Massively parallel identification of sequence motifs triggering ribosome-associated mRNA quality control

Katharine Chen

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Arvind Subramaniam, Chair

Toshio Tsukiyama

Chris Lapointe

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

Massively parallel identification of sequence motifs triggering ribosome-associated mRNA quality control

Katharine Chen

Chair of the Supervisory Committee:

Arvind Subramaniam

Department of Genome Sciences

Decay of mRNAs can be triggered by ribosome slowdown at stretches of rare codons or positively charged amino acids. However, the full diversity of sequences that trigger co-translational mRNA decay is poorly understood. To comprehensively identify sequence motifs that trigger mRNA decay, we use a massively parallel reporter assay to measure the effect of all possible combinations of codon pairs on mRNA levels in *S. cerevisiae*. In addition to known mRNA-destabilizing sequences, we identify several dipeptide repeats whose translation reduces mRNA levels. These include combinations of positively charged and bulky residues, as well as proline-glycine and proline-aspartate dipeptide repeats. Genetic deletion of the ribosome collision sensor Hel2 rescues the mRNA effects of these motifs, suggesting that they trigger ribosome slowdown and activate the ribosome-associated quality control (RQC) pathway. Deep mutational scanning of an mRNA-destabilizing dipeptide repeat reveals a complex interplay between the charge, bulkiness, and location of amino acid residues in conferring mRNA instability. Finally, we show that the mRNA effects of codon pairs are predictive of the effects of endogenous sequences. Our work highlights the complexity of sequence motifs driving co-translational mRNA decay in eukaryotes, and presents a high throughput approach to dissect their requirements at the codon level.

Contents

Acknowledgements

For Michael, Tenny, Momo, and Appa.

Chapter 1. Introduction

I. Eukaryotic gene expression - the central dogma of biology

Proteins, comprised of amino acid chains, are macromolecules that carry out the majority of cellular functions. The accurate and timely synthesis of proteins is vital for cell survival and proper cellular function. The flow of genetic information within a biological system starts with DNA, which is transcribed to messenger RNA (mRNA) and subsequently translated into protein[1]. Every step of this process, known as the central dogma of biology, is highly regulated to ensure proper protein output and thus proper cellular function. Genomic DNA is wrapped compactly in the nucleus around histone proteins, and the unwrapping or wrapping of DNA to form an open or closed conformation is one of the first layers of gene regulation[2]. At open regions of DNA, RNA polymerase II, with the assistance of transcription factors, initiates transcription by binding to a promoter region upstream of the gene[3]. RNA polymerase II then synthesizes the pre-mRNA transcript that is complementary to the genomic DNA sequence. This pre-mRNA transcript then undergoes multiple processing steps before it can be exported from the nucleus. The 5' end is linked to an N7-methylated guanosine via a triphosphate linkage with the first nucleotide, followed by methylation of the 2'O ribose of the first nucleotide to form the 5' 7-methylguanosine cap[4,5]. The spliceosome, comprised of small nuclear RNAs and associated proteins, removes non-coding introns and ligates coding exons together[6]. Finally, polyadenylation adds a string of adenosine residues to the 3' end, which contributes to mRNA stability and facilitates nuclear export to the cytoplasm where the final stage of the central dogma, translation, occurs[7]. The process of translating the mRNA message into protein is regulated by a diverse set of factors that mediate every step of this process, from initiation to elongation to termination. In addition to factors that regulate the stages of translation, various factors also mediate the stability, or longevity, of the mRNA message itself. All of these factors together contribute to how much protein is expressed off of a given mRNA transcript. This dissertation will describe the

regulation of protein expression in eukaryotic cells with particular focus on pathways that mediate mRNA degradation and factors that affect mRNA stability.

II. mRNA translation

Translation initiation

Ribosomes synthesize protein from mRNA in three key steps: initiation, elongation, and termination. The rate-limiting step of this process is translation initiation, which is strongly associated with the amount of protein synthesized from a transcript[8,9]. Eukaryotes exhibit two primary modes of translation initiation, the canonical m7G-cap-dependent mechanism and the non-canonical cap-independent mechanism. Before engaging with mRNA, the heterotrimeric eIF2-GTP complex recruits an initiator methionyl tRNA (Met-tRNA$_i$) and forms the eIF2 ternary complex, which is incorporated into the 40S ribosome subunit with eIF1/eIF1A/eIF3/eIF5 to form the 43S pre-initiation complex (PIC)[10]. Canonical translation initiation begins with recognition of the 5' m7G cap by the cap binding protein eIF4E, part of the eIF4F complex in association with eIF4G and eIF4A, in a process known as mRNA activation[10–13]. The interaction between eIF4G in the eIF4F complex and eIF3 in the 43S PIC allows recruitment of the 43S to the m7G cap[10]. The 43S PIC then scans the 5' untranslated region (5' UTR) of the mRNA via the ATP-dependent RNA helicase activity of eIF4A and interactions with eIF4G, eIF4B and eIF3[10,14,15]. Scanning factors, eIF1A and eIF1, maintain the 43S PIC in an open conformation during scanning in order to destabilize the interaction of non-cognate codons with the Met-tRNA$_i$ anticodon[10]. Various features of the 5' UTR, through which the 43S PIC scans, affects initiation efficiency and thus protein output. The presence of near cognate codons that differ from the AUG start codon by one base, upstream start codons resulting in upstream open reading frames (uORFs), strong secondary structure, or the length of the 5' UTR all impact the efficiency of translation initiation[16]. When the scanning 43S PIC encounters a stable

enough codon-anticodon base-pairing interaction, eIF2-bound GTP is hydrolyzed into GDP. As GTP hydrolysis occurs, eIF2 loses its affinity for the Met-tRNA$_i$ and is released from the 43S PIC to produce the stable 48S PIC, to which the large (60S) ribosome subunit joins to produce the 80S initiation complex that is able to begin protein synthesis[10].

Elongation and the protein coding sequence

Translation elongation begins with the 80S ribosome positioned at an AUG start codon with a methionyl-tRNA$_i$ in the P site. Elongation is comprised of three basic steps: tRNA selection or decoding, peptide-bond formation, and translocation. During tRNA selection, the aminoacyl-tRNA with an anticodon complementary to the mRNA codon is loaded into the A site of the ribosome. The GTPase eEF1A brings the aminoacyl-tRNAs to the ribosomal A site in a ternary complex with GTP[17]. When a cognate interaction between the codon and anticodon is sensed, eEF1A hydrolyses GTP to position the tRNA fully in the A site[18]. During peptide-bond formation, the amino group of the incoming amino acid attacks the ester linkage on the peptidyl-tRNA in the P site, and the elongating peptide is transferred to the tRNA in the A site[19,20]. As this peptide bond forms, the ribosome subunits rotate and the tRNAs adopt an altered conformation known as the 'hybrid' state[21]. In the hybrid state, the anticodon ends of the tRNAs remain essentially in the P and A sites of the small subunit, while the acceptor ends are positioned in the E site and P sites of the large subunit. This rotated hybrid state is the substrate eEF2, another GTPase, which translocates the mRNA–tRNA complex such that the uncharged tRNA now sits in the E site, while the tRNA with the peptide bond sits in the P site[22]. In yeast, an additional elongation factor, eEF3, is essential for elongation presumably by promoting tRNA release from the E site after translocation[23]. This cycle continues until the entire open reading frame (ORF) is translated and a stop codon is encountered.

Translation termination

Translation termination occurs when the ribosome encounters a stop codon in the A site of the ribosome. The eukaryotic release factor eRF1, whose shape and size bares strong resemblance to that of a tRNA, contains structural motifs that recognize the three stop codons, UAA, UAG and UGA, and binds to the ribosome A site[24,25]. eRF1 also has a Gly–Gly–Gln (GGQ) motif that extends into the peptidyl-transferase center to promote the release of the nascent peptide[26]. Similar to tRNA, eRF1 requires a specialized GTPase, eRF3, for proper function. eRF1, eRF3, and GTP form a ternary complex that engages the ribosome, which triggers GTP hydrolysis and releases eRF3 from the ribosome. Rli1 (ABCE1 in humans) then interacts with eRF1 to stimulate the hydrolysis of the nascent peptide from the peptidyl-transferase center by stabilizing the active eRF1 conformation[27,28]. The release of the peptide marks the completion of translation termination. Following translation termination, the 80S complex must be recycled into 40S and 60S subunits. The essential protein Rli1 uses ATP binding and hydrolysis in collaboration with eRF1 to dissociate the post-termination ribosome into the 40S and 60S subunits[27]. The split ribosomal subunits can then be bound by initiation factors to initiate subsequent rounds of translation. Following the stop codon on the mRNA molecule is the 3' untranslated region (3' UTR), which has been shown to affect mRNA expression in various ways[29]. The 3'UTR contains microRNA binding sites, which suppress expression and induce mRNA degradation[30]. RNA binding proteins (RBPs) have also been found to bind to 3'UTRs to affect protein abundance and half life[31]. Furthermore, AU-rich elements (AREs) in the 3'UTR direct poly(A) tail deadenylation, which leads to mRNA decay[29,32].

III. mRNA stability and co-translational quality control

Canonical mRNA turnover

mRNA degradation can be triggered by structurally compromising either of the mRNA's two main stability elements: the 3' poly(A) tail through deadenylation or the 5' 7-methylguanosine cap via decapping. Endonucleolytic cleavage within the mRNA can also initiate mRNA decay. The majority of eukaryotic mRNA decay is initiated by deadenylation via the poly(A) nuclease known as the Ccr4-Not complex[33]. Following deadenylation, the Dcp1-Dcp2 complex decaps the mRNA, which enables 5' to 3' exonucleolytic digestion by Xrn1[34]. In some instances, deadenylation can also trigger degradation of the mRNA body from the 3'end by the exosome[35]. While deadenylation is the trigger for the bulk of standard mRNA decay, various sequence features in the transcript also signal the need for mRNA degradation. A diverse set of co-translational pathways including nonsense-mediated decay (NMD), no-go decay (NGD), ribosome-associated quality control (RQC), and codon optimality-mediated decay (COMD) mediate the decay of mRNAs harboring these sequence features.

Nonsense-mediated mRNA decay

Nonsense-mediated mRNA decay (NMD) is a translation-dependent mRNA surveillance mechanism designed to degrade mRNAs harboring premature termination codons (PTCs). Two mechanisms have been described for NMD, the exon junction complex (EJC)-dependent mechanism and the EJC-independent mechanism, though the EJC-dependent mechanism is the most efficient form of NMD[36]. The EJC is a large protein complex deposited upstream of most spliced exon-exon junctions during pre-mRNA splicing. During translation, the processive translocation of 40S or 80S ribosomes through the 5' UTR or coding regions removes any proteins associating with the mRNA until the stop codon is reached and translation terminates. Thus, NMD is believed

to occur when a stop codon is reached more than 50-55 nucleotides upstream from an exon-exon junction or EJC[37], as this would represent an unusual occurrence, and is initiated after the first round of translation terminates. NMD is mediated by the RNA-dependent helicase and ATPase UPF1 and a phosphatidylinositol 3-kinase-related kinase SMG1, which together interact with the eRF1-eRF3 termination complex in the ribosome to form the SURF complex[38]. The SURF complex promotes phosphorylation of UPF1 to promote mRNA decay[38].

Non-Stop Decay

While nonsense-mediated mRNA decay is triggered by a translating ribosome terminating upstream of an EJC, non-stop decay (NSD) is characterized by a translating ribosome encountering a poly(A) tail in the 3'UTR in the event that the stop codon is missing[39,40]. NSD mRNA substrates are degraded primarily by the SKI complex with the exosome[41], and appear not to be degraded by Xrn1 like mRNAs harboring other ORF-internal aberrant sequences are[42,43]. Ribosomes that translate into poly(A) sequences experience sliding on iterated AAA codons that may result in frameshifting leading to NMD if a premature stop codon is encountered[44,45]. Structural studies subsequently found that the sliding was due to poly(A) sequences assuming a helical conformation within the mRNA channel that make decoding difficult[46,47]. Whether the key signaling event that triggers mRNA decay is the stalled ribosome, colliding ribosomes, or the particular structure of the poly(A) sequence within the mRNA channel is still to be determined.

No-Go Decay and ribosome-associated quality control

Another subset of mRNA decay pathways are triggered when translating ribosomes stall sufficiently within an ORF to cause trailing ribosomes to collide with it. Early studies found that stable secondary structure, damaged mRNA, truncations, and rare codons cause strong ribosome collisions that result in degradation of the aberrant mRNA[48–50], and the pathway was termed No-go Decay

(NGD). These ribosome collisions are detected by the E3 ubiquitin ligase, Hel2 (ZNF598 in humans), which ubiquitinates collided ribosomes and stabilizes the signal[51]. This triggers endonucleolytic cleavage of the mRNA by Cue2 and subsequent mRNA degradation by Xrn1[42,48–50]. Other mRNA sequences are now known to cause ribosome collisions that trigger Hel2-driven mRNA decay, including (non-poly-A) polybasic sequences and polytryptophan sequences[52,53]. Hel2 ubiquitination and Cue2 endonucleolytic cleavage is followed by recruitment of either the ribosome QC trigger complex (RQT) or, less frequently, Dom34/Hbs1 release factors to rescue stalled ribosomes and facilitate subsequent degradation of the stalled nascent peptide[43,52]. The process of mRNA and nascent peptide decay initiated by ribosome collisions and driven by Hel2 is often termed ribosome-associated quality control (RQC).

Codon optimality mediated mRNA decay

The final mRNA quality control pathway that is initated by the translating ribosome is codon optimality-mediated mRNA decay (COMD). While RQC is triggered by colliding ribosomes, COMD is triggered by ribosome slowdown. Codon optimality is broadly defined by the abundance of a codon's cognate tRNA and the strength of tRNA anticodon decoding, termed the tRNA adaptive index (tAI)[54]. Genome-wide studies of codon optimality found that mRNA half life correlates with optimality[55]. The protein Not5, part of the CCR4-NOT complex involved in canonical mRNA turnover, was recently found to detect slow-decoding ribosomes by binding directly to the tRNA exit site in order to trigger mRNA decay[56].

Sequences that cause ribosome stalling in *S. cerevisiae*

Elongating ribosomes translating through the CDS can stall due to several aberrations. Nonoptimal codons, or codons for which the cognate tRNA is in low cellular abundance, cause the ribosome to slow down and the ribosome A site to remain empty for longer than during translation of more

optimal codons[55,56]. The empty A site is sensed by Not5 (CNOT3 in humans), part of the Ccr4-Not deadenylase complex, which triggers deadenylation-dependent mRNA decay through Xrn1. Rare codons (CGA and CGG) are also considered nonoptimal in that their cognate tRNAs are in extremely low abundance, but cause very potent stalling because of the requirement for I:A wobble base pairing during decoding[50]. Because of this, rare codons trigger ribosome stalling and collision rather than just ribosome slowdown, and are regulated by Hel2-driven RQC[57]. In addition to rare codons, polybasic sequences (polylysine and polyarginine) stall ribosomes due to electrostatic interactions with the negatively charged ribosome exit tunnel in yeast[58–60]. Interestingly, polybasic sequences as a whole are not potent stalling sequences in human cells[61]. In human cells, the only polybasic sequence that potently stalls ribosomes is the polyA sequence, which was found to stall ribosomes due to ribosome sliding and inhibitory structures in the mRNA channel of the ribosome[46,47]. Aside from polybasic sequences, tryptophan sequences have also been found to stall ribosomes and trigger RQC. Greater than eight consecutive repeats of tryptophan triggers ribosome stalling and Hel2-driven RQC due to steric hindrance between polytryptophan residues and the discrimination gate of the ribosome exit tunnel[53]. Ribosomes can also stall due to structural abnormalities. mRNA truncation, which can occur due to chemical or enzymatic damaging or from abortively spliced mRNAs, are also known to potently stall translating ribosomes. Ribosomes that are stalled in these instances are mediated by Dom34/Hbs1 (PELOTA/HBS1 in humans), which recognize the empty A site of the ribosome at the 3' end of the truncated or damaged mRNA. Thus far, only two endogenous targets of RQC have every been identified. One endogenous substrate, Hac1, was found via ribosome profiling in the presence and absence of Dom34 in yeast[62]. Incorrect splicing ligation of the Hac1 mRNA results in truncation within the coding region, which is resolved by Dom34/Hbs1. The other known endogenous substrate is located in a gene of unknown function named Sdd1 and was found via a dual-luciferase reporter screen using endogenous sequences

containing CGA-CGA dicodons in yeast[63]. Cryo-EM of the Sdd1 stalling sequence ($SDD1_{196\text{-}212}$: FFYEDYLIFDCRAKRRK) revealed that there were a number of requisite interactions between the nascent peptide and the ribosome exit tunnel. First, there is a requirement for positive charge at positions 207 and 209, which are thought to perturb the peptidyl-transferase center of the ribosome, and a requirement for aromatic residues at position 201, which are thought to interact with the uL4/uL22 constriction point of the ribosome. Additionally, stalling appears to require the aspartate at position 200.

IV. Thesis Objectives

As more studies contribute to our understanding of mRNA quality control, it has become increasingly clear that the pathways described in the subsections above are often overlapping and can compensate for each other. While much work has been done in identifying the regulatory factors involved in these pathways, the mRNA sequences that trigger the various forms of quality control has not been extensively characterized. This dissertation aims to elucidate the full diversity of sequences that cause co-translational mRNA degradation in *S. cerevisiae* coding sequences, as well as the regulatory pathways that mediate this decay and the mechanisms for their instability. To this end, we use a variety of massively parallel reporter assays to experimentally test thousands of sequence motifs for their effect on mRNA stability in *S. cerevisiae*.

Using these approaches, our work, described in this dissertation, reports on the following key discoveries:

1. We identify a large set of dipeptide sequence motifs that trigger co-translational RNA decay.
2. We find that several of these motifs trigger mRNA decay through the ribosome-associated mRNA quality control (RQC) pathway.
3. We use a deep mutational scanning approach to identify the amino acid characteristics that

confer mRNA instability and RQC dependence.

4. Measurements using our massively parallel assay accurately predict the effect of endogenous sequences on mRNA stability.

The following chapters are in review at *Nucleic Acids Research*. Subsections named “Future directions” describe ongoing revision experiments.

Chapter 2: Massively parallel identification of sequence motifs triggering mRNA decay

Introduction

Translation and decay of mRNA are fundamental stages of gene expression whose interplay is crucial for determining steady-state protein levels in the cell. The protein coding region of mRNA has been recently recognized as an important determinant of mRNA stability[46,48,50,55,57,61,64–67]. Ribosome elongation rates can vary along the protein coding region, which is sensed by diverse regulatory factors to trigger mRNA decay[39,40,44,46–49,58,61,67–71]. Dysregulation of mRNA decay pathways has been linked to neurological diseases, autoinflammatory diseases, and cancer[72–77].

Several motifs in the protein coding region of eukaryotic mRNAs have been associated with changes in mRNA stability[39,40,48,49,53,55,57,66,67,78]. Nonoptimal codons decrease ribosome elongation rates and trigger Not5-dependent mRNA deadenylation and decay[55,56,79,80]. Strong ribosome stalls caused by polybasic residues, poly-tryptophan sequences, and rare codon repeats trigger ribosome collisions and Hel2-dependent ribosome-associated mRNA quality control (henceforth RQC)[42,53,57,61,67,70,71,81]. Poly-proline sequences decrease ribosome elongation rate, but such slowdowns are thought to be resolved by eIF5A and not trigger mRNA quality control[82,83]. Ribosome profiling studies have identified several dipeptide and tripeptide motifs that are enriched at sites of ribosome stalls and collisions[84–87]. However, whether such motifs are sufficient to trigger mRNA quality control is not known. Ribosome stalling motifs in endogenous protein coding sequences often depend on a complex combination of amino acid residues in the nascent peptide[63,88–91], and thus their relation to the simple repeat stalling sequences studied in reporter assays is not clear.

We recently developed a massively parallel reporter assay to identify coding sequence motifs triggering mRNA decay in human cells[78]. Using this assay, we found that translation of a diverse

set of dipeptide repeats composed of bulky and positively charged amino acids are sufficient to trigger mRNA decay in human cells. Nevertheless, the molecular mechanism by which translation of these dipeptide repeats triggers mRNA decay in human cells remains unknown. Further, the extent to which translation of bulky and positively charged residues serves as an evolutionarily conserved signal for mRNA decay in other eukaryotes is unclear. Since co-translational mRNA decay pathways have been extensively studied in the budding yeast *S. cerevisiae*[43,61,92–94], we sought to use this as an experimental model to dissect the molecular mechanism and sequence requirements of coding sequence-dependent mRNA decay. By extending our massively parallel reporter assay from human cells to *S. cerevisiae*, we identify several mRNA-destabilizing dipeptide motifs including combinations of bulky and positively charged residues, as well as proline-glycine and proline-aspartic acid dipeptide repeats. We define Hel2-dependent RQC as the major pathway regulating mRNA decay triggered by translation of these dipeptide repeats. Using deep mutational scanning, we further characterize the biochemical requirements at the codon level for bulky and positively charged dipeptide repeats in triggering Hel2-dependent mRNA decay. Together, our results highlight the diversity of coding sequence motifs triggering co-translational mRNA decay in *S.cerevisiae*, define the biochemical requirements for their mRNA-destabilizing effects, and reveal the extent of evolutionary conservation of these motifs across eukaryotes.

A massively parallel reporter assay for mRNA effects in *S. cerevisiae*

To study the effect of coding sequence motifs on mRNA levels in *S. cerevisiae* in an unbiased manner, we modified a pooled reporter assay that we previously developed in mammalian cells[78] (Fig. 1A). In our design for *S. cerevisiae*, a tandem 8× repeat of all possible codon pairs (4096 pairs in total) is inserted between the *PGK1* and *YFP* coding sequences. The 8× repetition amplifies the effect of each codon pair on mRNA levels. Each codon pair repeat is followed by a 24 nucleotide random barcode without stop codons, which enables their accurate quantification without

sequence-dependent biases. Barcode sequences linked to each codon pair insert are identified by sequencing the plasmid library. We integrated the plasmid library into a noncoding region of chromosome I of *S. cerevisiae*, extracted mRNA and genomic DNA, and counted barcodes by high throughput amplicon sequencing. Barcode counts in the cDNA normalized by corresponding counts in the genomic DNA provide a relative measure of the steady-state mRNA level of each codon pair insert in our library. We further normalized mRNA levels by the median value across all inserts in the library to account for different sequencing depths and to facilitate comparison across experiments.

We recovered a median of 20 barcodes linked to each codon pair insert in the cDNA and genomic DNA libraries out of the 100 barcodes per insert in the plasmid library (Fig. S1A). We identified barcodes linked to 97% of all codon pairs in the plasmid library and 91% in the cDNA and genomic DNA libraries (Fig. 1B), indicating our assay's ability to capture most of the codon pair motifs. Missing codon pairs in the plasmid library have a high GC content (Fig. S1B), suggesting that they are either resistant to cloning or toxic for *E. coli* growth. Many of the remaining missing codon pairs in the cDNA and genomic DNA from *S. cerevisiae* encode hydrophobic amino acids (Fig. S1C). Constitutive expression of such dipeptide repeats might be toxic due to their aggregation or membrane insertion.

To test whether our massively parallel assay recapitulates known codon and amino acid effects, we examined the average mRNA levels of individual codons and amino acids (Fig. 1C,E). To this end, we calculated the normalized ratio of barcode counts between cDNA and genomic DNA across all codon pairs containing each of the 64 codons or 20 amino acids. We observed a tight overlap of average mRNA levels of each codon or amino acid between positions 1 and 2 of the codon pair (Fig. 1C,E). This observation is consistent with the 8× repetitive nature of our codon pair library, due to which each codon pair insert is similar to its codon-reversed counterpart except for circular

permutation of a single codon.

Within several synonymous codon families, codons with lowest mRNA levels in our assay (Fig. 1C) correspond to the less frequent codons within that family in the *S. cerevisiae* transcriptome[54,95,96]. These include CGA, CGG, and AGG (Arg), ATA (Ile), and CCG (Pro) (Fig. 1C), all of which are known to reduce protein expression or trigger mRNA decay in *S. cerevisiae*[50,55,57,71,97,98]. In line with these observations, average mRNA levels of codons in our assay positively correlated with codon stability coefficients (CSCs) inferred from stability measurements on endogenous mRNAs in *S. cerevisiae*[55,66] (Fig. 1D, r=0.50, p<1e-4). This correlation with CSC is notable given that we vary only a 16 codon region within a 700 codon *PGK1-YFP* coding sequence in our assay.

At the amino acid level, arginine, lysine, and tryptophan had the lowest mRNA levels on average (Fig. 1E), consistent with the known role for these amino acids in triggering ribosome-associated quality control[40,53,57,59–61,71,93,99,100]. mRNA effects of these amino acids are comparable to that of stop codons, which trigger nonsense-mediated mRNA decay (NMD). In contrast to the codon effects, average mRNA levels of amino acids in our assay do not show significant correlation with amino acid stability coefficients (AASCs) inferred from stability measurements on endogenous mRNAs in *S. cerevisiae*[66] (Fig. 1F). This lack of correlation is in line with the limited role of amino acid identity in determining global mRNA stability in *S. cerevisiae*[55,66].

Overall, the average mRNA effects of codons and amino acids in our massively parallel reporter assay corroborate previously known stalling sequences in *S. cerevisiae* and show expected correlation with mRNA stability metrics inferred from endogenous mRNAs.

Figure 1

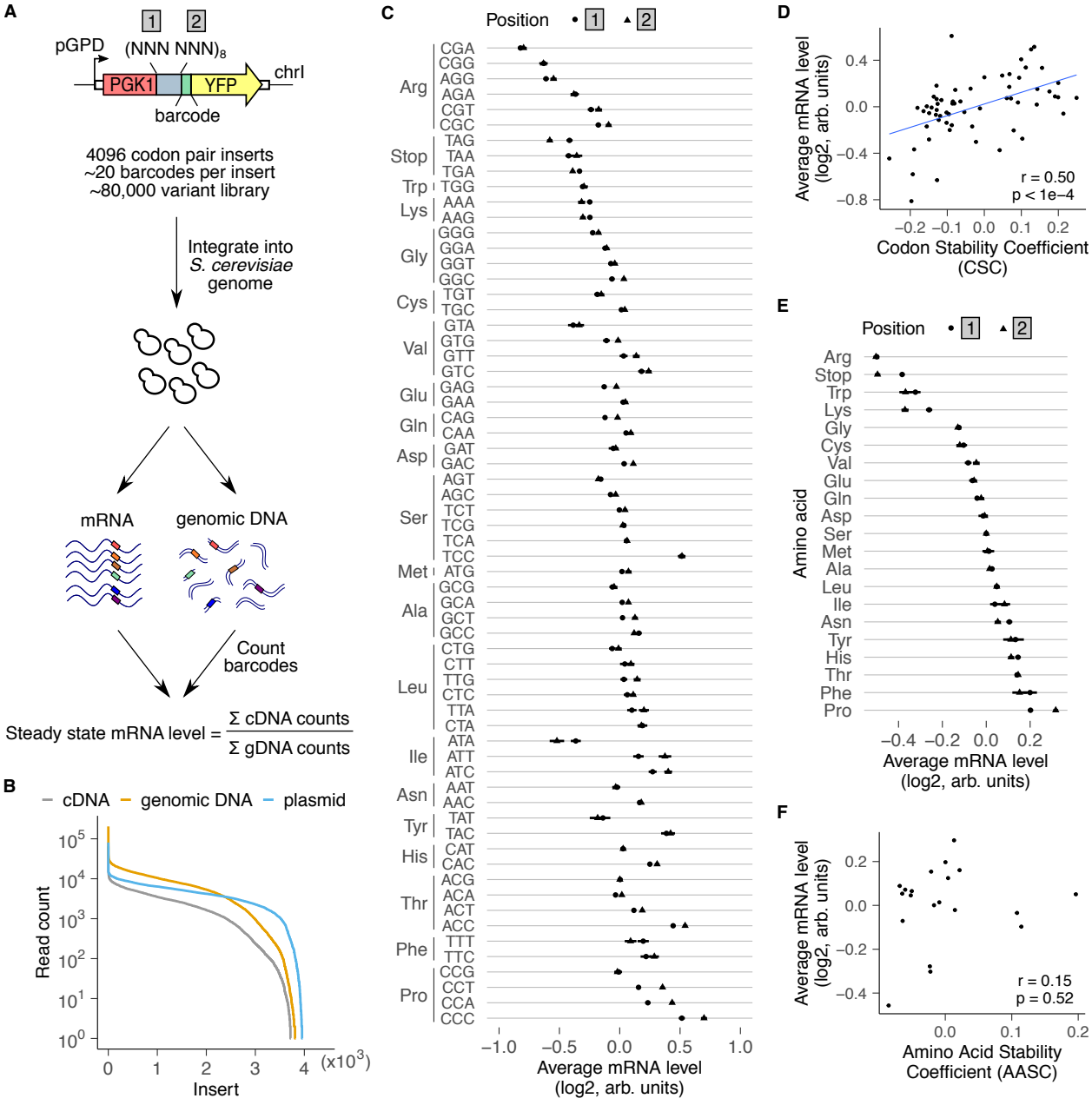

Fig.1: A massively parallel reporter assay for mRNA effects in *S. cerevisiae*. (A) Assay design. Each element in the library includes one of 4096 possible combinations of codon pairs repeated eight times. Each repeat is inserted in-frame between *PGK1* and *YFP*, and is followed by a random 24 nt barcode without in-frame stop codons (median of 20 barcodes/insert). The 80,000 variant library is integrated as a pool into a noncoding region of chromosome I. The barcodes in

cDNA and genomic DNA are counted by high throughput amplicon sequencing. Relative steady state mRNA effect of each insert is calculated by first normalizing cDNA counts by genomic DNA counts for all barcodes linked to that insert and then by median-normalizing across all codon pairs. **(B)** Distribution of reads per codon pair insert for cDNA, genomic DNA, and plasmid libraries. **(C)** Average mRNA level of reporters with indicated codons in position 1 (circles) or position 2 (triangles) of the codon pair. **(D)** Average mRNA effects of individual codons compared against codon stability coefficients derived from endogenous *S. cerevisiae* mRNAs[55]. **(E)** Average mRNA level of reporters encoding the indicated amino acid in position 1 (circles) or position 2 (triangles) of the codon pair. Error bars in C and E represent standard deviation over all variants containing the codon or amino acid at each position. Average mRNA levels in C and E are median-normalized over all codons or amino acids at each position. **(F)** Same as D, except for amino acids compared against amino acid stability coefficients[66].

Identification of codon pair repeats that reduce mRNA levels

Inclusion of all possible codon pair repeats in our library allowed us to next study the effect of pairwise codon and amino acid combinations on mRNA levels (Fig. 2A,B). We found a strong correlation (r=0.92, p<1e-10) between mRNA effects of codon pairs and their reverse counterparts, indicating the robustness of our measurements (Fig. S1D). We identified several families of synonymous codon pairs that consistently reduced mRNA levels relative to the remaining inserts in the library (black outlines, Fig. 2A,B). Among the most destabilizing codon families were those encoding lysine, arginine, and tryptophan repeats, in agreement with the average destabilizing effect of these amino acids (Fig. 1E).

Our assay revealed several dipeptide repeats that have not been previously associated with ribosome stalling or ribosome-associated quality control in *S. cerevisiae* (Fig. 2A,B). These include several combinations of bulky and positively charged amino acids such as phenylalanine-lysine (FK/KF), tryptophan-arginine (WR/RW), and tyrosine-lysine (YK/KY). Some combinations of hydrophobic and positively charged amino acids such as arginine-leucine (LR/RL) and arginine-isoleucine (IR/RI) were also destabilizing. Notably, we found similar mRNA-destabilizing combinations of positively charged amino acids with bulky and hydrophobic amino acids in human cells[78], indicating that these sequences may be broadly destabilizing across eukaryotes. We confirmed the requirement of bulkiness for reducing mRNA levels in a targeted experiment by replacing phenylalanine with the smaller non-polar glycine in combination with lysine (Fig. S2A). Using flow cytometry, we found FK dipeptide repeats reduced YFP reporter levels similar to the known RQC-inducing KK repeat (Fig. 2C,D). Moreover, protein levels of a control RFP reporter expressed from a different chromosomal locus was unaffected by FK repeat expression, indicating that it does not perturb global gene expression (Fig. 2C).

Proline-glycine (PG/GP) and proline-aspartic acid (PD/DP) repeats were also among the mRNA-

destabilizing codon pairs in our assay (black outlines, Fig. 2A,B). Unlike combinations of bulky and positively charged amino acids, these repeats did not reduce mRNA levels in human cells[78]. Conversely, amino acid combinations such as arginine-histidine and serine-phenylalanine that destabilize mRNAs in human cells[78] did not reduce mRNA levels in our assay in *S. cerevisiae*. Finally, dipeptides comprised of bulky and positively charged amino acids as well as proline-glycine and proline-aspartic acid dipeptides are enriched at sites of ribosome collisions in *S. cerevisiae* and mammalian cells[84–86]. This observation suggests that the mRNA-destabilizing effects of such dipeptide repeats in our assay arises from ribosome slowdown when these peptide motifs are synthesized during mRNA translation.

Figure 2

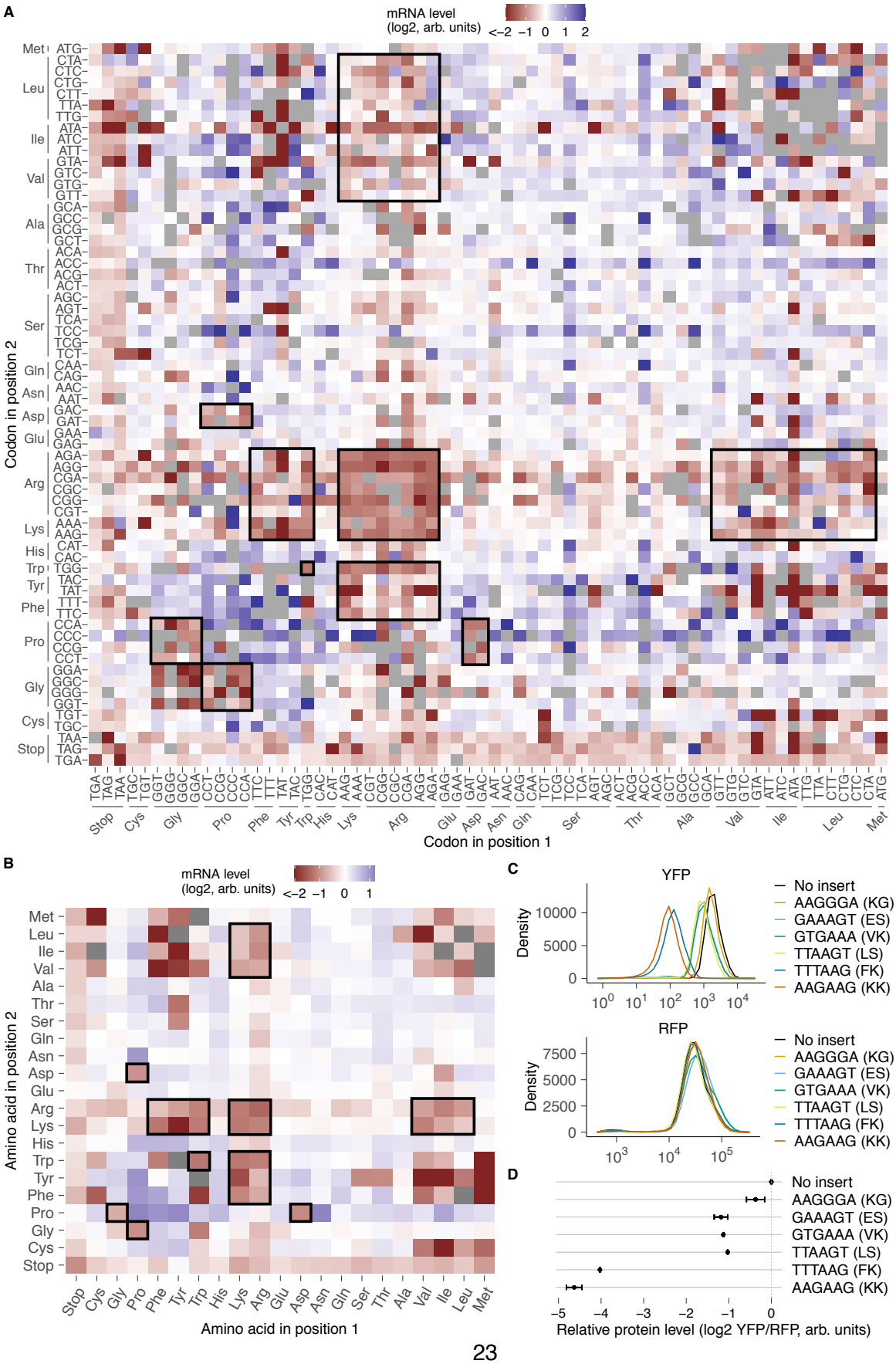
A
mRNA level
(log2, arb. units)
<–2 –1 0 1 2
Codon in position 2
Codon in position 1
B
mRNA level
(log2, arb. units)
<–2 –1 0 1
Amino acid in position 2
Amino acid in position 1
C
YFP
RFP
Density
No insert
AAGGGA (KG)
GAAAGT (ES)
GTGAAA (VK)
TTAAGT (LS)
TTTAAG (FK)
AAGAAG (KK)
D
Relative protein level (log2 YFP/RFP, arb. units)

Fig.2: Identification of codon pairs and dipeptides that reduce mRNA levels. (A) mRNA level of inserts encoding each codon pair repeat. Codons at the first or second position of each pair are shown along the horizontal or vertical axes, respectively. Missing codon pairs are in grey. Synonymous codon pair families with lower mRNA levels are outlined in black. **(B)** mRNA level of inserts encoding each dipeptide repeat. Amino acids at the first or second position of each dipeptide are shown along the horizontal or vertical axes, respectively. Missing dipeptides are in grey. Dipeptide groups with lower mRNA levels are outlined in black. **(C)** Protein expression from individual *PGK1-YFP* reporters measured by flow cytometry (Top). A control RFP reporter integrated at a different locus was also quantified (Bottom). **(D)** Quantification of median YFP signal in **C** relative to the constitutively expressed RFP reporter. Error bars represent standard error of the mean across 5 biological replicates. GAAAGT (ES) is a frameshift control for GTGAAA (VK), and TTAAGT (LS) is a frameshift control for TTTAAG (FK).

Dipeptide-induced mRNA destabilization requires translation

We used three different approaches to assay whether translation of dipeptide repeats is necessary for their mRNA-destabilizing effects.

First, we computationally tested whether the presence of codon pairs in the correct *PGK1-YFP* reading frame is necessary for the mRNA effects of the corresponding dipeptide repeats (Fig. 3A). mRNA effects of dipeptide repeats encoded in the correct +0 frame showed much lower correlation with the mRNA effects in the wrong +1 and +2 frames than with the correct +3 frame. We note that the +3 frameshift is essentially the same frame as the in-frame codon pair but with the codon positions interchanged. Thus, the simple presence of nucleotide sequences coding for destabilizing dipeptide repeats in the mRNA is not sufficient to reduce mRNA levels; they need to be present in the correct translated frame. Consistent with this observation, we found low correlation between mRNA levels of codon pair inserts and basic measures of nucleotide diversity such as GC content or GC3 content (Supplementary Fig. 2B).

Second, we tested whether global inhibition of translation is sufficient to rescue the mRNA-destabilizing effects of dipeptide repeats. Glucose deprivation is known to rapidly inhibit translation initiation in yeast[101,102]. Therefore, we grew *S. cerevisiae* cells containing the original codon pair library (Fig. 1A) in media without glucose for one hour, and quantified relative mRNA levels of inserts by high throughput sequencing as before. At the codon level, glucose deprivation increased the relative mRNA levels of inserts containing arginine and lysine codons, consistent with their mRNA effects arising at the translational level (Fig. 3B). Glucose deprivation also increased the relative mRNA levels of several dipeptide-encoding inserts that were destabilizing under normal growth (Fig. 3C). These include the known RQC-inducing polybasic sequences RR, RK, KR, and KK, as well as the novel destabilizing dipeptide repeats such as KW, FK, RW, PD, and PG that we identified in our original screen. Intriguingly, stop codon-containing inserts had

lower mRNA levels during glucose deprivation even though nonsense-mediated mRNA decay of these inserts requires translation. This might be because NMD occurs in processing bodies (P-bodies), whose formation is enhanced upon glucose deprivation[103–105].

Third, we tested whether experimentally altering the translated reading frame of codon pair inserts is sufficient to abrogate their mRNA-destabilizing effects, which would rule out transcription or RNA processing as possible mechanisms. Therefore, we inserted 2 base pairs upstream of the codon pair insert, leaving all other aspects of the reporter identical to the original library, and assayed for mRNA effects as before (Fig. 3D). The 2 base pair insertion shifts all codon pair inserts to the -1 frame, but does not introduce a stop codon upstream of the codon pair inserts. At the aggregate level, the -1 frameshifted library loses the previously observed correlation with codon stability coefficients (Fig. 3E, compare against Fig. 1D), consistent with the codon effects predominantly arising from translation. Similarly, most dipeptide repeats that destabilize mRNAs in the original library had higher relative mRNA levels in the -1 frameshifted library (Fig. 3F). Note that the WW dipeptide-coding repeat did not pass our read cutoff filter in both the glucose deprivation and the -1 frameshifting experiment (Fig. 3C,F).

In summary, our computational and experimental frameshifting assays, along with our glucose depletion experiment, establish the translation dependence of the mRNA effects of the destabilizing dipeptide repeats identified in our original screen.

Figure 3

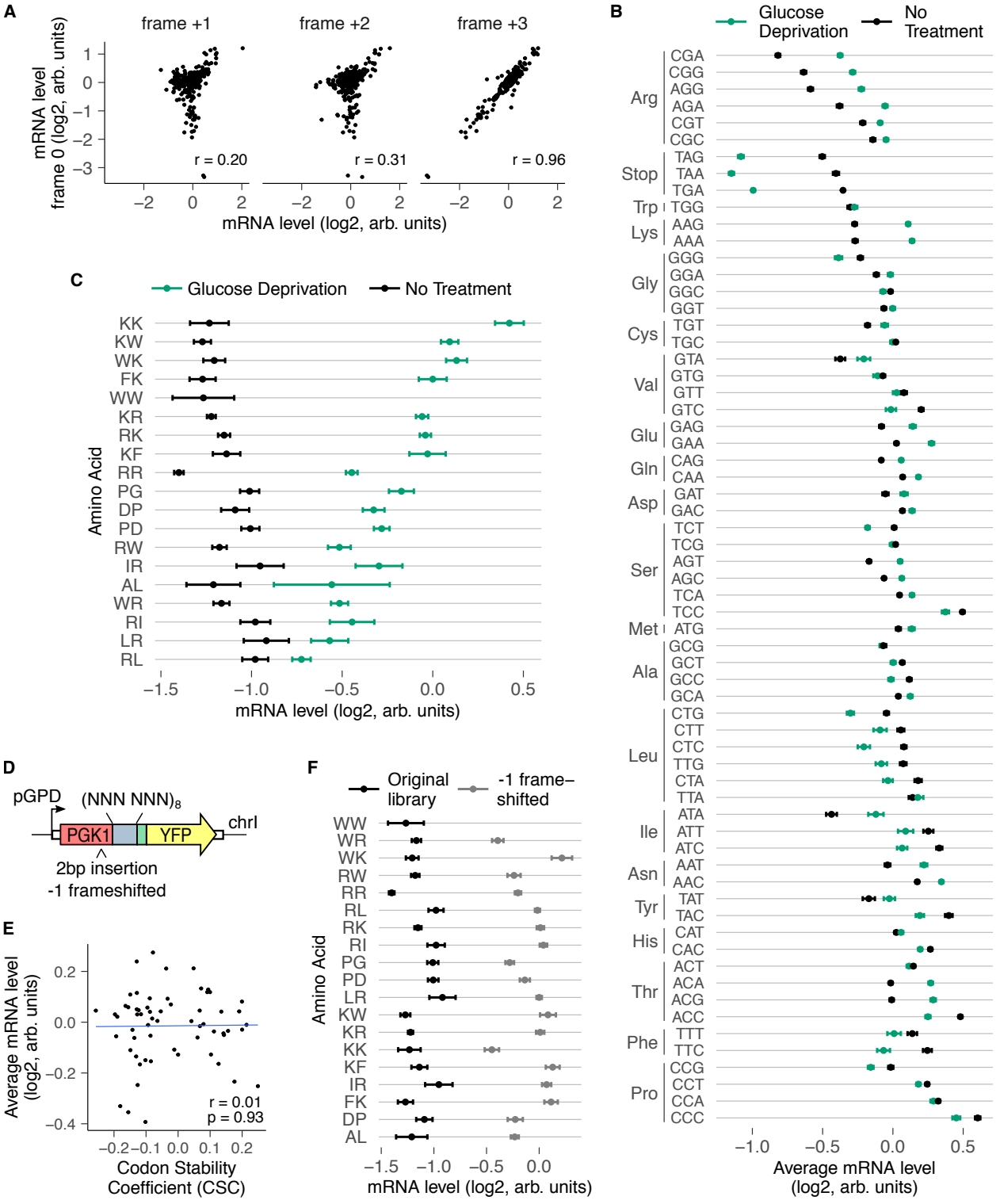

Fig. 3: mRNA effects of dipeptide repeats require in-frame translation. (A) mRNA level of reporters encoding 320 different dipeptide repeats (excluding stop codon-containing dipeptides and

pairs that did not pass read count cutoffs) compared between the correct reading frame (frame 0, vertical axis) and computationally-shifted +1, +2, or +3 reading frames (horizontal axes). *r* indicates Pearson correlation coefficient. **(B)** Average mRNA level of reporters with indicated codons averaged across positions 1 and 2 of the codon pair library during normal growth and glucose depletion. mRNA levels were median-normalized separately for each growth condition. Error bars represent standard deviation over all variants containing the codon at either position. **(C)** mRNA level of reporters encoding indicated dipeptides during normal growth and glucose deprivation. mRNA levels were median-normalized separately for each growth condition. Only dipeptide inserts with a minimum of 10 reads per barcode, 4 barcodes per insert, and low variability between barcodes are included here and in further analysis. Error bars represent standard deviation over barcodes linked to the indicated dipeptide repeat. **(D)** Schematic of frameshifted codon pair library. Two base pairs were inserted upstream of the codon pair insert in the 4096 codon pair library to create a -1 frameshift in the codon pair. Libraries were integrated and sequenced as in Fig. 1A. **(E)** Average mRNA effects of individual codons in the -1 frameshifted library compared against codon stability coefficients[55]. **(F)** mRNA levels of destabilizing dipeptides in the original in-frame library and in the -1 frameshifted library. Error bars calculated as in Fig. 3C.

Chapter 3: Identification of regulatory pathways mediating mRNA destabilization by codon pairs

Ribosome-associated quality control regulates mRNA destabilization by dipeptide motifs

Given the translational dependence of mRNA destabilization by dipeptide repeats, we sought to identify the co-translational regulatory pathways mediating these effects. Ribosome stalling at poly-lysine, poly-arginine, and poly-tryptophan repeats triggers ribosome-associated quality control (RQC) of nascent peptides and mRNAs in *S. cerevisiae*[53,57,60,99,100]. The E3 ubiquitin ligase Hel2 (*S. cerevisiae* homolog of human ZNF598), which binds collided ribosomes at extended ribosome stalls, is necessary for RQC induction at these sequences[60,67,81,100,106–108] (Fig. 4A). Syh1 (GIGYF2 in humans) has also been recently implicated in a Hel2-independent pathway of mRNA decay of reporters with repeats of the rare codon CGA[109–111] (Fig. 4A). To test the requirement for these factors in reducing the mRNA levels at the novel destabilizing dipeptide repeats identified in our screen, we integrated our original 4096-codon pair library into *S. cerevisiae* strains with *HEL2* or *SYH1* deletion, and measured relative mRNA levels as before (Fig. 4B).

We compared by linear regression the relative mRNA levels in the *hel2Δ* and *syh1Δ* strains against the wild-type strain to identify inserts with altered mRNA levels (Fig. 4C,D).

In the *hel2Δ* strain, 14 dipeptides had 1.5-fold or greater increase in relative mRNA levels compared to the wild-type strain (red points, Fig. 4C). These include the known RQC-inducing repeats, KK, RR, WW, RK, and KR. *HEL2* deletion also restored the mRNA levels of several bulky and positively charged dipeptide repeats (FK/KF, WR/RW, WK/KW) as well as proline-aspartic acid (PD/DP) and proline-glycine (PG/GP) repeats (Fig. 4E, Supplementary Fig. 3C). By contrast, *SYH1* deletion did not restore the mRNA levels of any dipeptide repeat (Fig. 4D,E). This is likely because Syh1 acts as a compensatory mechanism when Hel2-mediated RQC is inactive[109]. mRNA destabilization

by a few combinations of positively charged and hydrophobic amino acids (RL/LR, RI/IR) was not rescued by either *HEL2* or *SYH1* deletion. Together, these results reveal that Hel2-mediated RQC regulates most but not all mRNA-destabilizing effects of dipeptide repeats identified in our original screen.

Figure 4

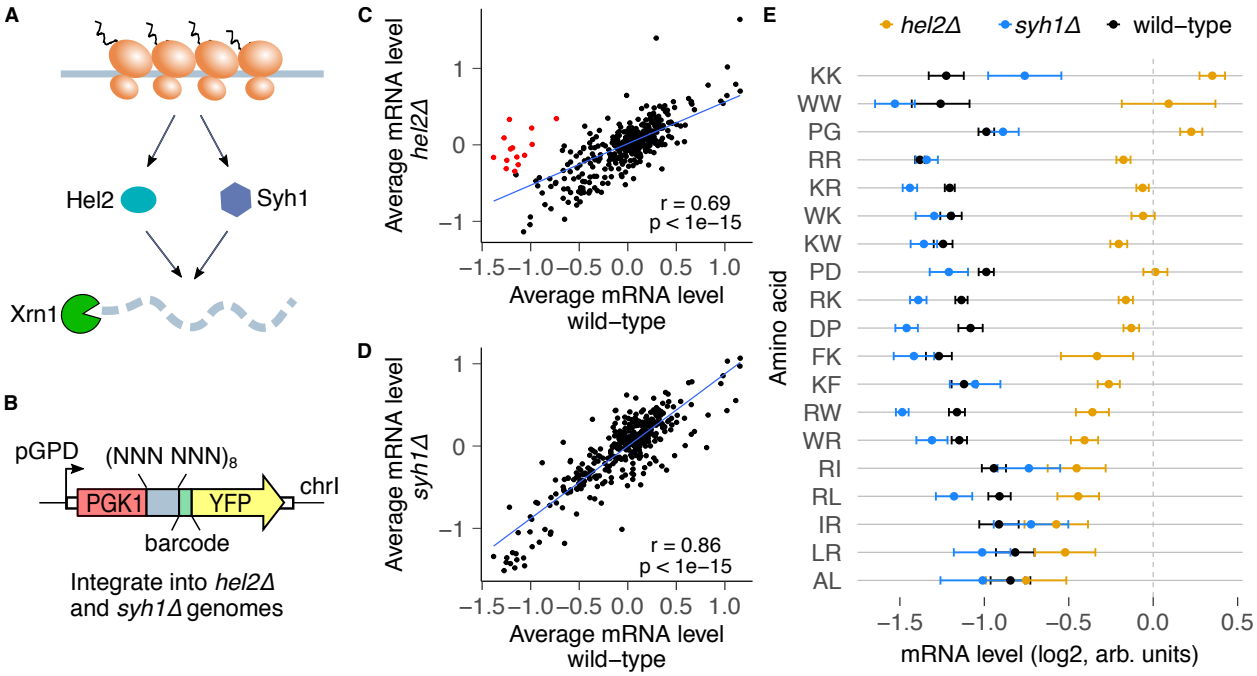

Fig.4: Ribosome collision sensor Hel2 regulates the mRNA effects of dipeptide repeats.

(A) The RQC factors Hel2 and Syh1 are known to respond to collided ribosomes and trigger mRNA decay through Xrn1. **(B)** The codon pair library in Fig. 1A was integrated into *hel2Δ* and *syh1Δ* cells, and mRNA levels were quantified as before. **(C)** mRNA levels for dipeptide repeats compared between *hel2Δ* and wild-type cells. mRNA levels were calculated as in Fig. 3C, and median-normalized separately for each strain. Dipeptide repeats with residuals less than -2 from the linear regression line are marked in red. **(D)** Same plot as in C, but for *syh1Δ* cells. No dipeptide repeats are preferentially stabilized in *syh1Δ* cells with residuals less than -2 from the linear regression line. **(E)** mRNA levels for wild-type mRNA-destabilizing dipeptides (from Fig. 3C) for *hel2Δ* and *syh1Δ*

cells. Error bars represent standard deviation over barcodes linked to the indicated dipeptide repeat.

Future directions

Another major mode of co-translational mRNA decay is codon optimality mediated decay, which is initiated by Not5. Future experimentation will involve performing the same analysis described above in *not5Δ* cells to determine which sequences are mediated COMD.

Chapter 4: Deep mutational scanning identifies critical residues mediating mRNA destabilization by dipeptide motifs

Deep mutational scanning of a positively charged and bulky repeat dipeptide

Ribosome-associated quality control often depends on interactions between specific residues in the nascent peptide and various regions of the ribosome such as the peptidyl-transferase center (PTC) and the uL4/uL22 constriction point in the exit tunnel[46,53,63,90,91]. To dissect the mechanism by which the FK dipeptide repeat triggers mRNA destabilization, we developed a deep mutational scanning assay using reporter mRNA level as a readout (Fig. 5A). Specifically, we mutated each location in the 16-codon insert encoding $(FK)_8$ to all 64 codons to generate a pooled library of 1024 variants. We cloned these variants between the *PGK1* and *YFP* coding sequences, integrated them into the genomes of wild-type and *hel2Δ* cells, and measured variant frequency in cDNA and genomic DNA by high throughput amplicon sequencing. We used the ratio of cDNA to genomic DNA to quantify the relative mRNA levels of each variant, and further normalized to spike-in control strains to enable comparison across different genotypes (see Methods). We confirmed reproducibility of mRNA levels between biological replicate transformations into *S. cerevisiae* of the same plasmid library (Fig. 5B).

Visualizing the relative mRNA levels of $(FK)_8$ mutants as a function of mutation identity and location yields several interesting observations (Fig. 5C). First, reporter mRNA levels increase sharply when stop codons are present at positions 11 through 16 in wild-type cells. Since translation of premature stop codons will trigger mRNA decay through the Hel2-independent NMD pathway, our results imply that a minimum of 10 residues of the FK dipeptide need to be translated in order to trigger Hel2-driven RQC over NMD when the $(FK)_8$ variants contain a stop codon. Interestingly, we also observe NMD suppression when stop codons are introduced after 10 residues of $(FK)_8$

in *hel2Δ* cells, suggesting that extended ribosome stalling or collisions on mRNAs are sufficient to suppress NMD. Second, mRNA levels for nearly all mutations from positions 1 to 6 are as low as the wild-type sequence. This observation is again consistent with 10 residues in $(FK)_8$ being the minimum RQC-inducing length, because mutating any of the first six residues will preserve this minimum length downstream of the mutated position. Pro is the only target mutation within the first six positions that consistently rescues mRNA levels, likely by limiting the conformational flexibility of the nascent peptide[112–114]. Third, location 12 (and to a lesser extent location 14) within $(FK)_8$ are the sole positions that require positively charged Arg or Lys to trigger Hel2-dependent RQC. At several other locations where the original amino acid is positively charged (such as at positions 6, 8, and 10), mutation to the bulkiest Trp residue can still trigger RQC, while mutations to other aromatic amino acids (Phe and Tyr) are insufficient. Fourth, at some locations where the original amino acid is bulky (such as at positions 9 and 11), mutating to the bulkier Trp or to positively charged Arg or Lys maintains RQC. The two preceding observations imply that positive charge and bulkiness play interchangeable roles at several locations within the $(FK)_8$ repeat in triggering RQC. Finally, at position 7, where the original amino acid is Phe, mutations to other aromatic amino acids (Trp or Tyr) or to a negatively charged residue (Glu or Asp) triggers RQC, while positive charge is insufficient. Thus, the interchangeability of bulkiness with positive charge in triggering RQC is not universal, but rather depends on the location within the stalling peptide.

We next compared the aggregate effect of all mutations at each location of the $(FK)_8$ repeat on mRNA levels between wild-type and *hel2Δ* cells (Fig. 5D). We excluded stop-codon containing mutants from this analysis to avoid convoluting NMD and RQC effects. The positions with the highest mutational effect differences between the two strains are at the ends of the stalling sequence: positions 1-6, 15, and 16 of $(FK)_8$. This observation is consistent with our earlier interpretation that translation of 10 residues of $(FK)_8$ is necessary to drive Hel2-dependent mRNA decay. Con-

versely, positions 10, 9, and 12 had the least mutational effect differences between the two strains, revealing that these positions are most important for triggering Hel2-dependent RQC. Finally, *HEL2* deletion did not fully rescue the mRNA effects of any $(FK)_8$ terminal mutants (positions 1, 15, 16), suggesting that Hel2-dependent RQC activity is saturated at longer repeat lengths, and mRNA decay proceeds through multiple compensatory pathways.

Figure 5

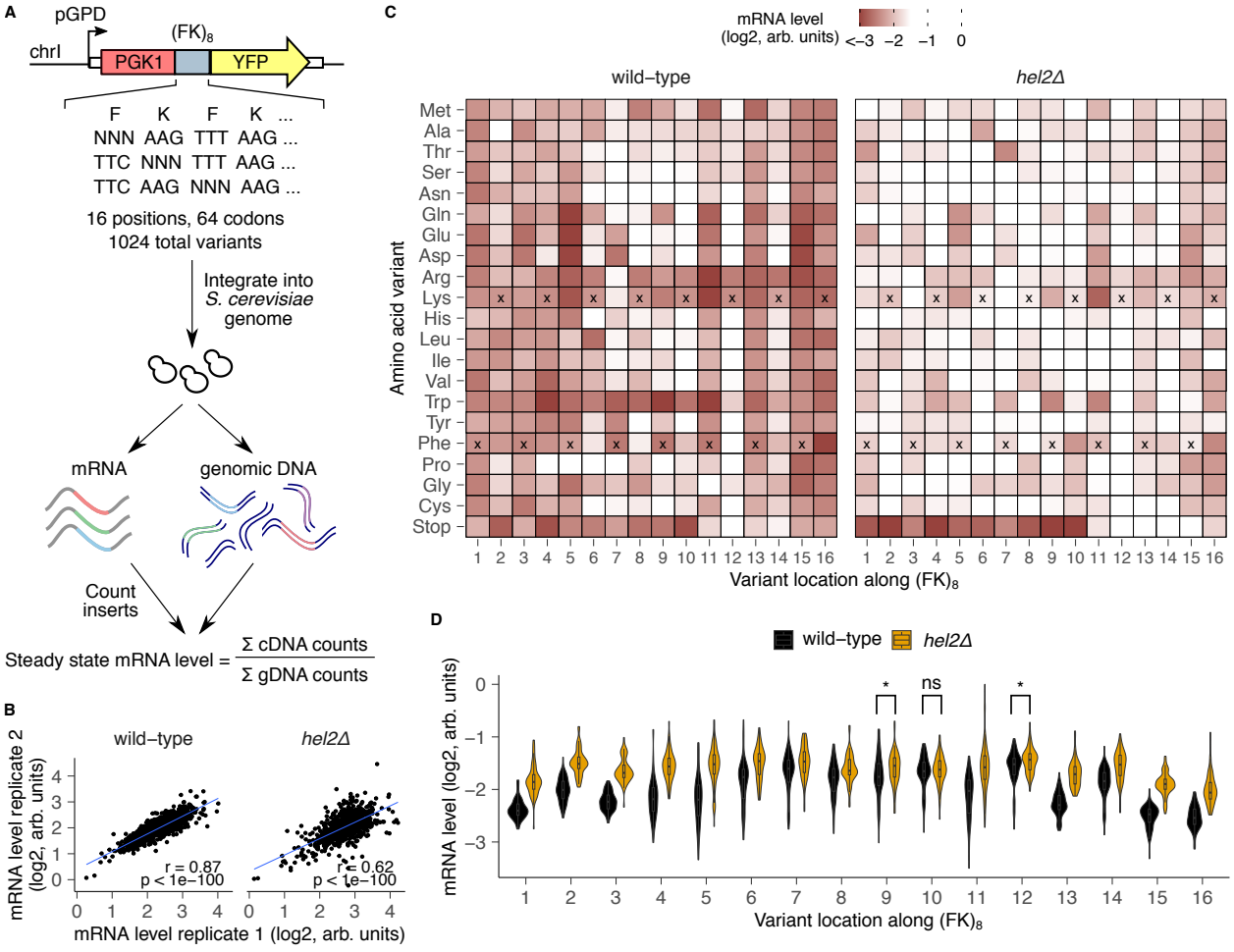

Fig. 5: Deep mutational scanning identifies amino acids critical for mRNA effects of a destabilizing dipeptide repeat.

(A) Schematic of deep mutational scanning (DMS) of the FK dipeptide repeat. Each location in an $(FK)_8$-encoding insert was randomized to all 64 codons. This 1024-variant library was cloned as a

pool between *PGK1* and *YFP*, and genomically integrated into wild-type and *hel2Δ* strains. Inserts were quantified in cDNA and genomic DNA by high throughput amplicon sequencing. **(B)** Pearson correlation between biological replicates for each variant in the $(FK)_8$ DMS library. **(C)** mRNA level for inserts containing the indicated amino acid mutation (vertical axis) at the indicated position (horizontal axis). mRNA levels are averaged across replicates and normalized within each genotype using spike-in control strains. The wild-type amino acid variant is marked with black crosses at each location. **(D)** Violin plots of mRNA level across all amino acid variants at each location in wild-type and *hel2Δ* cells for both replicates combined. Stop codon variants are excluded from this analysis. Any locations where distributions were not significantly different ($p>0.01$ by Wilcoxon rank sum test) are marked.

Future directions

Two specific proline containing motifs, $(PD)_8$ and $(PG)_8$, were also found to cause co-translational mRNA decay and be regulated by Hel2. However, the mechanism for stalling of these two motifs cannot be driven by positive charge or bulkiness as in $(FK)_8$. Future experimentation will involve deep mutational scanning performed as described above, but for the $(PD)_8$ and $(PG)_8$ dipeptides. Furthermore, our deep mutational scanning results unexpectedly identified an interesting interplay between regulatory pathways: Hel2-driven RQC and NMD. To further dissect the relationship between these pathways, subsequent experimentation will involve integration of these deep mutational scanning libraries in *upf1Δ* cells.

Chapter 5: Evaluation of mRNA effects in endogenous sequences

Codon pair library predicts mRNA effects of endogenous sequences

Though a few mRNA sequences are known to stall ribosomes and trigger RQC in reporter studies[63,88,115], the sequence motifs that underpin endogenous mRNA stability are not well understood. For example, the simple presence of polybasic stretches or rare codons is not sufficient to trigger quality control on endogenous yeast mRNAs[63,116]. Thus, we sought to test whether our codon pair assay could predict mRNA effects of sequence motifs in endogenous *S. cerevisiae* genes. To this end, we assayed 1904 fragments, each 48 nucleotides long, from endogenous mRNAs spanning a wide range of expression levels[117] using the same reporter design as the codon pair library (Fig. 6A). We integrated this endogenous fragment library into wild-type cells and counted barcodes by high throughput amplicon sequencing as before. Compared to the codon pair library, mRNA levels in the endogenous fragment library were more tightly distributed around the median, indicating more muted effects on mRNA stability (Fig. 6B). We next calculated the codon stability coefficient (CSC) values for each of the 64 codons using mRNA levels either from the codon pair library or the endogenous fragment library[55]. We found strong correlation ($r=0.67$, $p<1e-8$) between the two libraries, indicating that mRNA effects of codon pair repeats predict mRNA effects of endogenous sequence motifs in wild-type cells (Fig. 6C). We next integrated the endogenous fragment library into *hel2Δ* cells and tested how Hel2-dependent RQC affects the relationship between CSC values calculated from the codon pair and the endogenous fragment libraries. We found that *hel2Δ* cells still exhibited a significant correlation ($r=0.49$, $p<1e-4$) between the two libraries, though to a lesser extent than in wild-type cells (Fig. 6D), consistent with the majority of endogenous sequences not triggering prolonged ribosome slowdown or collisions.

Figure 6

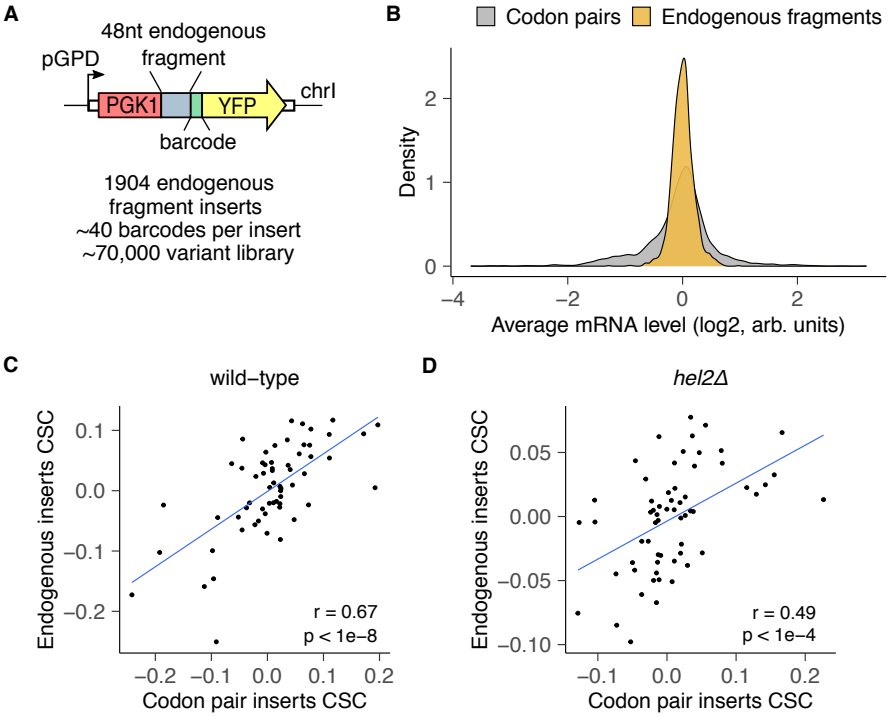

Fig.6: Codon pair measurements predict effects of endogenous mRNA sequences.

(A) Schematic of endogenous sequence insert library. Each element in the library includes one of 1904 possible 48nt endogenous fragments. Each sequence is inserted in-frame between *PGK1* and *YFP*, and is followed by a random 24nt barcode without in-frame stop codons (median of 40 barcodes/insert). The 70,000 variant library is genomically integrated into wild-type and *hel2Δ* cells, and mRNA levels are quantified as in Fig. 1A. **(B)** Distribution of mRNA levels for endogenous fragments vs codon pair inserts in wild-type cells. **(C)** Correlation between CSC values calculated for each codon from the endogenous fragment library against CSC values derived from the codon pair library in wild-type cells. Pearson correlation coefficient is reported as *r*. The CSC for each codon is calculated by taking the Pearson correlation coefficient between codon frequency of an insert and its steady state mRNA level. **(D)** Same plot as in C, but for *hel2Δ* cells.

Future directions

The strong correlation in CSCs observed in *hel2Δ* cells suggests that endogenous sequences may not trigger ribosome stalling strong enough to require Hel2 and RQC. Another co-translational decay pathway that has been described to act on endogenous mRNAs is the Not5 driven COMD pathway. Future experimentation will involve performing the same analysis described above in *not5Δ* cells.

Chapter 6: Materials and Methods

Parent vector construction

Plasmids constructed and used in this study are listed in table S1. Oligonucleotides used in this study are listed in table S3. Plasmid assembly was carried out using standard molecular biology techniques as described below. All polymerase chain reaction (PCR) reactions were performed using Phusion polymerase (Thermo Fisher F530S) or Phusion Flash High-Fidelity PCR Master Mix (Thermo Fisher F548L) according to manufacturer's instructions. Restriction enzymes were obtained from Thermo Fisher and FastDigest (FD) variants were used when available.

The chrI-integrating parent vector pHPSC1120 used for this study was constructed from pHPSC417 used in our previous work[71]. In comparison to pHPSC417, pHPSC1120 contains an additional Illumina Read1 primer binding site and T7 promoter sequences for deep sequencing of inserts and barcode sequences and for in vitro transcription from genomic DNA, respectively. The Illumina R1 sequencing primer binding and T7 promoter sequences were PCR-amplified using oHP558 as the forward primer, oHP530 as a bridge primer, and oHP529 as a reverse primer, and cloned into BamHI-linearized pHPSC417 using Gibson assembly. The -1 frameshifted parent vector pHPSC1114 was also constructed from pHPSC417 using the same strategy as for pHPHS1120 but with a different forward primer oHP528 that incorporates the frameshift. All plasmids were verified by Sanger sequencing.

Variable oligo pool design

Pool 1

Pool 1 includes the 8× dicodon library (4096 codon pair inserts) and the endogenous gene fragments library (1904 inserts). The 8× dicodon library (Fig. 1A) encodes all possible codon pair (6 nucleotide) combinations, for a total of 4096 codon pairs. Each codon pair is repeated eight times

to create 48 nucleotide (nt) inserts. The endogenous gene fragments library includes 1904 endogenous fragments, each 48 nt in length (Fig. 6A). Endogenous gene fragments were selected as 253 nt to 300 nt of each ORF. Only ORFs designated as "Verified" by the Saccharomyces Genome Database (SGD) in the R64-1-1 release were included (http://sgd-archive.yeastgenome.org/sequence/S288C_reference/genome_releases/). Every 2nd gene in descending order of RNA expression[117] was included in this library to encompass a wide range of expression levels. All 6000 inserts are flanked with the same 29 nt 5' homology arm and 24 nt 3' homology arm. The oligo pool (oAS385) was ordered from Twist Biotechnologies.

Pool 2

The FK_8 deep mutational scanning library (Fig. 5A) was constructed from a starting sequence composed of phenylalanine and lysine codons repeated eight times in tandem (48 nt inserts). The phenylalanine codons TTT and TTC and the lysine codons AAA and AAG were used interchangeably throughout the insert to avoid producing a repetitive mRNA sequence. At each of the 16 positions, an NNN sequence was used to randomize the codon. The oligo pool (oKC224) was ordered as an oPool from Integrated DNA Technologies.

Plasmid library construction

For the 8× dicodon library, oligo pool 1 (described above) was PCR-amplified with oKC97 and oHP531. For the -1 frameshifted 8× dicodon library, pool 1 was PCR-amplified with oHP532 and oHP531. As described above, oHP531 encodes a 24 nt random barcode region, comprised of 8× VNN repeats. Barcoded oligo pools were cloned into BamHI-linearized pHPSC1120 and pHPSC1114 by Gibson assembly. Assembled plasmid pools were transformed at high efficiency into NEB 10-Beta *E. coli* cells, and plated as 1:10 serial dilutions. 500,000 colonies were scraped from plates for extraction in order to bottleneck the number of unique variants.

Pool 2 was PCR-amplified with oKC97 and oKC225 and cloned into BamHI-linearized pHPSC1120 by Gibson assembly. The assembled plasmid pool was transformed at high efficiency into NEB 10-Beta *E. coli* cells. 70,000 colonies were scraped from plates for extraction in order to bottleneck the number of unique variants.

Individual plasmid construction

To generate the *PGK1-YFP* reporters used for flow cytometry of individually selected codon pairs, the desired codon pair inserts were amplified using two rounds of PCR from a pooled plasmid template pHPSC1136 not used in this study. Unique primers (oKC129-142) were used to amplify the six desired inserts. Homology arms were added to the six amplified inserts using oKC97 and oKC123 primers. Amplified inserts were cloned into BamHI-linearized pHPSC1120 by Gibson assembly to produce pHPSC1144, pHPSC1145, pHPSC1146, pHPSC1147, pHPSC1149, pHPSC1150 plasmids. All individual plasmids were verified by Sanger sequencing.

To create the small barcoded pool for mRNA measurement validation (Fig. S2A,E), oKC97 and oKC148 oligos were used to barcode and amplify inserts from the following plasmids (described above): pHPSC1144, pHPSC1145, pHPSC1146, pHPSC1147, pHPSC1149, pHPSC1150. oKC148 encodes a 24 nt random barcode region, comprised of 8× VNN repeats. Barcoded inserts were then pooled at equimolar concentrations and cloned into BamHI-linearized pHPSC1120 by Gibson assembly. The assembled plasmid pool was transformed at high efficiency into NEB 10-Beta *E. coli* cells. 2,000 colonies were scraped from plates for extraction in order to bottleneck the number of unique variants. Two colonies were picked and Sanger sequenced to obtain the identity of the insert and barcode pair of the two spike-in plasmids, pHPSC1159-sc2 and pHPSC1159-sc5.

Strain construction

All *S. cerevisiae* strains used in this study are listed in table S2. Integration of pooled plasmids into the *S. cerevisiae* genome was performed by transforming 30–200 µg of NotI-linearized plasmid library into 1–5e9 cells according to the LiAc/SS carrier DNA/PEG method[118]. Following heat shock, cells were transferred into a 5x volume of a 1:1 solution of 20% dextrose and synthetic complete (SC) media lacking uracil with 2% dextrose (SCD-URA) and spun at 1850*g* for 5 minutes. Cell pellets were gently resuspended in 100mL of fresh SCD-URA and allowed to recover overnight at 30˚C shaking at 200rpm. After 20–24 hours, 1e9 cells were passaged into 100mL fresh SCD-URA and grown overnight at 30˚C shaking at 200rpm. This process was repeated for a total of 72 hours of selection in SCD-URA before making glycerol stocks from saturated cultures. Integration of individual constructs into the *S. cerevisiae* genome was performed by transforming 0.5–1.0μg of linearized plasmid according to the LiAc/SS carrier DNA/PEG method[118]. Single yeast colonies were selected on SCD agar plates lacking uracil after 48 to 72 hours growth at 30˚C.

Harvesting pooled library cells

Glycerol stocks of cells containing pooled reporter strains were thawed and grown overnight in 20-50mL YEPD at starting OD_{600} between 0.1 and 0.5 at 30˚C with shaking at 200rpm. The saturated cultures were diluted approximately 200-fold (for starting OD_{600} of 0.1) and spike-in strains (scKC190 and scKC191) were introduced into each culture at a concentration approximately the same as each library variant based on OD_{600} density. Cultures were grown for 4–6 hours at 30˚C with shaking at 200rpm until mid-log phase (OD_{600} between 0.4-0.6), then transferred to ice-water baths. Each culture was split into 50mL aliquots (approximately >=200 million cells) in pre-chilled conical tubes and spun down at 3000*g*, 4˚C, for 5 minutes. The supernatant was removed and the cell pellets were flash-frozen in a dry ice-ethanol bath and stored at -80˚C.

Harvesting glucose-depleted cells

Glycerol stocks of cells containing the pHPSC1142 pooled reporter library were thawed and grown overnight as described above. Saturated cultures were diluted and spike-in strains (scKC190 and scKC191) were introduced as described above. Cells were grown for 4 hours at 30˚C with shaking at 200rpm until OD_{600} of 0.4. Cells were spun down at 3000rpm for 2 minutes and washed with 30mL H_2O twice, then resuspended into YEP (no glucose). Glucose-depleted cells were grown for 1 hour at 30˚C with shaking at 200rpm. After 1 hour of growth, cells were harvested by spinning in 50mL pre-chilled tubes at 3000*g*, 4˚C, for 5 minutes. The supernatant was removed and the cell pellets were flash-frozen in a dry ice-ethanol bath and stored at -80˚C.

Library genomic DNA extraction

For genomic DNA extraction, between 400 million to 1.2 billion cells (two to six flash-frozen pellets) were lysed and extracted using the YeaStar Genomic DNA kit (Zymo 11-323), following the manufacturer's instructions, with 240μL YD digestion buffer and 10μL R-Zymolyase per pellet. Extracted genomic DNA was sheared for 10 minutes (30 seconds on, 30 seconds off, on "High" setting) on ice using a Diagenode Bioruptor. Sheared gDNA was cleaned using DNA Binding Buffer (Zymo ZD4004-1-L) and UPrep Spin Columns (Genesee Scientific 88-143). Sheared and cleaned gDNA was then in vitro transcribed into RNA (denoted gRNA below and in analysis code) starting from the T7 promoter region in the insert cassette, similar to previous approaches[78,119], using the HiScribe T7 High Yield RNA Synthesis Kit (NEB E2040S). Transcribed gRNA was cleaned using the RNA Clean and Concentrator kit (Zymo R1013).

Library mRNA extraction

At least 200 million cells (one flash-frozen pellet) per sample was resuspended in 400μL Trizol (Thermo Fisher 15596-026) in a 1.5-ml tube and vortexed with 500µl of glass beads (Sigma G8772)

at 4°C for 10 min (2 minutes on, 1 minute on ice). RNA was extracted from the resulting lysate using the Direct-zol RNA Miniprep Kit (Zymo R2070) following manufacturer's instructions.

mRNA and genomic DNA barcode sequencing

For pHPSC1142, pHPSC1117, and pHPSC1160 libraries, between 0.5-10μg of mRNA and gRNA for each library was reverse transcribed into cDNA using SuperScript IV (Thermo Fisher 18090050) and a primer annealing to the Illumina R1 primer binding site (oPB354). A 170 nt region surrounding the 24 nt barcode was PCR-amplified from the resulting cDNA in two rounds. Round 1 PCRs used cDNA template comprising 1/5th of the PCR reaction volume and primers oPB354 and oHP534. Round 1 PCR cycle numbers were adjusted as needed to obtain adequate product concentration while avoiding overamplification (between 5 and 15 cycles), then cleaned using DNA Binding Buffer (Zymo ZD4004-1-L) and UPrep Micro Spin Columns (Genesee Scientific 88–343). Cleaned samples were then used as template for Round 2 PCR, and cycles were again adjusted to avoid overamplification (between 4 to 8 cycles). Round 2 PCRs used Round 1 PCR product comprising between 1/10th to 1/5th of the PCR reaction volume and oAS111 with indexed forward primers (oAS112-135 and oHP281-290). Amplified samples were run on a 2% agarose gel and fragments of the correct size were purified using ADB Agarose Dissolving Buffer (Zymo D4001-1-100) and UPrep Micro Spin Columns (Genesee Scientific 88–343). Concentrations of gel-purified samples were measured using a Qubit dsDNA HS Assay Kit (Q32851) with a Qubit 4 Fluorometer. Samples were sequenced using an Illumina NextSeq 2000 in 1×50, 2×50, or 1×100 mode (depending on other samples pooled with the sequencing library). For the pHPSC1142 libraries, samples were sequenced with standard Read 1, standard Read 2, and standard i7/i5 index sequencing primers. A subset of these libraries were sent for re-sequencing to obtain greater read depth and sequenced with standard Read 1, custom Read 2 oAS1638 (to maintain compatibility with other libraries in the pool), and standard i7/i5 index sequencing primers. For the pHPSC1117 libraries, samples were

sequenced with the standard Read 1 sequencing primer and standard index sequencing primers. For the pHPSC1160 libraries, samples were sequenced with standard Read 1, standard Read 2, and standard index sequencing primers.

For the FK_8 library (pHPSC1163), between 0.5-10μg of mRNA and gRNA were reverse transcribed into cDNA using SuperScript IV and a primer annealing to the Illumina R1 primer binding site that contains a 7 nt unique molecular identifier (UMI) (oKC235). A 195 nt region surrounding the 48 nt insert was PCR-amplified from the resulting cDNA in one round using oPN776 and indexed forward primers (oKC230-233, oKC239-242). PCR cycle numbers were adjusted as needed to obtain adequate product concentration while avoiding overamplification (between 10 to 17 cycles). Amplified samples were size-selected and quantified as described previously. Samples were sequenced using an Illumina NextSeq 2000 in 1×70 mode using standard Read 1, custom i7 sequencing primer oKC256, standard i5, and custom Read 2 sequencing primer oKC236.

The 8× dicodon library (pHPSC1142) in glucose-depleted cells was reverse transcribed following the same procedure and primer as pHPSC1163 described above. A 219 nt region surrounding the 48 nt insert and 24 nt barcode was PCR-amplified from the resulting cDNA in one round using oPN776 and indexed forward primers (oKC230-233, oKC239-242). PCR cycle numbers were adjusted as needed to obtain adequate product concentration while avoiding overamplification (between 8 to 16 cycles). Amplified samples were size-selected and quantified as described previously. Samples were sequenced using an Illumina NextSeq 2000 in 1×70 mode using standard Read 1, custom i7 sequencing primer oKC256, standard i5, and custom Read 2 sequencing primer oKC236.

Insert-barcode linkage sequencing

8–10 ng of plasmid pools (pHPSC1142, pHPSC1160, pHPSC1117) were used in PCR using Phusion polymerase (Thermo Fisher F530S) or Phusion Flash High-Fidelity PCR Master Mix (Thermo Fisher F548L). Round 1 PCR was carried out for up to 10 cycles, with 8-10 ng plasmid pool template comprising 1/5th of the PCR reaction volume, using primers oPB354 and oHP534. Round 1 PCRs were cleaned using DNA Binding Buffer (Zymo ZD4004-1-L) and UPrep Micro Spin Columns (Genesee Scientific 88–343). Cleaned samples were used as template for Round 2 PCR, carried out to between 4 to 8 cycles, using oAS111 and indexed forward primers (oAS112-135 and oHP281-290). Amplified samples were purified after size selection and quantified as described above. Samples were sequenced using an Illumina NextSeq 2000 in 2×50 or 1×100 mode. For the pHPSC1142 library, sequencing was performed using standard Read 1 sequencing primer, standard index sequencing primers, and custom Read 2 sequencing primer oAS1637. For the pHPSC1117 library, sequencing was performed using standard Read 1 sequencing primer and standard index sequencing primers. For the pHPSC1160 library, sequencing was performed using standard Read 1, standard Read 2, and standard index sequencing primers.

Flow cytometry

Five single *S. cerevisiae* colonies integrated with plasmids described above were inoculated into separate wells of 96-well plates containing 150 µl of SCD-URA medium in each well and grown overnight at 30°C with shaking at 800rpm. The saturated cultures were diluted 100-fold into 150µl of fresh SCD-URA medium and grown for 5-6 hours at 30°C with shaking at 800rpm. The plates were placed on ice and analyzed using the 96-well attachment of a BD FACS Aria or Symphony cytometer. Forward scatter (FSC), side scatter (SSC), YFP fluorescence (FITC), and RFP fluorescence (PE.Texas.Red) were measured for 10,000 cells in each well. The resulting data in individual

.fcs files for each well were combined into a single tab-delimited text file. YFP expression was first normalized to RFP expression per cell (henceforth referred to as YFP/RFP), then used to calculate the median value of each well. For the no-insert control, the median YFP/RFP values of all wells were averaged together. The median YFP/RFP value per replicate for all strains were then normalized to the average no-insert control value by taking the log2 difference. The average and standard error of this ratio across replicates were calculated (Fig. 2D).

Computational analyses

Pre-processing steps for high-throughput sequencing were implemented as Snakemake workflows run within Singularity containers on an HPC cluster. All container images used in this study are publicly available as Docker images at https://github.com/orgs/rasilab/packages. Python (v3.9.15) and R (v4.2.2) programming languages were used for all analyses unless mentioned otherwise.

Barcode to insert assignment

The raw data from insert-barcode linkage sequencing are in FASTQ format. All pertinent reads were concatenated into one FASTQ file using `fasterq-dump --concatenate-reads`, and inserts and barcodes were extracted and counted using `awk` (mawk implementation, v1.3.4). Only insert-barcode combinations where the insert matches a reference sequence in the list of reference sequences using `awk` were retained. Barcodes were aligned against themselves using `bowtie2` with options `-L 19 -N 1 --all --norc --no-unal -f`. This self-alignment was used to exclude barcodes that are linked to different inserts or that are linked to the same barcode but are aligned against each other by `bowtie2`. In the latter case, the barcode with the lower count is discarded in `filter_barcodes.ipynb`. The final list of insert-barcode pairs is written as a comma-delimited .csv file for aligning barcodes from genomic DNA and mRNA sequencing below.

Barcode counting in genomic DNA and mRNA

The raw data from sequencing barcodes in genomic DNA and mRNA is in FASTQ format. All pertinent reads were concatenated into one FASTQ file, and barcodes were extracted and counted using `awk`. For barcodes that are present in the filtered barcodes .csv file from linkage sequencing, the barcode count and associated insert are printed into a .csv file for subsequent analyses in R. For libraries containing both barcodes and UMIs, only distinct barcode-UMI combinations where the barcode is present in the filtered barcodes .csv file from linkage sequencing are retained. The number of UMIs per barcode and associated insert are printed into a .csv file for subsequent analyses in R.

mRNA quantification and statistical analyses for barcode sequencing

Only barcodes with a minimum of 10 reads and inserts with a minimum of 2–4 barcodes were included. The mRNA level for each insert was calculated as the mean log2 ratio of the summed mRNA barcode counts to the summed gRNA barcode counts using 100 bootstrap samples. The standard deviation was calculated across all barcodes for each insert using 100 bootstrap samples. For libraries with a large number of variants (e.g. >= 70,000) mRNA levels were median-normalized within each library. For libraries with a smaller number of variants (e.g. 1000-2000), libraries were normalized to spike-in strain barcode counts or library size (RPM). For all other experiments, the standard error of the mean was calculated using the `std.error` function from the `plotrix` R package. P-values for statistically significant differences were calculated using the `t.test` or `wilcox.test` R functions as appropriate for each figure (see figure captions).

Insert counting and mRNA quantification

For the FK_8 deep mutational scanning library, inserts were sequenced directly and thus barcodes were not counted or used for statistical analysis. Instead, inserts and UMIs were extracted and

counted using `awk`. Only insert-UMI combinations where the insert matches a reference sequence in the list of reference sequences using `awk` were retained. Subsequent insert-UMI counts were summed across the mRNA and gRNA samples. mRNA levels for each insert were calculated as the log2 ratio of the summed mRNA insert-UMI counts to the summed gRNA insert-UMI counts, and then averaged across the two biological replicates. Resultant mRNA levels were then normalized against mRNA levels of spike-in strains to allow comparison between wild-type and *hel2Δ* cells.

Data and Code Availability

The raw sequencing data generated in this study have been deposited in the Sequence Read Archive under BioProject accession number PRJNA974090, at https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA974090. Raw data from flow cytometry are available at http://flowrepository.org/id/FR-FCM-Z6QH. Code to reproduce figures in the manuscript starting from raw data is publicly available at https://doi.org/10.5281/zenodo.8365102 and https://github.com/rasilab/chen_2023. Software environments used to run the code in the above GitHub repository are publicly available as Docker containers at https://github.com/orgs/rasilab/packages. Biological reagents or methodology clarification can be publicly requested by opening an issue at https://github.com/rasilab/chen_2023/issues.

Conclusions

Here, we use a massively parallel approach to identify and dissect sequence motifs underlying mRNA instability in *S. cerevisiae*. In addition to validating known codon and amino acid effects on mRNA stability, we identify several sequence motifs that have not been previously associated with mRNA decay. These include combinations of bulky and positively charged amino acids, and proline with asparate and glycine, all of which trigger translation-dependent mRNA decay through the Hel2-dependent RQC pathway. By combining our massively parallel assay with deep mutational scanning, we dissect the codon-level biochemical requirements for triggering mRNA decay by a bulky and positively charged dipeptide repeat. Despite the apparent simplicity of the codon pair repeat library, we find that it captures the mRNA effects of endogenous coding sequence fragments from the *S. cerevisiae* transcriptome.

Our codon pair library confirms the role of codon optimality as a major determinant of mRNA stability in *S. cerevisiae*, and provides insights into the resulting hierarchy of effects. We observe several synonymous codon families within which aggregate mRNA levels differ based on the hierarchy of codon optimality[55,97] (Fig. 1C), but have different absolute effects. The nonoptimal codons ATA (Ile), GTA (Val), and TAT (Tyr) are highly destabilized relative to their optimal counterparts. By contrast, the optimal codon TCC (Ser) is preferentially stabilzed relative to its nonoptimal counterparts. Both the arginine and proline synonymous codon families are stratified based on codon optimality even though these amino acids have opposite average effects on mRNA stability (Fig. 1E, Arg – destabilizing, Pro – stabilizing). Thus, codon optimality effects on mRNA stability act in parallel and independent of amino acid identity. Consistent with codon optimality-mediated mRNA decay being a co-translational process[56,120,121], translational shutoff by glucose depletion rescues the mRNA-destabilizing effects of eight out of the 10 most non-optimal codons (ATA, CGA, AGG, GTA, ACG, AGT, AAA, AGC)[55] (Fig. 3B). Finally, the effects of codon optimality on mRNA stabil-

ity in our codon pair library are driven by mutations within a short 16 codon region despite being part of a 700 codon *PGK1-YFP* mRNA. This is likely because the *PGK1-YFP* region is efficiently translated[122], while the tandem and repetitive nature of the codon pairs amplifies their effect on ribosome slowdown and recruitment of mRNA-destabilizing factors.

While polybasic and poly-tryptophan sequences are known to trigger RQC in *S. cerevisiae*, our codon pair assay reveals combinations of bulky (Val, Ile, Leu, Phe, Tyr, Trp) and positively charged (Arg, Lys) amino acids as a general trigger of mRNA decay (Fig. 2A,B). Interestingly, combinations of Val, Ile, Leu, and Phe with Arg and Lys were also found to destabilize mRNA in human cells[78], indicating their evolutionary conservation as mRNA-destabilizing sequences across eukaryotes. Supporting these findings, ribosome profiling in human cells revealed an enrichment in disome occupancy at sites that followed an Arg-X-Lys pattern, with highest disome density occurring when X was Phe, Ile, or Leu[84]. We find that positively charged amino acids in combination with the bulkiest side chains (Phe, Trp) trigger RQC-dependent mRNA decay in *S. cerevisiae*, while less bulky side chains (Val, Ile, Leu) decrease mRNA levels in a Hel2-independent manner (Fig. 4E). We speculate that such mRNA motifs that stall ribosomes sufficiently to trigger mRNA decay but are less terminally stalling than Phe/Trp in combination with Arg/Lys may be acted on by compensatory pathways such as Syh1/Smy2[109], or through Not5-dependent decay[56].

In our codon pair assay, combinations of proline with aspartate and glycine (PD/DP, PG/GP) decrease mRNA levels in a Hel2-dependent manner (Fig. 2A,B, Fig. 4E, Supplementary Fig. 3). While poly-proline sequences stall ribosomes due to inefficient peptide bond formation, these sequences are not known to induce RQC and are instead translated with the assistance of eIF5A[82]. Consistent with these previous findings, proline-proline combinations, and all other proline-containing combinations except for with aspartate and glycine, are stabilizing in our assay. Conversely, no other aspartate or glycine containing codon pairs except the ones with proline

are destabilizing. While increased ribosome occupancy has been observed at proline, aspartate, and glycine codons in both *S. cerevisiae* and human cells[84,123,124], our results suggest that these effects may be driven by combinations of these amino acids rather than by their individual occurrence. Consistent with this idea, PD and PPD peptides have increased ribosome occupancy and are under-represented in the *S. cerevisiae* proteome, while PP and GG dipeptides also have increased ribosome occupancy but are over-represented[85]. Similarly, PD dipeptides in *E. coli*[125], and PD and PG motifs in mouse embryonic stem cells[86] have increased ribosome occupancy. Thus, PD and PG motifs may have evolutionarily conserved effects on ribosome slowdown through a mechanism distinct from poly-proline stalls, and can trigger Hel2-dependent mRNA decay in *S. cerevisiae*.

Our deep mutational scanning reveals complex codon-level requirements for the $(FK)_8$ repeat to confer mRNA instability in a Hel2-dependent manner (Fig. 5). Strikingly, these results also exhibit several similarities to the composite biochemical requirements for ribosome stalling observed at the known endogenous RQC substrate in *S. cerevisiae*, $SDD1_{196\text{-}212}$ (FFYEDYLIFDCRAKRRK)[63]. First, the strict requirement for positive charge at positions 12 and 14 of the $(FK)_8$ repeat to trigger mRNA decay matches the requirement for positive charge at positions 207 and 209 of $SDD1_{196\text{-}212}$, which are thought to perturb the petidyl-transferase center of the ribosome. Second, the requirement for bulky aromatic residues at position 7 of $(FK)_8$ is similar to the requirement for aromatic residues at position 201 of $SDD1_{196\text{-}212}$, which are thought to interact with the uL4/uL22 constriction point of the ribosome. Third, the ability of negatively charged aspartate, and to a lesser extent glutamate, at position 7 of $(FK)_8$ to preserve stalling resembles the requirement for aspartate at position 200 of $SDD1_{196\text{-}212}$, though in the *SDD1* case, the requirement for aspartate is strict. Our results show that bulkiness can be compensated by negative or positive charge in stall sequences depending on the position along the sequence. Specifically, aspartate's prevalence in stalling sequences

is evident in ribosome profiling studies from *S. cerevisiae* to humans, which show increases in monosome and disome occupancy at aspartate codons[84,123,124], presumably due to interactions with the negatively charged ribosome exit tunnel. Taken together, our deep mutational scanning results with a simple $(FK)_8$ repeat recapitulate and generalize the biochemical requirements for ribosome stalling and quality control observed with endogenous stall sequences.

While we did not intend to focus on NMD for this study, our assay nonetheless identified several patterns related to NMD. Surprisingly, we found that glucose depletion selectively destabilized stop codon-containing mRNAs for all three stop codons (Fig. 3B) even though NMD depends on mRNA translation. A possible basis for this observation is that glucose depletion increases the formation of P-bodies, which sequester translationally silenced mRNAs including those subject to NMD[103,104]. Thus, the co-localization of NMD substrates with NMD factors at P-bodies might enhance their decay during glucose depletion[103,105,126,127]. Deep mutational scanning of the $(FK)_8$ dipeptide also revealed the differential kinetics between NMD and RQC when in competition for the same substrates (Fig. 5C). Before 10 Phe and Lys residues are translated, stop-codon containing sequences are predominantly degraded by NMD. After this minimum stalling sequence is translated, RQC dominates as the primary regulatory mechanism. A minimum length of 10 Phe and Lys residues of RQC is consistent with 12 repeated tryptophan residues being sufficient to induce RQC, while greater than 8 residues were required[53]. Interestingly, in *hel2Δ* cells we observe that NMD is suppressed after $(FK)_5$ repeats are translated, even though Hel2-dependent RQC and NMD should presumably not be competing in these cells. This suggests that extended ribosome stalling and collisions is sufficient to prevent degradation of NMD substrates.

The results of our combinatorial codon pair and endogenous motif mRNA stability assays suggest that a wider diversity of mRNA sequences impact mRNA stability than previously appreciated. Poly-GP repeats, identified in our study to stall ribosomes and trigger RQC, are translated through repeat

associated non-ATG (RAN) translation of the pathogenic G_4C_2 repeat expansion in the *C9ORF72* gene and is a biomarker for C9ORF72-associated ALS[128]. Valine-arginine repeats, identified in our study to destabilize mRNAs in a Hel2-independent manner, are also translated through RAN in the mammalian TERRA sequence to form inclusions during disrupted telomere homeostasis[129]. Thus the sequences identified in our study have important implications in the maintenance of cellular homeostasis and disease progression.

References

1. Crick, F. Central Dogma of Molecular Biology. *Nature* **227**, 561–563 (1970).

2. Becker, P. B. & Workman, J. L. Nucleosome Remodeling and Epigenetics. *Cold Spring Harb Perspect Biol* **5**, a017905 (2013).

3. Woychik, N. A. & Hampsey, M. The RNA polymerase II machinery: structure illuminates function. *Cell* **108**, 453–463 (2002).

4. Proudfoot, N. J., Furger, A. & Dye, M. J. Integrating mRNA processing with transcription. *Cell* **108**, 501–512 (2002).

5. Ramanathan, A., Robb, G. B. & Chan, S.-H. mRNA capping: biological functions and applications. *Nucleic Acids Res* **44**, 7511–7526 (2016).

6. Lamond, A. I. The spliceosome. *Bioessays* **15**, 595–603 (1993).

7. Tudek, A. *et al.* Global view on the metabolism of RNA poly(A) tails in yeast Saccharomyces cerevisiae. *Nat Commun* **12**, 4951 (2021).

8. Sharma, A. K. *et al.* A chemical kinetic basis for measuring translation initiation and elongation rates from ribosome profiling data. *PLOS Computational Biology* **15**, e1007070 (2019).

9. Shah, P., Ding, Y., Niemczyk, M., Kudla, G. & Plotkin, J. B. Rate-Limiting Steps in Yeast Protein Translation. *Cell* **153**, 1589–1601 (2013).

10. Tidu, A. & Martin, F. The interplay between cis- and trans-acting factors drives selective mRNA translation initiation in eukaryotes. *Biochimie* (2023) doi:10.1016/j.biochi.2023.09.017.

11. Sonenberg, N., Rupprecht, K. M., Hecht, S. M. & Shatkin, A. J. Eukaryotic mRNA cap binding protein: purification by affinity chromatography on sepharose-coupled m7GDP. *Proceedings of the National Academy of Sciences* **76**, 4345–4349 (1979).

12. Haghighat, A. & Sonenberg, N. eIF4G Dramatically Enhances the Binding of eIF4E to the mRNA 5′-Cap Structure *. *Journal of Biological Chemistry* **272**, 21677–21680 (1997).

13. Rozen, F. *et al.* Bidirectional RNA Helicase Activity of Eucaryotic Translation Initiation Factors 4A and 4F. *Molecular and Cellular Biology* **10**, 1134–1144 (1990).

14. Rogers, G. W., Richter, N. J., Lima, W. F. & Merrick, W. C. Modulation of the Helicase Activity of eIF4A by eIF4B, eIF4H, and eIF4F *. *Journal of Biological Chemistry* **276**, 30914–30922 (2001).

15. Schütz, P. *et al.* Crystal structure of the yeast eIF4A-eIF4G complex: An RNA-helicase controlled by protein–protein interactions. *Proceedings of the National Academy of Sciences* **105**, 9564–9569 (2008).

16. De Nijs, Y., De Maeseneire, S. L. & Soetaert, W. K. 5′ untranslated regions: the next regulatory sequence in yeast synthetic biology. *Biological Reviews* **95**, 517–529 (2020).

17. Carvalho, M. D., Carvalho, J. F. & Merrick, W. C. Biological characterization of various forms of elongation factor 1 from rabbit reticulocytes. *Arch Biochem Biophys* **234**, 603–611 (1984).

18. Dever, T. E. & Green, R. The Elongation, Termination, and Recycling Phases of Translation in Eukaryotes. *Cold Spring Harb Perspect Biol* **4**, a013706 (2012).

19. Beringer, M. & Rodnina, M. V. The ribosomal peptidyl transferase. *Mol Cell* **26**, 311–321 (2007).

20. Schuller, A. P. & Green, R. Roadblocks and resolutions in eukaryotic translation. *Nat Rev Mol Cell Biol* **19**, 526–541 (2018).

21. Moazed, D. & Noller, H. F. Intermediate states in the movement of transfer RNA in the ribosome. *Nature* **342**, 142–148 (1989).

22. Spahn, C. M. T. *et al.* Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. *EMBO J* **23**, 1008–1019 (2004).

23. Andersen, C. B. F. *et al.* Structure of eEF3 and the mechanism of transfer RNA release from the E-site. *Nature* **443**, 663–668 (2006).

24. Brown, A., Shao, S., Murray, J., Hegde, R. S. & Ramakrishnan, V. Structural basis for stop codon recognition in eukaryotes. *Nature* **524**, 493–496 (2015).

25. Frolova, L. *et al.* A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor. *Nature* **372**, 701–703 (1994).

26. Frolova, L. Y. *et al.* Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *RNA* **5**, 1014–1020 (1999).

27. Shoemaker, C. J. & Green, R. Kinetic analysis reveals the ordered coupling of translation termination and ribosome recycling in yeast. *PNAS* **108**, E1392–E1398 (2011).

28. Preis, A. *et al.* Cryoelectron Microscopic Structures of Eukaryotic Translation Termination Complexes Containing eRF1-eRF3 or eRF1-ABCE1. *Cell Rep* **8**, 59–65 (2014).

29. Mignone, F., Gissi, C., Liuni, S. & Pesole, G. Untranslated regions of mRNAs. *Genome Biol* **3**, reviews0004.1–reviews0004.10 (2002).

30. O'Brien, J., Hayder, H., Zayed, Y. & Peng, C. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Frontiers in Endocrinology* **9**, (2018).

31. Tuller, T., Ruppin, E. & Kupiec, M. Properties of untranslated regions of the S. cerevisiae genome. *BMC Genomics* **10**, 391 (2009).

32. Chen, C.-Y. A. & Shyu, A.-B. Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip Rev RNA* **2**, 167–183 (2011).

33. Tucker, M. *et al.* The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in Saccharomyces cerevisiae. *Cell* **104**, 377–386 (2001).

34. Muhlrad, D., Decker, C. J. & Parker, R. Deadenylation of the unstable mRNA encoded by the yeast MFA2 gene leads to decapping followed by 5'-->3' digestion of the transcript. *Genes Dev.* **8**, 855–866 (1994).

35. Anderson, J. S. & Parker, R. P. The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. *EMBO J* **17**, 1497–1506 (1998).

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

37. Nagy, E. & Maquat, L. E. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci* **23**, 198–199 (1998).

38. Kashima, I. *et al.* Binding of a novel SMG-1–Upf1–eRF1–eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. *Genes Dev* **20**, 355–367 (2006).

39. Frischmeyer, P. A. *et al.* An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science* **295**, 2258–2261 (2002).

40. Ito-Harashima, S., Kuroha, K., Tatematsu, T. & Inada, T. Translation of the poly(A) tail plays crucial roles in nonstop mRNA surveillance via translation repression and protein destabilization by proteasome in yeast. *Genes Dev.* **21**, 519–524 (2007).

41. van Hoof, A., Frischmeyer, P. A., Dietz, H. C. & Parker, R. Exosome-Mediated Recognition and Degradation of mRNAs Lacking a Termination Codon. *Science* **295**, 2262–2264 (2002).

42. D'Orazio, K. N. *et al.* The endonuclease Cue2 cleaves mRNAs at stalled ribosomes during No Go Decay. *eLife* **8**, e49117 (2019).

43. D'Orazio, K. N. & Green, R. Ribosome states signal RNA quality control. *Molecular Cell* **81**, 1372–1383 (2021).

44. Koutmou, K. S. *et al.* Ribosomes slide on lysine-encoding homopolymeric A stretches. *eLife* **4**, e05534 (2015).

45. Arthur, L. L. *et al.* Translational control by lysine-encoding A-rich sequences. *Sci Adv* **1**, e1500154 (2015).

46. Chandrasekaran, V. *et al.* Mechanism of ribosome stalling during translation of a poly(A) tail. *Nat Struct Mol Biol* **26**, 1132–1140 (2019).

47. Tesina, P. *et al.* Molecular mechanism of translational stalling by inhibitory codon combinations and poly(A) tracts. *The EMBO Journal* **39**, e103365 (2020).

48. Doma, M. K. & Parker, R. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature* **440**, 561–564 (2006).

49. Tsuboi, T. *et al.* Dom34:Hbs1 Plays a General Role in Quality-Control Systems by Dissociation of a Stalled Ribosome at the 3′ End of Aberrant mRNA. *Molecular Cell* **46**, 518–529 (2012).

50. Letzring, D. P., Dean, K. M. & Grayhack, E. J. Control of translation efficiency in yeast by codon–anticodon interactions. *RNA* **16**, 2516–2528 (2010).

51. Ikeuchi, K. *et al.* Collided ribosomes form a unique structural interface to induce Hel2-driven quality control pathways. *The EMBO Journal* **38**, e100276 (2019).

52. Joazeiro, C. A. P. Mechanisms and functions of ribosome-associated protein quality control. *Nature Reviews Molecular Cell Biology* **20**, 368–383 (2019).

53. Mizuno, M. *et al.* The nascent polypeptide in the 60S subunit determines the Rqc2-dependency of ribosomal quality control. *Nucleic Acids Research* (2021) doi:10.1093/nar/gkab005.

54. Reis, M. dos, Savva, R. & Wernisch, L. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Research* **32**, 5036–5044 (2004).

55. Presnyak, V. *et al.* Codon Optimality Is a Major Determinant of mRNA Stability. *Cell* **160**, 1111–1124 (2015).

56. Buschauer, R. *et al.* The Ccr4-Not complex monitors the translating ribosome for codon optimality. *Science* **368**, (2020).

57. Letzring, D. P., Wolf, A. S., Brule, C. E. & Grayhack, E. J. Translation of CGA codon repeats in yeast involves quality control components and ribosomal protein L1. *RNA* **19**, 1208–1217 (2013).

58. Lu, J. & Deutsch, C. Electrostatics in the Ribosomal Tunnel Modulate Chain Elongation Rates. *Journal of Molecular Biology* **384**, 73–86 (2008).

59. Dimitrova, L. N., Kuroha, K., Tatematsu, T. & Inada, T. Nascent Peptide-dependent Translation Arrest Leads to Not4p-mediated Protein Degradation by the Proteasome *. *Journal of Biological Chemistry* **284**, 10343–10352 (2009).

60. Brandman, O. *et al.* A Ribosome-Bound Quality Control Complex Triggers Degradation of Nascent Peptides and Signals Translation Stress. *Cell* **151**, 1042–1054 (2012).

61. Arthur, L. L. & Djuranovic, S. PolyA tracks, polybasic peptides, poly-translational hurdles. *Wiley Interdiscip Rev RNA* e1486 (2018) doi:10.1002/wrna.1486.

62. Guydosh, N. R. & Green, R. Dom34 rescues ribosomes in 3' untranslated regions. *Cell* **156**, 950–962 (2014).

63. Matsuo, Y. *et al.* RQT complex dissociates ribosomes collided on endogenous RQC substrate SDD1. *Nature Structural & Molecular Biology* **27**, 323–332 (2020).

64. Hoof, A. van & Wagner, E. J. A brief survey of mRNA surveillance. *Trends in Biochemical Sciences* **36**, 585–592 (2011).

65. Shoemaker, C. J. & Green, R. Translation drives mRNA quality control. *Nat Struct Mol Biol* **19**, 594–601 (2012).

66. Forrest, M. E. *et al.* Codon and amino acid content are associated with mRNA stability in mammalian cells. *PLOS ONE* **15**, e0228730 (2020).

67. Simms, C. L., Thomas, E. N. & Zaher, H. S. Ribosome-based quality control of mRNA and nascent peptides. *Wiley Interdiscip Rev RNA* **8**, 10.1002/wrna.1366 (2017).

68. Meydan, S. & Guydosh, N. R. A cellular handbook for collided ribosomes: surveillance pathways and collision types. *Curr Genet* **67**, 19–26 (2021).

69. Simms, C. L., Yan, L. L. & Zaher, H. S. Ribosome Collision Is Critical for Quality Control during No-Go Decay. *Molecular Cell* **68**, 361–373.e5 (2017).

70. Guydosh, N. R. & Green, R. Translation of poly(A) tails leads to precise mRNA cleavage. *RNA* **23**, 749–761 (2017).

71. Park, H. & Subramaniam, A. R. Inverted translational control of eukaryotic gene expression by ribosome collisions. *PLOS Biology* **17**, e3000396 (2019).

72. Shu, H. *et al.* FMRP links optimal codons to mRNA stability in neurons. *Proc Natl Acad Sci U S A* **117**, 30400–30411 (2020).

73. Martin, R. *et al.* De novo variants in CNOT3 cause a variable neurodevelopmental disorder. *Eur J Hum Genet* **27**, 1677–1682 (2019).

74. De Keersmaecker, K. *et al.* Exome sequencing identifies mutation in CNOT3 and ribosomal genes RPL5 and RPL10 in T-cell acute lymphoblastic leukemia. *Nat Genet* **45**, 186–190 (2013).

75. Yang, K. *et al.* The mammalian SKIV2L RNA exosome is essential for early B cell development. *Science Immunology* **7**, eabn2888 (2022).

76. Tuck, A. C. *et al.* Mammalian RNA Decay Pathways Are Highly Specialized and Widely Linked to Translation. *Molecular Cell* **77**, 1222–1236.e13 (2020).

77. Yang, K. *et al.* Cytoplasmic RNA quality control failure engages mTORC1-mediated autoinflammatory disease. *J Clin Invest* **132**, (2022).

78. Burke, P. C., Park, H. & Subramaniam, A. R. A nascent peptide code for translational control of mRNA stability in human cells. *Nat Commun* **13**, 6829 (2022).

79. Hanson, G., Alhusaini, N., Morris, N., Sweet, T. & Coller, J. Translation elongation and mRNA stability are coupled through the ribosomal A-site. *RNA* **24**, 1377–1389 (2018).

80. Absmeier, E. *et al.* Specific recognition and ubiquitination of translating ribosomes by mammalian CCR4–NOT. *Nat Struct Mol Biol* 1–9 (2023) doi:10.1038/s41594-023-01075-8.

81. Garzia, A. *et al.* The E3 ubiquitin ligase and RNA-binding protein ZNF598 orchestrates ribosome quality control of premature polyadenylated mRNAs. *Nat Commun* **8**, 16056 (2017).

82. Gutierrez, E. *et al.* eIF5A Promotes Translation of Polyproline Motifs. *Molecular Cell* **51**, 35–45 (2013).

83. Pavlov, M. Y. *et al.* Slow peptide bond formation by proline and other N-alkylamino acids in translation. *Proceedings of the National Academy of Sciences* **106**, 50–54 (2009).

84. Han, P. *et al.* Genome-wide Survey of Ribosome Collision. *Cell Reports* **31**, 107610 (2020).

85. Sabi, R. & Tuller, T. Computational analysis of nascent peptides that induce ribosome stalling and their proteomic distribution in Saccharomyces cerevisiae. *RNA* **23**, 983–994 (2017).

86. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *Cell* **147**, 789–802 (2011).

87. Meydan, S. & Guydosh, N. R. Disome and Trisome Profiling Reveal Genome-wide Targets of Ribosome Quality Control. *Molecular Cell* (2020) doi:10.1016/j.molcel.2020.06.010.

88. Yanagitani, K., Kimata, Y., Kadokura, H. & Kohno, K. Translational Pausing Ensures Membrane Targeting and Cytoplasmic Splicing of XBP1u mRNA. *Science* **331**, 586–589 (2011).

89. Nakatogawa, H. & Ito, K. The Ribosomal Exit Tunnel Functions as a Discriminating Gate. *Cell* **108**, 629–636 (2002).

90. Bhushan, S. *et al.* SecM-Stalled Ribosomes Adopt an Altered Geometry at the Peptidyl Transferase Center. *PLOS Biology* **9**, e1000581 (2011).

91. Shanmuganathan, V. *et al.* Structural and mutational analysis of the ribosome-arresting human XBP1u. *eLife* **8**, e46267 (2019).

92. Sitron, C. S. & Brandman, O. Detection and Degradation of Stalled Nascent Chains via Ribosome-Associated Quality Control. *Annual Review of Biochemistry* **89**, 417–442 (2020).

93. Brandman, O. & Hegde, R. S. Ribosome-associated protein quality control. *Nature Structural & Molecular Biology* **23**, 7–15 (2016).

94. Bengtson, M. H. & Joazeiro, C. A. P. Role of a ribosome-associated E3 ubiquitin ligase in protein quality control. *Nature* **467**, 470–473 (2010).

95. Sharp, P. M. & Li, W. H. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* **15**, 1281–1295 (1987).

96. Wallace, E. W. J., Airoldi, E. M. & Drummond, D. A. Estimating Selection on Synonymous Codon Usage from Noisy Experimental Data. *Mol Biol Evol* **30**, 1438–1453 (2013).

97. Pechmann, S. & Frydman, J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat Struct Mol Biol* **20**, 237–243 (2013).

98. Gamble, C. E., Brule, C. E., Dean, K. M., Fields, S. & Grayhack, E. J. Adjacent Codons Act in Concert to Modulate Translation Efficiency in Yeast. *Cell* **166**, 679–690 (2016).

99. Kuroha, K. *et al.* Receptor for activated C kinase 1 stimulates nascent polypeptide-dependent translation arrest. *EMBO reports* **11**, 956–961 (2010).

100. Sitron, C. S., Park, J. H. & Brandman, O. Asc1, Hel2, and Slh1 couple translation arrest to nascent chain degradation. *RNA* **23**, 798–810 (2017).

101. Ashe, M. P., De Long, S. K. & Sachs, A. B. Glucose Depletion Rapidly Inhibits Translation Initiation in Yeast. *Mol Biol Cell* **11**, 833–848 (2000).

102. Gancedo, J. M. The early steps of glucose signalling in yeast. *FEMS Microbiology Reviews* **32**, 673–704 (2008).

103. Teixeira, D., SHETH, U., VALENCIA-SANCHEZ, M. A., BRENGUES, M. & PARKER, R. Processing bodies require RNA for assembly and contain nontranslating mRNAs. *RNA* **11**, 371–382 (2005).

104. Anderson, P. & Kedersha, N. RNA granules. *Journal of Cell Biology* **172**, 803–808 (2006).

105. Sheth, U. & Parker, R. Targeting of Aberrant mRNAs to Cytoplasmic Processing Bodies. *Cell* **125**, 1095–1109 (2006).

106. Juszkiewicz, S. & Hegde, R. S. Initiation of Quality Control during Poly(A) Translation Requires Site-Specific Ribosome Ubiquitination. *Mol Cell* **65**, 743–750.e4 (2017).

107. Sundaramoorthy, E. *et al.* ZNF598 and RACK1 Regulate Mammalian Ribosome-Associated Quality Control Function by Mediating Regulatory 40S Ribosomal Ubiquitylation. *Mol Cell* **65**, 751–760.e4 (2017).

108. Matsuo, Y. *et al.* Ubiquitination of stalled ribosome triggers ribosome-associated quality control. *Nature Communications* **8**, 159 (2017).

109. Veltri, A. J. *et al.* Distinct elongation stalls during translation are linked with distinct pathways for mRNA degradation. *eLife* **11**, e76038 (2022).

110. Hickey, K. L. *et al.* GIGYF2 and 4EHP Inhibit Translation Initiation of Defective Messenger RNAs to Assist Ribosome-Associated Quality Control. *Molecular Cell* (2020) doi:10.1016/j.molcel.2020.07.007.

111. Juszkiewicz, S., Speldewinde, S. H., Wan, L., Svejstrup, J. Q. & Hegde, R. S. The ASC-1 Complex Disassembles Collided Ribosomes. *Molecular Cell* **79**, 603–614.e8 (2020).

112. MacArthur, M. W. & Thornton, J. M. Influence of proline residues on protein conformation. *J Mol Biol* **218**, 397–412 (1991).

113. Richardson, J. S. & Richardson, D. C. The de novo design of protein structures. *Trends in Biochemical Sciences* **14**, 304–309 (1989).

114. Li, S. C., Goto, N. K., Williams, K. A. & Deber, C. M. Alpha-helical, but not beta-sheet, propensity of proline is determined by peptide environment. *Proceedings of the National Academy of Sciences* **93**, 6676–6681 (1996).

115. Sinha, N. K. *et al.* EDF1 coordinates cellular responses to ribosome collisions. *eLife* **9**, e58828 (2020).

116. Barros, G. C. *et al.* Rqc1 and other yeast proteins containing highly positively charged sequences are not targets of the RQC complex. *Journal of Biological Chemistry* **296**, 100586 (2021).

117. Weinberg, D. E. *et al.* Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation. *Cell Rep* **14**, 1787–1799 (2016).

118. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols* **2**, 31–34 (2007).

119. Muller, R., Meacham, Z. A., Ferguson, L. & Ingolia, N. T. CiBER-seq dissects genetic networks by quantitative CRISPRi profiling of expression phenotypes. *Science* **370**, (2020).

120. Radhakrishnan, A. *et al.* The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation by Monitoring Codon Optimality. *Cell* **167**, 122–132.e9 (2016).

121. Hanson, G. & Coller, J. Codon optimality, bias and usage in translation and mRNA decay. *Nature Reviews Molecular Cell Biology* **19**, 20–30 (2018).

122. Hoekema, A., Kastelein, R. A., Vasser, M. & de Boer, H. A. Codon replacement in the PGK1 gene of Saccharomyces cerevisiae: experimental approach to study the role of biased codon usage in gene expression. *Mol Cell Biol* **7**, 2914–2924 (1987).

123. Sabi, R. & Tuller, T. A comparative genomics study on the effect of individual amino acids on ribosome stalling. *BMC Genomics* **16**, S5 (2015).

124. Artieri, C. G. & Fraser, H. B. Accounting for biases in riboprofiling data indicates a major role for proline in stalling translation. *Genome Res.* **24**, 2011–2021 (2014).

125. Peil, L. *et al.* Distinct XPPX sequence motifs induce ribosome stalling, which is rescued by the translation elongation factor EF-P. *Proc Natl Acad Sci U S A* **110**, 15265–15270 (2013).

126. Sheth, U. & Parker, R. Decapping and Decay of Messenger RNA Occur in Cytoplasmic Processing Bodies. *Science* **300**, 805–808 (2003).

127. Teixeira, D. & Parker, R. Analysis of P-Body Assembly in Saccharomyces cerevisiae. *MBoC* **18**, 2274–2287 (2007).

128 Gendron, T. F. *et al.* Poly(GP) proteins are a useful pharmacodynamic marker for C9ORF72-associated amyotrophic lateral sclerosis. *Sci Transl Med* **9**, eaai7866 (2017).

129 Al-Turki, T. M. & Griffith, J. D. Mammalian telomeric RNA (TERRA) can be translated to produce valine–arginine and glycine–leucine dipeptide repeat proteins. *Proceedings of the National Academy of Sciences* **120**, e2221529120 (2023).