body of literature describing how expression of small peptide mediators of the immune response (i.e. cytokines) are regulated at the level of mRNA stability, typically through cis-acting elements in their long 3' UTRs[219], this only

provides a global mechanism of immune regulation by controlling production of effector molecules. In contrast, the ability to control expression of cytokine receptors through -1 PRF induced NMD, and how rates of -1 PRF in turn may be controlled by miRNAs, represents a way for individual recipient cells to modulate responses to cytokines; this would provide the means to fine tune immune responses at the local level, and suggests a novel molecular mechanism underlying immune desensitization.

### **Modes of -1 PRF Regulation.**

The central unanswered question in the field of -1 PRF centers on its regulation. The characterization of numerous mutants in yeast[220], and the demonstration that siRNA knockdown of eRF1 stimulated -1 PRF in human cells[221] suggests that PRF could be regulated through production of “specialized ribosomes”[218,222]. However, over the course of numerous studies spanning the past 20 years, we have observed that mutants and treatments that globally affect -1 PRF generally promote deleterious phenotypes[191], suggesting that this may not be the preferred way to regulate -1 PRF. Indeed, global increases in -1 PRF in human and mouse cells due to rRNA pseudouridylation defects suggests that dysregulation of -1 PRF may contribute to the pathology associated with X-linked dyskeratosis congenita and Hoyeraal-Hreidarsson syndrome[223]. Alternatively, regulation of -1 PRF could be effected in a sequence-specific manner by ncRNAs capable of interacting with individual -1 PRF signals. A major advantage of this strategy is that it could enable individual cells to rapidly regulate -1 PRF on specific mRNAs by synthesizing or releasing ncRNA species. Not only does this confer sequence specificity, but it is also more rapid and energetically less expensive than producing new or modifying pre-existing ribosomes. The demonstration that oligonucleotides capable of disrupting -1 PRF mRNA pseudoknot formation can inhibit -1 PRF *in vitro* provided proof-of-principle for this concept[224]. Similarly,

antisense RNAs have been shown to stimulate -1 PRF[209,212,225], while an antisense peptide nucleic acid has been used to inhibit SARS-CoV mediated -1 PRF[226]. Here, the demonstration that expression of hsa-miR-1224 and hsa-miR-141 specifically stimulated CCR5-promoted -1 PRF, and that hsa-miR-1224 can directly interact with the CCR5 -1 PRF signal, solves the central, heretofore unanswered question of how -1 PRF may be regulated in a sequence-specific manner. To our knowledge, this is also one of the few demonstrations of an miRNA affecting the expression of a cellular gene through an interaction with its ORF[227–229].

Little is known about hsa-miR-1224: its expression is induced by lipopolysaccharide[230], consistent with a role in modulating the immune response, and its dysregulation may be associated with bladder cancers[231] and lupus nephritis[232]. has-miR-141 is a member of the miR-200 family: its overexpression has been implicated in numerous cancers (see[233–235] and references therein), and post-transcriptional regulation of hsa-miR-141 has been implicated in cellular plasticity and remodeling in response to changes in cell adhesion[236]. We suggest that -1 PRF is used to control the expression of a significant number of mammalian genes through the NMD, is in turn regulated by ncRNAs such as miRNAs, and that of changes in rates of -1 PRF through dysregulation of miRNA expression may in part contribute to human disease phenotypes.

#### **Structural Plasticity.**

Inspection of (**Figure 42a**) suggests that the miRNAs may interfere with mRNA pseudoknot formation, and thus should inhibit CCR5-mediated -1 PRF. Thus, the observation that hsa-miR-1224 and hsa-miR-141 stimulated CCR5-mediated -1 PRF was unanticipated. The computational and experimental observations of multiple CCR5 -1 PRF signal conformers (**Figure 46** and **Figure 47**) suggest a solution to this problem.



While hsa-miR-1224 associates with the CCR5 -1 PRF signals (**Figure 43**), chemical protection experiments did not reveal any structural changes (**Figure 46**), suggesting that hsa-miR-1224 does not stimulate -1 PRF by creating any new RNA conformational state. Consistent with this, the “optical trap” monitored the folding/unfolding dynamics of the CCR5 -1 PRF signal revealed 4 major folding/unfolding pathways for this RNA (**Figure 48**). Importantly, addition of hsa-miR-1224 to this system did not create any new pathways, but rather altered their relative abundances (**Figure 49**). Additionally, both hsa-miR-1224 and hsa-miR-141 can potentially participate in multiple, non-exclusive base-pairing interactions with different segments of the CCR5 -1 PRF signal. While the experiments described in this study cannot define the specific binding sites for these miRNAs, the finding that hsa-miR-1224 may suppress pseudoknot formation (**Figure 49**) suggests that the slippery-site proximal binding site may be responsible for -1 PRF stimulation. In sum, we suggest that: A) the CCR5 -1 PRF signal is structurally plastic; B) different conformers promote greater or lesser rates of -1 PRF; and C) different miRNAs are used to drive the structural equilibria toward specific conformational states. Further, we hypothesize that RNA conformational plasticity is the underlying biophysical basis through which cells may utilize miRNAs to regulate CCR5-mediated -1 PRF. In support of this, pH-dependent switching between mRNA pseudoknot conformers was recently shown to control termination codon reassignment and -1 PRF in Murine Leukemia Virus and beet western yellows virus respectively[237]. To borrow a term from the prion field, conformational plasticity is ‘enciphered’ within i.e. is an inherent property of, the primary mRNA sequences of translational recoding elements such as these and the CCR5 -1 PRF signal. We also suggest that single nucleotide polymorphisms (SNPs) that may alter the slippery site or change the conformational dynamics of -1 PRF promoting pseudoknots could also affect -1 PRF efficiency, and thus mRNA stability, ultimately affecting gene expression. If so, this may account for disease phenotypes

associated with SNPs that do not alter the primary amino acid sequences of their encoded proteins.

#### **Effects on the Viral Context.**

RNA viruses such as retroviruses, coronaviruses, and totiviruses require stringent levels of -1 PRF for their propagation[65,238,239]. We suggest that their -1 PRF promoting structural elements have evolved in two different ways so as to ensure set rates of -1 PRF. First, their -1 PRF stimulatory elements should not interact with any ncRNAs present in the cells in which they replicate. Second, these elements should only have a single folding solution under physiological conditions. Thus, minimization of RNA conformational plasticity and avoidance of sequence-specific regulation of -1 PRF by ncRNAs may explain why -1 PRF-dependent RNA viruses are able to efficiently replicate in their hosts.

## Chapter 5

### Where do we go from here?

The amount of data available to those interested in mRNP dynamics is increasing at a rate which defies description. High-throughput SHAPE[240], next-generation sequencing of RNA and ribosome protected messages[241], and genome-wide measurements of RNA structure[242] provide three immediate examples of fantastically powerful new tools which have the potential to transform observations of mRNA structure and function from a view of individual molecules through an opaque, dirty lens to clear observations of every RNA molecule, in concert, throughout the life-cycle of the cell.

Data provided by these techniques will prove to be a tremendous boon to anyone searching for physiologically active RNA species in the cell; and the best part, we have not even figured out **what** to search for. There are huge new playgrounds of information just waiting for someone to come along and ask a new question. This of course has always been true, but at a personal level I never thought the exponential growth of information was so dramatic or immediately exciting. This project attempted to perform a predictive search for interesting mRNA secondary structures and develop the results of that search in the eukaryotic context. Transitioning from *Saccharomyces cerevisiae* to cultured mammalian cells provided an opportunity to change the focus of the work slightly and observe the effects of ncRNA on -1 PRF. Each of these three foci has the potential to be completely reworked given the new techniques and technologies available; and thus provides nearly infinite space for future experimentation. This text aims to provide a glimpse into a few of the possibilities for future work in the context of computational searches for interesting mRNA structures, different directions which may

be followed in *Saccharomyces cerevisiae*, and potential new avenues to explore in the mammalian context.

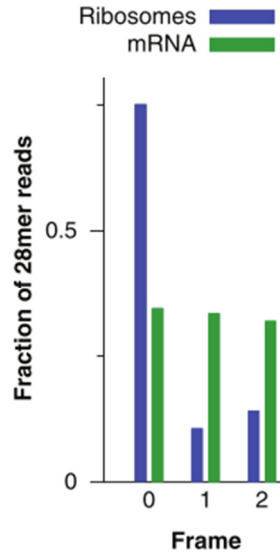
### **Computational Next Steps**

The current implementation of the PRFdb may be easily extended to: continue searching more sequence databases; make use of phylogenetic information, RNA alignments, and other statistical measurements to improve the existing set of analyses; cross-reference to other datasets to add new metrics of “interesting” to the existing predictive criteria; and search itself and other sequence databases for specific functional motifs.

In the first instance, this merely requires pointing the code at a list of GenBank accessions and letting it run; as long as database storage needs are met, search times should not become a problem until  $\sim 10^8$  sequences have been analysed. If one wishes instead to improve the sensitivity of the measurements provided by the PRFdb via a solid statistical framework or phylogenetic information, the list of already existing tools contains some excellent candidates. Indeed, some of these are already supported by Bioperl and so should be instantly available. In this context, the challenge lies in properly curating a training set of data, or devising a sensitive strategy to infer that specific sequences are related and therefore available to use as components of an RNA alignment. Similarly, it is possible with some small improvements to add other metrics of “significance;” a primary candidate is the Valley Index score mentioned in Freyhult et al.[148]. Another simple but potentially powerful improvement would be to extend the existing RNAMotif parser and descriptor generator to automatically perform database searches as per Lupták et al.[131]. This approach has the neat side-effect of turning the existing dataset into a search space for specific structures of interest.

## Correlation Studies

While all of the above approaches have merit, they do not leverage the new datatypes and approaches which are so exciting. The single most potentially sensitive search tool for substrates of No-Go decay, NSD, NMD, and functional genomic -1 PRF signals already exists in the dataset from Ingolia et al. (**Figure 52**). With minor

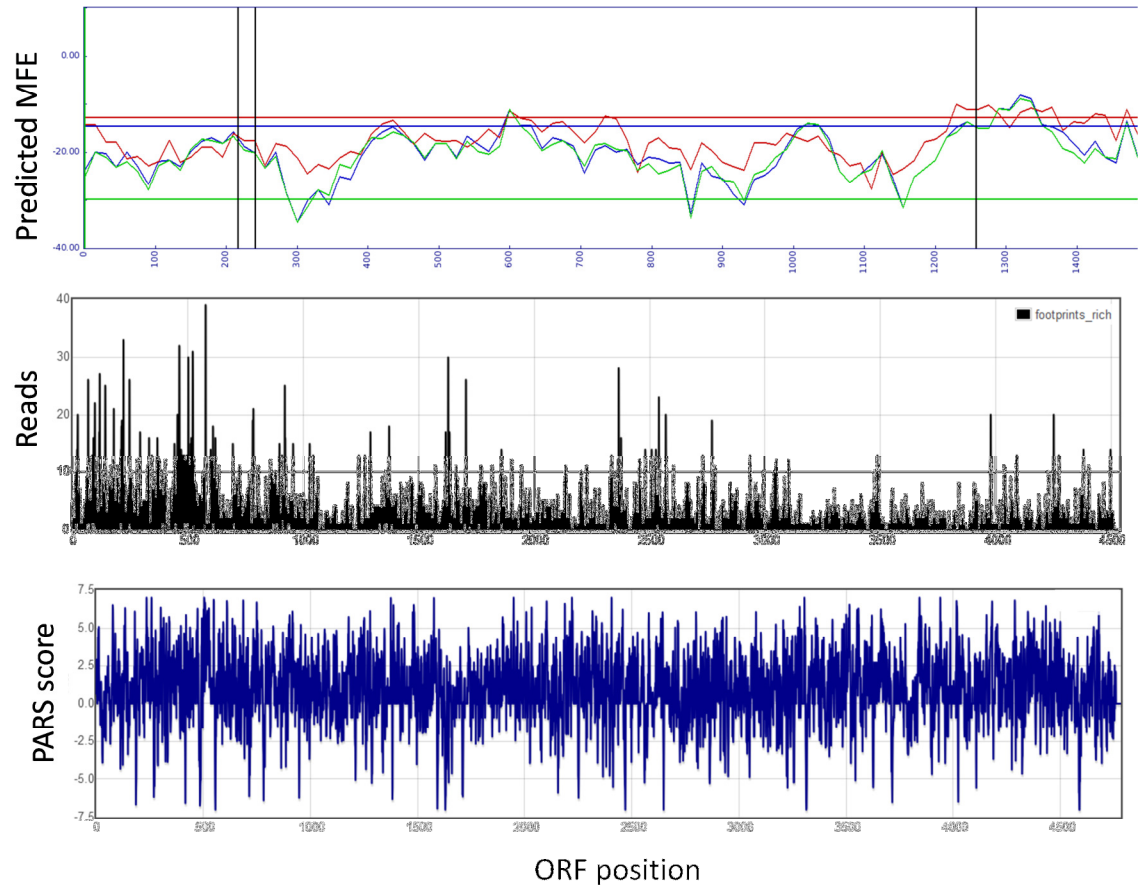


**Figure 52: Position of 28mer reads with respect to reading frame in ribosome profiling.**

One search strategy for functional PRF signals includes finding the positions where this ratio of ribosomes/mRNA is significantly increased in alternate reading frames. This figure is from Ingolia et al. [241].

improvements, and potentially more reads, it is possible to pinpoint every actively translated base in the yeast genome upon which a ribosome changes reading frame. Furthermore, this same data was recently made available for mouse[243] and human[244] cells.

Using the same data, it is also possible to compare ribosome profiles to predicted MFE with respect to ORF (**Figure 53**). A simple version of this was implemented as part of the PRFdb. If a given sequence of interest also has a significantly larger number of



**Figure 53: MFE prediction vs. ribosome footprint density vs. PARS score for *PDR5*.**

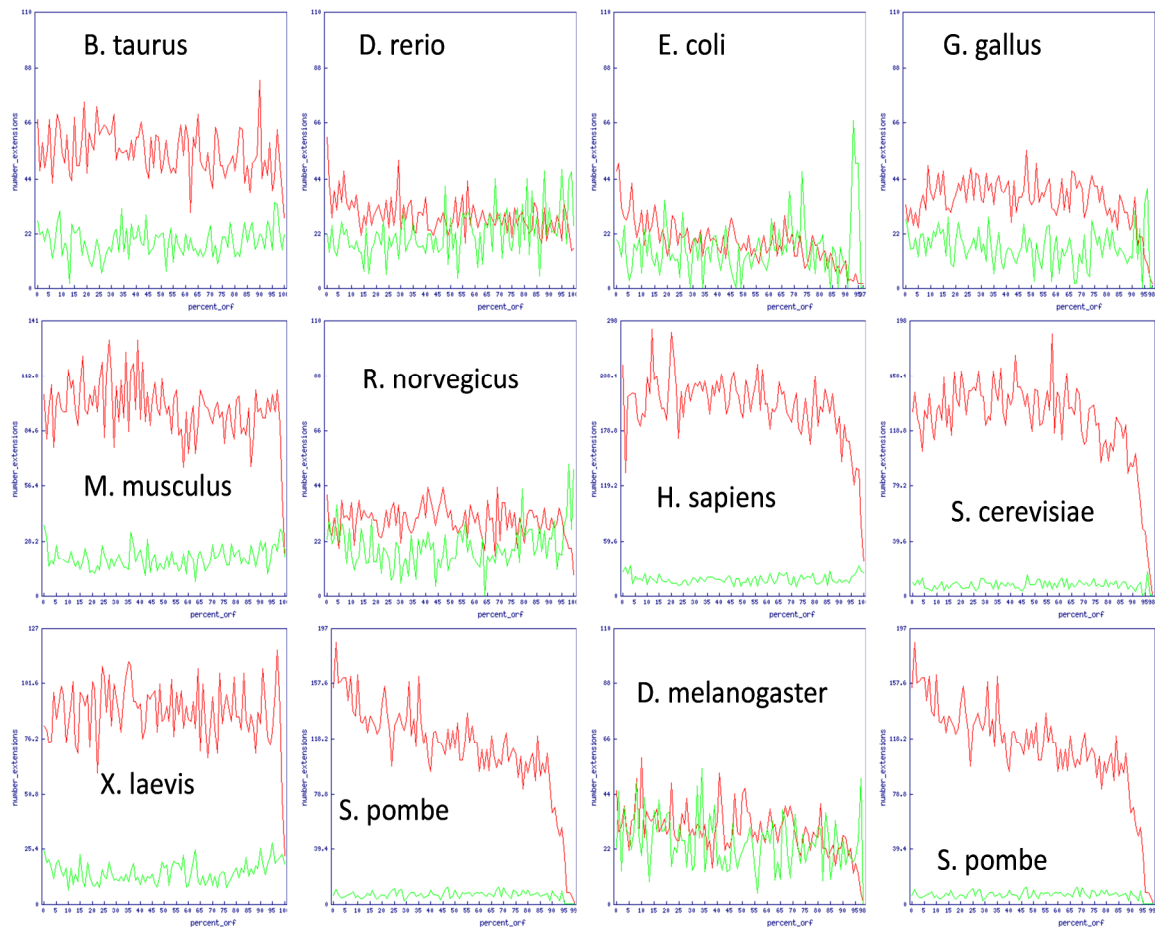
**Top:** The predicted MFE using pknots, Nupack, and RNAFold for a floating window of 115 bases along the *PDR5* ORF is plotted above. Only the first 1,400 bases are shown. **Middle:** Below is a histogram showing the number of ribosomal footprints detected along the *PDR5* ORF. Using this, it is possible to visualize the reads of full mRNA and footprints in yeast cells fed rich or amino-acid poor media. **Bottom:** PARS score at each position along the *PDR5* ORF. In this implementation, the score is:  $\log_2(\text{S1 reads}/\text{V1 reads})$  so that strongly dsRNA is less than 0 in the hopes of finding similar trends to the predicted MFE.

ribosomes protecting the same position, then it is reasonable to hypothesize that the mRNA is causing actively translating ribosomes to pause. In order to properly perform this type of analysis, these statistics will need to be more thoroughly developed, including metrics to properly exclude ORFs with a low signal/noise ratio as well as proper metrics to compare the ratio of footprint reads / mRNA reads. Another application of similar technology was introduced with the PARS score, which provides a

measurement of the observed ratio of reads observed after RNase V1 cleavage (thus measuring double-stranded bases) vs. the reads observed after RNase S1 (single-stranded). This data, like the ribosomal footprinting data, needs to be more fully developed before any real conclusions can be made, but its potential power is astonishing when taken in concert with these other data types. It is also possible to further interpret the information already existing in the database. For example, **Figure 54** plots the distribution of potential -1 PRF signals with respect to open reading frame for a series of genomes. In many cases, there is a simultaneous sharp decrease in the total number of potential -1 PRF signals and an increase in the percentage of those few that remain which are predicted to extend the -1 reading frame by more than 30 amino acids. This observation is one of many which may be found in the existing database by only looking more closely at the extant data.

### **Possible next steps in *Saccharomyces cerevisiae***

Improvements in the strategies for computationally searching for potential -1 PRF signals may yield a larger and more accurate set of signals to examine. The signals which have already been identified pose far more questions than they answer. When looking at the work performed in *Saccharomyces cerevisiae*, one candidate ORF stands out as particularly interesting: *EST2*. As previously noted, this gene harbors five computationally identified high-confidence -1 PRF signals. In addition, there is already a tremendous wealth of data which examines cellular homeostasis when telomere maintenance is dysregulated, including reporter systems and assays to directly assay steady-state telomere length, the rate of telomere shortening, and the proportion of cells in each phase of the cell cycle.

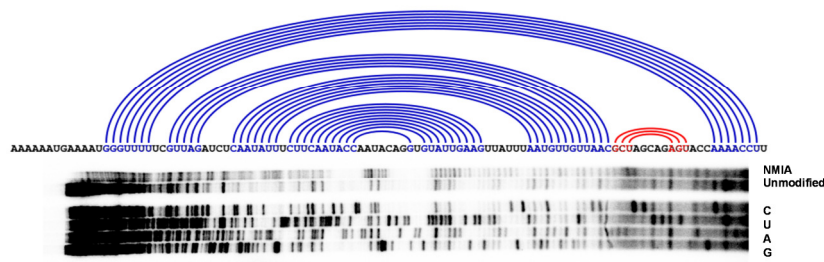


**Figure 54: Histograms of potential genomic -1 PRF signals**

Potential -1 PRF signals are plotted from 12 genomes. Red lines count the number of potential -1 PRF signals with respect to ORF. In green the percentage of -1 PRF signals which are predicted to result in a long -1 frame extension with respect to ORF. In many species there is an increase in the percentage instances which extend by more than amino acids at the 3' end (visible as a spike in the green plot). Simultaneously, these species exhibit fewer total -1 PRF signals at the end.

The *EST2* mRNA contains some interesting and strong secondary structures (Figure 55) and is illustrative of another potentially powerful future experiment. It should be technically trivial, but expensive, to perform a hybrid experiment of hSHAPE and next-generation sequencing. The results would be analogous to those observed in the PARS dataset, but query the relative flexibility of every base. Depending on the experimental conditions, it should be possible to search for nucleotides constricted by





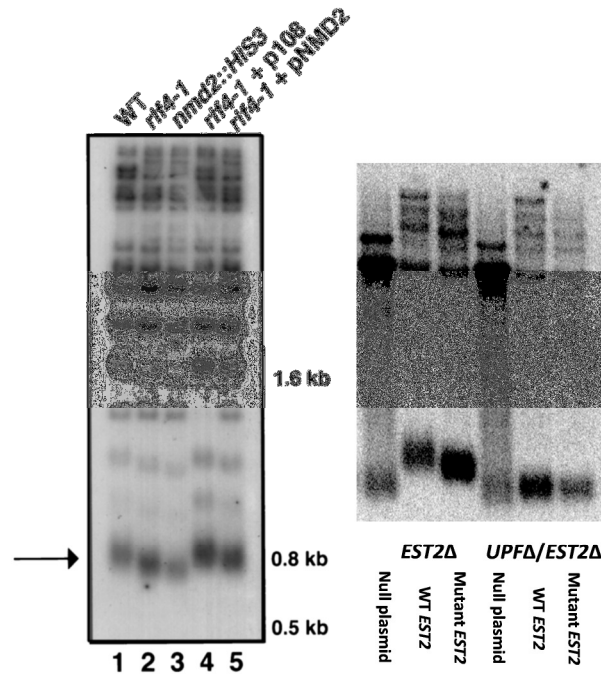
**Figure 55: The *EST2* mRNA is strongly protected position 1653 -1 PRF signal**

SHAPE was performed on *EST2* PRF signal containing mRNA showing strong protection from attack by NMIA. When the same assay was performed on sequence 100 bases downstream of this, little protection was observed (not shown).

protein binding (crosslinking as performed in Zhang et al.[245] to find binding partners for AUF1), directly observe mRNA dynamics through the cell cycle, or as a function of active translation (by adding cycloheximide to stop translation).

The *PGK1* reporter constructs containing the *EST2* -1 PRF signal was a strong substrate for NMD (**Figure 25 A** and **B**, lane 4, and **Figure 26**). Furthermore, the full length message with silent mutations in each of the 5 putative slippery sites was significantly stabilized (**Figure 28**). Therefore it is reasonable to hypothesize that the slippery site mutant containing full length mRNA should act similarly *in vivo* from wild-type to NMD deficient cells.

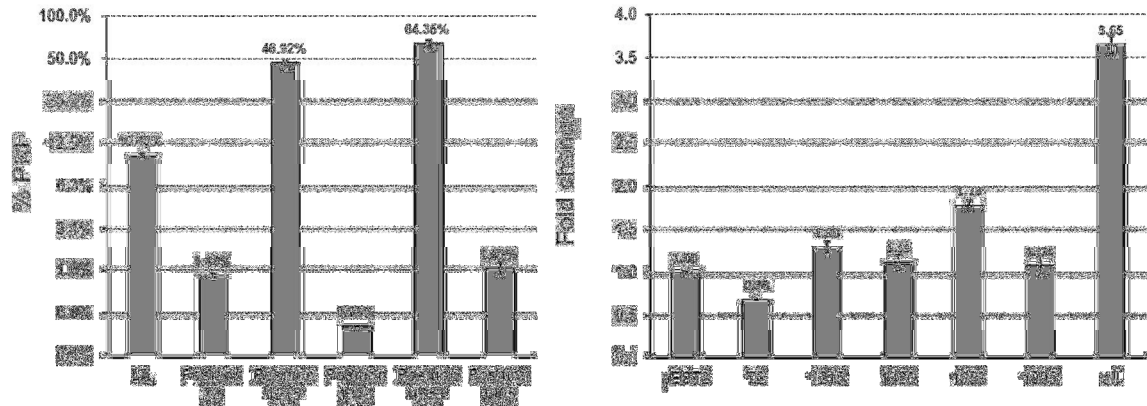
Testing this hypothesis is made possible because yeast telomeric regions include some well characterized and useful restriction sites[246]. These sites have been used to perform Southern blots which are sensitive to relatively small variations in telomere length, and therefore changes in aging of the cells. When this was performed in NMD deficient cells, the telomere size was shown to decrease; suggesting that an NMD mediated shift in the equilibrium of the telomere cap components leads to deficient telomere repair. The same assay was performed in *EST2* deficient cells supplemented with either a wild-type or mutant copy of *EST2* and a similarly counterintuitive result was observed (**Figure 56**).



**Figure 56: NMD deficient and *EST2* slipsite mutant cells have shorter telomeres.**

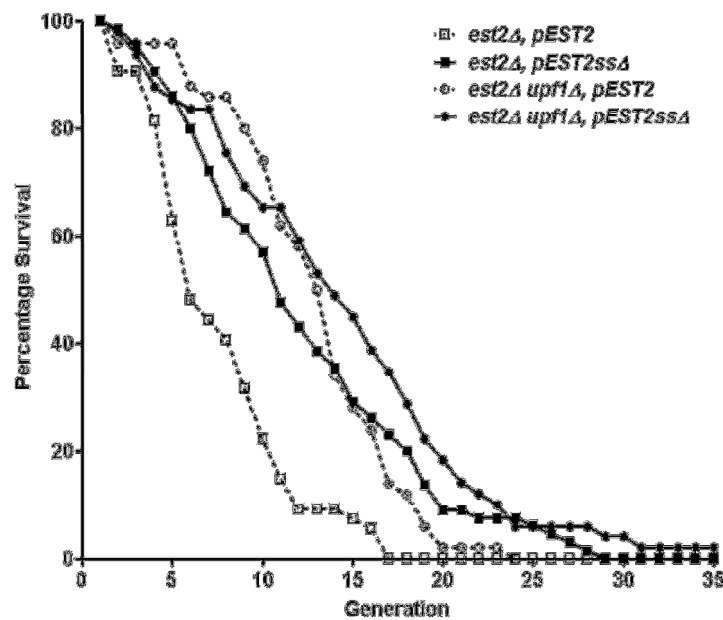
The left Southern blot is from Lew et al.[86], genomic DNA of wild-type and NMD deficient cells was probed with an end-labeled telomere specific probe and a decrease in telomere length was observed (lane 1 compared to lane 3). When a similar experiment was performed using *EST2* deficient cells supplemented with a low-copy plasmid borne copy of either wild-type or slippery site mutant *EST2*, a similar effect was observed. The *UPFΔ* and *UPFΔ/EST2Δ* lanes recapitulate the previous result and suggest that the *EST2* mutant mediated effects are hypostatic to *NMD* mediated shortening.

One avenue of future inquiry therefore includes expanding these analyses to attempt to understand the relative contributions of each -1 PRF signal from *EST2* (**Figure 57**), or the contributions of other telomere associated proteins (as speculated upon in the Discussion of Chapter 3 and **Figure 32**). It is also possible to ask about how -1 PRF mediated changes in *EST2* expression affect the yeast cell cycle; in this realm there are some wonderful possible experiments using FACS, cellular staining and microscopy, and even assays which observe the maximum generation number of yeast strains harboring specific *EST2* mutations (**Figure 58**).



**Figure 57: Initial attempts to quantify the relative contributions of each *EST2* PRF signal.**

**Left:** Dual luciferase assays were performed to assess the quality of each of the 5 potential *EST2* -1 PRF signals. Positions 1215 and 1653 (previously identified) were shown to be strong. **Right:** qPCR analyses followed and showed minor effects on full length *EST2* stability from positions 1215 and 1653 alone, but a much stronger effect when all were mutated.



**Figure 58: Counting the maximum number of generations of mutant *EST2* harboring cells**

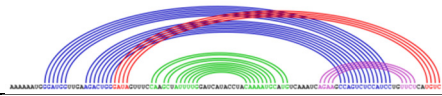
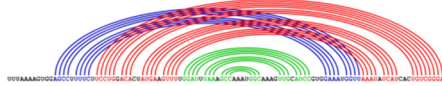
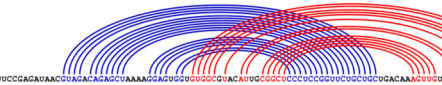
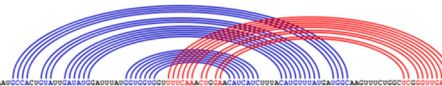
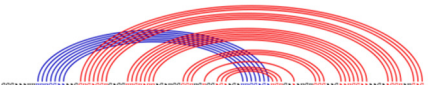
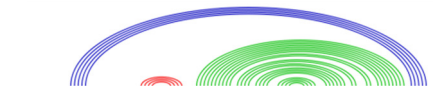
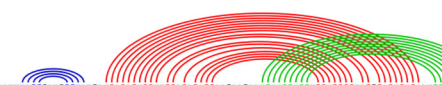
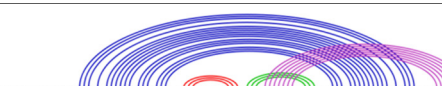
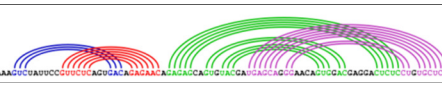
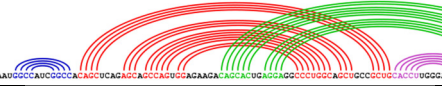

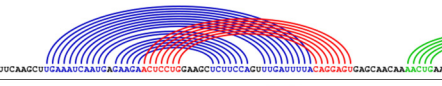
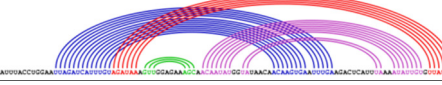
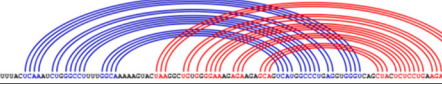
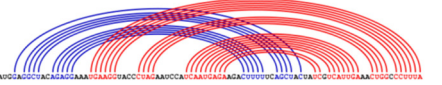
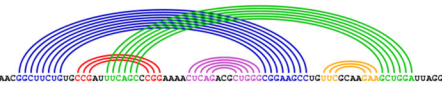
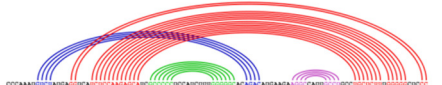
While silent mutation containing cells live longer in both the NMD knockouts and mutant harboring cells, the surviving cells are significantly sicker, have a slower doubling time, and display significant morphological differences from wild-type. This analysis was performed in the laboratory of Dr. Brittenbach-Koller.


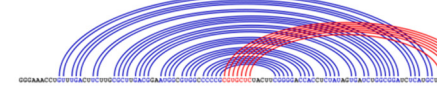
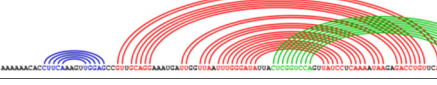
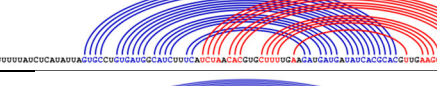
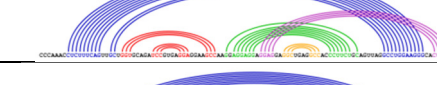
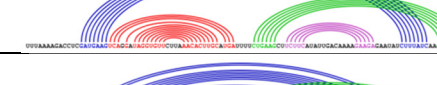

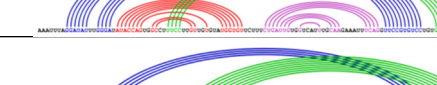
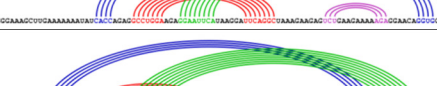
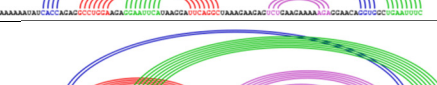
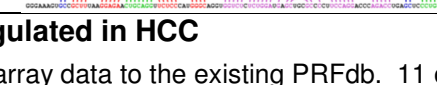
As these few examples illustrate, there are some exciting possibilities for future inquiry of -1 PRF in yeast. Some of the most interesting seek affects of -1 PRF upon cellular homeostasis or use recent technological advances to query the entire yeast transcriptome for strong candidates. The transition from yeast to the metazoan context provides an opportunity to reverse this logic and instead ask: how do other regulatory systems affect -1 PRF?

### **Future Mammalian Work**

We performed some initial experiments which implicate specific miRNA species as effectors of -1 PRF in mammalian cells. The possible questions this allows us to ask will certainly increase as the ncRNA field continues to mature. Likely candidate questions include: cell type specificity, effects during development, and changes via -1 PRF as miRNA species are dysregulated in disease states such as cancer.

One path we started following in the mammalian context is an attempt to correlate predictions from the PRFdb to microarray data from 200 patients with hepatocellular carcinoma. These patients donated one healthy and one cancerous liver tissue sample for analysis as well as diagnostic information including: tumor size, AFP, viral status for Hepatitis, lifespan post-diagnosis, etc. The resulting dataset provides approximately 14,000 points of expression between healthy and diseased liver cells per patient. Initial analyses focused upon improving the existing HCC subclasses[247]. A mixture of PLS and K-means clustering reduced the dataset from 14,000 genes to approximately 340 candidates which were then cross-referenced against the PRFdb, searching for strong candidates in both sets (**Table 5**).

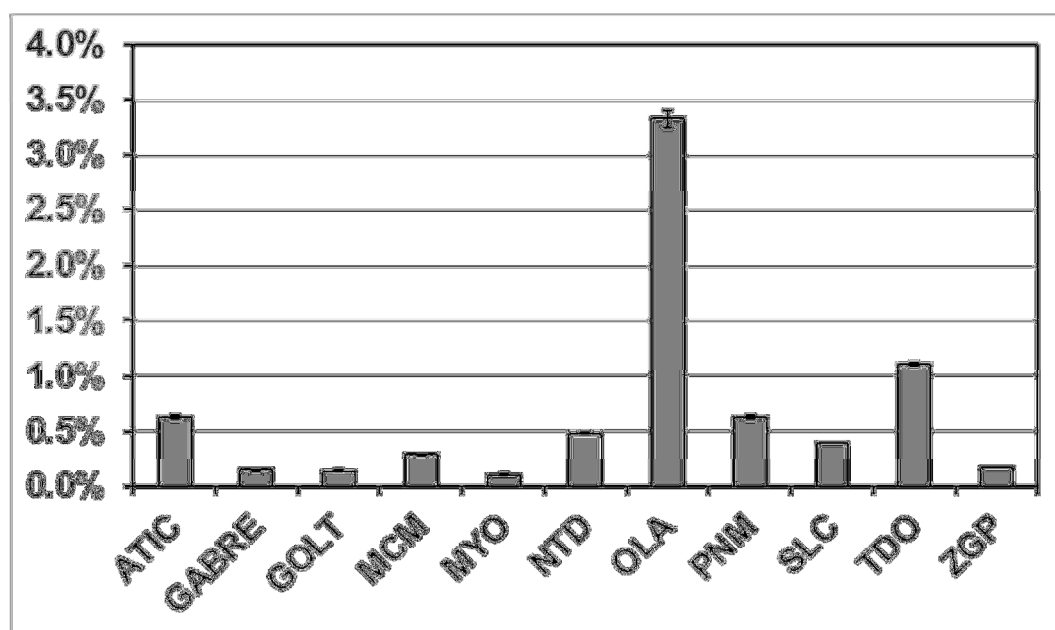
ORF	Position	MFE	Z	Prediction
ACTL6A (NM_178042)	485	-27.0	-2.13	
ASNS (BC014621)	649	-23.0	-0.54	
ATIC (BC008879)	1656	-33.2	-1.59	
BCHE (BC018141)	<b>503</b>	-25.1	-1.89	
CKAP5 (NM_014756)	212[3 6]	-20.1	-0.48	
CKAP5 (NM_014756)	2906	-32.2	-4.53	
CKAP5 (NM_014756)	2927	-36.1	-3.23	
DLK1 (BC014015)	509	-41.0	-0.99	
DNAJC9 (NM_015190)	264	-32.3	-1.20	
FAM21A (NM_001005751)	3307	-39.2	-0.30	
FOS (BC004490)	408	-46.8	-1.70	
GABRE (BC059376)	735	-27.8	-3.63	
GOLT1B (NM_016072)	437	-14.3	-0.43	
KIF20A (BC012999)	2668	-30.9	-1.66	
KIN (NM_012311)	1083	-18.9	0.42	
LPCAT1 (BC020166)	1110	-41.2	-1.75	
MCM5 (BC003656)	1004	-38.2	-2.56	

MYO5C (NM_018728)	3527	-23.3	-1.48	
NT5DC2 (BC014550)	841	-33.5	-0.07	
NUP37 (BC000861)	719	-29.1	-2.51	
OLA1 (NM_013341)	513	-27.4	-2.44	
PNMA1 (NM_006029)	1740	-38.6	-1.52	
PPP4R1 (NM_005134)	2197	-22.4	-1.65	
RNASEN (BC054003)	284[5 8]	-22.3	-0.94	
SLC38A (BC040342)	981	-23.0	-0.55	
TDO2 (BC005355)	732	-24.0	-3.79	
TDO2 (BC005355)	744	-24.0	-3.66	
ZGPAT (BC032612)	664	-37.5	-0.76	

**Table 5: -1 PRF Candidates Significantly Disregulated in HCC**

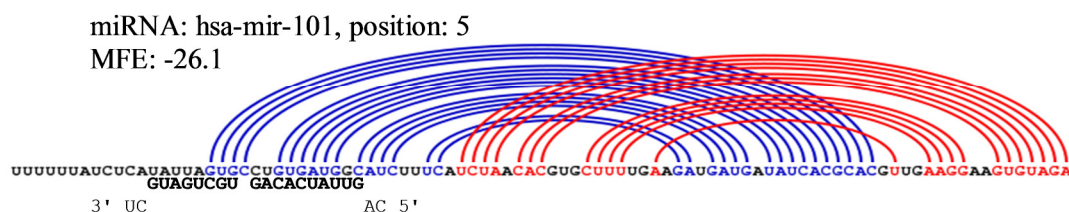
24 candidates were chosen by cross-referencing microarray data to the existing PRFdb. 11 of these were cloned into the dual-luciferase reporter system and analysed.

Thus far 11 of the 24 strongest candidates have been successfully cloned into the dual-luciferase reporter system. Two of them, OLA1 and TDO1 promote significant levels or frameshifting (**Figure 59**). A simple search with miRanda against the annotated *Homo sapiens* miRNA database showed that hsa-miR-101 may hybridize strongly with this potential -1 PRF signal (**Figure 60**). Interestingly, hsa-miR-101 has previously been implicated in HCC[248].



**Figure 59: Dual luciferase of HCC involved -1 PRF signals**

When the dual-luciferase assay was performed with the successful clones of HCC involved -1 PRF signals, two candidates displayed significant (> 1%) -1 PRF: OLA1 and TDO1.



**Figure 60: Possible interaction of hsa-miR-101 with the OLA1 -1 PRF signal**

hsa-miR-101, hsa-miR-1182, hsa-miR-1226, and hsa-miR-564 have potentially strong interactions with stem 1 of the *OLA1* -1 PRF signal (-26.1, -22.5, -28.3, and -27.9 kcal/mol respectively).

Taken as a whole, I find myself a little envious of new students who will get to play in the new fields of inquiry which are opening today. I am certain this experience is cliché, but it remains a bittersweet conclusion to a tremendously enjoyable learning experience.



## **Chapter 6**

### **Experimental Procedures**

#### **Introduction**

The techniques of Molecular Biology are fundamentally difficult for a new student to understand. It is my hope in this document to provide two apparently contradictory things at the same time: a concise summary for each process suitable for using in a scientific publication followed by a longer, explicit, and usable protocol. A much expanded version of this document containing examples, sample results, precursor protocols, and buffer recipes is separately maintained and periodically copied to the Dinman lab website: <http://dinmanlab.umd.edu/Information/protocols/>

#### **Cell Culture**

The HeLa cell line[249] (ATCC), its derivative cell lines: HeLa Tet-Off[250] (Clonetech), Tzm-BL[213] (aidsresearch.org and kindly provided by Dr. J. DeStefano); CHO[251] (kindly provided by Dr. D. Mosser); and Vero[252] (kindly provided by Dr. B. Frederickson) were cultured according to the manufacturers' instructions, which are summarized here.

All Vero and HeLa cell lines were cultured at 37° with 5% CO<sub>2</sub> using Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% irradiated fetal bovine serum (FBS), 1x non essential amino acids (NEAA), and [4 mM] glutamine. 1% Penicillin/Streptomycin was neither added when performing assays which are sensitive to changes in translation, nor when using Tet-Off cells. Instead Tet-Off cells received 200 µg/ml G418 in order to sustain the transactivator protein containing plasmid. CHO



cells were maintained in the same fashion except Modified Eagle's Medium  $\alpha$  (MEM  $\alpha$ ) supplemented with 1x proline, [4 mM] glutamine, and 1x NEAA.

These cells are all best passaged when growing exponentially. In most cases this suggests 70-80% confluency, however CHO cells will commonly grow to 100% confluency in less than 24 hours. HeLa derived cell lines will tire after 20-25 passages, while the existing Vero cell lines begin at earliest passage 85+.

Pre-warm Media, 0.05% trypsin (or EDTA containing PBS) and 1x PBS to 37°. Aspirate old medium from the T75 flask with a sterile pipette tip. Wash cells with 3-7 ml 1x PBS to remove residual media and cell debris.

Add 3-5 ml 0.05% (w/v) trypsin-EDTA and evenly disperse by rocking. (Storage note: After thawing, trypsin may be stored at 4° for up to 2 weeks, do not freeze thaw. In practice, trypsin may be stored significantly longer) Incubate at 37° 5-10 minutes to detach cells. If using non-enzymatic PBS-EDTA, a cell scraper may be required to detach cells, otherwise most (except Vero) cells are usually detached.

Add 5 ml fresh media to wash cells from the plate surface, pipette to dissociate cell clumps. (Avoid making foam, it somehow kills my cells) Remove to 15 ml conical tube and spin down 2 minutes at  $\leq 100$  g. Remove media and resuspend in 1 ml DMEM. Mix gently, add 1  $\mu$ l cells to hemacytometer with  $\leq 1$   $\mu$ l of trypan blue to count viable cells. 100 undiluted cells in the square on our hemacytometer correspond to 1,000 cells /  $\mu$ l.

Aliquot cells according to requirements: 30,000-40,000 / well of a 24 well plate is common; 100,000 in a T75 flask. Fewer cells should be used for longer-term transfections.

## **The Dual Luciferase Assay**

Dual luciferase assays were performed as previously described[204] using a Turner Biosystems GloMax-Multi Microplate Multimode Reader with some modifications. Data analysis follows the process outlined in Jacobs et al.[205].

The reference material provided by Promega is excellent. When working with mammalian cells, no changes are required. In yeast however, the provided passive lysis buffer should not be used. When cells are disrupted in the bead-beater, this buffer makes a foam which leads to large variance in the final results. Instead, use a PBS buffer with PMSF and (an optional) protease inhibitor cocktail.

150 mM PMSF: 0.2613 g of PMSF into 10 ml methanol. PMSF is extremely hydroscopic.

10x PBS: 30.47 g  $\text{Na}_2\text{HPO}_4$ , 0.155 g  $\text{NaH}_2\text{PO}_4$ , 2.19 g NaCl, pH this to 7.4 with HCl, add water to 250 ml and autoclave.

Protease inhibitor: 150  $\mu\text{l}$  Aprotinin (2mg/ml), 150  $\mu\text{l}$  leuprotinin (1mg/ml), 150  $\mu\text{l}$  Pepstatin A (1mg/ml), 550  $\mu\text{l}$  water.

3 ml 10x PBS, 0.2 ml protease inhibitor, 0.2 ml 150 mM PMSF, water to 30 ml.

Grow overnight cultures to exponential growth (This varies from an OD of 0.7-2.0 depending on strain. Different strains require different ODs, do a growth curve. Harvest cells by spinning down gently in the Sorvall (1200 rpm for 2 min) and removing media. Wash the cells with 1ml lysis buffer and move them to a fresh eppendorf. Spin them gently and remove lysis buffer. Add 500ul fresh lysis buffer and 500ul glass beads and chill on ice/cold room for 5 min. Vortex in the cold room at minimum 5 minutes or bead beat them 5 minutes. While the cells are lysing, start melting an aliquot of the Luciferase Assay Buffer.

## **Flow Cytometry, Fluorescence Activated Cell Sorting and Western Blotting**

### **FACS**

TZM-BL cells were harvested with a PBS dissociation buffer without enzyme. They were washed and stained with: primary antibody Mab hCCR5 (aidsreagent.org part numbers: 45531.111, 45523.111), which contains the IgG<sub>2a</sub> isotype from a Balb/c mouse; and secondary antibody Goat Anti-Mouse IgG<sub>2a</sub> (γ2a chain specific) conjugated to Fluorescein (Southern Biotech #1080-02). Fluorescence intensity was analyzed by flow cytometry of 10,000 events (using a BD FACS Canto II).

Perform transfections in 12 well plates as if repeating mRNA abundance: [10 nM] siRNA applied once per day for two days. Most printed protocols are for intracellular proteins. As a result they permeabilize the cells with Tween-20 or methanol, then fix them with 0.4-1.0% formaldehyde; while some protocols work in reverse order. In this particular case, the CCR5 antibody is theoretically specific to the second extracellular loop, but the antibody performs much better when applied to permeabilized cells.

After cells grow with siRNA for 48 hours, harvest them into 1.6 ml eppendorf tubes by rinsing with non-enzymatic PBS dissociation buffer; aliquot cells to 1 ml and let cells incubate 3-8 minutes at 37°. Fix cells by applying 0.2-0.6% formaldehyde at 25° for 5-10 minutes as per [203]. Pellet cells at 400 g for 2 minutes, suspend in 100 µl 5% FBS in PBS. Permeabilize cells with either 1 ml PBS-Tween (the same formulation used in Western blotting, in my hands both work equivalently) or 1 ml methanol. Sit on ice for 20-30 minutes as per[253].

Pre-rinse cells with 5% FBS PBS and pellet. Block with 10% FBS PBS for 15-30 minutes. Rinse with 5% FBS PBS and pellet. Incubate with 1:500 – 1:5,000 dilution primary antibody for 30-60 minutes on ice. Rinse three times in 5% FBS PBS. Incubate with 1:500 secondary antibody on ice in the dark for 15-60 minutes. Rinse three times in

5% FBS PBS. Aliquot cells for FACS analysis by resuspending in 500  $\mu$ l 5% FBS PBS in a 12x75 Fisher FACS polystyrene culture tube (14-956-3C) and keep on ice.

### **Western Blotting**

Cell lysates were prepared from TZM-BL cells transiently transfected with siRNA against hSmg1, hRent1, hUPF2 and analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and Western blotting as previously described [254] with changes as per [255]. Westerns were performed with antibodies against hCCR5(BioLegend Mouse IgG anti hCCR5 #321702) followed by secondary antibodies conjugated to horseradish peroxidase (HRP) (SantaCruz Goat IgG-HRP #sc-2005). The lysates were also probed with anti- tubulin antibodies (Mouse IgG anti Tubulin, a kind gift from the Song laboratory) followed by secondary antibodies against mouse IgG conjugated to HRP. (unsuccessfully with SantaCruz #sc-20227 Lot # A0307 Goat polyclonal IgG followed by SantaCruz #sc-2020 Donkey anti-goat IgG-HRP) After probing, blots were exposed using Luminol (SantaCruz #sc-2048) and a Fujifilm CCD camera.

### **Northern Blotting**

All RNA for northern blots was extracted with acid phenol/chloroform (pH = 4.5) from mid-logarithmic cell cultures [200], or with TRIzol<sup>®</sup> Reagent following the manufacturer's directions (Invitrogen). RNA (northern) blotting was performed as previously described[78]. Equal amounts of RNA (1  $\mu$ g, 2  $\mu$ g, or 4 $\mu$ g) were separated through 1% MOPS-formaldehyde-agarose. RNA samples were bottom-up transferred to Hybond-N (Amersham) membranes for 4-10 hours and UV cross linked. Blots were hybridized with  $\gamma$ [<sup>32</sup>P] 5'-end-labeled oligonucleotides specific for U3 snoRNA (loading control) and the exogenous *Renilla* fragment (experimental) for 4-24 hours at 52-60° in a rolling incubator. End-labelling was performed using the Roche T4 DNA kinase with 1  $\mu$ l

of enzyme and 30-50 mCi  $\gamma$ [ $^{32}\text{P}$ ]-ATP at 25° (this is explicitly in contradiction to the protocol listed in the manual for the enzyme which states 37°, but I found 5-10x more specific product at RT). After incubation, blots were rinsed 3-6 times in 2x SSC / 0.1% SDS for 15-30 minutes. Messenger RNAs were identified using a GeneStorm phosphoimager (Bio-Rad) and quantified using QuantifyOne (Bio-Rad). Due to the large size differences in the mRNAs observed, no stripping was performed, but blots were rinsed with 1x SSC / 0.2% SDS between probes. Blots were repeated three or more times and averaged to generate graphs.

### **Plasmid Construction**

Synthetic oligonucleotides (IDT DNA) used for plasmid construction are listed in appendices D and E. The plasmids generated are in appendices B and C. Insertions were amplified using PCR and ligated into appropriate backbone plasmids. Mutagenesis was performed using QuikChange (Agilent) site-directed mutagenesis kits. Clones were confirmed by sequencing (Genewiz). Examples of each cloning method and site-directed mutagenesis method follow:

#### **PCR Amplification, Restriction Enzyme Digestions, and Ligation.**

The PRF signal from Homo sapiens CCR5 was amplified from pCMV-XL4 (pJD819) containing the CCR5 open reading frame (Origene) using oligonucleotides with Bam HI and Sal I restriction sites. PCR products were ligated into p2luci (pJD175e)[204].

Choose oligonucleotides in two pieces: the first includes an appropriate spacer (often 4A), a restriction enzyme cutting site unique to both the cloned sequence and the multiple cloning site of the backbone plasmid; the second includes a region which is identical to the sequence of interest. Take note of the melting temperature ( $T_m$ ) of the identical sequence. This will define the PCR annealing temperature.

The 4-6 nucleotide spacer is important to ensure that restriction endonucleases efficiently cut the resulting dsDNA fragment. The length of the identical sequence is not very important, but should be long enough to be unique against the genome(s) of the organisms used to maintain the resulting plasmid. In practice, a  $T_m$  of 55-60° is often ideal. The most common mistake when choosing oligonucleotides is to inappropriately reverse or complement the 3' oligonucleotide. Dilute the lyophilized oligonucleotides to [100  $\mu$ M] when they arrive.

Perform a PCR reaction using 10-100 pMol of each oligonucleotide and 10-200 ng of template (usually plasmid). The annealing temperature should be 3-7° less than the  $T_m$  of the overlapping sequence. Annealing temperatures set too low result in non-specific products or primer-dimers, while too high results in no product. When using the Taq polymerase, the approximate rate of incorporation is 1 kb/min with an error rate of 1 base / 11,000 incorporated nucleotides. Thus products larger than 4 kb might require another polymerase.

The fragment should be gel-purified to remove unincorporated oligonucleotides. The gel-purification process uses TAE and may leave behind large amounts of EDTA; DNA ligase and some restriction enzymes require  $Mg^{+2}$ . Therefore, ethanol precipitate the PCR product and elute in 10-30  $\mu$ l water.

Digest the resulting dsDNA fragment and template plasmid with the appropriate restriction enzymes. Clean the resulting linearized plasmid and fragment before ligation. This may be done by PCR purification, gel purification, or ethanol precipitation. Ideally the final elution should result in a 3:1 insert:vector molar ratio (This ratio may change significantly depending on size and restriction site, blunt ends for example use 7:1 while very small fragments use 10:1). Using  $OD_{260}$  to quantify the insert:vector is unreliable if the fragment is very short (< 200 bp) due to a skewed extinction coefficient or

contaminants from gel purification. Thus it is often useful to quantify by electrophoresis and compare to a DNA ladder.

Ligate the fragment and plasmid and transform bacteria for amplification. While most protocols now use rapid ligation (5 minutes at 25°), I have found that difficult ligations work best when using T3 DNA ligase (NEB) overnight at 16°.

#### **Oligonucleotide Annealing and Ligation.**

Oligonucleotides encoding the -1 PRF signals from the hIL2 $\gamma$ , human and mouse IL7 $\alpha$ , hIL8 $\alpha$ , hIL8 $\beta$ , hIL22 $\alpha$ , and hIL27 $\alpha$  receptor chains were purchased containing 15-18 bases of internal overlap, PCR extended into a single Bam HI and Sal I containing product, and ligated into p2luci.

This process is effectively the same as above, but the oligonucleotides are chosen specifically to anneal to each other and the PCR product is explicitly intended to be “primer-dimer.” Therefore, the oligonucleotides chosen are longer and might need to be purified. Perform all ligations at a 10:1 molar ratio of insert:vector and use T3 DNA ligase (NEB) overnight.

#### **Site Directed Mutagenesis and “Mega-oligo” Mutagenesis.**

pTRE-R $\beta$  (pJD976), pTRE-R $\beta$ -ARE (pJD975), and pTET-Off (pJD979) were kind gifts from Dr. G. Brewer[207]. Insertions of the CCR5 -1 PRF signal were performed via a modified version of the “mega-oligo” site directed mutagenesis protocol[256].

Oligonucleotides (IDT) were chosen to include 23 nucleotides of the  $\beta$ -globin exon 1, 36 nucleotides of *Renilla* luciferase on the 5' side; and 23 nucleotides of  $\beta$ -globin exon 1, 15 nucleotides of firefly luciferase on the 3' side; and amplified from the CCR5 dual luciferase plasmid (pJD827). The resulting 330+ nucleotide oligos were used with pTRE-R $\beta$  template to generate pTRE-R $\beta$ -CCR5 and variants (pJD973, wt CCR5;

pJD974, wt CCR5 and ARE; pJD977, synonymous mutation of CCR5 slippery heptamer; and pJD1058, mutation of CCR5 slippery heptamer to GCGCGCG).

Choose primary oligonucleotides so that the 5' contains all of the mutagenized bases and the 3' is reverse complement of the template. Secondary oligonucleotide generation is performed via PCR using 10-100 pMoles each primary oligonucleotide and 10-100 ng template plasmid. 30-45 cycles of PCR are used with an appropriate annealing temperature (2-6° less than the  $T_m$  of the wild-type) and an appropriate extension time (often 30 seconds) at 64-68°.

The resulting short double-stranded DNA product is gel-purified, ethanol precipitated and eluted in 5-20  $\mu$ l water. 50-100% of this purified secondary oligonucleotide is used in the agilent Quikchange SDM protocol without modification.

### **Quantitative Real Time Reverse Transcription PCR.**

#### **Mammalian Cells**

For qPCR analyses of the  $\beta$ -globin based reporters, assays were performed as previously described [207] with the following modifications. The dual luciferase readthrough control was used for co-transfections rather than EGFP. RNA samples for qPCR were isolated using the RNAqueous kit (Ambion), digested with rDNase (Ambion) and assayed for DNA contamination using agarose gel electrophoresis and/or OD<sub>260/280</sub> measurements. The remaining samples were reverse transcribed using the iScript cDNA kit (Bio-Rad). The resulting cDNAs were diluted to 1:50-10,000 depending on mRNA concentration. Reactions were performed using 10  $\mu$ l of LightCycler 480 SYBR Green I Master mix (Roche), 0.2-0.3  $\mu$ M of each oligonucleotide, 2  $\mu$ l of cDNA, and water to 20  $\mu$ l/well. Best results were observed when all volumes were increased by 10%. All samples were assayed for genomic DNA contamination by performing the assay using wells containing 1-2  $\mu$ l of digested mRNA instead of cDNA. Reactions were



amplified using either a Roche 480 LightCycler or a Bio-Rad CFX 96 thermocycler as follows: 25°C for 10 seconds, 95°C for 5 min, followed by 45-60 cycles of 95°C for 10 seconds, 52°C for 15 seconds, and 72°C for 15 seconds. Melting curves were monitored by taking readings every 0.5°C from 55-95°C. The time-course qPCR analyses were performed with a 53°C and 54°C annealing temperatures and 20 second extension times with no significant changes in results. For qPCR analyses of the full length CCR5 mRNA, assays were performed as described for the  $\beta$ -globin assays, but using oligonucleotides specific for  $\beta$ -microglobulin and CCR5 [257]. Reactions were amplified using the same conditions as for the  $\beta$ -globin constructs except all reactions used 20 seconds at 55°C for extension.

### **Yeast Cells**

Full length *EST2* expression vectors (pJD641), *EST2* mutant vectors(pJD796), and null plasmids(pJD315) were transformed into WT(JD1281), *EST2* deletion(JD1287), *UPF2*(JD1288) and *EST2/UPF2*(JD1276) deletion strains. Total RNA was extracted with acid phenol/chloroform from mid-logarithmic cell cultures. To reduce nonspecific amplification, RNA was treated with DNase I before reverse transcription using Turbo DNase (Ambion). cDNA was generated using the Bio-Rad iScript cDNA synthesis kit and used in the LightCycler real-time PCR system. PCR was performed as in mammalian cell cultures with the following cycle conditions: 1 cycle of 95°C for 10 min; 40 cycles of 95°C for 10 s, 54°C for 20 s, and 72°C for 20 s. U3 snoRNA was chosen as a reference gene.

### **RNA Extraction**

#### **Mammalian Cells**

All RNA extractions of mammalian RNA were performed using either the Trizol reagent[258] or the Ambion RNAqueous kit with no modifications.

### **Yeast Cells**

RNA extractions from yeast require an initial step to disrupt the cell wall. This may be performed using a lysis buffer and agitating the cells in 200-500  $\mu$ l of glass beads for 1-5 minutes. Once this is complete, the rest of the extraction process continues with no modification.

### **SHAPE**

SHAPE was performed as previously described[215] with no known changes. RNA for SHAPE was generated by PCR amplifying T7 or T3 containing dsDNA from dual-luciferase plasmids and using the mMessage Machine transcription kit. The resulting RNA was passed through a sephadex G-25 column and ethanol precipitated before analysis.

### **Southern Blotting**

Southern blots were performed to observe telomere differences as previously described with changes[86]. Yeast genomic DNA was collected using the bust n' grab method[259]. Instead of end-label synthetic telomere-template oligonucleotides, the telomere template was PCR amplified from a plasmid (pJD972) which flanks the telomerase region with M13 and T7. ssRNA was transcribed using the mMessage Machine kit (Ambion) and  $\alpha$ [<sup>32</sup>P] ATP to generate probes with high specific activity. Unincorporated nucleotides were removed by passing the probe through a sephadex G-25 (or G-50) column. Yeast genomic DNA was cut with Pst I (Xho I when Pst I was exhausted) and electrophoresed through a 1% agarose gel. DNA was nicked in a crosslinker and denatured in 0.5 M NaOH, 1.5 M NaCl for 30-45 minutes. The gel was rinsed in water before placing in a blotting solution of 0.2 M NaOH, 1.5 M NaCl while preparing a bottom-up blotting apparatus (identical to that used for Northern blots) with a Hybond N+ membrane. Blots transferred overnight, were rinsed 2-3 times in 2x SSC

and crosslinked. After crosslinking, blots were pre-incubated with Church's buffer at 65° for 15-60 minutes before adding the probe. Blots hybridized overnight at 57-60°, washed 4 times with 1x SSC, 0.5% SDS followed by 1-3 washes with 0.1x SSC, 0.1% SDS. The blots were wrapped in a plastic baggie, sealed, exposed to a phosphorimager screen, and analyzed as per a Northern. Church's buffer contains: 2 ml 0.5 M EDTA pH 8.0 (1 mM), 2 ml 85% (w/v) H<sub>3</sub>PO<sub>4</sub>, 67.5 g Na<sub>2</sub>HPO<sub>4</sub>, 70 g SDS, and water to 1 L. On at least one occasion 200 µl ssDNA was added to the Church's buffer with no noticeable effect.

## **Time Course Assays**

### **Mammalian**

mRNA decay time course assays were performed as previously described with minor changes[207]. RNA isolations were performed immediately at each timepoint after transcriptional arrest using the RNeasy kit (Qiagen) rather than after freezing samples on dry ice.

### **Yeast**

mRNA decay time course assays were performed as previously described with minor changes[199]. The EST2 mRNA stability vector (pJD754), a readthrough control (pJD753) and premature stop codon containing control (pJD828) were transformed into *rpb1-1* (JD977) and *rpb1-1/Upf1Δ* cells.

Transformations of *rpb1* mutant cells must be performed entirely at room temperature, otherwise the temperature sensitive RNA polymerase may kill the cells. In order to accommodate this strain, transformations were performed with 2-3x more cell mass than usual, the ssDNA was boiled and then allowed to return partially to room temperature before addition, and the incubation with plasmid before plating was extended to 6-8 hours. Transformed cells were grown at 25° or 30° with no noticeable

difference. At collection time, media was pre-warmed to 52° and added so that the average temperature rose to 42°, the non-permissive temperature. Cells were collected and frozen immediately in liquid N<sub>2</sub>. mRNA was collected and Northern analyses/qPCR were performed as usual.

## **Transfections**

### **Plasmid**

All plasmid transfections were performed using the Fugene 6 (initially Roche, now Promega) reagent. Most efficient transfections were observed using the 3:1 ratio.

Allow cells to grow to 50-100% confluency, trypsinize and split into 6-24 well plates. Let them grow in the plates to 50-75% confluency. Assuming 24 well plates, refresh the media with 500 µl DMEM+ (DMEM supplemented with 10% irradiated fetal bovine serum (FBS), 1x non-essential amino acids (NEAA), and [4 mM] glutamine). In eppendorf tubes, mix 20 µl DMEM(without additives), 0.6 µl Fugene, and 300 ng plasmid, in that order for every well transfected. Let sit for 30-90 minutes and add dropwise to each well. Wait 1-2 days and assay.

### **miRNA**

Cells were transfected with the following miRNA precursors: hsa-miR-141, hsa-miR-711, hsa-miR-1224-5p, and hsa-miR-1205 using the siPORT reagent (Applied Biosystems/Ambion), or Lipofectamine for miRNA pulldowns (Invitrogen). When performing initial miRNA transfections for dual luciferase, [30 nM] was used. When performing the miRNA titration, four 1:10 dilutions were used starting at [5 nM]. Transfections were performed into 15,000-40,000 cells in 500 µl DMEM+ using 25 µl of DMEM and 1 µl of siPORT reagent after incubating for 20 minutes at room temperature per well. Media was replaced with fresh DMEM+ after 8-12 hours. Dual luciferase

plasmid transfections were either performed at the same time or 24 hours later using the FuGene 6 reagent.

### **siRNA**

Cells were transfected with synthetic RNA oligonucleotides specific to the mRNA of interest (AUF1, hRENT1, hUPF2, hSMG1, argonaute, hDKC1) or scrambled oligonucleotides using the HiPerFect transfection reagent (Qiagen). Initial transfections were performed at [1, 5, 10 nM], and [20 nM] for optimization. Most final transfections were performed at [10 nM], but some used [5 nM]. Transfections were performed into 30,000-40,000 cells (via hemacytometer) immediately after splitting cells in 500  $\mu$ l of DMEM+ using 100  $\mu$ l of DMEM and 3  $\mu$ l of HiPerFect reagent after incubating for 15-20 minutes at room temperature. Media was replaced with fresh DMEM+ after 8-12 hours. Assays were performed 48-72 hours after siRNA transfection. When other plasmids were also transfected, they were performed separately 24-48 hours after siRNA transfection using the Fugene 6 (Roche) reagent.

## **Transformations**

### **Bacteria**

Transformations of *E. coli* were performed as described previously using the calcium chloride method[196]. When transformations were performed for difficult plasmids (larger than 10,000 bp or SDM products), a slightly modified method was used as described in the QuikChange SDM manual. Incubate 20-50  $\mu$ l of cells with 1-4  $\mu$ l  $\beta$ -mercaptoethanol for 10 minutes on ice. Add plasmid, incubate on ice for 30-120 minutes. Heat shock at 42° for 30 seconds. Incubate again on ice for 10 minutes. Add 100-300  $\mu$ l pre-warmed LB, incubate at 37° for 1 hour, plate.

## Yeast

Yeast cells were transformed using the alkali cation method[197] with some changes. First check yeast genetic background and plasmid maps to ensure that the proper auxotrophies are maintained. Grow 2-3 ml yeast overnight in appropriate non-selective (for your plasmid) media. Start an aliquot of salmon sperm ssDNA boiling. Aliquot yeast into 1.6 ml eppendorf tubes and spin down at  $< 500$  g. (2,000 rpm on a desktop centrifuge, 1,200 rpm on the Sorval at maximum.) Remove media and resuspend in 200  $\mu$ l 0.1 M LiOAc/TE, repeat and resuspend in 100  $\mu$ l 0.1 M LiOAc/TE. Add 15  $\mu$ l ssDNA for each transformation (this is approximately double the original protocol). Add 500-2,000 ng plasmid and 500-700  $\mu$ l PEG/LiOAc/TE (this consists of 8 ml PEG 5,000, 1 ml 1.0 M LiOAc, and 1 ml 10x TE, which is more PEG than the default protocol due to a labeling mistake, but seems to work better). Incubate 90-900 minutes at 30°. Heat shock at 42° for 15 minutes. Pellet the cells, rinse and resuspend them in 100-400  $\mu$ l water and plate. Some strains (as noted) were performed at 25° without a heat shock.

## Appendices

### Appendix 1: Yeast Strains Used

Strain number	Phenotype	Purpose
JD1158	BY4742 <i>MAT<math>\alpha</math> his3<math>\Delta</math>1 leu2<math>\Delta</math>0 lys2<math>\Delta</math>0 ura3<math>\Delta</math>0</i>	WT strain for steady states assays and frameshifting
JD1181	BY4742 <i>MAT<math>\alpha</math> upf3::Kan<sup>R</sup> his3<math>\Delta</math>1 leu2<math>\Delta</math>0 lys2<math>\Delta</math>0 ura3<math>\Delta</math>0</i>	NMD deficient strain for steady states
JD1367	yRP2077 <i>MAT<math>\alpha</math> upf1::KanMX4 his3<math>\Delta</math> leu2<math>\Delta</math> met15<math>\Delta</math> ura3<math>\Delta</math></i>	NMD deficient strain for steady states, Parker lab
JD1170	<i>MAT<math>\alpha</math> xrn1::Kan<sup>R</sup> his3<math>\Delta</math>1 leu2<math>\Delta</math>0 ura3<math>\Delta</math>0</i>	5'→3' decay deficient strain for steady states
JD1345	<i>MAT<math>\alpha</math> ski2::Kan<sup>R</sup> his3<math>\Delta</math>1 leu2<math>\Delta</math>0 ura3<math>\Delta</math>0</i>	Exosome / 3' → 5' deficient for steady states
JD19	<i>MAT<math>\alpha</math> leu2<math>\Delta</math>0 ade2<math>\Delta</math>0 ura3<math>\Delta</math>0 PEP4::HIS3 NUC1::LEU2 ski3 [L-AHN M<sub>1</sub>]</i>	Exosome / 3' → 5' deficient for steady states
JD1122	<i>MAT<math>\alpha</math> lys2<math>\Delta</math>0 ura3<math>\Delta</math>0 ho::LYS2 leu2::hisG, his4b dcp1::hisG [L-AHN M<sub>1</sub>]</i>	Decapping deficient for steady states
JD1363	yRP2056 <i>MAT<math>\alpha</math>; his3<math>\Delta</math> leu2<math>\Delta</math> met15<math>\Delta</math> ura3<math>\Delta</math> dom34::KanMX4</i>	No-go decay deficient for steady states, Parker lab
JD977	<i>MAT<math>\alpha</math> ura3-52 trp1-<math>\Delta</math>1 his4-38 leu2-1 rpb1-1</i>	RNA pol2 $\beta$ , temperature sensitive for time courses
JD978	<i>MAT<math>\alpha</math> ura3-52 trp1-<math>\Delta</math>1 his4-38 leu2-1 rpb1-1 upf1<math>\Delta</math>::HISG</i>	RNA pol2 $\beta$ , temperature sensitive and NMD deficient
JD1263	<i>MAT a/<math>\alpha</math> trp1<math>\Delta</math> /trp1<math>\Delta</math> leu2<math>\Delta</math>/leu2<math>\Delta</math> can1-100/can1-100 ade2-1/ade2-1 est2<math>\Delta</math>::URA3/EST2 upf2<math>\Delta</math>::HIS3/UPF2 VR-ADE2-TEL/VR-TEL (YJB2659)</i>	Parental strain of <i>est2</i> mutant strains used in this study, Berman lab
JD1281	<i>MAT a trp1<math>\Delta</math> leu2<math>\Delta</math> can1-100 ade2-1 ura3-1 his3-11 VR-ADE2-TEL</i>	WT strain for assaying telomere activity
JD1287	<i>MAT <math>\alpha</math> trp1<math>\Delta</math> leu2<math>\Delta</math> can1-100 ade2-1 est2<math>\Delta</math>::URA3 his3-11 VR-ADE2-TEL</i>	Telomerase deficient strain to assay telomeres
JD1288	<i>MAT a trp1<math>\Delta</math> leu2<math>\Delta</math> can1-100 ade2-1 ura3-1 upf2<math>\Delta</math>::HIS3 VR-ADE2-TEL</i>	NMD deficient strain to assay telomeres
JD1276	<i>MAT a trp1<math>\Delta</math> leu2<math>\Delta</math> can1-100 ade2-1 est2<math>\Delta</math>::URA3 upf2<math>\Delta</math>::HIS3 VR-ADE2-TEL</i>	Telomerase and NMD deficient strain to assay telomeres

**Table 6: Yeast strains used in Chapter 3.**

## Appendix 2: Yeast Plasmids

Plasmid Name	Backbone Plasmid	Insertion
<b>pJD0375</b> [200]	pRS316[260]	Dual luciferase cassette in pRS316, genomic context (b)
<b>pJD0376</b> [200]	pJD0375	LA frameshift signal sequence inserted in dual luciferase reporter, viral context (b)
<b>pJD0520</b> [23]	pJD0375	SPR6 PRF sequence inserted in dual luciferase reporter, viral context (c,d)
<b>pJD0521</b> [23]	pJD0375	EST2 PRF sequence inserted in dual luciferase reporter, viral context (c,d)
<b>pJD0519</b> [23]	pJD0375	BUB3 PRF sequence inserted in dual luciferase reporter, viral context (c,d,a)
<b>pJD0518</b> [23]	pJD0375	TBF1 PRF sequence inserted in dual luciferase reporter, viral context (c,d)
<b>pJD0476</b> [23]	pJD0375	PPR1 PRF sequence inserted in dual luciferase reporter, viral context (e,a)
<b>pJD0477</b> [23]	pJD0375	NUP82 PRF sequence inserted in dual luciferase reporter, viral context (e)
<b>pJD0478</b> [23]	pJD0375	TBF1 PRF sequence inserted in dual luciferase reporter, viral context (e)
<b>pJD0522</b>	pJD0375	FLR1 PRF sequence inserted in dual luciferase reporter, viral context (e)
<b>pJD0523</b>	pJD0375	FKS1 PRF sequence inserted in dual luciferase reporter, viral context (e)
<b>pJD0624</b>	pJD0375	NUP82 PRF sequence inserted in dual luciferase reporter, genomic context (e)
<b>pJD0625</b>	pJD0375	CTS2 PRF sequence inserted in dual luciferase reporter, genomic context (e)
<b>pJD0626</b>	pJD0375	SPR6 PRF sequence inserted in dual luciferase reporter, genomic context (e)
<b>pJD0753</b> [202]	pJD0741[78]	Readthrough containing small amounts of Renilla and Firefly in PGK1 reporter (e)
<b>pJD0765</b> [202]	pJD0753	Premature termination codon in PGK1 reporter (e,a)
<b>pJD0748</b> [202]	pJD0753, pJD0520	SPR6 PRF sequence inserted in PGK1 reporter (a)
<b>pJD0938</b> [202]	pJD0753, pJD0520	SPR6 PRF sequence inserted in PGK1 reporter v2 (a,f)
<b>pJD0754</b> [202]	pJD0753, pJD0521	EST2 PRF sequence inserted in PGK1 reporter (a)
<b>pJD0755</b> [202]	pJD0753, pJD0519	BUB3 PRF sequence inserted in PGK1 reporter (a,e)
<b>pJD0756</b> [202]	pJD0753, pJD0518	TBF1 PRF sequence inserted in PGK1 reporter (a)
<b>pJD0818</b> [202]	pJD0753, pJD0520	SPR6 slipsite mutant in PGK1 reporter (a)
<b>pJD0806</b> [202]	pJD0753, pJD0521	EST2 slipsite mutant in PGK1 reporter (a)
<b>pJD0812</b> [202]	pJD0753, pJD0519	BUB3 slipsite mutant in PGK1 reporter (a)
<b>pJD0813</b> [202]	pJD0753, pJD0518	TBF1 slipsite mutant in PGK1 reporter (a)



<b>pJD0638</b> [188]	YEplac128	Full length EST2 high copy plasmid
<b>pJD0641</b>	pJD0638	Full length EST2 low copy plasmid (e)
<b>pJD0796</b> [202]	pJD0641	Mutated full length EST2 low copy plasmid (a,e)
<b>pJD0972</b> [188]	pTNT	Yeast telomeric sequence under T7 promoter control
<b>pJD0659</b>	pJD0375	EST2 PRF sequence (position 72) inserted in dual luciferase reporter, genomic context (e,a)
<b>pJD0660</b>	pJD0375	EST2 PRF sequence (position 1215) inserted in dual luciferase reporter, genomic context (e,a)
<b>pJD0661</b>	pJD0375	EST2 PRF sequence (position 1326) inserted in dual luciferase reporter, genomic context (a,e)
<b>pJD0662</b>	pJD0375	EST2 PRF sequence (position 1653) inserted in dual luciferase reporter, genomic context (e,a)
<b>pJD0667</b>	pJD0375	EST2 PRF sequence (position 1995) inserted in dual luciferase reporter, genomic context (a,e)
<b>pJD0673</b>	pJD0641	Full length EST2 low copy plasmid with mutations to the position 1215 PRF sequence (e)
<b>pJD0674</b>	pJD0641	Full length EST2 low copy plasmid with mutations to the position 1995 PRF sequence (e)
<b>pJD0685</b>	pJD0641	Full length EST2 low copy plasmid with mutations to the position 1326 PRF sequence (a)
<b>pJD0686</b>	pJD0641	Full length EST2 low copy plasmid with mutations to the position 72 PRF sequence (a)
<b>pJD0742</b>	pJD0375	PPR1 PRF sequence (position 660) inserted in dual luciferase reporter, viral context (a)
<b>pJD0743</b>	pJD0375	PPR1 PRF sequence (position 1188) inserted in dual luciferase reporter, viral context (a)
<b>pJD0744</b>	pJD0375	PPR1 PRF sequence (position 1260) inserted in dual luciferase reporter, viral context (a)
<b>pJD0745</b>	pJD0375	PPR1 PRF sequence (position 2094) inserted in dual luciferase reporter, viral context (a)
<b>pJD0746</b>	pJD0375	PPR1 PRF sequence (position 525) inserted in dual luciferase reporter, viral context (a)
<b>pJD0757</b>	pJD0753	CTS2 PRF sequence inserted in PGK1 reporter (a)
<b>pJD0758</b>	pJD0753	PPR1 PRF sequence (position 660) inserted in PGK1 reporter (a)
<b>pJD0759</b>	pJD0753	PPR1 PRF sequence (position 1188) inserted in PGK1 reporter (a)
<b>pJD0760</b>	pJD0753	PPR1 PRF sequence (position 1260) inserted in PGK1 reporter (a)
<b>pJD0761</b>	pJD0753	PPR1 PRF sequence (position 2094) inserted in PGK1 reporter (a)
<b>pJD0766</b>	pJD0753	EST2 PRF sequence (position 1215) inserted in PGK1 reporter (a)
<b>pJD0797</b>	pJD0375	SPR6 PRF sequence inserted in dual luciferase reporter, viral context, mutated slip site (a)
<b>pJD0806</b>	pJD0754	EST2 PRF sequence inserted in PGK1 reporter, slip site mutant (a)
<b>pJD0807</b>	pJD0757	CTS2 PRF sequence inserted in PGK1 reporter, slip site mutant (a)
<b>pJD0808</b>	pJD0521	EST2 PRF sequence inserted in dual luciferase reporter, viral context, mutated slip site (l)
<b>pJD0811</b>	pJD0754	EST2 PRF sequence inserted in PGK1 reporter, stem 1 mutant (a)
<b>pJD0812</b>	pJD0755	BUB3 PRF sequence inserted in PGK1 reporter, slip site mutant (a)
<b>pJD0813</b>	pJD0756	TBF1 PRF sequence inserted in PGK1 reporter, slip site mutant (a)
<b>pJD0814</b>	pJD0756	TBF1 PRF sequence inserted in PGK1 reporter, stem 1 mutant (a)
<b>pJD0815</b>	pJD0757	CTS2 PRF sequence inserted in PGK1 reporter, stem 1 mutant (a)
<b>pJD0816</b>	pJD0759	PPR1 PRF sequence inserted in PGK1 reporter, slip site mutant (a)
<b>pJD0817</b>	pJD0748	SPR6 PRF sequence inserted in PGK1 reporter, stem 1 mutant (a)
<b>pJD0818</b>	pJD0748	SPR6 PRF sequence inserted in PGK1 reporter, slip site mutant (a)
<b>pJD0859</b>	pJD0375	PDR5 PRF sequence (position 300) inserted in dual luciferase

		reporter, viral context (a)
<b>pJD0860</b>	pJD0837	PDR5 PRF sequence (position 3027) inserted in dual luciferase reporter, viral context (a)
<b>pJD0972</b>	pBC6	Yeast telomeric sequence flanked by M13
<b>pJD0984</b>	pJD0641	Full length EST2 low copy plasmid with mutations to the positions 72,1215 PRF sequence (a,j)
<b>pJD0985</b>	pJD0641	Full length EST2 low copy plasmid with mutations to the positions 72,1215,1326,1653 PRF sequence (a,j)
<b>pJD0986</b>	pJD0641	Full length EST2 low copy plasmid with mutations to the positions 72,1215,1326,1995 PRF sequence (a,j)
<b>pJD0987</b>	pJD0641	Full length EST2 low copy plasmid with mutations to the positions 72,1326,1653 PRF sequence (a,j)
<b>pJD0988</b>	pJD0641	Full length EST2 low copy plasmid with mutations to the positions 72,1215,1995 PRF sequence (a,j)
<b>pJD0989</b>	pJD0641	Full length EST2 low copy plasmid with mutations to the positions 72,1326 PRF sequence (a,j)
<b>pJD990</b>	pJD0641	Full length EST2 low copy plasmid with mutations to the positions 72,1215,1653 PRF sequence (a,j)
<b>pJD0991</b>	pJD0641	Full length EST2 low copy plasmid with mutations to the positions 72,1326,1653,1995 PRF sequence (a,j)
<b>pJD0992</b>	pJD0641	Full length EST2 low copy plasmid with mutations to the positions 1653 PRF sequence (a,j)
<b>pJD0993</b>	pJD0641	Full length EST2 low copy plasmid with mutations to the positions 72,1995 PRF sequence (a,j)
<b>pJD0994</b>	pJD0641	Full length EST2 low copy plasmid with mutations to the positions 1215,1326 PRF sequence (a,j)
<b>pJD0995</b>	pJD0984	Full length EST2 low copy plasmid with mutations to the positions 1215,1326 PRF sequence (a,j)
<b>pJD0996</b>	pJD0989	Full length EST2 low copy plasmid with mutations to the positions 72,1326,1995 PRF sequence (a,j)
<b>pJD0997</b>	pJD0674	Full length EST2 low copy plasmid with mutations to the positions 1215,1995 PRF sequence (a,j)
<b>pJD0998</b>	pJD0990	Full length EST2 low copy plasmid with mutations to the positions 72,1215,1653,1995 PRF sequence (a,j)
<b>pJD0999</b>	pJD0673	Full length EST2 low copy plasmid with mutations to the positions 72,1653 PRF sequence (a,j)
<b>pJD1000</b>	pJD0375	STN1 PRF sequence (position 885) inserted in dual luciferase reporter, genomic context (a)
<b>pJD1001</b>	pJD0375	STN1 PRF sequence (position 1203) inserted in dual luciferase reporter, genomic context (a)
<b>pJD1002</b>	pJD0375	EST1 PRF sequence (position 1203) inserted in dual luciferase reporter, genomic context (a)
<b>pJD1003</b>	pJD0375	EST1 PRF sequence (position 1272) inserted in dual luciferase reporter, genomic context (a)
<b>pJD1004</b>	pJD0375	EST1 PRF sequence (position 1920) inserted in dual luciferase reporter, genomic context (a)
<b>pJD1005</b>	pJD0375	CDC13 PRF sequence (position 1270) inserted in dual luciferase reporter, genomic context (a)
<b>pJD1010</b>	pRS316	CBF5 wild-type in pRS316 (k,a)
<b>pJD1011</b>	pRS316	CBF5 S36A in pRS316 (k,a)
<b>pJD1012</b>	pRS316	CBF5 D95A in pRS316 (k,a)
<b>pJD1013</b>	pJD0685	Full length EST2 low copy plasmid with mutations to the position 1326,1995 PRF sequence (a,j)

<b>pJD1014</b>	pJD0992	Full length EST2 low copy plasmid with mutations to the position 1326,1995 PRF sequence (j,a)
<b>pJD1015</b>	pJD0999	Full length EST2 low copy plasmid with mutations to the position 1326,1995 PRF sequence (a,j)
<b>pJD1016</b>	pJD0994	Full length EST2 low copy plasmid with mutations to the position 1215,1326,1995 PRF sequence (j,a)
<b>pJD1017</b>	pJD0659	EST2 PRF sequence (position 72) inserted in dual luciferase reporter, viral context (a)
<b>pJD1018</b>	pJD0660	EST2 PRF sequence (position 1215) inserted in dual luciferase reporter, viral context (a)
<b>pJD1019</b>	pJD0661	EST2 PRF sequence (position 1326) inserted in dual luciferase reporter, viral context (a)
<b>pJD1020</b>	pJD0667	EST2 PRF sequence (position 1995) inserted in dual luciferase reporter, viral context (a)
<b>pJD1021</b>	pJD0659	EST2 PRF sequence (position 72) inserted in PGK1 reporter (a)
<b>pJD1022</b>	pJD0992	Full length EST2 low copy plasmid with mutations to the positions 72,1215,1995 PRF sequence (j,a)
<b>pJD1023</b>	pJD1014	Full length EST2 low copy plasmid with mutations to the position 72,1326,1995 PRF sequence (a,j) CHECK THIS AGAIN
<b>pJD1024</b>	pJD1014	Full length EST2 low copy plasmid with mutations to the position 1215,1326,1995 PRF sequence (j,a)
<b>pJD1025</b>	pJD0994	Full length EST2 low copy plasmid with mutations to the positions 1215,1326,1653 PRF sequence (a,j)
<b>pJD1026</b>	pJD0992	Full length EST2 low copy plasmid with mutations to the positions 1326,1653 PRF sequence (j,a)
<b>pJD1027</b>	pJD1016	Full length EST2 low copy plasmid with mutations to the position 1215,1326,1653,1995 PRF sequence (j,a)
<b>pJD1028</b>	pJD0375	FKS1 PRF sequence inserted in dual luciferase reporter, genomic context (a)
<b>pJD1029</b>	pJD0375	FLR1 PRF sequence inserted in dual luciferase reporter, genomic context (a)
<b>pJD1030</b>	pJD0375	PPR1 PRF sequence (position 660) inserted in dual luciferase reporter, genomic context (a)
<b>pJD1031</b>	pJD0375	PPR1 PRF sequence (position 1260) inserted in dual luciferase reporter, genomic context (a)
<b>pJD1032</b>	pJD0375	PPR1 PRF sequence (position 2094) inserted in dual luciferase reporter, genomic context (a)
<b>pJD1036</b>	pRS314	CBF5 wild-type in pRS314 (a)
<b>pJD1037</b>	pRS314	CBF5 D95A in pRS314 (a)
<b>pJD1038</b>	pJD1000	STN1 PRF sequence (position 885) inserted in dual luciferase reporter, viral context (a)
<b>pJD1039</b>	pJD1001	STN1 PRF sequence (position 1203) inserted in dual luciferase reporter, viral context (a)
<b>pJD1040</b>	pJD1002	EST1 PRF sequence (position 1203) inserted in dual luciferase reporter, viral context (a)
<b>pJD1041</b>	pJD1003	EST1 PRF sequence (position 1272) inserted in dual luciferase reporter, viral context (a)
<b>pJD1043</b>	pJD1004	EST1 PRF sequence (position 1920) inserted in dual luciferase reporter, viral context (a)
<b>pJD1048</b>	pRS313	CBF5 wild-type in pRS313 (a,l)
<b>pJD1049</b>	pRS313	CBF5 D95A in pRS313 (a,l)
<b>pJD1070</b>	pJD0753	EST2 PRF sequence (position 72) inserted in PGK1 reporter (a)
<b>pJD1071</b>	pJD0753	EST2 PRF sequence (position 1326) inserted in PGK1 reporter (a)

<b>pJD1072</b>	pJD0753	EST2 PRF sequence (position 1995) inserted in PGK1 reporter (a)
<b>pJD1073</b>	pJD0753	STN1 PRF sequence (position 885) inserted in PGK1 reporter (l,a)
<b>pJD1074</b>	pJD0753	STN1 PRF sequence (position 1203) inserted in PGK1 reporter (l,a)
<b>pJD1075</b>	pJD0753	EST1 PRF sequence (position 1203) inserted in PGK1 reporter (l,a)
<b>pJD1076</b>	pJD0753	EST1 PRF sequence (position 1272) inserted in PGK1 reporter (l,a)
<b>pJD1077</b>	pJD0753	EST1 PRF sequence (position 1920) inserted in PGK1 reporter (l,a)

**Table 7: Plasmids used with yeast**

The complete sequences of these plasmids may be found at <http://github.com/abelew/plasmids/>. “Genomic context” notes that the downstream ORF is in the same reading frame as the upstream ORF; “viral context” notes that the downstream ORF is in the -1 frame with respect to the upstream ORF.

Created by: **a.** Ashton Belew; **b.** Jason Harger; **c.** Rasa Rakauskaitė; **d.** Ewan Plant; **e.** Jonathan Jacobs; **f.** Hamid-Reza Shahshahan; **g.** Sharmishtha Musalgaonkar; **h.** Bryan Fleming; **i.** Lara Hause; **j.** Curt Kugel; **k.** Rachel Niederer; **l.** Vivek Advani; **m.** Sergey Sulima; **n.** Jessica Neirermeier

### Appendix 3: Mammalian Plasmids

Plasmid Name	Backbone Plasmid	Insertion
pJD0175e	P2luc[204]	pJD175e is identical to p2luc
pJD0175f	P2luci [204]	pJD175f is identical to p2luci
pJD0187.wt[261]	P2luci	Insertion of the HIV-1 PRF signal into the dual luciferase plasmid
pJD0827	pJD0175e	Insertion of the <i>CCR5</i> PRF signal into the dual luciferase plasmid (a)
pJD0835	pJD0175e	Insertion of the <i>homo sapiens IL7<math>\alpha</math></i> PRF signal into the dual luciferase plasmid (g,a)
pJD0836	pJD0175e	Insertion of the <i>mus musculus IL7<math>\alpha</math></i> PRF signal into the dual luciferase plasmid (i,a)
pJD0844	pJD0827	Mutation of the <i>CCR5</i> slip site in the dual luciferase plasmid (n,a)
pJD0845	pJD0827	Mutation of the <i>CCR5</i> stem 2, 5' side in the dual luciferase plasmid (n,a)
pJD0846	pJD0827	Mutation of the <i>CCR5</i> stem 2, 3' side in the dual luciferase plasmid (n,a)
pJD0847	pJD0175e	Insertion of the <i>homo sapiens IL27</i> PRF signal into the dual luciferase plasmid (a)
pJD0848	pJD0827	Complementation mutation of the <i>CCR5</i> stem 2 in the dual luciferase plasmid (h,a)
pJD0850	pJD0827	Partial mutation (3 bases) of the <i>CCR5</i> stem 1, 3' side in the dual luciferase plasmid (h,a)
pJD0851	pJD0827	Partial mutation (3 bases) of the <i>CCR5</i> stem 1, 5' side in the dual luciferase plasmid (h,a)
pJD0852	pJD0827	Complementation mutation (3 bases) of the partial <i>CCR5</i> stem 1 in the dual luciferase plasmid (h,a)
pJD0854	pJD0827	Partial mutation (9 bases) of the <i>CCR5</i> stem 1, 5' side in the dual luciferase plasmid (h,a)
pJD0855	pJD0827	Complementation mutation (9 bases) of the partial <i>CCR5</i> stem 1 in the dual luciferase plasmid (h,a)
pJD0856	pJD0827	Partial mutation (12 bases) of the <i>CCR5</i> stem 2, 5' side in the dual luciferase plasmid (h,a)
pJD0857	pJD0827	Partial mutation (12 bases) of the <i>CCR5</i> stem 2, 3' side in the dual luciferase plasmid (h,a)
pJD0858	pJD0827	Complementation mutation (12 bases) of the partial <i>CCR5</i> stem 2 in the dual luciferase plasmid (h,a)
pJD0973	pTET	Rabbit <i>B-globin</i> reporter containing <i>CCR5</i> PRF signal in exon 1 (a,m)
pJD0974	pTET	Rabbit <i>B-globin</i> reporter containing <i>CCR5</i> PRF signal in exon 1 and the <i>TNF-<math>\alpha</math></i> ARE immediately after the stop (a)
pJD0975 [207]	pTET	Rabbit <i>B-globin</i> reporter containing the <i>TNF-<math>\alpha</math></i> ARE immediately after the stop
pJD0976 [207]	pTET	Rabbit <i>B-globin</i> reporter, readthrough
pJD0977	pTET	Rabbit <i>B-globin</i> reporter containing a mutated <i>CCR5</i> PRF signal in exon 1 (a)
pJD0978	pTET	Rabbit <i>B-globin</i> reporter containing mutated <i>CCR5</i> PRF signal in exon 1 and the <i>TNF-<math>\alpha</math></i> ARE immediately after the stop (a)

<b>pJD0979</b> [207]	pCMV	The Cytomegalovirus transactivator protein used to increase
<b>pJD1009</b>	pJD175e	Insertion of the <i>Dyskerin</i> PRF signal into the dual luciferase plasmid (j,a)
<b>pJD1033</b>	pEGFP	GFP plasmid from Clontech
<b>pJD1058</b>	pJD973	Rabbit <i>B-globin</i> reporter containing mutated <i>CCR5</i> PRF signal to GCGCGCG in exon 1 (a)
<b>pJD1059</b>	pJD1009	Rabbit <i>B-globin</i> reporter containing <i>dyskerin</i> PRF signal in exon 1 (j,a)
<b>pJD1060</b>	pJD827	Insertion of the <i>CCR5</i> PRF signal into the dual luciferase plasmid with slip site mutation to GCGCGCG (a)
<b>pJD1078</b>	pJD827	Insertion of the <i>CCR5</i> PRF signal into the dual luciferase plasmid with a stop codon immediately after <i>Renilla</i> (a)
<b>pJD1214</b>	pJD175f	Insert OLA1(position 513) PRF signal into dual luciferase plasmid
<b>pJD1215</b>	pJD175f	Insert PNMA1(position 1740) PRF signal into dual luciferase plasmid
<b>pJD1216</b>	pJD175f	Insert NT5DC(position 841) PRF signal into dual luciferase plasmid
<b>pJD1217</b>	pJD175f	Insert SLC38A(position 981) PRF signal into dual luciferase plasmid
<b>pJD1218</b>	pJD175f	Insert MCM5(position 1004) PRF signal into dual luciferase plasmid
<b>pJD1219</b>	pJD175f	Insert MYO5C(position 3527) PRF signal into dual luciferase plasmid
<b>pJD1220</b>	pJD175f	Insert TDO2(position 744) PRF signal into dual luciferase plasmid
<b>pJD1221</b>	pJD175f	Insert ZGPAT(position 664) PRF signal into dual luciferase plasmid
<b>pJD1224</b>	pJD175f	Insert ATIC(position 1656) PRF signal into dual luciferase plasmid
<b>pJD1225</b>	pJD175f	Insert GABRE(position 735) PRF signal into dual luciferase plasmid
<b>pJD1226</b>	pJD175f	Insert GOLT1B(position 437) PRF signal into dual luciferase plasmid

**Table 8: Mammalian plasmid list**

The complete sequences of these plasmids may be found at: <http://github.com/abelew/plasmids>

Created by: a. Ashton Belew; b. Jason Harger; c. Rasa Rakauskaitė; d. Ewan Plant; e. Jonathan Jacobs; f. Hamid-Reza Shahshahan; g. Sharmishtha Musalgaonkar; h. Bryan Fleming; i. Lara Hause; j. Curt Kugel; k. Rachel Niederer; l. Vivek Advani; m. Sergey Sulima; n. Jessica Neirermeir

## Appendix 4: Cloning Oligonucleotides

Order #	Name	Clone	Sequence
10882362	YOR026 W, Forward ( <i>BUB3</i> )	pJD0521	TCGACAAAAAATTTTCGCCAATTTAACGAAGACAGCGTGGTTA AAATTGCTTGTTTCGGACG
10882363	YOR026, reverse	pJD0521	GATCCGTCCGAACAAGCAATTTTAACCACGCTGTCTTCGTTA AATTGGCGAAATTTTTTG
10882364	YER115C, forward ( <i>SPR6</i> )	pJD0520	TCGACAAAAAAATAAGGAAACCAATCACTCTGGAGCATGG TTGCTTGTCAGGACCCGTGACTCTACGTTTCGGAAATTTTGCA GGAATCAGAGAG
10882365	YER115C, reverse	pJD0520	GATCCTCTCTGATTCTTGCAAAATTTCCGAAACGTAGAGTCA CGGGTCCTGACAAGCAACCATGCTCCAGAGTGATTGGTTTCC TTATTTTTTTTG
10882366	YLR318W , forward ( <i>EST2</i> )	pJD0521	TCGACAAAAAATCAAATGGGTTTTTCGTTAGATCTCAATATTT CTTCAATACCAATACAGGTGTATTGAAGTTATTTAATGTTGTT AACGCTG
10882367	YLR318W , reverse	pJD0521	GATCCAGCGTTAACAACATTAAATAACTTCAATACACCTGTA TTGGTATTGAAGAAATATTGAGATCTAACGAAAAACCCATT GATTTTTTG
10882368	YBR008C, forward ( <i>FLR1</i> )	pJD0522	TCGACAAAAAATCATCTTTTCAGGGTGGATTGGAACGGCCCCA GTGATCCTGAGAACCCACAAAACCTGGCCCCG
10882369	YBR008C, reverse	pJD0522	GATCCGGGCCAGTTTTGTGGGTTCTCAGGATCACTGGGGCCG TTCCAATCCACCCTGAAAGATGATTTTTTG
10882370	YLR342W , forward ( <i>FKS1</i> )	pJD0523	TCGACAAATTTCCACTACTAAGATTGGTGCTGGTATGGGTGA ACAAATGTTATCTCGTGAATATTATTATCTGGGTACCCAATTA CCAGTACG
10882371	YLR342W , reverse	pJD0523	GATCCGTACTGGTAATTGGGTACCCAGATAATAATATTCACG AGATAACATTTGTTACCCATACCAGCACCAATCTTAGTAGT GGAAATTTG
14892443	EST2 PRF1	pJD0659	CCCATTTTCGTTCTTCAGGGCATCCTTGAG
14892444	EST2 PRF2	pJD0660	GATACTTGGAATAAACTTATCACTCCATTTCATCGTAGAATATT TTAAGACG
14892445	EST2 PRF3	pJD0661	GCAAAATGAGGATTATACCTAAGAAGAGTAATAATGAGTTC AGG
14892446	EST2 PRF4	pJD0662	GAGGATACTCAAGGATGCGCTGAAGAACGAAAATGGG
14892447	EST2 PRF5	pJD0667	GCCAGTCCTAGCCAGGACACATTAATATTGAAGCTGGCTGAC GATTTC
14892448	GC - EST2 PRF1	pJD0659	CGAGGCCATTGAAGTGACCACACTTCAAGTTCTCTTTGTAAG TACTGTTGG
14892449	GC - EST2 PRF2	pJD0660	CGTCTTAAAATATTCTACGATGAATGGAGTGATAAGTTTATT CCAAGTATC
14892450	GC - EST2 PRF3	pJD0661	CCTGAACTCATTATTACTCTTCTTAGGTATAATCCTCATTTTG C
14892451	GC - EST2 PRF4	pJD0662	CCCATTTTCGTTCTTCAGCGCATCCTTGAGTATCCTC

14892452	GC - EST2 PRF5	pJD0667	GGAAATCGTCAGCCAGCTTCAATATTAATGTGTCTGGCTAG GACTGGC
17964081	ppr1_1_L	pJD0746	AAAAGTCGACCAAGTCTATGGCAAGCCCAC
16573903	ppr1_1_R	pJD0746	AAAAGGATCCGTCACCAGGAGATGTCCACAGTGTGCGGTG
15189677	ppr1_2_L	pJD0742	CCCCGTCGACCCGAGACAGGCAGCCCTATGAC
15611268	ppr1_2_R	pJD0742	CCCCGGATCCTGGATTGTTTCAGCCTCTGC
15611269	ppr1_3_L	pJD0743	CCCCGTCGACCAACATGCGTCCGATAGTTG
15611270	ppr1_3_R	pJD0743	CCCCGGATCCCAATGCCTCCAACCTGTCTG
16573905	ppr1_4_L	pJD0744	TTTTGTGCGACCATCACCACCAGGTATCCTAGCCTTTTGG
16573904	ppr1_4_R	pJD0744	AAAAGGATCCGATCTCGAGAGCTGGCGGATTGAAACCTAC
16573906	ppr1_5_L	pJD0745	AAAAGTCGACTGGTTACACGTGGGTAGCAGTTCAT
16573907	ppr1_5_R	pJD0745	AAAAGGATCCAACCTGATGATAAAATTTGTAACCTCGT
15452929	YJL061W Left	pJD0519	CCCCGTCGACCTGTATTAGTCCATGTGAACG
15452930	YJL061W Right	pJD0519	CCCCGGATCCTTTGCATGTCGATTGTAGTTGG
15452931	YPL128C Left	pJD0518	CCCCGTCGACTCAATCTCCAAATTCGTCAAC
15452932	YPL128C Right	pJD0518	CCCCGGATCCCATCCCATCTTCTAAATGAGG
15452933	YDR371W Left	pJD0625	CCCCGTCGACGGAGGGCATCTGGCCTTAC
15452934	YDR371W Right	pJD0625	CCCCGGATCCAGCTTCACCACATGACTCC
15452935	YOR026 W Left	pJD0519	CCCCGTCGACCCTATACACGGCTGGCTCTG
15452936	YOR026 W Right	pJD0519	CCCCGGATCCATCAGAAGTTGCCAGACATAG
15452937	YER115C Left	pJD0520	CCCCCGTCGACGATGAGTCCAAGTAGGAAG
15452938	YER115C Right	pJD0520	CCCCCGGATCCGGTAGCTTGCTGACATGCAC
15896741	FKS1 left	pJD1028	CCCCGTCGACGCTTCGTGGTGGTCGTATCAAG
15896742	FKS1 right	pJD1028	CCCAGATCTCAAATGGAAACCAGGATGGGC
15896743	FLR1 left	pJD1029	CCCCGTCGACATCGTGCTCTGAATCCTCTACC
16201602	FLR1 right	pJD1029	CCCAGATCTTGCCACTACATGACCAACGTG
17446869	LEFT KPN1 DLR	pJD0753	CCCCGGTACCTCGTTCGTTGAGCGAGTTC
17446870	RIGHT KPN1 DLR	pJD0753	CCCCGGTACCGGCGTCTTCCATGAGCTC
17447689	left DLR MCS KPN1 Amp.	pJD0753 based	CCCCGGTACCTCGTTCGTTGAGCGAGTTC
17447690	right DLR MCS KPN1 Amp.	pJD0753 based	CCCCGGTACCGGCGTCTTCCATGAGCTC
18409882	EST2 CDS Left	pJD0641	CCCCGGATCCCTGATTTATACTCATGAAAATCTTATCG



18409883	EST2 CDS Right	pJD0641	CCCCCTGCAGTCCTTATCAGCATCATAAGC
27608740	CCR5 PRF_451 Renilla_S al	pJD0827	AAAAAAGTCGACCTGTCGTCCATGCTGTGTTTGCTTTAAAG CCAGG
27608741	rev_CCR5 PRF_451 BamH1 firefly	pJD0827	AAAAAAGGATCCGACCTTCTTTTTGAGATCTGGTAAAGATGA TCCTGGGAG
32541225	oligo_ile7r clone_f_h s	pJD0835	CCCCGTCGACACCAAGAAAAATTTAAATGTGAGTTCAATCC TGAAAGTTTCCTGGACTGCCAGATTCATAGGGTGGATGACAT
32541226	oligo_ile7r clone_r_h s	pJD0835	CCCCGGATCCTTGCTGAGGAAACGTATCTTGCAGAAACCTT CCACTTCATCTCTAGCTTGAATGTCATCCACCCTATGAAT
32944228	Mouse ile7r forward	pJD0836	CCCCGTCGACTCACTTGAAAAAGAAATATTTAAAAAAGAA AGCATGATGTGGCCTACCGCCAGCAAGGGGTGAAAGCAAC TGGACG
32944229	Mouse ile7r reverse	pJD0835	CCCCGGATCCATACATTGCTTTTGGTCGTAGTTTCTCTGTGG GATTGTTGTTCTTGTGTGGAATAAAGATACATGCGTCCAGTT GCTTTCA
33735073	hs-ile27r 5prime	pJD0847	CCCCGTCGACAGGCAGGGGCGCCCCCTTCTGGCTGTGGCCGC TGCCCAAGCTGGCGCTGCTGCCTCTGTTGTGGGTGCTTTTCCA G
33735074	hs-ile27r 3prime	pJD0847	CCCCGGATCCCCCAAGGGTCCAACCTCCGTAGCACTGCAGTGG CCCGGCGCTGCCCTGGGGACGCGTCCGCTGGAAAAGCACCCA C
36621974	PDR5 300 3prime	pJD0859	CCCCGGATCCCTTCCAAGCGCAACCTAAGGAATAAGGCTTAT AAAAGTCAGGGTCTGCCGCACTTAGGTGAGCCATATTCTTAA CCCA
36621975	PDR5 300 5prime	pJD0859	CCCCGTCGACTCCAGGCTATGACCCAAAATGGACCCCAACTC CGAAAATTTTCTAGTGCCGCCTGGGTAAAGAATATGGCTCA CCTA
36621976	PDR5 3027 3prime	pJD0860	CCCCGGATCCCCAAGCAGTTTGAGAATCCAAACCAGAAGTAG GTTCACTCTAAAAAGACCAACAGTTTTGGTTTGGCAGTTAATT CAAC
36621977	PDR5 3027 5prime	pJD0860	CCCCGTCGACTGGTGTGCTGGTGAAGGTTTAACTTGAACA AAGAAAAAGATTAACCATTGGTGTGAATTAAGTCCAAACC AAAA
42384192	est2-72 fwd- cloning	pJD1017	AAAAGTCGACCATTGATCTACAGACCAACAGTACTTACAAAG
42384193	est2-72 rev- cloning	pJD1017	AAAAGGATCCGCCAAATGTTAGTACGTTGTTGTATAATTTCG
42384198	est2-1215 fwd- cloning	pJD1018	AAAAGTCGACTTACTTTAGACATGATACTTGAATAAACTTA TCA
42384199	est2-1215 rev- cloning	pJD1018	AAAAGGATCCTGTGAATTCTTCTTCGTCTGCCC
42384200	est2-1326 fwd- cloning	pJD1019	AAAAGTCGACGTTGTCCAATTTCAATCATAGCAAAATGAGG

42384201	est2-1326 rev- cloning	pJD1019	AAAAGGATCCTATTTGCGTTGGAGAATATATTTTAGTAAAC TAGTCG
42384204	est2-1995 fwd- cloning	pJD1020	AAAAGTCGACTTATAGCGAGTTTAAAGCCAGTCCTAGC
42384205	est2-1995 rev- cloning	pJD1020	AAAAGGATCCTATGTGCATTGCACAAAATTGAATAACCG
45726681	clone_5p stn1_885	pJD1000	AAAAGTCGACTGATCAAATTGACAACGGCAATGACG
45726682	clone_3p stn1_885	pJD1000	AAAAGGATCCCGATGCTAAGCTAGTTACTACGCTGCGTACTT CC
45726683	clone_5p stn1_1203	pJD1001	AAAGTCGACATTAAAAGATAAAACAAGTGAGACATTTGATTT GCTTCC
45726684	clone_3p stn1_1203	pJD1001	AAAAGGATCCCCACCAATTTTTTAGTACCTCAGGATACTGTTT CTTTGTC
45726685	clone_5p est1_1203	pJD1002	AAAAGTCGACCAAATATGCAGATTGAGTGAGCGCCAGG
45726686	clone_3p est1_1203	pJD1002	AAAAGGATCCGCTATATATATTTCTGAACAATTCAGTGGAC TATTATCAAGTCG
45726687	clone_5p est1_1272	pJD1003	AAAAGTCGACGACGTCGTCATCAAACCTCCTGGC
45726688	clone_3p est1_1272	pJD1003	AAAAGGATCCCCTGAAAATAATATCTTCTCTAAAAAGATAGC TTCTTTTCGG
45726689	clone_5p est1_1920	pJD1004	AAAAGTCGACGACAGAACTGGAAAAACAATTTGCAAATGTC CGG
45726690	clone_3p est1_1920	pJD1004	AAAAGGATCCTGGCACTTGGACGGTGATGTCCTC
45726691	clone_5p cdc13 1272	pJD1005	AAAAGTCGACGAGATCGTTATCCCGACGAGAGAGCGAATCT GTGAGC
45726692	clone_3p cdc13 1272	pJD1005	AAAAGGATCCGGGGTCTTTCCTTGCCATTTTGCTCATCC
46702125	Dyskerin clone 5prime	pJD1009	GCACGTCGACGAGAGTGATCATGGAGAGAGACACTTACCCTC GGAAGTGGGGTTTAGTCCAAAGGCAAGTCAGAAGAAGCTGA TGATCAAGCAGGGCC
46702126	Dyskerin clone 3prime	pJD1009	GCACGGATCCGTCACATACTCCTGCTTCCAGGTGGCAGGTG TGCTGTCTGTGGGCTTCCCATGCTTGTCAGAAGGCCCTGCTT GATCATCAGCTTC
48453484	dyskerin 0frame reverse	pJD1059	GCCTTTGGACCTAAACCCCACTTCCGAGGGT
48453483	dyskerin 0frame forward	pJD1059	ACCCTCGGAAGTGGGGTTTAGGTCCAAAGGC
79124738	>DNAJC9 264f		AAAAGTCGACCCAGATCCTGGGAAAATCTATTCCGTTCTCAG TGACAGAGAACAGAGAGCAGTGTACGATGAGCAGGGAACAG TGGACGAGGACTCTCC
79124739	>DNAJC9 264r		AAAAGAGCTCGTCCTCTAAAGATATCTTTTTAAAGAGTAGCC GCCAATACGCCTCCAGTCTCGGTCTTGGGTGAGCACAGGAG AGTCCTCGTCCACTGT
79124740	>GABRE		AAAAGTCGACCAAGTGGGAAAATTTAGCTTGAAATCAATG AGAAGAACTCCTGGAAGCTCTTCCAGTTTGATTTTACAGGAG

	735f		TGAGCAACAAAACCTGA
79124741	>GABRE 735r		AAAAGAGCTCGCCAAACCGCCTGCTCACATTGAAGAAAATC GTCATGACCATGAAGTCACCAACTGGGGTTGTGATTATTTCA GTTTGTGTGCTCACTCC
79124742	>SLC38A 2 981f		AAAAGTCGACGCTGTTTAGAAATTTAGATATTTGGGATATAC CAGTGGCCTTTCCCTTGTGTGTATGGTGTCTTTCTGATTGTG GTCATTTGCAAGAAA
79124743	>SLC38A 2 981r		AAAAGAGCTCTGGCTGTGTTAAGGTGGTGTATTATTGTTTCGTT AATTATCAAAGCAGCTTCCACAGGACACGGAACCTGAAATTT CTTGCAAATGACCAC
79124744	>IL18R1 714f		TCAAAAATGAACAAATGTGCGACGCTTAACCATGTTGCAGTGG AATTAGGAAAAAACGTAAGGCTAACTGCTCTGCTTTTGCTG
79124745	>IL18R1 714r		CGTCTTCCATGAGCTCCCCGGGGGATCCCATCCAATAAATTA CATCCTCTTCATTACAGCAAAGCAGAGCAGTTAGCCTTACGTTT TTCC
79124746	>TDO2 732f		AAAAGTCGACTAACTTCTGGGGAAAAGCTGAAAAAATATCA CCAGAGGCCTGGAAGAGGAATTCATAAGGATTACAGGCTAAA GAAGAGTCTGAAGAAAAA
79124747	>TDO2 732r		AAAAGAGCTCTTTCTCATCAAATAAGGACAGTAGCACCTCTT TTTGCTTCTGAAATTCAGCCACCTGTTCCTCTTTTCTTCAGAC TCTTCTTTAGCCTG
79124748	>TDO2 744f		AAAAGTCGACAAAGCTTGAAAAAATACACCAGAGGCCTGG AAGAGGAATTCATAAGGATTACAGGCTAAAGAAGAGTCTGAA GAAAAAGAGGAACAGGTG
79124749	>TDO2 744r		AAAAGAGCTCATGTTTCATGACGTTTCTCATCAAATAAGGACA GTAGCACCTCTTTTGTCTTCTGAAATTCAGCCACCTGTTCCTC TTTTCTTCAGA
79124750	>LPCAT1 1110f		AAAAGTCGACCGCCAATCCCAAACGCTTCTGTGCCGATTTC GCCCCGAAAACTCAGACGCTGGGCGGAAGCCTGTTTCGCAAG AAGCTGGATTAGGAC
79124751	>LPCAT1 1110r		AAAAGAGCTCGTCACTCGCAAAGAGGCTCATGGCGGTGATGT CCACGCGGGAGGGGCCGCTCTCTCCGCAACCCTGGGTCCTA ATCCAGCTTCTTGC
79124752	>OLA1 513f		AAAAGTCGACGGGGAATGCTTTTTATCCATATTAGTGCCTG TGATGGCATCTTTCATCTAACACGTGCTTTTGAAGATGATGAT ATCACGCACGTTGAA
79124753	>OLA1 513r		AAAAGAGCTCCATTTCTCATCTTTAAGCTGAAGCTCTTCATG TATTATTCTATATCTCGAATAGGATCTACACTTCCTTCAACG TGCGTGATATCATC
79124754	>KIF20A 2668f		AAAAGTCGACATTCTTCGAAATTTACTCCCCGAACACCAAC CTGCCAAAGCTCAACAGACTGCAGCCCTTATGCCCGGATCCT A
79124755	>KIF20A 2668r		AAAAGAGCTCGTACTTTTTGCCAAAAGGCCAGATTTGAGTA AAGGGGAACGCCGTGAGCGTAGGATCCGGGCATAAGGGCTG CAGTCTG
79124756	>CKAP5 2126f		AAAAGTCGACTGCCAGAAGGGAAATTTCCAAAACGTCAG CTCAGGTTGTATTAGATGGCCTTGTGGACAAGATTGGAGATG TGAAATGTGGGAACAAT
79124757	>CKAP5 2126r		AAAAGAGCTCCACAACCTGTTTCAGCAGTCCATGGTAACATAC AGGCTTCGGCTATTGCTGTCATAGCTTCTTTGCATTGTTCCC ACATTTACATCTC
79124758	>CKAP5 2906f		AAAAGTCGACAGCCATGGGCCCAAATATAAGCAACATGTAA AAAATTTAGGCATCCCTATCATCACAGTCCTTGGAGACAGCA AGAACAATGTTTCGAGCT
79124759	>CKAP5 2906r		AAAAGAGCTCATCTTCTCCTTCCAGCCATTCTTCATGCCAGT CTGTTCTGCCCAAGCATTACAGTCGCTAGGGCAGCAGCTCG AACATTGTTCTTGC
79124760	>CKAP5 2927f		AAAAGTCGACGCAACATGTAAAAAATTTGGCATCCCTATCAT CACAGTCCTTGGAGACAGCAAGAACAATGTTTCGAGCTGCTGC CCTAGCGACTGTGAAT

79124761	>CKAP5 2927r		AAAAGAGCTCTTCAGAAAGATCTTCTCCTTCCAGCCATTCTT CATGCCAGTCTGTTCTGCCCAAGCATTACAGTCGCTAGGGC AG
79124762	>MYO5C 3527f		AAAAGTCGACTCTGGAACATTTAAATGAGATGGAGAACTTTG GTTTGCTTATGAAGGACTAAAGAAAGCAACACGTGTTTTGGA GAGCCATTTCCAGTCT
79124763	>MYO5C 3527r		AAAAGAGCTCGATTTTCTTGACTGAGATGCACCACTTTGAAGT TCAAAGCTTCAATCTCCTTTTCATAGCAATCCTTCTGAGACTG GAAATGGCTCTCCA
79124764	>ZGPAT 664f		AAAAGTCGACCTTCTGGAGGGAAAGTGCCTTTAAGGAGA ACTGCAGGTTCTCCCATGGGCAGGTGGTCTCTCTGGATGAGC TGCGCCCCCTCCAGGAC
79124765	>ZGPAT 664r		AAAAGAGCTCTGCGTGCCAGAGGGCCATCCTGGTGCTTGGCCA GACACGCAGAGCCGGCTGCAGGGAGCTCAGGTCTGGGTCTT GGAAGGGGCGCAG
79124766	>DLK1 509f		AAAAGTCGACCGGGTACTCGGGAAGGATGCCAGAAAAAGG ACGGGCCCTGTGTGATCAACGGCTCCCCCTGCCAGCACGGAG GCACCTGCGTGGATGAT
79124767	>DLK1 509r		AAAAGAGCTCCACGATCTCGCAGAAATTGCCTGAGAAGCC AGGGGGGCACAGGCAGGAGGCATGGGAGGCCCGGCCCTCAT CATCCACGCAGGTGCCTC
79124768	>ATIC 1656f		AAAAGTCGACTGCCTTCTTCCCTTCCGGATAACGTAGACAG AGCTAAAAGGAGTGGTGTGGCGTACATTGCGGCTCCCTCCGG TTCTGCTGCTGACAAA
79124769	>ATIC 1656r		AAAAGAGCTCGTGGAAGAGCCGAAGGTTCTGATGAGCGAGG ATGATTCCCAGTTCTGTCGAGGCCTCAATCACAACCTTGTCA GCAGCAGAACCG
79124771	>PNMA1 1740f		AAAAGTCGACAGGGGCCAGCCCCAAACCTCTTCAGTTGCTGGT GCAGATCCGTGAGGAGGAAGCCAAGGAGG
79124772	>PNMA1 1740r		AAAAGAGCTCGAAGTGCCCTTCCAGGCCTAACTGCAGAAGG GTGGCCTCAGCCTCCTCCTCCTCCTTGGCTTCTCTCTCA
79124773	>ASNS 649f		AAAAGACTCCCTTTTTAAAAGGGAGCCTTTTCTTCTGGACA CTATGAAGTTTTGGATTAAAGCCAAATGGCAAAGTTGCATC CGTGGAATG
79124774	>ASNS 649r		AAAAGAGCTCAAAGAGTTTCTCCACATTGTCATAGAGGGCGT GCAGGGGTACATCCCGACAGTGATGATATTAAACCATTTC CGATGCAACTTTGC
79124775	>GOLT1B 437f		AAAAGTCGACATCCCTCCTAAATTTACTGGAATTAGATCATT TGATAGATAAAGTTGGAGAAAGCAACAATATGGTATTACAAC AAGTGAATTTGAAGACT
79124776	>GOLT1B 437r		AAAAGAGCTCGTAATTTAATTTTGTGCTGAATATTCTTCGAAT GACTTTATAAATAACACAATATTTTAAATGAGTCTTCAAATTC ACTGTTGT
79124777	>NT5DC2 841f		AAAAGTCGACCTATCGGCAGGGAAACCGTTTGACTTCTTGCG CTTGACGGAATGGCGTGGCCCCCGCGTGCTCTACTTCGGGGA CCACCTCTATAGTGAT
79124778	>NT5DC2 841r		AAAAGAGCTCCTCACGCTCCAGCTCGGGGATGATGGCGCCTG TGCGCCAGCCGTGCCGCAGCATGAGATCCGCCAGATCACTAT AGAGGTGGTCCCC
79124779	>MCM5 1004f		AAAAGTCGACGGCTGCCCTCCCAAATGCTATGAGGTTCATCTC CAAGAGCATCGCCCCCTCCATCTTTGGGGGCACAGACATGAA GAAGGCCATTGCCTGC
79124780	>MCM5 1004r		AAAAGAGCTCGTTGATGTCTCCTCGGCGAGTAAGTCCATCAG GGAGCCTCTTTCGGGAGCCCCAAAGAGCAGGCAGGCAATG GCCTTCTTCATGTCTG
79124781	>FOS- 408f		AAAAGTCGACAGCCCCCTACCCCTTTCGAGTCCCCGCCCTC CGTGGGGCTTACTCCAGGGCTGGCGTTGTGAAGACCATGAC AGGAGGCCGAGCGCAG
79124782	>FOS		AAAAGAGCTCGATTCTCCTTTTCTCTTCTTCTTCTGGAGATAA CTGTTCCACCTTGCCCTCCTGCCAATGCTCTGCGCTCGGCCT

	408r		CCTGTCAT
<b>79124783</b>	>NUP37 719f		AAAAGTCGACCTGGTGCTTAAAAAACACCTTCAAAGTTGGAG CCGTTGCAGGAAATGATTGGTTAATTTGGGATATTACTCGGT CCAGTTATCCTCAAAA
<b>79124784</b>	>NUP37 719r		AAAAGAGCTCCAGATTTTCTACTAATTGTGGACCACCTGAATA AGCAGGCTCGATCCATGTGAACAGGTCTCTTATTTTGAGGAT AACTGGACCGAG
<b>79124785</b>	>FAM21A 3270f		AAAAGTCGACCGCCATTTCCCCAAATGGCATCGGCCACAGCT CAGAGCAGCCAGTGGAGAAGACAGCACTGAGGAGGCCCTGG CAGCTGCCGCTGCACC
<b>79124786</b>	>FAM21A 3270r		AAAAGAGCTCGGAATGACCCAGAGACTTTGCAAAGGGGCTT CTGTCCACTCCAGGCACAGGACCACCTTCCCAAGGTGCAGCG GCAGCTGCC

**Table 9: Oligonucleotides used for cloning.**

Full specifications may be found by using the order number at <http://www.idtdna.com>

## Appendix 5: Site Directed Mutagenesis Oligonucleotide List

Order #	Name	Template	Sequence
15189665	Left EST2 ss72	pJD0641 based	CCCCGTCGACATTCGAGTTCATTCAAGACAAGC
15189666	Right EST2 ss72	pJD0641 based	CCCCGGATCCGTGGCTTAAGTCACCAGGAAGG
15189667	Left EST2 ss1215	pJD0641 based	CCCCGTCGACTCAATTGGCGATTTGTTC
15189668	Right EST2 ss1215	pJD0641 based	CCCCGGATCCATGATTGAAATTGGACAACGTG
15189669	Left EST2 ss1326	pJD0641 based	CCCCGTCGACCACGTTGTCCAATTTCAATCA
15189670	Right EST2 ss1326	pJD0641 based	CCCCGGATCCAGTCGGCCTTTTGTTCCTTAGG
15189671	Left EST2 ss1693	pJD0641 based	CCCCGTCGACCATACCAAGGATGGAATGTATG
15189672	Right EST2 ss1693	pJD0641 based	CCCCGGATCCTGATAAATGAACCGTCCTCAC
15189673	Left EST2 ss1995	pJD0641 based	CCCCGTCGACTAGCGAGTTTAAAGCCAGTCC
15189674	Right EST2 ss1995	pJD0641 based	CCCCGGATCCATCATCTGATTGGGAGCTTACG
22035998	CTS2 757 Stem1 Dis_F	pJD0815	GATGAAAAAATTCAATATACTCGAGTTATGATAACACTAAAT C
22035999	CTS2 757 Stem1 Dis_R	pJD0815	GATTTAGTGTTATCATAACTCGAGTATATTGAATTTTTTTCAT CAAAGC
22036002	EST2 754 Stem1 Dis_F	pJD0811	GGATGCGCTAAAAAATCAAATCCCAAATTCGTTAGATCTC
22036003	EST2 754 Stem1 Dis_R	pJD0811	GAGATCTAACGAATTTGGGATTTTCATTTTTTAGCGCATCC
22036006	PPR1 759 Stem1 Dis_F	Incomplete	CGACTTTTTTTTAAACATTTATAAACGAATTTGCGGATGCTA CGCAGG
22036006	PPR1 759 Stem1 Dis_R	Incomplete	CCTGCGTAGCATCCGCAAATTCGTTTATAAATGTTAAAAAAA AAGTCG
22036010	SPR6 748 Stem1	pJD0817	CAAATGTCGACAAAAAATCAAGGTTTGGATACACTCTGGAG CATGG

	Dis_F		
22036011	SPR6 748 Stem1 Dis_R	pJD0817	CCATGCTCCAGAGTGTATCCAAACCTTGATTTTTTTTGTGCGAC ATTTG
23810598	pJD759 SS SDM_L	pJD0816	CCGACAAAATTTTACATACCTTATTTCTTCTTGAATATTATAT TTGCTATTGG
23810599	pJD759 SS SDM_R	pJD0816	CCAATAGCAAATATAATATTCAAGAAGAAATAAGGTATGTGAA ATTTTGTGCGG
23810600	pJD754 SS SDM_L	pJD0806	GGATACTCAAGGATGCGCTGAAGAACGAAAATGGGTTTTTCG
23810601	pJD754 SS SDM_R	pJD0806	CGAAAAACCCATTTTCGTTCTTCAGCGCATCCTTGAGTATCC
23810602	pJD757 SS SDM_L	pJD0807	GCATATTGCTTTTGATGAGAAGAACTCAATATTTATCAGTTATG ATAACAC
23810603	pJD757 SS SDM_R	pJD0807	GTGTTATCATAACTGATAAATATTGAGTTCTTCTCATCAAAGC AATATGC
23810604	pJD748 SS SDM_L	pJD0797	CGTTAAGAAATTTTTGAAAAAGCAGAAGAAGTCAAGGAAACCA ATCACTCTGG
23810605	pJD748 SS SDM_R	pJD0797	CCAGAGTGATTGGTTTCCTTGACTTCTTCTGCTTTTTTCAAAA TTTCTTAACG
24059571	pJD521 SS_F	pJD0808	CAAATGTCGACGAAGAACCAAATGGGTTTTTTCGTTAGATCTC
24059572	pJD521 SS_R	pJD0808	GAGATCTAACGAAAAACCCATTTGGTTCTTCGTCGACATTTG
24059573	pJD521 S1_F	Incomplete d	CGACAAAAAATCAAATCCCAAATTCGTTAGATCTC
24059574	pJD521 S1_R	Incomplete d	GAGATCTAACGAATTTGGGATTTGATTTTTTGTGCG
24059575	pJD754 S1_F	pJD0811	GCTAAAAAATGAAAAATCCCAAATTCGTTAGATCTCAATATTTT
24059576	pJD754 S1_R	pJD0811	GAAATATTGAGATCTAACGAATTTGGGATTTTCATTTTTTAGC
24059577	pJD476 SS_F	Incomplete	CAAATGTCGACCTTCTTCTTGAATATATATTTGCTATTGGCC
24059578	pJD476 SS_R	Incomplete	GGCCAATAGCAAATATATATTCAAGAAGAAGGTCGACATTTG
24059581	pJD759 SS_F	pJD0816	CACATACCTTACTTCTTCTTGAATATTATATTTGCTATTGGCC
24059582	pJD759 SS_R	pJD0816	GGCCAATAGCAAATATAATATTCAAGAAGAAGTAAGGTATGTG
24059589	pJD757 SS_F	pJD0807	GCTTTGATGAGAAGAACTCAATATTTATCAGTTATGATAACAC
24059590	pJD757 SS_R	pJD0807	GTGTTATCATAACTGATAAATATTGAGTTCTTCTCATCAAAGC
24059591	pJD757 S1_F	pJD0815	GATGAAAAAATTCATATACTATAGTTATGATAACACTAAAT CAGTC
24059592	pJD757 S1_R	pJD0815	GACTGATTTAGTGTTATCATAACTATAGTATATTGAATTTTT TCATC
24059593	pJD520 SS_F	pJD0797	GAACAAATGTCGACAGAAGAAGTAAGGAAACCAATCACTC
24059594	pJD520 SS_R	pJD0797	GAGTGATTGGTTTTCTTACTTCTTCTGTCGACATTTGTTC
24059597	pJD748	pJD0818	GAAATTTTGAAGAAAGCAGAAGAAGTCAAGGAAACCAATCACT C

	SS_F		
24059598	pJD748 SS_R	pJD0818	GAGTGATTGGTTTCCTTGACTTCTTCTGCTTTTTCAAAAATTT C
24059599	pJD748 S1_F	pJD0817	GCAAAAAAATCAAGGTTTGGATACACTCTGGAGCATGG
24059600	pJD748 S1_R	pJD0817	CCATGCTCCAGAGTGTATCCAAACCTTGATTTTTTTTGC
24059605	pJD755 SS_F	pJD0812	CCCAGCAAGAAAATGAAGAACTTTGCCAAATTTAACGAAGACAG C
24059606	pJD755 SS_R	pJD0812	GCTGTCTTCGTAAATTTGGCAAAGTTCTTCATTTTCTTGCGG G
24059613	pJD756 SS_F	pJD0813	CTATTATGGAACAGAACTTGTCTCAGCATCCTTCATCAGC
24059614	pJD756 SS_R	pJD0813	GCTGATGAAGGATGCTGAGACAAGTTCTGTTCCATAATAG
24059615	pJD756 S1_F	pJD0814	GGAACAAAATTTATCTCACGATCGAAGTTCAGCTGCATCTGC
24059616	pJD756 S1_R	pJD0814	GCAGATGCAGCTGAACTTCGATCGTGAGATAAATTTGTTC
29714981	ccr5 slipsite mut forward	pJD0844	CGACTGTCGTCCATGCTGTGTTTGCCTTGAAGGCCAGGACGGT CACCTTGG
29714982	ccr5 slipsite mut reverse	pJD0844	CCAAAGGTGACCGTCTGGCCTTCAAGGCAAACACAGCATGGA CGACAGTCG
29714987	ccr5 stem2 5_comp forward	pJD0845	CCAGGACGGTCACCAAACCTCTACCATTCAAGTGTGATCACTTG GGTGGTGGC
29714988	ccr5 stem2 5_comp reverse	pJD0845	GCCACCACCCAAGTGATCACACTTGAATGGTAGAGTTTGGTGA CCGTCCTGG
29714989	ccr5 stem2 3_comp forward	pJD0846	GCGTCTCTCCAGGAATCATCTTAGTGGTGAGGGGTTAAGAA GGTCTTCATTACACC
29714990	ccr5 stem2 3_comp reverse	pJD0846	GGTGAATGAAGACCTTCTTAAACCCCTCACCCTAAGATGAT TCCTGGGAGAGACGC
34647145	ccr5 stem1core 5prime fwd	pJD0851 pJD0855	GCTGTGTTTGCTTTAAAGCCAGGACGCGGTCTTTGGGGTGG TGACAAGTGTG
34647146	ccr5 stem1core 5prime rev	pJD0851 pJD0855	CACACTTGTCAACACCCCAAGGACCGGTCTGGCTTTTAA GCAACACAGC
34647147	ccr5 stem1core 3prime fwd	pJD0852 pJD0855	GGTGGTGACAAGTGTGATCACTTGGGTGGACTGTGTGTTGCG TCTCTCCCAGG
34647148	ccr5 stem1core 3prime rev	pJD0852 pJD0855	CCTGGGAGAGACGCAACACACAGTCCACCCAAGTGATCACAC TTGTCACCACC
35774131	Stem_2_5' Proximal mut fwd	pJD0856	CCTTTGGGGTGGTGACAAGTTACTACAGGACCCTGGTGGCTGT GTTTGGCTCTCTCC
35774132	Stem_2_3' Proximal	pJD0856	GCTGTGTTTGGCTCTCTGGGTTACAGTGCTTTACCAGATCTC AAAAAGAAGG



	mut fwd		
36894175	hs-spr6 mutagen fwd	pJD0947	CTCAAAAATGAACAAATGTCGACAAAAAAATCAAGGAAACCA ATCACTC
36894176	hs-spr6 mutagen rev	pJD0947	GAGTGATTGGTTTCCTTGATTTTTTTTGTGACATTGTTTCAT TTTTGAG
45726676	SDM_5p est2 pos72	pJD1070	GTACTTACAAAGAAAATTTAAAAGTGGTCACTTCAATGGCCTC G
45726677	SDM_5p est2 pos1326	pJD1071	GCAAAAATGAGGATTATACCAAAAAAAGAATAATGAGTTCAGG
45726678	SDM_5p est2 pos1215	pJD0766	GGAATAAACTTATCACCCCTTTTTCGTAGAATATTTAAGACG
45726680	SDM_5p est2 pos1995	pJD1070	GCCAGGACACATTAATTTTAAAATGGCTGACGATTCC
46659196	SDM_3p est2 pos72 del1bp	pJD1017	CGAGGCCATTGAAGTGACCACTTTTAAATTTCTTTGTAAGTA C
46659197	SDM_3p est2 pos1326 del1bp	pJD1019	CCTGAACTCATTATCTTTTTTTTGGTATAATCCTCATTTTGC
46659198	SDM_3p est2 pos1215 del1bp	pJD1018	CGTCTTAAATATTCTACGAAAAAGGGTGATAAGTTTATTCC
46659199	SDM_3p est2 pos1995 del1bp	pJD1020	GGAAATCGTCAGCCATTTTAAAATTAATGTGTCCTGGC
49375506	est1 1203 minus1bp	pJD1040	GAGTGAGCGCCAGGTTTTTTTTTTAGATTGAGCTTTGATTTTA TTGC
49375507	est1 1272 minus1bp	pJD1041	CCCTCCTGGCAAAAAACAGGAAGACTTTCGATATCTAGCC
49375508	est1 1920 minus1bp	pJD1043	GCAATGTCCGGAGAACAAAAAATGTCTCCGCTCCCAGAAAA AGATGG
49375509	stn1 885 minus1bp	pJD1038	GCAATGACGAACAGTTAAAAAATAGAATATCAAAGCGCTAAT CTACC
49375510	stn1 1203 minus1bp	pJD1039	GATTTGCTTCCTCTAAAAAATTTTTGAATATGCTGAAAAACG
51722417	ccr5 slipsite mut_v2	pJD1060, pJD1058	CCATGCTGTGTTTGCCGCGCGGCCAGGACGGTCACCTTTGGG GTGG
52856887	CCR5 IRES_test	pJD1078	GAACAAATGTCGACTTAAGTCGTCCATGCTGTG

**Table 10: Oligonucleotides used for Site Directed Mutagenesis**

Full specifications may be found by using the order number at <http://www.idtdna.com>

## Appendix 6: Oligonucleotides Used for Sequencing

Order #	Name	Template	Sequence
16027528	DL MCS Seq_R	Dual luciferase	CGTACGTGATGTTACCC
14892453	EST2 Seq1	pJD0641 based	CTTAACCATAACTAACACGCCCTC
14892454	EST2 seq2	pJD0641 based	TGGTCGGTACATACGCATTTCG
14892455	EST2 seq3	pJD0641 based	GCAATCACCAAAGGAACGAGTC
14892456	EST2 seq4	pJD0641 based	TCACAAAAATGCTATCCAGCCCAC
14892457	EST2 seq5	pJD0641 based	CAACAGACCAACAGCAAGTG
14892458	EST2 seq6	pJD0641 based	CAGCGGTTGTCCAATTACGAAATG
14892459	EST2 seq7	pJD0641 based	CCTGATTAAATGTGCCCGGTCTC
15728281	ppr1_seq_1	pJD0746	CGAAGATGATGATTAAATCATG
15728282	ppr1_seq_2	pJD0742	GAGCGGAAGAGCTAC
15728283	ppr1_Seq_3	pJD0743	GCTATTGGCCATGCTAC
15728284	ppr1_seq_4	pJD0744	GACGTTCCCAAAAACCTTG
15728285	ppr1_seq_5	pJD0745	CGAACTGGAAAAACACAG
17446871	LEFT PGK1/PRF	pJD0753 based	GTCGGTCCAGAAGTTGAAGC
17446872	RIGHT, PGK1/PRF	pJD0753 based	TGAGAACTCGCTCAACGAAC
17447687	PGK1 3'-5' Seq. Primer	pJD0753 based	AACCGACCATAGAAGAGTGAGC
17447688	PGK1 5'-3' Seq Primer	pJD0753 based	AAGGTCAAGGCTTCCAAGG
17858744	PGK1-3UTR Sequencing Primer	pJD0741	GGAATTGCCAGGTGTTGC
17858745	PGK1-5UTR Sequencing Primer	pJD0741	TGGAAGCTGCAATCAATAGG
42164785	brewer-5p_ampr reading_5pr	$\beta$ -globin plasmids	CGTGAACCATCACCTAATCAAG
42164786	brewer-3p_ampr reading_3pr	$\beta$ -globin plasmids	GTTTGCCGGATCAAGAGCTACC
45341167	est2_rev sequencing	EST2	GATTCTAGTGTTAAACAGCG
48871874	cbf5 seq555_fwd	CBF5	TTCTGGGCTTCTCTGTGAAGCTGGTACTT
48871875	cbf5 seq555_rev	CBF5	AAGTACCAGCTTCACAGGAAGCCAGAA
48871876	cbf5 seq_m6_fwd	CBF5	AGTTCGTTACAACCTACACAGAGGACCGAT

<b>48871877</b>	cbf5 seq_m6_rev	<i>CBF5</i>	ATCGGTCCTCTGTGTAGGTTGTAACGAACT
<b>51122299</b>	Bglob seq_exon2 5p	$\beta$ -globin plasmids	GCACGTGGATCCTGAGAACT
<b>51122300</b>	cole1 seq_3p	$\beta$ -globin plasmids	ACGCCAGCAACGCGGC
<b>51122301</b>	Bglob exon3_seq 3p	$\beta$ -globin plasmids	CACCAACTTCTTCCACATTACCC

**Table 11: Oligonucleotides used for sequencing**

Full specifications may be found by using the order number at <http://www.idtdna.com>

## Appendix 7: Oligonucleotides Used for qPCR and Northern Analyses

Order #	Name	Template	Sequence
16743856	EST2 qPCR Left	Yeast cDNA	TGGTCGGTACATACGCATTTC
16743857	EST2 qPCR Right	Yeast cDNA	CGGCAGATGAGGTTTCGTTAC
17425701	18S qPCR Left	Yeast cDNA	GGAATTCTTAGTAAGCGCAAG
17425702	18S qPCR Right	Yeast cDNA	GCCTCACTAAGCCATTCAATC
17470908	Fluc qPCR Left	Dual lux cDNA	AACATCACGTACGCGGAATAC
17470909	Fluc qPCR Right	Dual lux cDNA	TCACTGCATACGACGATTCTG
17858742	PGK1-DLR qPCR Right	Exogenous <i>PGK1</i>	GTTCGTTGAGCGAGTTCTCA <b>These give heterogeneous products.</b>
17858743	PGK1-DLR qPCR Left	Exogenous <i>PGK1</i>	GGTACCGGCGTCTTCCAT <b>These give heterogeneous products.</b>
19855886	Left, RDN18 1 qRT-PCR	Yeast cDNA	GGAATTCCTAGTAAGCGCAAG
19855887	Right, RDN18 1 qRT-PCR	Yeast cDNA	GCCTCACTAAGCCATTCAATC
19855888	LEFT PGK1/PRF	Yeast RNA	GTCGGTCCAGAAGTTGAAGC
19855889	RIGHT, PGK1/PRF	Yeast RNA	TGAGAACTCGCTCAACGAAC
20699366	PGK1 5' AntiS probe	Yeast RNA	TTGACAGCGGCTTCAACTTCTGGACCGACACAGTCG
20699367	PGK1 AntiS probe	Yeast RNA	ACACCGTACTTCTTAGCGACAGTGGCAGTGTCAACCAC CA
20699368	RenillaMCS Antisense Probe	Yeast RNA	CATTTTGTGAGAACTCGCTCAACGAACGAGGTACC
20699369	FireflyMCS Antisense Probe	Yeast RNA	TACCGGCGTCTTCCATGAGCTCCC
22036012	U3_qRT-PCR Forward	Yeast cDNA	CGACGTACTTCAAGTATGTAATATACCCAA
22036013	U3_qRT-PCR Reverse	Yeast cDNA	TTGTCAGACTGCCATTGTACCCA
23277727	U3 Forward	Yeast cDNA	TCCAAGTGGTTGATGAGTCC
23277728	U3 Reverse	Yeast cDNA	CGAACCGCTAAGGATTGC
23277729	UPF1 Forward	Yeast cDNA	TACTCTGGCATGCAACATCC
23277730	UPF1 Reverse	Yeast cDNA	ATGTGAATGTGTCTTGGGAAGC
23277731	UPF2 Forward	Yeast cDNA	ACACCGAACACAGAGTCAGC
23277732	UPF2 Reverse	Yeast cDNA	CATCGTCATCGTCATCATCC
23277733	UPF3 Forward	Yeast cDNA	TGGTTGGAAGTGGTGATAAGG
23277734	UPF3 Reverse	Yeast cDNA	GCTTCTTCTTGCTCTGTTC
26782501	U3_northern probe	Yeast mRNA	CCAAGTTGGATTCAAGTGGCTCTTTGAAGAGTCAAAG AGTGACGATTCC
27163193	Renilla northern probe	Yeast mRNA	TCGACATTTGTTTCAATTTTGTGAGAACTCGCTCAACGAA CG
27163194	Firefly northern probe	Yeast mRNA	GGCGTCTTCCATGAGCTCCCCGGGGGATCC
30469750	est2 northern probe	Yeast mRNA	GGATGATCGTTGGACCCATTGGGC

46469611	Actin qrtPCR 5p	<i>H.s.</i> cDNA	CCAATTTACGCTGGTTTCTCTCTACC
46469612	Actin qrtPCR 3p	<i>H.s.</i> cDNA	CCTTGATGTCACGGACAATTTCTC
46469614	Bglobin qrtPCR 3p-changed	<i>H.s.</i> cDNA	ATGATGAGACAGCACATAACCAG
46469618	Bglobin qrtPCR 3p-original	<i>H.s.</i> cDNA	ATGAGTAGACAGCACATAACCAG
46469613	Firefly qrtPCR 5p	<i>H.s.</i> cDNA	GGTTTGGGAATGTTTACTACTCTCG
46469615	Firefly qrtPCR 3p	<i>H.s.</i> cDNA	CCTGAAGGGATCGTAAAAACAGC
46469616	Renilla qrtPCR 5p	<i>H.s.</i> cDNA	GCGTTGATCAAATCTGAAGAAGG
46469617	Renilla qrtPCR 3p	<i>H.s.</i> cDNA	GGTTCTAACTTTCTCATGATTTTTGATGG
46659195	Bglobin qrtPCR-5p	<i>H.s.</i> cDNA	GTGAAGTGCAGTGTGACAAGC
50876065	PGK1 luciferase 3p_v2	Yeast cDNA	CCGGGGAGCTCATGGAAGACGCCGTACCGCTCACAGAGCTCACTC
50876064	PGK1 luciferase 5p_v2	Yeast cDNA	CATCAACGATGCCTTCGGTACCTCGTTGAGCGAGTTCTC
49103336	auf1 qPCR_5p	<i>H.s.</i> cDNA	GCGAAGATTGACGCCAGTAAG
49103337	auf1 qPCR_3p	<i>H.s.</i> cDNA	CTGTGATAGGATCTAACTTCAGAGTG
49103338	upf1 qPCR_5p	<i>H.s.</i> cDNA	AGCTCGCAGACTCTCACTTTC
49103339	upf1 qPCR_3p	<i>H.s.</i> cDNA	CGTCTGGCTAGGAAGAGTAAAG
49629004	Hs brf1 qPCR_3p	<i>H.s.</i> cDNA	CTTGAGGCTGCTGAGGAG
49629005	Hs brf1 qPCR_5p	<i>H.s.</i> cDNA	GACAGAAAGGCAGTGGGC
49629006	Hs upf2 qPCR_5p	<i>H.s.</i> cDNA	CAGTGAGCAGCAAGGAGAG
49629009	Hs upf2 qPCR_3p	<i>H.s.</i> cDNA	CTTTTCTCTTATCATCTTCCAGTCTC
50158852	Gapdh 5p_qPCR forward	<i>H.s.</i> cDNA	TCGGAGTCAACGGATTGGTTCG
50158853	Gapdh 5p_qPCR reverse	<i>H.s.</i> cDNA	TAAACCATGTAGTTGAGGTCAATGAAGG
50158854	Gapdh 3p_qPCR forward	<i>H.s.</i> cDNA	AAGCTCATTTCTGGTATGACAACG
50158855	Gapdh 3p_qPCR reverse	<i>H.s.</i> cDNA	TCTTCCTCTTGCTCTTGCTGG
50312511	Renilla 5p_qPCR_v2	<i>H.s.</i> cDNA	AGGTGAAGTTCGTCGTCCAACATTATC
50312512	Renilla 3p_qPCR_v2	<i>H.s.</i> cDNA	GAACTTCTTGGCACCTTCAACAATAGC
50312513	eGFP 5p_qPCR	<i>H.s.</i> cDNA	AGAAGAACGGCATCAAGGTGA
50312514	eGFP 3p_qPCR	<i>H.s.</i> cDNA	CGGACTGGGTGCTCAGGTAG
50312515	Hs u6 5p_qPCR	<i>H.s.</i> cDNA	CGCTTCGGCAGCACATATAC
50312516	Hs u6 3p_qPCR	<i>H.s.</i> cDNA	AAAATATGGAACGCTTCACGA
52335636	CCR5 forward	<i>H.s.</i> cDNA	GTCCCTTCTGGGCTCACTAT
52335637	CCR5 reverse	<i>H.s.</i> cDNA	CCCTGTCAAGAGTTGACACATTGTA
59501271	BAG1 qPCR_F	<i>H.s.</i> cDNA	TGCCGGGTCATGTTAATTGGG
59501272	BAG1 qPCR_R	<i>H.s.</i> cDNA	GCAGAGAGCTTCAGCTTGCAATCC

**Table 12: Oligonucleotides used as probes for qPCR and Northern**

Full specifications may be found by using the order number at <http://www.idtdna.com>

## Appendix 8: RNAi Oligonucleotides

Order #	Name	Sequence
59828675	1224-5p	5' GUG AGG ACU CGG GAG GUG G 3'
78135739	mir1224_r	5' CGA CCU CCC GAG UCC UCA C 3'
78135740	mir1224_f biotin	5' /Biotin/ GUG AGG ACU CGG GAG GUG G 3'
	hUPF2	Target: 5' CA CCA TGA GCG TGG AGG CGT A 3' Sense: 5' CCA UGA GCG UGG AGG CGU Att 3' Anti: 3' gt GGU ACU CGC ACC UCC GCA U 5'
	hHNRPD (AUF1)	Target: 5' AA CAG CCA AGG TTA CGG TGG T 3' Sense: 5' CAG CCA AGG UUA CGG UGG Utt 3' Anti: 3' tt GUC GGU UCC AAU GCC ACC A 5'
	hBRF1	Target: 5' CA CCA GTC AGT TGA CCA TTG A 3' Sense: 5' CCA GUC AGU UGA CCA UUG Att 3' Anti: 3' gt GGU CAG UCA ACU GGU AAC U 5'
	hMAPK	Target: 5' CC CCG GTA CCT CGT TCG TTG AGC GAG TTC 3' Sense: 5' CCG GUA CCU CGU UCG UUG AGC GAG UUC tt 3' Anti: 5' gg GGC CAU GGA GCA AGC AAC UCG CUC AAG 5'
	hSMG1	Target: 5' CA CCA TGG TAT TAC AGG TTC A 3' Sense: 5' CCA UGG UAU UAC AGG UUC Att 3' Anti: 5' gt GGU ACC AUA AUG UCC AAG U 5'
	hEIF2C1 (argonaute 1)	Target: 5' TA GTC TTA ACA TAA AGC CGA A 3' Sense: 5' GUC UUA ACA UAA AGC CGA Att 3' Anti: 3' at CAG AAU UGU AUU UCG GCU U 5'
	hRENT1 (UPF1)	Target: 5' CA CCA TGA GCG TGG AGG CGT A 3' Sense: 5' CCA UGA GCG UGG AGG CGU Att 3' Anti: 3' gt GGU ACU CGC ACC UCC GCA U 5'
	Scrambled	Target: 5' AAT TCT CCG AAC GTG TCA CGT 3' Sense: 5' UUC UCC GAA CGU GUC ACG Utt 3' Anti: 3' ACG UGA CAC GUU CGG AGA Att 5'

**Table 13: Oligonucleotides used in RNAi studies**

Synthetic dsRNA oligonucleotides include the sense and antisense strands with 2 base pair overhangs. Full specifications for the hsa-miR-1224 oligonucleotides may be found by using the order number at <http://www.idtdna.com>. The details for the overhang containing nucleotides may be found at <https://www.qiagen.com/geneglobe/> by searching for the ORF name.

## Appendix 9: Oligonucleotides Used for SHAPE / *in-vitro*

### Translation

Order #	Name	Template	Sequence
27598536	RenillaKozak	Dual luciferase plasmids	TAATACGACTCACTATAGGGAGACCACCATGTTGCCATCAAAAA TC
27598537	FireflypolyA	Dual luciferase plasmids	TTTTTTTTTTTTTTTTTTTTTGTTCCTGTCATACGACGTC
39376817	est2 mrna oligo t7kozak 5prime	pJD0641 etc	CAATATTTTCCTTATCAGCATCATAAGCTGTCAGTA
39376818	est2 mrna oligo 3prime	pJD0641 etc	CAATATTTTCCTTATCAGCATCATAAGCTGTCAGTA
39376819	<i>EST2</i> mRNA	est2 shape 10bp-down	TTGAGATCTAACGAAAA
39376820	<i>EST2</i> mRNA	est2 shape 56bp-down	CAACATTAAATAACTTC
39376821	<i>EST2</i> mRNA	est2 shape 113bp-down	CCGTCCTCACATTATCT
39376822	<i>EST2</i> mRNA	est2 shape 170bp-down	CTGTTTTAAATATTTCC
39376823	<i>EST2</i> mRNA	est2 shape 223bp-down	GAGCCCTGAAAAAGACC
39714802	Yeast genomic	pdr5 5prime with t7	TAATACGACTCACTATAGGGAGAATGCCCGAGGCCAAGCTTAAC AATAACG
39714803	Yeast genomic	pdr5 3prime fulllength	CCAAGAAATAATAGAATTTTGAATTTGGTTAAGAAAAGAACTT ACC
39714804	Yeast genomic	pdr5 3prime 3kb	GGATAGAGCCAACAATAACATAGAAGCGGGAACCATAGC
39938945	<i>PDR5</i> mRNA	pdr5 3prime revcomp	GTAAGTTTCTTTCTTAACCAAATTCAAAATTCTATTATTTCTT G
39997805	<i>PDR5</i> mRNA	pdr5 200 revcomp	GGTTTCAAGATTGGAAGTATTGGTTTCTTTAG
39997806	<i>PDR5</i> mRNA	pdr5 150 revcomp	GGATTTTGTATGGAATATTAACC
39997807	<i>PDR5</i> mRNA	pdr5 100 revcomp	GCGGAAGCACCAGAAGCACTTAAGTTC
39997808	<i>PDR5</i> mRNA	pdr5 50 revcomp	GGAATAAGGCTTATAAAAGTCAGGGTCTGCC
39997809	<i>PDR5</i> mRNA	pdr5 0 revcomp	CCCAGGCGGCACTAGAAAAATTTTCGG
42258371	Dual luciferase mRNA	5prime of Renilla RToligo	GGGCCTTTCTTTATGTTTTTGGC
42258372	pJD0835 mRNA	ile7r-RToligo	GCTTGAATGTCATCCACCCTATG
42258373	Dual luciferase plasmid	5p-renilla t7 oligo	TAATACGACTCACTATAGGGAGAAAGGTGAAGTTCGTCGTCC

<b>42258374</b>	Dual luciferase plasmid	3p-firefly RT oligo	TCCCCGACTTCTTTCGAAAGAGGTGCGCCC
<b>48668279</b>	<i>CCR5</i> plasmids	pJD827 5p_T7	TAATACGACTCACTATAGGTTTGCTTTAAAAGCCAGG
<b>48668280</b>	<i>CCR5</i> plasmids	pJD827_3p	GAGATCTGGTAAAGATGATT
<b>56107905</b>	<i>CCR5</i>	CCR5 Tweezer T7_5prime	TAATACGACTCACTATAGGGGGAAAATATATCAAATCGTTCGTT GAGCG
<b>56107906</b>	<i>CCR5</i>	CCR5 Tweezer 3prime	GATACTGACTGTATGGAAAATGAGAGCTGC

**Table 14: Oligonucleotides used for SHAPE and *in-vitro* transcription**

Full specifications may be found by using the order number at <http://www.idtdna.com>



## Appendix 10: Code Summary

***pymol/plugin\_helices.py***: Top level pymol plugin, reads ribosome PDB publication spreadsheet, parses pdb files, labels chains, reads data files containing ribosomal helices and base modifications.

***pymol/movies/render.py***: Submits pdb sessions for rendering on the University of Maryland High Performance Computing cluster.

***prfdb/prf\_daemon***: Top level perl interface to the prfdb. Handles importing sequences, running them through the filtering pipeline, randomization, and data storage.

***prfdb/lib/PRFConfig.pm***: Primary configuration object. This combines configuration file parsing and command line options into a single namespace. In addition it ensures that webserver and compute nodes follow the same configuration.

***prfdb/lib/PRFdb.pm* and *prfdb/lib/PRFdb/\*.pm***: Database abstraction layer. Contains interfaces for all database activities and schema.

***prfdb/lib/PRFGraph.pm***: Graphing library. Takes data from PRFdb and uses GD::Graph, SVG::TT::Graph, JSON, etc to create graphs.

***prfdb/lib/PRFBlast.pm***: Layer for Bio::Tools::Run::Blast\*, handles running various blast searches against the local database as well as ncbi.

***prfdb/lib/SeqMisc.pm***: Reads in sequences and provides various nucleotide frequencies, randomizations, translations, etc. Simpler than Bio::SeqIO.

***prfdb/lib/HTMLMisc.pm***: Commonly reused html fragments.

***prfdb/handler.pl***: Defines how the apache webserver responds to requests to the PRFdb, initializes connections to the databases.

***prfdb/lib/MyDeps.pm***: Perl dependency resolver. Uses CPAN to locally install any missing dependencies without user intervention.

***prfdb/lib/PRFsnp.pm***: Integrates ncbi SNP data into the PRFdb. Initially written by Nic Hepler.

***prfdb/lib/PkParse.pm***: Parser for RNA secondary structure output, provides bpseq, parentheses, fasta outputs.

***prfdb/lib/Bootlace.pm***: Randomization and refolding library.

***prfdb/lib/Agree.pm***: Measures amount of agreement among various RNA structure prediction programs.

***prfdb/lib/MicroRNA.pm***: Integrates microRNA data into PRFdb.

***prfdb/lib/RNAMotif.pm***: Provides interfaces to *rnasamotif*. (Should be superseded by Bio::Tools::Run::RnaMotif.pm)

***prfdb/lib/RFolders.pm***: Provides interfaces to various secondary structure programs including: Vienna's *RNAfold*, *pknobs*, *Nupack*, *Hotknots*, *mfold*, *ilm*.

***prfdb/lib/MyGenbank.pm***: Wrapper for Bio::DB::Universal and Bio::SeqIO to place all of the annotation data therein into a single namespace.

***prfdb/lib/Overlap.pm***: Computes length of +1 and -1 frame ORF extensions.

***prfdb/\*.sh***: Maintenance tasks including pbs queues, database backup, etc.

***prfdb/\*.html***: HTML::Mason objects defining the web interface to the PRFdb.

***prfdb/contrib/\*.pl***: Various testing scripts, database modification, graphing, etc.

## Appendix 11: Partial Summary of the PRFdb

Species	ORFs	Windows	1std_mfe	1std_z	1std_both
<b>Caenorhabditis elegans</b>	50,914	46,725	5,731	5,158	4,694
<b>Branchiostoma floridae</b>	50,729	164	20	24	16
<b>Strongylocentrotus purpuratus</b>	42,275	66	10	7	7
<b>Homo sapiens</b>	37,337	62,272	8,428	8,197	6,205
<b>Danio rerio</b>	33,997	8,731	1,249	1,242	979
<b>Ricinus communis</b>	28,270	91	13	12	7
<b>Xenopus laevis</b>	26,537	26,689	3,543	3,514	2,852
<b>Drosophila yakuba</b>	16,095	21,741	3,056	1,952	1,605
<b>Mus musculus</b>	15,683	28,948	4,101	3,913	3,193
<b>Drosophila willistoni</b>	15,426	5,025	649	698	561
<b>Ciona intestinalis</b>	13,921	34	4	7	4
<b>Kluyveromyces waltii</b>	10,888	0	0	0	0
<b>Nasonia vitripennis</b>	9,495	71	7	6	4
<b>Apis mellifera</b>	9,266	77	11	9	9
<b>Bos taurus</b>	9,225	16,449	2,242	2,260	1,628
<b>Saccharomyces paradoxus</b>	8,955	37,274	4,481	4,310	3,891
<b>Drosophila virilis</b>	7,725	9,497	1,339	52	46
<b>Saccharomyces cerevisiae</b>	6,354	37,423	4,441	4,133	3,991
<b>Kluyveromyces polysporus</b>	5,526	80,234	4,451	3,537	3,254
<b>Rattus norvegicus</b>	5,376	9,035	1,308	1,307	1,020
<b>Bacillus anthracis</b>	5,287	6,067	771	801	664
<b>Candida glabrata</b>	5,237	1	3	1	3
<b>Kluyveromyces thermotolerans</b>	5,171	0	0	0	0
<b>Kluyveromyces lactis</b>	5,138	10	2	3	2
<b>Xenopus tropicalis</b>	5,126	14,742	1,968	1,905	1,509
<b>Schizosaccharomyces pombe</b>	5,091	33,760	3,916	3,620	3,505
<b>Zygosaccharomyces rouxii</b>	5,051	2	3	1	3
<b>Drosophila melanogaster</b>	5,013	9,847	1,372	1,216	1,104
<b>Saccharomyces bayanus</b>	4,970	23,438	2,868	2,782	2,474
<b>Ashbya gossypii</b>	4,753	0	0	0	0
<b>Saccharomyces castellii</b>	4,684	24,289	2,936	2,842	2,616
<b>Gallus gallus</b>	4,450	10,467	1,328	1,361	1,027
<b>Escherichia coli</b>	4,322	5,446	820	804	633
<b>Xenopus silurana</b>	3,901	0	0	0	0
<b>Saccharomyces kudriavzevii</b>	3,778	15,769	2,050	1,937	1,758
<b>Haloarcula marismortui</b>	3,412	425	69	58	42

<b>Saccharomyces mikatae</b>	3,109	13,048	1,729	1,664	1,509
<b>Salmo salar</b>	2,986	32	7	7	7
<b>Saccharomyces kluyveri</b>	2,977	11,217	1,433	1,406	1,232
<b>Arabidopsis thaliana</b>	2,903	73	11	10	10
<b>Staphylococcus aureus</b>	2,538	10,515	1,298	1,338	1,137
<b>Thermus thermophilus</b>	1,973	319	45	42	34
<b>Streptococcus pyogenes</b>	1,939	6,369	848	808	708
<b>Oncorhynchus mykiss</b>	844	0	0	0	0
<b>Pan troglodytes</b>	774	1,522	213	210	154
<b>Oryzias latipes</b>	472	20	6	5	6
<b>Viral</b>	599	5,850	600	17,700	300
<b>Total</b>	500,492	583,774	69,380 (11.9%)	80,859 (13.9%)	54,403 (9.3%)

**Table 15: A partial summary of the PRFdb**

This summary of the PRFdb counts only those species for which there are more than 450 open reading frames. It shows the number of ORFs, the number of significant folds with respect to MFE, Z-score, and both. Some species have been imported but not examined, notable examples include: *Kluyveromyces thermotolerans* and *Candidata glabrata*.

## Appendix 12: Categorized Recoding Signals

	Genes/Proteins	Occurrence	Recoding type	Recoding Site	Stimulators
Chromosomal Recoding Events (Non-viral)	<i>oaz</i> : antizyme	From <i>S. pombe</i> to invertebrates	+1 FS	YCC UGA UUU UGA	Stop codon, polyamines
	<i>oaz1/2</i> : antizymes 1, 2	Mammals	+1 FS	UCC UGA	Stop codon, polyamines
	<i>oaz3</i> : antizyme 3	Mammals	+1 FS	UCC UGA	Unknown
	<i>p45</i> : telomerase component	<i>Euplotes</i>	+1 FS	AAA UAA	Unknown
	<i>est3</i> : telomerase component	<i>S. cerevisiae</i>	+1 FS	CUU AGU	Hungry codon in A site
	Actin-filament binding protein	<i>S. cerevisiae</i>	+1 FS	CUU AGU	Hungry codon in A site
	Retrotransposons Ty1, Ty2, Ty4	<i>S. cerevisiae</i>	+1 FS	CUU AGG	Hungry codon
	<i>prfB</i> : Peptide release factor 2	Most bacteria	+1 FS	CUU AGU U	Shine-Dalgarno
	<i>dnaX</i> : DNA pol III	<i>E. coli</i>	-1 FS	A AAA AAG	Shine-Dalgarno
	<i>dnaX</i> : DNA pol III	<i>T. thermophilus</i>	slippage (- n)	UUUUUUUUU	None!
	<i>cdd</i> : Cytidine deaminase	<i>B. subtilis</i>	-1 FS	A CGA AAG	Shine-Dalgarno
	<i>arg1</i> : Ornithine carbamoyltransferase	<i>E. coli</i>	+1 FS	UUU C	Shift into termination at 5' end of ORF
	Multiple viral mobile elements	Bacteria	-1 FS	Various	Pseudoknots
	<i>kel</i> : Kelch	<i>D. melanogaster</i>	suppression	UGA	Developmental regulation
	<i>oaf</i> : Out at first	<i>D. melanogaster</i>	suppression	UGA	Developmental regulation
	<i>hdc</i> : Headcase	<i>D. melanogaster</i>	suppression	UAA	Developmental regulation
	<i>topA</i> : DNA topoisomerase I	<i>B. firmus</i>	suppression	UGA	Unknown
	Adhesion factors	<i>E. coli</i>	suppression	UAG	Unknown
	Selenocysteine incorporation	All kingdoms	suppression	UGA	Multiple mechanisms, stem-loops, tRNAs with long anticodon loops
Viral Recoding	<i>gag-pol</i> , <i>gag-pro-pol</i>	HIV, MMTV, etc	-1 FS	N NNW WWH	3' pseudoknot/stem-loop
	<i>gag-pol</i> , <i>gag-pro-pol</i>	MuLV, TMV	suppression	UAG	Pseudoknot
	<i>pol</i> : RNA polymerase	BYDV	-1 FS	G GGU UUU	Stimulatory element 3kb downstream!
	<i>gene 60</i> : Topoisomerase subunit	Phage T4	Sequence bypassing (shunting)	GGA (47 bases) GGA	Stem-loop
	Coat protein	RNA phage Q $\beta$	suppression	UGA	Unknown
	Coat lysis hybrid	RNA phage MS2	+1 FS	Unknown	Stem loop
	Capsid-RNA replicase	Sindbis	suppression	UGA	
	Genes <i>g-t</i> tail assembly proteins	Lambdoid phages	-1 FS	G GGA AAG	
	<i>Gene 10</i> : Major coat protein	Phage T7	-1 FS	G GUU UUC	3' stimulatory element in UTR

**Table 16: A partial bestiary of recoding signals**

This is a short listing of characterized recoding signals sorted from chromosomal to viral elements. This is mostly transcribed from Baranov et al.[158]. Many hundreds of new examples may be found at the much improved Recode2 database[6].

### Appendix 13: Observed Viral Origin -1 PRF Signals.

Family/Group	Genus	Virus	Gene Overlap	References
<b>Retroviridae</b>	Lentivirus	Human Immunodeficiency Virus type 1	gag-pol	Jacks et al (1998)
<b>Retroviridae</b>	Lentivirus	Feline Immunodeficiency Virus	gag-pol	Morikawa and Bishop (1992)
<b>Retroviridae</b>	ALSV	Rous Sarcoma Virus	gag-pol	Jacks and Varmus (1985)
<b>Retroviridae</b>	B-type	Mouse Mammary tumor virus	gag-pro	Moore et al (1987), Jacks et al(1987)
<b>Retroviridae</b>	D-type	Simian Retrovirus Type 1	gag-pro	ten Dam et al.(1994)
<b>Retroviridae</b>	HTLV	Human T cell leukemia virus Type 1	gag-pro	Nam et al.(1988)
<b>Retroviridae</b>	HTLV	Human T cell leukemia virus Type 1	pro-pol	Nam et al.(1993)
<b>Retroviridae</b>	HTLV	Human T cell leukemia virus Type 2	gag-pro-pol	Mador et al.(1989)
<b>Coronaviridae</b>	Coronavirus	Infectious Bronchitis Virus	orf1a-orf1b	Brierley et al.(1988)
<b>Coronaviridae</b>	Coronavirus	Mouse Hepatitis Virus	orf1a-orf1b	Bredenbeek et al.(1990)
<b>Coronaviridae</b>	Coronavirus	Human Coronavirus	orf1a-orf1b	Herold et al.(1993)
<b>Coronaviridae</b>	Coronavirus	Transmissible gastroenteritis virus	orf1a-orf1b	Eleouet et al.(1995)
<b>Coronaviridae</b>	Torovirus	Berne Virus	orf1a-orf1b	Snijder et al.(1990)
<b>Coronaviridae</b>	Arterivirus	Equine Arteritis Virus	orf1a-orf1b	den Boon et al.(1991)
<b>Astroviridae</b>	Astrovirus	Human astrovirus serotype 1	orf1a-orf1b	Marczinke et al.(1994)
<b>Totiviridae</b>	Totivirus	Giardia lamblia virus	orf1-orf2	Wang et al.(1993)
<b>Totiviridae</b>	Totivirus	Saccharomyces cerevisiae L-A	gag-pol	Dinman et al.(1991)
<b>Totiviridae</b>	Totivirus	Saccharomyces cerevisiae L-1	cap-pol	Diamond et al.(1989)
<b>Podoviridae</b>	T7 phage	Bacteriophage T7	10A-10B	Condron et al.(1991)
<b>Siphoviridae</b>	λ phage	Bacteriophage λ	gpG-T	Levin et al.(1993)
<b>Luteoviridae</b>	Luteovirus	Barley Yellow Dwarf Virus	39K-60K	Brault and Miller(1992)
<b>Luteoviridae</b>	Luteovirus	Beet Western Yellow Virus	orf2-orf3	Garcia et al.(1993)
<b>Luteoviridae</b>	Luteovirus	Potato Leaf Roll	orf2a-orf2b	Prufer et al.(1992)
<b>Dianthoviridae</b>	Dianthovirus	Red clover necrotic mosaic virus	P27-p57	Xiong et al.(1993)
Family/Group	Number of Signals	Non-viral Genus	Number	
<b>Potyvirus</b>	59	<b>Drosophila</b>	20	
<b>Alphavirus</b>	20	<b>E. coli</b>	7	

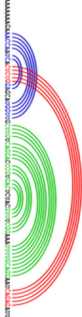


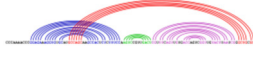
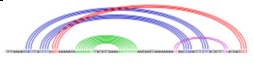

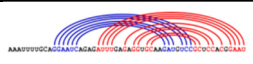
<b>Poleovirus</b>	12	<b>Salmonella</b>	4	
<b>Lentivirus</b>	12	<b>Neisseria</b>	3	
<b>Deltaretrovirus</b>	10	<b>Mus</b>	2	
<b>Coronavirus</b>	9	<b>Bacillus</b>	1	
<b>Betaretrovirus</b>	8	<b>Bombyx</b>	1	
<b>P2-like viruses</b>	7	<b>Ceratitidis</b>	1	
<b>Flavivirus</b>	7	<b>Homo</b>	1	
<b>Luteovirus</b>	6	<b>Vibrio</b>	1	
<b>Bymovirus</b>	4			
<b>Tritimovirus</b>	4			
<b>Umbravirus</b>	4			
<b>Rymovirus</b>	3			
<b>Dianthovirus</b>	3			
<b>Ipomovirus</b>	3			
<b>Sobemovirus</b>	3			
<b>Giardiavirus</b>	3			
<b>Mamastrovirus</b>	3			
<b>Totivirus</b>	3			
<b>Alpharetrovirus</b>	2			
<b>Avastrovirus</b>	2			
<b>Lambda-like virus</b>	1			
<b>T-7 like viruses</b>	1			
<b>Torovirus</b>	1			
<b>Okavirus</b>	1			

**Table 17: A catalog of known viral origin -1 PRF signals**

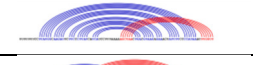
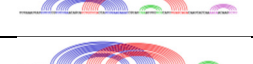
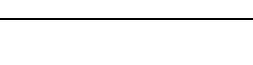
Categorizing the various viral origin -1 PRF signals[6,13]

## Appendix 14: -1 PRF Signals are not Conserved across Yeast Species


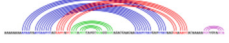
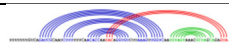


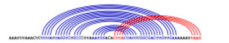
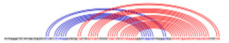
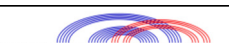
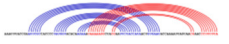
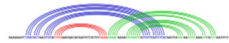
### SPR6 -1 PRF signals in *Saccharomyces* species

<p><i>S. cerevisiae</i> SPR6 PRF signal @ 279</p> <p>MFE: -20.33 Kcal/mol</p> <p>Slippery Site: AAAAAAA</p> 	Yeast species	Slippery site location of similar frameshifting signal	Slippery site	Predicted MFE (Minimum Free Energy) Kcal/mol	Predicted psuedoknot structure
	<i>S. paradoxus</i>	243	AAAAAAA	-10.7	
		348	AAAUUUU	-31.5	
		417	CCCAAAA	-29.7	
	<i>S. mikatae</i>	No strong			
	<i>S. bayanus</i>	No predicted			
	<i>S. kudriavzevii</i>	216	UUUAAAA	-11.5	
		279	AAAAAAA	-9.6	
		348	AAAUUUU	-8.6	

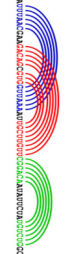

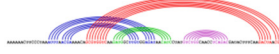


### EST2 -1 PRF signals in *Saccharomyces* species

<p><i>S. cerevisiae</i> EST2 PRF signal @ 1653</p> <p>MFE: -16.9 Kcal/mol</p> <p>Slippery Site: AAAAAAU</p>	Yeast species	Slippery site location of similar frameshifting signal	Slippery site	Predicted MFE (Minimum Free Energy) Kcal/mol	Predicted psuedoknot structure
	<i>S. paradoxus</i>	585	UUUUUUU	-13	
		744	UUUAAAU	-19.7	
		1131	UUUUUUU	-20.3	


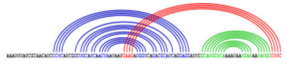

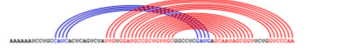
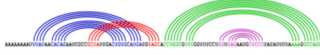

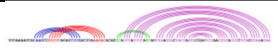


		1320	AAAAAAA	-20	
		1662	UUUUUUU	-12.3	
	<i>S. mikatae</i>	426	UUUUUUU	-13.75	
		1995	UUUAAAA	-12.19	
	<i>S. bayanus</i>	933	AAAUUUA	-12.9	
		936	UUUAAAC	-15.6	
		1689	UUUUUUC	-18.6	
	<i>S. castellii</i>	453	AAAUUUA	-16	
		2022	AAAAAAU	-5.83	
	<i>S. kudriavzevii</i>	No predicted			

### BUB3 -1 PRF signals in *Saccharomyces* species

<i>S. cerevisiae</i> BUB3 PRF signal @ 858 MFE: -11.71 Kcal/mol Slippery Site: AAAAAAU 	Yeast species	Slippery site location of similar frameshifting signal	Slippery site	Predicted MFE (Minimum Free Energy) Kcal/mol	Predicted pseudoknot structure (nupack)
	<i>S. paradoxus</i>	858	AAAAAAC	-10.6	
	<i>S. mikatae</i>	No predicted			
	<i>S. bayanus</i>	864	AAAAAAC	-22.3	
	<i>S. castellii</i>	138	UUUAAAA	-9.96	
	<i>S. kudriavzevii</i>	417	AAAAAAC	-3.4	
	<i>S. kluyveri</i>	None predicted			

### TBF1 -1 PRF signals in *Saccharomyces* species

<i>S. cerevisiae</i> TBF1 PRF signal @ 771  MFE: -9.5 Kcal/mol  Slippery Site: AAAUUUA  	<b>Yeast species</b>	<b>Slippery site location of similar frameshifting signal</b>	<b>Slippery site</b>	<b>Predicted MFE (Minimum Free Energy) Kcal/mol</b>	<b>Predicted pseudoknot structure</b>
	<i>S. paradoxus</i>	None predicted			
	<i>S. mikatae</i>	1521	AAAUUUA	-18.5	
	<i>S. bayanus</i>	None predicted			
	<i>S. kudriavzevii</i>	495	GGGAAAC	-12.4	
		1080	AAAAAAU	-27	
	<i>S. kluyveri</i>	273	AAAAAAA	-27.33	
		1293	AAAAAAC	-22.6	
	<i>S. castellii</i>	537	UUUAAAA	-20.1	

**Table 18: -1 PRF signals are not conserved across yeast species**

The predicted -1 PRF signals for the ORF homologs in this study are listed; including the location of slippery sites, predicted minimum free energies (MFE) and downstream pseudoknots (3' of the slippery site) from *Saccharomyces cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. bayanus*, *S. castellii*, *S. kudriavzevii* and *S. kluyveri*. *S. cerevisiae* genes were used as inputs for the PRFdb (<http://prfdb.umd.edu/>) local BLAST database. The resulting homologous genes were queried for potentially significant -1 PRF signals as previously described[12]. Downstream stimulatory structures were predicted using the hotknots, unpack, or pknots algorithms. Note that no homologs were identified in the *S. castellii* and *S. kluyveri* genomes for SPR6, in the *S. kluyveri* genome for EST2, or for TBF1 in the *S. castellii* genome. Hyperlinked URLs lead to all potential folding solutions for each potential -1 PRF signal. These analyses were performed by Vivek Advani.

## Appendix 15: Interleukin Receptor Genes Containing -1 PRF

### Signals

- **CCR5** is the co-receptor for HIV-1, and individuals who do not express CCR5 are refractory to infection by the virus[262,263].
- Expression of the **IL7RA** chain is critical for mounting an effective cellular immune response, and for maintaining long term T-cell memory (reviewed in [264]).
- Interleukin-8 is a proinflammatory cytokine that is involved in chemoattraction and activation of neutrophils. The **IL8  $\alpha$ -chain** receptor (CXCR1) is a member of the G-protein-coupled receptor family that binds to IL8 with high affinity and transduces the signal through a G-protein activated second messenger system[265]. CXCR1 cleavage represents an important pathophysiologic mechanism in Cystic Fibrosis and other neutrophilic airway diseases[266], and expression of the CXCR1-Ha allele was found to protect patients against rapid progression to AIDS[267]. While CXCR1 only binds IL8, the IL8  $\beta$ -chain receptor (CXCR2) also binds GRO and NAP-2[268]. Gene knockout studies in mice revealed a profound increase in the neutrophil and B-cell populations accompanied by lymphadenopathy and splenomegaly, suggesting that this receptor is the major mediator of neutrophil migration to sites of inflammation[269].
- Interleukin-27 triggers expansion of antigen-specific naive CD4-positive T cells and promotes polarization towards a Th1 phenotype, which is critical for cell-mediated immunity. The interleukin 27 receptor  $\alpha$ -chain (**IL27R**) is widely expressed in the immune system, and IL27R deficient mice were less able to control bacterial and parasite infections[270–272]. The hyper-susceptibility of IL27R knockout mice to experimental autoimmune encephalomyelitis suggested that this receptor may regulate production of proinflammatory cytokines and antagonize T cell-mediated immune hyperactivity[271,273].
- The interleukin-2  $\gamma$ -chain receptor (IL2RG) is shared among receptors for numerous interleukins involved in T-cell memory generation (IL2, IL4, IL7, IL15 and IL21), and defects in its expression can result in Severe Combined Immune Deficiency syndrome (SCID)[264].
- Interleukin-22 is structurally related to IL10[274], and the IL22 receptor  $\alpha$ -1 chain (IL22R1) complexes with the IL10 receptor  $\beta$ -chain to serve as a common receptor chain for both IL10 and IL22[275]. IL22R1 expression is restricted to skin and respiratory and digestive tissues[276]; skin from patients with psoriasis or atopic dermatitis expressed high levels of IL22 and  $\beta$ -defensins, and it has been proposed that IL22 directly promotes the innate, nonspecific immunity of tissues[277].

## References

1. Baranov PV, Gesteland RF (2002) Recoding: translational bifurcations in gene expression. *Gene* 286: 187–201. doi:10.1016/S0378-1119(02)00423-7.
2. Hao B, Gong W, Ferguson TK, James CM, Krzycki J a, et al. (2002) A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* 296: 1462–1466. doi:10.1126/science.1069556.
3. Anderson JC, Magliery TJ, Schultz PG (2002) Exploring the limits of codon and anticodon size. *Chemistry & Biology* 9: 237–244.
4. Xie J, Schultz PG (2005) An expanding genetic code. *Methods* 36: 227–238. doi:10.1016/j.ymeth.2005.04.010.
5. Muhlrads D, Decker C (1995) Turnover mechanisms of the stable yeast PGK1 mRNA. *Molecular and Cellular Biology* 15: 2145.
6. Bekaert M, Firth AE, Zhang Y, Gladyshev VN, Atkins JF, et al. (2010) Recode-2: new design, new search tools, and many more genes. *Nucleic Acids Research* 38: D69–D74. doi:10.1093/nar/gkp788.
7. Liao PY, Choi YS, Dinman JD, Lee KH (2010) The many paths to frameshifting: kinetic modelling and analysis of the effects of different elongation steps on programmed -1 ribosomal frameshifting. *Nucleic acids research* 39: 300–312. doi:10.1093/nar/gkq761.
8. Maquat LE (2006) Nonsense-Mediated mRNA Decay. doi:10.1016/j.gde.2011.03.008.
9. Fabian MR, Sonenberg N, Filipowicz W (2010) Regulation of mRNA translation and stability by microRNAs. *Annual Review of Biochemistry* 79: 351–379. doi:10.1146/annurev-biochem-060308-103103.
10. Jacks T, Varmus HE (1985) Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. *Science* 230: 1237–1242. doi:10.1126/science.2416054.
11. Craigen WJ, Cook RG, Tate WP, Caskey CT (1985) Bacterial peptide chain release factors: conserved primary structure and possible frameshift regulation of release factor 2. *Proceedings of the National Academy of Sciences* 82: 3616.
12. Clare JJ, Belcourt M, Farabaugh PJ (1988) Efficient translational frameshifting occurs within a conserved sequence of the overlap between the two genes of a yeast Ty1 transposon. *Proceedings of the National Academy of Sciences* 85: 6816.
13. Brierley I (1995) Ribosomal frameshifting on viral RNAs. *Journal of General Virology* 76: 1885. doi:10.1099/0022-1317-76-8-1885.

14. Cobucci-Ponzano B, Conte F, Benelli D (2006) The gene of an archaeal  $\alpha$ -fucosidase is expressed by translational frameshifting. *Nucleic Acids Research* 34: 4258. doi:doi:10.1093/nar/gkl574.
15. Cobucci-Ponzano B, Rossi M, Moracci M (2005) Recoding in archaea. *Molecular microbiology* 55: 339–348. doi:10.1111/j.1365-2958.2004.04400.x.
16. Namy O, Rousset J-pierre, Naphthine S, Brierley I (2004) Reprogrammed genetic decoding in cellular gene expression. *Molecular Cell* 13: 157–168. doi:10.1016/S1097-2765(04)00031-0.
17. Dinman JD (2006) Programmed ribosomal frameshifting goes beyond viruses: organisms from all three kingdoms use frameshifting to regulate gene expression, perhaps signaling a paradigm shift. *Microbe* 1: 521. doi:10.1016/j.bbi.2008.05.010.
18. Giedroc DP, Cornish PV (2009) Frameshifting RNA pseudoknots: structure and mechanism. *Virus Research* 139: 193–208. doi:10.1016/j.virusres.2008.06.008.
19. Chang Y-F, Imam JS, Wilkinson MF (2007) The nonsense-mediated decay RNA surveillance pathway. *Annual Review of Biochemistry* 76: 51–74. doi:10.1146/annurev.biochem.76.050106.093909.
20. Conti E, Izaurralde E (2005) Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. *Current Opinion in Cell Biology* 17: 316–325.
21. Lewis BP, Green RE, Brenner SE (2003) Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proceedings of the National Academy of Sciences* 100: 189. doi:10.1073/pnas.0136770100.
22. Johansson MJO, He F, Spatrick P (2007) Association of yeast Upf1p with direct substrates of the NMD pathway. *Proceedings of the National Academy of Sciences* 104: 20872. doi:10.1073/pnas.0709257105.
23. Jacobs JL, Belew AT, Rakauskaitė R, Dinman JD (2006) Identification of functional, endogenous programmed -1 ribosomal frameshift signals in the genome of *Saccharomyces cerevisiae*. *Nucleic Acids Research* 35: 165–174. doi:10.1093/nar/gkl1033.
24. Hammell AB, Taylor RL, Peltz SW, Dinman JD (1999) Identification of putative programmed -1 ribosomal frameshift signals in large DNA databases. *Genome Research* 9: 417–427. doi:10.1101/gr.9.5.417.
25. Price DH (2008) Poised polymerases: on your mark...get set...go! *Molecular Cell* 30: 7–10. doi:10.1016/j.molcel.2008.03.001.

26. Fialcowitz EJ, Brewer BY, Keenan BP, Wilson GM (2005) A hairpin-like structure within an AU-rich mRNA-destabilizing element regulates trans-factor binding selectivity and mRNA decay kinetics. *Journal of Biological Chemistry* 280: 22406. doi:10.1074/jbc.M500618200.
27. Chen CYA, Shyu AB (1995) AU-rich elements: characterization and importance in mRNA degradation. *Trends in Biochemical Sciences* 20: 465–470. doi:10.1016/S0968-0004(00)89102-1.
28. Mayr C, Bartel DP (2009) Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 138: 673–684. doi:10.1016/j.cell.2009.06.016.
29. Hook BA, Goldstrohm AC, Seay DJ, Wickens M (2007) Two yeast PUF proteins negatively regulate a single mRNA. *The Journal of Biological Chemistry* 282: 15430–15438. doi:10.1074/jbc.M611253200.
30. Bhattacharyya SN, Habermacher R, Martine U, Closs EI, Filipowicz W (2006) Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell* 125: 1111–1124. doi:10.1016/j.cell.2006.04.031.
31. Bartel DP (2004) MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* 116: 281–297. doi:10.1016/S0092-8674(04)00045-5.
32. Hamilton A, Voinnet O, Chappell L, Baulcombe D (2002) Two classes of short interfering RNA in RNA silencing. *The EMBO Journal* 21: 4671–4679. doi:10.1093/emboj/cdf464.
33. Marshall RA, Aitken CE, Dorywalska M, Puglisi JD (2008) Translation at the single-molecule level. *Annual Review of Biochemistry* 77: 177–203. doi:10.1146/annurev.biochem.77.070606.101431.
34. Ogle JM, Brodersen DE, Clemons WM, Tarry MJ, Carter AP, et al. (2001) Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* 292: 897. doi:10.1126/science.1060612.
35. Beringer M, Rodnina MV (2007) The ribosomal peptidyl transferase. *Molecular cell* 26: 311–321. doi:10.1016/j.molcel.2007.03.015.
36. Dunkle JA, Wang L, Feldman MB, Pulk A, Chen VB, et al. (2011) Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science* 332: 981. doi:10.1126/science.1202692.
37. Rheinberger HJ, Sternbach H, Nierhaus KH (1981) Three tRNA binding sites on *Escherichia coli* ribosomes. *Proceedings of the National Academy of Sciences* 78: 5310.

38. Budkevich T, Giesebrecht J, Altman RB, Munro JB, Mielke T, et al. (2011) Structure and Dynamics of the Mammalian Ribosomal Pretranslocation Complex. *Molecular Cell* 44: 214–224. doi:10.1016/j.molcel.2011.07.040.
39. Mendez R, Richter JD (2001) Translational control by CPEB: a means to the end. *Nature Reviews Molecular Cell Biology* 2: 521–529. doi:10.1038/35080081.
40. Matlin AJ, Clark F, Smith CWJ (2005) Understanding alternative splicing: towards a cellular code. *Nature Reviews Molecular Cell Biology* 6: 386–398. doi:10.1038/nrm1645.
41. Behm-Ansmant I, Izaurralde E (2006) Quality control of gene expression: a stepwise assembly pathway for the surveillance complex that triggers nonsense-mediated mRNA decay. *Genes & Development* 20: 391. doi:10.1101/gad.1407606.
42. Maquat LE, Kiledjian M (2008) RNA turnover in eukaryotes: analysis of specialized and quality control RNA decay pathways. Elsevier/Academic Press. p. doi:10.1016/S0076-6879(08)02403-8.
43. Frischmeyer PA, van Hoof A, O'Donnell K, Guerrerio AL, Parker R, et al. (2002) An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science* 295: 2258. doi:10.1126/science.1067338.
44. Vasudevan S, Peltz SW, Wilusz CJ (2002) Non-stop decay--a new mRNA surveillance pathway. *BioEssays* 24: 785–788. doi:10.1002/bies.10153.
45. Doma MK, Parker R (2006) Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature* 440: 561–564. doi:10.1038/nature04530.
46. Baird SD, Turcotte M, Korneluk RG, Holcik M (2006) Searching for IRES. *RNA* 12: 1755–1785. doi:10.1261/rna.157806.
47. Liao PY, Gupta P, Petrov AN (2008) A new kinetic model reveals the synergistic effect of E-, P-and A-sites on +1 ribosomal frameshifting. *Nucleic Acids Research* 36: 2619. doi:10.1093/nar/gkn100.
48. Komine Y, Kitabatake M, Yokogawa T, Nishikawa K, Inokuchi H (1994) A tRNA-like structure is present in 10Sa RNA, a small stable RNA from *Escherichia coli*. *Proceedings of the National Academy of Sciences* 91: 9223.
49. Williams KP, Martindale KA, Bartel DP (1999) Resuming translation on tmRNA: a unique mode of determining a reading frame. *The EMBO Journal* 18: 5423–5433. doi:10.1093/emboj/18.19.5423.
50. Zuo X, Wang J, Yu P, Eyler D, Xu H, et al. (2010) Solution structure of the cap-independent translational enhancer and ribosome-binding element in the 3' UTR of turnip crinkle virus. *Proceedings of the National Academy of Sciences* 107: 1385. doi:10.1073/pnas.0908140107.

51. Filbin ME (2009) Toward a structural understanding of IRES RNA function. *Current Opinion in Structural Biology* 19: 267–276. doi:10.1016/j.sbi.2009.03.005.
52. Hatta M, Kohlmeier CK, Hatta Y, Ozawa M, Kawaoka Y (2009) Region required for protein expression from the stop-start pentanucleotide in the M gene of influenza B virus. *Journal of Virology* 83: 5939. doi:10.1128/JVI.00180-09.
53. Wills NM, Gesteland RF, Atkins JF (1991) Evidence that a downstream pseudoknot is required for translational read-through of the Moloney murine leukemia virus gag stop codon. *Proceedings of the National Academy of Sciences* 88: 6991.
54. Felsenstein KM, Goff SP (1992) Mutational analysis of the gag-pol junction of Moloney murine leukemia virus: requirements for expression of the gag-pol fusion protein. *Journal of Virology* 66: 6601.
55. Macejak DG, Sarnow P (1991) Internal initiation of translation mediated by the 5 leader of a cellular mRNA. *Nature* 353: 90–94. doi:10.1038/353090a0.
56. Hundsdoerfer P, Thoma C, Hentze MW (2005) Eukaryotic translation initiation factor 4G1 and p97 promote cellular internal ribosome entry sequence-driven translation. *Proceedings of the National Academy of Sciences* 102: 13421–13426. doi:10.1073/pnas.0506536102.
57. Johnson ZI, Chisholm SW (2004) Properties of overlapping genes are conserved across microbial genomes. *Genome Research* 14: 2268–2272. doi:10.1101/gr.2433104.
58. Hinnebusch AG (2005) Translational regulation of GCN4 and the general amino acid control of yeast. *Annual Review of Microbiology* 59: 407–450. doi:10.1146/annurev.micro.59.031805.133833.
59. Vatter KM, Wek RC (2004) Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proceedings of the National Academy of Sciences* 101: 11269–11274. doi:10.1073/pnas.0400541101.
60. Atkins JF, Elseviers D, Gorini L (1972) Low activity of beta-galactosidase in frameshift mutants of *Escherichia coli*. *Proceedings of the National Academy of Sciences* 69: 1192–1195.
61. Curran JF, Yarus M (1986) Base substitutions in the tRNA anticodon arm do not degrade the accuracy of reading frame maintenance. *Proceedings of the National Academy of Sciences* 83: 6538–6542.
62. Harger JW, Meskauskas A, Dinman JD (2002) An “integrated model” of programmed ribosomal frameshifting. *Trends in Biochemical Sciences* 27: 448–454. doi:10.1016/S0968-0004(02)02149-7.



63. Yueh A (2000) Translation by ribosome shunting on adenovirus and hsp70 mRNAs facilitated by complementarity to 18S rRNA. *Genes & Development*: 414–421. doi:10.1101/gad.14.4.414.
64. Dinman JD, Icho T, Wickner RB (1991) A-1 ribosomal frameshift in a double-stranded RNA virus of yeast forms a gag-pol fusion protein. *Proceedings of the National Academy of Sciences* 88: 174–178.
65. Dinman JD, Wickner RB (1992) Ribosomal frameshifting efficiency and gag/gag-pol ratio are critical for yeast M1 double-stranded RNA virus propagation. *Journal of Virology* 66: 3669.
66. Manktelow E, Shigemoto K, Brierley I (2005) Characterization of the frameshift signal of Edr, a mammalian example of programmed -1 ribosomal frameshifting. *Nucleic Acids Research* 33: 1553–1563. doi:10.1093/nar/gki299.
67. Wills NM, Moore B, Hammer A, Gesteland RF, Atkins JF (2006) A functional -1 ribosomal frameshift signal in the human paraneoplastic Ma3 gene. *The Journal of Biological Chemistry* 281: 7082–7088. doi:10.1074/jbc.M511629200.
68. Plant EP, Dinman JD (2006) Comparative study of the effects of heptameric slippery site composition on -1 frameshifting among different eukaryotic systems. *RNA* 12: 666. doi:10.1261/rna.2225206.
69. Kollmus H, Hentze MW, Hauser H (1996) Regulated ribosomal frameshifting by an RNA-protein interaction. *RNA* 2: 316.
70. Yu C-H, Noteborn MH, Pleij CW, Olsthoorn RCL (2011) Stem-loop structures can effectively substitute for an RNA pseudoknot in -1 ribosomal frameshifting. *Nucleic Acids Research*: 1–8. doi:10.1093/nar/gkr579.
71. Su L, Chen L, Egli M, Berger JM, Rich A (1999) Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot. *Nature Structural Biology* 6: 285–292. doi:10.1038/6722.
72. Jacks T, Madhani HD, Masiarz FR, Varmus HE (1988) Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell* 55: 447–458. doi:10.1016/0092-8674(88)90031-1.
73. Plant EP, Jacobs KLM, Harger JW, Meskauskas AM, Jacobs JL, et al. (2003) The 9-Å solution: How mRNA pseudoknots promote efficient programmed-1 ribosomal frameshifting. *RNA* 9: 168–174. doi:10.1261/rna.2132503.
74. Namy O, Moran SJ, Stuart DI, Gilbert RJC, Brierley I (2006) A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature* 441: 244–247. doi:10.1038/nature04735.
75. Léger M, Dulude D, Steinberg SV, Brakier-Gingras L (2007) The three transfer RNAs occupying the A, P and E sites on the ribosome are involved in viral

- programmed-1 ribosomal frameshift. *Nucleic Acids Research* 35: 5581–5592. doi:10.1093/nar/gkm578.
76. Cleland W (1975) Partition analysis and concept of net rate constants as tools in enzyme kinetics. *Biochemistry* 14: 3220–3224. doi:10.1021/bi00685a029.
  77. Clark AC, Raso SW, Sinclair JF, Ziegler MM, Chaffotte AF, et al. (1997) Kinetic mechanism of luciferase subunit folding and assembly. *Biochemistry* 36: 1891–1899. doi:10.1021/bi962477m.
  78. Plant EP, Wang P, Jacobs JL, Dinman JD (2004) A programmed -1 ribosomal frameshift signal can function as a cis-acting mRNA destabilizing element. *Nucleic acids research* 32: 784–790. doi:10.1093/nar/gkh256.
  79. Losson R, Lacroute F (1979) Interference of nonsense mutations with eukaryotic messenger RNA stability. *Proceedings of the National Academy of Sciences* 76: 5134–5137.
  80. Arnold TE, Yu J, Belasco JG (1998) mRNA stabilization by the ompA 5' untranslated region: two protective elements hinder distinct pathways for mRNA degradation. *RNA* 4: 319–330.
  81. Hogg JR, Goff SP (2010) Upf1 Senses 3'UTR Length to Potentiate mRNA Decay. *Cell* 143: 379–389. doi:10.1016/j.cell.2010.10.005.
  82. Alberghina L, Albermann K, Albers M, Aldea M, Alexandraki D, et al. (1997) The yeast genome directory. *Nature* 387: 5.
  83. Spingola M, Grate L, Haussler D, Ares M, Ares Jr M (1999) Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA* 5: 221–234.
  84. Ares Jr M, Grate L, Pauling MH (1999) A handful of intron-containing genes produces the lion's share of yeast mRNA. *RNA* 5: 1138.
  85. He F, Li X, Spatrick P, Casillo R, Dong S, et al. (2003) Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5'to 3'mRNA decay pathways in yeast. *Molecular Cell* 12: 1439–1452. doi:10.1016/S1097-2765(03)00446-5.
  86. Lew JE, Enomoto S, Berman J (1998) Telomere length regulation and telomeric chromatin require the nonsense-mediated mRNA decay pathway. *Molecular and Cellular Biology* 18: 6121.
  87. Czaplinski K, Ruiz-Echevarria MJ, Paushkin SVV, Han X, Weng Y, et al. (1998) The surveillance complex interacts with the translation release factors to enhance termination and degrade aberrant mRNAs. *Genes & Development* 12: 1665–1677. doi:10.1101/gad.12.11.1665.

88. Kervestin S, Mangus D a DA, Ghosh S, Jacobson A, Amrani N, et al. (2004) A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature* 432: 112–118. doi:10.1038/nature03060.
89. Ishigaki Y, Li X, Serin G (2001) Evidence for a Pioneer Round of mRNA Translation:: mRNAs Subject to Nonsense-Mediated Decay in Mammalian Cells Are Bound by CBP80 and CBP20. *Cell* 106: 607–617. doi:10.1016/S0092-8674(01)00475-5.
90. Lejeune F, Ranganathan A., Maquat LE (2004) eIF4G is required for the pioneer round of translation in mammalian cells. *Nature Structural & Molecular Biology* 11: 992–1000. doi:10.1038/nsmb824.
91. González C, Ruiz-Echevarría M, Vasudevan S, Henry MF, Peltz SW (2000) The yeast hnRNP-like protein Hrp1/Nab4 marks a transcript for nonsense-mediated mRNA decay. *Molecular cell* 5: 489–499. doi:10.1016/S1097-2765(00)80443-8.
92. Zhang S, Ruiz-Echevarria MJ, Quan Y, Peltz SW, Hagan KW (1995) Identification and characterization of a sequence motif involved in nonsense-mediated mRNA decay. *Molecular and Cellular Biology* 15: 2231.
93. Hosoda N, Kobayashi T, Uchida N, Funakoshi Y, Kikuchi Y, et al. (2003) Translation termination factor eRF3 mediates mRNA decay through the regulation of deadenylation. *Journal of Biological Chemistry* 278: 38287. doi:10.1074/jbc.C300300200.
94. Brogna S, Wen J (2009) Nonsense-mediated mRNA decay (NMD) mechanisms. *Nature Structural & Molecular Biology* 16: 107–113. doi:10.1038/nsmb.1550.
95. Kim YK, Furic L, DesGroseillers L, Maquat LE (2005) Mammalian Staufen1 recruits Upf1 to specific mRNA 3'UTRs so as to elicit mRNA decay. *Cell* 120: 195–208. doi:10.1016/j.cell.2004.11.050.
96. Applequist SE, Selg M, Raman C, Jäck HM (1997) Cloning and characterization of HUPF1, a human homolog of the *Saccharomyces cerevisiae* nonsense mRNA-reducing UPF1 protein. *Nucleic Acids Research* 25: 814–821. doi:10.1093/nar/25.4.814.
97. Cheng Z, Muhlrads D, Lim MK, Parker R, Song H (2007) Structural and functional insights into the human Upf1 helicase core. *The EMBO Journal* 26: 253–264. doi:10.1038/sj.emboj.7601464.
98. Ivanov PV, Gehring NH, Kunz JB, Hentze MW, Kulozik AE (2008) Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. *The EMBO Journal* 27: 736–747. doi:10.1038/emboj.2008.17.
99. Clerici M, Mourão A, Gutsche I, Gehring NH, Hentze MW, et al. (2009) Unusual bipartite mode of interaction between the nonsense-mediated decay factors,

UPF1 and UPF2. The EMBO Journal 28: 2293–2306.  
doi:10.1038/emboj.2009.175.

100. Chakrabarti S, Jayachandran U, Bonneau F, Fiorini F, Basquin C, et al. (2011) Molecular Mechanisms for the RNA-Dependent ATPase Activity of Upf1 and Its Regulation by Upf2. *Molecular Cell* 41: 693–703.  
doi:10.1016/j.molcel.2011.02.010.
101. Shirley RL, Ford AS, Richards MR, Albertini M, Culbertson MR (2002) Nuclear Import of Upf3p Is Mediated by Importin- $\alpha$ - $\beta$  and Export to the Cytoplasm Is Required for a Functional Nonsense-Mediated mRNA Decay Pathway in Yeast. *Genetics* 161: 1465.
102. Leshin JA, Heselpoth R, Belew AT, Dinman JD (2011) High throughput structural analysis of yeast ribosomes using hSHAPE. *RNA Biology* 8.
103. Hir L, Nguyen LS, Huang L, Ge J, Chan W-K, et al. (2009) A UPF3-mediated regulatory switch that maintains RNA surveillance. *Nature Structural & Molecular Biology* 16: 747–753. doi:10.1038/nsmb.1612.
104. Cali BM, Kuchma SL, Latham J, Anderson P (1999) smg-7 Is Required for mRNA Surveillance in *Caenorhabditis elegans*. *Genetics*: 605–616.
105. Houseley J, LaCava J, Tollervey D (2006) RNA-quality control by the exosome. *Nature Reviews Molecular Cell Biology* 7: 529–539. doi:10.1038/nrm1964.
106. Kim VN, Yong J, Kataoka N, Abel L, Diem MD, et al. (2001) The Y14 protein communicates to the cytoplasm the position of exon-exon junctions. *The EMBO Journal* 20: 2062–2068. doi:10.1093/emboj/20.8.2062.
107. Lau C-kong, Diem MD, Dreyfuss G, Duyne GDV (2003) Structure of the Y14-Magoh Core of the Exon Junction Complex. *Current Biology* 13: 933–941.  
doi:10.1016/S.
108. Collier JM, Tucker M, Sheth U, Valencia-Sanchez MA, Parker R (2001) The DEAD box helicase, Dhh1p, functions in mRNA decapping and interacts with both the decapping and deadenylase complexes. *RNA* 7: 1717–1727.
109. Orban TII, Izaurralde E (2005) Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. *RNA* 11: 459. doi:10.1261/rna.7231505.
110. Büttner K, Wenig K, Hopfner K-P (2005) Structural framework for the mechanism of archaeal exosomes in RNA processing. *Molecular Cell* 20: 461–471.  
doi:10.1016/j.molcel.2005.10.018.
111. Symmons MF, Jones GH, Luisi BF (2000) A duplicated fold is the structural basis for polynucleotide phosphorylase catalytic activity, processivity, and regulation. *Structure* 8: 1215–1226.

112. Parker R, Song H (2004) The enzymes and control of eukaryotic mRNA turnover. *Nature Structural & Molecular Biology* 11: 121–127. doi:doi:10.1038/nsmb724.
113. Kloosterman WP, Plasterk RH (2006) The diverse functions of microRNAs in animal development and disease. *Developmental Cell* 11: 441–450. doi:10.1016/j.devcel.2006.09.009.
114. Bushati N, Cohen SM (2007) microRNA functions. *Annual Review of Cell and Developmental Biology* 23: 175–205. doi:10.1146/annurev.cellbio.23.090506.123406.
115. Bao N, Lye K-W, Barton MK (2004) MicroRNA binding sites in Arabidopsis class III HD-ZIP mRNAs are required for methylation of the template chromosome. *Developmental Cell* 7: 653–662. doi:10.1016/j.devcel.2004.10.003.
116. Henke JI, Goergen D, Zheng J, Song Y, Schüttler CG, et al. (2008) microRNA-122 stimulates translation of hepatitis C virus RNA. *The EMBO Journal* 27: 3300–3310. doi:10.1038/emboj.2008.244.
117. Nathans R, Chu C-ying, Serquina AK, Lu C-C, Cao H, et al. (2009) Cellular microRNA and P bodies modulate host-HIV-1 interactions. *Molecular Cell* 34: 696–709. doi:10.1016/j.molcel.2009.06.003.
118. Bennasser Y, Le S-Y, Yeung ML, Jeang K-T (2004) HIV-1 encoded candidate micro-RNAs and their cellular targets. *Retrovirology* 1: 43. doi:10.1186/1742-4690-1-43.
119. Nair V, Zavolan M (2006) Virus-encoded microRNAs: novel regulators of gene expression. *Trends in Microbiology* 14: 169–175. doi:10.1016/j.tim.2006.02.007.
120. Cai X, Hagedorn CH, Cullen BR (2004) Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* 10: 1957–1966. doi:10.1261/rna.7135204.
121. Yekta S, Shih I-hung, Bartel DP (2004) MicroRNA-directed cleavage of HOXB8 mRNA. *Science* 304: 594–596. doi:10.1126/science.1097434.
122. Djuranovic S, Nahvi A, Green R (2011) A parsimonious model for gene regulation by miRNAs. *Science* 331: 550–553. doi:10.1126/science.1191138.
123. Humphreys DT, Westman BJ, Martin DIK, Preiss T (2005) MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proceedings of the National Academy of Sciences* 102: 16961–16966. doi:10.1073/pnas.0506482102.
124. Petersen CP, Bordeleau M-E, Pelletier J, Sharp P (2006) Short RNAs repress translation after initiation in mammalian cells. *Molecular Cell* 21: 533–542. doi:10.1016/j.molcel.2006.01.031.

125. Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120: 15–20. doi:10.1016/j.cell.2004.12.035.
126. Krützfeldt J, Rajewsky N, Braich R, Rajeev KG, Tuschl T, et al. (2005) Silencing of microRNAs in vivo with “antagomirs”. *Nature* 438: 685–689. doi:10.1038/nature04303.
127. Altschul SF, Madden TL (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25: 3389–3402.
128. Eddy SR (2011) Accelerated Profile HMM Searches. *PLoS Computational Biology* 7: e1002195. doi:10.1371/journal.pcbi.1002195.
129. Macke TJ, Ecker DJ, Gutell RR, Gautheret D, Case DA, et al. (2001) RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Research* 29: 4724–4735. doi:10.1093/nar/29.22.4724.
130. Webb CHT, Riccitelli NJ, Ruminski DJ, Lupták A (2009) Widespread occurrence of self-cleaving ribozymes. *Science* 326: 953.
131. Riccitelli NJ, Lupták A (2010) Computational discovery of folded RNA domains in genomes and in vitro selected libraries. *Methods* 52: 133–140. doi:10.1016/j.ymeth.2010.06.005.
132. Hammann C, Westhof E (2007) Searching genomes for ribozymes and riboswitches. *Genome Biology* 8: 210. doi:10.1186/gb-2007-8-4-210.
133. Hamilton RS, Hartwood E, Vendra G, Jones C, Van De Bor V, et al. (2009) A bioinformatics search pipeline, RNA2DSearch, identifies RNA localization elements in *Drosophila* retrotransposons. *RNA* 15: 200–207. doi:10.1261/rna.1264109.
134. Mathews DH, Sabina J, Zuker M, Turner DH (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure1. *Journal of Molecular Biology* 288: 911–940.
135. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* 31: 3406–3415. doi:10.1093/nar/gkg595.
136. Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL (2008) The Vienna RNA websuite. *Nucleic Acids Research* 36: W70–W74. doi:10.1093/nar/gkn188.
137. Nawrocki EP, Kolbe DL, Eddy SR (2009) Infernal 1.0: inference of RNA alignments. *Bioinformatics* 25: 1335–1337. doi:10.1093/bioinformatics/btp157.

138. Durbin R, Eddy S, Krogh A, Mitchison G (1998) Biological sequence analysis: Probabilistic models of proteins and nucleic acids. Cambridge UK, New York: Cambridge university press. p.
139. Rivas E, Lang R, Eddy SR (2011) A range of complex probabilistic models for RNA secondary structure prediction that include the nearest-neighbor model and more. *RNA* 18: 193–212. doi:10.1261/rna.030049.111.
140. Knudsen B (2003) Pfold: RNA secondary structure prediction using stochastic context-free grammars. *Nucleic Acids Research* 31: 3423–3428. doi:10.1093/nar/gkg614.
141. Yao Z, Weinberg Z, Ruzzo WL (2006) CMfinder--a covariance model based RNA motif finding algorithm. *Bioinformatics* 22: 445–452. doi:10.1093/bioinformatics/btk008.
142. Lyngsø RB, Pedersen CNS (2000) Pseudoknots in RNA secondary structures. *Proceedings of the Fourth Annual International Conference on Computational Molecular Biology*: 201–209. doi:10.1145/332306.332551.
143. Rivas E, Eddy SR (1999) A dynamic programming algorithm for RNA structure prediction including pseudoknots1. *Journal of Molecular Biology* 285: 2053–2068. doi:10.1006/jmbi.1998.2436.
144. Freier SM, Kierzek R, Jaeger J a, Sugimoto N, Caruthers MH, et al. (1986) Improved free-energy parameters for predictions of RNA duplex stability. *Proceedings of the National Academy of Sciences* 83: 9373–9377.
145. Ruan J, Stormo GD, Zhang W (2003) An iterative loop matching approach to the prediction of RNA secondary structures with pseudoknots. *Computational Systems Bioinformatics*: 519–520. doi:10.1109/CSB.2003.1227394.
146. Dirks RM, Pierce NA (2004) An algorithm for computing nucleic acid base-pairing probabilities including pseudoknots. *Journal of Computational Chemistry* 25: 1295–1304. doi:10.1002/jcc.20057.
147. Bon M, Orland H (2011) TT2NE: a novel algorithm to predict RNA secondary structures with pseudoknots. *Nucleic Acids Research* 39. doi:10.1093/nar/gkr240.
148. Freyhult E, Gardner PP, Moulton V (2005) A comparison of RNA folding measures. *BMC Bioinformatics* 6: 241. doi:10.1186/1471-2105-6-241.
149. Workman C, Krogh A (1999) No evidence that mRNAs have lower folding free energies than random sequences with the same dinucleotide distribution. *Nucleic Acids Research* 27: 4816.
150. Clote P, Ferré F, Kranakis E, Krizanc D (2005) Structural RNA has lower folding energy than random RNA of the same dinucleotide frequency. *RNA* 11: 578–591. doi:10.1261/rna.7220505.

151. Bonnet E, Wuyts J, Rouzé P, Van de Peer Y (2004) Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics* 20: 2911–2917. doi:10.1093/bioinformatics/bth374.
152. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz S a, et al. (2002) The Bioperl toolkit: Perl modules for the life sciences. *Genome Research* 12: 1611–1618. doi:10.1101/gr.361602.
153. Wiese KC, Glen E (2006) jViz. Rna-An interactive graphical tool for visualizing RNA secondary structure including pseudoknots. In: *Computer-Based Medical Systems, 2006. CBMS 2006. 19th IEEE International Symposium*. IEEE. pp. 659–664.
154. Byun Y, Han K (2009) PseudoViewer3: generating planar drawings of large-scale RNA structures with pseudoknots. *Bioinformatics* 25: 1435–1437. doi:10.1093/bioinformatics/btp252.
155. Gesteland RF, Atkins JF (1996) Recoding: dynamic reprogramming of translation. *Annual Review of Biochemistry* 65: 741–768. doi:10.1146/annurev.bi.65.070196.003521.
156. Farabaugh PJ (1996) Programmed translational frameshifting. *Microbiology and Molecular Biology Reviews* 60: 103.
157. Baranov PV, Gurvich OL, Hammer AW, Gesteland RF, Atkins JF (2003) RECODE 2003. *Nucleic Acids Research* 31: 87–89. doi:10.1093/nar/gkg024.
158. Baranov PV, Gurvich OL, Fayet O, Prère MF, Miller WA, et al. (2001) RECODE: a database of frameshifting, bypassing and codon redefinition utilized for gene expression. *Nucleic Acids Research* 29: 264–267. doi:10.1093/nar/29.1.264.
159. Moon S, Byun Y, Han K (2007) FSDB: a frameshift signal database. *Computational Biology and Chemistry* 31: 298–302. doi:10.1016/j.compbiolchem.2007.05.004.
160. Han K, Byun Y (2003) PseudoViewer2: visualization of RNA pseudoknots of any type. *Nucleic Acids Research* 31: 3432–3440. doi:10.1093/nar/gkg539.
161. Firth AE (2006) Detecting overlapping coding sequences in virus genomes. *BMC bioinformatics* 7: 75. doi:doi:10.1186/1471-2105-7-75.
162. Bruford E, Lush MJ, Wright MW, Sneddon TP, Povey S, et al. (2008) The HGNC Database in 2008: a resource for the human genome. *Nucleic Acids Research* 36: D445–8. doi:10.1093/nar/gkm881.
163. Dirks RM, Pierce NA (2003) A partition function algorithm for nucleic acid secondary structure including pseudoknots. *Journal of Computational Chemistry* 24: 1664–1677. doi:10.1002/jcc.10296.



164. Ren J, Rastegari B, Condon A, Hoos HH (2005) HotKnots: heuristic prediction of RNA secondary structures including pseudoknots. *RNA* 11: 1494–1504. doi:10.1261/rna.7284905.
165. Mathews DH, Turner DH (2006) Prediction of RNA secondary structure by free energy minimization. *Current Opinion in Structural Biology* 16: 270–278. doi:10.1016/j.sbi.2006.05.010.
166. Lopinski JD, Dinman JD, Bruenn JA (2000) Kinetics of ribosomal pausing during programmed-1 translational frameshifting. *Molecular and Cellular Biology* 20: 1095. doi:10.1128/MCB.20.4.1095-1103.2000.
167. Kontos H, Naphthine S, Brierley I (2001) Ribosomal pausing at a frameshifter RNA pseudoknot is sensitive to reading phase but shows little correlation with frameshift efficiency. *Molecular and Cellular Biology* 21: 8657. doi:10.1128/MCB.21.24.8657-8670.2001.
168. Shah AA, Giddings MC, Parvaz JB, Gesteland RF, Atkins JF, et al. (2002) Computational identification of putative programmed translational frameshift sites. *Bioinformatics* 18: 1046. doi:10.1093/bioinformatics/18.8.1046.
169. Gao X, Havecker ER, Baranov PV, Atkins JF, Voytas DF (2003) Translational recoding signals between gag and pol in diverse LTR retrotransposons. *RNA* 9: 1422–1430. doi:10.1261/rna.5105503.
170. Gurvich OL, Baranov PV, Zhou J, Hammer AW, Gesteland RF, et al. (2003) Sequences that direct significant levels of frameshifting are frequent in coding regions of *Escherichia coli*. *The EMBO Journal* 22: 5941–5950. doi:10.1093/emboj/cdg561.
171. Moon S, Byun Y, Kim HJ, Jeong S (2004) Predicting genes expressed via –1 and +1 frameshifts. *Nucleic Acids Research* 32: 4884–4892.
172. Belew AT, Hepler NL, Jacobs JL, Dinman JD (2008) PRFdb: A database of computationally predicted eukaryotic programmed -1 ribosomal frameshift signals. *BMC Genomics* 9: 339. doi:10.1186/1471-2164-9-339.
173. Counter CM, Meyerson M, Eaton EN, Weinberg RA (1997) The catalytic subunit of yeast telomerase. *Proceedings of the National Academy of Sciences* 94: 9202–9207.
174. Lendvay TS, Morris DK, Sah J (1996) Senescence mutants of *Saccharomyces cerevisiae* with a defect in telomere replication identify three additional EST genes. *Genetics* 144: 1399.
175. Taggart AKP (2003) Telomerase: what are the Est proteins doing? *Current Opinion in Cell Biology* 15: 275–280. doi:10.1016/S0955-0674(03)00040-1.

176. Osterhage JL, Friedman KL (2009) Chromosome end maintenance by telomerase. *The Journal of Biological Chemistry* 284: 16061–16065. doi:10.1074/jbc.R900011200.
177. Dahlseid JN, Lew-Smith J, Lelivelt MJ, Enomoto S, Ford A, et al. (2003) mRNAs encoding telomerase components and regulators are controlled by UPF genes in *Saccharomyces cerevisiae*. *Eukaryotic Cell* 2: 134–142. doi:10.1128/EC.2.1.134-142.2003.
178. Smogorzewska A, de Lange T (2004) Regulation of telomerase by telomeric proteins. *Annual review of biochemistry* 73: 177–208. doi:10.1146/annurev.biochem.73.071403.160049.
179. Isken O, Maquat LE (2007) Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes & development* 21: 1833–1856. doi:10.1101/gad.1566807.
180. Michiels PJA, Versleijen AAM, Verlaan PW, Pleij CWA, Hilbers CW, et al. (2001) Solution structure of the pseudoknot of SRV -1 RNA, involved in ribosomal frameshifting. *Journal of Molecular Biology* 310: 1109–1123. doi:10.1006/jmbi.2001.4823.
181. Cornish PV, Hennig M, Giedroc DP (2005) A loop 2 cytidine-stem 1 minor groove interaction as a positive determinant for pseudoknot-stimulated -1 ribosomal frameshifting. *Proceedings of the National Academy of Sciences* 102: 12694. doi:10.1073/pnas.0506166102.
182. Chou MY, Lin SC (2010) Stimulation of -1 programmed ribosomal frameshifting by a metabolite-responsive RNA pseudoknot. *RNA* 16: 1236–1244. doi:10.1261/rna.1922410.
183. Olsthoorn RCL, Reumerman R (2010) Functional analysis of the SRV -1 RNA frameshifting pseudoknot. *Nucleic Acids Research* 38: 7665–7672. doi:10.1093/nar/gkq629.
184. Brierley I, Jenner AJ, Inglis SC (1992) Mutational analysis of the slippery-sequence component of a coronavirus ribosomal frameshifting signal. *Journal of Molecular Biology* 227: 463–479. doi:10.1016/0022-2836(91)90361-9.
185. Morris DK, Lundblad V (1997) Programmed translational frameshifting in a gene required for yeast telomere replication. *Current Biology* 7: 969–976. doi:10.1016/S0960-9822(06)00416-7.
186. Mozdy AD, Cech TR (2006) Low abundance of telomerase in yeast: implications for telomerase haploinsufficiency. *RNA* 12: 1721–1737. doi:10.1261/rna.134706.
187. Koering CE, Fourel G, Binet-Brasselet, Laroche T, Klein F, et al. (2000) Identification of high affinity Tbf1p-binding sites within the budding yeast genome. *Nucleic Acids Research* 28: 2519–2526.

188. Enomoto S, Glowczewski L, Lew-Smith J, Berman JG (2004) Telomere cap components influence the rate of senescence in telomerase-deficient yeast cells. *Molecular and Cellular Biology* 24: 837. doi:10.1128/MCB.24.2.837-845.2004.
189. Liao XH, Zhang ML, Yang CP, Xu LX, Zhou JQ (2005) Characterization of recombinant *Saccharomyces cerevisiae* telomerase core enzyme purified from yeast. *Biochemical Journal* 390: 169. doi:10.1042/BJ20050208.
190. Grandin N, Damon C, Charbonneau M (2000) Cdc13 cooperates with the yeast Ku proteins and Stn1 to regulate telomerase recruitment. *Molecular and Cellular Biology* 20: 8397. doi:10.1128/MCB.20.22.8397-8408.2000.Updated.
191. Dinman JD, Connor MO, Farabaugh PJ (2010) Recoding: Expansion of Decoding Rules Enriches Gene Expression. *Nucleic Acids and Molecular Biology* 24: 321–344. doi:10.1007/978-0-387-89382-2.
192. Dinman JD, Wickner RB (1994) Translational Maintenance of Frame: Mutants of *Saccharomyces cerevisiae* with Altered -1 Ribosomal Frameshifting Efficiencies. *Genetics* 136: 75–86.
193. Askree SH, Yehuda T, Smolikov S, Gurevich R, Hawk J, et al. (2004) A genome-wide screen for *Saccharomyces cerevisiae* deletion mutants that affect telomere length. *Proceedings of the National Academy of Sciences* 101: 8658–8663. doi:10.1073/pnas.0401263101.
194. Watanabe M, Watanabe D, Nogami S, Morishita S, Ohya Y (2009) Comprehensive and quantitative analysis of yeast deletion mutants defective in apical and isotropic bud growth. *Current Genetics* 55: 365–380. doi:10.1007/s00294-009-0251-0.
195. Kallal L a, Bhattacharyya M, Grove SN, Iannacone RF, Pugh T a, et al. (1990) Functional analysis of the sporulation-specific SPR6 gene of *Saccharomyces cerevisiae*. *Current Genetics* 18: 293–301.
196. Sambrook J, Russell DW (2001) *Molecular cloning: a laboratory manual*. Cold Spring Harbor, NY: Cold spring harbor laboratory press. p.
197. Ito H, Fukuda Y, Murata K, Kimura A (1983) Transformation of intact yeast cells treated with alkali cations. *Journal of Bacteriology* 153: 163–168.
198. Wickner RB, Leibowitz MJ (1976) Two chromosomal genes required for killing expression in killer strains of *Saccharomyces cerevisiae*. *Genetics* 82: 429.
199. Cui Y, Hagan KW, Zhang S, Peltz SW (1995) Identification and characterization of genes that are required for the accelerated degradation of mRNAs containing a premature translational termination codon. *Genes & Development* 9: 423. doi:10.1101/gad.9.4.423.

200. Harger JW, Dinman JD (2003) An in vivo dual-luciferase assay system for studying translational recoding in the yeast *Saccharomyces cerevisiae*. *RNA* 9: 1019–1024. doi:10.1261/rna.5930803.
201. Atkins JF, Baranov PV (2010) The distinction between recoding and codon reassignment. *Genetics* 185: 1535–1536. doi:10.1534/genetics.110.119016.
202. Belew AT, Advani VM, Dinman JD (2011) Endogenous ribosomal frameshift signals operate as mRNA destabilizing elements through at least two molecular pathways in yeast. *Nucleic Acids Research* 39: 2799. doi:10.1093/nar/gkq1220.
203. Atchison RE, Gosling J, Monteclaro FS, Franci C, Digilio L, et al. (1996) Multiple extracellular elements of CCR5 and HIV-1 entry: dissociation from response to chemokines. *Science* 274: 1924–1926.
204. Grentzmann G, Ingram JA, Kelley PJ, Gesteland RF, Atkins JF (1998) A dual-luciferase reporter system for studying recoding signals. *RNA* 4: 479–486.
205. Jacobs JL, Dinman JD (2004) Systematic analysis of bicistronic reporter assay data. *Nucleic Acids Research* 32: e160. doi:10.1093/nar/gnh157.
206. Guhaniyogi J (2001) Regulation of mRNA stability in mammalian cells. *Gene* 265: 11–23. doi:10.1016/S0378-1119(01)00350-X.
207. Ysla RM, Wilson GM, Brewer G, Riza YM (2008) Assays of adenylate uridylate-rich element-mediated mRNA decay in cells. *Methods in Enzymology* 449: 47–71. doi:10.1016/S0076-6879(08)02403-8.
208. Yepiskoposyan H, Aeschimann F, Nilsson D, Okoniewski M, Muhlemann O, et al. (2011) Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. *RNA* 17: 2108–2118. doi:10.1261/rna.030247.111.
209. Henderson CM, Anderson CB, Howard MT (2006) Antisense-induced ribosomal frameshifting. *Nucleic Acids Research* 34: 4302–4310. doi:10.1093/nar/gkl531.
210. Vickers TA, Ecker DJ (1992) Enhancement of ribosomal frameshifting by oligonucleotides targeted to the HIV gag-pol region. *Nucleic Acids Research* 20: 3945–3953.
211. Howard MT, Gesteland RF, Atkins JF (2004) Efficient stimulation of site-specific ribosome frameshifting by antisense oligonucleotides. *RNA* 10: 1653–1661. doi:10.1261/rna.7810204.
212. Yu C-H, Noteborn MH, Olsthoorn RC (2010) Stimulation of ribosomal frameshifting by antisense LNA. *Nucleic Acids Research* 38: 8277–8283. doi:10.1093/nar/gkq650.

213. Platt E, Wehrly K, Kuhmann S (1998) Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1. *Journal of Virology* 72: 2855.
214. Orom UA, Lund AH (2007) Isolation of microRNA targets using biotinylated synthetic microRNAs. *Methods* 43: 162–165. doi:10.1016/j.ymeth.2007.04.007.
215. Wilkinson KA, Merino EJ (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nature protocols* 1: 1610–1616. doi:10.1038/nprot.2006.249.
216. Marko JF, Siggia ED (1995) Stretching DNA. *Macromolecules* 28: 8759–8770. doi:10.1021/ma00130a008.
217. de Messieres M, Brawn-Cinani B, La Porta A (2011) Measuring the folding landscape of a harmonically constrained biopolymer. *Biophysical Journal* 100: 2736–2744. doi:10.1016/j.bpj.2011.03.067.
218. Clark MB, Jänicke M, Gottesbühren U, Kleffmann T, Legge M, et al. (2007) Mammalian gene PEG10 expresses two reading frames by high efficiency -1 frameshifting in embryonic-associated tissues. *The Journal of biological chemistry* 282: 37359–37369. doi:10.1074/jbc.M705676200.
219. Khabar KS a, Young H a (2007) Post-transcriptional control of the interferon system. *Biochimie* 89: 761–769. doi:10.1016/j.biochi.2007.02.008.
220. Dinman JD (2009) The eukaryotic ribosome: current status and challenges. *The Journal of Biological Chemistry* 284: 11761–11765. doi:10.1074/jbc.R800074200.
221. Kobayashi Y, Zhuang J, Peltz S, Dougherty J (2010) Identification of a cellular factor that modulates HIV-1 programmed ribosomal frameshifting. *Journal of Biological Chemistry* 285: 19776.
222. Shigemoto K, Brennan K, Walls E, Watson CJJ, Stott DW, et al. (2001) Identification and characterisation of a developmentally regulated mammalian gene that utilises 21 programmed ribosomal frameshifting. *Nucleic Acids Research* 29: 4079–4088.
223. Jack K, Bellodi C, Landry DM, Niederer RO, Meskauskas AM, et al. (2011) rRNA pseudouridylation defects affect ribosomal ligand binding and translational fidelity from yeast to human cells. *Molecular Cell* 44: 660–666. doi:10.1016/j.molcel.2011.09.017.
224. Plant EP, Dinman JD (2005) Torsional restraint: a new twist on frameshifting pseudoknots. *Nucleic Acids Research* 33: 1825. doi:10.1093/nar/gki329.
225. Olsthoorn RC, Laurs M, Sohet F, Hilbers CW, Heus HA, et al. (2004) Novel application of sRNA: stimulation of ribosomal frameshifting. *RNA* 10: 1702–1703.

226. Ahn D-G, Lee W, Choi J-K, Kim S-J, Plant EP, et al. (2011) Interference of ribosomal frameshifting by antisense peptide nucleic acids suppresses SARS coronavirus replication. *Antiviral Research* 91: 1–10. doi:10.1016/j.antiviral.2011.04.009.
227. Forman JJ, Legesse-Miller A, Collier H a (2008) A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. *Proceedings of the National Academy of Sciences* 105: 14879–14884. doi:10.1073/pnas.0803230105.
228. Moretti F, Thermann R, Hentze MW (2010) Mechanism of translational regulation by miR-2 from sites in the 5' untranslated region or the open reading frame. *RNA* 16: 2493–2502. doi:10.1261/rna.2384610.
229. Duursma AM, Kedde M, Schrier M, le Sage C, Agami R (2008) miR-148 targets human DNMT3b protein coding region. *RNA* 14: 872–877. doi:10.1261/rna.972008.
230. Niu Y, Mo D, Qin L, Wang C, Li A, et al. (2011) Lipopolysaccharide-induced miR-1224 negatively regulates tumour necrosis factor- $\alpha$  gene expression by modulating Sp1. *Immunology* 133: 8–20. doi:10.1111/j.1365-2567.2010.03374.x.
231. Dudzic E, Miah S, Choudhry HMZ, Owen HC, Blizzard S, et al. (2011) Hypermethylation of CpG islands and shores around specific microRNAs and mirtrons is associated with the phenotype and presence of bladder cancer. *Clinical Cancer Research* 17: 1287–1296. doi:10.1158/1078-0432.CCR-10-2017.
232. Te JL, Dozmorov IM, Guthridge JM, Nguyen KL, Cavett JW, et al. (2010) Identification of unique microRNA signature associated with lupus nephritis. *PloS One* 5: e10344. doi:10.1371/journal.pone.0010344.
233. Snowdon J, Zhang X, Childs T, Tron V a, Feilotter H (2011) The microRNA-200 family is upregulated in endometrial carcinoma. *PloS One* 6: e22828. doi:10.1371/journal.pone.0022828.
234. Cheng H, Zhang L, Cogdell DE, Zheng H, Schetter AJ, et al. (2011) Circulating plasma MiR-141 is a novel biomarker for metastatic colon cancer and predicts poor prognosis. *PloS One* 6: e17745. doi:10.1371/journal.pone.0017745.
235. Liu X-G, Zhu W-Y, Huang Y-Y, Ma L-N, Zhou S-Q, et al. (2011) High expression of serum miR-21 and tumor miR-200c associated with poor prognosis in patients with lung cancer. *Medical Oncology*. doi:10.1007/s12032-011-9923-y.
236. Kim Y-K, Yeo J, Ha M, Kim B, Kim VNN (2011) Cell Adhesion-Dependent Control of MicroRNA Decay. *Molecular Cell* 43: 1005–1014. doi:10.1016/j.molcel.2011.07.031.
237. Houck-Loomis B, Durney M a., Salguero C, Shankar N, Nagle JM, et al. (2011) An equilibrium-dependent retroviral mRNA switch regulates translational recoding. *Nature* 480: 561–564. doi:10.1038/nature10657.

238. Plant EP, Rakauskaite R, Taylor DR, Dinman JD (2010) Achieving a golden mean: mechanisms by which coronaviruses ensure synthesis of the correct stoichiometric ratios of viral proteins. *Journal of Virology* 84: 4330–4340. doi:10.1128/JVI.02480-09.
239. Biswas P, Jiang X, Pacchia AL, Dougherty JP, Peltz SW (2004) The human immunodeficiency virus type 1 ribosomal frameshifting site is an invariant sequence determinant and an important target for antiviral therapy. *Journal of Virology* 78: 2082.
240. Wilkinson KA, Gorelick RJ, Vasa SM, Guex N, Rein A, et al. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biology* 6: e96.
241. Ingolia NT, Ghaemmamghami S, Newman JRS, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324: 218–223. doi:10.1126/science.1168978.
242. Kertesz M, Wan Y, Mazor E, Rinn JL, Nutter RC, et al. (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467: 103–107. doi:10.1038/nature09322.
243. Ingolia NT, Lareau LF, Weissman JS (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147: 789–802. doi:10.1016/j.cell.2011.10.002.
244. Guo H, Ingolia NT, Weissman JS, Bartel DP (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466: 835–840. doi:10.1038/nature09267.
245. Zhang W, Wagner BJ, Ehrenman K, Schaefer a W, DeMaria CT, et al. (1993) Purification, characterization, and cDNA cloning of an AU-rich element RNA-binding protein, AUF1. *Molecular and Cellular Biology* 13: 7652–7665. doi:doi: 10.1128/MCB.13.12.7652.
246. Horowitz H, Haber JE (1984) Subtelomeric regions of yeast chromosomes contain a 36 base-pair tandemly repeated sequence. *Nucleic Acids Research* 12: 7105–7121.
247. Hoshida Y, Nijman SMB, Kobayashi M, Chan J a, Brunet J-P, et al. (2009) Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. *Cancer Research* 69: 7385–7392. doi:10.1158/0008-5472.CAN-09-1089.
248. Su H, Yang J-R, Xu T, Huang J, Xu L, et al. (2009) MicroRNA-101, down-regulated in hepatocellular carcinoma, promotes apoptosis and suppresses tumorigenicity. *Cancer Research* 69: 1135–1142. doi:10.1158/0008-5472.CAN-08-2886.

249. Gey GO, Coffman W, Kubicek MT (1952) Tissue culture studies of the proliferative capacity of cervical carcinoma and normal epithelium. *Cancer Research* 12: 264–265.
250. Gossen M, Bujard H (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences* 89: 5547–5551.
251. Puck TT, Cieciura SJ, Robinson A (1958) Genetics of somatic mammalian cells. *The Journal of Experimental Medicine* 108: 945. doi:10.1084/jem.108.6.945.
252. Sheets R (2000) History and Characterization of the Vero Cell Line. *The Vaccine and Related Biological Product Advisory*.
253. Melar M, Ott DE, Hope TJ (2007) Physiological levels of virion-associated human immunodeficiency virus type 1 envelope induce coreceptor-dependent calcium flux. *Journal of Virology* 81: 1773–1785. doi:10.1128/JVI.01316-06.
254. Vila-Coro A, Mellado M, Martin de Ana A, Lucas P, Del Real G, et al. (2000) HIV-1 infection through the CCR5 receptor is blocked by receptor dimerization. *Proceedings of the National Academy of Sciences* 97: 3388. doi:10.1073/pnas.050457797.
255. Ribardo DA, Lambert TJ, Kevin S, Mciver KS (2004) Role of *Streptococcus pyogenes* Two-Component Response Regulators in the Temporal Control of Mga and the Mga-Regulated Virulence Gene emm Role of *Streptococcus pyogenes* Two-Component Response Regulators in the Temporal Control of Mga and the Mga-Regulated. *Infection and Immunity* 72: 3668–3673. doi:10.1128/IAI.72.6.3668.
256. Ke SH, Madison EL (1997) Rapid and efficient site-directed mutagenesis by single-tube “megaprimer” PCR method. *Nucleic acids research* 25: 3371–3372.
257. Mattapallil JJ, Douek DC, Hill B, Nishimura Y, Martin M, et al. (2005) Massive infection and loss of memory CD4 T cells in multiple tissues during acute SIV infection. *Nature* 434: 1093–1097. doi:10.1038/nature03501.
258. Chomczynski P, Sacchi N (2006) The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nature Protocols* 1: 581–585. doi:10.1038/nprot.2006.83.
259. Harju S, Fedosyuk H, Peterson KRR (2004) Rapid isolation of yeast genomic DNA: Bust n’ Grab. *BMC Biotechnology* 4: 8. doi:10.1186/1472-6750-4-8.
260. Sikorski RS (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* 122: 19–27.
261. Dinman JD, Richter S, Plant EP, Taylor RC, Hammell AB, et al. (2002) The frameshift signal of HIV-1 involves a potential intramolecular triplex RNA structure.



Proceedings of the National Academy of Sciences 99: 5331.  
doi:10.1073/pnas.082102199.

262. Deng HK, Liu R, Ellmeier W, 1S, Unutmaz D, et al. (1996) Identification of a major co-receptor for primary isolates of HIV-1. *Nature* 381: 20.
263. Dragic T, Litwin V, Allaway GP, Martin SR, Huang Y, et al. (1996) HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature* 381: 667–673.
264. Schluns KS, Lefrançois L (2003) Cytokine control of memory T-cell development and survival. *Nature Reviews Immunology* 3: 269–279.
265. Holmes W, Lee J, Kuang W, Rice G, Wood W (1991) Structure and functional expression of a human interleukin-8 receptor. *Science* 253: 1278–1280.  
doi:10.1126/science.1840701.
266. Hartl D, Latzin P, Hordijk P, Marcos V, Rudolph C, et al. (2007) Cleavage of CXCR1 on neutrophils disables bacterial killing in cystic fibrosis lung disease. *Nature Medicine* 13: 1423–1430. doi:10.1038/nm1690.
267. Vasilescu a, Terashima Y, Enomoto M, Heath S, Poonpiriya V, et al. (2007) A haplotype of the human CXCR1 gene protective against rapid disease progression in HIV-1+ patients. *Proceedings of the National Academy of Sciences* 104: 3354–3359. doi:10.1073/pnas.0611670104.
268. Murphy P, Tiffany H (1991) Cloning of complementary DNA encoding a functional human interleukin-8 receptor. *Science* 253: 1280–1283.  
doi:10.1126/science.1891716.
269. Cacalano G, Lee J, Kikly K, Ryan A, Pitts-Meek S, et al. (1994) Neutrophil and B cell expansion in mice that lack the murine IL-8 receptor homolog. *Science* 265: 682–684. doi:10.1126/science.8036519.
270. Chen Q, Ghilardi N, Wang H, Baker T, Xie MH, et al. (2000) Development of Th1-type immune responses requires the type I cytokine receptor TCCR. *Nature* 407: 916–920. doi:10.1038/35038103.
271. Yoshida H, Hamano S, Senaldi G, Covey T, Faggioni R, et al. (2001) WSX-1 is required for the initiation of Th1 responses and resistance to L. major infection. *Immunity* 15: 569–578.
272. Hamano S, Himeno K, Miyazaki Y, Ishii K, Yamanaka A, et al. (2003) WSX-1 is required for resistance to *Trypanosoma cruzi* infection by regulation of proinflammatory cytokine production. *Immunity* 19: 657–667.
273. Batten M, Li J, Yi S, Kljavin NM, Danilenko DM, et al. (2006) Interleukin 27 limits autoimmune encephalomyelitis by suppressing the development of interleukin 17-producing T cells. *Nature Immunology* 7: 929–936. doi:10.1038/ni1375.

- 274. Dumoutier L, Van Roost E, Colau D, Renauld JC (2000) Human interleukin-10-related T cell-derived inducible factor: molecular cloning and functional characterization as an hepatocyte-stimulating factor. *Proceedings of the National Academy of Sciences* 97: 10144–10149. doi:10.1073/pnas.170291697.
- 275. Kotenko SV, Izotova LS, Mirochnitchenko OV, Esterova E, Dickensheets H, et al. (2001) Identification of the functional interleukin-22 (IL-22) receptor complex: the IL-10R2 chain (IL-10Rbeta ) is a common chain of both the IL-10 and IL-22 (IL-10-related T cell-derived inducible factor, IL-TIF) receptor complexes. *The Journal of biological chemistry* 276: 2725–2732. doi:10.1074/jbc.M007837200.
- 276. Wolk K, Witte E, Wallace E, Döcke W-D, Kunz S, et al. (2006) IL-22 regulates the expression of genes responsible for antimicrobial defense, cellular differentiation, and mobility in keratinocytes: a potential role in psoriasis. *European Journal of Immunology* 36: 1309–1323. doi:10.1002/eji.200535503.
- 277. Wolk K, Kunz S, Witte E, Friedrich M, Asadullah K, et al. (2004) IL-22 increases the innate immunity of tissues. *Immunity* 21: 241–254. doi:10.1016/j.immuni.2004.07.007.
- 278. Cherry JM (1998) *Saccharomyces cerevisiae* Gene Nomenclature Conventions: <http://www.yeastgenome.org/help/community/nomenclature-conventions>.
- 279. Wain HM, Bruford E a, Lovering RC, Lush MJ, Wright MW, et al. (2002) Guidelines for human gene nomenclature. *Genomics* 79: 464–470. doi:10.1006/geno.2002.6748.