Investigating the influence of codon composition and configuration on mRNA stability in *Saccharomyces cerevisiae*

A Thesis

submitted to

Indian Institute of Science Education and Research Pune in partial fulfilment of the requirements for the BS-MS Dual Degree Programme

by

Misaal Bedi

Indian Institute of Science Education and Research Pune

Dr. Homi Bhabha Road,

Pashan, Pune 411008, INDIA.

Date: October, 2023

Under the guidance of

Dr Attila Becskei

Biozentrum, University of Basel

Certificate

This is to certify that this dissertation entitled "Investigating the influence of codon composition and configuration on mRNA stability in *Saccharomyces cerevisiae*", submitted towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune, represents work carried out by Misaal Bedi at the Biozentrum, University of Basel, under the supervision of Dr Attila Becskei, Professor, Department Biozentrum, during the academic year January 2023 to October 2023.

Committee:

Dr. Attila Becskei

Professor

Biozentrum, University of Basel

Misaal Bedi

Dr. M.S. Madhusudhan

Professor

IISER, Pune

Declaration

I hereby declare that the matter embodied in the report entitled "Investigating the influence of codon composition and configuration on mRNA stability in *Saccharomyces cerevisiae*" are the results of the work carried out by me at the Biozentrum, University of Basel under the supervision of Dr. Attila Becskei and the same has not been submitted elsewhere for any other degree.

Dr. Attila Becskei

Professor

Biozentrum, University of Basel

Misaal Bedi

Date: October 19, 2023

Contents

1 List of Figures

2 List of Tables

3 Acknowledgements

I am immensely grateful to Dr. Attila Becskei for the incredible opportunity to conduct research in his lab and for his guidance throughout the past year. I am deeply thankful to the lab members, Nicole, Samuel, Sayanur, Simone and Sylvia, for guiding and supporting me in the lab and outside of it. Thank you to the Biozentrum community for the invaluably stimulating and collaborative scientific environment.

Special thanks to Sylvia Voegeli for providing the plasmids used in this thesis and to Davide Bonfanti, whom I supervised over the summer, for his assistance in measuring the half-lives of constructs in the *ski7*Δ background. I would also like to acknowledge the wonderful members of the Flow Cytometry facility for their technical assistance.

Thank you to the friends I made and the members of the Borri for being so warm and welcoming during my stay in Basel. To my parents and my sister, thank you for your unwavering love and support; it means the world to me. Thank you to Srishti and Gaurav for being amazing friends who enriched my life and were always there for me and kept me going.

To everybody who was a part of my journey, I will be forever grateful for your help, support and generosity; this thesis would not be what it is without you.

4 Abstract

Ribosomes decode some codons termed ‘optimal’ with greater speed and efficiency than ‘non-optimal’ codons. The codon composition of mRNAs thus plays a vital role in regulating gene expression by modulating mRNA stability and translation efficiency. The codon stabilisation coefficient (CSC) measures the influence of codons on mRNA stability. Optimal codons have positive CSCs and stabilise mRNAs, while non-optimal codons have negative CSCs and destabilise mRNAs. mRNA stability and translation efficiency were thought to exhibit a negative linear correlation with the proportion of non-optimal codons in an mRNA. While this linear relationship can explain about half the variance in mRNA half-lives, non-linear dependencies between codon composition, mRNA stability, and translation efficiency are becoming increasingly evident. In this thesis, we characterise the role of codon configuration in modulating mRNA stability in *S. cerevisiae*. Using variants of a tandem EGFP-mScarlet reporter, we find that clustering non-optimal codons together destabilises mRNAs and reduces translation efficiency in most cases, regardless of the length or position of the cluster, independently of mRNA surveillance pathways. We show that short non-optimal codon stretches are only able to induce modest position-dependent effects on stability despite variations in codon composition. In contrast, long non-optimal clusters or short aberrant codon stretches, such as CGA codon repeats, can substantially alter mRNA stabilities based on their position. Not accounting for non-optimal codon clustering, we show that while the proportion of non-optimal codons in mRNAs can alter mRNA half-lives by as much as 7-fold, its influence on translation efficiency is much less dramatic (approximately 1.5-fold). While no particular position in the mRNA consistently stabilises or destabilises mRNAs upon inserting non-optimal codon clusters, placing such clusters near the start of genes increases translation efficiencies compared to when they are placed farther from the start codon. We extend our findings to endogenous genes in yeast, showing that heterogeneities in the CSC profiles of highly optimal genes can destabilise mRNAs by as much as 2-fold. Our results reveal that codon configuration modulates mRNA stability much more intricately than previously realised.

5 Introduction

Regulating gene expression is a fundamental and widespread strategy organisms employ for diverse functions such as cell differentiation and responding to external stimuli. Cells possess a vast repertoire of gene regulatory strategies capable of controlling virtually every step of gene expression, from transcription to final protein function (Alberts et al., 2002). Modulating messenger RNA (mRNA) stability is a critical post-transcriptional strategy used to influence steady-state transcript levels, determined by the ratio of mRNA synthesis and degradation rates. Regulating mRNA stability plays a vital role in antiviral defence mechanisms and maintaining gene expression fidelity through specialised degradation pathways targeting aberrant mRNAs (Coller and Parker, 2004). Altering degradation rates allows for fast initial responses to new environments and stress conditions, crucial for short-term cellular adaptation (Balagopal et al., 2012). All the literature presented in this thesis pertains to *Saccharomyces cerevisiae* and may not directly apply to other model systems unless explicitly mentioned.

mRNA stability, typically quantified by its half-life ($T_{1/2}$), is the time required for 50% of the existing mRNA molecules to be degraded and can be tuned by different pathways. Extensive studies of degradation pathways in *S. cerevisiae* revealed that most ‘normal’ mRNAs are primarily degraded in the cytoplasm by two major exonucleolytic pathways (Parker, 2012). Thus, the modifications commonly found in most eukaryotic mRNAs, particularly the 5’ cap (7mGpppN) and the 3’ poly(A) tail, play an essential role in determining transcript stability as they confer protection against exonucleases (Balagopal et al., 2012).

Degradation is initiated for both pathways by deadenylation (the rate-limiting step), which refers to the shortening of the poly(A) tail (Decker and Parker, 1993; Garneau et al., 2007; Parker, 2012). This is followed either directly by 3’ to 5’ degradation via the exosome complex or by decapping (removal of the 5’ cap) and subsequent 5’ to

3’ co-translational degradation by Xrn1. Despite their functional redundancy, the decapping-mediated pathway predominates in yeast[1] (Hu et al., 2009; Parker, 2012).

Although most yeast mRNAs are targeted by a limited number of decay pathways, mRNA half-lives display substantial variation, spanning an order of magnitude or more (Parker, 2012).

The question of what factors influence mRNA stability has been a longstanding and extensively studied issue in gene regulation. While particular features in the untranslated regions (UTRs) have been linked to mRNA stability (Garneau et al., 2007; Geisberg et al., 2014), their influence is mediated by transcript-specific regulatory proteins. As such, these mechanisms cannot fully explain the range of half-lives spanned by the yeast transcriptome.

Hence, it is plausible that there are broader, more general features influencing mRNA stability. One potential clue may lie in the fundamental similarity among all mRNAs: their codon-based sequence.

5.1 Understanding codon decoding variability through codon bias and optimality

The mRNA sequence comprises three-letter codons decoded by the ribosome into amino acids to form proteins. The genetic code's degeneracy results in 61 codons coding for the 20 canonical amino acids, which allows an amino acid to be encoded by multiple ‘synonymous’ codons.

Over 50 years ago, it was observed that synonymous codons were used with non-random frequencies in genes (Goel et al., 1972). This non-uniformity in codon usage frequencies was termed ‘codon usage bias’ (CUB) (Parvathy et al., 2022). Numerous hypotheses have been proposed to explain the role of CUB, which have been extensively discussed in other excellent reviews (Plotkin and Kudla, 2011). In the context of this thesis, however, we limit ourselves to discussing the role of codon usage with regard to mRNA stability and translation.

[1] Some mRNAs take less common routes to degradation. These involve pathways initiated by deadenylation-independent decapping, and by endonucleolytic cleavage. The interested reader can find these non-canonical degradation pathways reviewed by (Garneau et al., 2007).

CUB is most evident in highly expressed genes, which is attributed to selection acting on codon usage aimed at optimising translational efficiency and fidelity (Hanson and Coller, 2018, 2018; Hia and Takeuchi, 2021; Sharp and Li, 1987). A related concept, codon optimality refers to the efficiency with which codons are translated, arising from the balance between tRNA supply and demand by ribosomes (Carneiro et al., 2019; Hia and Takeuchi, 2021). The efficiency of decoding codons, which depends on the tRNA concentration and the codon-anticodon pairing strength, amongst other factors, allows some codons, termed 'optimal', to be decoded faster and more efficiently compared to 'non-optimal' codons. Thus, the codon composition of a transcript plays a crucial role in determining its translational speed (Carneiro et al., 2019; Hia and Takeuchi, 2021; Presnyak et al., 2015). Other factors like sequence context (the identity of the codons or amino acids surrounding the codon being decoded), mRNA structure, peptide composition in the ribosome exit tunnel, etc., can also impact translation elongation rates (Passmore and Coller, 2022). Several metrics have been developed to characterise the influence of codon composition on translation and mRNA decay[2]. We discuss two such metrics: the tRNA adaptation index (tAI) and the codon stabilisation coefficient (CSC).

5.2 The tRNA Adaptation Index (tAI) characterises the optimality of codons

The tRNA adaptation index (tAI) is directly applicable to the study of translation efficiency and protein synthesis as it considers the influence of tRNA abundance in the cell on translation, which is thought to be a major factor in influencing translational efficiency (Bahiri-Elitzur and Tuller, 2021; Reis et al., 2004). Highly expressed genes in yeast use optimal codons based on the levels of the major isoacceptor tRNAs in the cell and display a high degree of CUB. This gave rise to the idea that codons may be adapted to the cellular tRNA pool, leading to the development of the tAI. The metric uses the fact that in some genomes, like that of *S. cerevisiae,* the copy number of genes encoding specific tRNAs positively correlates with tRNA concentrations in the cell (Reis et al., 2004). By accounting for the gene

[2] A wide array of metrics exist to study codon bias and optimality, which goes beyond the scope of this thesis. For readers seeking a more comprehensive discussion of the topic, (Bahiri-Elitzur and Tuller, 2021) offer an extensive review.

copy number of cognate tRNAs and codon-anticodon pairing efficiencies (accounting for wobble interactions), the tAI broadly assesses the translational efficiency of genes based on their adaptation to the intracellular tRNA pool (Hia and Takeuchi, 2021; Reis et al., 2004). The tAI can thus be considered a measure of tRNA usage efficiency by the ribosome (Carneiro et al., 2019). Based on the tAI, (Pechmann and Frydman, 2013) proposed a criteria for classifying codons as optimal or non-optimal. According to their system, codons with a tAI > 0.47 are classified as optimal, while those with lower values are considered non-optimal. Their criteria also accounts for CUB in determining the optimality of a codon.

5.3 The relationship between codon composition and mRNA stability

The relationship between the codon composition of mRNAs and their stabilities was initially explored in a study by (Herrick et al., 1990), which showed that unstable mRNAs had a significantly higher proportion of non-optimal codons (or so-called 'rare' codons – codons with low abundance cognate tRNAs) as compared to stable mRNAs.

A study by (Presnyak et al., 2015) further established the influence of codon optimality on mRNA stability in *S. cerevisiae*. By analysing transcriptome-wide mRNA stability data, they observed that some codons were preferentially enriched in either stable ($T_{1/2}$ > 20 minutes) or unstable mRNAs ($T_{1/2}$ < 5 minutes).

5.3.1 The Codon Stabilisation Coefficient (CSC)

To measure the contribution of individual codons on modulating mRNA stability, the authors devised the codon stabilisation coefficient (CSC). They computed the proportion of each codon within each mRNA of the transcriptome and calculated the Pearson correlation between these codon proportions and mRNA half-lives. The resulting R-value of this correlation was designated as the CSC for a given codon. An example calculation of the CSC values of the 'GGT' and 'AGG' codons is shown in Figure 1.

Codons with positive CSC values were designated as stable codons since their occurrence in transcripts positively correlated with mRNA half-lives. In contrast, codons with negative CSC values were termed unstable, indicating their negative

correlation with half-lives. Remarkably, the classification of codons as stable or unstable closely paralleled their assignments as optimal or non-optimal based on the tAI (Figure 2A).

Given the high correlation between the tAI and CSC metrics and the consistent classification of codons as optimal/stable or non-optimal/unstable, we use these terms interchangeably throughout this thesis.

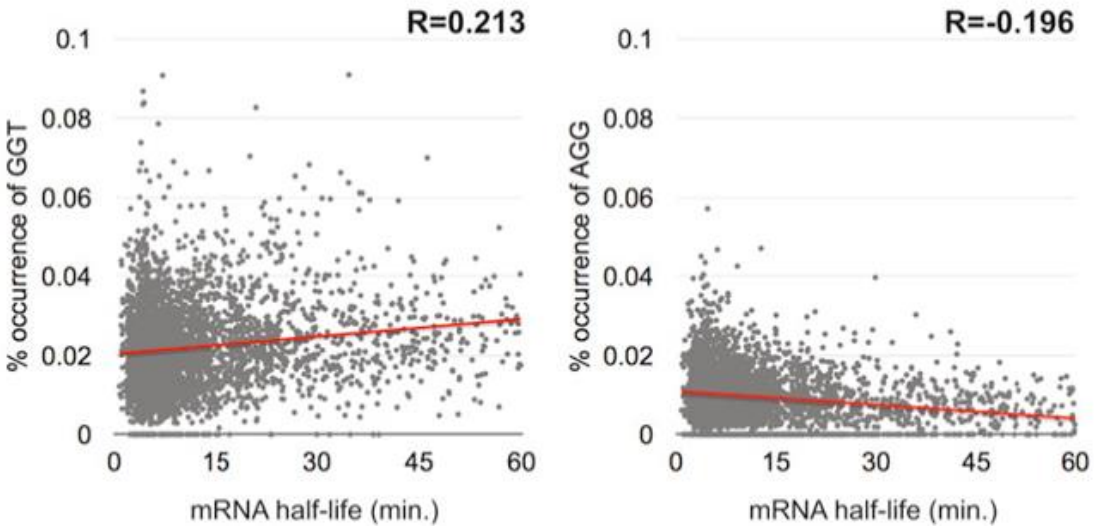

Figure 1: Calculation of CSC values for specific codons

Example CSC calculations for the 'GGT' and 'AGG' codons. Figure adapted from (Presnyak et al., 2015).

5.3.2 The influence of codon optimality on mRNA stability

The correlation between codon occurrence and mRNA half-lives may be influenced by specific codons with a substantial impact on mRNA stability. For example, CGA codon repeats have been identified as potent translational inhibitors that also destabilise mRNAs (Veltri et al., 2022). However, (Presnyak et al., 2015) observed that rather than being enriched in any particular codons, the overall codon composition of an mRNA determined its stability. Specifically, the percentage of optimal codons positively correlated with mRNA stability (Figure 2B). This was experimentally validated by substituting the codons of a stable endogenous mRNA

with synonymous non-optimal codons, which resulted in a reduction of its half-life. Conversely, substitution with optimal codons in an unstable mRNA led to an increase in its half-life. Importantly, this general trend was shown to persist in different UTR contexts.

The authors determined that codon optimality influences mRNA stability via translation. They observed a similar ribosome density on synonymous variants of the *HIS3* gene despite differences in protein output. This suggested a decrease in ribosome translocation rates for variants composed of non-optimal codons, which was validated by ribosome run-off experiments through polysome profiling. Polysome profiling, a technique based on differential sedimentation using a sucrose density gradient, was used to measure ribosome density. This method fractionates mRNAs based on their weight, distinguishing between transcripts not linked to the translation machinery (found in the ribonucleoprotein (RNP) complex fraction), transcripts associated with a single ribosome (found in the 80S fraction) and transcripts associated with multiple ribosomes, termed 'polysomes' (found in the polysome fraction) (Hanson and Coller, 2018).

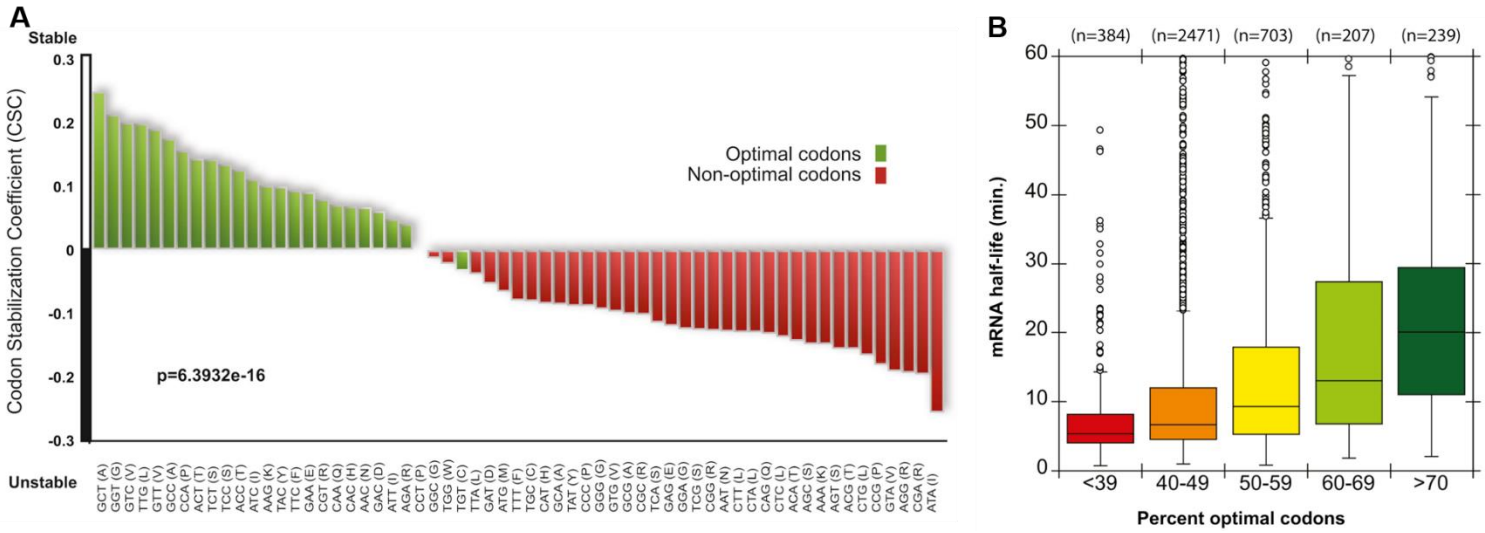

Figure 2: The relationship between codon composition and mRNA stability

*(**A**) The CSC values and optimality assignments of codons superimposed on each other. (**B**) Box plot depicting the relationship between the percentage of optimal codons in transcripts and their half-lives in S. cerevisiae. Figure adapted from (Presnyak et al., 2015).*

5.3.3 The Codon Stabilisation Coefficient (CSC) consistently quantifies codon contributions to mRNA stability

Importantly, (Presnyak et al., 2015) ruled out the possibility that the CSC was sensing other features of mRNAs that may influence mRNA stability, such as their GC content, by introducing +1 and +2 frameshifts in open reading frames (ORFs) *in silico,* which abrogated the correlation between the CSC and tAI.

The CSC also correlates well with other metrics of codon optimality and translation efficiency, such as the inverse of the ribosome density at the A site (a proxy for ribosome residence time on codons), as determined by ribosome footprinting/profiling (see section 9.2 for a description of ribosome profiling). Since ribosomes spend less time translating optimal codons, the ribosome occupancy on such codons is expected to be lower compared to non-optimal codons.

It has been noted that mRNA half-lives obtained by different methods differ in their absolute values and exhibit low correlations with each other. This inconsistency has led to instances where an mRNA could be considered stable or unstable depending on the method used. This presents a substantial challenge in identifying patterns in mRNA sequences that contribute to mRNA stability (Baudrimont et al., 2017; Carneiro et al., 2019). As shown by (Carneiro et al., 2019), the CSC consistently measures a codon's contribution to mRNA stability irrespective of the specific method used to measure mRNA half-lives and can be used to evaluate similar half-life datasets.

The mean CSC value of all the codons in an mRNA (μCSC) is a reliable indicator of mRNA stability. The μCSC provides a straightforward measure to compare genes based on their stability (Carneiro et al., 2019). This thesis uses the μCSC as a coarse-grained measure of a transcript's stability. Transcripts with high μCSCs are generally expected to be more stable than those with low μCSCs. However, it is important to note that this may not account for the non-linear effects of sequence composition on mRNA stability.

5.4 The relationship between codon configuration and mRNA stability

While the codon composition of an mRNA is a critical determinant of its stability, it accounts for approximately 50% of the variance in mRNA half-lives (Cheng et al., 2017). There is growing recognition of the various non-linear dependencies between codon composition and mRNA stability, involving factors such as the local sequence context surrounding the codon being decoded, the decoding of repetitive sequences and translation speed (Jaquet et al., 2022). Another such feature is the arrangement of codons along the mRNA.

A study by (Radhakrishnan et al., 2016) probed the relationship between codon configuration and mRNA stability by introducing a stretch of ten non-optimal codons in the highly stable yeast gene *PGK1* at different positions in the ORF (Figure 3A).

Strikingly, they observed a polarity in mRNA half-lives dependent on the distance of the non-optimal codon stretch from the start codon. Specifically, the *PGK1* variants displayed an inverse relationship between their half-lives and the distance of the non-optimal codon stretch from the start codon (Figure 3B). The authors attributed the polarity of half-lives to the increased accumulation of slow-moving or paused ribosomes upstream of the non-optimal codon stretch as it was placed farther from the start codon. Critically, this phenomenon was independent of well-established mRNA surveillance and ribosome-associated quality control pathways, as evidenced by the persistence of the polarity in various deletion backgrounds involved in these pathways. *DHH1P*, a known decapping regulator also involved in translational repression, was implicated in driving the polarity effect, which was eliminated in a *dhh1*∆ background.

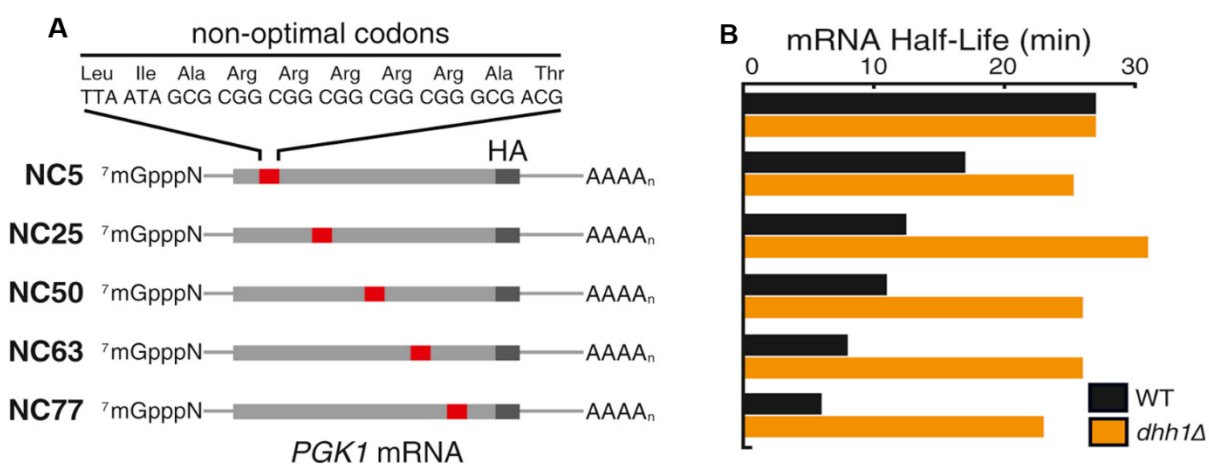

Figure 3: The influence of codon configuration on *PGK1* stability

*(**A**) PGK1 variants featuring a ten-codon non-optimal stretch at various locations within the ORF. (**B**) Half-lives of the PGK1 variants depicted in (**A**) in wild-type (in black) and dhh1Δ (in orange) cells. Figure adapted from (Radhakrishnan et al., 2016).*

The polarity of half-lives in *PGK1* observed by (Radhakrishnan et al., 2016) was replicated by (Sharma et al., 2021*)* using simulations of mRNA translation and degradation.

However, a bioinformatics analysis by (Carneiro et al., 2019) could not extend the findings of (Radhakrishnan et al., 2016) to the entire yeast transcriptome. The authors found that while ten-codon non-optimal stretches were concentrated near the start codon (Figure 4A), there was no correlation between the position of these stretches and mRNA half-lives (Figure 4B). The authors reasoned that since *PGK1* is a highly stable gene, such a non-optimal codon stretch would present as an aberration since most genes in *S. cerevisiae* are non-optimal with negative µCSC values. Thus, the observations from *PGK1* variants may not reflect the behaviour of most endogenous genes.

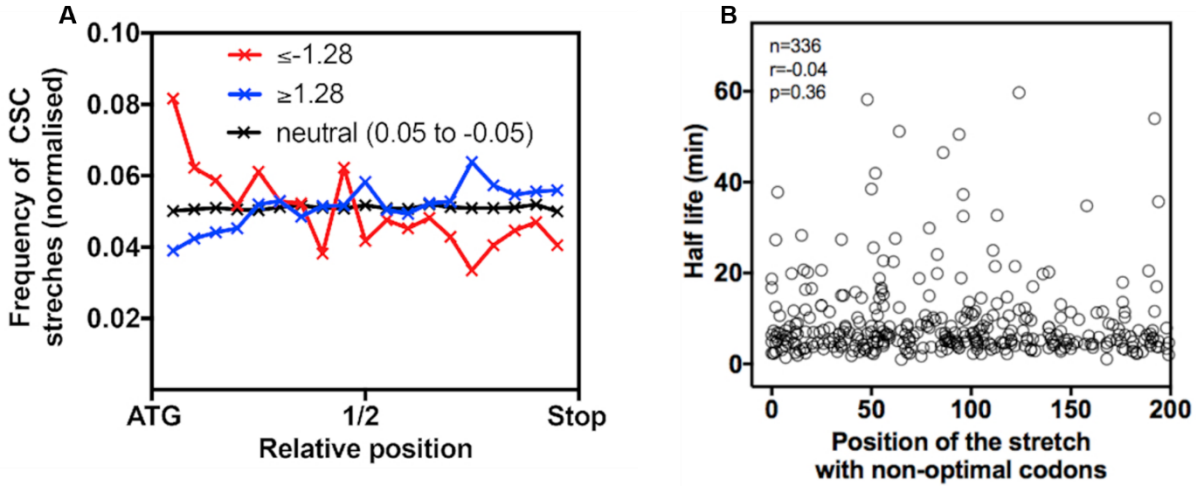

Figure 4: The relationship between the position of non-optimal codon stretches and mRNA half-lives across the *S. cerevisiae* transcriptome

*(**A**) The frequency distribution of ten-codon non-optimal stretches with the same sum of CSC values (-1.28) as the ten-codon stretch used by (Radhakrishnan et al., 2016) in Figure 3A. Neutral stretches with sum CSC values between -0.05 and 0.05 and optimal stretches with sum CSCs > 1.28 were used as controls. (**B**) The mRNA half-lives of genes with a single non-optimal ten-codon stretch whose sum of CSC values was less than 1.28 plotted against the position of the stretch. Figure adapted from (Carneiro et al., 2019).*

5.4.1 Multiple non-optimal codon stretches modulate the relationship between codon configuration and mRNA stability

(Sharma et al., 2021) investigated the influence of two non-optimal codon stretches on mRNA stability, considering that 78% of transcripts in *S. cerevisiae* contain multiple such stretches.

The authors generated synonymous variants of three endogenous *S. cerevisiae* transcripts with high initiation rates. They found two stretches of eight optimal codons in the genes. In the first variant, the stretch near the 3’ end was substituted with synonymous non-optimal codons, while in the second variant, both stretches were substituted. For all genes, the latter variant was more stable. A similar observation was made in two endogenous genes with median initiation rates.

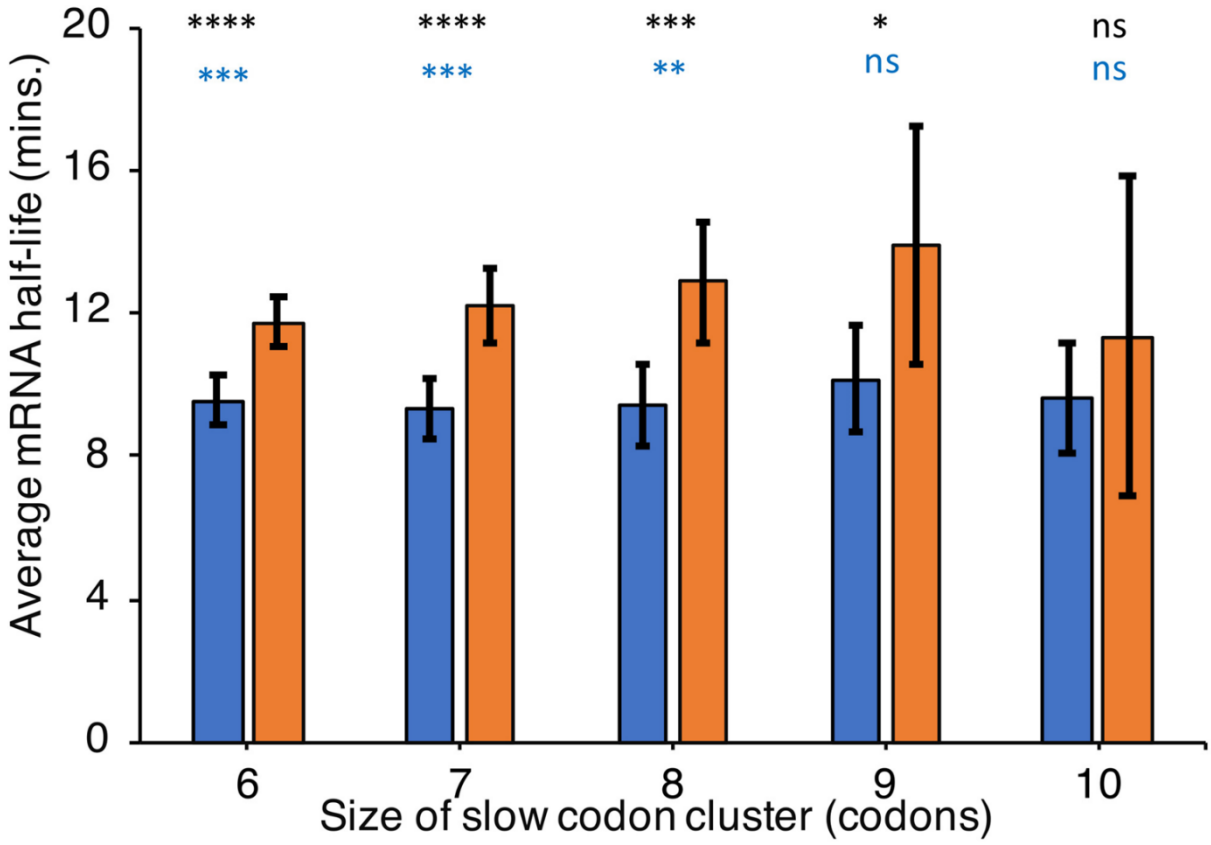

Figure 5: mRNA half-lives of endogenous *S. cerevisiae* transcripts with multiple versus a single non-optimal codon stretch

The average mRNA half-lives of transcripts with a non-optimal codon stretch in the first and last 10% of the transcript (orange bars), compared to those with a non-optimal codon stretch only in the last 10% of the transcript (blue bars). Figure adapted from (Sharma et al., 2021).

These results were examined across the yeast transcriptome using the half-life dataset generated by (Presnyak et al., 2015). Genes were divided into two groups: those with a non-optimal codon stretch in the first and last 10 or 20% of the transcript, and those with a non-optimal codon stretch only in the last 10 or 20% of the transcript. For non-optimal codon stretches of sizes 6-8, a significant increase in average mRNA half-lives was observed in the first group compared to the second for both the 10% and 20% cases (Figure 5). This finding was attributed to differences in the lengths of ribosome queues that could form upstream of the non-optimal codon stretches. The authors postulated that the 5' end stretch reduces the rate of ribosome entry into the mRNA body, thereby reducing the formation of traffic jams in the region between the two stretches. This could facilitate the existence of downstream non-optimal codon stretches, allowing them to maintain their biological benefits, such as

aiding in co-translational protein folding while mitigating their impact on mRNA stability.

5.5 mRNA surveillance pathways targeting aberrant mRNAs

The coupling of codon optimality to mRNA stability is primarily thought to operate through translation (Hia and Takeuchi, 2021; Monaghan et al., 2023; Presnyak et al., 2015; Radhakrishnan et al., 2016; Shoemaker and Green, 2012).

Ribosomes decode non-optimal codons more slowly than optimal ones, leading to ribosomal 'pausing' during translation. While such pauses are thought to play critical roles in regulating protein folding, localization and activity (Hia and Takeuchi, 2021; Pechmann and Frydman, 2013), prolonged pausing can be detrimental to the cell. Ribosomes sequestered on transcripts, engaged in unproductive translation, deplete the free ribosome pool, which in turn may affect the initiation rates of other transcripts (Raveh et al., 2016). Such situations may be exacerbated by ribosome collisions, where fast-translating ribosomes can collide with downstream ribosomes stuck on translationally challenging sequences.

Furthermore, proteotoxic stress due to the accumulation of misfolded proteins and the harmful effects of erroneously translated mRNAs presents a significant threat to cellular homeostasis, and a considerable amount of regulation exists at the codon level to balance the need for accurate protein folding with fast elongation speeds (Hia and Takeuchi, 2021).

This presents a crucial challenge for cells: distinguishing between detrimental and beneficial ribosomal pausing. One solution to this problem may lie in the use of mRNA surveillance pathways that actively monitor and degrade mRNAs with translational aberrations. An overview of these surveillance pathways is shown in Figure 6.

5.5.1 Nonsense-mediated decay (NMD)

This pathway targets mRNAs with premature stop codons. NMD targets are typically recognized by the absence of molecular factors that usually associate downstream of the stop codon, the presence of an extended 3' UTR or inefficient premature

translation termination upstream of the poly(A) (Peccarelli and Kebaara, 2014; Shoemaker and Green, 2012). In *S. cerevisiae,* three main proteins are involved in the functioning of this pathway: Upf1, Upf2, and Upf3. Inactivation of any of these factors rescues the mRNA stability of NMD targets. Upf1, in its phosphorylated form, serves as the central regulator in the pathway, while Upf2 and Upf3 regulate the function of Upf1p (Peccarelli and Kebaara, 2014). Numerous sequence and structural features within an mRNA arising from its codon composition can trigger NMD during translation.

One example involves frameshifting, a phenomenon where ribosomes shift into a different reading frame, potentially giving rise to stop codons in the mRNA body. Ribosomal frameshifting signals are widespread in the *S. cerevisiae* genome, which induce degradation either by NMD or the no-go decay pathway (through pseudoknot formation). Additionally, translating a polylysine tract composed of AAA codons can lead to frameshifting and the induction of NMD, as it can cause ribosomal sliding. Interestingly, studies in yeast revealed that mRNAs targeted by NMD share several characteristics, such as reduced translation efficiency, lower average codon optimality, and a propensity to contain stretches of non-optimal codons. It is postulated that NMD targets these mRNAs due to their increased likelihood of making errors or frameshifting during translation (Embree et al., 2022).

5.5.2 No-go decay (NGD)

No-go decay (NGD) serves to resolve deleterious ribosome stalls[3]. NGD is believed to be triggered by stalled or slowly moving ribosomes, leading to collisions with trailing ribosomes. Thus, multiple ribosomes are typically required to initiate NGD (Simms et al., 2017). Damaged bases, secondary structures like stem-loops or pseudoknots, or the translation of specific codon sequences or peptides can all cause ribosome stalling (Shoemaker and Green, 2012).

Collisions have two major consequences: they trigger translational repression by preventing translation re-initiation on the mRNA, and they induce mRNA degradation via NGD, along with the degradation of the peptide via the ribosome-associated

[3] The kinetic details of ribosome pausing, and the differences between transient “pauses” and pauses that can induce NGD remain to be determined. We thus follow the convention used by (Shoemaker and Green, 2012), referring to pauses that are sufficient to induce NGD specifically as stalls.

quality control (RQC) pathway. NGD and RQC can occur independently or be coupled (Ikeuchi et al., 2019; Monaghan et al., 2023). When two ribosomes collide, they form a ‘disome’, the minimal collision unit required for NGD and RQC. The stalled leading ribosome blocks the path of trailing ribosomes, preventing them from translocating forward. The disome is characterised by a unique ‘rotated’ 40S-40S interface, recognized by Hel2 in yeast. Hel2 polyubiquitinates ribosomal proteins to initiate RQC on the leading ribosome (Ikeuchi et al., 2019). This ubiquitination signals Cue2, an endonuclease that cleaves the mRNA in different locations depending on whether NGD is coupled or uncoupled to RQC. However, it is hypothesised that Cue2-mediated endonucleolytic degradation represents a secondary pathway that may become active when the canonical degradation pathways are compromised or overwhelmed, especially during stress conditions (D’Orazio et al., 2019; Monaghan et al., 2023; Veltri et al., 2022). Ribosomes stalled at cleavage sites are dissociated into their constituent subunits, and mRNA fragments are degraded through the canonical Xrn1 and exosome complex mediated pathways. Peptide degradation is simultaneously initiated by the RQC trigger complex (RQT) (D’Orazio et al., 2019; Monaghan et al., 2023; Simms et al., 2017). The Syh1 and Smy2 proteins also partake in mRNA degradation and repress translation via a pathway complementary to Hel2-mediated NGD (Monaghan et al., 2023; Veltri et al., 2022).

Furthermore, ribosome collisions are thought to induce frameshifting when trailing ribosomes exert a pull on the mRNA, causing the leading ribosomes to slip over the A site codon. The likelihood of such frameshifting events increases upon the inactivation of NGD, indicating that the NGD pathway serves additional roles in mitigating the consequences of incorrectly translated mRNAs by degrading them (Simms et al., 2019). Surprisingly, while ribosome collisions are widespread in yeast, recognition of such collisions does not always induce NGD and RQC (Meydan and Guydosh, 2020).

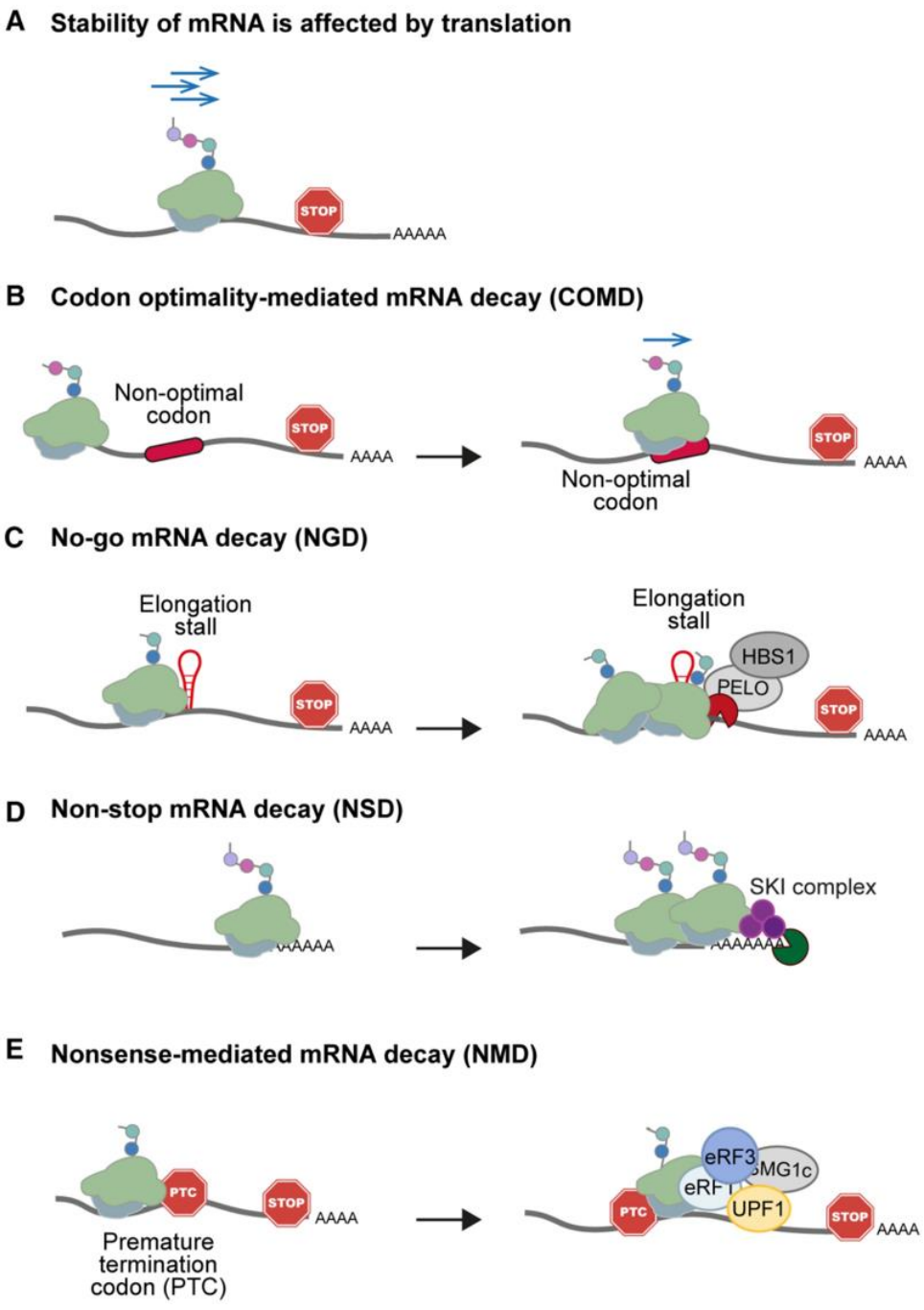

Figure 6: An overview of the major mRNA surveillance pathways in yeast

*(**A**) mRNA stability can be regulated by translation through various factors such as translation speed and mRNA secondary structure formation. (**B**) Ribosomes slow down or 'pause' on non-optimal codons, which can decrease mRNA stability. (**C**) Deleterious stalling events trigger no-go decay (NGD). (**D**) The absence of a stop codon triggers nonstop decay (NSD). (**E**) The presence of a premature stop codon in the mRNA body triggers nonsense-mediated decay (NMD). Figure from (Monaghan et al., 2023)*

5.5.3 Nonstop decay (NSD)

This pathway degrades mRNAs lacking a stop codon. Several conditions can trigger NSD, which ultimately involves the translation of a poly(A) stretch either due to a missing stop codon or the presence of a poly(A) stretch within the mRNA sequence (Monaghan et al., 2023). The exosome complex, which degrades mRNAs in a 3' to 5' direction, primarily degrades such transcripts. Ski7, a member of the SKI complex, recruits the exosome to NSD substrates. Translating sequences of positively charged amino acids can stall ribosomes since the ribosome exit tunnel is negatively charged (Monaghan et al., 2023; Shoemaker and Green, 2012). This complicates the distinction between NSD and NGD, as poly(A) translation can also trigger NGD. Moreover, degradation via NSD involves the same initial steps of ubiquitination and endonucleolytic cleavage as NGD, differing in the downstream degradation process. The SKI complex ultimately engages the exosome, coupling mRNA degradation and ribosome recycling (Monaghan et al., 2023).

5.6 General sensors of codon optimality

The coupling between translation and mRNA degradation is a tightly regulated process. Emerging evidence is uncovering an important aspect of this control: the crosstalk between deadenylation and decapping, converging at the ribosome. (Passmore and Coller, 2022).

The Ccr4-Not complex is the primary deadenylase in *S. cerevisiae,* containing two exonucleases: Ccr4 and Caf1 (Garneau et al., 2007). Canonically, decapping can only be initiated after deadenylation and release of the final poly(A)-binding protein from the mRNA. Pab1 is preferentially enriched/stabilised on fast-translating mRNAs. The two exonucleases have distinct substrates: while Ccr4 can shorten the poly(A) tail regardless of Pab1, Caf1 can only target naked poly(A). Consequently, in transcripts with high codon optimality and fast translation rates, the enrichment or stabilisation of Pab1 leads to a slower, Ccr4-dependent deadenylation. Moreover, highly-translated mRNAs feature shorter poly(A) tails, extensively protected by Pab1, leaving little room for Caf1 activity. In contrast, transcripts with slow translation rates, characterised by longer poly(A) tails and a higher amount of naked poly(A), undergo

rapid deadenylation due to the concerted action of both Ccr4 and Caf1, resulting in shorter mRNA half-lives (Passmore and Coller, 2022; Webster et al., 2018).

Dhh1p, involved in both translational repression and decapping activation, is thought to sense translational speed and couple it to degradation, evidenced by the preferential stabilisation of low codon optimality transcripts in *dhh1*Δ backgrounds. Dhh1p dynamically samples the mRNA, binding to slowly translating ribosomes and degrading the respective mRNAs (Radhakrishnan et al., 2016).

Recently, Not5, a subunit of the Ccr4-Not complex, was identified as a general sensor of codon optimality. Ribosomes decoding non-optimal codons experience prolonged delays in A site tRNA accommodation compared to optimal codons. Since the E-site tRNA can exit the ribosome independent of A-site tRNA accommodation, there is an increased likelihood of the E-site becoming empty before the A-site is filled when ribosomes decode non-optimal codons. Ribosomes in the post-translocation state with empty E and A sites adopt a specific conformation recognized by Not5. Thus, Not5 indirectly senses the optimality of the codon in the ribosome A-site. Once bound to the ribosome, Not5 promotes deadenylation and Dhh1-dependent decapping of the mRNA (Buschauer et al., 2020).

5.7 Distinct ribosome pauses trigger different degradation pathways

A recent study by (Veltri et al., 2022) highlighted the role of different types of non-optimal codons in triggering distinct degradation pathways. They introduced a $(CGA)_{12}$ stretch (known to cause ribosomal stalling and collisions) in the *HIS3* ORF in yeast, which dramatically destabilized the mRNA. The half-life of this construct was substantially rescued in a *syh1*Δ*hel2*Δ background. However, in a *not5*Δ background, a modest rescue was observed due to the presence of non-optimal codons. Furthermore, they recoded the *HIS3* ORF by replacing optimal codons with synonymous non-optimal codons. The half-life of this reporter was substantially rescued in a *not5*Δ background; however, the *syh1*Δ*hel2*Δ background failed to rescue its half-life. These results emphasize the role of the nature of the ribosomal pause in dictating the choice of the degradation pathway and, ultimately, the half-life of the mRNA.

Furthermore, a study by (Hou et al., 2023) demonstrated similar effects in reporters containing CGA repeats and non-optimally substituted leucine codons. The CGA reporter was targeted via NGD, while a non-Hel2-mediated mechanism targeted the non-optimal reporter. Importantly, their results revealed that non-optimal codons could trigger different degradation pathways despite stalling elongation for similar durations.

5.8 Measuring mRNA half-lives using Multiplexed Gene Control (MGC)

While mRNA stability can be measured indirectly through changes in steady-state transcript levels, this approach cannot accurately estimate mRNA half-lives since steady-state mRNA levels can change due to changes in transcription or other processes involved in responses to environmental stimuli (Chen et al., 2008). In an ideal scenario, where there is a complete shut-down of transcription *in vivo*, mRNA decay follows first-order kinetics, allowing half-lives to be estimated by fitting an exponential decay to a time series of mRNA levels (Chen et al., 2008; Ross, 1995).

Three main classes of methods are used to determine mRNA half-lives: regulatable promoter systems for transcriptional control, *in vivo* metabolic labelling and global transcriptional inhibition (Passos and Parker, 2008).

In metabolic labelling, transcripts are radiolabelled *in vivo* by brief exposure to radioactively labelled nucleic acid precursors, which can be incorporated into newly transcribed RNAs. This is followed by a ‘chase’, where unlabelled nucleotides wash out the labelled nucleotides. The half-life is determined by observing the decay of the radioactive signal from labelled mRNAs. This method is scalable for transcriptome-wide measurements, requires no genetically modified strains or specially cloned constructs and minimally perturbs cells. However, it suffers from a poor signal-to-noise ratio for low copy number mRNAs, requires a large amount of radioactive material, and cannot provide information on mRNA integrity (Herrick et al., 1990; Passos and Parker, 2008).

Global transcriptional inhibition can be achieved using drugs or genetically modified strains containing a temperature-sensitive allele of RNA polymerase II. Exposure to the drug or the restrictive temperature shuts down transcription genome-wide and

allows for half-life measurement by northern blotting or microarrays. The advantage of this method is its general applicability and information on mRNA integrity. However, using drugs or heat shock can alter the gene expression landscape of the cell and have other nonspecific effects that can affect mRNA half-lives (Passos and Parker, 2008).

Transcription can also be shut off for genes of interest by expressing them from regulatable promoters. A popular promoter choice is the GAL promoter, which can be induced by growing cells in media with galactose as the carbon source. Upon shifting cells to glucose-containing media, transcription is shut off, and mRNA decay can be followed in time. While this method does circumvent some of the issues associated with global transcriptional inhibition and causes minimal cell perturbation, it is not easily scalable to the entire transcriptome. Most importantly, however, a change in the carbon source might affect mRNA decay by affecting cell physiology in unanticipated ways (Passos and Parker, 2008).

The ‘Tet-OFF’ system can also be used to control gene transcription. The promoter contains *tet* operators, which are bound by the tetracycline transactivator protein (tTA), driving gene expression under normal conditions. Gene expression is shut off upon adding tetracycline or doxycycline as tTA dissociates from the operator sites (Garí et al., 1997; Passos and Parker, 2008). The main advantage of this system is that it leaves cell physiology unaltered.

To ensure minimal cell perturbation and precise transcriptional control of genes, we used the multiplexed gene control (MGC) method to measure mRNA half-lives (Baudrimont et al., 2017).

The method uses the Tet-OFF system by employing a synthetic promoter, a modified *GAL1* promoter containing four *tet* operator binding sites instead of the native Gal4p sites. To allow for multiplexing of half-life measurements of different mRNAs, individual cell cultures containing an mRNA of interest are pooled together in equal proportions. Subsequently, transcription is shut off by the addition of doxycycline, and samples are collected at defined time points to track mRNA levels over time. Importantly, doxycycline does not affect gene expression in most yeast strains. The key advantage of multiplexing lies in its substantial reduction of time required for measuring the half-lives of multiple mRNAs since the RNA extraction and reverse

transcription steps are performed on pooled strains (see section 7.10 in Materials and Methods for more details).

5.9 Aims of this thesis

The linear relationship between codon optimality and mRNA stability accounts for approximately 50% of the variance in mRNA half-lives (Cheng et al., 2017). Various factors, including codon configuration, can introduce non-linear dependencies between codon composition and mRNA stability. However, the general relationship between codon configuration and mRNA stability is unclear. While (Radhakrishnan et al., 2016) demonstrated an inverse relationship between the distance of a ten-codon non-optimal stretch from the start codon and mRNA stability (Figure 3B) in *PGK1*, this relationship could not be extended to the yeast transcriptome (Figure 4B) (Carneiro et al., 2019). Furthermore, multiple non-optimal codon stretches can modulate mRNA stability differently (Sharma et al., 2021) (Figure 5). The precise mechanism through which codon configuration may influence mRNA stability also remains to be determined.

This thesis explores the influence of codon composition and configuration within mRNAs on mRNA stability and translation efficiency. We investigate this using synthetic gene reporters featuring various proportions and arrangements of optimal and non-optimal codons. While previous studies have primarily focused on the effects of short, non-optimal codon stretches on mRNA stability, we aim to explore the impact of clustering many non-optimal codons on mRNA stability and translation efficiency. Additionally, we examine the characteristics of sequences that may induce polarity effects when introduced into optimal mRNAs. We seek to understand the mechanism by which codon arrangement can influence mRNA stability by analysing the reporters in various gene deletion backgrounds involved in mRNA surveillance and codon-optimality-mediated degradation pathways. Given that non-optimal codon stretches can hamper translation, these pathways may be relevant in determining the stabilities of such genes.

The configuration of optimal and non-optimal codon stretches is much more intricate in endogenous genes than the well-defined arrangements in the synthetic reporters. We thus intend to explore the influence of heterogeneities in the CSC profiles of

endogenous genes on their mRNA stabilities. This will allow us to assess the significance of codon configuration as a regulatory feature affecting mRNA stability in native gene sequences in yeast.

6 Results and Discussion

6.1 Overview of constructs

We designed four classes of constructs to systematically investigate the influence of codon composition and configuration on mRNA stability and translation efficiency, as outlined below.

6.1.1 Tandem fluorescent reporters

We used variants of a tandem fluorescent reporter composed of EGFP and mScarlet featuring various proportions and arrangements of optimal and non-optimal codons to study their influence on mRNA stability and translation efficiency.

6.1.1.1 Asymmetric EGFP-mScarlet constructs

Previous studies examining the relationship between codon configuration and mRNA stability analysed the effect of non-optimal codon stretches composed of ten or fewer codons (Carneiro et al., 2019; Radhakrishnan et al., 2016; Sharma et al., 2021). To explore the influence of longer non-optimal codon clusters, we designed the ‘asymmetric EGFP-mScarlet’ constructs. These constructs feature non-optimal codon clusters, ranging from approximately 3% to 75% of the sequence positioned at different locations (Figure 7A). For a more detailed description of these constructs, refer to section 6.2.1.

6.1.1.2 Polarity-assessment constructs

A study by (Radhakrishnan et al., 2016) revealed an inverse relationship between the distance of a ten-codon non-optimal stretch from the start codon and mRNA stability in the highly stable *PGK1* gene in yeast (Figure 3B). In alignment with their results, we use the term ‘polarity effects’ to describe a monotonic relationship between the distance of non-optimal codon stretches from the 5’ end of ORFs and mRNA stabilities. To investigate the types of sequences that may induce polarity effects, we designed constructs featuring non-optimal codon stretches with diverse codon compositions inserted at different locations within the ‘optimised’ EGFP-mScarlet reporter (Figure 13). The optimised gene comprised codons with the maximum CSC value available for each amino acid, with a µCSC similar to *PGK1*. For a more detailed description of these constructs, refer to section 6.4.1.

6.1.2 Truncated fluorescent reporters

The ‘ramp’ is a region of slow translation at the beginning of mRNAs, influenced partly by the codon composition in this region (Weinberg et al., 2016). Interestingly, the ramp length coincides with a recently identified gene length threshold (Rahaman et al., 2023). The mRNA stability of genes whose length is below this threshold is unaffected by codon composition. To study the interaction between the ramp and threshold lengths, we designed truncated versions of an EGFP gene with non-optimal codons in various configurations (Figure 19B). For a more detailed description of these constructs, refer to section 6.6.1.

6.1.3 Endogenous reporters

In contrast to the well-defined arrangements of optimal and non-optimal codon stretches in the aforementioned synthetic genes, endogenous genes exhibit more heterogeneous CSC profiles. To investigate whether such heterogeneities can influence the mRNA stabilities of endogenous genes, we designed recoded versions of selected genes to ‘smoothen’ their CSC profiles and disperse non-optimal codons throughout the sequence (Figure 17). For a more detailed description of these constructs, refer to section 6.5.1.

6.2 The influence of non-optimal codon clustering on mRNA stability

6.2.1 Overview of asymmetric EGFP-mScarlet reporters

To explore the influence of non-optimal codon clustering on mRNA stability, we used variants of a tandem EGFP-mScarlet reporter. Within the reporter, we substituted codons with synonymous non-optimal codons having the minimum CSC values available for a given amino acid. These substitutions were introduced in varying proportions at different positions along the sequence. We denote optimal or stable codon clusters/stretches as ‘S’ and non-optimal or unstable ones as ‘U’.

We categorised these constructs into six groups, each with different percentages of unstable or non-optimal codons: 75% (3/4U), 50% (1/2U), 25% (1/4U), 12.5% (1/8U), 6.25% (1/16U) and 3.125% (1/32U). Except for the 3/4U and 1/32U constructs, each group comprised four variants, with the non-optimal codon clusters positioned at

various locations along the ORF. Figure 7A illustrates an example featuring the positional variants of the 1/2U construct. We included two additional constructs comprising 87.5% (7/8U) and 93.75% (15/16U) non-optimal codons to cover the entire spectrum of non-optimal codon proportions. We refer to these reporters as 'asymmetric EGFP-mScarlet' variants due to the asymmetries in the arrangement of optimal and non-optimal codon clusters among the positional variants within a group (Figure 7A).

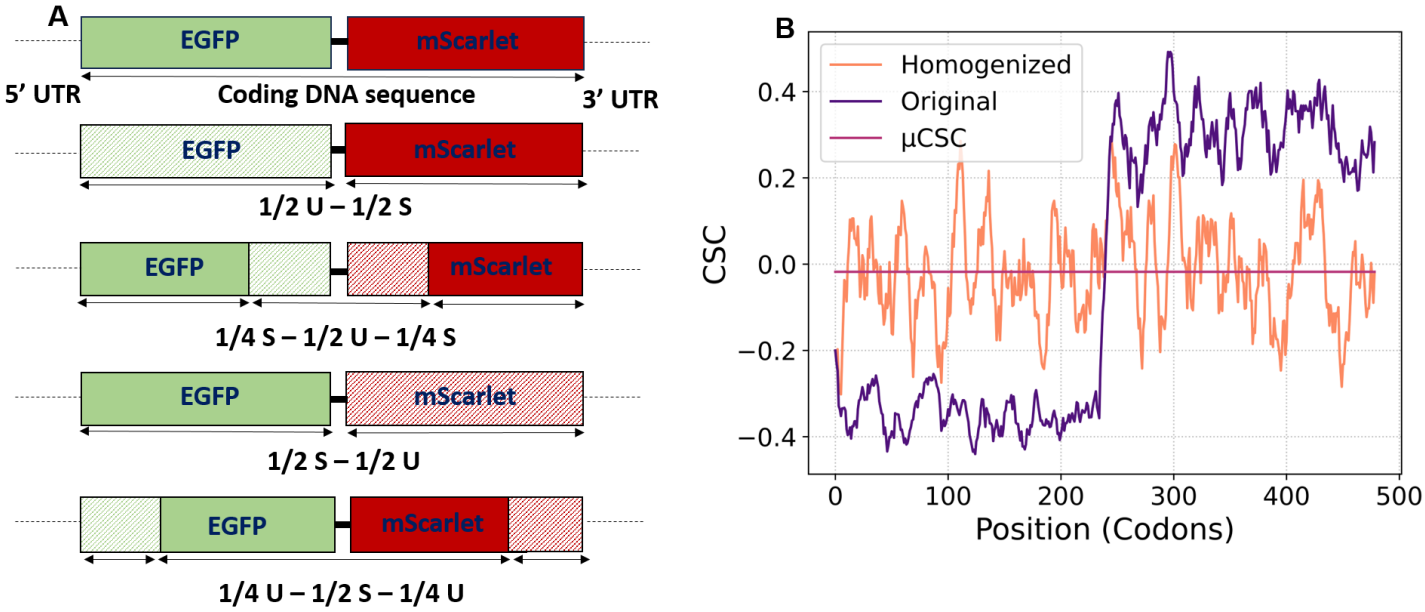

Figure 7: Schematic and CSC profiles of the asymmetric EGFP-mScarlet positional variants

*(**A**) Representation of the positional variants within the 1/2U category. All variants of the asymmetric EGFP-mScarlet constructs are 1467 bps in length. (**B**) The CSC profiles of the original (in purple) and the homogenised (in orange) versions of the 1/2U – 1/2S construct. The pink line indicates the μCSC of both constructs. The CSC profile is represented as the moving average of 10-codon windows along the ORF.*

For each category, we 'homogenised' the XU – X*S variant to study the general impact of non-optimal codon clustering on mRNA stability irrespective of the cluster location. Here, X represents the proportion of the sequence composed of non-optimal codons, while X* denotes the proportion comprising optimal codons. Figure 7B illustrates the CSC profiles of the original and homogenised versions of the 1/2U – 1/2S construct.

As previously mentioned, we use the μCSC of a transcript as a coarse-grained measure of its stability. Thus, transcripts with high μCSCs are expected to be more stable than those with low μCSCs.

6.2.2 Non-optimal codon clusters of different lengths exhibit variable position-dependent effects on mRNA stability

Our constructs confirm the previously observed inverse relationship between the proportion of non-optimal codons within an mRNA and its half-life (Presnyak et al., 2015). However, this relationship is not strictly linear (Figure 8A) due to the non-linear influences of codon configuration on mRNA stability. For example, the most stable positional variant of the 1/4U category and the least stable positional variant of the 1/8U category exhibit similar half-lives. A similar pattern can be observed for the 1/8U and 1/16U categories (Figure 8B). This highlights how the placement of non-optimal codon clusters at various positions within the mRNA can modulate mRNA stability, resulting in deviations from the expected linear relationship.

Except for the 1/32U category, all constructs exhibited a statistically significant difference in half-lives between the most and least stable positional variants (Figure 8B). Additional positional variants are required for the 1/32U category to determine if it also displays position-dependent effects on mRNA half-life.

However, these differences in mRNA half-lives don’t follow consistent patterns across all six categories of constructs (Figure 8B). For instance, consider the most stable variants from each category - the location of the non-optimal codon cluster is not identical. To illustrate, in the 1/16U category, placing the non-optimal codon cluster near the 5’ end of the ORF (1/16U – 15/16S) results in the most stable mRNA compared to the other variants. In the 1/4U construct, the most stable variant has the non-optimal codon cluster in the middle (3/8S – 1/4U – 3/8S) of the sequence. Conversely, in the 1/8U construct, the most stable variant has the optimal codon cluster in the middle (1/16U – 7/8S – 1/16U). Nevertheless, in all constructs except the 1/32U category, one positional variant consistently stands out as either the most or least stable with a statistically significantly different half-life compared to the remaining variants, which display similar half-lives (Figure 8B). Thus, while the

position of non-optimal codon clusters can significantly alter mRNA stability, their position-dependent effect varies with the length of the cluster.

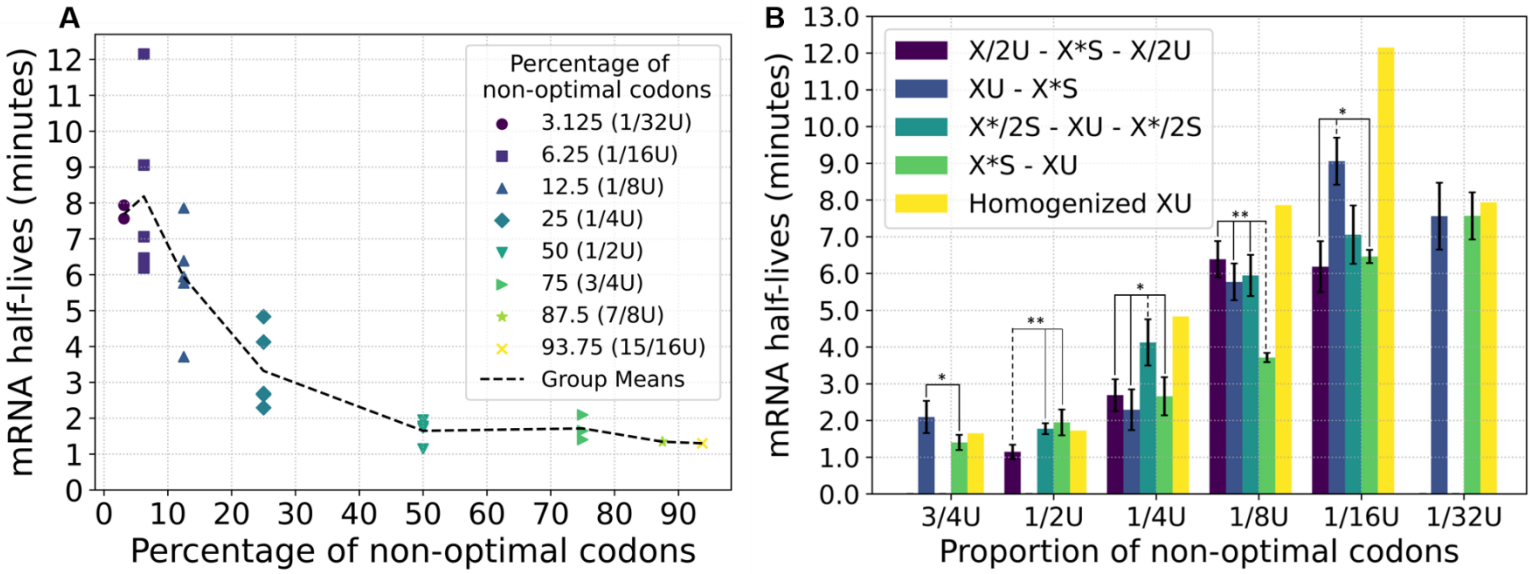

Figure 8: Influence of codon composition and configuration on the mRNA stability of the asymmetric EGFP-mScarlet constructs in WT cells

*(**A**) The inverse relationship between the percentage of non-optimal codons in the asymmetric EGFP-mScarlet constructs and their respective half-lives. The dashed line connects the mean half-lives of each group of constructs with the same percentage of non-optimal codons. (**B**) The mRNA half-lives of the asymmetric EGFP-mScarlet constructs in WT cells. All genes except the homogenised constructs have three to five replicate measurements. 'S' represents optimal codon clusters, 'U' represents non-optimal ones. X and X* represent the proportion of the sequence composed of non-optimal and optimal codons, respectively. Bars represent the means of replicates, while error bars indicate standard deviations. Pairwise comparisons amongst positional variants within each category were performed using the unpaired t-test (α = 0.05), and p-values were adjusted using the Benjamini-Hochberg correction. For multiple comparisons, the dashed line represents the construct whose mean differs significantly from the other positional variants. The symbols '*' and '**' indicate p-values less than 0.05 and 0.01, respectively.*

6.2.3 Position-dependent effects of non-optimal codon clusters on mRNA stability are independent of mRNA surveillance pathways

With these findings in hand, we sought to investigate whether the variations in half-lives among constructs with equal proportions of non-optimal codons but different codon configurations were due to their degradation by different mRNA surveillance pathways. To test this, we introduced a subset of the asymmetric EGFP-mScarlet constructs into three gene deletion backgrounds to assess their ability to rescue mRNA half-lives. Specifically, we examined the constructs in the following genetic backgrounds: *syh1*Δ*hel2*Δ (associated with no-go decay), *ski7*Δ (associated with nonstop decay) and *upf1*Δ (associated with nonsense-mediated decay). We also intended to introduce the constructs into a *not5*Δ background (general sensor of codon optimality); however, this strain had a substantial growth defect, and we encountered considerable difficulties transforming plasmids into this strain. Due to time constraints, we could not measure half-lives in this deletion background, which is planned as the next step.

While most constructs showed no substantial rescue (less than 1.2-fold) in their half-lives compared to the wild-type (WT) (Figure 9), indicating that mRNA surveillance pathways were not degrading them, an intriguing observation emerged.

Specifically, the half-life of the 3/4U – 1/4S construct decreased in the deletion backgrounds by as much as 1.5-fold (in the *syh1*Δ*hel2*Δ background) compared to the WT (Figure 9). We refer to this counterintuitive phenomenon as ‘squelching’, describing the destabilisation of mRNAs upon the inactivation of specific degradation pathways. The term was initially coined to describe the effect that transcriptional activators in high concentrations could have in repressing the transcription of other genes (Ptashne, 1988).

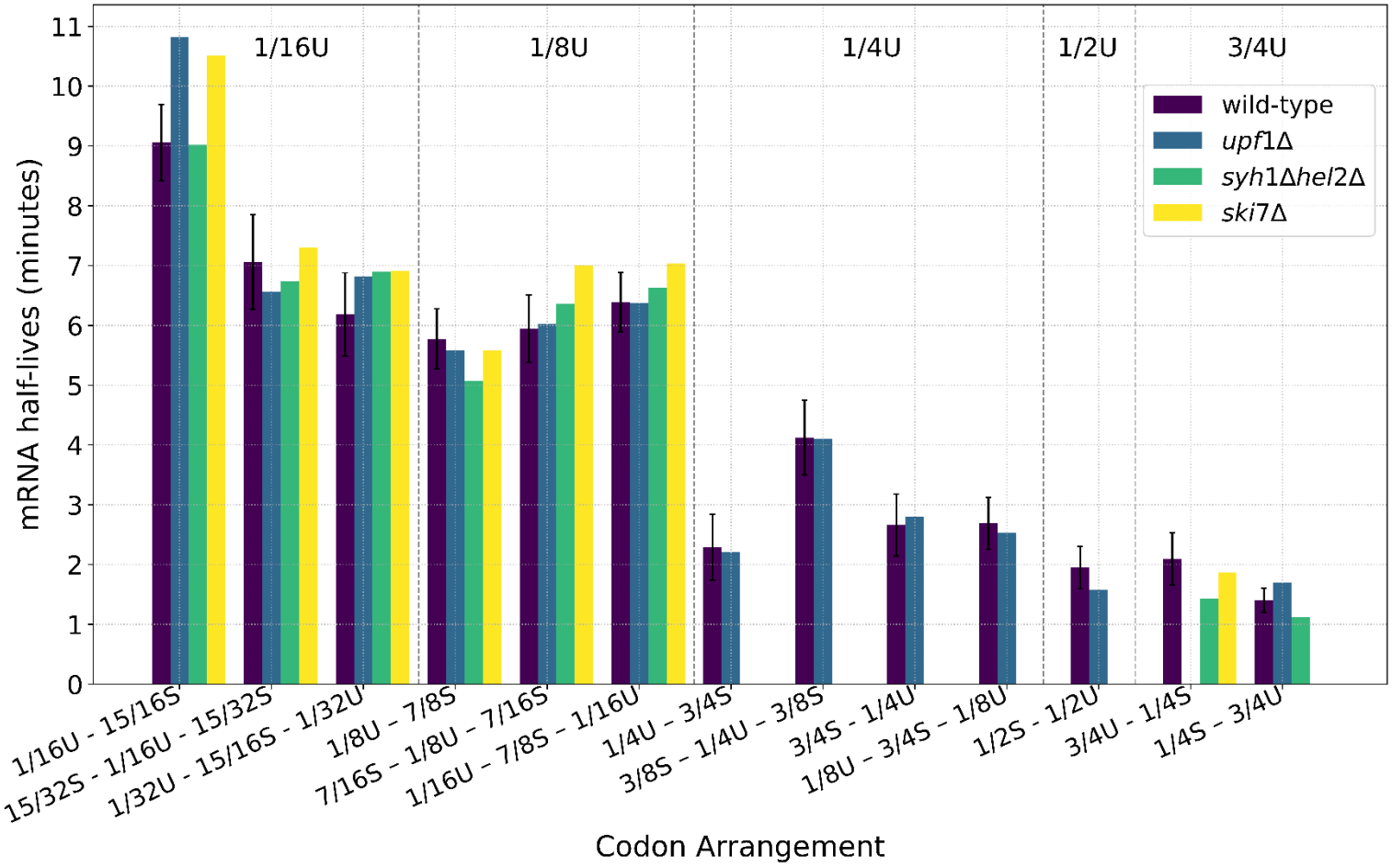

Figure 9: mRNA half-lives of the asymmetric EGFP-mScarlet constructs in various gene deletion backgrounds

The half-lives of a subset of the asymmetric EGFP-mScarlet constructs in three deletion backgrounds: syh1Δhel2Δ (in green), ski7Δ (in yellow) and upf1Δ (in blue). Half-lives of genes in the WT background (in purple) have three to five replicate measurements, while the rest have one measurement.

6.2.4 Non-optimal codon clusters destabilise mRNAs irrespective of length or position

In all cases, except for the construct subject to squelching, the homogenised constructs were the most stable among the positional variants in WT cells, with some variants closely approaching the half-life of the homogenised constructs (Figure 8B).

Surprisingly, for the 3/4U – 1/4S construct, the half-life of the homogenised construct was similar to that in the deletion backgrounds, which was less than the WT half-life (Figure 8B versus Figure 9).

These results imply that the clustering of non-optimal codons, regardless of the length or the position of the cluster, destabilises mRNAs compared to if the non-optimal codons are distributed throughout the mRNA. On the other hand, the clustering of non-optimal codons can stabilise mRNAs with a high proportion of non-optimal codons potentially due to squelching effects (3/4U – 1/4S).

To quantify the deviation of the half-lives of the positional variants compared to the homogenised constructs, we developed a metric termed the 'clustering coefficient'. We fit a sigmoid function to the half-lives and μCSCs of the homogenised constructs due to the expected saturation in half-lives at extremely low and high μCSC values (Figure 10A). We define the clustering coefficient (C) as follows:

$$C = \ T_{1/2}\ expected - \ T_{1/2}\ observed$$

where $T_{1/2}\ expected$ denotes the half-life of a gene without any clustering effect (as predicted by the sigmoid function), and $T_{1/2}\ observed$ is the measured half-life of the gene.

We examined the relationship between the constructs' clustering coefficients and μCSCs (Figure 10B). Similar to the lack of consistent position-dependent effects of non-optimal codon clusters on mRNA stability (Figure 8B), we observed a considerable variation in the clustering coefficients of the positional variants. These observations underscore the complexities of the relationship between codon configuration and mRNA stability. By expanding our dataset to span a broader range of μCSC values, we postulate that this coefficient may help guide the *in silico* design of sequences with target half-lives.

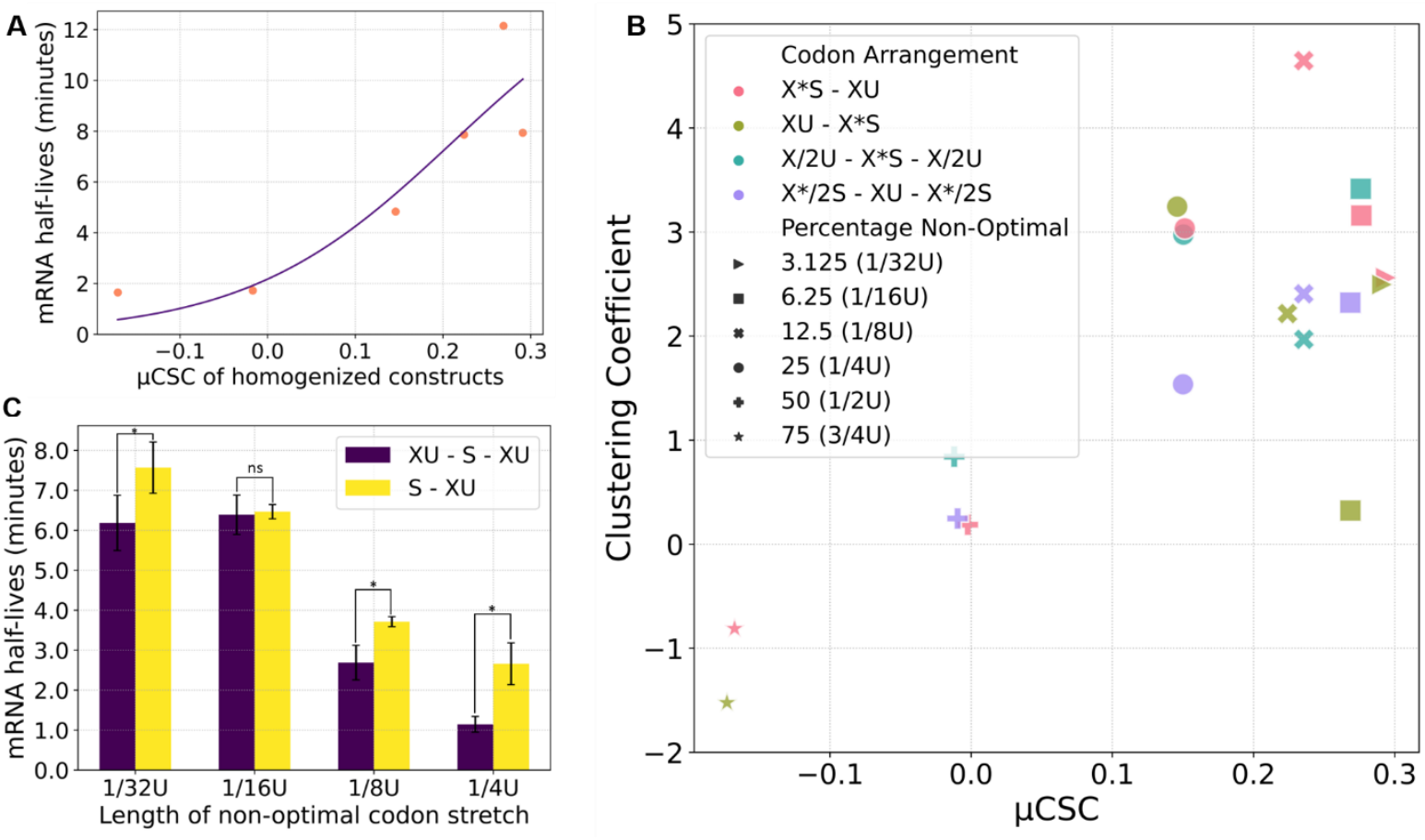

Figure 10: Clustering coefficient determination and the influence of multiple non-optimal codon clusters on mRNA stability

*(**A**) The sigmoid fit characterising the relationship between mRNA half-lives and μCSCs of the homogenised EGFP-mScarlet constructs. (**B**) Scatter plot of the clustering coefficients of positional variants of the asymmetric EGFP-mScarlet constructs. 'S' represents optimal codon clusters, 'U' represents non-optimal ones. X and X* represent the proportion of the sequence composed of non-optimal and optimal codons, respectively. (**C**) Comparison of half-lives of asymmetric EGFP-mScarlet constructs with two non-optimal codon clusters at the 5' and 3' ends of the ORF (in purple) and those with the same non-optimal codon cluster at the 3' end of the ORF (in yellow). Bars represent the means of three to five replicates, while error bars indicate standard deviations. Comparisons were made using a two-tailed t-test (α = 0.05). The symbol '*' indicates p-values less than 0.05. 'ns' stands for not significant.*

6.2.5 The influence of multiple non-optimal codon clusters on mRNA stability

78% of genes in *S. cerevisiae* contain more than one non-optimal codon stretch, which can influence mRNA stability in unanticipated ways. (Sharma et al., 2021) found that genes featuring two non-optimal eight-codon stretches near the 5' and 3' ends of the ORF were more stable compared to genes with a single eight-codon non-optimal stretch near the 3' end of the ORF. This was hypothesised to be due to the reduced influx of ribosomes into the mRNA mediated by the non-optimal codon stretch near the 5' end, consequently decreasing the formation of ribosomal traffic jams upstream of the 3' non-optimal codon stretch, which could stabilise the mRNA. However, for the asymmetric EGFP-mScarlet constructs, we found that constructs of the X*S – XU type exhibited significantly higher stability or no significant difference in stability compared to the XU – X*S – XU type (Figure 10C).

Notably, the smallest non-optimal codon cluster belongs to constructs of the 1/32U category, comprising 16 codons - already twice the length of the stretch used by (Sharma et al., 2021). It is thus possible that as the length of a non-optimal codon cluster increases, having two clusters compared to one could be detrimental to mRNA stability, potentially due to a higher proportion of non-optimal codons in the former.

6.3 The influence of non-optimal codon clustering on translation efficiency

Translation efficiency is the amount of protein produced per mRNA (Hanson and Coller, 2018). Given that the influence of codon composition on mRNA stability is thought to be mediated via translation (Hanson and Coller, 2018; Hia and Takeuchi, 2021; Liu et al., 2021; Presnyak et al., 2015; Radhakrishnan et al., 2016), we wanted to explore how the codon composition and configuration of the asymmetric EGFP-mScarlet constructs impact translation efficiency. We used the fluorescence intensities of the reporters as a proxy for their protein levels. Calculating translation efficiencies with this approach is particularly challenging for genes with a high proportion of non-optimal codons since they exhibit low fluorescence intensities above a negative background, making it difficult to estimate their protein levels

reliably. As a result, we could not calculate the translation efficiencies of constructs containing 50% or more non-optimal codons.

6.3.1 Non-optimal codon clusters of different lengths exhibit consistent position-dependent effects on translation efficiency

We observed a consistent position-dependent effect of non-optimal codon clusters on translation efficiency across all categories of constructs for which we calculated translation efficiencies (Figure 11). We found that the variant with the non-optimal codon cluster at the start of the mRNA had a higher translation efficiency (>1.5-fold) than the variant with the non-optimal codon cluster in the middle across all categories, as assessed by both EGFP and mScarlet. In the 1/4U category, the variant with the non-optimal codon cluster at the 3' end displayed the lowest translation efficiency. Additionally, in all categories, the X/2U - X*S - X/2U variant had a greater translation efficiency than variants with non-optimal codon clusters farther from the start codon.

Previous research into the sequence determinants of translation efficiency had proposed the existence of a 'ramp' sequence at the beginning of the ORF, which is thought to play an important role in regulating translation initiation and elongation. One proposed function of the ramp is to slow down translation near the start of the mRNA, consequently limiting the flux of ribosomes into the mRNA body. This is thought to increase translation efficiency by reducing the formation of ribosomal traffic jams, collisions and other events that decrease translation efficiency (Hanson and Coller, 2018; Hia and Takeuchi, 2021; Tuller et al., 2010). Ribosome pausing and queueing can influence translation efficiency through several mechanisms. Since translation initiation is typically regarded as the rate-limiting step in protein synthesis, modulating translation speed alone may only affect the rate of ribosome run-off rather than translation efficiency itself. It has been suggested that ribosomal queueing upstream of non-optimal codon stretches can interfere with translation initiation and, consequently, reduce the translation efficiency of transcripts (Hanson and Coller, 2018; Liu et al., 2021). Moreover, paused ribosomes engaged in unproductive translation can lead to premature translation termination, reducing translation efficiency (Liu et al., 2021). Thus, our findings align with the idea of slow translation at the beginning of the mRNA in increasing translation efficiency.

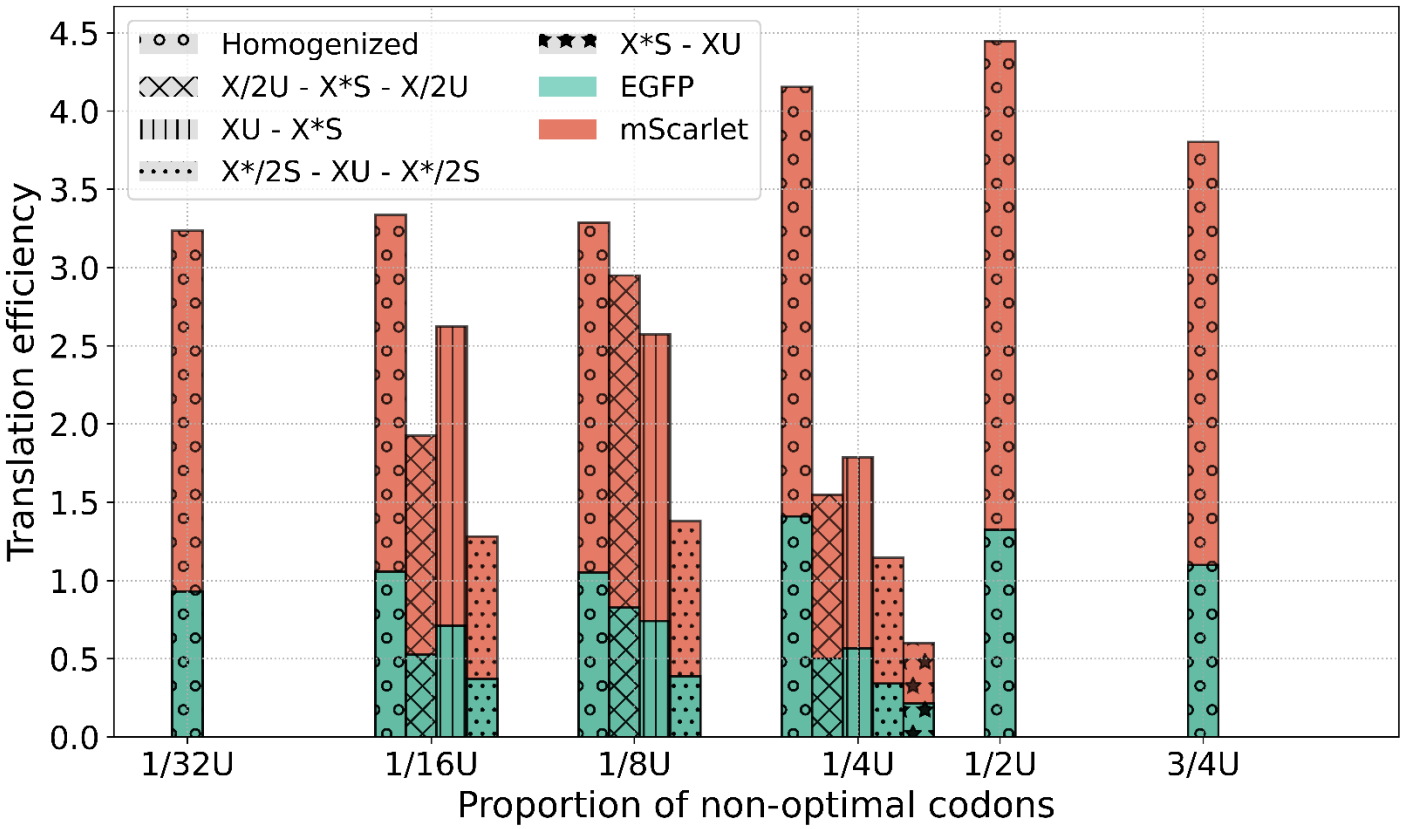

Figure 11: Translation efficiencies of the asymmetric EGFP-mScarlet constructs in WT cells

Translation efficiencies as determined by EGFP (in green) and mScarlet (in red) for a subset of the asymmetric EGFP-mScarlet constructs. Bars represent the mean of two replicate measurements. 'S' represents optimal codon clusters, 'U' represents non-optimal ones. X and X represent the proportion of the sequence composed of non-optimal and optimal codons, respectively.*

We also introduced a subset of the asymmetric EGFP-mScarlet constructs into the *syh1*Δ*hel2*Δ background to verify that it did not impact translation, as was the case for mRNA stability (Figure 9). Indeed, we found that the translation efficiencies in the deletion and WT backgrounds were very similar, differing by less than 1.2-fold (Figure 12).

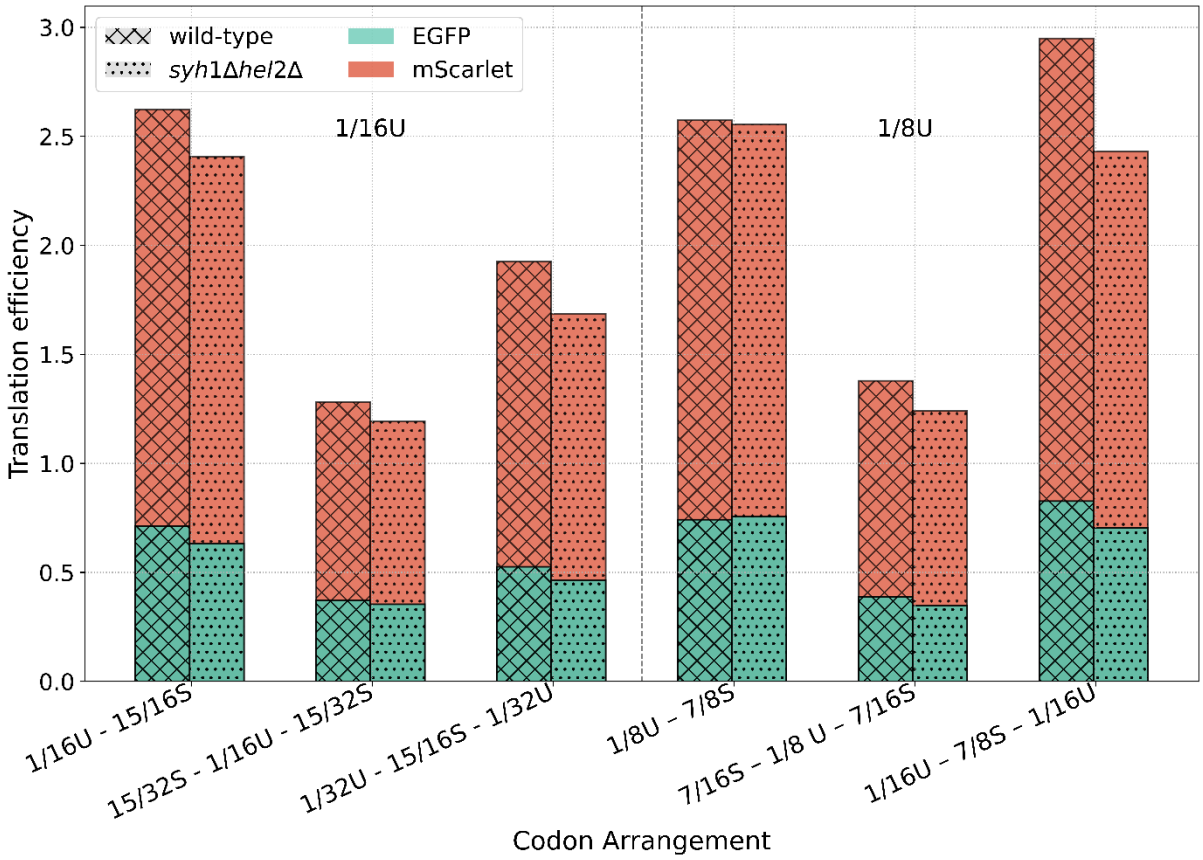

Figure 12: Translation efficiencies of the asymmetric EGFP-mScarlet constructs in *syh1*Δ*hel2*Δ cells

Translation efficiencies as determined by EGFP (in green) and mScarlet (in red) for a subset of the asymmetric EGFP-mScarlet constructs in WT and syh1Δhel2Δ backgrounds. Bars represent the mean of two replicate measurements for WT cells and one measurement for syh1Δhel2Δ cells.

6.3.2 Non-optimal codon clusters decrease translation efficiency irrespective of length or position

In line with the observed trend for the mRNA stabilities of the asymmetric EGFP-mScarlet constructs, we found that the homogenised constructs exhibited higher translation efficiencies than any of the positional variants across all categories for which we calculated translation efficiencies (Figure 11). While we could reliably measure the fluorescence intensities of the homogenised constructs across all proportions of non-optimal codons, the positional variants of mRNAs with more than 50% non-optimal codons yielded weak fluorescence signals (data not shown). This further underlines the impact of non-optimal codon clustering in hampering translation

and protein production for highly non-optimal genes. Thus, clustering non-optimal codons can destabilise the mRNA and reduce translation efficiency compared to if they are distributed across the mRNA body.

6.3.3 Codon composition differentially modulates mRNA stability and translation efficiency

We were particularly surprised to find that despite the percentage of non-optimal codons ranging from approximately 3% to 75% of the sequence, the translation efficiencies of the homogeneous constructs were confined to a narrow range. This range of variation in translation efficiencies (approximately 1.5-fold for EGFP and 1.3-fold for mScarlet) between the most and least efficiently translated homogeneous constructs, starkly contrasts with the approximately 7-fold change in their half-lives, ranging from less than 2 minutes to over 12 minutes (Figure 8B and Figure 11). Interestingly, it was previously shown that there was a positive correlation between the proportion of optimal codons in variants of the *HIS3* gene and their translation efficiencies (Hanson et al., 2018). It is important to note that we used protein fluorescence as a proxy for protein levels, which may not be perfectly correlated. Thus, we aim to verify the relationship between fluorescence intensities and actual protein levels in the future.

In summary, our results indicate that the proportion of non-optimal codons substantially impacts mRNA stability but minimally influences translation efficiency. In contrast, clustering non-optimal codons reduces both mRNA stability and translation efficiency, with the position of the cluster modulating both properties.

6.4 The characteristics of sequences inducing polarity effects on mRNA stability

6.4.1 Overview of polarity assessment constructs

To explore the characteristics of sequences that may induce polarity effects, we designed constructs featuring four types of non-optimal codon stretches inserted at different locations within the ‘optimised’ EGFP-mScarlet reporter (Figure 13). The optimised version of the gene was composed of codons with the maximum CSC

value available for each amino acid, resulting in a µCSC of 0.3, similar to that of *PGK1* (µCSC = 0.32).

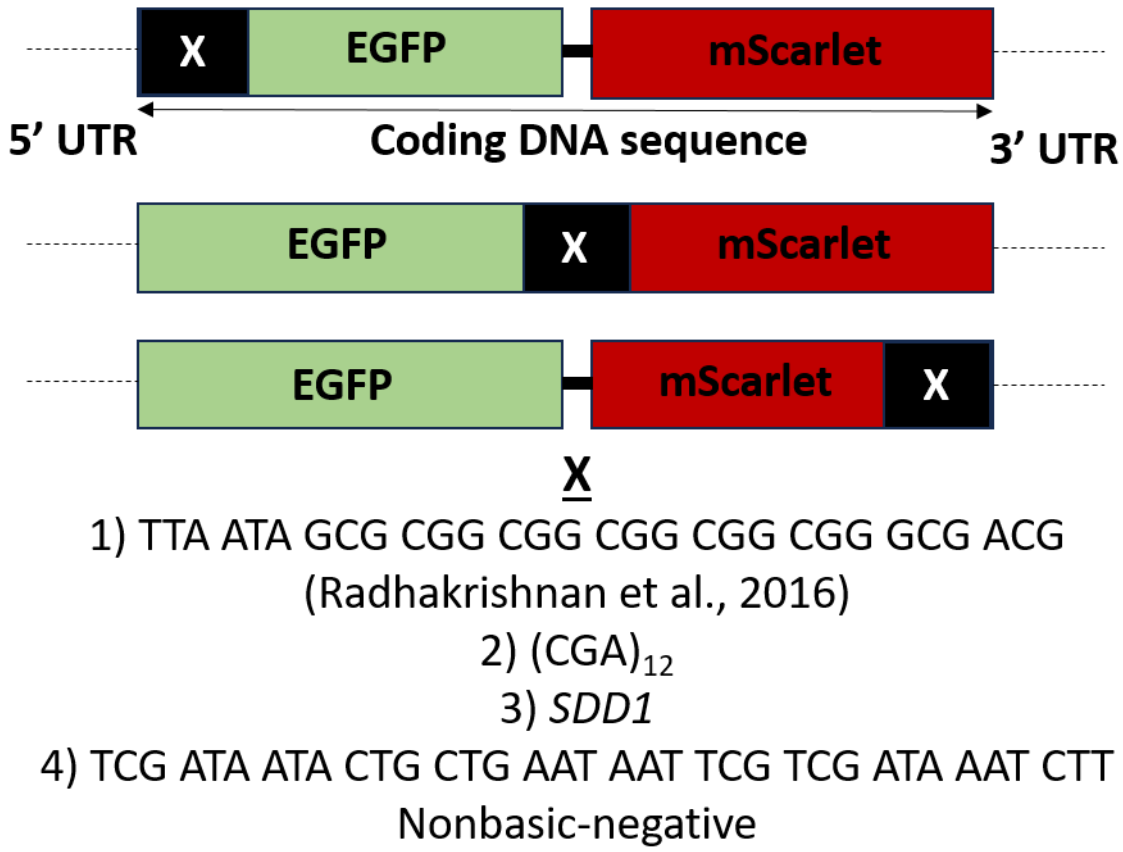

Figure 13: Schematic of the polarity assessment constructs

Schematic depicting the four types of non-optimal codon stretches placed at three positions within the optimised EGFP-mScarlet gene to study their position-dependent effects on mRNA stability.

We introduced the ten-codon non-optimal stretch used by (Radhakrishnan et al., 2016) (Figure 3A) into the optimised EGFP-mScarlet gene at the 5' end, the middle, and the 3' end of the ORF. Additionally, we introduced a $(CGA)_{12}$ codon stretch at the same positions, known to induce severe ribosomal stalling and collisions. CGA codons are decoded by inefficient wobble pairing, and CGA codon repeats are known to be potent translational inhibitors that substantially destabilise mRNAs through degradation by the no-go decay (NGD) pathway (Letzring et al., 2013; Veltri et al., 2022). Furthermore, we incorporated the 'arrest sequence' from the *SDD1*

gene in *S. cerevisiae* into the optimised reporter. *SDD1* is an endogenous substrate of the ribosome quality control (RQC) pathway, containing a stretch of polybasic amino acids known to cause ribosomal stalling due to interactions with its negatively charged exit tunnel (Matsuo et al., 2020; Monaghan et al., 2023; Shoemaker and Green, 2012). Lastly, we introduced what we term the 'nonbasic-negative' stretch into the optimised gene. The stretch comprises twelve codons characterised by exceptionally low CSC values and no positively charged amino acids. These constructs allowed us to assess the impact of non-optimal codon stretches on mRNA stability without the concurrent effect of positively charged amino acids. All constructs were also introduced into the *syh1*Δ*hel2*Δ background to examine whether position-dependent effects on mRNA stability arising from these stretches were due to induction of the NGD pathway.

6.4.2 Most non-optimal codon stretches exhibit modest partial polarity effects, except for CGA repeats

Among all constructs, those featuring non-optimal codon stretches in the middle and at the 3' end of the sequence displayed remarkably similar half-lives, differing by less than 1.1-fold (Figure 14). The variations in half-lives among the positional variants were introduced by the construct with the non-optimal codon stretches at the 5' end. Modest, partial polarity effects were observed in the (Radhakrishnan et al., 2016), *SDD1* and the nonbasic-negative constructs, displaying less than a 1.3-fold change between the most and least stable variants (Figure 14). For the (Radhakrishnan et al., 2016) constructs, the least stable variant had the non-optimal codon stretch at the 5' end, in contrast to the *SDD1* and nonbasic-negative constructs where the most stable variant belonged to that category.

The most substantial position-dependent effect was observed in the $(CGA)_{12}$ constructs, exhibiting a 1.7-fold change between the most and least stable variants (Figure 14). Surprisingly, the $(CGA)_{12}$ variants displayed a partial polarity effect, which was the inverse of the trend observed by (Radhakrishnan et al., 2016) (Figure 3B). The variant with the non-optimal codon stretch at the 5' end of the ORF was the least stable, while those with the non-optimal codon stretches in the middle and at the 3' end had similar half-lives.

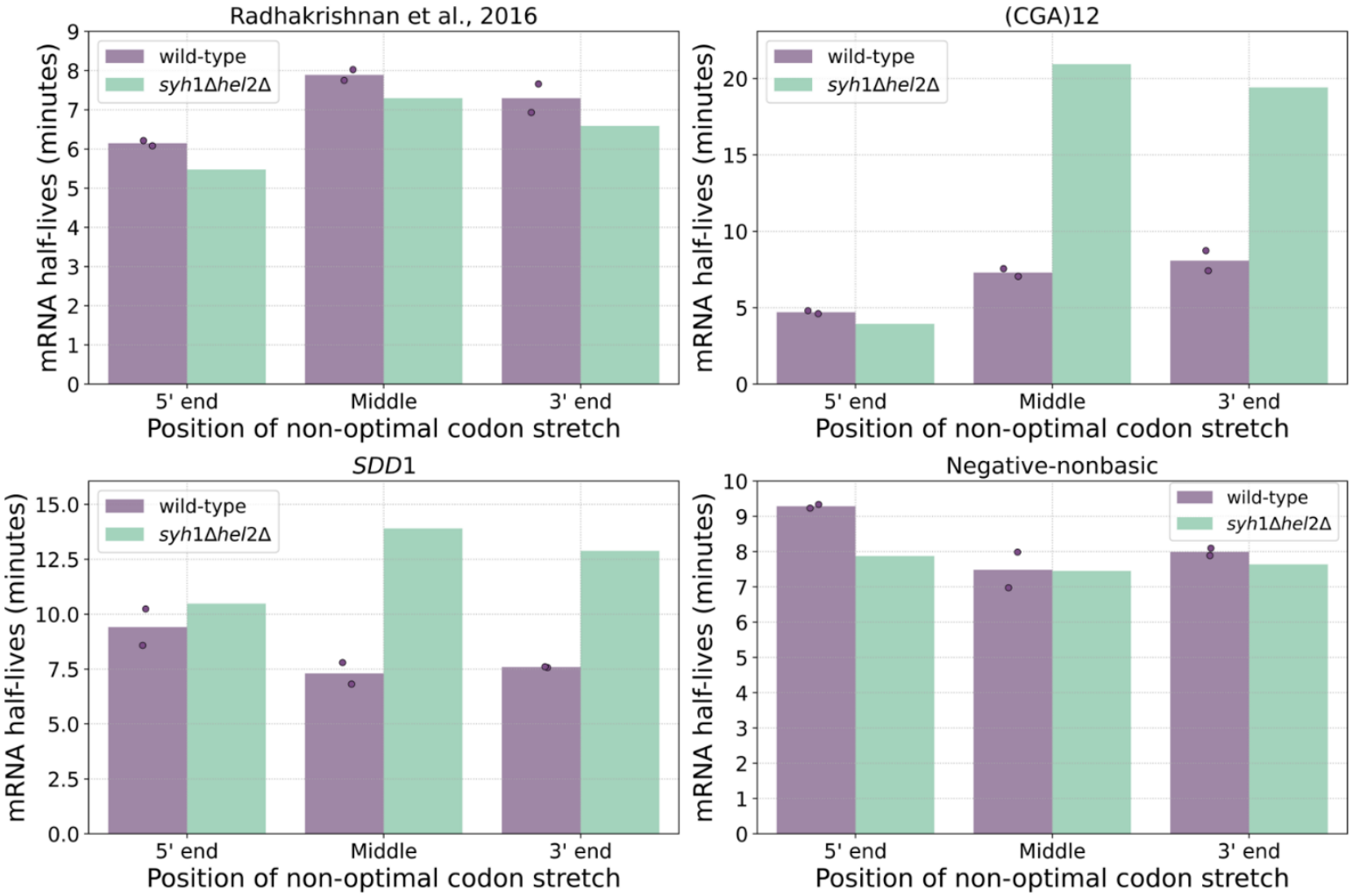

Figure 14: mRNA half-lives of the polarity assessment constructs in WT and *syh1*Δ*hel2*Δ cells

The half-lives of the polarity assessment constructs shown in Figure 13, in WT (in purple) and syh1Δhel2Δ (in green) cells. Bars represent the mean of two replicate measurements in the WT background and one measurement in the syh1Δhel2Δ background. Dots on the bars indicate individual replicates.

6.4.3 No-go decay targets some polarity assessment constructs for degradation

Both the *SDD1* and the $(CGA)_{12}$ variants, featuring the non-optimal codon stretches in the middle and at the 3' end of the sequence, were subject to degradation by NGD, as evidenced by their >2.4-fold and >1.7-fold rescue in the *syh1*Δ*hel2*Δ background respectively (Figure 14). This suggests the presence of ribosome collisions in these constructs, as they are required to trigger NGD (Monaghan et al., 2023; Simms et al., 2017).

Nonetheless, we cannot exclude the presence of collisions in other constructs, as it has been observed that recognition of collided ribosomes does not always induce NGD (Hia and Takeuchi, 2021; Meydan and Guydosh, 2020). As expected, constructs with non-optimal codon stretches at the 5' end of the ORF did not induce NGD since multiple ribosomes are typically required to trigger this pathway. Consequently, stall sequences near the start codon tend not to induce NGD since ribosome queues cannot extend upstream of the initiation site (Simms et al., 2017).

It has recently been shown that the decay rate of the NGD pathway saturates at relatively short translational stall lengths in constructs with as few as 3 CGA repeats (Hou et al., 2023). This could potentially explain the similar stabilities observed in the *SDD1* and $(CGA)_{12}$ variants with the non-optimal codon stretch in the middle and at the 3' end, despite their potential to form ribosome queues of varying lengths.

All constructs exhibit considerable variations in the codon composition of their non-optimal codon stretches in terms of their CSC values and the proportion of positively charged amino acids. While the nonbasic-negative stretch has the lowest µCSC value amongst all four stretches and lacks positively charged amino acids, the *SDD1* stretch has a µCSC of almost zero, with a third of the stretch comprising positively charged amino acids. The (Radhakrishnan et al., 2016) stretch has the third lowest µCSC value, with half its amino acids being positively charged. Despite these variations in their properties, they all exhibit modest partial polarity effects of similar magnitudes but different patterns. In contrast, the $(CGA)_{12}$ constructs with the second lowest µCSC value amongst all four stretches and all amino acids being positively charged display the most substantial position-dependent effect, a partial polarity

effect, which is the inverse of the trend observed by (Radhakrishnan et al., 2016) (Figure 3B).

These results suggest that short stretches of non-optimal codons with exceptionally low CSC values, when embedded in highly optimal CSC contexts, may not induce strong polarity effects even with the inclusion of positively charged amino acids. In contrast, short yet highly aberrant stretches like the $(CGA)_{12}$ stretch, which is entirely positively charged and decoded by wobble pairing, have the potential to introduce substantial partial polarity effects.

6.5 The influence of spatial heterogeneities in codon optimality on mRNA stability of endogenous genes

6.5.1 Overview of the gene selection procedure

We extended our analysis beyond synthetic gene sequences to explore how heterogeneities in the CSC profiles of endogenous genes in yeast might influence their mRNA stability.

Previous work by (Carneiro et al., 2019) examined the distribution of the positions of ten-codon non-optimal stretches with μCSCs identical to the ten-codon stretch used by (Radhakrishnan et al., 2016) across the yeast transcriptome (Figure 3A). They observed that such stretches were predominantly concentrated near the start codon (Figure 4A). We conducted a similar analysis of ten or twenty codon stretches with the minimum μCSC in 4214 genes in yeast. We used moving windows of ten or twenty codons, shifting one codon at a time. Subsequently, we computed the μCSC for each window and noted the starting position along the gene. We then identified the window with the lowest μCSC, referred to as the ‘most non-optimal window’, and normalised its starting position with the length of the gene to account for potential length-dependent effects. Finally, we analysed the frequency distribution of the most non-optimal windows across the transcriptome and observed a trend similar to that reported by (Carneiro et al., 2019), with the most non-optimal windows tending to be located near the 5’ end of the ORF (Figure 15A). Moreover, the most non-optimal windows were concentrated near the start codon for a higher percentage of highly optimal genes (μCSC > 0.15) in contrast to genes with the lowest μCSCs across the transcriptome (Figure 15B).

We also observed a positive correlation between the µCSCs of the most non-optimal windows and the respective genes across the yeast transcriptome (Figure 16), suggesting that highly non-optimal codon stretches are predominantly present in non-optimal genes (µCSC < 0).

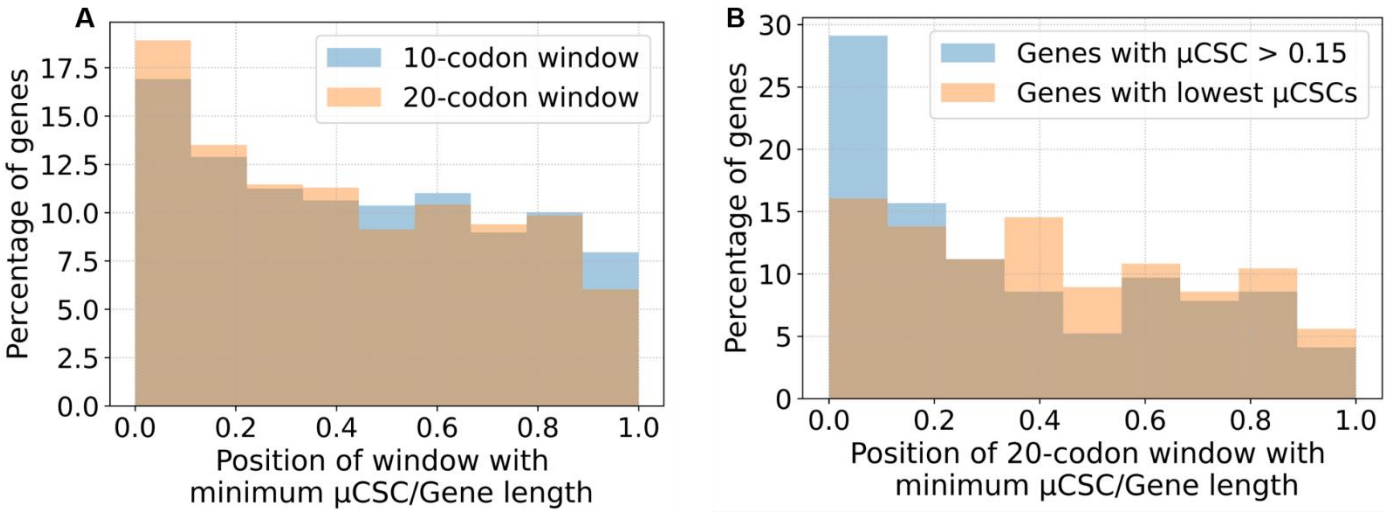

Figure 15: Frequency distribution of the most non-optimal windows in the yeast transcriptome

*(**A**) Frequency distribution of the positions of the most non-optimal windows normalised by gene length for ten (in blue) and twenty (in orange) codon windows for 4214 genes. (**B**) Frequency distribution of the positions of the most non-optimal twenty codon windows normalized by gene length for highly optimal genes (µCSC > 0.15) (in blue) and genes with the lowest µCSCs (in orange). We selected an equal number of genes with the lowest µCSCs in the transcriptome as the highly optimal genes.*

Interestingly, (Carneiro et al., 2019) noted that approximately two-thirds of *S. cerevisiae* genes were non-optimal, with negative µCSCs. We speculated that the configuration of non-optimal codon stretches would modulate the mRNA stability of optimal (µCSC >0) and non-optimal (µCSC <0) genes differently due to events hindering translation such as the formation of ribosomal queues and collisions, particularly in the former. This led us to explore the influence of heterogeneities in the CSC profiles of highly optimal genes on their mRNA stabilities.

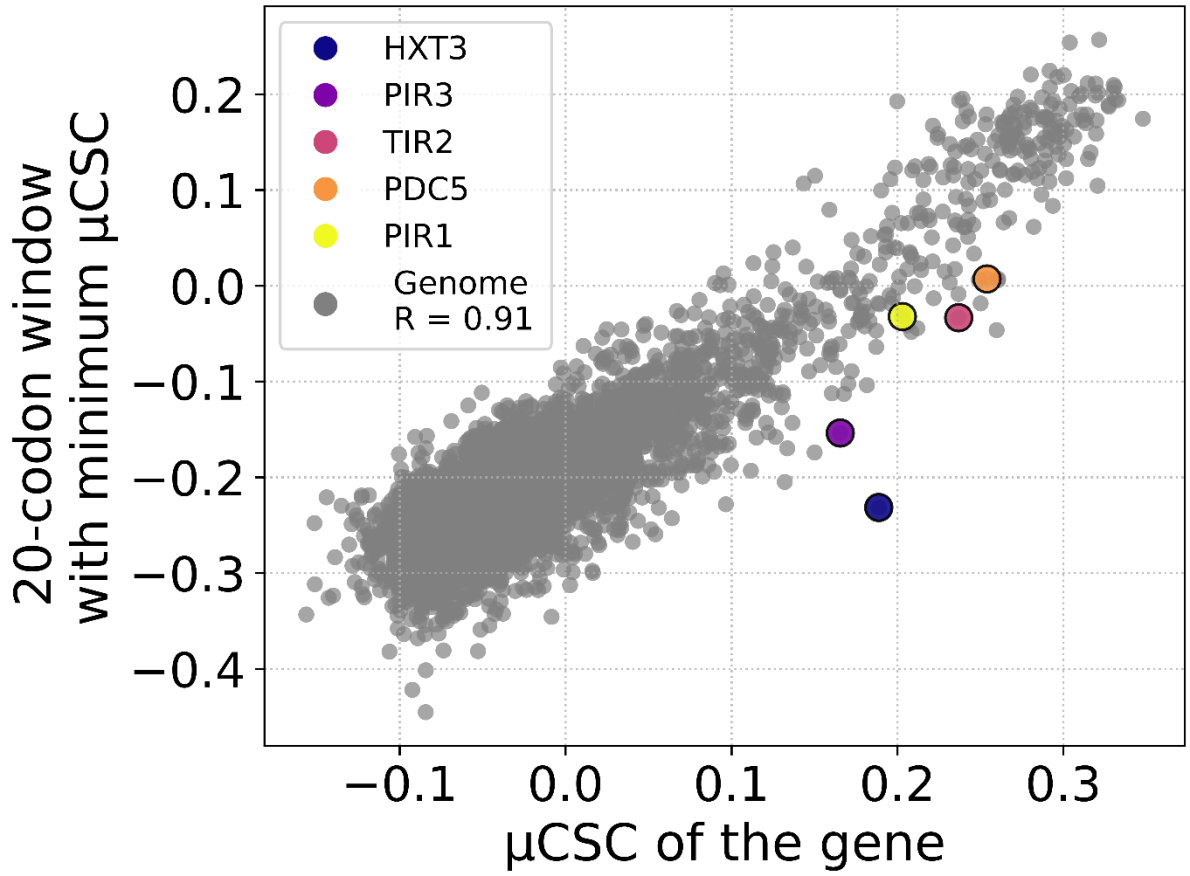

Figure 16: Linear relationship between the µCSC of the most non-optimal windows and genes in yeast

The µCSCs of the most non-optimal twenty-codon windows and the respective genes are linearly correlated. The endogenous genes selected for recoding are highlighted in different colours, while the remaining genes in the genome are grey. 'R' denotes the Pearson correlation coefficient.

To quantify the deviations of mRNA half-lives from the linear relationship between codon composition and mRNA stability, (Rahaman et al., 2023) performed a linear regression on the half-lives of mRNAs against their µCSCs. They calculated the residuals of mRNA half-lives from this linear fit, referred to as the ultracodonic stabilisation factor (UCSF). The UCSF encompasses all factors affecting mRNA stability beyond codon composition. While several factors could contribute to high UCSF values, such as the influence of different UTRs or codon sequence context, we were interested in exploring whether heterogeneities in CSC profiles could also contribute to the deviation from the linear relationship between codon composition and mRNA stability in endogenous genes.

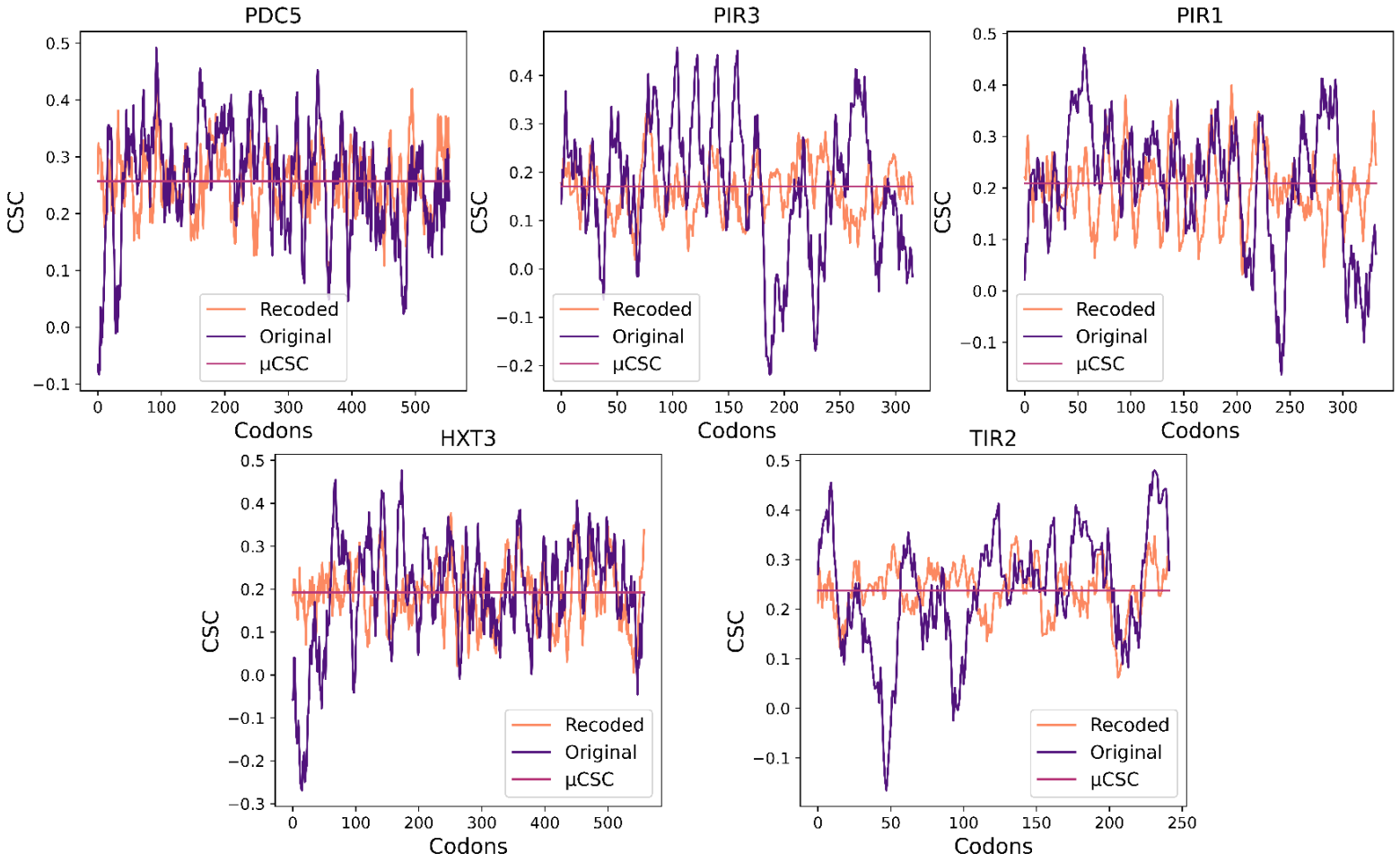

Figure 17: CSC profiles of the original and recoded endogenous genes

CSC profiles of the genes selected for recoding. The native sequence profile is in purple, while the recoded version is in orange. The μCSC of the native gene is in pink. The CSC profile is represented as the moving average of 10-codon windows along the ORF.

If such contributions exist, we may be able to change the mRNA half-life by 'smoothening' the CSC profiles of genes to disperse local heterogeneities. Among the 4214 genes in our dataset, we applied a filter to select genes with a μCSC > 0.15, which can be characterised as highly optimal genes (Carneiro et al., 2019). Subsequently, we sorted these genes based on the difference between the μCSC of the most non-optimal window and that of the gene for window lengths of ten, twenty or thirty codons. Since the top-ranking genes with the highest absolute differences were consistent among all three datasets, we used the twenty-codon window dataset for further analysis. We arbitrarily chose five genes from this dataset*: PDC5, PIR3, PIR1, HXT3* and *TIR2*, with high ultracodonic stabilisation factor (UCSF) values from

among the top 30 genes in the list. We recoded the gene sequences as previously described (Rahaman et al., 2023). The CSC profiles of the original and recoded 'smoothened' versions of the chosen genes are illustrated in Figure 17.

6.5.2 Heterogeneous CSC profiles can destabilise endogenous mRNAs

We observed that *PDC5*, *PIR1* and *TIR2* exhibited a modest change in the half-lives between the native and recoded versions, differing by less than 1.2-fold. However, the half-lives of *HXT3* and *PIR3* were substantially stabilised in the recoded versions, with a 2-fold and 1.6-fold greater half-life compared to the native genes (Figure 18).

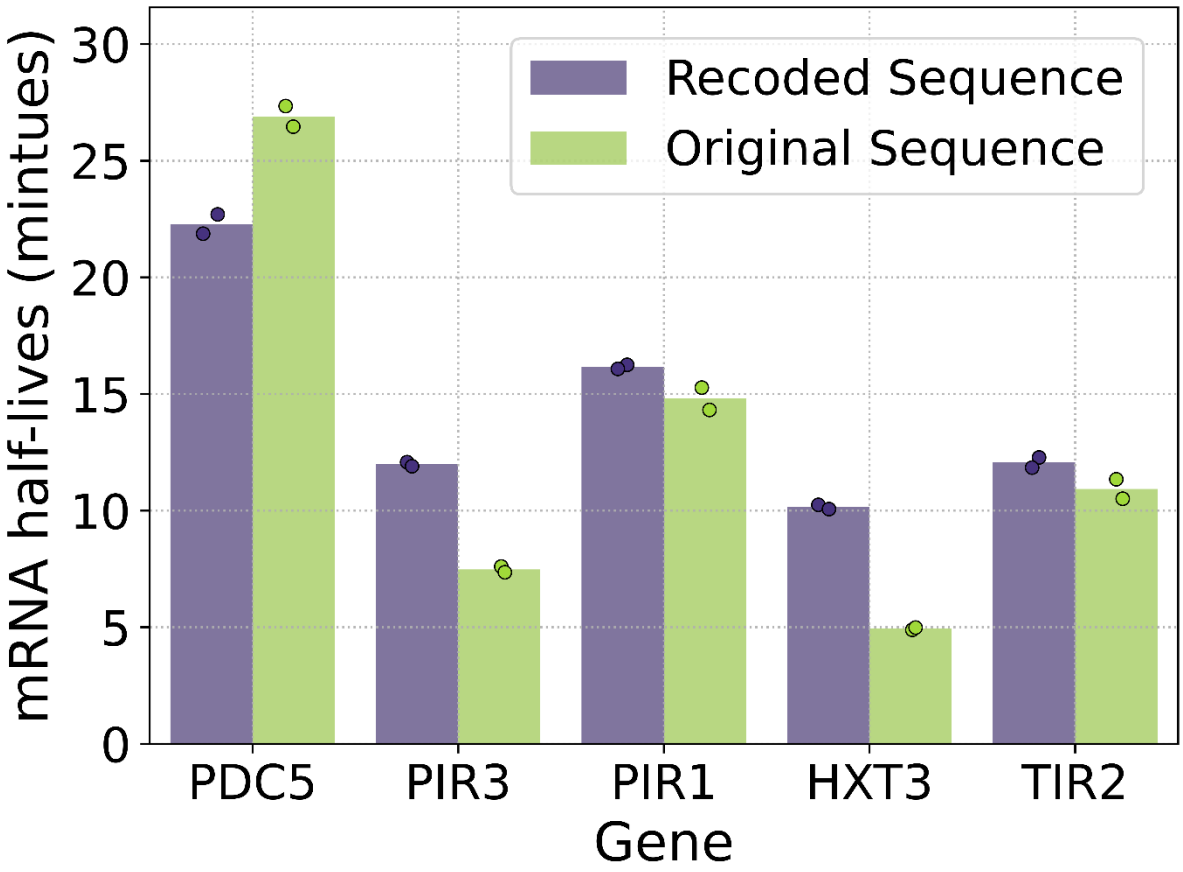

Figure 18: mRNA half-lives of the original and recoded endogenous genes

mRNA half-lives of the recoded (in purple) and original (in green) endogenous genes. Bars represent the mean of two replicate measurements, and the dots on the bars indicate individual replicates.

Among the genes with a µCSC > 0.15, *HXT3* had the largest difference between the µCSC of the most non-optimal window and the gene, while *PIR3* ranked third. These

results imply that heterogeneities in the CSC profiles of highly optimal endogenous genes can destabilise the mRNA and can introduce non-linear dependencies between codon composition and mRNA stability. Thus, tuning mRNA stability through the strategic arrangement of non-optimal and optimal codon stretches within genes may offer an additional mechanism for gene expression regulation.

6.6 The impact of length thresholds in coupling codon composition to mRNA stability

6.6.1 Overview of truncated fluorescent reporters

Several explanations have been proposed for the slow translation and consequently the high ribosome density (as determined by ribosome profiling, see section 9.2) in the ramp region. While (Tuller et al., 2010) attributed the slow translation in this region to the preferential use of rare codons, (Weinberg et al., 2016) found that differential codon usage alone could not explain the excess ribosome density on the ramp compared to the rest of the ORF. Their analysis revealed that, on average, the mean density of ribosome-protected fragments (RPFs) whose ribosome A sites mapped to a codon when it was present within the first 200 codons was 33% higher than if it were present in the remainder of the ORF. Notably, (Rahaman et al., 2023) recently identified a gene length threshold below which mRNA stability is unaffected by codon composition (Figure 19A). The threshold length was found to vary between 200 to 700 nucleotides depending on the specific UTR used, coinciding closely with the length of the ramp. We were thus interested in examining whether the threshold and ramp lengths could interact with each other to influence mRNA stability.

We considered two potential models for the interaction between the lengths of the ramp and the threshold in influencing mRNA stability: an ‘integration’ model and a ‘ramp’ model. In the integration model, the threshold and ramp lengths are independent, with the threshold acting on the entire gene length. In this scenario, the overall codon optimality, or the μCSC of the mRNA, would determine its half-life since the threshold would not mask the codon optimality of genes longer than the threshold length. In contrast, the ‘ramp’ model suggests an overlap between the threshold and the ramp lengths. In this model, the threshold masks the impact of codon composition in the ramp region, allowing for slow translation regardless of its

codon composition. This potentially uncouples the codon composition within this region from mRNA stability.

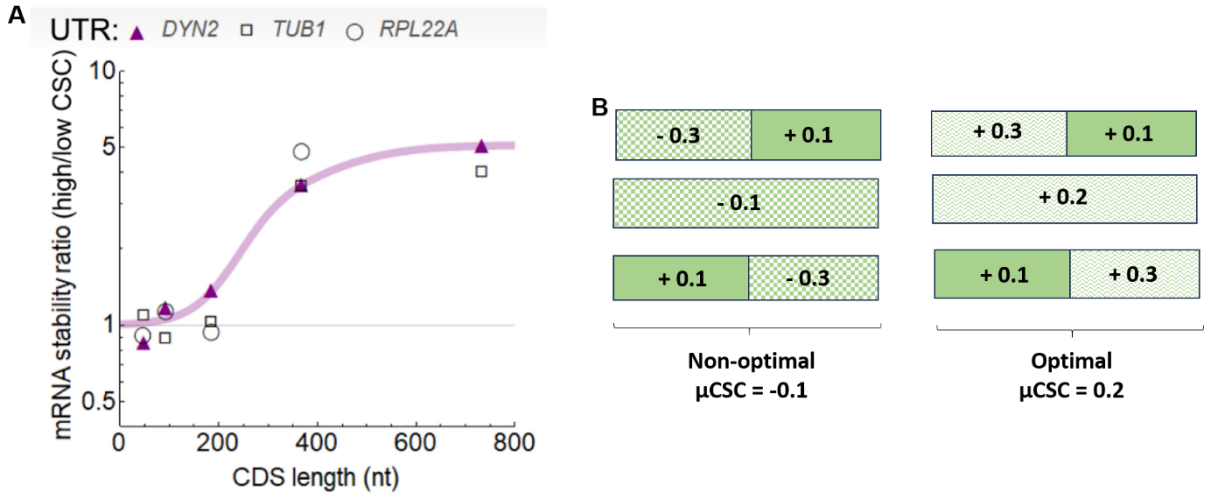

Figure 19: Gene length thresholds and design of truncated fluorescent reporters

*(**A**) The mRNA stability ratio for optimal (high µCSC) and non-optimal (low µCSC) reporters of different lengths and UTRs. The ratio approaches 1 below a specific length threshold for each UTR, indicating an uncoupling between codon composition and mRNA stability. Figure adapted from (Rahaman et al., 2023) (**B**) Schematic of the constructs designed to test the relationship between the threshold and ramp lengths. Constructs in the optimal category have a µCSC of 0.2, while constructs in the non-optimal category have a µCSC of -0.1. The numbers on the constructs represent the µCSC of a particular gene segment. The illustrated gene segments are half the total gene length. All genes are 369 bps in length, representing truncated EGFP sequences.*

We designed six truncated versions of the EGFP sequence to test this (Figure 19B)**.** The constructs have a length of 369 nucleotides, which exceeds the threshold length determined for the *RPL22A* UTR used in our constructs (Rahaman et al., 2023) (Figure 19A). Consequently, we expect them to be subject to codon optimality-mediated degradation.

There are two categories of constructs, with µCSCs of -0.1 (non-optimal) and +0.2 (optimal). Each category includes three types of constructs: one with a homogeneous CSC profile, one with a 0.1 µCSC codon stretch at the 3’ end, and its mirror image with the 0.1 µCSC codon stretch at the 5’ end (Figure 19B).

If the ‘ramp’ model is valid, we would expect constructs featuring the 0.1 µCSC codon stretch at the 3’ end in both the optimal and non-optimal categories to exhibit similar half-lives since the codon composition within the ramp region would not impact mRNA stability. Conversely, we would expect the half-lives of these constructs' mirror images to differ from themselves and each other. In contrast, if the ‘integration’ model is valid, we would expect all constructs in the non-optimal category to be less stable compared to those in the optimal category. We previously showed that heterogeneities in the CSC profile along the mRNA can have non-linear effects on mRNA stability (Figure 8A and Figure 8B). Therefore, it is unlikely that a complete integration of codon composition exists for the mRNA, given the influence of codon configuration.

Including the homogeneous constructs with optimal (µCSC = 0.2) or non-optimal (µCSC = -0.1) codon compositions allowed us to assess the impact of codon clustering within the mRNA body on its stability.

6.6.2 Codon composition is partially integrated and non-linearly influences mRNA stability in short genes

Our results revealed the presence of a partial integration effect since the constructs in the non-optimal category with codon clustering were less stable than constructs in the optimal category. We observed a more than 3.5-fold destabilisation of the non-optimal construct (µCSC = -0.1) with the 0.1 µCSC codon stretch at the 3’ end compared to its optimal counterpart (µCSC = 0.2) (Figure 20). Similarly, the mirror image of the non-optimal construct exhibited a 1.8-fold destabilisation relative to the corresponding variant in the optimal category. Notably, the non-optimal construct with the 0.1 µCSC codon stretch at the 3’ end was less stable than its mirror image (Figure 20). Thus, beyond the partial integration effect, codon optimality is differentially weighted at the 5’ and 3’ ends in these constructs. As expected, constructs in the optimal category with a µCSC of 0.2 all had similar half-lives, as the clustering of optimal codons is not likely to substantially impact mRNA stability in optimal genes. While the homogeneous non-optimal construct was more stable than the variants with codon clustering, we were extremely surprised to observe that the homogeneous optimal and non-optimal constructs displayed similar half-lives (~1.1-fold change) despite the difference in their µCSCs.

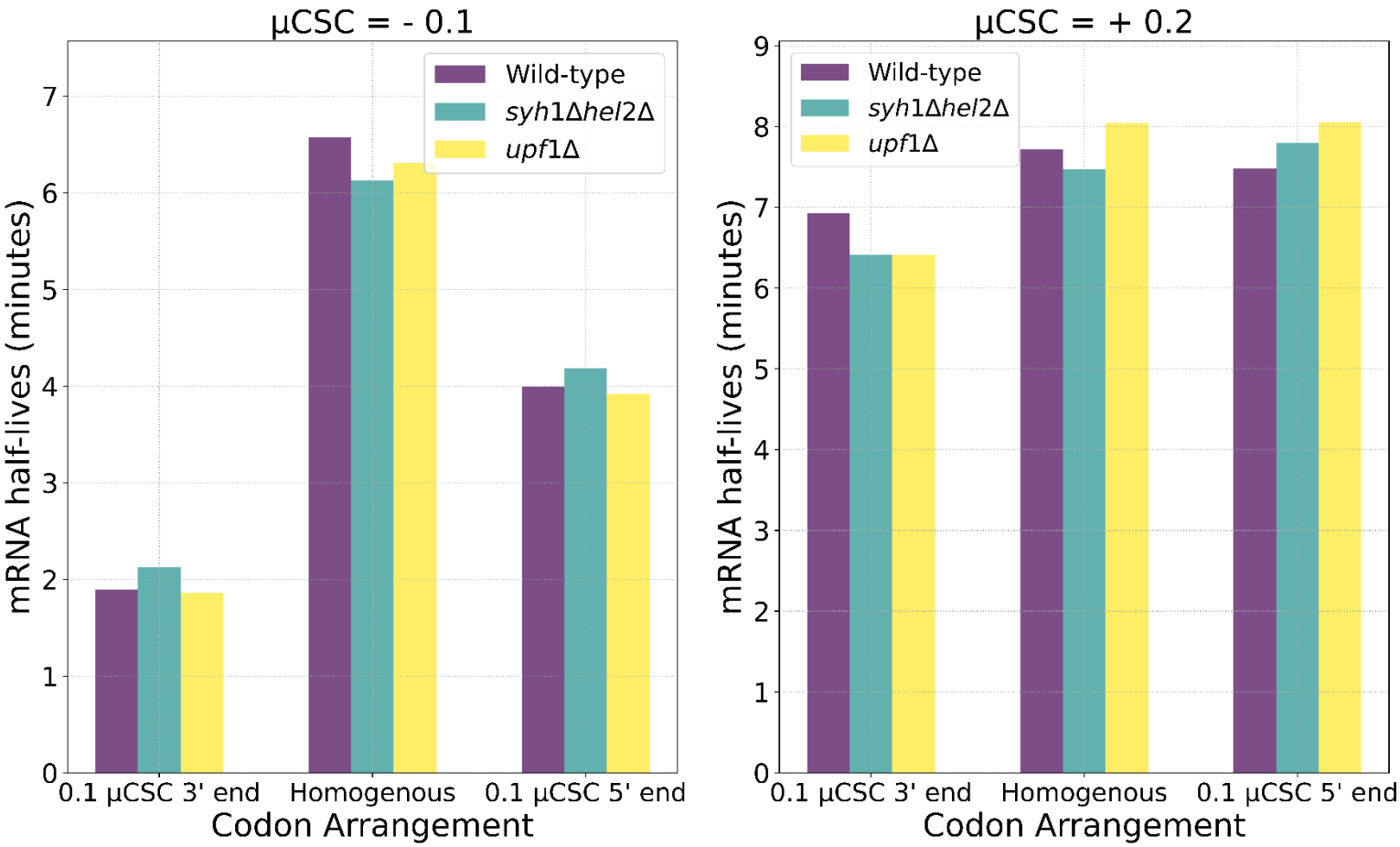

Figure 20: Half-lives of the truncated fluorescent reporters in WT and gene deletion backgrounds

The half-lives of the constructs in the optimal and non-optimal categories, depicted Figure 19B, measured in WT (in purple), syh1Δhel2Δ (in yellow), and upf1Δ (in green) cells.

We investigated the potential presence of squelching effects that might contribute to the stability of these constructs in the WT background. We thus introduced these constructs into the *syh1Δhel2Δ* and *upf1Δ* backgrounds to examine their impact on mRNA stability. We observed that the half-lives of all constructs in both deletion backgrounds were very similar to the WT half-lives, thus ruling out the stabilising influence of squelching in these mRNAs (Figure 20).

While the influence of spatial heterogeneities on mRNA stability has been established, the similarity in half-lives between the homogeneous constructs with µCSCs of -0.1 and 0.2 is unexpected. Importantly, the length threshold determined by (Rahaman et al., 2023) was established using optimal and non-optimal constructs

with a µCSC difference of ~0.6. In contrast, our constructs exhibit a µCSC difference of 0.3, which may modulate the threshold differently.

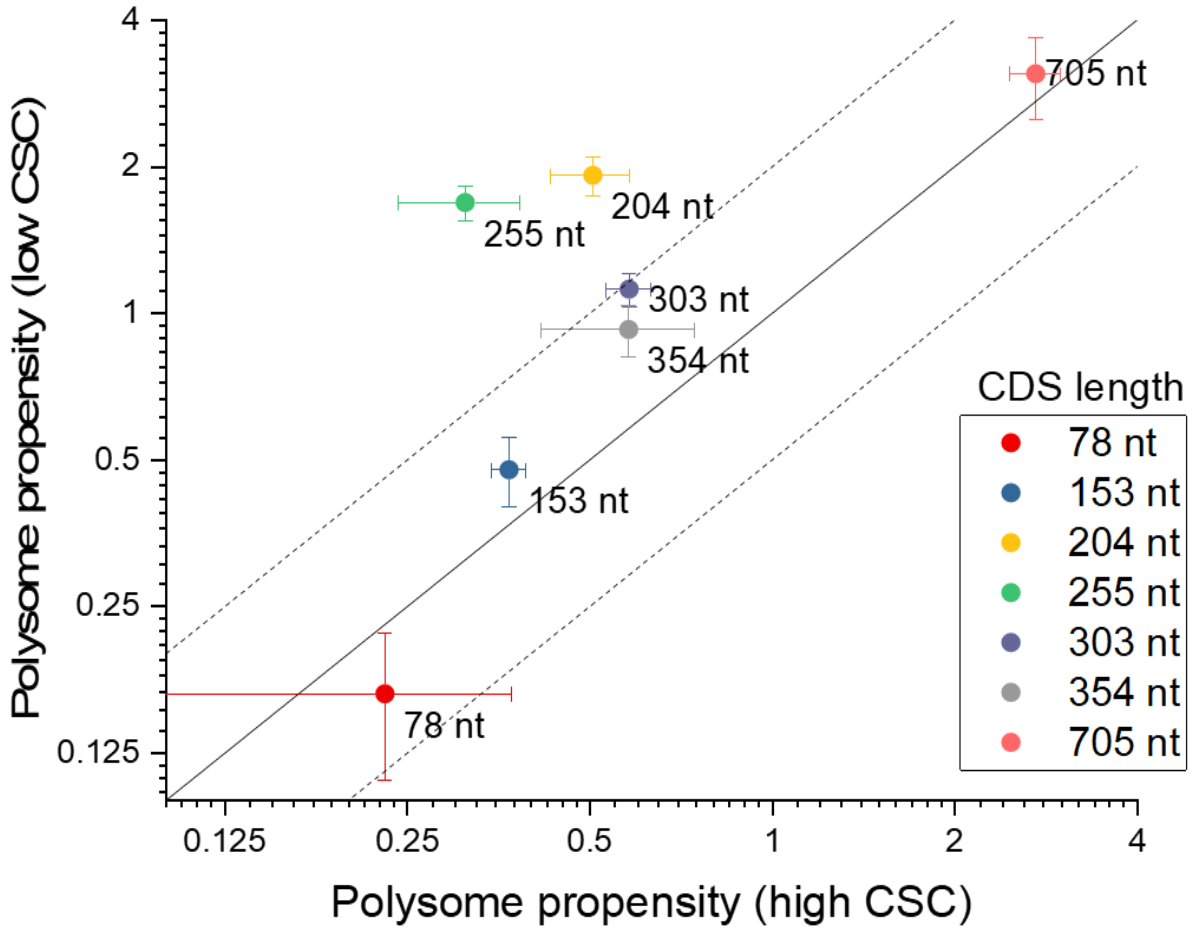

Figure 21: Relationship between polysome propensities of optimal and non-optimal genes

The relationship between the polysome propensities of optimal (high µCSC) and non-optimal (low µCSC) reporters of different lengths with the RPL41A UTR. The solid line represents the y=x line, while the dashed lines represent y = 0.5x and y = 2x. Polysome propensity is defined as the ratio of mRNAs in the polysome fraction occupied by three or more ribosomes compared to the monosome fraction, as determined by polysome profiling. Figure adapted from (Rahaman et al., 2023).

Additionally, it is known that mRNAs composed primarily of optimal or non-optimal codons have similar polysome profiles and ribosome densities, differing in ribosome

translocation rates (Presnyak et al., 2015). However, deviations from this equality have been observed in the case of shorter genes (<400 nucleotides) (Rahaman et al., 2023) (Figure 21). Thus, the nonlinearities in polysome profiles for genes of comparable lengths to our constructs, coupled with the influence of codon optimality, may lead to complex interactions, potentially resulting in unanticipated effects on mRNA stability in this gene length range.

6.7 Conclusions

Understanding the determinants of mRNA stability and translation efficiency has been an extensively studied topic in gene regulation. In this thesis, we aimed to explore the influence of codon composition and configuration on mRNA stability and translation efficiency.

While our results confirmed the previously established inverse relationship between the percentage of non-optimal codons in mRNAs and their half-lives (Presnyak et al., 2015) (Figure 2B), we showed that this relationship was not linear due to the non-linear influences of codon arrangement on mRNA stability (Figure 8A). Using the asymmetric EGFP-mScarlet constructs, we found that long non-optimal codon clusters could significantly alter mRNA stabilities across positional variants (Figure 8B), independent of mRNA surveillance pathways (Figure 9). However, no particular position could consistently stabilise or destabilise mRNAs when non-optimal codon clusters of varying lengths were placed there (Figure 8B). Furthermore, (Sharma et al., 2021) had previously reported that genes with two non-optimal codon stretches near the 5’ and 3’ ends of the ORF were more stable than those with a single non-optimal codon stretch near the 3’ end (Figure 5). However, we observed that mRNAs with a single non-optimal codon cluster at the 3’ end of the ORF exhibited either significantly higher stability or no significant difference in stability compared to mRNAs with two non-optimal codon clusters at the 5’ and 3’ ends (Figure 10C). Given that our non-optimal codon clusters are considerably longer than those used by (Sharma et al., 2021), we speculate that there may be disadvantages associated with the presence of two such clusters compared to one, potentially due to an increase in the proportion of non-optimal codons, whose contribution to mRNA stability may become more relevant as the length of the cluster increases.

Several constraints on mRNA sequences may underlie the destabilising or stabilising effects observed when non-optimal codon clusters of different lengths are positioned at specific locations. For example, the secondary structure of mRNAs, which is known to influence mRNA stability (Mauger et al., 2019), may be shaped by the configuration of codons along the mRNA body. Multiple factors, including the requirements for accurate translation and protein folding, preventing ribosomes from stalling in unproductive translation and avoiding inhibitory secondary structures, among others, likely interact in intricate ways to collectively determine mRNA stability, consequently favouring specific codon configurations over others for a given codon composition.

In contrast to their inconsistent effect on mRNA stability, we observed a consistent position-dependent effect of non-optimal codon clusters on translation efficiency, independent of the cluster length. We found that genes were translated more efficiently if such clusters were present at the 5' end of the ORF, compared to if they were located farther from the start codon (Figure 11). Thus, our results agree with the previously proposed role of the ramp region at the beginning of mRNAs in increasing translation efficiency by decreasing the speed of translation in this region.

We found that, in general, clustering non-optimal codons destabilises mRNAs and reduces translation efficiency, regardless of the cluster's length or position (Figure 8B and Figure 11), compared to when these codons are distributed across the mRNA. There are exceptional cases where clustering non-optimal codons can stabilise mRNAs. For example, in the 3/4U – 1/4S construct, we observe a 'squelching' effect, where the half-life of the construct decreases upon inactivation of mRNA degradation pathways. In this case, the half-life of the homogenised version of the construct matches its half-life in specific deletion backgrounds, which is less than the WT half-life (Figure 8B and Figure 9). Clustering non-optimal codons may decrease translation efficiency due to ribosome drop-off events, a phenomenon in which ribosomes prematurely terminate translation. Various factors, such as translational pauses or stress conditions, can trigger ribosome drop-off. Intriguingly, it was noted that the impact of ribosome drop-off on translation efficiency depends on codon composition and configuration (Bonnin et al., 2017). Moreover, it is possible that the presence of multiple adjacent slow-moving ribosomes on the cluster may

enhance the engagement of codon-optimality-mediated degradation pathways compared to individual ribosomes. However, the precise mechanism by which non-optimal codon clustering may destabilise mRNAs and reduce translation efficiency remains to be determined. As mentioned earlier, we could not measure mRNA half-lives in a not5Δ background due to time constraints and the challenges of working with this strain. Given that *NOT5* is a general sensor of codon optimality, it will be interesting to explore whether this codon optimality-based degradation pathway mediates the effects of codon clustering and configuration on mRNA stability.

While the influence of codon optimality on mRNA stability is evident, we were surprised to find that this was not the case for translation efficiency. The range of variation in translation efficiencies among the homogenised EGFP-mScarlet constructs was considerably narrower compared to the variation in the half-lives of these constructs (approximately 1.5-fold versus a 7-fold difference, respectively) (Figure 8B and Figure 11). Furthermore, our experiments demonstrated that clustering non-optimal codons, particularly in genes with a high proportion of non-optimal codons, adversely affects translation. This was evident from the substantially reduced fluorescence intensity observed in the positional variants of genes, with more than 50% of the sequence consisting of non-optimal codons compared to their homogenised counterparts (data not shown)**.**

We did not observe an inverse correlation between the distance of a non-optimal codon stretch from the start codon and mRNA stability, as reported by (Radhakrishnan et al., 2016) (Figure 3B and Figure 8B). We introduced the ten-codon non-optimal stretch used in their study (Figure 3A) into the optimised EGFP-mScarlet gene with a comparable μCSC and length as *PGK1*; however, we were unable to obtain the observed polarity of half-lives scaling with the distance of the non-optimal codon stretch from the start codon (Figure 14)**.** The lack of correlation between the position of ten-codon non-optimal stretches in endogenous yeast mRNAs and their half-lives was also previously demonstrated by (Carneiro et al., 2019). Specific features unique to *PGK1*, not captured by our constructs, may contribute to the observed polarity effect. This suggests that the CSC context in

which the non-optimal codon stretch is embedded alone may be insufficient to induce polarity effects.

We also tested other non-optimal codon stretches with a range of µCSCs and proportions of positively charged amino acids for their potential to induce polarity effects in the optimised EGFP-mScarlet gene (Figure 13). We found that constructs with stretches other than the $(CGA)_{12}$ stretch displayed modest partial polarity effects of different patterns. However, the $(CGA)_{12}$ constructs displayed a substantial partial polarity effect, which was the inverse of the trend observed by (Radhakrishnan et al., 2016) (Figure 3B). In this case, the construct with the stretch at the 5' end of the ORF was the least stable, while those with the stretch in the middle and at the 3' end of the ORF had similar stabilities (Figure 14).

While our previous findings were based on synthetic genes, we demonstrated that heterogeneities in the CSC profile along the gene body could also destabilise highly optimal endogenous mRNAs *in vivo*. Notably, for *HXT3* and *PIR3,* we observed a 2-fold and 1.5-fold rescue of half-lives, respectively, in the recoded version of the gene with a 'smoothened' CSC profile compared to the original sequence (Figure 18). An important consideration is our use of a modified 3' UTR for measuring the half-lives of these genes, which could affect their stability since UTRs are known to modulate mRNA stability (Rahaman et al., 2023). Thus, measuring mRNA half-lives with the native UTRs is crucial to rule out UTR-dependent regulation of mRNA stability. Modulating mRNA stability via the CSC profile could provide genes with an additional mechanism to use their codon composition as a regulatory element of gene expression.

Furthermore, we observed unexpected relationships between the µCSC of genes and their half-lives for short gene lengths (<370 bps). Specifically, we found a differential spatial weighting, where constructs with non-optimal codon stretches at the 3' end of the ORF were more stable compared to those with non-optimal codon stretches at the 5' end (Figure 20). However, the stability of homogeneous constructs with µCSCs of -0.1 and 0.2 were similar (~1.1-fold change). This may be due to non-linearities in the polysome profiles of optimal

and non-optimal genes in this length range (Figure 21), combined with the influence of codon composition (Rahaman et al., 2023).

Overall, our results reveal that codon configuration modulates mRNA stability and translation efficiency more intricately than previously realised.

7 Materials and Methods

The CSC values of codons used in this thesis were obtained from the study by (Jaquet et al., 2022).

7.1 Plasmids, UTRs and yeast growth conditions

Plasmids were chromosomally integrated into wild-type *S. cerevisiae* strains, BY4742 (mating type α) and BY4741 (mating type A), both descendants of the S288C strain, or into appropriate gene deletion strains. All gene deletion strains, except *syh1*Δ*hel2*Δ*,* were sourced from EUROSCARF. As a standard practice, we introduced the plasmid carrying the tTA gene (constitutively expressed) into mating type A strains of the relevant genetic background. In contrast, the plasmid carrying the target gene for analysis was introduced into mating type α strains of the same genetic background.

All genes except the endogenous genes featured 5’ and 3’ UTRs derived from *RPL22A*. This UTR was selected due to its relatively neutral impact on mRNA stability (Rahaman et al., 2023). We retained the native 5’ UTR of the endogenous genes while substituting their 3’ UTR with that of *RPL22A*. This allowed us to avoid any potential effects on translation initiation due to a non-native 5’ UTR. Replacing the 3’ UTR enabled the quantification of mRNA levels of the endogenous genes under the modified *tet* promoter without potential cross-reactions with native gene mRNAs.

Strains expressing a gene of interest were grown overnight in a shaker incubator for approximately 16 hours in synthetic complete media (2% raffinose, 0.005% glucose) at 30°C. The overnight cultures were refreshed to an OD_{600} of 0.15 and grown at 30°C till they reached an OD_{600} of ~0.5 (mid-log phase). If different strains were being pooled together for multiplexed gene control (MGC), they were combined in a single tube to ensure that each strain contributed an equal number of cells to the pool, resulting in a pooled OD_{600} of 0.15.

7.2 Knockout strain generation using homology-directed repair

Gene deletion was performed using the homology-directed repair mechanism in *S. cerevisiae.* A *hel2*Δ background strain (EUROSCARF accession: Y03625 and

Y13625) was used to generate the *syh1*Δ*hel2*Δ double deletion mutant. Templates for homologous recombination containing homology arms of 55-60 bps upstream and downstream of the *SYH1* ORF flanked the *URA3* ORF. These templates were PCR-amplified using Phusion DNA polymerase. >10 µg of PCR product was purified using ethanol precipitation and transformed into the *hel2*Δ background strains. Transformants were selected on -URA plates and screened for gene deletion by colony PCR and RT-qPCR, confirming the absence of both *SYH1* from its genomic locus and its mRNA.

7.3 DNA purification using ethanol precipitation

PCR-amplified homology template DNA was purified by ethanol precipitation for transformation in yeast. 7 M ammonium acetate (equivalent to 10% of the PCR reaction volume) was mixed with the PCR reaction. 2.5 volumes of ice-cold 100% ethanol were added to the solution before placing it at -80 ºC for 30 minutes. Afterwards, the solution was centrifuged at 20817 g (maximum speed) for 30 minutes at 4 ºC. The supernatant was discarded, and the pellet was washed with 70% ethanol (without resuspension), followed by centrifugation at 20817 g for 10 minutes at 4 ºC. The pellets were dried and resuspended in nuclease-free water and stored at -20 ºC.

7.4 Yeast transformation

Cells were grown overnight in YPAD media (2% glucose, 1% yeast extract, 2% peptone and 0.004% adenine sulphate) and refreshed to an OD_{600} of 0.2 in 50 ml, totalling eight transformations. Plasmids to be transformed were linearised (0.5 µg) with the appropriate restriction enzyme and heat-inactivated if necessary.

Exponentially growing cells were collected once they reached an OD_{600} of ~1 and centrifuged at 3000 g for 5 minutes at room temperature. The resulting pellet was washed with dH_2O and centrifuged again, followed by resuspension in 25 ml of LiT buffer (10 mM Tris-HCl pH 7.5, 100 mM lithium acetate). The cells were then incubated at room temperature for 40 minutes with gentle shaking. After incubation, the cells were centrifuged, and the supernatant was discarded, followed by resuspension in 750 µl of LiT buffer, yielding competent cells.

For each transformation, 50 µl LiT buffer, 5 µl salmon sperm DNA (pre-incubated at 95 ºC for 10 minutes) and 0.5 µg plasmid DNA were mixed with 100 µl of competent cells and incubated at room temperature for 10 minutes. Next, 300 µl of polyethylene glycol (PEG) solution (equal parts PEG 3350 and LiT buffer) was added to each tube and incubated at room temperature for 10 minutes. The cells were then subjected to a heat shock at 42 ºC for 15 minutes, followed by centrifugation at 20817 g for 1 minute. The supernatant was discarded, and the pellet was resuspended in 1 ml of YPAD media, followed by incubation at 30 ºC for one hour. Afterwards, the cells were harvested again by centrifugation at 20817 g for 1 minute and resuspended in 100 µl of 10 mM Tris-HCl (pH 7.5). The cells were then plated on an appropriate selection plate and incubated for 2-3 days at 30 ºC before assessing transformants for correct plasmid integration by colony PCR.

7.5 Colony PCR

Colony PCR was used to screen transformants for genomic integration of the desired plasmid at the correct location in a single copy. Colonies growing on the appropriate selection plates were resuspended in 10 µl of 0.02 M NaOH. The samples were then boiled at 99 ºC for 10 minutes to lyse the cells. The resulting lysates were kept on ice or at -20 ºC for extended storage. 1 µl of the lysate was amplified with *Taq* DNA polymerase (1M betaine was added to the PCR reaction).

Plasmids introduced into yeast can integrate into the genome in single or multiple copies, referred to as concatenation. Thus, primers were used to test plasmid integration and concatenation at the appropriate genomic locus. PCR products were visualised using agarose gel electrophoresis. Colonies with the correct sized bands for integration and the absence of bands for concatenation were selected for further experimentation.

7.6 Flow cytometry-based mating partner selection

The dynamic range of a gene measures how efficiently transcription is inhibited by doxycycline addition. It is quantified as the ratio of a gene's mRNA or protein levels in the absence of doxycycline to those in its presence. We selected a common mating partner with the tTA gene for all genes in a particular genetic background that drove a high dynamic range.

For dynamic range optimisation, multiple clones of the A strain with the tTA plasmid, including single and multiple copy integrands, were mated with randomly chosen clones of an α strain containing an EGFP-mScarlet reporter in a single copy. The α strain was chosen such that it exhibited a high fluorescence intensity (a proxy for EGFP-mScarlet protein levels).

Cells were mated anywhere between 6 hours to overnight on non-selective YPAD plates. Subsequently, they were streaked onto appropriate selection plates and incubated for 2-3 days at 30 ºC. Overnight cultures of the resulting diploids were grown and refreshed as previously described (see section 7.1). Each diploid strain was grown in two conditions: one with doxycycline (c_{final} = 10 µg/ml) and one without. Doxycycline was also added during the refresh at the same concentration. 500 µl of mid-log phase cells were mixed with phosphate-buffered saline (PBS) in a 1:1 ratio for flow cytometry. Fluorescence intensities were measured using the BD LSRFortessa™ system (see section 7.12 for more details). The tTA clone yielding the maximum dynamic range for both EGFP and mScarlet was selected as the mating partner for all strains of that specific genetic background.

7.7 Clone selection for dynamic range optimisation

There is considerable variability in dynamic ranges among different clones, even when using a common mating partner. Therefore, we aimed to identify a clone for each strain with single-copy gene integration exhibiting a high dynamic range. We randomly selected clones based on the results of the colony PCR and mated them with the common mating partner of the same genetic background. The resulting diploids were grown overnight with or without doxycycline (c_{final} = 10 µg/ml). Doxycycline was also added during the refresh at the same concentration. Cells from different strains were combined in a single tube for the refresh if they exhibited no cross-reaction (see section 7.8). 5 ml of exponentially growing cells were harvested in an equal volume of chilled methanol on dry ice. Cells were then centrifuged at 3000 g for 10 minutes, and the supernatant was discarded. The resulting pellet was stored at -80 ºC prior to cell lysis. RNA extraction and RT-qPCR were performed with gene-specific primers, as detailed in section 7.10. Dynamic ranges greater than ten were deemed acceptable. The clone exhibiting the highest dynamic range for each strain was selected for further experimentation.

7.8 Cross-reaction analysis for multiplexing

The measurement of mRNA half-lives using MGC (Baudrimont et al., 2017) involves common RNA extraction and reverse transcription (RT) steps for all genes of interest. Thus, we substantially reduce sample processing times by pooling or multiplexing cells from different strains together.

Since gene-specific primers are used to amplify individual genes from the pooled RNA samples, it is critical to ensure that primers for a target gene don't inadvertently amplify other genes in the pooled sample. Such off-target amplification or 'cross-reaction' can lead to an overestimation of mRNA levels for the target gene. Thus, eliminating cross-reactions within pools is crucial for accurately estimating mRNA levels and, consequently, mRNA half-lives.

Primers were designed using Primer-BLAST (Amplicon size between 75 to 150 nucleotides, melting temperatures between 58 and 60 ºC) (Ye et al., 2012). Furthermore, we assessed the propensity of the primer pairs for dimerisation and stem-loop formation using the OligoAnalyzer™ tool by Integrated DNA Technologies.

Pairwise nucleotide alignments of the genes of interest were analysed using BLAST (Altschul et al., 1990). The primer design strategy involved a range of approaches to prevent cross-reactivity: we selectively targeted regions in genes with limited sequence similarity to others in the same pool for amplification. In cases where such regions were unavailable, we exploited the requirement for double-stranded DNA for fluorescence emission by SYBR Green I, ensuring that, at most, one of the primers in a pair was complementary to off-target genes. For genes exhibiting substantial sequence similarity, we designed primers lacking 3' end complementarity with off-target genes, featuring partial complementarity at the 5' ends of the primers, with low melting temperatures. If none of these strategies were effective, we considered alternative pooling strategies to separate particularly challenging genes into distinct pools.

Upon designing a set of primers, we conducted a cross-reaction analysis. We linearised plasmids with the appropriate enzymes (New England Biolabs) to ensure robust amplification. We had previously observed multiple cases of non-amplification of target genes with the appropriate primers, hypothesised to be due to plasmid

supercoiling interfering with amplification efficiency. Subsequent troubleshooting was able to resolve this issue by linearising plasmids. As a result, we adopted a general practice of linearising all plasmids for the cross-reaction analysis.

Cross-reaction analysis involved testing the amplification of each primer pair against every gene to be pooled together using quantitative PCR (qPCR). We considered a cycle threshold (Ct) value difference of six or greater between the target gene and an off-target gene to indicate an absence of cross-reaction. The error in mRNA level estimation for the gene of interest is given by: $\frac{1}{2^{\Delta Ct}}$ where ΔCt is the difference between the Ct values of the target and off-target gene. Thus, a Ct value difference of six or more contributes an error of 1.5% or less in the estimation of mRNA levels of the target gene.

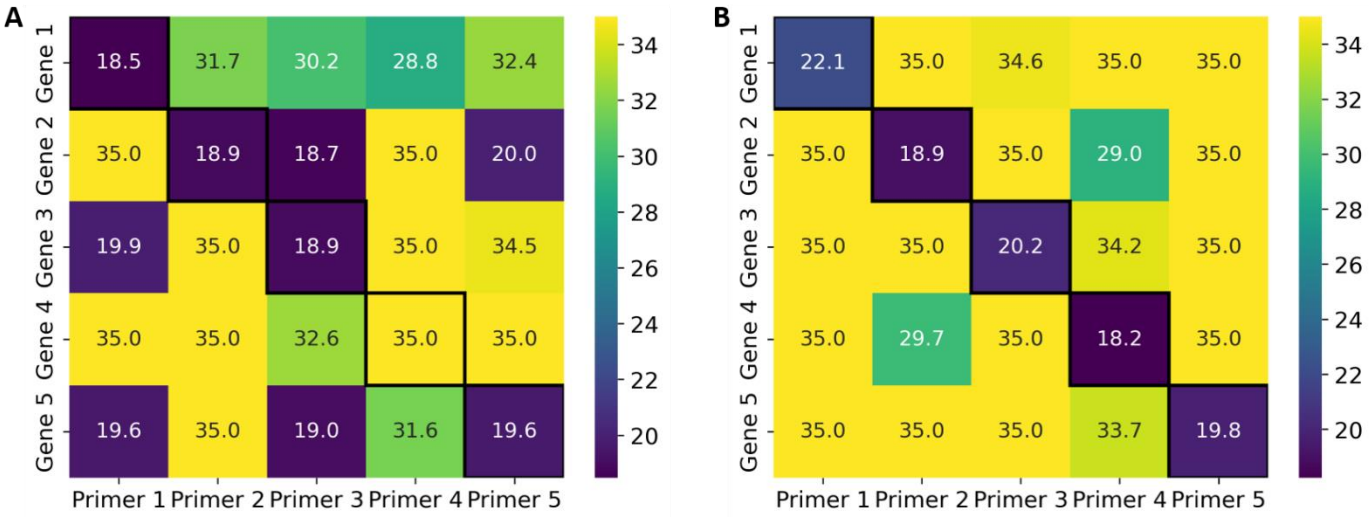

Figure 22: Cross-reaction analysis for multiplexing

*(**A**) Representative heat map of a cross-reaction analysis depicting cross-reactions and non-amplifications within a pool. (**B**) An example pool with no cross-reactions. Every primer pair is tested for amplification against every gene in a pool. Primer 1 targets gene 1 and so on. Cells represent the mean Ct values of technical duplicates for each amplification.*

A heat map illustrating an example cross-reaction analysis, showing cross-reactions and non-amplifications within a pool, is shown in Figure 22A. Figure 22B shows a pool without any cross-reactions.

7.9 mRNA decay

Cells were grown overnight and refreshed as previously described (see section 7.1). Cells from different strains were combined in a single tube for the refresh if they exhibited no cross-reaction (see section 7.8). Sampling for half-life estimation was performed at 30 ºC. Doxycycline was added (c_{final} = 10 µg/ml) to shut off transcription of genes of interest. Cells were harvested by collecting 5 ml of culture in chilled methanol on dry ice in a 1:1 ratio at the following time points: 0 minutes (before doxycycline addition), 0 minutes (immediately after doxycycline addition), 2, 4, 6, 8, 12, 18, 24, 36, 48 and 64 minutes. These time points were chosen to cover the typical range of half-lives observed in the yeast transcriptome. Cells were subsequently centrifuged at 3000 g for 10 minutes, and the supernatant was removed. The pellet was stored at -80 ºC prior to cell lysis. RNA extraction and RT-qPCR were performed with gene-specific primers, as detailed in section 7.10.

The following exponential decay curve was fit to the relative mRNA levels with Mathematica (Wolfram Research) to determine mRNA half-lives as previously described (Baudrimont et al., 2017):

$$R(t) = R_0 e^{-kt} + B$$

where $R(t)$ is the amount of RNA at a given time t, R_0 is the amount of RNA at the initial time point, and B is the amount of RNA due to basal expression. The function was fit from the 2-minute time point to account for the initial lag period following doxycycline addition, during which transcriptional repression occurs and after which mRNA levels start to decline.

The mean half-life was obtained by averaging half-lives calculated independently for biological replicates. All replicate experiments were performed with cells grown on different days.

7.10 RNA extraction, reverse transcription and quantitative PCR (RT-qPCR)

Cells were lysed at 4 ºC using the FastPrep-24™ 5G bead beating grinder and lysis system in Lysing Matrix Y tubes (MP Biomedicals #116960050-CF). Total RNA was

extracted from the lysate using the Rneasy Mini kit (Qiagen #74104) per the manufacturer's instructions, including the optional DNase digestion step. Reverse transcription (RT) to obtain complementary DNA (cDNA) from the extracted RNA was performed using the SuperScript III reverse transcriptase (Thermo Fisher Scientific #18080044) with gene-specific primers. The reaction was carried out using 1 µg of total RNA in a volume of 20 µl. Since we were using gene-specific primers, we diluted the recommended quantity of enzyme ten-fold using 0.1 µl per reaction.

The cDNA obtained was quantified with gene-specific primers using qPCR, in a reaction volume of 10 µl, using the KAPA SYBR Fast master mix and the Roche LightCycler 480. Technical duplicates were performed for each qPCR reaction. Ct values were calculated by the LightCycler 480 SW 1.5.1, using the maximum of the second derivative of the amplification curves. Relative mRNA levels were computed by normalizing average gene Ct values using the geometric mean of two housekeeping genes, TFC1 and UBC6, as follows:

$$R = \frac{\sqrt[2]{(1+TFC1_e)^{TFC1_{Ct}} \times (1+UBC6_e)^{UBC6_{Ct}}}}{(1+GOI_e)^{GOI_{Ct}}}$$

where R is the relative amount of mRNA for a gene of interest (GOI); $TFC1_e$, $UBC6_e$ and GOI_e are the primer efficiencies of the primers targeting *TFC1, UBC6* and the *GOI,* respectively. $TFC1_{Ct}$, $UBC6_{Ct}$ and GOI_{Ct} are the average Ct values of the technical duplicates of *TFC1, UBC6* and the *GOI,* respectively.
The specific housekeeping genes were chosen since they exhibit little variation in their mRNA levels across different environmental conditions (Baudrimont et al., 2017).

7.11 Primer efficiency determination

cDNA obtained either from the first time point of an mRNA decay or from steady-state mRNA samples was used to measure primer efficiencies. 0.25 µg of cDNA was used as the starting concentration for five 4-fold serial dilutions. 2 µl of each dilution was added in a 10 µl qPCR reaction. A linear regression was performed between the diluted samples' average Ct values and the sample quantities' logarithm. Finally, the efficiency was calculated as follows:

$$E = (Dilution\ factor)^{\frac{-1}{slope}} - 1$$

where E is the primer efficiency of a primer pair, $Dilution\ factor$ is the fold-dilution and $slope$ is the slope of the aforementioned linear regression. Data points with average Ct values >34 or residuals with absolute values > 1 were not considered for the efficiency calculation.

7.12 Translation efficiency determination using flow cytometry

Translation efficiency is the amount of protein produced per mRNA (Hanson and Coller, 2018). We used the fluorescence intensities of genes as a proxy for their protein levels. We did not pool cells of different strains due to the use of common primers.

Cells were grown overnight and refreshed as previously described (see section 7.1). Samples of exponentially growing cells were collected for quantification of steady-state mRNA levels using common primers as previously described (see sections 7.9 and 7.10). 500 µl of the remaining sample was mixed with phosphate-buffered saline (PBS) in a 1:1 ratio for flow cytometry. Fluorescence intensities were measured using the BD LSRFortessa™ system. The Ex488_505LP_512/25-A filter was used for EGFP and Ex561_600LP_610/20-A for mScarlet. The fluorescence of a negative control strain containing no fluorescent proteins – BY4743 was used to normalise the measured fluorescence intensities. RNA extraction and RT-qPCR were performed with common primers, and relative steady-state mRNA levels were calculated as previously described (see section 7.10). All replicates were performed using cells grown on different days.

Common primers targeting a shared region for all genes of interest were used to account for differences in amplification efficiencies that might affect steady-state mRNA level calculations. In cases where a single primer could not be used for all genes, we performed a 'cross-normalization' of steady-state mRNA levels. We used genes that could be amplified by different primer sets and calculated the ratios of mRNA levels obtained from them. These ratios were then used to normalise the mRNA levels of genes amplified by different sets of primers.

For genes with low fluorescence intensities relative to the background, random fluctuations and measurement noise make it difficult to estimate their protein levels reliably. Thus, we only calculated translation efficiencies for genes with a 5-fold or

greater fluorescence intensity over the negative background for both EGFP and mScarlet.

FlowJo (V10.9.0) was used for flow cytometry data analysis. After identifying yeast cells by SSC-A v/s FSC-A gating, we performed two rounds of doublet elimination, using SSC-W v/s SSC-A and FSC-w v/s FSC-A gating, to obtain single yeast cells. We then calculated the mean fluorescence intensity for EGFP and mScarlet from the single cells and normalised them with the fluorescence intensity of the negative control strain in the EGFP and mScarlet filters, respectively. Finally, we divided the relative fluorescence intensities by the relative steady-state mRNA levels to obtain the translation efficiency.

7.13 Recoding sequences to homogenise CSC distributions

For the asymmetric EGFP-mScarlet constructs, we homogenised the variant of the XU – X*S type (where X represents the proportion of the sequence composed of non-optimal codons and X* represents the remaining proportion composed of optimal codons). We aimed to disperse the non-optimal codons in the XU region evenly throughout the sequence for homogenisation. To achieve this, we calculated the ideal spacing between the non-optimal codons of each amino acid in the XU region to be the length of the mRNA divided by the number of occurrences of the amino acid in the XU region. For each amino acid in the XU region, we identified all the positions in the sequence where it occurred and generated all $\binom{n}{u}$ permutations, where n represents the number of occurrences of the amino acid in the entire sequence and u is the number of occurrences of the amino acid in the XU region. We then selected those configurations of optimal and non-optimal codons from the permutations of each amino acid, which minimised the difference between the mean and ideal inter non-optimal codon distances. For amino acids where computing all possible permutations was computationally prohibitive, we manually determined the positions of non-optimal codons, maintaining the percentage error between the mean and ideal inter non-optimal codon distance to be less than 5%. From the previously selected configurations, we finally chose the ones with the minimum standard deviation of inter non-optimal codon distances. In cases where multiple configurations met the aforementioned criteria, we selected one randomly. When there was only one

occurrence of an amino acid in the XU region, we randomly positioned the non-optimal codon along the sequence where that amino acid was available. After determining the positions of non-optimal codons for amino acids in the XU region, we recoded the entire sequence, ensuring that it matched the original sequence's amino acid sequence and codon composition. We retained the original linkers and restriction sites and made minor manual adjustments to the sequence to remove restriction sites involved in cloning or plasmid linearisation.

8 References

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). An Overview of Gene Control. In Molecular Biology of the Cell. 4th Edition, (Garland Science), p.

Bahiri-Elitzur, S., and Tuller, T. (2021). Codon-based indices for modeling gene expression and transcript evolution. Computational and Structural Biotechnology Journal *19*, 2646–2663. https://doi.org/10.1016/j.csbj.2021.04.042.

Balagopal, V., Fluch, L., and Nissan, T. (2012). Ways and means of eukaryotic mRNA decay. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms *1819*, 593–603. https://doi.org/10.1016/j.bbagrm.2012.01.001.

Baudrimont, A., Voegeli, S., Viloria, E.C., Stritt, F., Lenon, M., Wada, T., Jaquet, V., and Becskei, A. (2017). Multiplexed gene control reveals rapid mRNA turnover. Science Advances *3*, e1700006. https://doi.org/10.1126/sciadv.1700006.

Bonnin, P., Kern, N., Young, N.T., Stansfield, I., and Romano, M.C. (2017). Novel mRNA-specific effects of ribosome drop-off on translation rate and polysome profile. PLoS Comput Biol *13*, e1005555. https://doi.org/10.1371/journal.pcbi.1005555.

Buschauer, R., Matsuo, Y., Sugiyama, T., Chen, Y.-H., Alhusaini, N., Sweet, T., Ikeuchi, K., Cheng, J., Matsuki, Y., Nobuta, R., et al. (2020). The Ccr4-Not complex monitors the translating ribosome for codon optimality. Science *368*, eaay6912. https://doi.org/10.1126/science.aay6912.

Carneiro, R.L., Requião, R.D., Rossetto, S., Domitrovic, T., and Palhano, F.L. (2019). Codon stabilization coefficient as a metric to gain insights into mRNA stability and codon bias and their relationships with translation. Nucleic Acids Research *47*, 2216–2228. https://doi.org/10.1093/nar/gkz033.

Chen, C.A., Ezzeddine, N., and Shyu, A. (2008). Chapter 17 Messenger RNA Half-Life Measurements in Mammalian Cells. In Methods in Enzymology, (Academic Press), pp. 335–357.

Cheng, J., Maier, K.C., Avsec, Ž., Rus, P., and Gagneur, J. (2017). Cis-regulatory elements explain most of the mRNA stability variation across genes in yeast. RNA *23*, 1648–1659. https://doi.org/10.1261/rna.062224.117.

Coller, J., and Parker, R. (2004). Eukaryotic mRNA Decapping. Annual Review of Biochemistry *73*, 861–890. https://doi.org/10.1146/annurev.biochem.73.011303.074032.

Decker, C.J., and Parker, R. (1993). A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. Genes Dev. *7*, 1632–1643. https://doi.org/10.1101/gad.7.8.1632.

D'Orazio, K.N., Wu, C.C.-C., Sinha, N., Loll-Krippleber, R., Brown, G.W., and Green, R. (2019). The endonuclease Cue2 cleaves mRNAs at stalled ribosomes during No Go Decay. eLife *8*, e49117. https://doi.org/10.7554/eLife.49117.

Embree, C.M., Abu-Alhasan, R., and Singh, G. (2022). Features and factors that dictate if terminating ribosomes cause or counteract nonsense-mediated mRNA decay. Journal of Biological Chemistry *298*, 102592. https://doi.org/10.1016/j.jbc.2022.102592.

Garí, E., Piedrafita, L., Aldea, M., and Herrero, E. (1997). A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in Saccharomyces cerevisiae. Yeast *13*, 837–848. https://doi.org/10.1002/(SICI)1097-0061(199707)13:9<837::AID-YEA145>3.0.CO;2-T.

Garneau, N.L., Wilusz, J., and Wilusz, C.J. (2007). The highways and byways of mRNA decay. Nat Rev Mol Cell Biol *8*, 113–126. https://doi.org/10.1038/nrm2104.

Geisberg, J.V., Moqtaderi, Z., Fan, X., Ozsolak, F., and Struhl, K. (2014). Global Analysis of mRNA Isoform Half-Lives Reveals Stabilizing and Destabilizing Elements in Yeast. Cell *156*, 812–824. https://doi.org/10.1016/j.cell.2013.12.026.

Goel, N.S., Subba Rao, G., Yčas, M., Bremermann, H.J., and King, L. (1972). A method for calculating codon frequencies in DNA. Journal of Theoretical Biology *35*, 399–457. https://doi.org/10.1016/0022-5193(72)90143-9.

Hanson, G., and Coller, J. (2018). Codon optimality, bias and usage in translation and mRNA decay. Nat Rev Mol Cell Biol *19*, 20–30. https://doi.org/10.1038/nrm.2017.91.

Hanson, G., Alhusaini, N., Morris, N., Sweet, T., and Coller, J. (2018). Translation elongation and mRNA stability are coupled through the ribosomal A-site. RNA *24*, 1377–1389. https://doi.org/10.1261/rna.066787.118.

Herrick, D., Parker, R., and Jacobson, A. (1990). Identification and comparison of stable and unstable mRNAs in Saccharomyces cerevisiae. Mol Cell Biol *10*, 2269–2284. https://doi.org/10.1128/mcb.10.5.2269-2284.1990.

Hia, F., and Takeuchi, O. (2021). The effects of codon bias and optimality on mRNA and protein regulation. Cell. Mol. Life Sci. *78*, 1909–1928. https://doi.org/10.1007/s00018-020-03685-7.

Hou, W., Harjono, V., Harvey, A.T., Subramaniam, A.R., and Zid, B.M. (2023). Quantification of elongation stalls and impact on gene expression in yeast. 2023.03.19.533377. https://doi.org/10.1101/2023.03.19.533377.

Hu, W., Sweet, T.J., Chamnongpol, S., Baker, K.E., and Coller, J. (2009). Co-translational mRNA decay in Saccharomyces cerevisiae. Nature *461*, 225–229. https://doi.org/10.1038/nature08265.

Ikeuchi, K., Tesina, P., Matsuo, Y., Sugiyama, T., Cheng, J., Saeki, Y., Tanaka, K., Becker, T., Beckmann, R., and Inada, T. (2019). Collided ribosomes form a unique structural interface to induce Hel2-driven quality control pathways. The EMBO Journal *38*, e100276. https://doi.org/10.15252/embj.2018100276.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. Science *324*, 218–223. https://doi.org/10.1126/science.1168978.

Jaquet, V., Wallerich, S., Voegeli, S., Túrós, D., Viloria, E.C., and Becskei, A. (2022). Determinants of the temperature adaptation of mRNA degradation. Nucleic Acids Res *50*, 1092–1110. https://doi.org/10.1093/nar/gkab1261.

Letzring, D.P., Wolf, A.S., Brule, C.E., and Grayhack, E.J. (2013). Translation of CGA codon repeats in yeast involves quality control components and ribosomal protein L1. RNA *19*, 1208–1217. https://doi.org/10.1261/rna.039446.113.

Liu, Y., Yang, Q., and Zhao, F. (2021). Synonymous but Not Silent: The Codon Usage Code for Gene Expression and Protein Folding. Annual Review of Biochemistry *90*, 375–401. https://doi.org/10.1146/annurev-biochem-071320-112701.

LJUNG, G.M., and BOX, G.E.P. (1978). On a measure of lack of fit in time series models. Biometrika *65*, 297–303. https://doi.org/10.1093/biomet/65.2.297.

Matsuo, Y., Tesina, P., Nakajima, S., Mizuno, M., Endo, A., Buschauer, R., Cheng, J., Shounai, O., Ikeuchi, K., Saeki, Y., et al. (2020). RQT complex dissociates ribosomes collided on endogenous RQC substrate SDD1. Nat Struct Mol Biol *27*, 323–332. https://doi.org/10.1038/s41594-020-0393-9.

Mauger, D.M., Cabral, B.J., Presnyak, V., Su, S.V., Reid, D.W., Goodman, B., Link, K., Khatwani, N., Reynders, J., Moore, M.J., et al. (2019). mRNA structure regulates protein expression through changes in functional half-life. Proceedings of the National Academy of Sciences *116*, 24075–24083. https://doi.org/10.1073/pnas.1908052116.

Meydan, S., and Guydosh, N.R. (2020). Disome and Trisome Profiling Reveal Genome-wide Targets of Ribosome Quality Control. Molecular Cell *79*, 588-602.e6. https://doi.org/10.1016/j.molcel.2020.06.010.

Monaghan, L., Longman, D., and Cáceres, J.F. (2023). Translation-coupled mRNA quality control mechanisms. The EMBO Journal *n/a*, e114378. https://doi.org/10.15252/embj.2023114378.

Park, K.I. (2017). Fundamentals of Probability and Stochastic Processes with Applications to Communications (Springer).

Parker, R. (2012). RNA Degradation in Saccharomyces cerevisae. Genetics *191*, 671–702. https://doi.org/10.1534/genetics.111.137265.

Parvathy, S.T., Udayasuriyan, V., and Bhadana, V. (2022). Codon usage bias. Mol Biol Rep *49*, 539–565. https://doi.org/10.1007/s11033-021-06749-4.

Passmore, L.A., and Coller, J. (2022). Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. Nat Rev Mol Cell Biol *23*, 93–106. https://doi.org/10.1038/s41580-021-00417-y.

Passos, D.O., and Parker, R. (2008). Chapter 20 Analysis of Cytoplasmic mRNA Decay in Saccharomyces cerevisiae. In Methods in Enzymology, (Academic Press), pp. 409–427.

Peccarelli, M., and Kebaara, B.W. (2014). Regulation of Natural mRNAs by the Nonsense-Mediated mRNA Decay Pathway. Eukaryot Cell *13*, 1126–1135. https://doi.org/10.1128/EC.00090-14.

Pechmann, S., and Frydman, J. (2013). Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. Nat Struct Mol Biol *20*, 237–243. https://doi.org/10.1038/nsmb.2466.

Plotkin, J.B., and Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. Nat Rev Genet *12*, 32–42. https://doi.org/10.1038/nrg2899.

Presnyak, V., Alhusaini, N., Chen, Y.-H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R., et al. (2015). Codon optimality is a major determinant of mRNA stability. Cell *160*, 1111–1124. https://doi.org/10.1016/j.cell.2015.02.029.

Radhakrishnan, A., Chen, Y.-H., Martin, S., Alhusaini, N., Green, R., and Coller, J. (2016). The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation by Monitoring Codon Optimality. Cell *167*, 122-132.e9. https://doi.org/10.1016/j.cell.2016.08.053.

Rahaman, S., Faravelli, S., Voegeli, S., and Becskei, A. (2023). Polysome propensity and tunable thresholds in coding sequence length enable differential mRNA stability. Science Advances *9*, eadh9545. https://doi.org/10.1126/sciadv.adh9545.

Raveh, A., Margaliot, M., Sontag, E.D., and Tuller, T. (2016). A model for competition for ribosomes in the cell. Journal of The Royal Society Interface *13*, 20151062. https://doi.org/10.1098/rsif.2015.1062.

Reis, M. dos, Savva, R., and Wernisch, L. (2004). Solving the riddle of codon usage preferences: a test for translational selection. Nucleic Acids Res *32*, 5036–5044. https://doi.org/10.1093/nar/gkh834.

Ross, J. (1995). mRNA stability in mammalian cells. Microbiol Rev *59*, 423–450. https://doi.org/10.1128/mr.59.3.423-450.1995.

Sharma, A.K., Venezian, J., Shiber, A., Kramer, G., Bukau, B., and O'Brien, E.P. (2021). Combinations of slow-translating codon clusters can increase mRNA half-life in Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences *118*, e2026362118. https://doi.org/10.1073/pnas.2026362118.

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

Shoemaker, C.J., and Green, R. (2012). Translation drives mRNA quality control. Nat Struct Mol Biol *19*, 594–601. https://doi.org/10.1038/nsmb.2301.

Simms, C.L., Yan, L.L., and Zaher, H.S. (2017). Ribosome Collision Is Critical for Quality Control during No-Go Decay. Molecular Cell *68*, 361-373.e5. https://doi.org/10.1016/j.molcel.2017.08.019.

Simms, C.L., Yan, L.L., Qiu, J.K., and Zaher, H.S. (2019). Ribosome Collisions Result in +1 Frameshifting in the Absence of No-Go Decay. Cell Rep *28*, 1679-1689.e4. https://doi.org/10.1016/j.celrep.2019.07.046.

Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., and Pilpel, Y. (2010). An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation. Cell *141*, 344–354. https://doi.org/10.1016/j.cell.2010.03.031.

Veltri, A.J., D'Orazio, K.N., Lessen, L.N., Loll-Krippleber, R., Brown, G.W., and Green, R. (2022). Distinct elongation stalls during translation are linked with distinct pathways for mRNA degradation. eLife *11*, e76038. https://doi.org/10.7554/eLife.76038.

Webster, M.W., Chen, Y.-H., Stowell, J.A.W., Alhusaini, N., Sweet, T., Graveley, B.R., Coller, J., and Passmore, L.A. (2018). mRNA Deadenylation Is Coupled to Translation Rates by the Differential Activities of Ccr4-Not Nucleases. Molecular Cell *70*, 1089-1100.e8. https://doi.org/10.1016/j.molcel.2018.05.033.

Weinberg, D.E., Shah, P., Eichhorn, S.W., Hussmann, J.A., Plotkin, J.B., and Bartel, D.P. (2016). Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation. Cell Reports *14*, 1787–1799. https://doi.org/10.1016/j.celrep.2016.01.043.

9 Appendix

9.1 Identification of periodic patterns in CSC profiles: autocorrelation analysis

Various features of the CSC profiles of mRNAs can impact mRNA stability. We were particularly interested in investigating whether periodicity in the CSC profiles was a feature of the yeast transcriptome that could influence mRNA stability. To explore this, we performed an autocorrelation analysis on 4214 genes to detect the presence of periodic patterns in the CSC profiles along these genes. Autocorrelation is defined as the correlation, generally the Pearson correlation coefficient, between a variable and a lagged version of itself as a function of the lag. This allows us to assess periodic patterns since high correlations at a given lag imply that the values of the variable are strongly related to its past values in time or space and are thus not random (Park, 2017).

We represented each gene as a vector of the CSC values of its codons to conduct the autocorrelation analysis. An example of the autocorrelation calculation for the *HSP60* gene is shown in Figure 23A for fifty codon lags. To assess the statistical significance of these autocorrelations, we used the Ljung-Box test. This test evaluates whether the lagged values of a variable exhibit autocorrelation. It tests the null hypothesis that the data are independently distributed and the alternate hypothesis that the data exhibit autocorrelation. The test computes a Q statistic based on autocorrelations at different lag values and compares it to a chi-squared distribution, whose degrees of freedom are dependent on the number of lags being tested (LJUNG and BOX, 1978).

Across different lag values tested, we noted that the maximum correlation for a given lag for genes displaying significant autocorrelation as determined by the Ljung-box test was quite low (Figure 23B, C and D). Moreover, the mean correlation across all observed lags was less than 0.1. Since the correlation for most lags is low, a significant autocorrelation at a few lags in the considered range can make the Q statistic statistically significant. However, considering the low mean and maximum correlations observed across all lags, we concluded that autocorrelation was likely not a prevalent feature of the CSC profile of endogenous genes and might not offer

substantial insights into general mechanisms modulating mRNA half-life through codon composition.

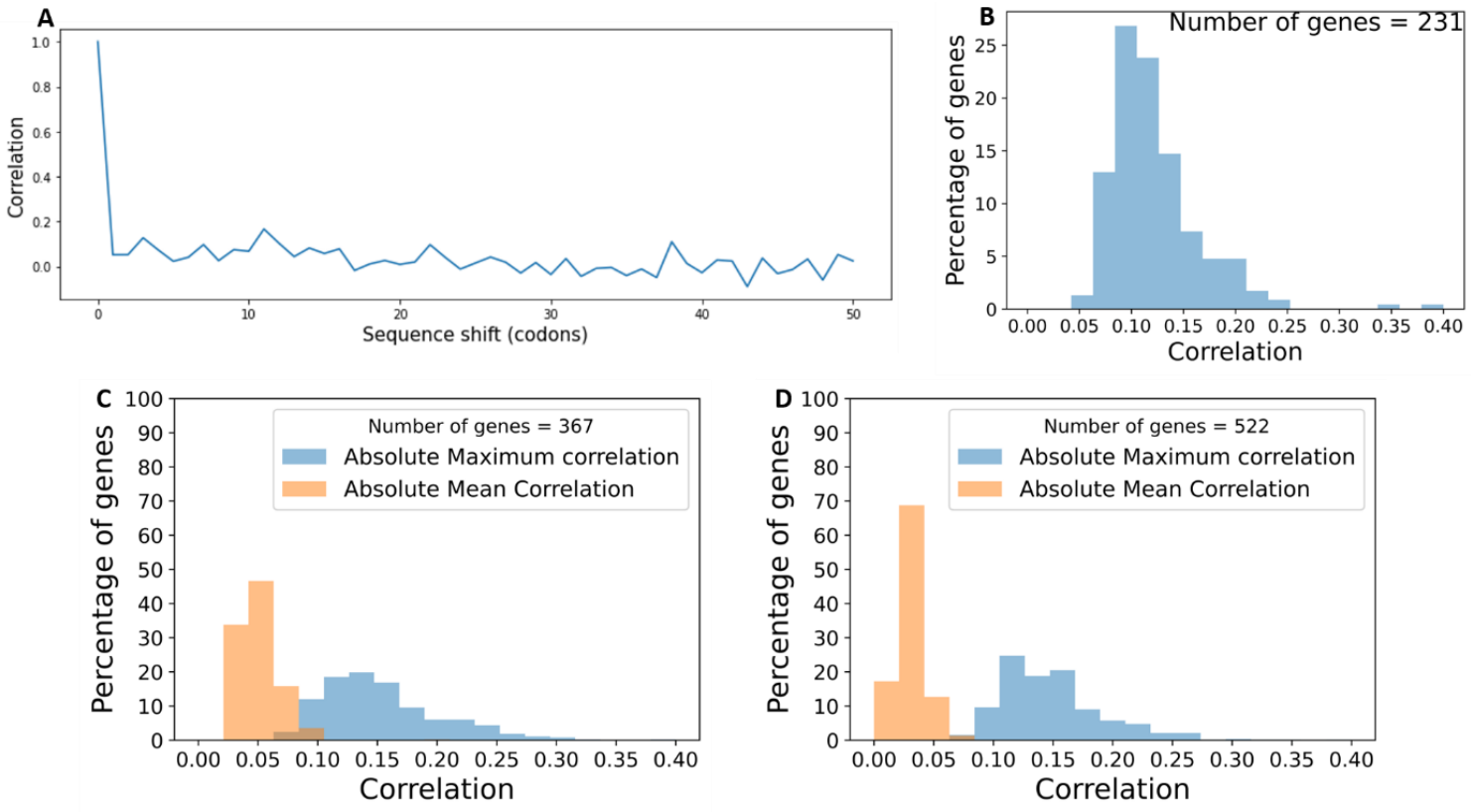

Figure 23: Autocorrelation analysis of the yeast transcriptome

*(**A**) An example autocorrelation calculation for the HSP60 gene, calculated up to fifty codon lags. (**B**) Frequency distribution of the autocorrelation for all genes displaying significant autocorrelation for a lag value of one. (**C**) Frequency distribution of absolute mean and absolute maximum autocorrelation for all genes displaying significant autocorrelation up to fifty lags. (**D**) Frequency distribution of the absolute mean and absolute maximum autocorrelation for all genes displaying significant autocorrelation up to a lag value equal to the gene length. The absolute maximum represents the maximum magnitude of correlation observed across all calculated lags, and the absolute mean represents the mean of the absolute correlations for all calculated lags. Statistical analysis was performed using the Ljung-Box test ($\alpha < 0.05$).*

9.2 Ribosome profiling

Ribosome profiling is a method used to map, with nucleotide resolution, the position of ribosomes along transcripts. The technique uses the fact that ribosomes shield a region of more than one codon (known as its ‘footprint’) from RNases. After blocking translation elongation using elongation inhibitors like cycloheximide or flash-freezing

cells, mRNAs are digested with RNase, and the protected fragments are captured and sequenced, allowing one to pinpoint the location of ribosomes on a transcript at a given moment (Hia and Takeuchi, 2021; Ingolia et al., 2009).

9.3 Plasmids and qPCR primers used in this study

Table 1: Plasmids and qPCR primers

Plasmid	Description	Forward Primer	Reverse Primer	Primer efficiency
pSV_840	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(1/4 U – 3/4 S)**	TGCCGGAGGGATATGTACAG	ACCGTCTTCCTTGAAGTCGA	0.75
pSV_841	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(1/4 S – 3/4 U)**	GTCAAGTTCGAAGGTGATACACT	TCCATTTTTCTGTTTATCCGCCA	0.95
pSV_842	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(1/8 U – 3/4 S – 1/8 U)**	GTACCCAGAAGACGGTGTCT	TATACGCTCCCGGCATCTG	1.15
pSV_843	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-**	GCTGACAAGCAAAAGAACGG	CGGTCCATCTCCTATCGGTG	1.07

	BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(3/8 S – 1/4 U – 3/8 S)**			
pSV_844	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(1/8 U – 7/8 S)**	TGCACAACAGGAAAACTTCCG	TGGACGTAACCTTCTGGCAT	0.92
pSV_845	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(1/8 S – 7/8 U)**	GTAAGTTGCCAGTCCCGTG	CTCTCCTGTACATATCCCTCCG	0.87
pSV_846	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(1/16 U – 7/8 S – 1/16 U)**	AATGCCAGGTGCTTACAACG	TCCTGTTGAATGCCTTCCCT	1.17
pSV_847	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(7/16 S – 1/8 U – 7/16 S)**	GTCTTGTTGCCAGACAACCA	CGCCGCTGTTACAAACTCA	1.05
pSV_848	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-**	ACCGATACTTGTAGAGCTTGATG	CATGGGACTGGCAACTTACC	0.91

	Stop_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(1/16 U – 15/16 S)**			
pSV_849	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(1/16 S – 15/16 U)**	GTCAACGGTCACAAATTTTCAGT	GTTACAAGTGTCGGCCACG	0.95
pSV_850	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(1/32 U – 15/16 S – 1/32 U)**	AACGAAGACTACACCGTCGT	ATAAAGCTCATCCATTCCTCCTG	1.13
pSV_851	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(15/32 S – 1/16 U – 15/32 S)**	GTTGGAGTTCGTAACAGCGG	CATGTGGACCTTGAAACGCA	1.12
pSV_902	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(1/2 S - 1/2 U)**	CTGCTGGTATCACCTTGGGT	TGTCCATTCATTGATCCCTCCA	1.12
pSV_903	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp	TTGAGTTTGTAACAGCGGCG	GTGGACCTTGAAACGCATGA	0.99

	YLR061W_XmaI-Ter-NotI **(1/2 U - 1/2 S)**			
pSV_904	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(3/4 S – 1/4 U)**	CGTCACCCAAGACACCTCT	TCCATCCTCCGGATAAAGCC	0.93
pSV_905	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(3/4 U – 1/4 S)**	AGCGGTAACAGTAACACAGGA	TTTGCATGACTGGACCGTCT	0.98
pSV_906	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(1/4 S – 1/2 U – 1/4 S)**	ACTACAAGACCCGTGCTGAA	TGTCCAAGTATATTTCCATCCTCTT	1.07
pSV_907	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(1/4 U – 1/2 S – 1/4 U)**	GGAGGGATATGTACAGGAGAGG	ACCGTCTTCCTTGAAGTCGAT	1.03
pSV_975	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp	AGGGTGACATCAAGATGGCT	TTATACGCTCCCGGCATCTG	0.96

	YLR061W_XmaI-Ter-NotI **(7/8 S – 1/8 U)**			
pSV_976	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(15/16 S – 1/16 U)**	AAGGCTAAGAAGCCAGTCCA	CCTCTGACCTCTCATACTGCT	0.94
pSV_977	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(31/32 S – 1/32 U)**	CAACGAAGACTACACCGTCG	AGCTCATCCATTCCTCCTGT	0.97
pSV_978	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(1/32 U – 31/32 S)**	GGAGAGGAGCTTTTTACAGGAG	AGAGAACTTGTGACCGTTGAC	0.95
pSV_979	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_Radhakrishnan2016_BamHI_eGFP-stable_EcoRI_mScarlet-I-stable_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI	ACAATGGGATCCTTAATAGCGC	CACCGTCCAATTCGACCAAG	0.93
pSV_980	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP-stable_EcoRI_Radhakrishnan2016_EcoRI_mScarlet-I-**	CTTGTTGGAGTTCGTCACCG	GAACCGAATTCCGTCGCC	0.96

	stable_KasI-Stop_3'UTR+100bp YLR061W_XmaI-Ter-NotI			
pSV_981	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP-stable_EcoRI_mScarlet-I-stable_KasI-Radhakrishnan2016-KasI_Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI	TGGCGCCTTAATAGCGCG	GAATCATTCATTAAACATATTTATTAGGTTTAGAAG	0.75
pSV_982	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_CGA12_BamHI_eGFP-stable_EcoRI_mScarlet-I-stable_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI	AAACAATGGGATCCCGACGAC	TTCGACCAAGATTGGGACGA	0.88
pSV_983	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP-stable_EcoRI_CGA12_EcoRI_mScarlet-I-stable_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI	GTTCTTCTGAATTCCGACGACG	CCTTGATGACAGCTTCACCC	0.99
pSV_984	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP-stable_EcoRI_mScarlet-I-stable_KasI-CGA12-KasI_Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI	GTTCTTCTGGCGCCCGAC	GGTTTAGAAGATTTTAATTATACAAAACTTATGCTC	0.86
pSV_985	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_SDD1_BamHI_eGFP-stable_EcoRI_mScarlet-I-**	GATTGCAGAGCGAAACGACG	CACCGTCCAATTCGACCAAG	0.94

	stable_KasI-Stop_3'UTR+100bp YLR061W_XmaI-Ter-NotI			
pSV_986	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP-stable_EcoRI_SDD1_EcoRI_mScarlet-I-stable_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI	TTATTTGATTTTCGATTGCAGAGCG	GACAGAAGAACCGAATTCTTTCCTC	0.91
pSV_987	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP-stable_EcoRI_mScarlet-I-stable_KasI-SDD1-KasI_Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI	TGAAGGTCGTCACTCTACCG	GAATTTAGGCGCCTTTCCTCTTCA	0.95
pSV_988	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_Nonbasic_negative_BamHI_eGFP-stable_EcoRI_mScarlet-I-stable_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI	CAATGGGATCCTCGATAATACTGC	TTCGACCAAGATTGGGACGA	0.88
pSV_989	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP-stable_EcoRI_Nonbasic_negative_EcoRI_mScarlet-I-stable_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI	CTTGTTGGAGTTCGTCACCG	GAACCGAATTCAAGATTTATCGACG	0.92
pSV_990	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP-stable_EcoRI_mScarlet-I-**	TGAAGGTCGTCACTCTACCG	TGAATTTAGGCGCCAAGATTTATCG	0.97

	stable_KasI-Nonbasic_negative-KasI_Stop_3'UTR+100bp YLR061W_XmaI-Ter-NotI			
pSV_991	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR-**PDC5**-STOP_KasI_3'UTR+100bp YLR061W_XmaI-Ter-NotI	TGTTGCCAGTCTTTGATGCTC	GGTTTAGAAGATTTTAATTATACAAAACTTATGCTC	0.90
pSV_992	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR-**PIR3**-STOP_KasI_3'UTR+100bp YLR061W_XmaI-Ter-NotI	CACATTGGTTCTCAGTGCCATG	GGTTTAGAAGATTTTAATTATACAAAACTTATGCTC	0.91
pSV_993	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR-**PIR1**-STOP_KasI_3'UTR+100bp YLR061W_XmaI-Ter-NotI	ATGCAGTCCACCTACAAGCTATC	GGTTTAGAAGATTTTAATTATACAAAACTTATGCTC	0.91
pSV_994	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR-**HXT3**-STOP_KasI_3'UTR+100bp YLR061W_XmaI-Ter-NotI	GGTGCTAACTACGATGCTGATG	GGTTTAGAAGATTTTAATTATACAAAACTTATGCTC	0.88
pSV_995	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR-**TIR2**-STOP_KasI_3'UTR+100bp YLR061W_XmaI-Ter-NotI	GCCAAGGCTGTCATCGGT	GGTTTAGAAGATTTTAATTATACAAAACTTATGCTC	0.92
pSYN115	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_ 369 bps truncated eGFP (non-optimal: -0.3 5' ORF, 0.1 3'ORF)_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI	GGAGGAGTTATTTACGGGGGT	CCTCTCCCTCCCCTGATACT	0.98

pSYN116	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_ 369 bps truncated eGFP (optimal: 0.3 5' ORF, 0.1 3'ORF)_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI	ACAAGTTCTCCGTTTCCGGC	CGCCGTAAGTCAAAGTAGTCACT	0.91
pSYN117	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_ 369 bps truncated eGFP (non-optimal: homogeneous)_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI	GCGTGGTGCCCATTTTAGTG	CAGGGCACGGGTAATTTGC	0.92
pSYN118	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_ 369 bps truncated eGFP (optimal: homogeneous)_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI	CGATGCCTGAAGGCTACGTT	TCGAACTTAACTTCCGCACGG	0.86
pSYN119	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_ 369 bps truncated eGFP (non-optimal: 0.1 5' ORF, -0.3 3'ORF)_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI	TACAGTGCTTTTCACGCTATCC	CCGCTCCTGTACATACCCC	0.93
pSYN120	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_ 369 bps truncated eGFP (optimal: 0.1 5' ORF, 0.3 3'ORF)_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI	ATTCACTGGCGTGGTGCC	AGGCACAGGTAACTTGCCA	0.88

pSYN121	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR-**PDC5 (recoded)**-STOP_KasI-Stop_3'UTR+100bp YLR061W_XmaI-Ter-NotI	TTGCCAGTTTTCGACGCC	GGTTTAGAAGATTTTAATTATACAAAACTTATGCTC	0.91
pSYN122	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR-**PIR3 (recoded)**-STOP_KasI-Stop_3'UTR+100bp YLR061W_XmaI-Ter-NotI	CACATTGGCTCCCAATGTCAC	GGTTTAGAAGATTTTAATTATACAAAACTTATGCTC	0.76
pSYN123	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR-**PIR1 (recoded)**-STOP_KasI-Stop_3'UTR+100bp YLR061W_XmaI-Ter-NotI	GCACTCAATGTAACGCTGTTC	GGTTTAGAAGATTTTAATTATACAAAACTTATGCTC	0.91
pSYN124	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR-**HXT3 (recoded)**-STOP_KasI-Stop_3'UTR+100bp YLR061W_XmaI-Ter-NotI	CGAACTACGACGCAGACG	GGTTTAGAAGATTTTAATTATACAAAACTTATGCTC	0.92
pSYN125	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR-**TIR2 (recoded)**-STOP_KasI-Stop_3'UTR+100bp YLR061W_XmaI-Ter-NotI	ATGGGCGCTGGCGTTATG	GGTTTAGAAGATTTTAATTATACAAAACTTATGCTC	0.93
pSYN126	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(homogenised 1/32U)**	AAGGTAAAGTTGCGTGGTACC	CTTCTGGGTACAAACGCTCG	1.10

pSYN127	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(homogenised 1/16U)**	AAGGTCAAGTTGCGTGGTACA	ACCGTCTTCTGGGTACAAACG	0.77
pSYN128	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(homogenised 1/8U)**	ACATCTTGGGACACAAACTTGAA	CCAATTGGACTGAACCGTCCTC	0.89
pSYN129	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(homogenised 1/4U)**	CTTATAACGTCGACAGGAAGCTTGA	TCCATACCACCTGTTGAATGACG	0.76
pSYN130	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(homogenised 1/2U)**	GACGGACCAGTATTGTTGCC	GATCACGCTTCTCATTCGGG	0.95
pSYN131	pRS303-ApaI-Fig1-XhoI-Tgal7-AvrII-[tetO2]4inPgal1-SphI-5'UTR YLR061W-**ATG-BamHI_eGFP_EcoRI_mScarlet-I_KasI-Stop**_3'UTR+100bp YLR061W_XmaI-Ter-NotI **(homogenised 3/4U)**	CCGGGAGCGTACAATGTAGA	CTTCAGAACGCTCGTACTGC	0.77