

Diversity in mRNA closed-loop formation during translation

A thesis submitted for the degree of

Doctor of Philosophy

of

The Australian National University

Yu-Ting (Jennifer) Hung

September 2023



**Australian
National
University**

Originality Statement

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at ANU or any other educational institution, except where due acknowledgement is made in this thesis. Any contribution made to the research by others, with whom I have worked at ANU or elsewhere, is explicitly acknowledged in this thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

Signed ...Yu-Ting Hung.....

Date ...26th September 2023.....

Acknowledgements

This PhD experience and the writing of this thesis have been made possible by the many cherished mentors, friends and family that have supported me through this journey.

Professor Thomas Preiss

Thank you for the time, energy and resources you have dedicated as my supervisor to helping me finish this thesis and PhD. Your scientific expertise, wise advice and practical supervision has taught me how to think critically, ask good scientific questions and interpret results thoroughly. Thank you as well for graciously putting up with our conference antics and never complaining about all the parties we held while you were trying to sleep.

Dr Nikolay Shirokikh

Thank you for your co-supervision, your valuable input and scientific advice. Your scientific and technical understanding has been instrumental in the development of this project.

My panel – Prof Eduardo Eyras and Prof Leonie Quinn

Thank you for your encouragement, important insights, advice and suggestions.

Dr Attila Horvath

The bioinformatics of this project would not have been possible without your expertise and biological understanding. Thank you for your time and dedication in making this project happen, and for always contributing great ideas and suggestions.

Dr Yoshika Janapala

I cannot thank you enough for the technical guidance, encouragement and mentoring in and out of the lab. You have taught me all the lab skills I needed to know for my project, answered all my questions, and have done all this with the patience of a saint. Thank you for all the words of encouragement and advice outside of the lab too.

Dr Lithin Karmel Louis

What would we do without the wise advice of our resident grandpa? Thank you for being such a calming presence, always ready to give us practical advice and remind us to ‘just go home’ when it’s been a rough day.

Ms Madhu Kanchi

Thanks for keeping the lab running and always looking after me in the lab. Your infectious laugh and bubbly personality have made the lab so much livelier.

Mr Marco Guarnacci

The best PhD buddy I could’ve asked for! I’ve been lucky to have you to go through this PhD with. Thanks for all the support, rant sessions and procrastination chats. You’ve made the sometimes-lonely PhD journey much less so.

The rest of the Preiss Lab

Thank you to everyone that has come and gone from the Preiss lab during my time here – Yiang, Sona, Sarah, Kamel, Shen, Hanrong and Mojdeh for making the lab a fun and pleasant place to be.

Dr Hardip Patel

Thank you for the fun lunchtime conversations, drinks at Badger, support and wise advice about science and academia.

The many FROGS (Friends of Genome Science)

Yoona, Renzo, Pablo, AJ, Agin, Kat, Nazmul, some already mentioned and many others too – thank you for the fun times and great memories. I know that when I look back on this PhD, it's the crazy parties, laughter-filled hangouts and mutual support that I will remember most.

My family

Thank you for your never-ending love and support that has gotten me here today. Mum, dad, my sister Wendy and brother Kenny – together you have given me all the best opportunities in my academic pursuits and I can't thank you enough.

爸爸、媽媽、哥哥、姐姐，感謝你們一直以來的愛與支持。我今天能完成我的博士研究，完全都是因為你們。

My favourite person, Sam

Throughout this entire PhD journey, your unwavering love and support has been my biggest source of joy and comfort. Thank you for everything you have done for me – from the late-night pick-ups from the lab, to patiently listening to me talk about my day, to always being there for me through the difficult moments. You've been my biggest supporter and you've always believed in me even when I didn't. I'm beyond lucky to have you. Thank you as well for putting up with thesis-writing Jen, I promise she's gone now.

Abstract

A key feature of eukaryotic translation initiation is the ‘closed-loop’ model, which describes a process where the ends of an mRNA molecule, namely the 5’ cap and the 3’ poly(A) tail, are circularised by a bridge of protein factors - eIF4E, eIF4G and Pab1. Formation of this structure has long been proposed as an explanation for how mRNA end proximity might enhance translation. However, although initial studies concluded an important involvement of the closed-loop in productive translation, this has been challenged by more recent *in vivo* findings reporting low levels of closed-loop formation on some translating mRNAs and the potential for mRNA-dependent closed-loop formation.

To better understand characteristics of the mRNA closed-loop, we have developed an *in vivo* approach in *Saccharomyces cerevisiae* to probe the closed-loop transcriptome-wide. Combining formaldehyde crosslinking, immunoprecipitation, limited nuclease digestion and short-read RNA-sequencing, we captured the closed-loop and associated mRNAs. We then devised a bioinformatic method to define mRNA extremities and compare the relative association of mRNA ends to the closed-loop factors as a measure for an intact closed-loop protein bridge.

From our transcriptome-wide analyses, we find that the closed-loop likely can form *in vivo*. We are able to observe examples of mRNAs with significant enrichments at both ends upon immunoprecipitation with the 5’ cap-associated closed-loop factor eIF4E. However, it is noteworthy that the closed-loop does not appear to be a consistently observed phenomenon at steady state. Instead, our research demonstrates that the extent of closed-loop formation varies greatly across the transcriptome. Some mRNAs exhibit a ‘high closed-loop forming’ pattern, characterised by nearly equal association of both ends with the closed-loop factors, while others demonstrate a ‘low closed-loop forming’ profile, with a more unequal distribution of their ends in relation to these factors.

Correlating the extent of closed-loop formation propensity to physical and functional mRNA features such as translation efficiency, we do not find a strong relationship between the various features tested and extent of closed-loop formation. Therefore, our findings prompt a more nuanced consideration of the role of the closed-loop in translation, and provide a methodology for further investigation of the closed-loop’s biological functions.

Abbreviations

4E-BP	4E binding protein
APA	Alternative polyadenylation
CAT	Chloramphenicol acetyltransferase
CDS	Coding sequence
CHO	Chinese hamster ovary
CPSF	Cleavage and polyadenylation specificity factor
DSN	Double strand nuclease
eIF4E	Eukaryotic translation initiation factor 4E
eIF4G	Eukaryotic translation initiation factor 4G
eIFs	Eukaryotic initiation factors
IFs	Initiation factors
IRES	Internal ribosome entry site
LSU	Large subunit
LUC	Luciferase
miRNA	microRNA
mRNP	Messenger ribonucleoprotein
MW	Molecular weight
ORF	Open reading frame
Pab1	Poly(A) binding protein
PIC	Pre-initiation complex
RRM	RNA recognition motif
ssRNA	Single-stranded RNA
SSU	Small subunit
TCP-seq	Translation complex profile sequencing
TE	Translation efficiency
TSS	Transcription start site
TES	Transcription end site
WCL	Whole cell lysate
WT	Wild type

1	INTRODUCTION	12
1.1	Overview of mRNA Translation	12
1.1.1	Translation initiation	13
1.1.2	Yeast as a model organism for studying eukaryotic translation	16
1.2	Development of the closed-loop model	17
1.2.1	Key mRNA components involved in the closed-loop	18
1.2.2	The protein components making up the closed-loop and their interactions.....	22
1.2.3	Hypothesised functions of the closed-loop in translation initiation.....	28
1.3	Challenges to the view of the closed-loop being a predominant feature of translation	29
1.3.1	Viability of strains with closed-loop mutations or deletions	30
1.3.2	Proposed alternative closed-loop conformations	30
1.3.3	Single molecule studies visualising ends.....	31
1.3.4	Ends being intrinsically close together.....	31
1.3.5	Summary of various alternative models compared to the closed-loop model	32
1.4	Could the closed-loop formation be mRNA dependent and how?	32
1.4.1	Differential closed-loop association and formation between mRNAs.....	33
1.4.2	Possibility for mRNA-specific regulation of the closed-loop.....	34
1.4.3	Potential features determining closed-loop formation	35
1.5	Hypothesis and Aims.....	37
2	MATERIALS AND METHODS.....	39
2.1	Materials	39
2.1.1	Chemicals/reagents.....	39
2.1.2	Buffers and solutions.....	41
2.1.3	Antibodies	42
2.1.4	Yeast strains	43
2.1.5	Primers and oligonucleotides.....	44
2.1.6	Depletion probes for probe-directed degradation	46
2.2	Methods.....	47
2.2.1	Commonly used methods	47
2.2.2	Yeast tagging and mutation	51
2.2.3	Growth curve measurements	55
2.2.4	Bead binding capacity	55
2.2.5	Crosslinking optimization	57
2.2.6	Rnase I digestion optimization	57
2.2.7	Growing, crosslinking and grindate preparation.....	58
2.2.8	Immunoprecipitation and RNA extraction	58
2.2.9	Initial sequencing	60
2.2.10	Library preparation of final libraries.....	60
2.2.11	Sequencing	65
2.2.12	Initial bioinformatic processing of sequencing files	65
2.2.13	De novo transcription unit predictions.....	66
2.2.14	Calculation of end and middle signals	66
2.2.15	Filtering of individual mRNAs	66
2.2.16	Calculation of closed-loop score.....	67
2.2.17	Clustering	67
2.2.18	Correlation analyses.....	67

3	DEVELOPMENT OF AN EXPERIMENTAL METHOD TO CAPTURE THE CLOSED-LOOP IN VIVO	70
3.1	Ensuring visualisation of the closed-loop proteins	71
3.1.1	Lack of antibodies against the closed-loop proteins in yeast	72
3.1.2	Tagging of closed-loop components with 3xFLAG	74
3.1.3	Confirmation of 3xFLAG tag incorporation	79
3.1.4	Growth curves	81
3.2	Optimising immunoprecipitation conditions	81
3.2.1	Bead selection	82
3.2.2	Formaldehyde crosslinking optimisation	85
3.2.3	RNase I digestion optimisation	88
3.3	Recreation of results from original study	91
3.3.1	Co-immunoprecipitation of closed-loop components	91
3.3.2	mRNA end enrichment analysis by qPCR	92
3.4	Chapter 3 Conclusion	95
4	DEVELOPMENT OF AN RNA-SEQ METHOD AND BIOINFORMATIC PIPELINE FOR TRANSCRIPTOME-WIDE ANALYSIS OF CLOSED-LOOP FORMATION	97
4.1	Choice of library preparation and sequencing strategy	98
4.1.1	Initial sequencing strategy	98
4.1.2	New library preparation approach for RNA-seq	103
4.2	Defining transcription units de novo	109
4.2.1	Collating list of ‘high confidence’ examples	109
4.2.2	Development of workflow	110
4.2.3	Validation of de novo transcription unit predictions	115
4.3	Reproducibility of replicates	116
4.3.1	Reproducibility between replicates	118
4.3.2	Comparison within replicates	121
4.4	Chapter 4 conclusions	122
5	INVESTIGATING MRNA FEATURES AFFECTING THE CLOSED-LOOP ...	124
5.1	Snapshot of the closed-loop in living cells	125
5.1.1	Metagene plots – by read	126
5.1.2	Metagene plots, by gene	128
5.2	Development of closed-loop score	130
5.2.1	Start and end signals	131
5.2.2	Choice of middle region signal	132
5.2.3	Exclude low expressed genes + flat genes	134
5.2.4	Final score and split into groups (high 5’ UTR, u-shape, high 3’ UTR)	134
5.3	Variety in closed-loop between mRNAs	135
5.4	mRNA features affecting the closed-loop	137
5.4.1	Physical attributes of mRNA – CDS, UTR and overall transcript lengths	137

5.4.2	Regulatory and functional features vs closed-loop formation	140
5.5	Chapter 5 conclusions	146
6	DISCUSSION	150
6.1	Summary	150
6.2	Study approach and method development.....	151
6.3	The closed-loop is likely able to form in vivo	153
6.4	The pervasiveness and necessity of CL during translation	154
6.5	mRNA-specific formation of the closed-loop	155
6.6	Alternative roles for the closed-loop	157
6.7	Future directions.....	158
6.8	Conclusion and outlook.....	159
7	REFERENCES.....	161

List of Figures

Figure 1.1.	Overview of mRNA translation stages, assuming the closed-loop model, and with a focus on translation initiation.....	13
Figure 1.2.	The closed-loop model components and structure.....	18
Figure 1.3.	Binding of closed-loop proteins to each other.....	23
Figure 2.1.	Overall yeast tagging strategy for 3xFLAG insertion	52
Figure 3.1.	Overview of experimental method	71
Figure 3.2.	Trial of antibodies against endogenous yeast proteins.....	73
Figure 3.3.	Summary of double-tagged strains generated	74
Figure 3.4.	Schematic of overall tagging strategy	76
Figure 3.5.	Generation of 3xFLAG cassettes for transformation	77
Figure 3.6.	Example colony PCR to check for correct insertion of 3xFLAG tag.....	79
Figure 3.7.	Yeast strains before and after 3xFLAG tag incorporation	80
Figure 3.8.	Comparison of growth curves between WT and eIF4E-TAP, eIF4G1-3xFLAG strains	81
Figure 3.9.	Sheep vs Mouse IgG bead binding comparison	83
Figure 3.10.	Immunoprecipitation using the 3xFLAG tag	84
Figure 3.11.	Effect of formaldehyde percentage on capture of polysomes	86
Figure 3.12.	Effect of crosslinking percentage on co-immunoprecipitation of closed-loop components	88
Figure 3.13.	Effect of RNase I digestion on polysomal fraction	89
Figure 3.14.	Effect of RNase I digestion on mRNA recovery at ends.....	90
Figure 3.15.	Coimmunoprecipitation of closed-loop components.....	92
Figure 3.16.	mRNA enrichment at 5' and 3' ends upon 5' immunoprecipitation	94
Figure 4.1.	Initial quality control of sequencing libraries.....	99
Figure 4.2.	Sequencing statistics pre- and post- adapter trimming.....	100
Figure 4.3.	Overrepresented sequences as determined by FastQC.....	102

Figure 4.4. Mapping statistics	103
Figure 4.5. Overview of final library preparation approach taken	104
Figure 4.6. Final libraries obtained for sequencing	105
Figure 4.7. Overview of sequencing pre-processing steps	106
Figure 4.8. Number and distribution of reads mapping to different RNA species	108
Figure 4.9. The focus of the study is on sequences at the starts and ends of mRNAs.....	109
Figure 4.10. List of 'high confidence' individual examples	110
Figure 4.11. Overview of <i>de novo</i> transcription unit prediction approach	111
Figure 4.12. Comparing signal threshold values for defining signal islands.....	112
Figure 4.13. Comparison between a minimum length of 20 and 50 with a signal threshold of 5.....	112
Figure 4.14. 5' and 3' UTR length distribution from Nagalakshmi et al. (2008)	114
Figure 4.15. Removal of introns from transcription unit predictions	114
Figure 4.16. Lengths of predicted transcription units compared with CDS lengths.....	115
Figure 4.17. Comparison of predicted vs reference transcription unit lengths.....	116
Figure 4.18. Clustered spearman's correlation matrix of samples between replicates	118
Figure 4.19. Individual correlations between replicates	120
Figure 4.20. Comparison of samples within replicates.....	121
Figure 5.1. Summation of RNA-seq coverage at a per-read level, with all reads given equal weighting.....	127
Figure 5.2. Summation of RNA-seq coverage at a per-gene level, where normalised coverage profiles are first determined for each gene individually before aggregation	129
Figure 5.3. Approach for comparing mRNAs by their extent of relative end enrichment, as proxy for closed-loop formation.	130
Figure 5.4. Methodology for defining start and end signals	131
Figure 5.5. Examples of start and end signal predictions.	132
Figure 5.6. Options tested for defining the 'middle' signal.	133
Figure 5.7. Distribution of closed-loop measurements across the transcriptome	136
Figure 5.8. Correlation between closed-loop measure and mRNA features.....	139
Figure 5.9. Correlation between gene expression and closed-loop measure	141
Figure 5.10. Correlation between closed-loop factor associations and closed-loop measure	143
Figure 5.11. Comparison between mRNA half-life and closed-loop measure	144
Figure 5.12. Comparison between protein synthesis rate and closed-loop measure	146

List of Tables

Table 1. Chemicals and Reagents	39
Table 2. Buffers and Solutions.....	41
Table 3. Antibodies	42
Table 4. Yeast strains.....	43
Table 5. Tagging primers	44
Table 6. qPCR primers.....	44
Table 7. Primers for library preparation	45
Table 8. rRNA depletion probes	46
Table 9. qPCR master mix components.....	50

Chapter 1 – Introduction

1 Introduction

This introduction will begin with a general overview of mRNA translation, with a focus on eukaryotic translation. Then I will introduce the concept of the closed-loop and its various components. Finally, I will discuss various challenges to the closed-loop and the potential for closed-loop formation to be mRNA-dependent.

1.1 Overview of mRNA Translation

Understanding the closed-loop necessitates a general understanding of mRNA translation – the act of protein synthesis from mRNA. mRNA translation is a complex, fundamental process conserved across the kingdoms of life (Kapp and Lorsch, 2004; Sonenberg and Hinnebusch, 2009).

Briefly, mRNA translation consists of three main stages – initiation, elongation and termination (Cooper, 2000). Translation initiation refers to the recruitment of the ribosomal small subunit (SSU) to the mRNA by, and with, a host of initiation factor (IFs), followed by identification of the start codon and recruitment of the ribosomal large subunit (LSU) (Kozak, 1999; Jackson, Hellen and Pestova, 2010). Elongation then ensues, with the elongation-competent ribosome progressively moving along the mRNA. As the ribosome translocates, each mRNA codon is recognised by a specific tRNA through complementary base pairing, allowing the ribosome to progressively catalyse the sequential addition of amino acids to a growing polypeptide chain (Voorhees and Ramakrishnan, 2013). Upon the ribosome reaching the stop codon, termination occurs, in which the newly synthesised polypeptide is released, followed by sequential disassembly of the ribosome subunits in preparation for re-initiation of translation (Laurberg *et al.*, 2008; Hellen, 2018).

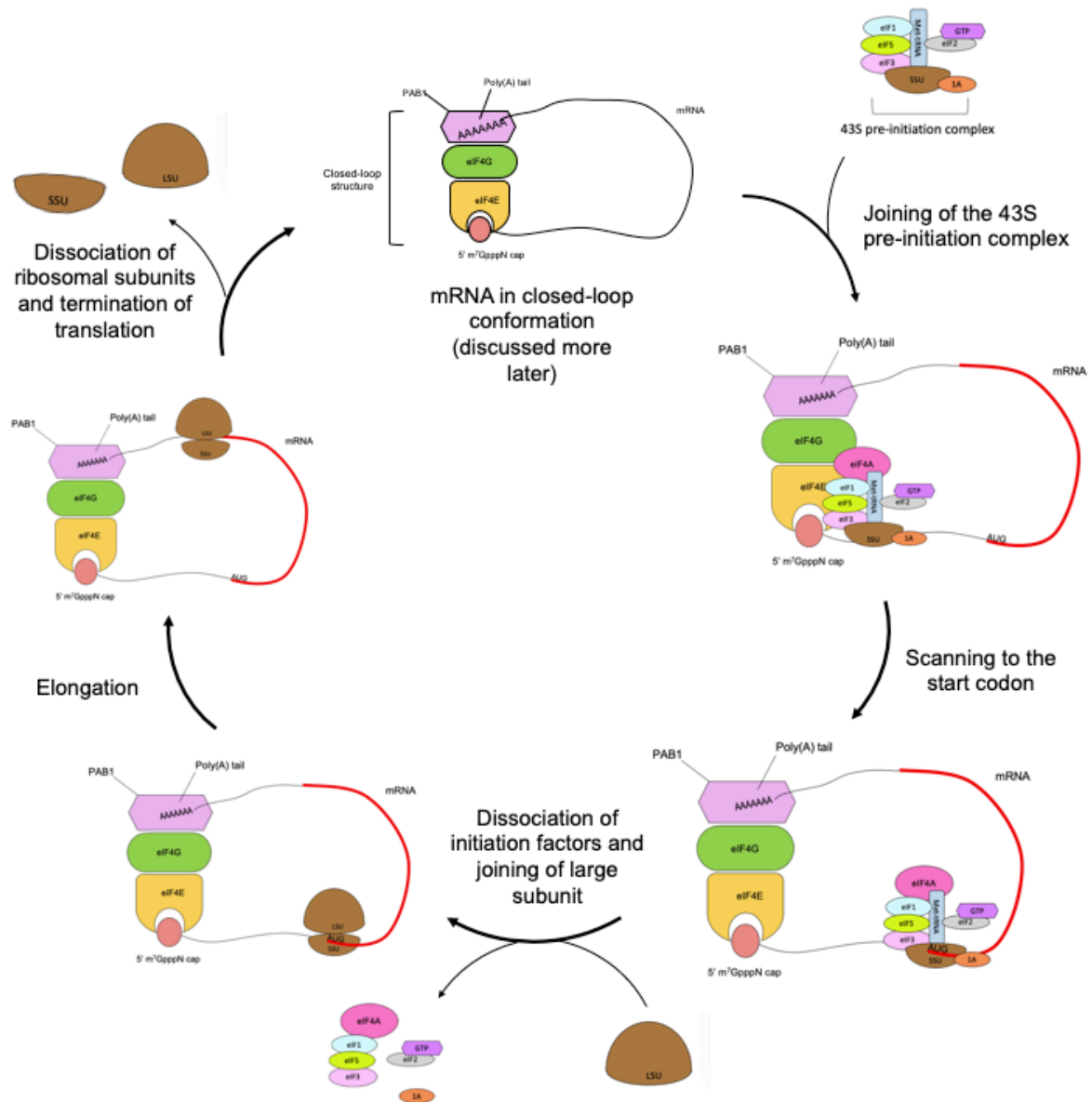


Figure 1.1. Overview of mRNA translation stages, assuming the closed-loop model, and with a focus on translation initiation.

1.1.1 Translation initiation

Of these stages, it is the stage of initiation which contains the greatest diversity across the kingdoms of life (Marintchev and Wagner, 2004). One of the most prominent examples of this diversity is in the mechanism of start codon recognition. In prokaryotes, an upstream, purine-rich sequence of the mRNA, named the Shine-Dalgarno sequence, is identified by direct base pairing with the 16S rRNA component of the SSU, allowing for SSU recognition of the downstream start codon (Dalgarno and Shine, 1973; Steitz and Jakes, 1975; Kozak,

1999). Eukaryotic start codon recognition, on the other hand, is thought to occur via a ‘scanning’ mechanism, in which the SSU is recruited to the mRNA 5’ end and begins to scan the 5’ UTR in a 5’ to 3’ manner until complementary base-pairing of the Met-tRNA_i^{Met} CAU anticodon is achieved (Hinnebusch, 2014; Shirokikh and Preiss, 2018).

In addition to being the most diverse stage of mRNA translation, translation initiation – especially in eukaryotes – is also the most complex, with a number of tightly orchestrated steps and a host of factors involved in its regulation (Sonenberg and Hinnebusch, 2009). For example, translation initiation in eukaryotes requires up to nine initiation factors (eIFs), many of which are individually subject to extensive regulation in response to cellular or external signals (Jackson, Hellen and Pestova, 2010). As a result of this complexity and regulation, translation initiation is naturally placed as a cellular process of interest in the context of health and disease.

1.1.1.1 Cap-dependent translation initiation in eukaryotes

The predominant translation initiation pathway in eukaryotes is via a cap-dependent mechanism. Owing to the complex and dynamic nature of translation initiation, there are many uncertainties in the exact sequence of events or factors involved. However, taking into account the body of biochemical and structural studies on cap-dependent translation initiation, a consensus model has developed, with the main steps of the process being (1) attachment to the cap; (2) scanning; and (3) start codon recognition with large subunit joining (Sonenberg and Hinnebusch, 2009; Jackson, Hellen and Pestova, 2010; Shirokikh and Preiss, 2018) (Figure 1.1).

The process of translation initiation in this canonical model starts with recruitment of the ribosomal SSU to the mRNA cap. The mRNA cap is a 5’ terminal structure consisting of a 7-methyl guanosine connected to the rest of the mRNA via a 5’ to 5’ triphosphate link (Muthukrishnan *et al.*, 1975; Furuichi, LaFiandra and Shatkin, 1977). As the mRNA cap does not intrinsically have affinity for the SSU, the recruitment of the SSU to the cap is mediated by a host of eukaryotic initiation factors (eIFs) that associate with the capped mRNA (Shirokikh and Preiss, 2018). These eIFs include eIF4E, which contacts the cap (Sonenberg *et al.*, 1978, 1980), eIF4G, a large scaffold protein (Goyer *et al.*, 1993) and eIF4A, an ATP-

dependent RNA helicase (Rogers, Lima and Merrick, 2001), which are often together known as the eIF4F complex (Grifo *et al.*, 1983; Merrick, 2015). These factors, together known as the eIF4F complex, interact with the 43S pre-initiation complex (PIC) – consisting of the SSU, the aminoacylated initiator Met-tRNA_i^{Met}, GTP, initiation factors 1, 1A, 2, 3 and likely 5 – to bring the SSU to the cap (Jackson, Hellen and Pestova, 2010).

The PIC then scans the mRNA 5' UTR nucleotide by nucleotide in an ATP-dependent, 5' to 3' manner until a match for the anticodon of the initiator Met-tRNA_i^{Met} lands in the P site of the SSU (Kozak, 1989; Lomakin and Steitz, 2013; Hinnebusch, 2014). Due to the common presence of RNA secondary structures in the 5' UTR, this process is thought to also require the function of eIFs such as eIF4F, 4A and 4B to help unwind and facilitate ATP-dependent movement along the mRNA (Pestova and Kolupaeva, 2002; Marintchev and Wagner, 2004; García-García *et al.*, 2015).

Upon reaching a start codon perfectly paired to the anticodon of the initiator Met-tRNA_i^{Met}, with fidelity of pairing thought to be aided by eIF1 and eIF1A (Grifo *et al.*, 1983)(Fekete *et al.*, 2007; Passmore *et al.*, 2007), the next step of initiation takes place. The 43S complex is arrested at the start codon and the complex commits to initiation through eIF5-mediate hydrolysis of GTP from eIF2, resulting in the release of eIF2 and the other eIFs (Merrick, 1979; Peterson, Safer and Merrick, 1979; Lee *et al.*, 2002). Finally, this allows for joining of the 60S large subunit (LSU), resulting in a full ribosome ready for elongation (Sonenberg and Hinnebusch, 2009; Lorsch and Dever, 2010).

1.1.1.2 Cap independent translation

While cap-dependent translation is the predominant pathway in eukaryotes, it should be noted that it is not the only mechanism of translation initiation possible. Many uncapped viruses, such as encephalomyocarditis virus (EMCV) (Jang and Wimmer, 1990), poliovirus (Pelletier and Sonenberg, 1988) and human hepatitis C virus (Brown *et al.*, 1992) have been observed to contain highly-structured regions in their mRNA 5' UTRs, named IRESes (for independent ribosome entry site), that allow for recruitment of host ribosomes for subsequent translation. Similar structures have since been discovered in some eukaryotic cellular mRNAs, and cap-independent ribosome recruitment has also been observed at such sites (Hellen and Sarnow,

2001). Some prominent examples include vascular endothelial growth factor (VEGF) (Akiri *et al.*, 1998), c-myc (Nanbru *et al.*, 1997) and eIF4G1 (Johannes and Sarnow, 1998). It has been proposed that IRES-mediated translation of cellular mRNAs may represent a mechanism for select mRNAs to bypass cap-dependent translational shutdown during events such as hypoxia, viral infection or apoptosis (Sarnow, 1989; Coldwell *et al.*, 2000; Hellen and Sarnow, 2001).

Although cap-independent translation of cellular mRNAs does appear to play a role in eukaryotic protein synthesis, it is currently only expected to occur in a small handful of situations, with the bulk of translation still occurring via a cap-dependent mechanism (Merrick, 2004). In discussing the closed-loop model of translation, it is interesting to keep in mind the possibility of alternative translation mechanisms, especially as components of the closed-loop have been associated with cap-independent translation. For example, eIF4G1 and eIF4G2/DAP5 have been reported to mediate translation in cap-independent conditions (López de Quinto and Martínez-Salas, 2000; de Breyne *et al.*, 2009; Haizel *et al.*, 2020). Ultimately however, the closed-loop model of translation is largely studied in the context of cap-dependent translation, and for this reason, the rest of this thesis will focus on the cap-dependent pathway of mRNA translation.

1.1.2 Yeast as a model organism for studying eukaryotic translation

One common model organism for studying eukaryotic translation is yeast – in particular *Saccharomyces cerevisiae* (also known as the common brewer's yeast) – due to significant conservation of major biological features and processes with higher eukaryotes (Linder, 1992; Botstein and Fink, 2011). It was the first eukaryotic genome to be fully sequenced (Goffeau *et al.*, 1996), and has become a valuable tool in the understanding of molecular processes such as transcription and translation in eukaryotes (Botstein, Chervitz and Cherry, 1997). It has been estimated that a yeast ortholog exists for approximately 23% of known disease-associated genes in humans, rendering yeast as a suitable model organism in many gene function studies (Jiang, Lin and Wang, 2006; Mohammadi *et al.*, 2015; Liu *et al.*, 2017). It also presents further advantages in its complexity-reducing small genome size, short life cycle and ease of genetic manipulation with modern molecular biology approaches (Linder, 1992; Botstein, Chervitz and Cherry, 1997; Mohammadi *et al.*, 2015).

S. cerevisiae (from here on referred to as ‘yeast’) has been particularly instrumental in contributing to our understanding of eukaryotic translation, with many early studies into translation utilising cell-free mRNA translation systems derived from yeast (Gasior *et al.*, 1979; Hartley *et al.*, 1996). Many translation initiation, elongation and termination factors have been isolated and characterised owing to their initial discovery in *S. cerevisiae* and subsequent identification in higher eukaryotes (Linder, 1992). In particular, a host of eukaryotic translation initiation factors, such as eIF4E, eIF4A and eIF4G1 have been extensively studied and characterised in yeast model systems to elucidate their roles and interactions during initiation (Sachs, Davis and Kornberg, 1987; Tarun and Sachs, 1996; Haghighat and Sonenberg, 1997). Finally, and central to this thesis, a majority of studies on the closed-loop were conducted in yeast and will be discussed in detail in the following subchapter 1.2. Similarly, for the research presented in this thesis, yeast has also been chosen as the model organism for studying the closed-loop model of mRNA translation.

1.2 Development of the closed-loop model

The closed-loop model of translation describes the functional circularisation of an mRNA molecule by proteins eIF4E, eIF4G and Pab1, resulting in an mRNA with ends brought into close proximity by this protein bridge. Under the model, eIF4E contacts the cap, PAB1 contacts the poly(A) tail and eIF4G contacts both eIF4E and PAB1 to complete the circularisation (Figure 1.2).

The closed-loop model arose as a culmination of studies detailing the individual and combined stimulatory effects of the various components of the closed-loop on translation. From these studies, it was largely considered a distinctive feature of eukaryotic translation initiation. This section will first introduce the mRNA components of the closed-loop and their effect on translation initiation, followed by the protein components and their roles, and finally detail the way the protein factors and mRNA interact.

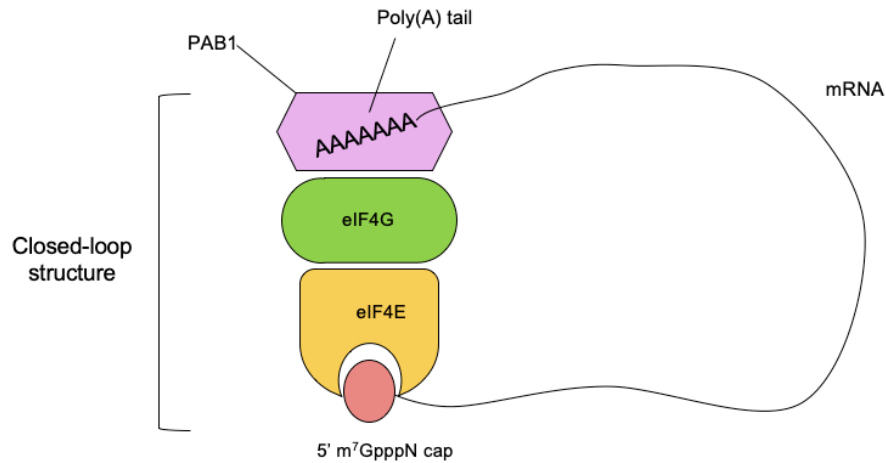


Figure 1.2. The. closed-loop model components and structure.

eIF4E (yellow) contacts the mRNA 5' cap (red). *Pab1* (pink) contacts the poly(A) tail. *eIF4G* (green) contacts both *eIF4E* and *Pab1*.

1.2.1 Key mRNA components involved in the closed-loop

1.2.1.1 The mRNA 5' cap and its involvement in translation

The m⁷GpppN cap refers to a 7-methyl guanosine connected to the start of the mRNA 5' UTR via a 5' to 5' triphosphate link (Muthukrishnan *et al.*, 1975; Furuichi, LaFiandra and Shatkin, 1977). This 5' terminal structure is unique to, and conserved across all eukaryotic mRNAs. Broadly, mRNAs are capped co-transcriptionally in the nucleus by an evolutionarily conserved enzymatic pathway involving removal of the γ -phosphate, capping by addition of guanosine monophosphate and subsequent N7-methylation (Furuichi, 2015; Ramanathan, Robb and Chan, 2016). The mRNA cap has been shown to play many roles, such as in nuclear export, as a signature of 'self RNA' and in preventing 5' $\rightarrow$ 3' exonuclease-mediated degradation (Ramanathan, Robb and Chan, 2016; Galloway and Cowling, 2019; Drazkowska *et al.*, 2022).

Amongst its many functions, the mRNA cap's role in translation initiation – and later in the context of the closed-loop – has been widely studied. As described in 1.1.1.1, the main pathway of translation initiation in eukaryotes depends on presence of a 5' cap. Early *in vitro* studies compared the translation of uncapped mRNAs, such as those from reovirus (Held, West and Gallagher, 1977), vaccinia virus (Muthukrishnan *et al.*, 1978) and tobacco mosaic virus (Wodnar-Filipowicz *et al.*, 1978), to their capped equivalent in a range of cell-free

protein synthesis systems. These studies found that although the ionic concentration and source of the cell-free extract influenced the degree of cap-dependence, the uncapped mRNAs were consistently poorer templates for protein synthesis in *in vitro* systems than the equivalent capped mRNA (Banerjee, 1980). Another study which synthetically capped naturally cap-less prokaryotic λ *cro* mRNA found that this mRNA could be translated in wheat germ extract, but in a manner completely dependent on addition of the cap structure (Paterson and Rosenberg, 1979). Capped mRNAs were also shown to be more efficient than uncapped mRNAs in binding ribosomes (Banerjee, 1980). Finally, *in vitro* studies using capped and uncapped reporter mRNAs reported greater translational output from capped mRNAs compared to their equivalent uncapped counterparts (Gallie, 1991; Preiss and Hentze, 1998).

1.2.1.2 *The poly(A) tail and its effect on mRNA translation*

The poly(A) tail is a string of adenosines found at the ends of almost all eukaryotic mRNAs (Passmore *et al.*, 2007; Neve *et al.*, 2017). Co-transcriptionally, and aided by a host of other 3' end processing factors, cleavage of the newly-synthesised pre-mRNA is catalysed by the multi-subunit cleavage and polyadenylation specificity factor (CPSF) upon recognition of a poly(A) signal (Murthy and Manley, 1995; Mandel *et al.*, 2006). This signal is variable, and transcripts can often have more than one at their 3' ends, leading to alternative polyadenylation (APA) (Tian and Manley, 2017). Following cleavage, the addition of adenosines is catalysed by PAP with the help of other factors (Gershon *et al.*, 1991; Dickson *et al.*, 2001).

All eukaryotic mRNAs have a poly(A) tail, with the only known exception being metazoan replication-dependent histone mRNAs (Marzluff, 2005). Modal steady-state poly(A) tail lengths are thought to vary from ~30 nt in yeast, to 50-100 nt in other eukaryotes (Passmore and Collier, 2021). The poly(A) is thought to play a role in many aspects of mRNA regulation, for example in the transport of mRNA from the nucleus to cytoplasm (Fuke and Ohno, 2008), protection of mRNA from 3' $\rightarrow$ 5' degradation (Ford, Bagga and Wilusz, 1997) and overall mRNA metabolism (Sachs, 1990).

The poly(A) tail's involvement with the closed-loop model began with multiple observations that presence of the tail could stimulate and enhance translation *in vitro*. In 1983, Jacobson and Favreau added increasing concentrations of poly(A) ribopolymer to a polyadenylated mRNA mix in rabbit reticulocyte lysate and observed a dose-dependent inhibition of translation with increasing poly(A) concentration. They also measured the rate of protein synthesis in the presence and absence of competing poly(A), finding that without addition of poly(A), the rate of synthesis was much faster, suggesting that the inhibition by poly(A) was due to its effect on the rate of translation and not other factors. Finally, they also observed that increasing exogenous poly(A) in a pool of poly(A)+ and poly(A)- mRNAs inhibited translation of the poly(A)+ mRNAs but increased translation of poly(A)- mRNAs, concluding that by inhibiting the translation of poly(A)+ mRNA, the exogenous poly(A) allowed the more poorly translated poly(A)- mRNA to be better translated. From these findings, it was concluding that exogenous poly(A) is a competitive inhibitor of translation, suggesting that native poly(A) tails of mRNAs were involved in their translation through a yet unknown mechanism (Jacobson and Favreau, 1983).

Follow-on studies by Munroe and Jacobson in 1990 *in vitro* transcribed mRNAs with and without poly(A) tails and observed their translation by measuring their incorporation of [³⁵S]methionine in rabbit reticulocyte lysates. It was observed that poly(A)+ mRNAs were translated more efficiently, and that this was not due to protective effects of the tail on degradation. They also observed by polysome profiling that poly(A)+ mRNAs were recruited into polysomes more than their poly(A)- mRNAs, and that this specifically affected translation initiation over the other translation stages. The stage of translation initiation affected was also narrowed down to 80S formation by using inhibitors of the different translation stages followed by comparing sedimentation profiles of poly(A)-/+ mRNAs (Munroe and Jacobson, 1990).

Unlike the more intuitive observation that the 5' cap at the beginning of the mRNA plays a role in translation initiation, the role of the poly(A) tail in translation initiation, given its location at the 3' end of the message, is perhaps more unexpected. However, with the above findings suggesting a role for the poly(A) tail in translation initiation, a model began to develop for how this might be explained. It was proposed that the poly(A) tail, despite its location at the 3' end of the message, had some involvement or 'cross-talk' with the cap at

the 5' end of the message, particularly during translation initiation (Jackson and Standart, 1990; Munroe and Jacobson, 1990; Sachs, 1990).

1.2.1.3 Synergistic effects of the cap and tail in enhancing translation

With the knowledge that polyA tails and mRNA caps each stimulate translation, various studies sought to determine whether their dual existence resulted in an additive or synergistic increase in translation efficiency.

Gallie (1991) reported a synergistic effect of the cap and poly(A) tail in the translation of reporter firefly luciferase (LUC) mRNA electroporated into tobacco, with the poly(A) tail increasing expression by 1.5-fold on uncapped mRNA but over 21-fold on capped mRNA. Similarly, the cap increased expression by 21-fold on poly(A)- mRNA and by 297-fold on poly(A)+ mRNA. Similar observations were made when the constructs were electroporated into Chinese hamster ovary (CHO) cells, suggesting a similar phenomenon in higher eukaryotes. However, they failed to recreate the synergism in rabbit reticulocyte lystate *in vitro*, citing incomplete recreation of the *in vivo* environment as a likely reason. mRNA half-lives were also measured to account for effects of stability on protein synthesis. Overall, it was concluded that the synergy of the cap and tail on translation, when observed, was from increasing translation efficiency rather than message stability (Gallie, 1991). Similar studies by Tarun and Sachs (1995) found a synergistic, rather than additive, effect of the cap and tail in translation of firefly luciferase mRNA in yeast extracts, with a synergistic interaction in the range of 2.5- to 8-fold in comparison to either cap or tail alone. Similar to Gallie (1991), the extent of this effect could not be explained by differences in chemical or functional mRNA stability.

Studies into the mechanism of synergy between the poly(A) and 5' cap by Preiss and Hentze (1998) further echoed the above findings. Chloramphenicol acetyltransferase (CAT) mRNA translated in yeast extract showed an additive effect of the cap and tail. Capped-only and poly(A)-only mRNA resulted in 30% and 53% of the protein output of the capped and polyadenylated version respectively. However, when these constructs were electroporated into living cells and protein synthesis quantified by immunoprecipitation, the synergistic effect was much greater, with capped-only and poly(A) only constructs giving 19% and 6%

of the translation product of the capped and polyadenylated construct respectively. Interestingly, the poly(A) tail's ability to stimulate translation was also diminished *in vivo* compared to *in vitro*. To explain these differences between *in vitro* and *in vivo* data, it was hypothesised that this may be due to mRNA competition, with the nuclease-treated *in vitro* extract containing little competing mRNA and the *in vivo* condition containing more. Translation of the CAT mRNA in yeast extract preincubated with or without micrococcal nuclease (to remove competing mRNAs) was therefore compared. It was observed that in the presence of competing mRNA, the synergistic effect of the cap and poly(A) was more prominent and similar to that observed *in vivo*, whereas in a non-competitive environment (micrococcal-treated), the synergy was abolished, suggesting that mRNA competition enforces combined use of both the cap and poly(A) tail as translational 'promoters' (Preiss and Hentze, 1998).

From these various studies into the effects of the cap and poly(A) tail on translation, it was starting to appear that both the cap and poly(A) tail had a role in stimulating translation, and that at least in some situations and likely *in vivo*, the two could work together in a synergistic manner to increase translational output. For these two elements to exert a similar function despite being on opposite ends of the mRNA, it was proposed that there might be a physical interaction to bring the ends into close proximity (Gallie, 1991; Tarun and Sachs, 1995; Preiss and Hentze, 1998). As this hypothesis developed, the protein factors that could mediate this interaction were simultaneously discovered and characterised, and will be discussed in detail in the next section.

1.2.2 The protein components making up the closed-loop and their interactions

For the mRNA ends to be found in close proximity, it was hypothesised that this interaction may have been mediated by one or multiple protein factors (Tarun and Sachs, 1995; Preiss and Hentze, 1998). Biochemical and structural studies subsequently identified three proteins, eIF4E, eIF4G and PAB1, that could bind to each other and to the mRNA ends in a manner leading to effective circularisation (Figure 1.3). These interactions provide a plausible explanation for how the ends could be brought into closed proximity. However, it is important also to note that this series of documented interactions does not exist to the exclusion of other possibilities for how the proximity of the ends might occur. Nevertheless,

this section will focus on the proteins and interactions in the canonical closed-loop model, and explain each protein factor and their role in forming the closed-loop and in translation initiation.

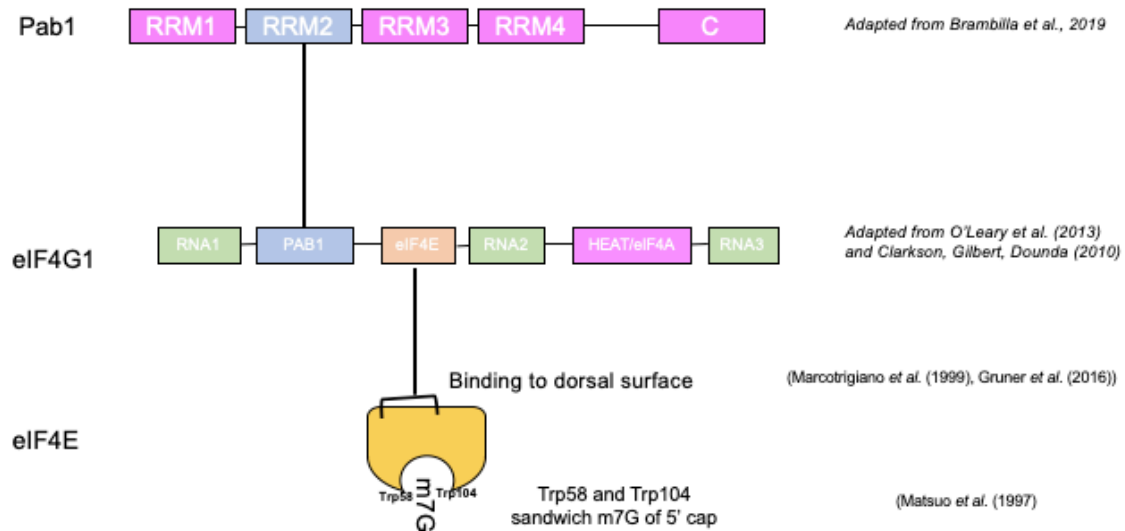


Figure 1.3. Binding of closed-loop proteins to each other.

Pab1 bind *eIF4G1* through its second RNA recognition motif (RRM2). It has 4 RRMs and a c-terminal region (C). *eIF4G1* contacts *Pab1* through its *Pab1* binding domain in the first third of the protein. Downstream, *eIF4G1* has a binding domain to *eIF4E*'s dorsal surface. Two conserved tryptophans, W58 and W104, sandwich the m7G of the 5' cap. *Pab1* domain structure was adapted from (Brambilla et al., 2019), *eIF4G1* structure was adapted from (Clarkson, Gilbert and Doudna, 2010; O'Leary et al., 2013).

1.2.2.1 *eIF4E*

Eukaryotic translation factor 4E (eIF4E) is a 24 kDa protein encoded by gene *CDC33* (Altmann, Handschin and Trachsel, 1987; Brenner *et al.*, 1988). It is found in all eukaryotes and is characterised by its role as the mRNA 5' cap binding protein (Sonenberg *et al.*, 1978; Ross-Kaschitzka and Altmann, 2020). eIF4E is an essential gene, with deletion found to be lethal (Altmann, Handschin and Trachsel, 1987). It was first identified as a protein binding to the 5' cap of synthetically capped viral mRNAs in a preparation of rabbit reticulocyte and mouse ascite ribosomes. Specifically, chemical crosslinking of the cap and interacting proteins, followed by polyacrylamide gel electrophoresis, resolved a single 24 kDa polypeptide (Sonenberg *et al.*, 1978). It was subsequently purified and characterised in HeLa cells (Sonenberg *et al.*, 1979) and yeast (Altmann *et al.*, 1985), with the gene responsible for coding it later identified as *CDC33* (Altmann, Handschin and Trachsel, 1987).

From resolution of its structure in yeast (Matsuo *et al.*, 1997) and mouse (Marcotrigiano *et al.*, 1997), eIF4E is known to have a concave face responsible for interacting with the mRNA cap. Importantly, two evolutionarily conserved tryptophans within this pocket stack with the methylated base of the cap, greatly increasing affinity of 4E to the methylated cap (Marcotrigiano *et al.*, 1997; Matsuo *et al.*, 1997). These two tryptophans, along with a further 6, are found in an evolutionarily conserved 182 amino acid c-terminal portion, resulting in a conserved cap binding pocket structure (Marcotrigiano *et al.*, 1997; Ross-Kaschitza and Altmann, 2020).

Since its discovery, the role of eIF4E in stimulating cap-dependent translation has been consistently observed. eIF4E was observed to stimulate translation of capped Sindbis virus, reovirus and rabbit globin mRNAs by 2-4 fold, but had no effect on naturally uncapped mRNAs from mouse encephalomyocarditis and satellite tobacco necrosis virus (Sonenberg *et al.*, 1979). eIF4E was shown to increase translation of synthetically capped viral mRNA, and also shifted translation in favour of capped mRNAs in a pool of capped and uncapped mRNAs (Sonenberg *et al.*, 1980).

eIF4E has been shown to both be a rate-limiting regulator of cap-dependent translation due to its low molar concentration (Marcotrigiano *et al.*, 1997; Sonenberg and Gingras, 1998), while also having its own activity regulated by other factors. eIF4E is bound by eIF4G at its dorsal surface (Hershey *et al.*, 1999; Grüner *et al.*, 2016; Miras *et al.*, 2017), and has been purified with eIF4G and eIF4A in a complex termed eIF4F (Grifo *et al.*, 1983). Together, this complex was shown to stimulate translation (Grifo *et al.*, 1983), with the binding of eIF4G1 to eIF4E shown to increase cap-dependent translation (Haghighat and Sonenberg, 1997; Marcotrigiano *et al.*, 1999; Grüner *et al.*, 2016). A group of other proteins, generally termed 4E-binding proteins (4E-BPs), also compete with eIF4G for eIF4E binding through a conserved YXXXXLΦ (with Φ being any hydrophobic amino acid) motif, resulting in inhibition of cap-dependent translation initiation by sequestering eIF4E from eIF4G (Sonenberg and Gingras, 1998; Marcotrigiano *et al.*, 1999; Richter and Sonenberg, 2005; Grüner *et al.*, 2016). eIF4E overexpression has also been observed in many cancers, leading to its designation as a proto-oncogene (Sonenberg, 2008). eIF4E therefore directly and indirectly plays a key role in the regulation of cap-dependent translation.

1.2.2.2 *eIF4G1 and eIF4G2*

Eukaryotic translation initiation factor 4 gamma (eIF4G) is a protein factor playing a pivotal role in translation initiation. At least in yeast and humans, two isoforms of eIF4G, eIF4G1 (~107 kDa in yeast) and eIF4G2 (~104 kDa in yeast) are encoded by *Tif4631* and *Tif4632* respectively, with the two proteins sharing 53% identity overall and 80% identity over a specific 320 amino-acid stretch (Goyer *et al.*, 1993). eIF4G1 and eIF4G2 are thought to play both distinct and functionally overlapping roles in translation initiation, with eIF4G1 mutants exhibiting a slow growing phenotype and impaired translation, while eIF4G2 mutants exhibited no growth defect (Goyer *et al.*, 1993; Clarkson, Gilbert and Doudna, 2010). However, deletion of both led to unviable strains (Goyer *et al.*, 1993). Although eIF4G2 has also shown to bind eIF4E and Pab1 (Tarun and Sachs, 1996), eIF4G1 has been studied in greater detail and appears to be more strongly implicated in translation and formation of the closed-loop, so will be the focus of this section going forward, and of the thesis more broadly.

eIF4G1 has been shown to contact multiple initiation factors, including eIF4E (Marcotrigiano *et al.*, 1999), Pab1 (Kessler and Sachs, 1998), eIF4A (Imataka and Sonenberg, 1997), eIF3 (LeFebvre *et al.*, 2006) and also RNA (Berset *et al.*, 2003). Therefore, it is considered a ‘scaffolding’ or ‘adapter’ protein providing docking for the various initiation factors and RNA during translation (Hentze, 1997; Prévôt, Darlix and Ohlmann, 2003).

Specifically focusing on the closed-loop interactions, eIF4G1’s binding to eIF4E occurs through a 7-amino acid YXXXXLΦ (with Φ being any hydrophobic amino acid) motif that is conserved with other 4E binding proteins which compete with eIF4G1 for binding to eIF4E (Sonenberg and Gingras, 1998; Marcotrigiano *et al.*, 1999; Richter and Sonenberg, 2005; Grüner *et al.*, 2016). eIF4G1 also contacts Pab1 through a necessary and sufficient 114 amino-acid region conserved between eIF4G1 and eIF4G2 (Tarun and Sachs, 1996; Salvador Z Tarun *et al.*, 1997). It was observed that this interaction with Pab1 is responsible for eIF4G1’s role in mediating synergistic effects on translation by the cap and tail. Interestingly, it was also observed that the interaction between eIF4G1 and Pab1 became essential when eIF4E’s interaction with eIF4G1 was disrupted, and coupled with the observation that strains with both Pab1 and eIF4E mutations were unviable, it was hypothesised that eIF4G1 played a

central role in coordinating the two ends of an mRNA message (Salvador Z. Tarun *et al.*, 1997). As eIF4G1's role in the closed-loop is largely in the context of its interactions with eIF4E and Pab1, to avoid repetition, further details of eIF4G1's specific interactions with eIF4E and Pab1 are detailed in section 1.2.2.1 and section 1.2.2.3 respectively.

1.2.2.3 *Pab1*

Poly(A)-binding protein (Pab1) is an essential, highly abundant, 72 kDa protein encoded by the gene *PAB1*, best characterised by its cellular function of binding the poly(A) tail (Tarun *et al.*, 1997). While it has been found in the nucleus in a truncated 55 kDa form thought to be involved in mRNA export and 3' end processing, Pab1 is predominantly found in the cytoplasm as the full-length 72 kDa protein (Brune *et al.*, 2005), and it is this cytoplasmic form of Pab1 which will be the focus of this section.

Pab1 was first discovered as a protein bound to the poly(A) region of globin mRNAs translated in rabbit reticulocyte (Blobel, 1971), and was since also purified and characterised in liver (rat) and L cells (mouse) (Blobel, 1973) and yeast (Adam *et al.*, 1986). Although the specific sequence differs, the function of Pab1 in binding the poly(A) tail, and the presence of a RNA recognition motif (RRM) domain to facilitate this is an evolutionarily conserved feature in eukaryotes (Grange *et al.*, 1987; Melamed *et al.*, 2013). Pab1 has a footprint of approximately 25-27 nt on the poly(A) tail (Baer and Kornberg, 1983) and it has been observed that multiple Pab1 proteins can bind a single poly(A) tail (Sawazaki *et al.*, 2018). Pab1 is involved in mediating multiple roles of the poly(A) tail including, and not limited to, mRNA biogenesis, export, polyadenylation, mRNA decay and translation (Brune *et al.*, 2005).

Before the interaction between Pab1 and eIF4G was characterised, it was already observed that Pab1 was implicated in translation initiation. Neutralisation or depletion of Pab1 in yeast removed the stimulatory effect of the poly(A) tail on translation in yeast and disrupted the synergy of the cap and poly(A) tail. Re-introducing Pab1 restored translation to normal levels seen before Pab1 depletion, suggesting a role for Pab1 in mediating the translation enhancing properties previously attributed to the poly(A) tail (Tarun and Sachs, 1995). It was also

observed that addition of excess poly(A) inhibited the *in vitro* translation of mRNA, but this effect was overcome by addition of purified Pab1 (Sachs, Sarnow and Hentze, 1997).

Following these studies, the mechanism by which Pab1 might influence translation initiation at the 5' end was explored and found to be through association with eIF4G. Tarun and Sachs (1996) observed the eIF4G-dependent copurification of Pab1 with the eIF4F complex. Pab1 and eIF4G co-purified in stoichiometric quantities, and if eIF4G was removed, Pab1 no longer co-purified with eIF4E. Pab1 was also shown to bind specifically and directly to eIF4G1 and eIF4G2 in an RNA-dependent manner, in that binding of RNA to Pab1 was required for Pab1's association with eIF4G (Tarun and Sachs, 1996). In yeast, Pab1 contains four RRM domains for associating with the RNA molecule (Adam *et al.*, 1986; Sachs, Bond and Kornberg, 1986) and it was later determined that the RRM2 domain was responsible for the association with eIF4G. By removing the different RRM domains independently and testing these Pab1 mutants for translational stimulation, it was observed that removal of RRM2 resulted in a lack of poly(A) dependent translation, while removal of the other domains did not (Kessler and Sachs, 1998). Taking these various interactions into account, along with the earlier finding that Pab1 played a role in recruitment of the 40S subunit to the mRNA (Tarun and Sachs, 1995), it was hypothesised that eIF4E, eIF4G and Pab1 together effectively circularise mRNA molecules, leading to their efficient translation (Sachs, Sarnow and Hentze, 1997).

1.2.2.4 Visualisation of the closed-loop interactions together

With the interactions between all the closed-loop components characterised, a model had begun to develop regarding the complete closed-loop and its role in translation. However, the simultaneous binding of the cap-eIF4E-eIF4G-Pab1-poly(A) tail, and this proposed closed-loop structure's functional role in translation had yet to be demonstrated.

Interestingly, before conceptualisation of the closed-loop model occurred, circular polysomes had already been described in the literature. Christensen, Kahn and Bourne (1987) studied the shape and distribution of polysomes on the rough endoplasmic reticulum of somatotropes and mammatropes in the rat pituitary by electron microscopy. They observed that approximately 80% of the polysomes observed in these two cell types were in circular conformations of 6-7

ribosomes, with a further 17-20% found in spiral formations (Christensen, Kahn and Bourne, 1987).

After the identification and characterisation of the closed-loop proteins, atomic force microscopy was utilised to study whether the eIF4E/eIF4G1/Pab1 complex could circularise mRNA. Reconstituted eIF4E, Pab1 and an eIF4G1 peptide containing Pab1 and eIF4E binding domains were added to RNA in different combinations. In the absence of all three factors, RNA molecules were observed to be in a linear conformation 94% of the time. However, when the three components were added to the RNA together, ~54% of the molecules were in a circulate conformation and ~46% were linear. Incubations with mutated eIF4G1 peptides which diminish binding to either Pab1 or eIF4E resulted in less circularisation than WT eIF4G1, but still resulted in more circularisation than addition of no factors at all. From these findings, it was concluded that the closed-loop components – eIF4E, eIF4G1 and Pab1 – had the ability to together circularise mRNA.

1.2.3 Hypothesised functions of the closed-loop in translation initiation

Although each of the mRNA features and protein factors of the closed-loop had individually or together been shown to increase translation efficiency, it was still not directly shown how or why the formation of a circular structure specifically contributed to this observation, or why the 3' end of the message was important to translation. Multiple biological reasons for the stimulatory effect of this circular structure on translation have been hypothesised with varying levels of supporting evidence. Some of the prominent theories are discussed below.

1.2.3.1 Ribosome recycling

Possibly the most prominent theory explaining the role of the closed-loop is that it may allow for efficient ribosome recycling after a ribosome has finished translation. With ribosomes shown to be capable of scanning beyond the termination codon to re-begin translation at a downstream initiation codon (Kozak, 1984), and the observation that polysomes can exist in a circular structure (Christensen, Kahn and Bourne, 1987), it is possible that the close proximity of the 5' UTR with its own 3' UTR could facilitate ribosome re-joining after one round of translation. Functional studies also reported an interaction between Pab1 and the

translation termination factor complex eRF1-GSPT. When this GSPT-Pab1 interaction was disrupted, translation of capped and polyadenylated mRNA was significantly attenuated despite the disruption having no effect on *de novo* formation of 80S ribosomal complexes. It was therefore suggested that Pab1 may mediate the recycling of ribosomes within a molecule through its interactions with GSPT and the eIF4E/eIF4G complex (Uchida *et al.*, 2002).

1.2.3.2 Mutually stabilising interactions

Another theory for why the closed-loop may be biologically significant is due to the mutually stabilising interactions between the components. For example, binding of eIF4E to the cap was greatly increased upon eIF4G1 binding to eIF4E (Haghighat and Sonenberg, 1997; Ptushkina *et al.*, 1998; von Der Haar, Ball and McCarthy, 2000). eIF4E binding to the cap was also stabilised by the binding of eIF4G1 binding to Pab1 (Wei *et al.*, 1998; Borman, Michel and Kean, 2000; von Der Haar, Ball and McCarthy, 2000). A strong interaction and conformational change was further shown to occur at the eIF4E and eIF4G interface upon binding of the two proteins to each other, resulting in a stable complex and by extension, a possibly more stable mRNP capable of multiple rounds of translation (Gross *et al.*, 2003). Taken together, it is plausible that by forming the closed-loop, the individual components of the complex are stabilised on, or in close proximity to the mRNA (Thompson and Gilbert, 2017; Vicens, Kieft and Rissland, 2018). The components themselves are implicated in translation initiation, for example with eIF4G1's ability to contact other initiation factors such as eIF4A and eIF3 (Imataka and Sonenberg, 1997; LeFebvre *et al.*, 2006) and the eIF4F complex's ability to recruit the 43S pre-initiation complex (Jackson, Hellen and Pestova, 2010). Therefore, formation of this stabilised loop may serve as a way for the required initiation factors to all be present at the mRNA in a manner leading to effective translation initiation.

1.3 Challenges to the view of the closed-loop being a predominant feature of translation

Despite the body of evidence supporting the existence of a closed-loop structure and its ability to enhance translation, evidence for it playing a central role in translation initiation was not so strongly established outside of structural studies, *in vitro* translation systems and

in vivo experiments with introduced reporter constructs. Studies moving beyond studying the mechanistic possibility of this closed-loop forming and its effect on the few reporter constructs tested came to less definitive conclusions regarding the closed-loop's biological prevalence or function. This raised the possibility that the closed-loop might not always be a defining feature of translation initiation.

1.3.1 Viability of strains with closed-loop mutations or deletions

One early indication that the closed-loop may not be essential for translation in its canonical form came in the way of mutation and gene disruption studies of the closed-loop components. In yeast, for example, eIF4G1 mutant strains with varying truncations or point mutations to disrupt binding with Pab1 were still viable, despite some having a slow-growing or temperature sensitive phenotype. Furthermore, it was observed that the Pab1-eIF4G1 association could compensate for loss of eIF4E-cap binding *in vivo*. However, simultaneous Pab1-eIF4G1 and eIF4E-eIF4G1 disruption mutations resulted in synthetic lethality, suggesting some form of functional redundancy between Pab1 and eIF4E stimulated translation mediated by eIF4G1 (Salvador Z. Tarun *et al.*, 1997; Kessler and Sachs, 1998). Park *et al.* (2011) did not observe the same synthetic lethality as the previous study, but identified other RNA binding domains on eIF4G1, which when removed in addition to the Pab1-eIF4G1 and eIF4E-eIF4G1 mutations, did result in lethality. This led them to conclude that closed-loop formation via Pab1-eIF4G1 is not essential *in vivo*, and that other contacts between eIF4G1 and the mRNA may compensate for lack of Pab1-eIF4G1 binding (Park, Walker, *et al.*, 2011). Despite the slight inconsistencies between the two studies presented, both concluded that there appears to be alternative pathways and mechanisms for translation to take place which do not necessitate formation of the closed-loop in its canonical form.

1.3.2 Proposed alternative closed-loop conformations

The possibility of the closed-loop existing in states other than the canonical eIF4E-eIF4G-Pab1 one has also been suggested by as a result of directly studying translation complexes. Analytical centrifugation coupled with fluorescent detection was used to understand the stoichiometry of the closed-loop factors in translating complexes. It was observed that translating complexes varied in their composition regarding the closed-loop factors

depending on translation state, with eIF4E and eIF4G often dissociating from the translation complex during transition from initiation to elongation. Two predominant forms of translating monosomes and polysomes were also observed, one containing eIF4E/eIF4G and Pab1, and another containing only eIF4E and eIF4G (Wang, Xi, Toomey, Chiang, Hasek, Thomas M Laue, *et al.*, 2016).

1.3.3 Single molecule studies visualising ends

Single molecule visualisation studies have also indicated that translating mRNAs do not show close ends to the extent expected if the closed-loop was a constant feature of translation. Adivarahan *et al.* (2018) hybridised fluorescent probes to the 5' and 3' ends of mRNAs and used single molecule resolution fluorescent in situ hybridization (smFISH) to observe their co-localisation. Measuring the distribution of co-localisation distances between the 5' and 3' probes, they observed that average distances for the mRNAs tested were consistently larger than would be expected from mRNAs in a closed-loop conformation. To rule out the possibility that these mRNAs with large 5' and 3' distances were ones which were not actively translated, they repeated this study with translation inhibitors. The rationale was as follows – if the small subset showing good co-localisation (short 5' to 3' distances) were the 'translating' mRNAs, this subset would decrease in proportion upon disruption of translation. However, this was not observed. Furthermore, it was also observed that upon disassociation of polysomes following puromycin treatment, 5' to 3' distances decreased. From these observations, it was suggested that mRNAs with translating mRNPs are rarely arranged in a closed-loop conformation, and that disassembly of polysome resulted in compact mRNAs. Finally, a correlation between ribosome occupancy and the separation of the ends was reported, suggesting the act of translation was negatively associated with the closed-loop conformation (Adivarahan *et al.*, 2018). Similar observations were reported by Khong and Parker (2018), finding that in translating conditions, mRNA 5' and 3' ends were separated, whereas in non-translating conditions, the ends became closer.

1.3.4 Ends being intrinsically close together

As RNA is single-stranded, it has the ability to base-pair with itself, forming RNA secondary structures (Higgs, 2000). It has therefore been proposed that mRNAs may have the ability to

base pair with itself to form intrinsically close ends, rather than requiring the activities of the closed-loop components. Bioinformatic modelling predicted that the ends of RNA molecules should be close together independent of length and sequence due to the average proportion of paired bases in RNA molecules (Yoffe *et al.*, 2011). As experimental validation of this prediction, and of other bioinformatics studies with similar conclusions, single molecule Fluorescence Resonance Energy Transfer (smFRET) was used to measure the distance between 5' and 3' ends for a number of single-stranded RNA (ssRNA). RNAs were found to have short end-to-end distances despite differences in length, origin and secondary structure (Leija-Martínez, Casas-Flores, Rubén D. Cadena-Nava, *et al.*, 2014). Another study with an expanded set of bioinformatic predictions and experimental validations also concluded the same (Lai *et al.*, 2018).

1.3.5 Summary of various alternative models compared to the closed-loop model

The studies presented above differ widely in their conclusions regards mRNA circularisation and translation. Studies investigating closed-loop interaction mutants largely suggest that there may be some redundancy in the closed-loop interactions and the effect of the individual components on translation stimulation. Analytical ultracentrifugation studies suggest different closed-loop complexes can form, while single molecule studies conclude that the closed-loop is rarely found on translating mRNAs and is rather a feature of non-translating mRNAs. Finally, bioinformatic and FRET analyses have suggested mRNAs intrinsically form close ends without the requirement for protein bridging factors to facilitate this circularisation. As translation initiation is known to be complex, dynamic and kinetic, these findings may not all be mutually exclusive. Regardless, together they suggest that the role of the closed-loop in translation may not be as straightforward as initially hypothesised (Thompson and Gilbert, 2017; Vicens, Kieft and Rissland, 2018).

1.4 Could the closed-loop formation be mRNA dependent and how?

On one hand, a large body of work has detailed how the closed-loop can form and that it stimulates translation. On the other hand, there are also documented examples where formation of the closed-loop is not required for translation to take place. Although initially they may seem incongruent, these two bodies of work do not necessarily need to contradict

each other. One possible explanation of these two seemingly competing findings is that the closed-loop is forming, but only under certain conditions or on certain mRNAs. Some studies, in particular those relating to the regulation of the closed-loop components, suggest that it is possible the closed-loop functions at a more individual level, rather than at a global translation-wide level.

1.4.1 Differential closed-loop association and formation between mRNAs

One study lending weight to the idea of closed-loop formation being mRNA dependent is a transcriptome-wide study comparing the association of different mRNAs to the closed-loop factors in yeast (Costello *et al.*, 2015). The closed-loop proteins (eIF4E, eIF4G1, Pab1) were individually tagged and immunoprecipitated. Co-immunoprecipitating mRNAs were sequenced and the degree of enrichment above input was calculated to determine their extent of association with the each closed-loop protein. Differential association to the closed-loop factors was observed between mRNAs, with a class of mRNAs showing stronger preference for eIF4E and eIF4G1 compared to other mRNAs. Association with Pab1 was much less mRNA-dependent, consistent with the role of Pab1 as a general poly(A) binding protein. While increased association with the closed-loop factors is not direct evidence for these mRNAs forming more closed-loop, it is nevertheless an interesting finding, given that association is inevitably a pre-requisite for closed-loop formation.

To more directly demonstrate closed-loop formation, our research group developed a method to immunoprecipitate closed-loop complexes and use the relative enrichment of 5' and 3' mRNA sequences associated with the closed-loop as a means of comparing closed-loop formation (Archer *et al.*, 2015). Differences were observed between mRNAs in the extent of their relative end enrichments, suggesting a differential formation of the closed-loop. However, this study was limited by the size of the mRNA panel, consisting only of approximately 20 mRNAs, so it was difficult to draw transcriptome-wide observations and conclusions. This thesis focuses on extending this method for transcriptome-wide analysis, so the original study will be revisited in later sections, particularly in chapter 3.

1.4.2 Possibility for mRNA-specific regulation of the closed-loop

Given the extensive regulation of translation initiation, and the strong association of the closed-loop factors with this stage of translation, a host of regulatory elements are known to act on the closed-loop either directly or indirectly. Of these, some of these regulatory pathways have the potential to act in an mRNA-dependent manner.

For example, a group of eIF4E binding proteins, 4E-BPs, inhibit translation by competing with eIF4G1 for binding with eIF4E through a conserved binding motif (Sonenberg and Gingras, 1998). This is regulated by the mechanistic target of rapamycin (mTOR) pathway, in which mTOR phosphorylates 4E-BP1, resulting in its dissociation from eIF4E, allowing eIF4E to associate with eIF4G1 and stimulate translation (Gingras *et al.*, 1999). This regulation by mTOR may affect the translation of mRNAs to different extents depending on their reliance on forming the closed-loop for translation initiation. mTOR is generally downregulated under stress (Su and Dai, 2017), resulting in 4E-BPs remaining hypophosphorylated, and eIF4E sequestered from eIF4G1. If mRNAs do indeed have a tendency to form the closed-loop to varying degrees, the effect of stress may therefore affect different mRNAs differently, allowing for complex fine-tuning of protein expression in response to external stimuli.

eIF4G1 has also been reported as a target of regulation. It has been observed that eIF4G1 phosphorylation by Ksp1 during glucose starvation in yeast results in the degradation of specific mRNAs, notably glycolytic mRNAs, likely through recruitment of Dhh1 (Chang and Huh, 2018a). While the study did not observe Ksp1-mediated repression of global translation, the finding that phosphorylation state of eIF4G1 can recruit mRNA degradation-promoting factors such as Dhh1 is consistent with a potential role for the closed-loop in translational control more broadly. Another study sought to investigate the change in translation genome wide upon eIF4G depletion (Park, Zhang, *et al.*, 2011). Interestingly, depletion of eIF4G did not change translation efficiency (TE) for the majority of mRNAs. However, depletion of eIF4G was shown to narrow the range of TEs genome wide – mRNAs with higher than average TEs were generally translated worse upon eIF4G depletion, while mRNAs with lower than average TEs were translated better upon depletion. This suggested that eIF4G and the closed-loop is not essential for translation *in vivo*, but works to differentiate translation efficiencies between different mRNAs (Park, Zhang, *et al.*, 2011). This is also consistent with

the possibility of closed-loop formation being mRNA dependent and a means of translational regulation.

1.4.3 Potential features determining closed-loop formation

For formation of the closed-loop to regulate translation in an mRNA dependent manner, this would imply that there are defining features of mRNAs, which when recognised, would result in more or less formation of the closed-loop. These features could relate to mRNA function and the proteins they encode, to the structure or sequence of the mRNA, or with physical features of the mRNA molecule such as translation state or stage of the mRNA life cycle.

Some possible features that could determine closed-loop formation have been suggested. For example, mRNAs with short open reading frame (ORFs) lengths were reported to have greater association to the closed-loop factors (Thompson and Gilbert, 2017). However, as this correlation was based on a dataset measuring the association of these mRNAs to the individual closed-loop factors, namely eIF4E and eIF4G (Costello *et al.*, 2015), it is difficult to definitely conclude that this would translate to greater closed-loop formation. An eIF4G dependence for translation has also been reported for mRNAs with shorter 5' UTRs (Park, Zhang, *et al.*, 2011). This conclusion has a similar limitation to the previous, in that eIF4G dependence may not translate to dependence on formation of the whole closed-loop structure, especially considering eIF4G has additional roles beyond mRNA circularisation, for example in contacting RNA, eIF4A and eIF3 (Hentze, 1997; Prévôt, Darlix and Ohlmann, 2003).

Other possibilities remain to be tested. For example, it has been reported that the binding of microRNAs to mRNA can inhibit translation initiation, with both the cap and tail required to be present for full repression to be observed. eIF4E was also identified as a target of this miRNA-mediated repression (Humphreys *et al.*, 2005). Given the central role of these components in forming the closed-loop, it is possible that miRNAs can affect circularisation of the mRNA by the closed-loop, but this remains to be tested directly. mRNA secondary structure has also been presented as a potential determinant of closed-loop formation, with the possibility for some mRNAs to already have close ends from base pairing of their 5' and 3' regions, thus aiding the protein factors to overcome the steric cost of circularisation, resulting in more closed-loop formation (Leija-Martínez, Casas-Flores, Rubén D Cadena-

Nava, *et al.*, 2014). Finally, average poly(A) tail lengths of mRNAs could also play a role, as these are shown to differ between mRNAs (Tudek *et al.*, 2021), and Pab1 is known to require a minimum tail length for binding (Baer and Kornberg, 1983). A close examination of mRNA features, structures and sequence motifs in relation to their closed-loop forming status could help elucidate the roles of these features, if any, on formation of the closed-loop, and potentially also on any translation regulation taking place through its formation.

1.5 Hypothesis and Aims

Taking into account both the body of structural and biochemical evidence for the closed-loop's existence and role in stimulating translation, and the various studies suggesting the closed-loop is not essential for translation *in vivo*, we hypothesise that formation of the closed-loop is variable across the transcriptome. It is possible that instead of being a constant feature of cap-dependent translation, the closed-loop may be more nuanced in its presentation in living cells.

The literature already contains transcriptome-wide data of association with the closed-loop factors. It also contains single-molecule studies on individual mRNAs and their circularisation or lack thereof. However, there is currently a gap of transcriptome-wide information on closed-loop formation measured more directly than through individual factor associations. Previous findings from our research group suggest that the closed-loop may form to different extents on different mRNAs based on comparison of mRNA end enrichments, but this analysis was only performed for a panel of less than 20 mRNAs (Archer *et al.*, 2015). The overarching aim of this thesis was therefore to extend these observations to the transcriptome, with the hope of correlating mRNA-intrinsic and/or functional features to close-loop formation propensity.

Specifically, the aims of this project were to –

- 1) Capture the closed-loop complex and associated mRNAs *in vivo* using an immunoprecipitation-based experimental technique
- 2) Bioinformatically investigate the extent of the closed-loop across the transcriptome at steady state
- 3) Investigate mRNA features affecting differences in closed-loop formation between mRNAs

Chapter 2 - Materials and Methods

2 *Materials and Methods*

2.1 *Materials*

2.1.1 *Chemicals/reagents*

Table 1. Chemicals and Reagents

Chemical/Reagent	Manufacturer	Catalog number
1 M Tris pH 8.0	Invitrogen	AM9855G
10% IPEGAL (NP-40)	Sigma	CA-630
1x Complete mini EDTA free protease inhibitor	Roche	11836170001
20% SDS	Sigma	05030-500ML-F
25:24:1 Phenol:Chloroform:Isoamyl Alcohol (pH 7-8)	Sigma Aldrich	77617-100ML
3M Sodium Acetate pH 5.5	Invitrogen	AM9740
Acidic 25:24:1 Phenol:Chloroform:Isoamyl Alcohol (pH 4-5)	Sigma Aldrich	77619-100ML
Adenine sulfate dihydrate	Amresco	0607-50G
Agar	BD	214010
Bolt LDS sample buffer (4x)	Novex	B0007
Chloroform		
CircLigase II Kit	Lucigen	CL9021K
D-glucose	Sigma	49139-1KG
DL-Dithiothreitol solution (DTT)	Sigma- Aldrich	43816
Dulbecco's Phosphate Buffered Saline (DPBS)	Gibco	14190-144
DSN	Evrogen	
Dynabeads Antibody Coupling Kit	Thermofisher	14311D
Dynabeads M-280 Sheep Anti-Mouse IgG	Invitrogen	11202D
E. coli Poly(A) polymerase (EPAP)	NEB	M0276S

EDTA	Sigma- Aldrich	E7889-100ML
Ethyl alcohol, Pure	Sigma- Aldrich	E7023-1L
Formaldehyde	Sigma- Aldrich	F1635-500ML
Freeze and Squeeze DNA Gel Extraction Spin Columns	BioRad	7326165
Generuler 50 bp ladder	Thermo Scientific	SM0371
Generuler 1 kb plus ladder	Thermo Scientific	SM1331
Glycine	Sigma- Aldrich	G7126
Glycogen (5mg/mL)	Invitrogen	AM9510
HEPES	Sigma	H3375-500G
High Resolution Agarose for molecular biology	Sigma Aldrich	A4718-100G
iBlot 2 PVDF Mini Stacks	Invitrogen	IB24002
Intercept Blocking Buffer PBS	Li-cor	927-70001
Invitrogen 1 kb plus ladder	Invitrogen	10787018
Lithium Acetate Dihydrate	Sigma	L4158-250G
M-270 Epoxy Dynabeads	ThermoFisher	14301
Magnesium Chloride	Sigma	M1028-100ML
MES SDS Running Buffer	Invitrogen	B000202
MicroSpin G-25 Columns	Cytiva	27532501
Peptone	Sigma Aldrich	70178-500G
Platinum SuperFi DNA polymerase	Thermofisher Scientific	12351010
Poly(ethylene glycol)	Aldrich	P2139-500G
Potassium Chloride	Sigma	60128-1KG-F
Precision Plus Protein™ Kaleidoscope™	BioRad	1610375

Prestained Protein Standards		
RNase I, cloned, 100 U/μl	Invitrogen, Ambion	AM2294
RNaseOUT Recombinant Ribonuclease Inhibitor (40 U/μl)	Invitrogen	10777019
Safe Image Green DNA Stain	Abcam	ab286890
Sodium Chloride	Merck	1.06404.3000
SuperScript Reverse Transcriptase IV	Thermofisher Scientific	18090010
SYBR green PCR master mix	Applied Biosystems	4309155
T4 Polynucleotide Kinase (PNK)	NEB	M0201S
Trichloroacetic acid	Sigma	T0699-100ML
TURBO DNA-free kit	Thermofisher Scientific	AM1907
TWEEN 20	Sigma	P2287-500ML
Yeast extract	Sigma- Aldrich	70161

2.1.2 Buffers and solutions

Table 2. Buffers and Solutions

Name	Composition
10x Annealing buffer	250 mM HEPES-KOH pH 7.4, 1 M NaCl, 50 mM MgCl ₂
10X TE	100 mM Tris pH 7.5, 10 mM EDTA
1x THE	100 mM Tris, 200 mM HEPES (pH 5.5), 10 mM EDTA
Breaking buffer	2% (v/v) Triton X-100, 1% (v/v) SDS, 0.1 M NaCl, 0.01 M Tris-HCl pH 8, 1 mM EDTA

Cold buffer A (base)	25 mM HEPES pH 7.4, 140 mM KCl, 1.8 mM MgCl ₂ , 0.1% IPEGAL
Cold buffer A (supplemented)	Base buffer A supplemented with 0.5 mM DTT, 35 mM NaCl, 1x Complete EDTA-free protease inhibitor cocktail
HBB	20 mM HEPES pH 7.4, 100 mM KCl, 2 mM MgCl ₂
HBB-DRP	HBB buffer supplemented with 1x Complete mini EDTA free protease inhibitor cocktail (Roche), 0.5 mM DTT, 200 U RnaseOUT per sample)
HBB-glycine	250 mM glycine in 1xHBB
HE	10 mM HEPES pH 7.4, 0.1 mM EDTA
IP wash buffer 1	base buffer A with 2 mM DTT
IP wash buffer 2	base buffer A with 4 mM DTT
PBST	0.1% Tween®-20 in 1x PBS.
SETG	1% SDS, 10 mM EDTA, 10 mM Tris-HCl pH 2, 10 mM glycine
STOP solution 1	10% w/v SDS, 100 mM EDTA
STOP solution 2	20 mM EDTA in 1x HE (10 mM HEPES pH 7.4, 0.1 mM EDTA)
Trafo-1	100 mM LiAc in 1x TE (10 mM Tris pH 7.4, 1 mM EDTA)
Trafo-2	100 mM LiAc in 40% (w/v) PEG
YEPD	1% w/v yeast extract, 2% w/v peptone, 2% w/v glucose, 40mg/l adenine sulfate

2.1.3 Antibodies

Table 3. Antibodies

Antibody	Dilution	Manufacturer	Catalog number
Mouse IgG2a, Kappa Monoclonal antibody	1:1000	Abcam	Ab18413
Anti-FLAG	1:1000	Abcam	F1804-200UG

Anti-protein A	1:1000	Abcam	Ab6026
Anti-GAPDH	1:1000	Abcam	Ab125247
Goat Anti-Rabbit	1:20000	Licor	926-68071
Secondary			
Goat Anti-Mouse	1:20000	Licor	926-32210
Secondary			
Anti-Pab1	1:100	Santa-Cruz	Sc-166381 (A-4)
eIF4E	1:100	Santa-Cruz	Sc-9976 (P-2)
eIF4G	1:100	Santa-Cruz	Sc-133155 (A-10)
Beta-actin-Alex	1:500	Santa-Cruz	Sc-4778 AF790
Fluor 790			

2.1.4 Yeast strains

Table 4. Yeast strains

Name	Background	Genotype
Wild type (WT)	W303-1A	leu2-3,112; trp1-1; can1-100; ura3-1; ade2-1; his3-11,15
eIF4E-TAP	W303-1A	WT + CDC33::TAP
eIF4E-TAP, eIF4G1-FLAG	W303-1A	WT + CDC33::TAP; TIF4631::3XFLAG
eIF4E-TAP, Pab-FLAG	W303-1A	WT + CDC33::TAP; PAB1::3XFLAG
eIF4G1-pro(A)	W303-1A	WT + TIF4631::PRO(A)
eIF4G1-pro(A), eIF4E-FLAG	W303-1A	WT + TIF4631::PRO(A); CDC33::3XFLAG
eIF4G1-pro(A), Pab-FLAG	W303-1A	WT + TIF4631::PROA; PAB1::3XFLAG
Pab1-pro(A)	W303-1A	WT + PAB1::PRO(A)
Pab1-pro(A), eIF4E-FLAG	W303-1A	WT + PAB1::PRO(A); CDC33::3XFLAG

Pab1-pro(A), eIF4G1-FLAG	W303-1A	WT + PAB1::PROA; TIF4631::3XFLAG
-------------------------------------	---------	----------------------------------

2.1.5 Primers and oligonucleotides

Table 5. Tagging primers

Name	Used for	Sequence (5' → 3')
eIF4E cassette F	Generate 3xFLAG cassette for	AAATTCAAGCAAGTTTAAAATTAACCGATGACGGG CATTTGGAATTCTTTCCACATTCCAGTGCCAATGGTA GACACCCTCAACCATCAATCACCTTGCGGATCCCCGG GTTAATTAAC
eIF4E cassette R	insertion into <i>CDC33</i> (eIF4E)	GTGCATCCTAACTAGTAGAATGTATATATATTATATT ATGATTATATATATAATTTTGATTAAAATAACAATTA TCTTAAGAAAAATTTCAGACTATCTTACATCGATGAAT TCGAGCTCG
eIF4G1 cassette F	Generate 3xFLAG cassette for	AGCCAAAGGGCTCCTCCTCCAAAGGAAGAACCAGCT GCACCAACTTCTACCGCAACAAATATGTTTCAGTGCAT TAATGGGAGAAAGTGATGACGAAGAGCGGATCCCCG GGTTAATTAAC
eIF4G1 cassette R	insertion into <i>Tif4631</i> (eIF4G1)	ATTTGTGTTGTACTATTAATTTATTATTTTATTAGACT TTCTACCAACATCCTTGTATCCAAGTGACATTTTCGA TACTTAACATGATCTATTCATGTTACATCGATGAATT CGAGCTCG
PAB1 cassette F	Generate 3xFLAG cassette for	TTGTTGGAAAGTGATGAATTGTTTCGAACAACACTAC AAAGAAGCTTCTGCTGCCTATGAGTCTTTCAAAAAG GAGCAAGAACAACAACTGAGCAAGCTCGGATCCCC GGGTTAATTAAC
PAB1 cassette R	insertion into <i>PAB1</i> (PAB1)	AGGATAATAACTAGTATTTAAGACAAAAAAGG GAAGGGAGAAAAAAGATGATAAGTTTGTTGAGTA GGGAAGTAGGTGATTACATAGAGCATTACATCGATG AATTCGAGCTCG

Table 6. qPCR primers

Name	Used for	Sequence (5' → 3')
SSC1 5' F	End enrichment analysis	TACGTTGCTGTCCTCCTCTG
SSC1 5' R	End enrichment analysis	TGTCACTCTGGCAAAGGTTG

SSC1 mid F	End enrichment analysis	CCTGCCATTTATCACTGCTG
SSC1 mid R	End enrichment analysis	TGAGCCCTGGAGAACTTCAT
SSC1 3' F	End enrichment analysis	ATATACACGCCCCAAACATA
SSC1 3' R	End enrichment analysis	ATACATGGCTTTCCCGCTTA
RPS20 5' F	End enrichment analysis	ACAAAATGTCTGACTTTCAAAAGGA
RPS20 5' R	End enrichment analysis	TTGTTGTTCTTGTTCTTCAACCTT
RPS20 mid F	End enrichment analysis	GGTCAAGAAGGGTCCAGTCAG
RPS20 mid R	End enrichment analysis	TTGGAGTCTTTCTGGTGGAGA
RPS20 3' F	End enrichment analysis	TCTCATACTTTTGTTGATTCTTGTGT
RPS20 3' R	End enrichment analysis	ACGCTTGCTTTGTTTTGTCTTT
GUS1 5' F	End enrichment analysis	GATGCCATCTACCTTGACTATT
GUS1 5' R	End enrichment analysis	TTAGTTCAGCATAAGCCACAA
GUS1 mid F	End enrichment analysis	CCGAAGAAATGAAAAACGGTA
GUS1 mid R	End enrichment analysis	ATCGATCTTGGCACGAACAC
GUS1 3' F	End enrichment analysis	TGGTGCAAAGAAATAAGCATCACA
GUS1 3' R	End enrichment analysis	CAAAAGGTACTGGTTGAGAAATAACA

Table 7. Primers for library preparation

Name	Used for	Sequence (5' → 3')
p155	Split adapter	5'-phosphate-GA TCG TCG GAC TGT AGA ACT CTG AAC G /9-carbon spacer/ G TGA CTG GAG TTC CTT GGC ACC CGA GAA TTC CAT TTT TTT TTT TTT TTT TTT TVN-3'
p156	Empty adapter circularisation blocker	5'-TTN BAA AAA AAA AAA AAA AAA AAA /iSuper-dT/GG AAT TCT CGG GTG CGT GTG T/3BioTEG/-3'
p157	Circularisation block remover	5'-ACA CAC GCA CCC GAG AAT TCC ATT TTT TTT TTT TTT TTT TTT T/iSuper-dT/T VNAA/3BioTEG/-3'
F	Library amplification barcode primer	CAA GCA GAA GAC GGC ATA CGA GAT XXX XXX GTG ACT GGA GTT CCT TGG CAC CCG AGA ATT CCA
R	Library amplification reverse primer	5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACG TTC AGA GTT CTA CAG TCC GA-3'

2.1.6 Depletion probes for probe-directed degradation

Note: This table was copied directly from Janapala *et al.* (2022), which was adapted from (Archer *et al.*, 2016).

Table 8. rRNA depletion probes

DNA probe name	Target	Sequence 5' → 3'
18S-1	18S-rRNA	CTGGTTGATCCTGCCAGTAGTCATATGCTTGTC
18S-2	18S-rRNA	CAAAGATTAAGCCATGCATGTCTAAGTATAAGC
18S-3	18S-rRNA	CAGTGAAACTGCGAATGGCTCATTAAATCAG
18S-4	18S-rRNA	CATGGTATAACTGTGGTAATTCTAGAGCTAATACATGC
18S-5	18S-rRNA	TGGAAGAGATGTATTTATTAGAT
18S-6	18S-rRNA	TGATGATTCATAATAACTTTTCGAATCGCATGGC
18S-7	18S-rRNA	TGGCGATGGTTCATTCAAATTTCTGCCCTATCAACTT
18S-8	18S-rRNA	TCCTAATTCAGGGAGGTAGTGACAATAAATAACGA
18S-9	18S-rRNA	TACCTTAACGAGGAACAATTGGAGGGCAAGTC
18S-10	18S-rRNA	GGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCC
18S-11	18S-rRNA	TTCCAACGGGGCCTTT
18S-12	18S-rRNA	AAAATTAGAGTGTTCAAAGCAGGCGT
18S-13	18S-rRNA	TTGCTCGAATATATTAGCATGGAATAATAGAAT
18S-14	18S-rRNA	ATGATTAATAGGGACGGTCGGGGGCAT
18S-15	18S-rRNA	AAGGACGTTTTTCATTAATCAAGAACGA
18S-16	18S-rRNA	ACTATGCCGACTAGGGATCGGGTGGTGTT
18S-17	18S-rRNA	CCCACTCGGCACCTTACGAGAAATCAAAGTCTT
18S-18	18S-rRNA	AGGTCCAGACACAATAAGGATTGACAGAT
18S-19	18S-rRNA	AGTTGGTGGAGTGATTTGT
18S-20	18S-rRNA	CGCTACACTGACGGAGCCAGCGAGT
18S-21	18S-rRNA	GATTGAATGGCTTAGTGAGGCCTCAGGATC
18S-22	18S-rRNA	AGCGGAGAATTTGGACAACTTGGTCATT
18S-23	18S-rRNA	ACCTGCGGAAGGATCATTAA
18S-24	18S-rRNA	ACCTGCGGAAGGATCATTAAAAAA
25S-1	25S-rRNA	GACCTCAAATCAGGTAGGAGTACCCGCTGAACT
25S-2	25S-rRNA	GGAAAAGAAACCAACCGGGATTGCCTTAG
25S-3	25S-rRNA	CTTGGAACAGGACGTCATAGAGGGTGAGAATC

25S-4	25S-rRNA	AGAGTCGAGTTGTTTGGGAATGCAGCTCTAAGT
25S-5	25S-rRNA	CAGTGATGGAAAGATGAAAAGAACTTTGAAAAG
25S-6	25S-rRNA	GCCAGCATCAGTTTTGGTGGCAGGAT
25S-7	25S-rRNA	AATACTGCCAGCTGGGACTGAGGACTG
25S-8	25S-rRNA	TCTAACGTCTATGCGAGTGTTTGGGTGTAAAC
25S-9	25S-rRNA	CAAGAGGTGCACAATCGACCGATCCTGA
25S-10	25S-rRNA	ATCGAACCATCTAGTAGCTGGTTCCTGCCGAAGT
25S-11	25S-rRNA	ACGTAGAGTTAAGGTGCCGGAATACACGCTC
25S-12	25S-rRNA	GGACGTGGGTTAGTCGATC
25S-13	25S-rRNA	GGATTCTTCACGGTAACGTAAGTGAATG
25S-14	25S-rRNA	CAGCTTATCACCCCGGAATTGGTTTA
25S-15	25S-rRNA	CCGGAGATGGGGTCTT
25S-16	25S-rRNA	GCTTGTGACGGCCCGTGAAAATCCACAGGAAGG
25S-17	25S-rRNA	TCCACAGGAAGGAATAGTTTTTCATGCCAGGTCGTAC
25S-18	25S-rRNA	TCAAAGTGAAGAAATTCAACCAAGCGCGGGTAA
25S-19	25S-rRNA	GAGGGTGTAGAATAAGTGGGAGCTTCGGCGC
25S-20	25S-rRNA	CACGTTCTAGCATTCAAGGTCCC
25S-21	25S-rRNA	ATACCGAAGCAGAATTCGGTAAGCGTTGGATTGT
25S-22	25S-rRNA	GATGGATACGAATAAGGCGTC
25S-23	25S-rRNA	TGCTGGCGAATTGCAATGTC
25S-24	25S-rRNA	AATCATTTGTATACGACTTAGATGTACAACGGG
5.8S-1	5.8S-rRNA	GGGCATGCCTGTTTGAGCGTCATT
5S-1	5S-rRNA	ACGCGAAACTCAGGTGCTGCAATCT

2.2 *Methods*

2.2.1 *Commonly used methods*

2.2.1.1 *Ethanol precipitation of RNA and DNA*

Unless otherwise specified, purified RNA and DNA were obtained by phenol-chloroform extraction and ethanol precipitation. Aqueous solutions containing the nucleic acids were added to an equal volume of 25:24:1 Phenol:Chloroform:Isoamyl Alcohol (Sigma Aldrich), at either pH 4-5 (RNA) or pH 7.7-8.3 (DNA). The mixture was vortexed at maximum speed

for 2 min, then centrifuged at 13,000×g for 10 min at RT on the Corning LSE High Speed microcentrifuge. The upper aqueous phase was carefully removed, leaving behind a small amount to avoid protein and phenol-chloroform contamination from the interphase. To precipitate the nucleic acid, the aqueous layer was added to 2.5 volumes 100% ethanol (Sigma-Aldrich), 1/10 volume 3M sodium acetate (Invitrogen) and 20 µg glycogen (Invitrogen), vortexed for 15 s at max. speed and incubated overnight at -80 °C or until required. After incubation at -80 °C, samples were thawed at RT briefly to minimise salt co-precipitation, then centrifuged at 13,000×g for 30 min at RT to collect the pellet. Supernatant was removed and pellets were washed twice by addition of 1 mL 80% (v/v) ethanol (Sigma-Aldrich) and further centrifugation at 13,000×g for 5 min at RT. The wash supernatant was removed and pellet dried at RT for 15 min to evaporate residual ethanol. Pellets were resuspended in 30 µl 1x HE buffer (10 mM HEPES pH 7.4, 0.1 mM EDTA) and stored at -80 °C (RNA) or -20 °C (DNA) until required.

2.2.1.2 Agarose gel electrophoresis

Agarose gels were prepared by addition of High-resolution Agarose (Sigma-Aldrich) to 1x THE buffer (100 mM Tris, 200 mM HEPES (pH 5.5), 10 mM EDTA). Agarose was dissolved by boiling until clear, and the resultant gel mixture was poured into a casting tray and set for at least 45 min at 4 °C. In general, final agarose concentrations of 0.8% (w/v) were used for visualisation of DNA fragments above 2.5 kb, 1-1.5% (w/v) for fragments between 500 bp to 2.5 kb and 2% (w/v) for fragments under 500 bp. 1 µl of Safe Image Green DNA Stain (Abcam) was added per 5 µl of sample and to 5 µl of either the 1kb plus DNA ladder (Invitrogen) or Generuler 1kb plus DNA ladder (Thermo Scientific). For library preparation (see chapter 3), the 50 bp generuler ladder (Thermo Scientific) was used instead. Samples and ladders were run on gels in the Bio-Rad electrophoresis tank connected to the BioRad POWERPAC 300, typically at 80-100V for 1.5 h.

2.2.1.3 Western blotting

Western blotting was performed by adding 4x Bolt LDS Sample buffer (Novex) to samples to a final concentration of 1x loading dye and DTT to a final concentration of 0.1 M. Samples were heated at 95 °C for 2-3 min and loaded on 4-12% Bolt gels in the Mini Gel Tank

(Invitrogen), alongside 3 µl of the pre-stained protein ladder (BioRad). Protein gels were typically run in 1x MES running buffer (Invitrogen) at 80V for approximately 1-1.5 h. Upon completion of run, gels were transferred using the iBlot dry transfer system (Invitrogen) and iBlot 2 PVDF Mini stacks (Invitrogen). Membranes were blocked for 1 h with Intercept Blocking Buffer (PBS) (Licor), then incubated overnight at 4 °C with the primary antibody at a 1:1000 dilution in a mixture of 1:1 PBST (0.1% Tween®-20 in 1x PBS) and Intercept blocking buffer (Licor). Following primary antibody incubation, membranes were washed with 1x PBST (0.1% Tween®-20 in 1x PBS) for 3x 10-minute washes. Membranes were incubated in light-protected boxes for 1 h with Licor secondary antibodies at dilutions of 1:2000 in 1:1 PBST/blocking buffer. Finally, membranes were washed 3x for 5 min each with PBST, then imaged using the Odyssey CLx Imaging System (Licor) at the appropriate wavelength (680 or 800 nm). If quantification was needed, the 'lane' analysis in ImageStudio was used. If incubation of the same membrane with another antibody was required, membrane was stripped for 15 min with 1x Reblot Plus Strong stripping solution (Millipore) and blocked for 1 h before repeating primary and secondary antibody incubations.

2.2.1.4 DNase digestion, reverse transcription and qPCR

Up to 10 µg of extracted RNA was transferred to a new 0.5 mL tube. Turbo DNase buffer (ThermoFisher Scientific) was added to a final concentration of 1x in 50 µl. 2 units of TURBO DNase (ThermoFisher Scientific) was added to each and volume adjusted to 50 µl. Samples were incubated at 37 °C for 30 min. 0.1 volumes DNase inactivation resin was added (ThermoFisher Scientific). Tubes were incubated for 5 min at RT with gentle flicking. Samples were centrifuged at 10,000 ×g for 1.5 min at RT and RNA transferred to a new tube.

A 2 µM gene specific primer mix was made from 20 µM working stocks of the relevant reverse primers. 1 µl 2 µM primer mix and 1 µl 10 mM dNTP mix was added to up to 5 µg DNase-treated RNA and final volume adjusted to 13 µl. Tubes were incubated at 65 °C for 5 min then cooled on ice for 1 min. Each sample was split into 2 for reverse transcription and no RT reactions. SuperScript Reverse Transcriptase IV (ThermoFisher Scientific) reaction buffer was added to a final concentration of 1x, followed by 5 mM DTT (Sigma-Aldrich), 1 µl RNase OUT and 200 U superscript reverse transcriptase (or 1 µl milliQ for no RT

reaction). Reaction mixes were incubated at 50 °C for 10 min, then at 80 °C for 10 min to inactivate the enzyme.

cDNA template samples were diluted with 1xHE buffer (10 mM HEPES pH 7.4, 0.1 mM EDTA). A mastermix was prepared from SYBR green reaction mix (Applied Biosystems), forward primer, reverse primer and milliQ in the following proportions:

Table 9. qPCR master mix components

Component	Stock concentration	Volume per reaction [μl]
SYBR green reaction mix	2X	5
Forward primer	10 μM	0.375
Reverse primer	10 μM	0.375
H2O	NA	0.25
Total		6

4 μl of the diluted cDNA template was added to 6 μl of master mix in technical triplicates, with the addition of a technical triplicate of 4 μl milliQ as a no template control in a 384 well MicroAmp optical reaction plate (Applied Biosystems). qPCR reactions were performed on the QuantStudio 12k Flex Realtime PCR with the SYBR reagent setting, and a thermal cycling profile of 60 s at 95 °C, followed by 40x cycles of [95 °C for 15 s, 60 °C for 30 s], followed by a final melt curve.

2.2.1.5 Polysome profiling

Polysome profiling was performed to compare translation profiles between different strains. To prepare gradients, linear 10%-40% sucrose gradients were made using the freeze-thaw method. First, two sucrose solutions, one at 10% (w/v) and one at 40% (w/v), were each buffered to a final concentration of 25 mM HEPES pH 7.6, 0.1 mM EDTA (Sigma-Aldrich), 100 mM KCl (Sigma) 5.1 mM MgCl₂ (Sigma) and 2 mM DTT (Sigma-Aldrich). The two buffered sucrose solutions were mixed in appropriate proportions to create four sucrose solutions of 10% (w/v), 20% (w/v), 30% (w/v) and 40% (w/v). 3 mL of buffered 40% sucrose solution was dispensed into Thinwall Ultra-Clear ultracentrifuge tubes (14 mL, Beckman Coulter), tubes were covered in parafilm and frozen at -80 °C for 15 min. Gradients were

removed, and keeping on ice, the 30% sucrose solution was dispensed above the frozen 40% layer as described. Tubes were returned to -80 °C again for 15 min, and this process was repeated until all sucrose solutions were dispensed. Gradients were stored at -80 °C and thawed at 4 °C overnight prior to use. To prepare cell lysates for polysome profiling, approximately 100 mg of yeast grindate prepared according to section 2.2.7 was thawed and resuspended in HBB-DRP buffer (HBB buffer supplemented with 1x Complete mini EDTA free protease inhibitor cocktail (Roche), 0.5 mM DTT, 200 U RnaseOUT (Invitrogen) per sample) then clarified 2x by centrifugation at 12,000×g for 5 min at 4 °C. Absorbance was quantified for each sample by Nanodrop. Typically, 10-12 Absorbance units (absorbance at 260 nm x volume (in mL)) of each sample was carefully loaded at the top of each thawed gradient. Tubes were balanced with milliQ, then ultracentrifuged in a SW 41 Ti rotor at 41,000 rpm for 1 h 30 min at 4 °C. Absorbance profiles of the sedimentation-separated samples were read at 254 nm, 1.5 mL/min using the Gradient Fractionator (Brandel) according to manufacturer's instructions.

2.2.2 Yeast tagging and mutation

Overall, yeast strains were tagged or mutated by exploiting the ability of *S. cerevisiae* to integrate exogenous linear DNA into its genome via homologous recombination (see Figure 2.1). A cassette containing both the sequence to be inserted and a selection marker conferring antibiotic resistance was amplified by PCR. Primers for this amplification were designed with 50-100 bp overhangs homologous to the region around the proposed insertion site. The final cassettes (containing the target sequence, antibiotic selection marker and two regions of homology immediately before and after the target insertion site) were transformed into the target strains, allowed to grow on selective antibiotic plates, and colonies were screened by colony PCR and sanger sequencing for successful fragment incorporation.

2.2.2.1 Generating cassette for 3xFLAG tagging

3xFLAG tag cassettes for transformation were generated by amplification of the 3xFLAG tag and NatNT resistance cassette from plasmid pFA6a-3xFLAG-NatNT, using overhang primers with 100 bp homology to the region either side of the planned insertion site, on either *CDC33* (eIF4E), *Tif4631* (eIF4G1), or *PAB1* (PAB1) (see Table 5 for primer information,

Figure 2.1 for overall strategy). Cassettes amplification reactions were carried out using Phusion DNA polymerase (NEB) according to manufacturer's instructions (1x Phusion buffer, 0.2 mM dNTPs, 0.5 μ M each F and R primer, ~10-50 ng plasmid, 2 U Phusion DNA polymerase), with addition of 1 μ l DMSO (NEB) per reaction. Thermal cycling conditions consisted of 2 min at 98 $^{\circ}$ C, followed by 32x cycles of [98 $^{\circ}$ C for 10 s, 61 $^{\circ}$ C for 45 s, 72 $^{\circ}$ C for 2 min], then a final extension of 5 min at 72 $^{\circ}$ C. The reaction was purified using the NucleoSpin Gel and PCR clean up kit (Macherey-Nagel) according to manufacturer's instructions. To avoid false positive transformants induced by transformation of the intact plasmid into yeast cells, the original plasmid template was digested by addition of 20 U *DpnI* (NEB) in 1x CutSmart buffer (NEB) followed by incubation at 37 $^{\circ}$ C for 1h. The reaction mixture was immediately purified again with the PCR clean up kit (Macherey-Nagel) and the final cassette concentration determined on the Nanodrop One. Approximately 1 μ g of each cassette was used for transformation according to the protocol in section 2.2.2.2.

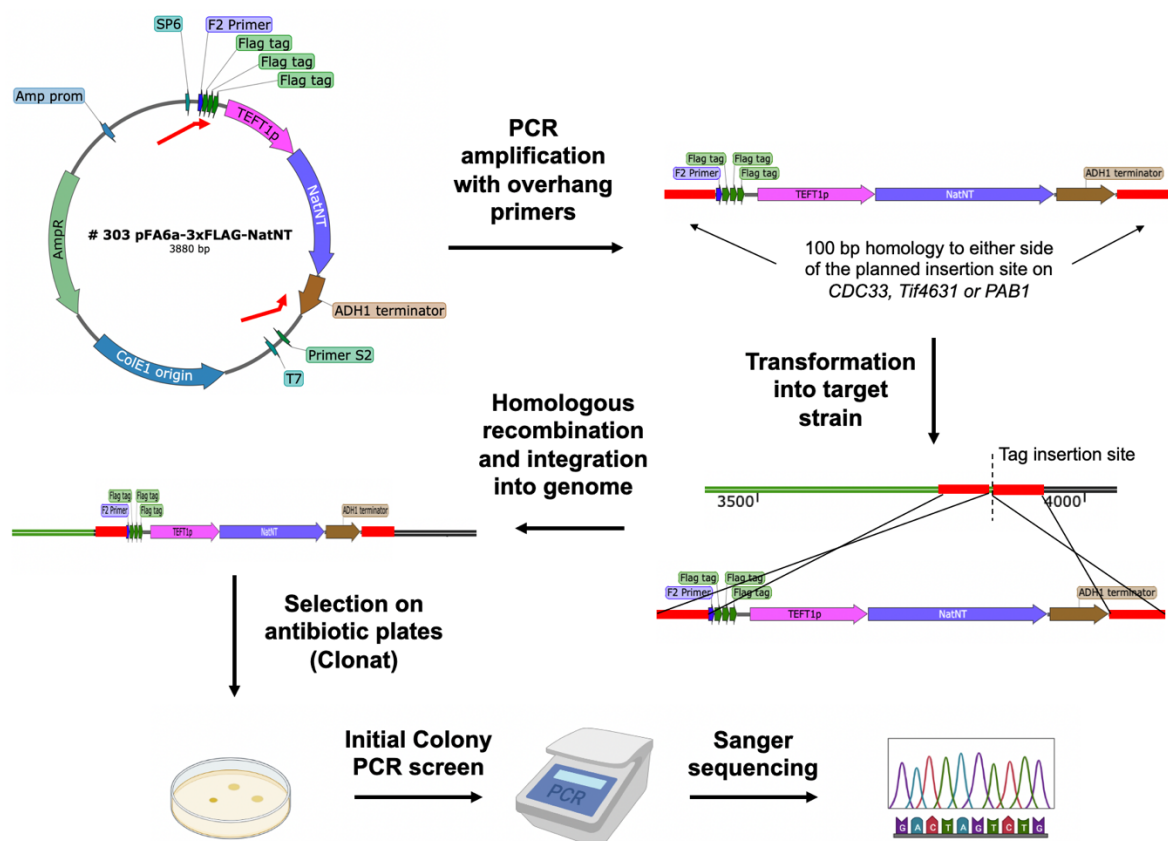


Figure 2.1. Overall yeast tagging strategy for 3xFLAG insertion

2.2.2.2 *Common transformation, colony PCR and genomic DNA extraction protocol*

Although the specifics of each cloning procedure varied, all cloning shared the same protocol below for transformation, colony PCR and gDNA extraction for sanger sequencing. This follows from cassette generation as described in previous sections.

In preparation for transformation, YPD media (1% w/v yeast extract, 2% w/v peptone, 2% w/v glucose, 40mg/l adenine sulfate) was inoculated overnight with the yeast strain to be targeted for transformation. On the day of transformation, the yeast culture was diluted with fresh YPD media to an OD₆₀₀ of 0.2 in 50 mL and allowed to grow at 30 °C with gentle shaking until an OD of 0.7. Cultures were centrifuged at 600 ×g for 2 min at RT and supernatant was removed. Cells were washed with 20 mL autoclaved water, then centrifuged again at 600×g for 2 min at RT. Cells were resuspended in 1 mL trafo-1 (100 mM LiAc in 1x TE (10 mM Tris pH 7.4, 1 mM EDTA)), transferred to a new 1.5 mL Eppendorf, centrifuged again at 2600 rpm for 2 min at RT, then resuspended in 250 µl of trafo-1. In parallel, herring sperm single-stranded DNA carrier was heated up to 95 °C for 5 min then snap-chilled on ice for at least 2 min in preparation for transformation. The transformation mixture was set up with 300 µl of trafo-2 solution (100 mM LiAc in 40% (w/v) PEG), 50 µl of the washed cells, 20 µl of ssDNA carrier and approximately 1-1.5 µg of each cassette bearing the feature to be inserted, antibiotic resistance and flanking homology regions. A negative control reaction was assembled in parallel with the same components, except for substitution of the DNA cassette with an equivalent volume milliQ. The transformation reactions were mixed by gentle pipetting and incubated on a turning wheel for 1 h at 30 °C. Reactions were heat shocked for 17 min at 42 °C, then supplemented with 1 mL YPD. Tubes were centrifuged at 1200×g for 2 min at RT, supernatant was removed and cells resuspended in 1 mL YPD. Cells were returned to the turning wheel for 3 h at 30 °C, then centrifuged at 600×g, and finally resuspended in 100 µl YPD. The transformation reaction and the negative reaction were each divided and spread between one YPD-only plate and one YPD plate supplemented with 100 µg/mL nourseothricin. Plates were grown at 30 °C for 2-3 days and growth was checked. The viability of the transformed cells was confirmed by presence of transformed cells on YPD-only plates (positive control) and the selectivity of the antibiotic selection plates was confirmed by lack of growth of non-transformed cells on antibiotic-positive plates (negative

control). Typically, up to 8 colonies of transformed cells on the antibiotic plate were picked and screened by colony PCR and sanger sequencing.

To screen colonies for successful transformation, a pipette tip was used to pick cells from each colony and added to individual wells of a PCR strip containing 5 μ l zymomix (1.2 M sorbitol, 0.5 mg/mL zymolyase 100 T, 100 mM NaH_2PO_4). A mix of the parental strain (untransformed strain) with zymolyase was also included alongside colonies as a negative control. The mix was incubated for 15 min at 37 °C, then zymolyase deactivated by heating to 98 °C for 2 min. The reaction was stopped by immediate transfer to ice and addition of 95 μ l milliQ. 2 μ l of each mixture was used as template for PCR and amplified according to manufacturer's instructions (1x Taq buffer, 0.2 μ M dNTPs (each), 0.5 μ M each forward and reverse colony PCR primers, 0.5 U HotStart Taq (NEB)). The PCR thermal cycling profile consisted of 95 °C for 2 min, 35x cycles of [95 °C for 30 s, 49-55 °C for 30 s (depending on recommended annealing temperature of primers), 72 °C for 1 min], then final extension of 10 min at 72 °C.

5 μ l of each reaction was run on an agarose gel as described in section 2.2.1.2 and imaged to check for correct insertion of cassette. Positive colonies were re-streaked and grown overnight at 30 °C for genomic DNA extraction and sanger sequencing. To extract gDNA for sanger sequencing, one small inoculation loop of cells was harvested from each individual re-streaked colony and added to 1 mL of milliQ in a 2 mL screw-top tube (Sarstedt), then centrifuged at 1200 $\times$ g for 1 min at RT. 200 μ l of silica beads were added to each tube, followed by 200 μ l breaking buffer (2% (v/v) Triton X-100, 1% (v/v) SDS, 0.1 M NaCl, 0.01 M Tris-HCl pH 8, 1 mM EDTA) and 200 μ l 25:24:1 Phenol:Chloroform:Isoamyl Alcohol (pH 7-8). Cells were lysed by 2x 20 s vortex at 5000 rpm with the Precelly's 24 bead lysis machine, with 1 min rest on ice in between. The layers were separated by centrifugation at 12,000 $\times$ g for 10 min at 4 °C. The upper aqueous layer was removed and added to an equal volume chloroform (Merck), vortexed for 1 min and centrifuged at maximum speed for 1 min. The supernatant was again removed and added to 1 volume chloroform, vortexed for 1 min, and centrifuged at maximum speed for 1 min. The upper, aqueous layer was removed and added to a fresh DNA low-bind tube with 2.5 volumes 100% ethanol, 1/10 volume sodium acetate and 20 μ g glycogen. Ethanol precipitation took place at -80 °C overnight, then tubes were briefly thawed and centrifuged at 13,000 $\times$ g for 10 min at 4 °C. The

supernatant was removed and pellet was washed with 500 μ l cold 80% (v/v) ethanol, then centrifuged again at 13,000 $\times$ g for 10 min at 4 °C. Ethanol was removed completely and pellets were allowed to air dry for 15-30 min, before resuspension in 30 μ l 1xTE (10 mM Tris pH 7.5, 1 mM EDTA). The resuspended gDNA was incubated with 1 μ l Rnase A/T1 (2 mg/mL Rnase A, 5000 U/mL Rnase T1) at 37 °C for 1 h to remove any residual RNA, then quantified using the Nanodrop One. DNA samples were submitted to the BRF Sanger sequencing facility at JCSMR with the relevant primers. Sanger sequencing results were compared to the expected sequence to screen for any deleterious (frameshift, nonsense, substitution etc) mutations. Finally, strains were also screened by western blotting according to section 2.2.1.3 to check for correct protein size and/or tag incorporation.

2.2.3 Growth curve measurements

S. cerevisiae strains to be measured were inoculated from yeast plates stored at 4 °C into liquid cultures of YPD (1% w/v yeast extract, 2% w/v peptone, 2% w/v glucose, 40mg/l adenine sulphate) and allowed to grow overnight at 30 °C with gentle shaking. Cultures were diluted to an OD600 of 0.2 in 50 mL fresh YPD the next morning. 2 mL of culture was removed every hour and immediately measured with the V-1200 spectrophotometer (VWR), with care taken to minimise the time flasks spent outside of the incubator during measurement. Measurements were taken over 7 h.

2.2.4 Bead binding capacity

2.2.4.1 Initial beads trialed

Either 50 μ l, 100 μ l or 150 μ l of 10 mg/mL Dynabeads M-280 Sheep Anti-Mouse IgG was aliquoted into DNA low bind tubes and commercial buffer was exchanged for 1 mL cold base buffer A (25 mM HEPES pH 7.4, 140 mM KCl, 1.8 mM MgCl₂, 0.1% IPEGAL). Beads were resuspended by gentle inversion, collected again, then resuspended in 80 μ l cold supplemented buffer A (base buffer A supplemented with 0.5 mM DTT, 35 mM NaCl, 1x Complete EDTA-free protease inhibitor cocktail (Roche)). 100 μ l of supplemented closed buffer A and 200 units Rnase OUT was added to 100 mg of grindate from either the PAB pro(A) tagged, eIF4E-3xflag stain or the untagged WT strain. Grindate tubes were vortexed and resuspended. Cell debris was pelleted by centrifugation at 12,000 $\times$ g for 5 min at 4 °C.

Supernatant was transferred to a new tube and clarified again by centrifugation as described. The final supernatant was transferred to a new tube and absorbance at 260 nm was measured on a nanodrop spectrophotometer (Thermo Scientific). 20 μ l of lysate was added to each tube of pre-prepared with equilibrated beads. 20 μ l of lysate from each strain was added to another tube containing 80 μ l of supplemented cold buffer A as the input sample. All input aliquots and bead-lysate reactions were rotated for 30 min at 22 °C. After 30 min of incubation, all tubes were placed immediately on ice. Input aliquots were divided into two for protein (10%) and RNA (10%) analysis. The aliquot for protein analysis was frozen, while the aliquot for RNA extraction was extracted with method 1 (see section 2.2.8.1 below). Meanwhile, the bead-lysate IP mixture was placed on a magnetic rack to collect beads. The reaction buffer was removed and beads were rinsed with 1 mL ice-cold wash buffer 1 (base buffer A with 2 mM DTT). Beads were again washed with wash buffer 1 by gentle rotation for 5 min at 4 °C then again with 1 mL of wash buffer 2 (base buffer A with 4 mM DTT). The beads were split into two new tubes in an 80/20 split for RNA analysis and protein analysis respectively. The final wash was taken off and beads were placed on ice. Beads set aside for protein analysis were frozen. Protein analysis was performed as described in 2.2.1.3.

2.2.4.2 *FLAG bead pulldown*

50 μ l of Anti-FLAG M2 magnetic beads (50% suspension in glycerol, Sigma) was aliquoted into a DNA low bind tube and commercial buffer was exchanged for 1 mL cold base buffer A (25 mM HEPES pH 7.4, 140 mM KCl, 1.8 mM MgCl₂, 0.1% IPEGAL). Beads were resuspended by gentle inversion, collected again, then resuspended in 160 μ l cold supplemented buffer A (base buffer A supplemented with 0.5 mM DTT, 35 mM NaCl, 1x Complete EDTA-free protease inhibitor cocktail (Roche)). 100 μ l of supplemented cold buffer A and 200 units RNase OUT was added to 100 mg of grindate from the Pab1-3xFLAG, eIF4E-TAP strain. Grindate was vortexed and resuspended. Cell debris was pelleted by centrifugation at 12,000 $\times$ g for 5 min at 4 °C. Supernatant was transferred to a new tube and clarified again by centrifugation as described. The final supernatant was transferred to a new tube and absorbance at 260 nm was measured on a nanodrop spectrophotometer (Thermo Scientific). 40 μ l of lysate was added to each tube pre-prepared with equilibrated beads. 40 μ l of lysate from each strain was added to another tube containing 80 μ l of supplemented cold buffer A as the input sample. All input aliquots and bead-lysate reactions were rotated overnight at 4 °C. All tubes were placed immediately on ice. Input

aliquots were divided into two for protein (10%) and RNA (10%) analysis. Flow through was retained. The bead-lysate IP mixture was placed on a magnetic rack to collect beads. The reaction buffer was removed and beads were rinsed with 500 µl ice-cold wash buffer 1 (base buffer A with 2 mM DTT). Beads were again washed with wash buffer 1 by gentle rotation for 5 min at 4 °C then again with 1 mL of wash buffer 2 (base buffer A with 4 mM DTT). The final wash was taken off and beads were placed on ice. To elute from FLAG beads, 200 µl of 200 µg/mL 3xFLAG peptide working stock was added to the beads and rotated at 4 °C for 30 min. Elution was repeated and supernatants were combined. Beads were also retained for protein analysis. Protein analysis was performed as described in 2.2.1.3.

2.2.4.3 New beads – Mouse IgG2a conjugated M270 epoxy Dynabeads

Either 80 or 160 µg of Mouse IgG2a, Kappa Monoclonal antibody (ab18413, abcam) was conjugated to 16 mg of M-270 Epoxy Dynabeads (ThermoFisher) using the Dynabeads Antibody Coupling Kit (Thermofisher) according to manufacturer's instructions. ¼ of the final conjugated and equilibrated beads were added to clarified lysate equivalent to 40 mg grindate from the Pab1-pro(A), eIF4E-TAP or the WT untagged strain. All other experimental methods are as described in **Error! Reference source not found..**

2.2.5 Crosslinking optimization

The Pab1-pro(A), eIF4E-3xFLAG strain was grown as described in 2.2.7 and crosslinked with 37% formaldehyde (Sigma-Aldrich) to a final concentration of either 0%, 1%, 2% or 3% in parallel. Immunoprecipitation was performed as described in the section 2.2.4.3. Protein analysis by western blotting was performed as described in section 2.2.1.3. Polysome profiling of lysates with different crosslinking percentages was performed as described in section 2.2.1.5.

2.2.6 Rnase I digestion optimization

The Pab1-pro(A), eIF4E-3xFLAG strain was grown as described in 2.2.7 and crosslinked with 37% formaldehyde (Sigma-Aldrich) to a final concentration of 2.2%. For polysome profiling, one sample of clarified lysate was split into 6, with either 0, 1, 2, 3, 4 or 5 U/Au of

RNAse I (Ambion) added to each tube. Lysates were rotated for 30 min at RT, then 1x Suprase-In was added to stop reactions. Equal absorbance units of lysates were added to different ultracentrifugation sucrose gradients and method in section 2.2.1.5 was performed.

For RNA analysis, the clarified lysate was split between 4 tubes, and either 0, 1, 2 or 3 units RNAse I (Ambion) was added per tube. Immunoprecipitation was performed as described in section 2.2.4.3. RNA analysis by qPCR was performed as described in 2.2.1.4.

2.2.7 *Growing, crosslinking and grindate preparation*

S.cerevisiae was grown to mid-log phase (to an optical density at 600 nm of 0.7-0.8) under standard conditions (30 °C with shaking) in YEPD media (1% w/v yeast extract, 2% w/v peptone, 2% w/v glucose, 40mg/l adenine sulfate). Liquid cultures were flash-chilled by addition of 25% (w/v) crushed ice. 37% (w/v) formaldehyde (Sigma-Aldrich) was added to a final concentration of 2.2% (v/v). Cultures were transferred to polypropylene centrifuge tubes and allowed to cross-link for 10 min on ice. Cultures were transferred to a pre-cooled SLA-3000 rotor and pelleted at 6000 g for 5 min at 4 °C. Supernatant was discarded and pellet was resuspended in 40 mL HBB buffer (20 mM HEPES pH 7.4, 100 mM KCl, 2 mM MgCl₂) supplemented with 10% (v/v) 2.5 M glycine. Cells were pelleted at 4000 g for 8 min at 4 °C, then washed twice again with HBB buffer (-glycine) with pelleting after each wash as described. Pellets were weighed and 1 mL HBB-DRP buffer (HBB buffer supplemented with 1x Complete mini EDTA free protease inhibitor cocktail (Roche), 0.5 mM DTT, 200 U RnaseOUT per sample) was added to each pellet per 1 g weight. The resuspended pellet was snap frozen by dripping into liquid nitrogen and homogenised by two 1-minute rounds at 27 Hz using the MM 400 Homogeniser (Retsch). Steel canisters were kept cold before, between and after homogenisation by submerging in liquid nitrogen. The final powder was divided into ~100 mg aliquots and stored at -80 °C until required.

2.2.8 *Immunoprecipitation and RNA extraction*

160 µg of Mouse IgG2a, Kappa Monoclonal antibody (ab18413, abcam) was conjugated to 16 mg of M-270 Epoxy Dynabeads (ThermoFisher) using the Dynabeads Antibody Coupling Kit (Thermofisher) according to manufacturer's instructions. Conjugated beads were collected using a magnetic rack and storage buffer was replaced with 1 mL cold base buffer

A (25 mM HEPES pH 7.4, 140 mM KCl, 1.8 mM MgCl₂, 0.1% IPEGAL). Beads were resuspended by gentle inversion, collected again, then resuspended in 1 mL cold supplemented buffer A (base buffer A supplemented with 0.5 mM DTT, 35 mM NaCl, 1x Complete EDTA-free protease inhibitor cocktail (Roche)). Beads were rotated for 5 min at 4 °C, then split into 4x 1.5 mL protein low-bind Eppendorf tubes and kept on ice. 400 µl cold supplemented buffer A and 200 units RnaseOUT was added to each tube of 100 mg grindate and vortexed until thawed and resuspended, then for another 5 s. Cell debris was pelleted by centrifugation at 12,000×g for 5 min at 4 °C. Supernatant was transferred to a new tube and clarified again by centrifugation as described. The final supernatant was transferred to a new tube and absorbance at 260 nm was measured on a nanodrop spectrophotometer (Thermo Scientific). 1 U Rnase I (Ambion) was added per 1 absorbance unit ($A_{260} \times \text{volume in mL}$) of lysate. An aliquot equivalent to 20% of the immunoprecipitation (IP) reaction (80 µl) was retained as an input sample, and the remainder (400 µl) was divided between two tubes of pre-equilibrated and aspirated beads. Both input aliquots and bead-lysate reactions were rotated for 20 min at 22 °C. After 20 min of incubation, all tubes were placed immediately on ice. Input aliquots were divided into two for protein (10%) and RNA (10%) analysis. The aliquot for protein analysis was frozen, while the aliquot for RNA extraction was extracted either with method 1 (see section 2.2.8.1 below) or with method 2 (see section 2.2.8.2 below). Meanwhile, the bead-lysate IP mixture was placed on a magnetic rack to collect beads. The reaction buffer was removed and beads were rinsed with 1 mL ice-cold wash buffer 1 (base buffer A with 2 mM DTT). Beads were again washed with wash buffer 1 by gentle rotation for 5 min at 4 °C. Wash buffer 1 was collected, then the two identical bead-lysate reactions were combined into the one tube using 1 mL of wash buffer 2 (base buffer A with 4 mM DTT). The beads were split into two new tubes in an 80/20 split for RNA analysis and protein analysis respectively. The wash buffer 2 used for transferring beads was removed and replaced with fresh wash buffer 2, then beads were rotated for a final wash for 5 min at 4 °C. The final wash was taken off and beads were placed on ice. Beads set aside for protein analysis were frozen.

2.2.8.1 Decrosslinking and RNA extraction method 1 (initial strategy)

Input aliquots set aside for RNA analysis were made up to 400 µl with buffer SETG (1% SDS, 10 mM EDTA, 10 mM Tris-HCl pH 2, 10 mM glycine) and a further 400 µl of acidic 25:24:1 Phenol:Chloroform:Isoamyl Alcohol (pH 4-5) was added. Immunoprecipitation

eluate samples were treated similarly, with addition of 400 µl buffer SETG and 400 µl acidic 25:24:1 Phenol:Chloroform:Isoamyl Alcohol (pH 4-5) to the beads after the final wash step. Samples were de-crosslinked by incubation in a thermomixer at 65 °C for 25 min with gentle shaking. Once input and IP aliquots had been de-crosslinked, the mixture was centrifuged at 13,000×g for 10 min to separate the layers, and the aqueous layer was ethanol precipitated as described in section 2.2.1.1. Extracted RNA was DNase digested, reverse transcribed and analysed by qPCR as described in 2.2.1.4.

2.2.8.2 Decrosslinking and RNA extraction method 2 (updated for library preparation of final sequencing)

Input aliquots set aside for RNA analysis were made up to 400 µl with buffer SETG (1% SDS, 10 mM EDTA, 10 mM Tris-HCl pH 2, 10 mM glycine). Immunoprecipitation eluate samples were treated similarly, with addition of 400 µl buffer SETG to the beads after the final wash step. Samples were de-crosslinked by incubation in a thermomixer at 65 °C for 25 min with gentle shaking. Once input and IP aliquots had been de-crosslinked, RNA was extracted, precipitated and resuspended as described in section 2.2.1.1.

2.2.9 Initial sequencing

Library preparation and sequencing for the initial RNA-seq experiments were both performed by Azenta Life Sciences (formerly Genewiz). Libraries were prepared with the VAHTS Total RNA-seq (H/M/R) Library prep kit for Illumina® (with rRNA depletion) and barcoded to allow lane multiplexing. 150 nt, paired-end sequencing was performed on the HiSeq 2500 (Illumina). Samples were sequenced to a depth of at least 100M paired-end reads per individual sample.

2.2.10 Library preparation of final libraries

The library preparation approach was based off one previously established in the lab (Archer *et al.*, 2016), with further refinements by Janapala (2022). Broadly, the workflow consists of 3' polyadenylation of input or immunoprecipitated fragments, oligo-d(T) primed reverse transcription, cDNA circularisation and probe-directed degradation of ribosomal RNA

(rRNA). An expanded set of rRNA depletion probes were designed and detailed by Janapala (2022) and were subsequently used in this study.

2.2.10.1 End repair and polyadenylation

From input and immunoprecipitated eluates, 1.5 µg of each sample was transferred to new DNA low-bind tubes and adjusted to a volume of 20 µl in 1x HE buffer (10 mM HEPES pH 7.4, 0.1 mM EDTA). Samples were heated for 2 min at 70 °C and immediately cooled on ice for 5 min.

To prepare for polyadenylation, samples were first end-repaired by addition of 1 U/µl PNK enzyme (New England Biolabs, NEB) and 2 U/µl RnaseOUT (Invitrogen) in 1x PNK buffer (without ATP, NEB), followed by incubation for 2 h at 37 °C. 3' polyadenylation was carried out by addition of 0.1 U/µl EPAP (NEB) and incubation at 37 °C for 1 h, under the recommended conditions for use (1x EPAP buffer, 1 mM ATP, 1 mM DTT, 0.8 U/µl RnaseOUT). The polyadenylation reaction was stopped by addition of 15 µl STOP solution 1 (10% w/v SDS (Sigma), 100 mM EDTA (Sigma-Aldrich)) and 280 µl milliQ. RNA was extracted, precipitated and resuspended in 25 µl 1xHE (10 mM HEPES pH 7.4, 0.1 mM EDTA) as described in section 2.2.1.1.

2.2.10.2 Reverse transcription

To reverse transcribe the RNA, 2 pmol of the oligo-d(T) split-adapter primer (p155, 5'-phosphate-GA TCG TCG GAC TGT AGA ACT CTG AAC G /9-carbon spacer/ G TGA CTG GAG TTC CTT GGC ACC CGA GAA TTC CAT TTT TTT TTT TTT TTT TTT TVN-3') was added to each RNA sample, along with 0.5 mM (each) dNTPs mixture (NEB). Primer was annealed by incubation for 2 min at 65 °C, followed by a slow ramp (3 °C/s) to 4 °C and transferred to ice. While on ice, the reaction was supplemented with 5 mM DTT, 2 U/µl RNaseIN Plus (Promega) and 10 U/µl Superscript IV Reverse Transcriptase (Invitrogen by Thermo Fisher Scientific). The reaction mixture was then incubated for 10 min at 30 °C, followed by 10 min at 35 °C and 10 min at 40 °C. cDNA was immediately purified by addition of 1.2x volume Ampure XP beads and incubation for 5 min at RT. Beads were washed twice with 500 µl of fresh 80% (v/v) ethanol, then dried at RT for 10 min. cDNA was

eluted off the beads by addition of 20 µl milliQ and incubation for 10 min at RT, with gentle vortexing for 2-3 seconds at 1 minute intervals. 9 µl was used for the following circularisation step and the rest was stored at -20 °C as backup.

2.2.10.3 Circularisation

In preparation for cDNA circularisation, 9 µl of each sample was heated at 85 °C for 3 min, then snap-chilled on ice for 5 min. Samples were incubated with 1 µl of RNase A/T1 mix (2 mg/mL RNase A, 5000 U/mL RNase T1 (Merck)) for 30 min at 37 °C. To prevent circularisation of empty, unextended primers, 80 pmol of empty split adapter circularisation blocker (p156, 5'-TTN BAA AAA AAA AAA AAA AAA AAA /iSuper-dT/GG AAT TCT CGG GTG CGT GTG T/3BioTEG/-3') was added to each sample along with 10x annealing buffer (250 mM HEPES-KOH pH 7.4 (Merck), 1 M NaCl (Merck), 50 mM MgCl₂ (Sigma)) to a final concentration of 1x. p156 was annealed to empty split adapter primers by incubation for 2 min at 80 °C, ramping up to 95 °C for 1 min, cooling down to 50 °C for 1 min and final slow ramp (3 °C/min) to 40 °C. Samples were kept in the PCR machine while immediately proceeding to the circularisation reaction.

The RNA-depleted cDNA was circularised for 3 h at 40 °C by addition of 100 U CircLigase II ssDNA ligase (Lucigen) according to manufacturer's recommendations (supplemented with 1x CircLigase II buffer (Lucigen), 2.5 mM MnCl₂, 100 mM Betaine). Reaction was stopped by addition of 10 µl 100% STOP solution I. Circularised cDNA was extracted by addition of 1x volume 25:24:1 Phenol:Chloroform:Isoamyl Alcohol (pH 7-8) (Sigma-Aldrich), vortexing at maximum speed for 2 min, followed by centrifugation at 12,000×g for 10 min at RT. The aqueous phase (~50 µl) was added to Microspin G-25 columns (Amersham/Cytiva) pre-equilibrated with 1xHE and centrifuged at 700×g for 2 min for further purification. Two empty reactions were performed in parallel as a reference for the size of empty library fragments (for use during gel extraction). Empty 1 contained 9 µl 1x HE, 1 µl 10 µM split adapter (p155), 1 µl 10 µM split adapter circularisation blocker (p156) and 2 µl 10x annealing buffer. Empty 2 contained the same components as empty 1, except with substitution of the split adapter circularisation blocker (p156) for 1 µl 1x HE. Empty reactions were treated identically to the samples up to and including the purification step just described.

In order to remove the empty split adapter circularisation blocker (p156), 23 µl of the gel-filtered material was mixed with 3 µl of 10x annealing buffer and 40 pmol of the circularisation block removal oligonucleotide (p157, 5'-ACA CAC GCA CCC GAG AAT TCC ATT TTT TTT TTT TTT TTT TTT T/iSuper-dT/T VNAA/3BioTEG/-3'). Reaction was incubated for 2 min at 80 °C, 95 °C for 1 min, 80 °C for 10s, followed by slow ramping (3 °C/s) to 50 °C, then a further incubation for 1 min at 50 °C and cooled to 40 °C. The reaction was transferred to a new DNA low-bind tube (Eppendorf) containing streptavidin beads (NEB) pre-equilibrated with 1x annealing buffer and equivalent to 320 µl of the original bead suspension. The mixture was incubated with the beads for 10 min at RT, with gentle flicking of the tube at 1-minute interval to keep beads in suspension. The mixture was incubated for 2 min at 35 °C and immediately separated on a magnetic rack. The supernatant containing unbound, circularised cDNA was collected.

2.2.10.4 Probe-Directed Degradation (PDD) of rRNA sequences

To deplete samples of cDNA corresponding to abundant rRNA sequences, probe-directed degradation was performed using the depletion probes listed in Table 8. Depletion probes were first mixed to a final concentration of 1.5 µM each, then 1 µl of this mixture was added to 15 µl of each sample of circularised cDNA. The probes were annealed to their targets by heating to 95 °C followed by slow ramping (3 °C/s) to 75 °C. While keeping samples in the machine, 2x DSN buffer (Evrogen) was added to a final concentration of 1x, followed by slow ramping (3 °C/s) to 60 °C, 1 min hold at 60 °C and finally addition of 0.5 U Duplex-specific Nuclease (DSN, Evrogen). The temperature was slow ramped (3 °C/s) to 48 °C, then incubated for 20 min. Probe-directed degradation was stopped by addition of 57 µl STOP solution 2 (20 mM EDTA (Sigma-Aldrich) in 1x HE (10 mM HEPES pH 7.4, 0.1 mM EDTA). Non-degraded cDNA was extracted by addition of an equal volume (~70 µl) 25:24:1 Phenol:Chloroform:Isoamyl Alcohol (pH 7-8) (Sigma-Aldrich), followed by 2 min vortex at maximum speed and centrifugation at 12,000×g for 10 min at RT. The aqueous layer was again gel-purified with pre-equilibrated Microspin G-25 columns (Amersham/Cytiva) as previously described.

2.2.10.5 Library amplification

Following probe-directed degradation of rRNA sequences, SuperFi Platinum DNA polymerase (Thermofisher scientific) and custom primers compatible with Illumina TruSeq Small RNA sequencing (see Table 7) were used to amplify the resultant cDNA for sequencing. Forward primers had the sequence 5'-CAA GCA GAA GAC GGC ATA CGA GAT XXX XXX GTG ACT GGA GTT CCT TGG CAC CCG AGA ATT CCA-3', where XXX XXX is replaced with the Illumina hexa-nucleotide barcoding sequence unique to each library. The reverse primer was common for all libraries with the sequence 5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACG TTC AGA GTT CTA CAG TCC GA-3'. Each library amplification reaction was set up with 10 µl of the purified, rRNA depleted cDNA and a final concentration of 0.2 mM dNTP mix (NEB), 1x Platinum SuperFi buffer (Thermo Fisher Scientific), 0.5 µM forward primer, 0.5 µM reverse primer, 4.2 ng/µl extreme thermostable single-stranded binding protein (ET SSB, NEB), and 0.02 U/µl Platinum SuperFi DNA polymerase (Thermo Fisher Scientific). The reaction mix was multiplied as needed to amplify all of the cDNA obtained from the previous step. Reactions were amplified by a thermal program of 98 °C for 5 min, 3x cycles of [98 °C for 30 s, 62 °C for 30 s, 72 °C for 45 s], then 22x cycles of [98 °C for 30 s, 76 °C for 30 s, 72 °C for 45 s], followed by a final extension of 72 °C for 1 min. The empty reactions from the circularisation reaction were also set up and amplified as above.

2.2.10.6 Size selection and purification of libraries

Libraries were pre-stained with 6x GRGreen loading buffer (Excellgen) to a final GRgreen concentration of 1x, then size-separated on a 2% native agarose gel as described in section 2.2.1.2, with the only modification that voltage was never set above 80V to minimize DNA degradation during separation. To allow for accurate size quantification, samples were run alongside 5 µl of the 50 bp Generuler DNA ladder (Merck), diluted 3/100 by addition of 3 µl ladder to 69.4 µl milliQ, 2.4 µl 10 mM dNTP (NEB), 24 µl 5 x Platinum SuperFi buffer (Thermo Fisher Scientific), 1.2 µl ET SSB (NEB) and 20 µl 6x GRGreen). 1/10 diluted empty 1 & empty 2 were also run alongside samples to determine the size of the empty, insert-less library fragments. Library fragments between size 50-700 nt were excised from the gel with a sterile scalpel and transferred to Freeze 'N Squeeze DNA gel extraction tubes (BioRad), then incubated at -20 °C for 30 min. To elute the library fragments, tubes were centrifuged at RT for 3 min at 13,000×g. The eluted DNA was extracted with 25:24:1

Phenol:Chloroform:Isoamyl Alcohol (Sigma Aldrich) (pH 7-8), precipitated with 100% Ethanol (Sigma-Aldrich) and resuspended in 1x HE as described in section 2.2.1.1. The size distribution and concentration of the final libraries were checked for consistency and quality using the TapeStation High Sensitivity D1000 ScreenTape (Agilent) and directed for sequencing.

2.2.11 Sequencing

150 base-pair, paired-end sequencing was performed using the Illumina NovaSeq6000 short-read sequencing system with hexanucleotide demultiplexing barcodes and sequencing primers designed to be compatible with Illumina Tru-Seq sequencing. Libraries were sequenced to a minimum requirement of 36G, or approximately 120 million paired-end reads each.

2.2.12 Initial bioinformatic processing of sequencing files

Strategy was devised in conjunction with Dr Attila Horvath. However, data-processing and coding of scripts was performed by Dr Attila Horvath.

150 bp paired-end sequencing libraries were received in fastq format. The cutadapt poly(A) trimming function was used to identify continuous poly(A) tracts of 20 (A)s artificially introduced during library preparation, so that these tracts and any immediately downstream sequences are removed taking into account bp. Poly(A)-removed sequences were subject to quality trimming using Trimmomatic (SLIDINGWINDOW of 7 nt, phred quality cut off of 24, minimum fragment length of 14 nt). Reads mapping to rRNA (sacCer3 rRNA) were discarded, followed by reads mapping to tRNA (GtRNAdb) and finally structural RNA (sacCer3). Remaining reads were mapped to the budding yeast genome using STAR (outFilterMultimapNmax 20000). The final BAM alignment file was split into strand-specific Forward (F) and Reverse (R) bigwig coverage files. Library size normalisation was performed by addition of both F and R bigwigs to determine total coverage per sample, followed by division of coverage at each location in F and R files by the total coverage, then finally multiplication of coverage signal by 10^8 for visualisation.

2.2.13 *De novo transcription unit predictions*

Yeast CDS annotations were downloaded from yeastgenome.org. For each sample, average, length-normalised signal across each annotated CDS was calculated. The genes in the bottom 5% by expression were discarded.

From final IP coverage files, continuous stretches of IP signal with minimum length of 20 nt and average signal value of at least 5 (with a gap of 1 nt allowed) were designated a ‘signal island’. Signal islands with coordinates overlapping at least partially with annotated CDS coordinates were assigned to that CDS. Signal islands having coordinates within at least 300 nt of an annotated CDS (measured by the distance between the signal island and CDS’s closest edges to each other) were also assigned to that CDS. The outermost coordinates of all assigned IP signal islands were merged with the coordinates of its corresponding CDS annotations and input signal islands to extend the CDS annotation. All actions were performed in python and R using in-house scripts.

2.2.14 *Calculation of end and middle signals*

From the newly defined *de novo* transcript unit (Tx) predictions, regions with a distance equal to 30% of the length of that annotation were identified, starting from the outermost 5’ and 3’ coordinates of that prediction. Within each region, the highest window of peak signal, measuring 50% of the radius length, was determined. The average coverage over this window was calculated and taken as the start or end signal for that gene. A 50 bp region from the direct middle of the transcript unit annotation was taken as the middle region, and signal from this region was average over 50 bp to determine final middle signal value.

2.2.15 *Filtering of individual mRNAs*

With start, end and middle signals calculated, individual mRNAs were filtered for certain criteria. Low-expressing mRNAs where signal could not be detected in at least one peak – either start, middle or end – were discarded. Genes where the average input signal was below 2 were also discarded. Genes with introns were placed aside for separate analysis. Genes with ‘dubious’ or ‘retrotransposon’ tags were discarded. ‘Flat’ genes were also removed, with

genes requiring an enrichment of at least 2-fold of start signal/middle signal or end signal/middle signal to be included for further analysis.

2.2.16 Calculation of closed-loop score

Start and end enrichments over middle were calculated by taking the $\log_2(\text{average start signal/average middle signal})$ and $\log_2(\text{average end signal/average middle signal})$ respectively. The ‘closed-loop’ measure was determined by subtraction of the $\log_2$ start and end enrichments values, leaving $\log_2(\text{start/end})$. For correlation analyses, the closed-loop score was transformed to $1/|\text{closed-loop measure}|$, referred to from here as ‘transformed close-loop score’.

2.2.17 Clustering

Genes were sorted into three categories, with ‘closed-loop’ mRNAs defined as those meeting the criteria of $-1 < \log_2(\text{start/end}) < 1$. ‘High 5’ UTR’ group was defined as $\log_2(\text{start/end}) > 1$. ‘High 3’ UTR group’ was defined as $\log_2(\text{start/end}) < -1$. K-mean Clustering was performed on mRNAs in R and visualised using the pheatmap package.

2.2.18 Correlation analyses

2.2.18.1 Tx length, CDS length, UTR lengths

Full transcript, 5’ UTR and 3’ UTR lengths as defined *de novo* previously were compared to the transformed closed-loop score using our in-house pipeline. CDS lengths were collected from yeastgenome.org.

2.2.18.2 Association with closed-loop factors

Association data of individual mRNAs with eIF4E, eIF4G1 and Pab1 were taken from Costello *et al.* (2015). These values were compared to the transformed closed-loop score.

2.2.18.3 *Gene Ontology analysis*

The Gene Ontology term finder function (GO term finder, version 0.86) on Saccharomyces Genome database was used. The ‘u-shaped’ category of gene names was probed for enriched molecular function, biological process or cellular component categories, with the full list of ‘u-shaped’, ‘high 5’ UTR’ and ‘high 3’ UTR’ genes as the reference set.

2.2.18.4 *Translation efficiency*

‘Protein synthesis rates per mRNA’ measure by SILAC were obtained from Riba *et al.* (2019). Genes were binned by closed-loop score (un-transformed) in 0.5 increments and compared to protein synthesis rates using our in-house pipeline.

2.2.18.5 *mRNA half life*

mRNA half-life measurements in minutes were taken from Geisberg *et al* (2014). Genes were binned by closed-loop score (un-transformed) in 0.5 increments and compared to mRNA half-life data using our in-house pipeline.

Chapter 3 – Development of an experimental method to capture the closed- loop *in vivo*

3 Development of an experimental method to capture the closed-loop in vivo

In order to study formation of the closed-loop transcriptome-wide, the first step was to develop a method to capture the closed-loop and associated mRNAs in a way which would allow comparison between mRNAs as to the extent of their closed-loop association. To ensure probing of the closed-loop in conditions as representative as possible of normal cellular conditions, an *in vivo* approach was also preferred.

With these conditions and objectives in mind, our research group had developed a method to immunoprecipitate closed-loop complexes and use the relative enrichment of 5' and 3' mRNA sequences as a means of comparing closed-loop formation (Archer *et al.*, 2015) (Figure 3.1). This method consisted of formaldehyde-fixing actively growing yeast cells, then immunoprecipitating tagged closed-loop factors concurrently with mild RNase I digestion. By stringently controlling RNase I digestion, the mRNA was digested randomly, with the digested portions not retained. Left with the immunoprecipitated complex and remaining associated mRNAs, qPCR probes were used to compare the mRNA enrichment at the 5' and 3' ends for pulldowns via eIF4E or Pab1. As an example, it was expected that many 5' fragments would be detected in the eIF4E-tagged immunoprecipitation, given that eIF4E contacts the cap. However, given the RNase I digestion to remove middle portions of the mRNA, mRNA enrichment detected at the 3' end would signal an intact bridge between the ends, presumably formed by the closed-loop factors. By comparing the extent of 3' end enrichment relative to 5' enrichment for a panel of mRNAs, differences were observed between mRNAs in the extent of their relative end enrichments, suggesting a differential formation of the closed-loop.

Though informative, this study was limited by the size of the mRNA panel, consisting only of approximately 20 mRNAs, so it was difficult to draw transcriptome-wide observations and conclusions. Therefore, the first aim of this thesis was to adapt the previously-established method for use with transcriptome-wide RNA-sequencing as a readout (Figure 3.1). Although conceptually straightforward, this was a challenge technically. Various experimental conditions needed to be tightly controlled and optimised in order to faithfully represent the state of the closed-loop in living cells, and elements of the experimental workflow required

careful examination before proceeding with RNA-seq. Availability of reagents and instruments had also changed in the period between the previous study and the start of this PhD project, so these also needed re-optimising. Briefly, we needed to achieve a few main objectives before proceeding with RNA sequencing –

- Ensure a method of detecting whether the closed-loop proteins are immunoprecipitated together
- Optimise immunoprecipitation conditions for maximal recovery at the mRNA ends while balancing specificity and yield
- Recreate findings from the original study using qPCR, both in terms of co-immunoprecipitation of the closed-loop components, and relative end enrichments for certain mRNAs

In this chapter I present the development of a technique to achieve the above objectives.

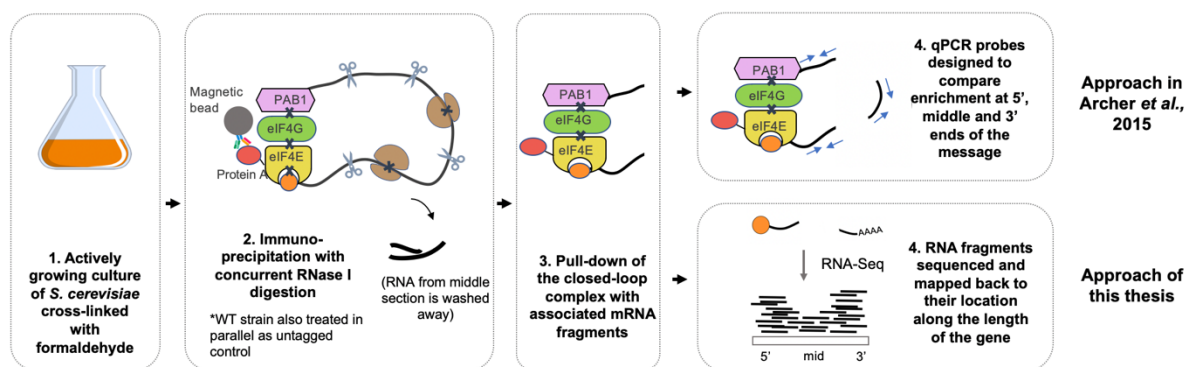


Figure 3.1. Overview of experimental method

Differences between the original study (Archer et al., 2015) and current approach are highlighted.

3.1 Ensuring visualisation of the closed-loop proteins

As a first step before development of an immunoprecipitation method to capture the closed-loop components and associated mRNAs, I first needed to ensure that the closed-loop components could be detected simultaneously. This was important for later validating that the closed-loop components co-immunoprecipitate together under our chosen immunoprecipitation conditions, ensuring that successful capture of the closed-loop complex has been achieved. For example, if the closed-loop complex was pulled down through eIF4E, then detection of eIF4G1 and Pab1 in the eluate would be used as indication that the

crosslinking and immunoprecipitation conditions were adequate to capture the whole closed-loop.

3.1.1 Lack of antibodies against the closed-loop proteins in yeast

The original study detected the primary immunoprecipitated protein (the protein immunoprecipitated via the magnetic bead – either eIF4E, eIF4G1 or Pab1) by detection of its protein(A) tag used for the pulldown. The other two components were detected by antibodies to the endogenous protein. However, at the time of the present study, the endogenous antibodies previously used were no longer in production and could not be obtained. Therefore, I aimed to find new antibodies which could detect the proteins in yeast. I also aimed to find an antibody unrelated to the closed-loop for use as a negative control. No closed-loop antibodies from the major antibody retailers were listed specifically for use in yeast, but some were listed for use across multiple species, although not specifically including yeast. I therefore tested some of these multi-species antibodies for their ability to detect the yeast closed-loop components and negative controls in yeast, in case yeast had not been tested as a species.

Antibodies were ordered and tested against different amounts of clarified whole-cell lysate (WCL) from an untagged WT strain, according to the manufacturer's recommendations. No bands were detected with the antibodies tested for Pab1 (expected molecular weight (MW) = 64.3 KDa) or eIF4E (expected MW = 24.2 KDa), apart from non-specific bands of the wrong size for eIF4E (Figure 3.2a,b). Decrease in antibody dilution or optimisation of dilution medium was still unable to yield detectable bands, suggesting these antibodies do not detect yeast Pab1 and eIF4E. A faint signal of the correct size was obtained for eIF4G1 (expected MW = 107.1 KDa) (Figure 3.2c). However, this signal was lost upon stripping and re-probing, suggesting a very faint detection ability. Both beta-actin and GAPDH antibodies were able to detect bands of the correct sizes (expected beta-actin MW = 41.7 KDa, expected GAPDH MW = 36 KDa) (Figure 3.2d,e). I decided to continue with the GAPDH antibody, as this gave a slightly cleaner signal with no non-specific bands, as opposed to beta-actin which resulted in a non-specific band just above 25 KDa. However, in the case of the closed-loop proteins, it was determined that an alternative approach would be required to detect these proteins.

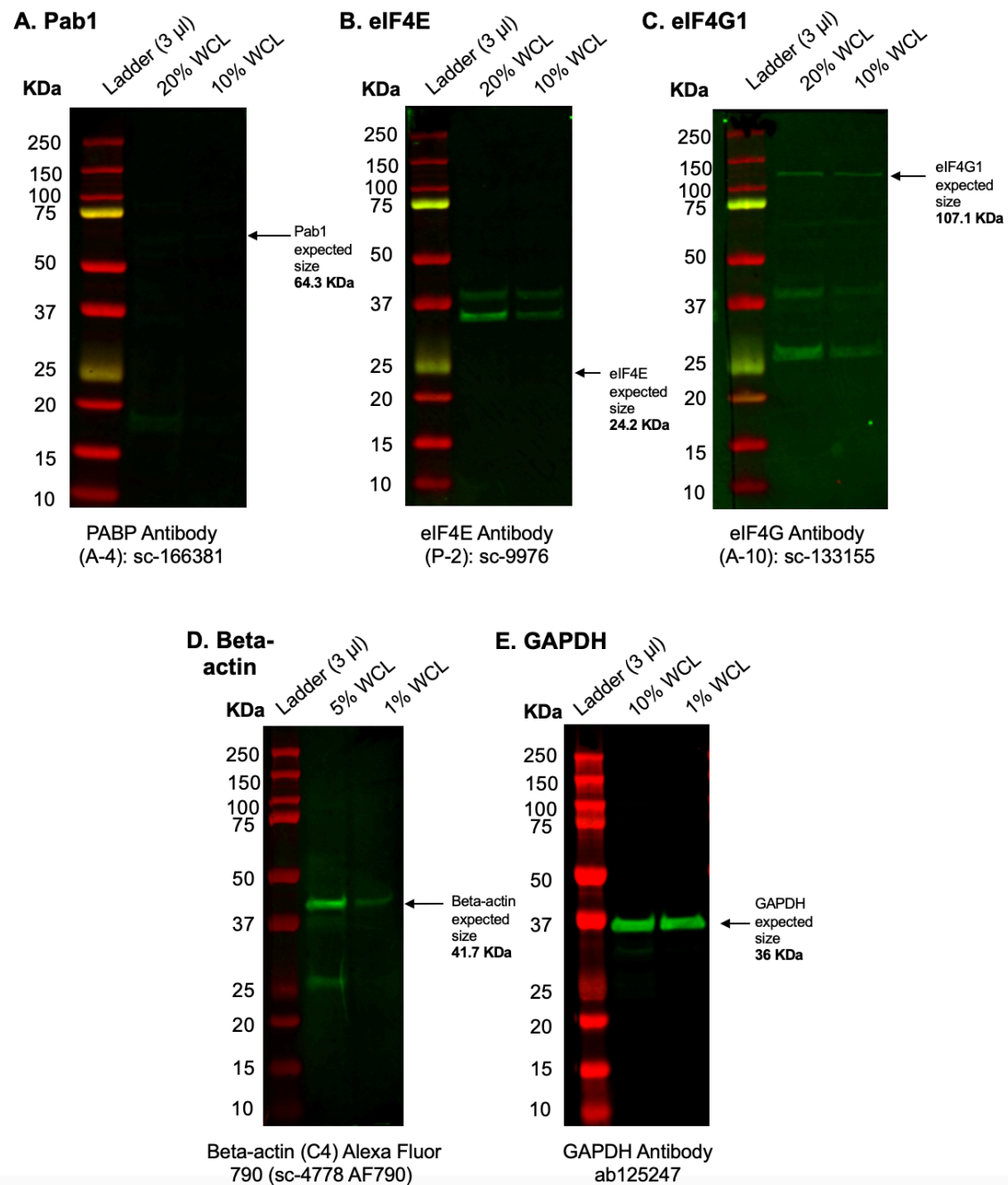


Figure 3.2. Trial of antibodies against endogenous yeast proteins

Western blots of SDS-page gels where 20% and 10% whole cell lysates (WCL) were probed with primary antibodies (diluted 1:100) for the closed loop proteins – Pab1, eIF4E and eIF4G1. Li-cor secondary antibodies (diluted 1:20,000) were used to detect the proteins using the Odyssey CLx Imaging System. A 1:500 dilution of primary antibody against beta-actin conjugated with Alexa Fluor 790 was also tested against 5% and 1% WCL for future use as an unrelated control. Similarly, a 1:1000 dilution of primary antibody against GAPDH was tested against 10% and 1% WCL, followed by a 1:20,000 dilution of Li-cor secondary antibody, for use as an additional negative control.

3.1.2 Tagging of closed-loop components with 3xFLAG

As endogenous antibodies against the yeast closed-loop proteins did not result in successful or consistent detection of said proteins, it was decided a new strategy would be required. I chose to C-terminal tag the closed-loop proteins with the 3xFLAG tag, consisting of three slightly varying repeats of the FLAG motif. This 3xFLAG tag is a widely used protein tag easily detectable by anti-FLAG antibodies, and is often used to enhance the detection of lowly-expressed proteins (Hernan, Heuermann and Brizzard, 2000). Therefore, it was a promising candidate for tagging of the closed-loop proteins in a manner comparable with the protein(A) tags already present in the yeast strains used. The aim was to generate 6 tagged strains covering all combinations of protein(A)-tagged and 3xFLAG-tagged closed-loop proteins, with each strain containing one protein(A)-tagged and one 3xFLAG-tagged closed-loop factor (Figure 3.3).

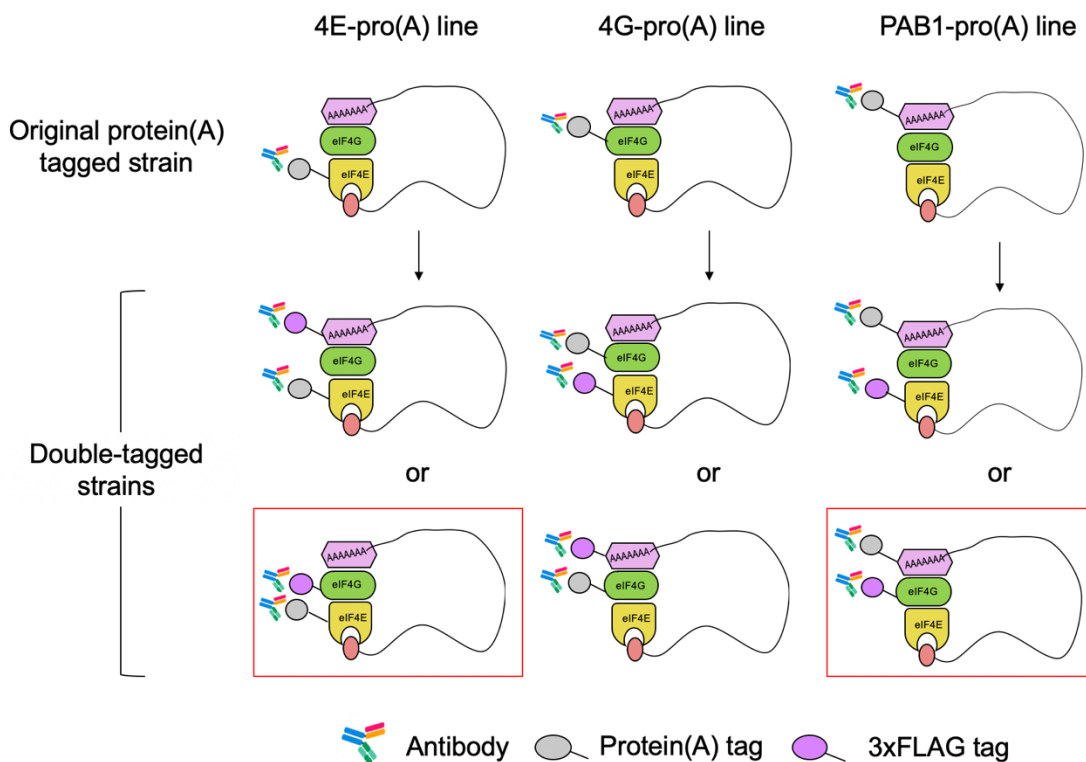


Figure 3.3. Summary of double-tagged strains generated

Strains already bearing a protein(A) tag on one closed-loop protein were further tagged with a 3xFLAG tag on one other closed-loop protein, to generate 6 double-tagged strains. The two strains used most commonly in this thesis are circled in red.

3.1.2.1 Overall strategy

The strategy for tagging each strain with a 3xFLAG tag on either eIF4E, eIF4G1 or Pab1 was similar. Yeast is able to recombine a linear sequence of DNA into its genome by homologous recombination, provided the linear DNA contains an overlap to the genome at both ends with overlap lengths of at least 30 bp (although over 60bp is preferred) (Orr-Weaver, Szostak and Rothstein, 1981; Hua *et al.*, 1997). The gene for antibiotic resistance to nourseothricin, *nat1* from *Streptomyces noursei*, can also be introduced into yeast to confer resistance to nourseothricin (CloNAT) in YPD growth media. In the context of selection after genetic manipulation, antibiotic selection has an advantage over auxotrophic mutations in its flexibility and reduction in deleterious effects (Goldstein and McCusker, 1999). Taking into account these two features of yeast, PCR was used to amplify cassettes containing the 3xFLAG epitope followed by a CloNAT-resistance marker, flanked by two 100 bp regions homologous to region directly up- and downstream of the planned insertion site at the C-terminal of either eIF4E, eIF4G1 or Pab1. These cassettes were transformed into their corresponding strains already bearing a protein(A) tag on another closed-loop protein. Colonies growing on antibiotic-enriched YPD plates were selected and verified by colony PCR for the correct insertion, then sequenced by Sanger sequencing to ensure there were no mutations arising as a result of the insertion (Figure 3.4).

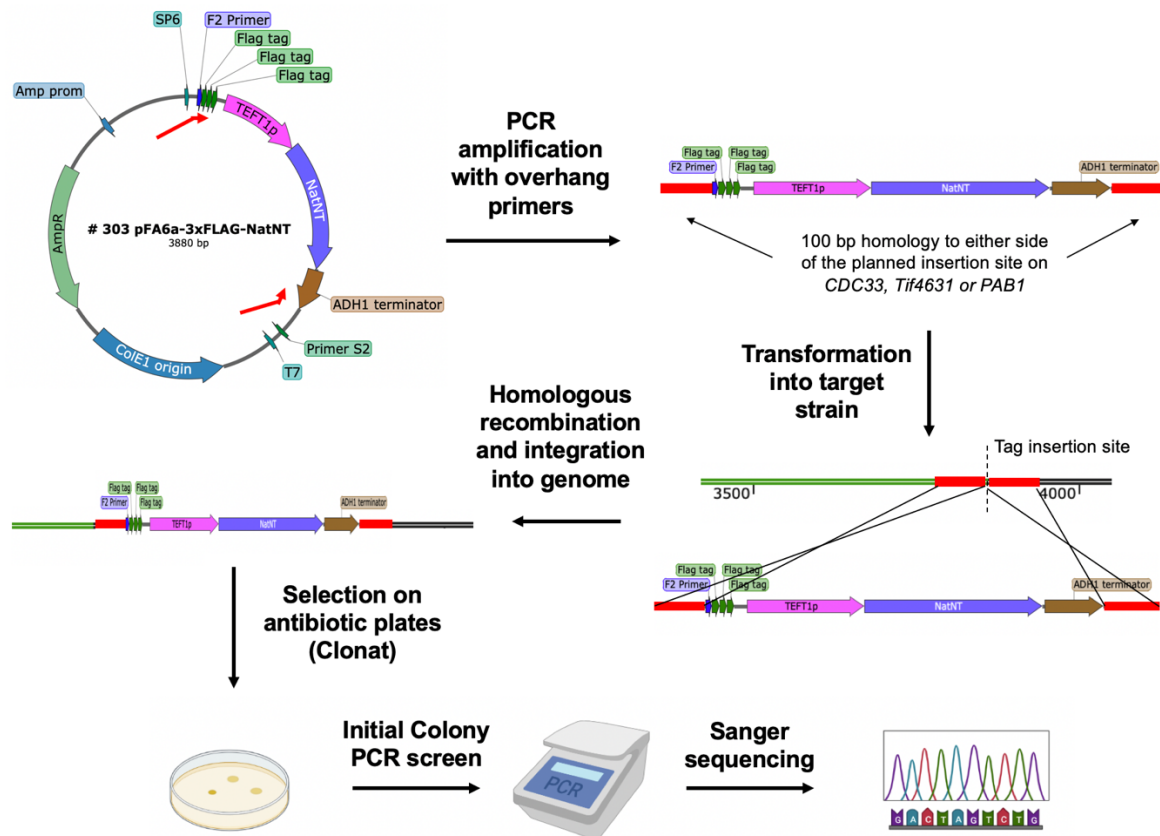


Figure 3.4. Schematic of overall tagging strategy

The 3xFLAG epitope followed by an antibiotic-resistance cassette (CloNAT, or nourseothricin) was amplified from plasmid #303 pFA6a-3xFLAG-NatNT using primers containing 100bp overhangs. Overhangs on forward and reverse primers corresponded to 100bp sequences of homology to region directly upstream or downstream of the planned insertion site respectively. Amplified 3xflag and antibiotic selection cassettes were purified and transformed into the desired target strain, allowing for homologous recombination-mediated insertion into the yeast genome. Colonies were selected for successful recombination by growth on CloNAT plates, with viable colonies further tested by colony PCR for successful recombination at the correct location. Positive colonies were verified by sanger sequencing and western blotting to ensure correct insertion of the 3xFLAG tag.

3.1.2.2 Making of cassettes for transformation

To make the 3xFLAG cassettes for incorporation into the yeast genome, classic PCR techniques were used to amplify the approximately 1.5 kb fragment from a plasmid using overhang primers homologous to the planned insertion site. To decrease occurrences of sequence mutations introduced by PCR, a high-fidelity polymerase, Phusion, was used instead of the more common Taq polymerase. DMSO was added to inhibit secondary structures and increase efficiency of amplification. Finally, the time and temperature of each of the PCR steps was optimised sequentially for maximum yield. After PCR, the methylated plasmid template was digested with *DpnI* while the amplified cassette remained undigested due to lack of methylation. This was to decrease false-positive transformants that take up the

plasmid without genomic recombination. Running of the final three cassettes on an agarose gel resulted in visualisation of three distinct bands of the correct size just above the 1.5 kb ladder band (Figure 3.5). No noticeable primer dimers were detected, as shown by lack of other bands in both the cassette reaction and the no template control lanes. Lack of a faint band above the ladder bands also suggests no significant amounts of leftover plasmid backbone were present. Therefore, it was concluded that the PCR reactions resulted in three cassettes of the correct size and composition for transformation.

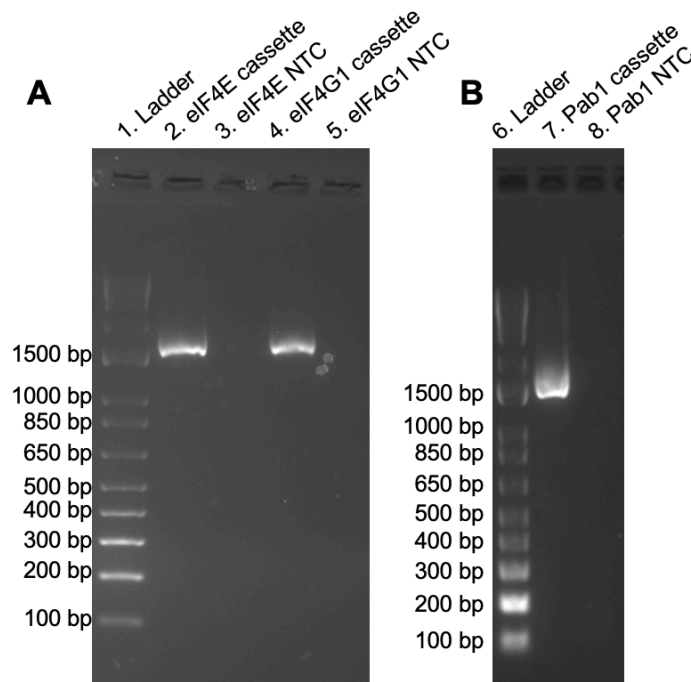


Figure 3.5. Generation of 3xFLAG cassettes for transformation

Cassettes bearing the 3xFLAG epitope and a CloNAT selection marker were amplified from plasmid pFA6a-3xFLAG-NatNT using primers with 100bp overhangs to the planned insertion site in either A) CDC33 (eIF4E, lane 2), Tif4631 (eIF4G1, lane 4) or B) PAB1 (Pab1, lane 7). No template controls (NTC) (lanes 3, 5, 8) with all components apart from the plasmid template were also done in parallel to test for non-specific amplification and primer dimers. Agarose gel electrophoresis was performed on PCR reactions to detect presence of a successful amplification product,

3.1.2.3 Transformation, colony PCR and sanger sequencing

Transformation of the cassettes into their target strains was performed as described in methods section 2.2.2.2. A negative reaction with no cassette was always performed in parallel. Both reactions, with and without the cassette, were plated onto YPD plates with and without CloNAT in the medium. After three days of growth, control plates were checked for appropriate growth. The negative reaction was assessed for growth on the antibiotic-free

plate, suggesting cells did not die during the transformation process. Lack of growth on the antibiotic plate signaled adequate levels of antibiotics were used. If either condition was not met, the transformation was repeated. Positive colonies from the cassette-containing reaction on antibiotic plates were picked and further tested by colony PCR.

To check for correctly-placed insertion of the cassettes into positive colonies, colony PCR followed by agarose gel electrophoresis was performed on colonies after zymolyase treatment to degrade cell walls. Primers were designed to span both homologous recombination junctions for each transformation, to ensure the fragment had recombined in the correct location and not somewhere else in the genome. One example colony PCR is shown below in Figure 3.6. One primer is placed in the coding region of eIF4G1 upstream of the correct recombination site. Another primer is placed inside the cassette sequence, at the end of the 3xFLAG sequence (Figure 3.6a). Presence of a band suggests correct insertion at this location. A single band can be seen in lanes 2-9 of Figure 3.6b, corresponding to colonies 1-8 obtained after the Pab1-pro(A) strain is transformed with the cassette bearing a 3xFLAG for insertion at the C-terminus of eIF4G1. As expected, the parental strain in lane 10, which had undergone no transformation, did not show the same band, suggesting this is not a non-specific amplification. Similarly, the NTC failed to yield a band, suggesting the band is specific. Similar results were obtained with an eIF4E-pro(A) strain, with the difference that not all colonies had successful insertions at the location tested (Figure 3.6b, lanes 12-20). Similar results for the parental eIF4E-pro(A) strain and NTC were obtained as previously discussed. From these colony PCR reactions, and others not shown, positive colonies were picked for further validation.

The colony PCR primers were re-used to generate PCR products for sanger sequencing. Obtained sequences and the expected sequences were compared in Clustal Omega to check for correct insertion and lack of frameshift or non-silent mutations.

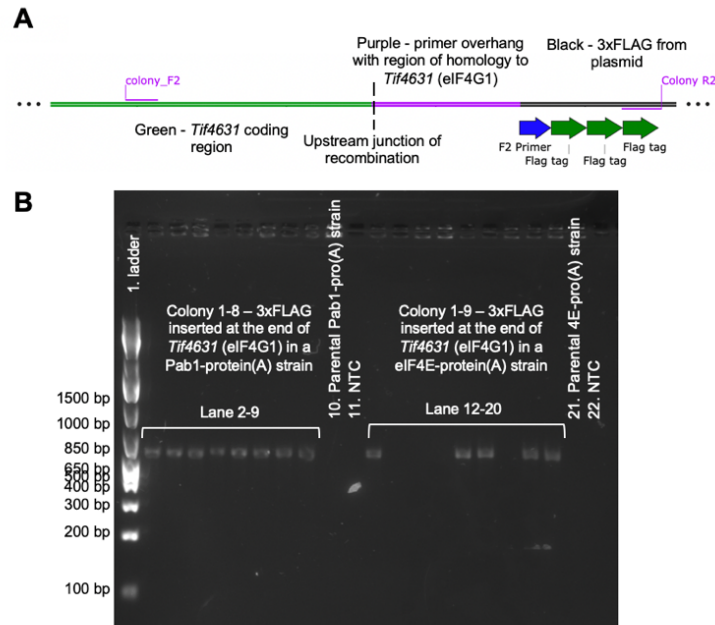


Figure 3.6. Example colony PCR to check for correct insertion of 3xFLAG tag

A) Schematic to depict location of colony PCR primers – the forward primer (colony_F2) was placed upstream of the recombination junction, in this case in the coding region of Tif4631 (eIF4G1). The reverse primer was placed within the 3xFLAG cassette prepared for insertion. B) All colonies picked from transformation of the Pab1-pro(A) strain with eIF4G1-3xFLAG were positive for the insertion at the site tested (lanes 2-9). 5 of 9 colonies picked from transformation of the eIF4E-pro(A) strain with eIF4G1-3xFLAG were positive (lanes 12, 16, 17, 19, 20). Parental strains and NTCs for both strains did not show bands (lanes 10, 11, 21, 22).

3.1.3 Confirmation of 3xFLAG tag incorporation

As another confirmation of correct tag insertion, western blotting was performed on whole cell lysates of the final colonies selected after sanger sequencing. The strains subject to the 3xFLAG tag incorporation each originally had one closed-loop factor tagged with protein(A), or in the case of eIF4E, the TAP tag inclusive of protein(A). These strains showed bands corresponding to the molecular weight of the tagged protein plus their respective protein(A) tag when probed with an anti-protein(A) antibody (Figure 3.7a,b, lanes 2 and 4), or no bands when probed with only the anti-flag antibody (Figure 3.7c, lanes 2 and 4). This remained unchanged after 3xFLAG-tagging (Figure 3.7a,b,c, lanes 3 and 5).

The 3xFLAG tag has a molecular weight of approximately 2.7 kDa, so on a western blot this should not result in a visibly large increase in molecular weight. As eIF4E, eIF4G1 and Pab1 have molecular weights of 24 kDa, 107 kDa and 64 kDa respectively, the size of the proteins after 3xFLAG addition should be 27 kDa, 110 kDa and 67 kDa respectively. From visual

inspection of bands corresponding to the 3xFLAG epitope (green), all proteins were of expected sizes after 3xFLAG tagging (Figure 3.7a,b,c, lanes 3 and 5). The size of all proteins in Figure 3.7b were slightly larger than expected, but as this effect was seen across all proteins and a different ladder was used, this effect can be explained by inconsistencies between the running of the ladder and proteins. It was therefore concluded that for all 6 double-tagged strains, the 3xFLAG tag insertions had resulted in accurately sized proteins detectable via western blotting, ready for use in downstream steps of the immunoprecipitation experiment. From these 6 double-tagged strains, two strains – eIF4E-TAP, eIF4G1-3xFLAG and Pab1-pro(A), eIF4G1-3xFLAG were chosen as the focus strains for optimisation of immunoprecipitation. These were chosen as they were the two strains that together could allow for immunoprecipitation from both the 5' end and the 3' end, while allowing for detection of a common middle factor – eIF4G1. The other strains were stored for future use.

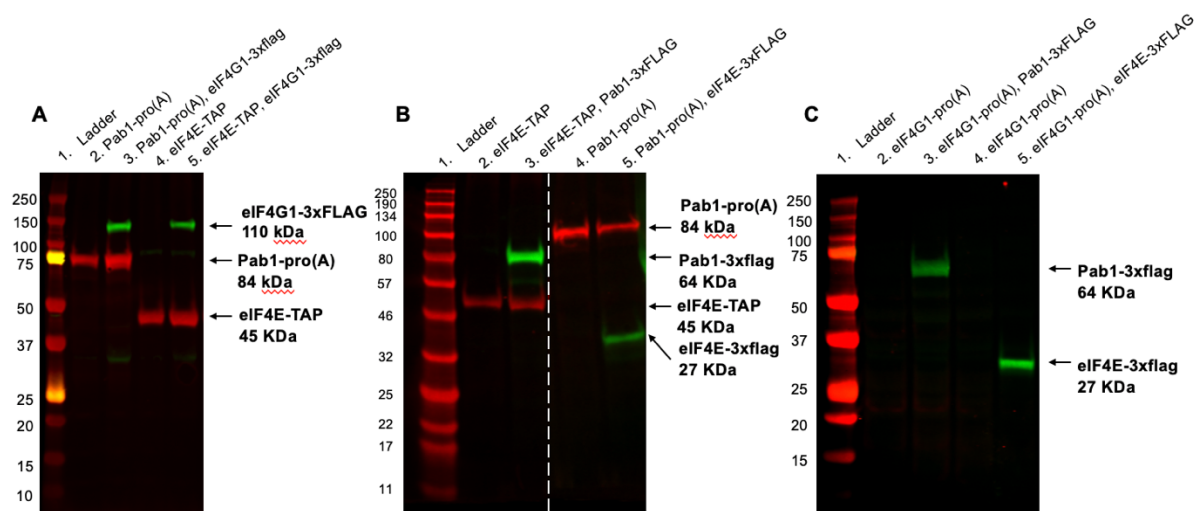


Figure 3.7. Yeast strains before and after 3xFLAG tag incorporation

Equal amounts of whole cell lysates - from strain before and after transformation with the 3xFLAG tag – were loaded onto a 4-12% SDS-PAGE gel and electrophoretically resolved, transferred onto a PVDF membrane and probed for either the protein(A) epitope (red) or the 3xFLAG epitope (green) using antibodies against the respective proteins (dilution 1:1000). The membrane was then probed with the corresponding Licor secondary antibody (1:20,000 dilution) and imaged using the Odyssey CLx Imaging system.

3.1.4 Growth curves

Of the strains generated, eIF4E-TAP, eIF4G1-3xFLAG was the main strain used in the studies described in this thesis, and is the strain used for 5' end pulldown and sequencing. This strain was therefore compared to the untagged WT strain, to ensure that the 3xFLAG tagging had not introduced any major growth defects or a slow-growing phenotype. OD600 measurements were recorded over a 7-hour period for a WT culture and a culture of the eIF4E-TAP, eIF4G1-3xFLAG strain in parallel. This was performed for three replicates from three separate -80 °C freezer stocks. Broadly, all replicates grew similarly for each strain (Figure 3.8a) and there was no significant difference ($p > 0.05$ for all timepoints) in OD600 at each timepoint between the WT and eIF4E-TAP, eIF4G1-3xFLAG strain (Figure 3.8b). It was concluded that the tagging of eIF4G1 had not significantly impaired growth or resulted in a slow-growing phenotype.

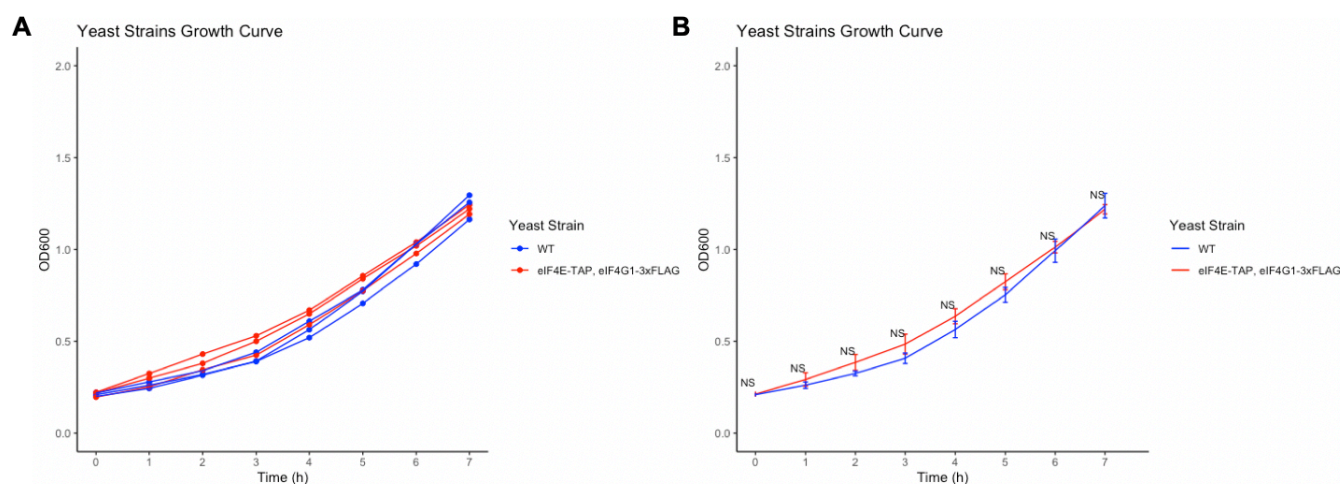


Figure 3.8. Comparison of growth curves between WT and eIF4E-TAP, eIF4G1-3xFLAG strains

A) Individual OD600 measurements for WT and eIF4E-TAP, eIF4G1-3xFLAG strains are shown. Strains were measured in parallel at hourly intervals over a 7-hour period, on three separate occasions. B) Mean $\pm$ SD for each strain at each timepoint is displayed ($n=3$). NS denotes non-significant differences between strains at each timepoint ($p < 0.05$, two-tailed t-test).

3.2 Optimising immunoprecipitation conditions

After additional 3xFLAG tags were introduced at the C-terminals of the closed-loop proteins, the next step was to use these new strains to optimise immunoprecipitation conditions, namely choice of immunoprecipitation beads, extent of crosslinking and extent of RNase I

digestion. This was an iterative process, as all factors need to be controlled and balanced with each other for maximal yield and specificity, while allowing for detection of both the closed-loop proteins and the mRNAs fragments at each end.

3.2.1 Bead selection

In selecting the immunoprecipitation beads used to pulldown the protein(A)-tagged closed-loop factors, the main consideration was yield. Three different types of beads were selected for trial with the Pab1 as the main protein to be immunoprecipitated. As Pab1 is the most abundant of the three closed-loop proteins, conditions optimised to give a good recovery of Pab1 in the eluate would be expected to also give a good recovery of the other two factors.

3.2.1.1 Sheep Ig vs mouse IgG bead comparison

The first beads trialled were magnetic sheep anti-mouse IgG beads, which were the most readily-available at the time. As protein(A) is known for binding the Fc region of antibodies, it was thought that the protein(A) tag on the closed-loop proteins would also bind the Fc region of the sheep antibodies conjugated to these beads. A standard immunoprecipitation reaction was performed using clarified whole cell lysate (WCL) from the Pab1-pro(A), eIF4E-3xFLAG strain, according to the conditions described in the original study (Archer *et al.*, 2015), but without the crosslinking or RNase I digestion steps. It was expected that at minimum, Pab1-pro(A) would be detected in the input and in the eluate, with variable amounts detected in the immunoprecipitation flow through depending on bead:lysate ratio. GAPDH, a negative control, was expected to be present in the input (WCL) and flow throughs (FT), but not present in the eluate, as the GAPDH protein is not protein(A) tagged or expected to significantly associate with Pab1. These expectations were met in the case of GAPDH (Figure 3.9a). However, in the case of Pab1-pro(A), very little Pab1-pro(A) was observed in the eluate. Increasing the ratio of beads from 0.5 mg beads per 10 mg of original grindate to 1 mg or 1.5 mg beads per 10 mg grindate did not increase binding efficiency, with the majority of the Pab1-pro(A) remaining in the FT, suggesting overall low binding efficiency.

This result prompted a more extensive literature search, revealing a much higher affinity of protein(A) binding to mouse IgG compared to sheep Igs (Lindmark, Thorén-Tolling and Sjöquist, 1983). Thus, the same magnetic beads coupled to mouse IgG were tested under the conditions described above. Both bead:grindate ratios of 0.5 mg bead : 10 mg grindate and 1 mg bead : 10 mg grindate gave improved recovery of Pab1-pro(A) in the eluate compared to previous beads, with very little in the FT lanes. Presence of GAPDH was similar to before, with detectable amounts in the input and FT, but little in the IP (Figure 3.9b). As the mouse IgG beads resulted in much better protein binding, experiments proceeded with these beads.

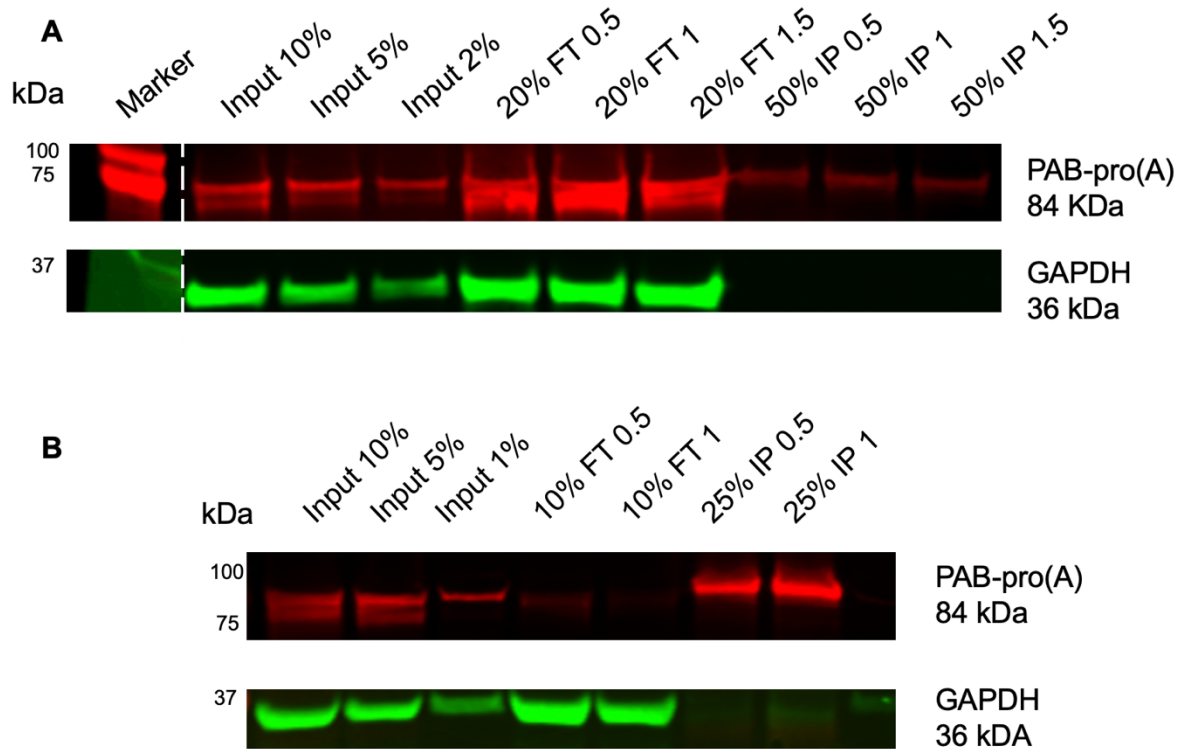


Figure 3.9. Sheep vs Mouse IgG bead binding comparison

Whole cell lysates from the Pab1-pro(A), eIF4E-3xFLAG strain were immunoprecipitated with magnetic beads coupled to either A) sheep Ig or B) mouse IgG2a, according to experimental conditions described in Archer et al., 2015. Inputs = WCL kept aside before immunoprecipitation, FT = WCL leftover after selection with magnetic beads, IP = material eluted from IgG conjugated beads after immunoprecipitation. Percentages refer to percentage of each fraction loaded in that lane. 0.5, 1 and 1.5 after fraction name refer to bead to grindate ratio – either 0.5, 1 or 1.5 mg of beads per 10 mg grindate. As grindate aliquots were prepared consistently but the clarified lysate volumes differed between experiments, beads used were calculated per 10 mg grindate.

3.2.1.2 FLAG beads

As the 3xFLAG epitope can also be used for immunoprecipitation and elution can be performed by a competing 3xFLAG peptide, I explored whether it would be possible to also immunoprecipitate with the FLAG tag. This would give the advantage of being able to do a double pull-down, first with a protein(A) pulldown for one closed-loop factor, then with the FLAG-epitope for another closed-loop factor, or vice versa, providing another confirmation that the proteins were indeed co-immunoprecipitating, increasing specificity. FLAG beads were used with a Pab1-3xFLAG strain, with overall method as described in the previous section 3.2.1.1. Immunoprecipitation with FLAG beads was efficient and specific, as demonstrated by the lack of GAPDH in the IP, and the high recovery of Pab1-3xFLAG in the IP, despite some protein remaining in the FT and on the beads after peptide (Figure 3.10, top and bottom rows). However, when probed for eIF4E-pro(A), a band can be seen in the beads lane, which corresponds to the beads boiled with detergent after peptide elution (to determine if elution was efficient) (Figure 3.10, middle row). Upon consideration, it was determined that as the FLAG-beads also contained an Ig, the protein(A) tagged factor can bind directly to that antibody rather than through Pab1. It would not be possible to discern whether co-detection of a protein(A) tagged factor upon 3xFLAG pulldown was due to association with the 3xFLAG factor, or by direct binding of the protein(A)-tagged factor to the immunoprecipitation antibody. Therefore, immunoprecipitation using the 3xFLAG tag was no longer considered.

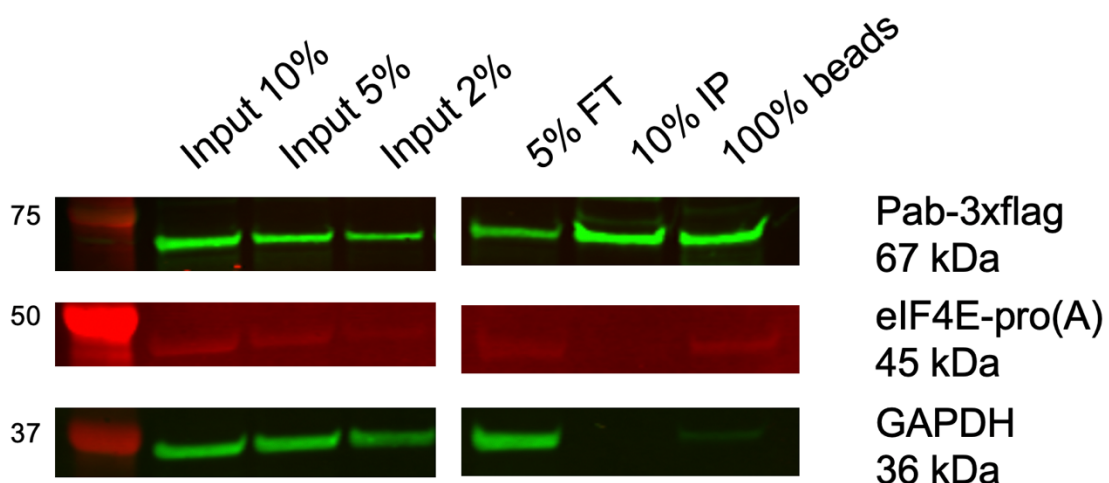


Figure 3.10. Immunoprecipitation using the 3xFLAG tag

Immunoprecipitation of eIF4E-TAP, Pab1-3xFLAG WCL was performed with FLAG beads in method similar to that described in the previous section. Inputs = WCL kept aside before immunoprecipitation, FT = WCL

leftover after selection with magnetic beads, IP = material eluted from IgG conjugated beads after immunoprecipitation. Percentages refer to percentage of each fraction loaded in that lane.

3.2.2 Formaldehyde crosslinking optimisation

A key feature of the experimental protocol was formaldehyde fixation, which rapidly crosslinks macromolecules separated by ~ 2 Å. This technique enables the study of closely spaced protein-protein, protein-DNA, and protein-RNA interactions (Thavarajah *et al.*, 2012; Hoffman *et al.*, 2015). The speed of formaldehyde crosslinking is crucial for investigating translation complexes, as it swiftly permeates cell membranes and fixes these dynamic complexes, preserving their near-native state.

Despite its advantages, formaldehyde crosslinking conditions need careful optimisation. The percentage of formaldehyde needs to be high enough to adequately capture complexes in proportions reflective of *in vivo* conditions, but also low enough to avoid decreased specificity and yield from over-fixation. Therefore, to determine the optimal level of crosslinking to use, different crosslinking percentages were tested for their ability to – 1) capture translation complexes, as determined by polysome profiling, and 2) allow simultaneous detection of two closed-loop proteins by western blotting after immunoprecipitation.

3.2.2.1 Polysome profiling

Utilising the ability of formaldehyde to capture translation complexes, polysome profiling was employed as a method of determining the optimal extent of fixation using different percentages of formaldehyde. Specifically, as formaldehyde had already been used to successfully fix polysomal complexes (Archer *et al.*, 2016), the increase in capture of polysomes by polysome profiling was used as an indication of increasing crosslinking. Furthermore, due to the lack of preference with which formaldehyde fixes different types of biomolecules together, fixation of polysomes was deemed an appropriate readout for overall extent of crosslinking. As shown in Figure 3.11, few polysomes were captured with 0% formaldehyde. Polysome capture increased substantially when percentage of crosslinking was increased to 1%. Further increases in crosslinking percentage to 2% and 3% resulted in slight increases in polysomes captured (Figure 3.11a). When each profile was normalised to its own

80S peak, the increase in polysome capture with 2% and 3% formaldehyde was more prominent (Figure 3.11b). However, after normalisation, the difference in polysome capture between 2% and 3% crosslinking was still minimal, suggesting little benefit in further increasing formaldehyde percentage. Therefore, the midway between the two conditions was settled on at 2.5% formaldehyde.

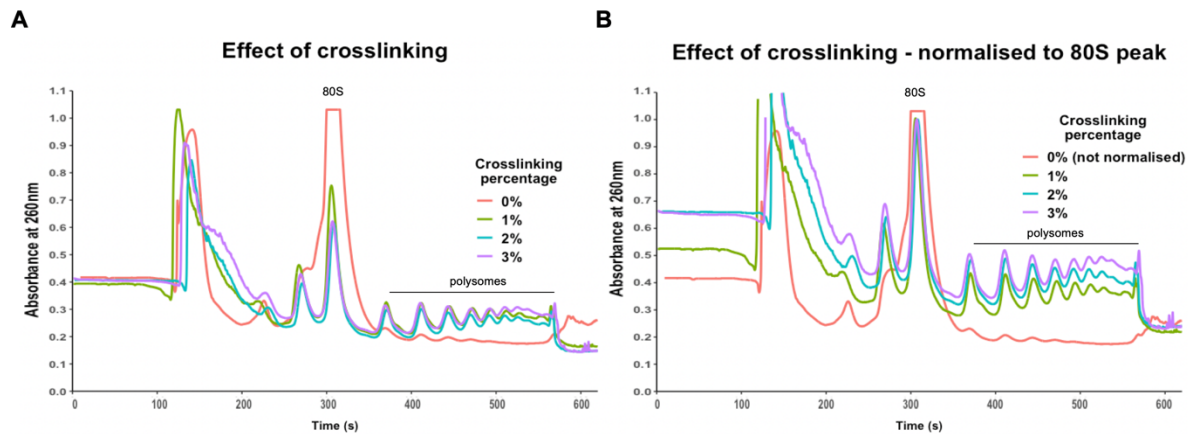


Figure 3.11. Effect of formaldehyde percentage on capture of polysomes

Exponentially growing yeast cells were fixed to a final percentage of 0, 1, 2 or 3% formaldehyde by snap-chilling and addition of formaldehyde, followed by incubation on ice for 10 mins. Clarified lysate from such cells were subject to sucrose gradient sedimentation by ultracentrifugation, followed by fractionation to determine polysome profiles. Polysome profiles of samples crosslinked with 0-3% formaldehyde are compared either unnormalised (A), or normalised to their 80S (monosome) peak (B).

3.2.2.2 Western blotting of closed-loop components

The same samples crosslinked as above were also tested for their ability to show co-immunoprecipitation of the closed-loop components by western blotting.

Immunoprecipitation from the whole cell lysate of the Pab1-pro(A), eIF4E-3xFLAG strain was performed with previously optimised mouse IgG beads, in order to pull down the closed-loop complex via the protein(A)-tagged component. It was expected that increasing crosslinking percentage would increase co-immunoprecipitation of the 3xFLAG tagged component relative to the protein(A) tagged component.

Interestingly, the best recovery of Pab1-pro(A) in the IP appeared to be from the sample with 0% crosslinking. Samples with 1, 2, and 3% crosslinking were not markedly different in their recovery of Pab1-pro(A) (Figure 3.12, top plot, row 1). A possible decrease of Pab1-pro(A) recovery with increasing crosslinking could be argued, although this is very subtle, and it is difficult to say whether it is a true difference or due to differences in loading amounts. In

regards to recovery of the FLAG-tagged factor, eIF4E, the 0% crosslinking also appeared to give best recovery of eIF4E-3xFLAG, but it should be noted that Pab1-pro(A) also had a greater recovery in this sample, so the recovery relative to the pulled down factor may not be greater than the other crosslinking conditions. Crosslinking percentages 1-3% did not appear to vary greatly in eIF4E-3xFLAG recovery (Figure 3.12, top plot, row 2). GAPDH did not appear prominently in any IP samples (Figure 3.12, top plot, row 3). Levels of eIF4E-3xFLAG in the flow throughs also did not differ substantially between crosslinking percentages once differences in loading were taken into account, as evidenced by the same differences mirrored in the corresponding inputs (Figure 3.12, bottom plot, row 2).

Overall, it was concluded that 0% crosslinking gave better yield during immunoprecipitation, at least in this instance, possibly due to greater availability of the protein(A)-tagged Pab1 to the IgG beads by virtue of not being crosslinked to other biomolecules in complexes. Interestingly, eIF4E-3xFLAG could also be detected despite no crosslinking, potentially because of some degree of preservation of protein-protein interactions by the immunoprecipitation conditions. 1-3% crosslinking did not appear to noticeably increase recovery of eIF4E-3xFLAG with Pab1. GAPDH was also not observed co-immunoprecipitating with Pab1-pro(A) for the conditions tested, a positive indicator that non-specific over-fixation was unlikely to be happening. Although there did not seem to be a large improvement in co-immunoprecipitation of the closed-loop components with increasing formaldehyde, crosslinking was still determined to be required for preservation of the protein-mRNA interactions. Therefore, the formaldehyde percentage was set at 2.5% considering the previous polysome profiling results.

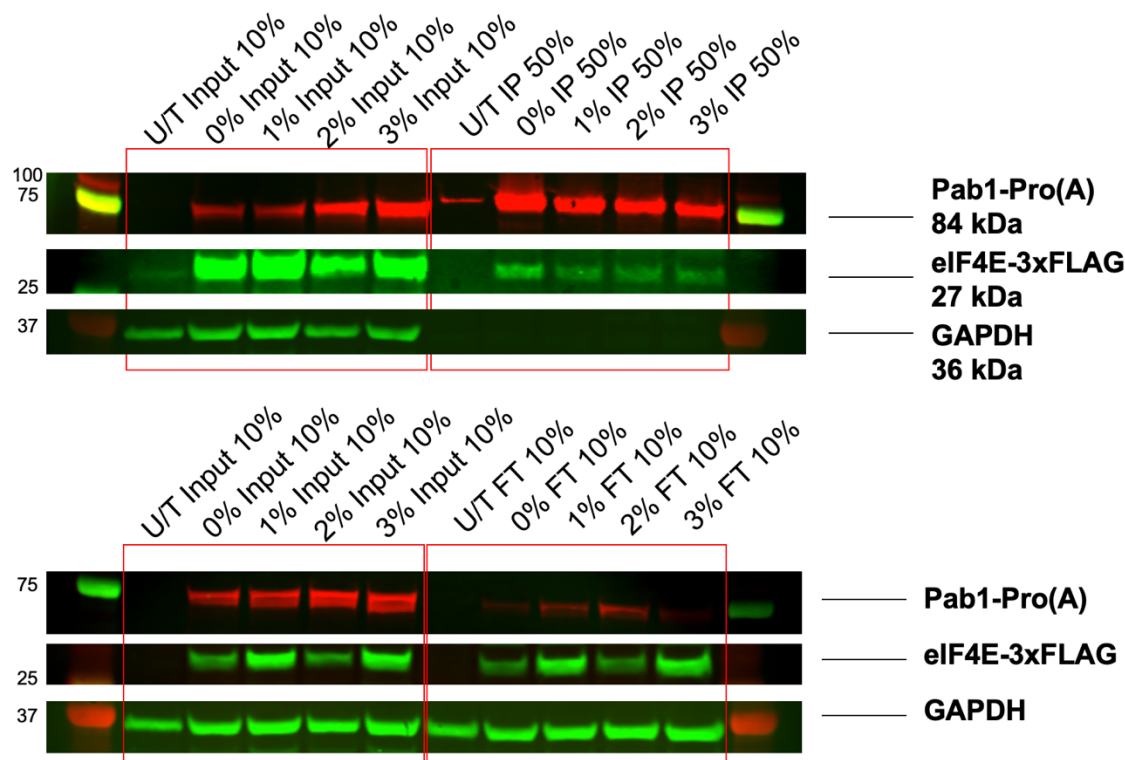


Figure 3.12. Effect of crosslinking percentage on co-immunoprecipitation of closed-loop components

Exponentially growing cells were fixed to a final percentage of 0,1,2 or 3% formaldehyde by snap-chilling and addition of formaldehyde, followed by incubation on ice for 10 mins. Clarified lysate from such cells were loaded onto a 4-12% SDS-PAGE gel and electrophoretically resolved, transferred onto a PVDF membrane and probed for either the protein(A) epitope (red) or the 3xFLAG epitope (green) using antibodies against the respective proteins (dilution 1:1000). The membrane was then probed with the corresponding Licor secondary antibody (1:20,000 dilution) and imaged using the Odyssey CLx Imaging system. U/T = untagged WT control, input = WCL set aside before immunoprecipitation, FT = flow through, IP = immunoprecipitation eluate. Percentages in labels denote percentage of each fraction loaded.

3.2.3 RNase I digestion optimisation

Next, the extent of RNase I digestion needed optimisation in order to balance adequate depletion in the middle of the mRNA and detecting enough mRNA recovery at the 5' and 3' ends. If digestion is too mild and the middle sections of mRNAs are not depleted, then both ends of the mRNA may be present despite no closed loop forming. If digestion is too strong, the ends may no longer be intact for detection of the closed-loop. Two techniques were utilised to determine optimal RNase I digestion – polysome profiling to measure collapse in polysome peaks and qPCR to measure enrichment at the ends and middle of mRNAs.

3.2.3.1 Polysome profiling

To accurately compare the different RNase I concentrations, polysome profiling was performed on samples digested in parallel with different concentrations of RNase I. The metric focused on was on the collapse of the polysome peaks, especially the disome peak, to suggest that the mRNA had been cut between ribosomes.

With no RNase I digestion, a prominent polysome fraction can be observed (Figure 3.13, red). With even 1 U of RNase I per absorbance unit of lysate, a collapse of the polysome fraction can already be observed. Increasing RNase I concentration collapsed the polysomal fraction even further, although the biggest difference was between 0 U/Abs and 1 U/Abs. Here a very minimal digestion was targeted, so the minimum condition to result in a collapse of the polysome fraction was chosen, being 1 U RNase I per 1 absorbance unit of lysate.

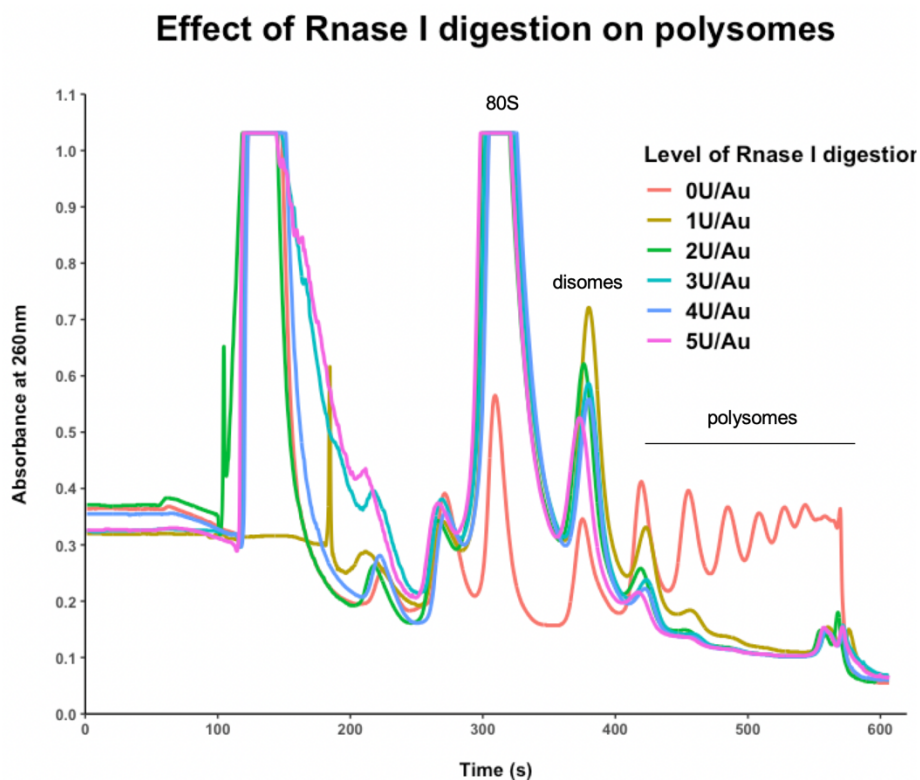


Figure 3.13. Effect of RNase I digestion on polysomal fraction

One preparation of WCL from the Pab1-pro(A), eIF4E-3xFLAG strain was split into 6 tubes of equal volume. Each tube was incubated in parallel with the specified units of RNase I, then stopped at the same time with SupraseIN. Digested lysates were separated by sucrose gradient sedimentation, followed by fractionation to determine polysome profiles.

3.2.3.2 qPCR validation

Here, the Pab1-pro(A), eIF4E-3xFLAG strain was again used (referred to as L2 in Figure 3.14 legend). To ensure that mRNA signals could be detected at the 5' and 3' ends for different RNase I digestion conditions, qPCR was performed with primers designed for 5', middle and 3' sequences of two mRNAs tested in the original study – *RPS20* and *GUS1*. When pulled-down with Pab1 in the original study, a high IP signal enrichment above input was observed for the 3' amplicon, for both *RPS20* and *GUS1*. A slight signal was detected at the 5' end for *RPS20*, and less was reported for *GUS1*. Very little signal enrichment was detected from the middle amplicon for both. Comparing those results to the ones below in Figure 3.14, an RNase I concentration of 1 U/absorbance unit of lysate appeared to give the profile most similar to those from the original study. 0 units of RNase I resulted in a profile containing all three signals – 3', middle and 5', as expected from intact, undigested mRNA. Increasing RNase I digestion past 1U/Abs decreased the signal at the 3' end (the pulldown end) substantially, and also abolished detection of recovery at the 5' end (opposite to pulldown). Therefore, from the qPCR results, 1 U RNase I per 1 absorbance unit of lysate was chosen as it appeared to be most appropriate from the mRNA perspective. This was in accordance with the conclusion derived from polysome profiling in the previous section.

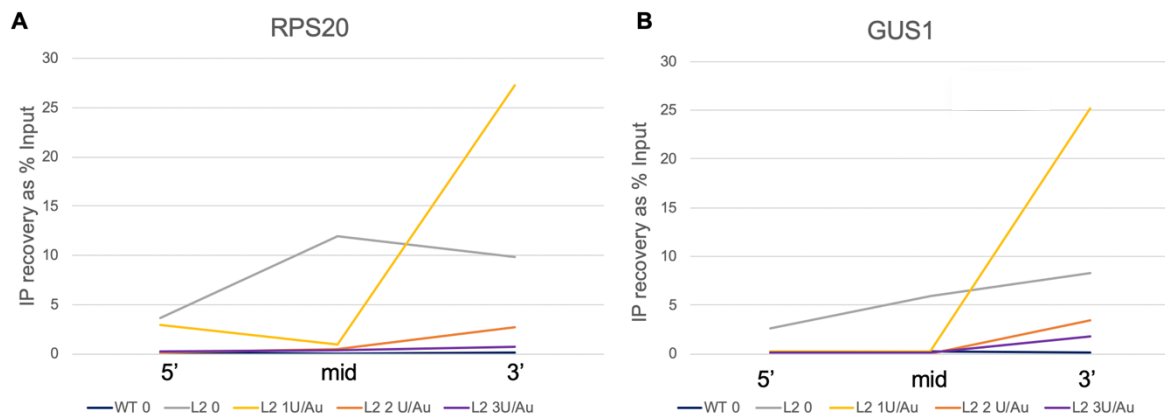


Figure 3.14. Effect of RNase I digestion on mRNA recovery at ends

One preparation of WCL from the Pab1-pro(A), eIF4E-3xFLAG strain was split into 4 tubes of equal volume. An untagged WT control was prepared in parallel. The specified units of RNase I were added to each tube and immunoprecipitation was performed in parallel for all tubes. mRNA was extracted, reverse transcribed and assayed by qPCR using primers specific for 5', middle (mid) and 3' sequences for each of *RPS20* (A) and *GUS1* (B).

3.3 Recreation of results from original study

With each individual condition optimised, one more immunoprecipitation was performed to ensure all individually optimised conditions (Mouse IgG beads, 2.5% crosslinking, 1 U/Abs RNase I) remained ideal when taken together. The main focus was to ensure the closed-loop proteins co-immunoprecipitated together, and that mRNA signal could be detected at the 5' and 3' ends. Western blotting and qPCR, respectively, were employed for testing that these two conditions were met.

3.3.1 Co-immunoprecipitation of closed-loop components

Two strains, eIF4E-TAP, eIF4G1-3xFLAG and Pab1-pro(A), eIF4G1-3xFLAG were used for immunoprecipitation to observe whether the middle factor, eIF4G1, could be detected in the eluate when immunoprecipitation was performed via protein(A) tagged eIF4E or Pab1. When eIF4E was pulled down, a noticeable recovery of both eIF4E and eIF4G1 can be detected in the eluate by western blotting (Figure 3.15a), reflecting the tendency of eIF4G and eIF4E to copurify as the eIF4F complex (Merrick, 2015). GAPDH – the negative control not expected to associate with the closed-loop factors – was not detected in the eluate, reflecting specificity of pulldown. A fraction of eIF4E and eIF4G1 was detected in the flow through, indicating some level of incomplete binding or saturation of bead binding capacity, but recovery in the IP was still adequate to detect. Pab1- pro(A) was also detected in significant amounts in the IP compared to input, with very little unbound protein detected in the flow through (Figure 3.15b). Co-immunoprecipitation of eIF4G1 with Pab1 was not as prominent, consistent with Pab1's function as an abundant general poly(A) binding protein with roles independent of its association with eIF4G1. Again, GAPDH was not observed in the IP, suggesting specificity of the pull down. Overall, it was concluded that under the experimental conditions specified, the closed-loop components co-immunoprecipitate.

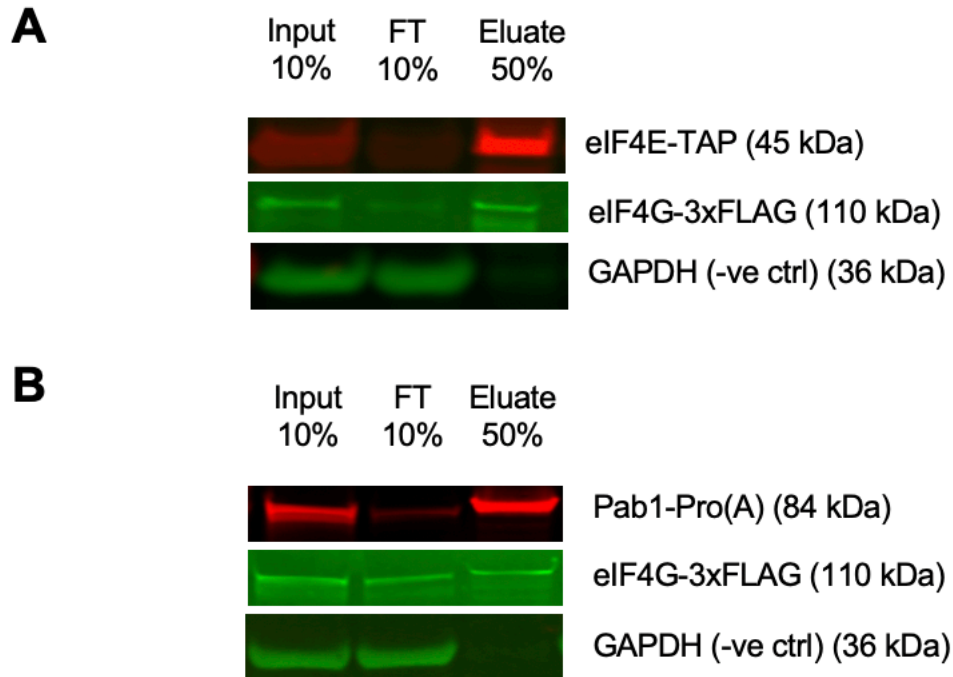


Figure 3.15. Coimmunoprecipitation of closed-loop components

Immunoprecipitation was performed on crosslinked WCLs of strains *eIF4E-TAP*, *eIF4G1-3xFLAG* (A) and *Pab1-pro(A)*, *eIF4G1-3xFLAG* (B). Cell lysates were incubated with mouse IgG immunoprecipitation beads with concurrent RNase I digestion, and each fraction was retained (input = WCL set aside before immunoprecipitation, FT = flow through, IP = immunoprecipitation eluate). Percentages in labels denote percentage of each fraction loaded. Samples were loaded onto a 4-12% SDS-PAGE gel and electrophoretically resolved, transferred onto a PVDF membrane and probed for either the protein(A) epitope (red), the 3xFLAG epitope (green) or GAPDH, the negative (-ve) control (green), using antibodies against the respective proteins (dilution 1:1000). The membrane was then probed with the corresponding Licor secondary antibody (1:20,000 dilution) and imaged using the Odyssey CLx Imaging system. Western blot shown is a representative example from $n = 3$.

3.3.2 mRNA end enrichment analysis by qPCR

Using the same methodology as described in the *qPCR for RNase I digestion* section (3.2.3.2), sequences at the 5' and 3' end of example mRNAs were probed by qPCR to ensure retention of signals from both ends of the message after immunoprecipitation and RNase I digestion. Here, the *eIF4E-TAP*, *eIF4G1-3xFLAG* strain was used as it was the strain to be used for subsequent RNA-sequencing. A 5' pulldown through *eIF4E* was favoured as the first immunoprecipitation to perform, as it was the one with best mRNA recovery at the other end (3' end). From the original study, 5' signal was generally low upon 3' end pulldown through *Pab1*, likely due to the abundance of *Pab1* and its existence in contexts other than as part of the closed-loop.

qPCR was performed with primers designed for 5', middle and 3' sequences of two mRNAs tested in the original study – *RPS20* and *SSC1*. When pulled-down with eIF4E-TAP in the original study, a high IP signal enrichment above input was observed for the 5' amplicon, for both *RPS20* and *SSC1*. A high signal was detected at the 3' end for *RPS20* suggesting strong closed-loop formation, while less was reported for *SSC1*. Very little signal enrichment was detected from the middle amplicon for both (Figure 3.16a,b, right panel, orange).

In the current study, *RPS20* showed a recovery of the 5' amplicon in the IP with an average value of ~6% of input signal. This value dropped for the middle signal, but increase again to ~13% at the 3' end (Figure 3.16a, left panel). Although it appears recovery at the 3' end was higher than at the pulled down 5' end, this result was skewed by a single replicate with one low value in the 5' end and one high value in the 3' end, reflected by the large error bar. It is anticipated that an increased replicate number would narrow the discrepancy between the ends. The recovery at the 3' end was not as pronounced for *SSC1*, with an enrichment of ~4% of input compared to ~12% of input at the 5' end. Middle enrichment was lower than both ends (Figure 3.16b, left panel). The IP material from the untagged WT control was also tested in parallel for the same amplicons, and very minimal signal was detected at any location, suggesting minimal non-specific pulldown of proteins and associated mRNAs. qPCR results from this current study broadly reflected previously reported observations. Therefore, it was determined that conditions required to obtain the results seen in Archer *et al.* (2015) had been faithfully recreated.

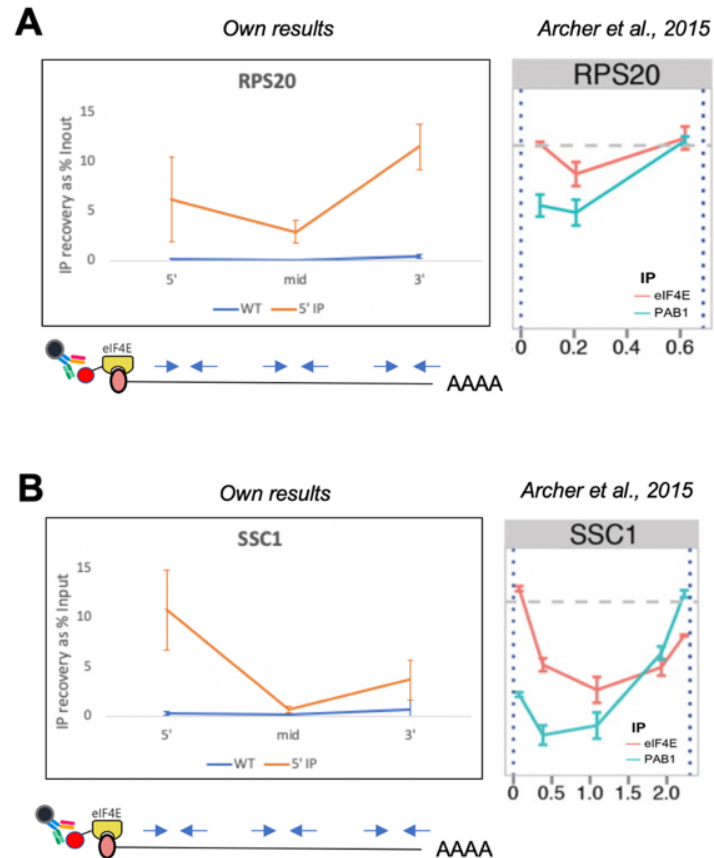


Figure 3.16. mRNA enrichment at 5' and 3' ends upon 5' immunoprecipitation

Immunoprecipitation was performed on crosslinked WCLs of strain *eIF4E-TAP*, *eIF4G1-3xFLAG* (5' IP) and an untagged *WT* strain (*WT*). Cell lysates were incubated with mouse IgG immunoprecipitation beads with concurrent RNase I digestion, and a fraction of WCL was also RNase I digested in parallel and retained as input. mRNA from inputs and IPs was extracted, reverse transcribed and probed by qPCR using probes specific for signal at the 5', middle and 3' regions of *RPS20* (A) and *SSC1* (B). Profiles obtained are displayed next to results from Archer et al. (2015).

3.4 Chapter 3 Conclusion

To briefly conclude, this chapter describes the development of a method to capture the closed-loop and associated mRNAs *in vivo*. First, 3xFLAG tagging of the closed-loop proteins was performed to ensure a means of detecting the proteins simultaneously after immunoprecipitation. Next, optimal conditions for immunoprecipitation were determined, namely the choice of immunoprecipitation beads, extent of formaldehyde crosslinking and extent of RNase I digestion. A combination of western blotting, polysome profiling and qPCR techniques were utilised to ensure adequate crosslinking, specificity of immunoprecipitation, depletion of mRNA middle sections, and adequate enrichment at 5' and 3' ends. Finally, immunoprecipitation under the optimised conditions was performed to confirm co-immunoprecipitation of the closed-loop proteins and detectable signal enrichments at the 5' and 3' ends of example mRNAs upon 5' pulldown. Results reflecting the previous study by Archer *et al.* (2015) were recreated. Therefore, it was concluded that a method to capture the closed-loop and associated mRNAs had been achieved, and that this method was ready for use with RNA-sequencing to analyse closed-loop prevalence across the transcriptome.

Chapter 4 – Development of an RNA-seq method and bioinformatic pipeline for transcriptome-wide analysis of closed-loop formation

Note on contribution

I performed all laboratory work, bioinformatic processing/coding and figure generation in section 4.1.1 (initial sequencing experiments). The pipeline for initial processing and filtering of final replicates (section 4.1.2) was adapted from Archer et al. (2016) and coding and data processing was performed by Dr Attila Horvath. Adaptions from the original pipeline were decided in collaboration between myself and Dr Horvath.

The approach detailed in section 4.2 was decided upon by collaboration between myself and Dr Horvath. Coding and data processing were performed by Dr Horvath. Visual verification was performed by myself. All figures and schematics in this section were generated by myself.

The reproducibility heatmap and plots in section 4.3 were produced by Dr Horvath.

4 Development of an RNA-seq method and bioinformatic pipeline for transcriptome-wide analysis of closed-loop formation

After development of an experimental method to capture the closed-loop, the next step was to perform short-read RNA sequencing on the material collected, followed by processing of the sequencing results for further transcriptome-wide analysis of the closed-loop. At this stage, one replicate consisted of four samples – wild-type input (WT IN), eIF4E-TAP tagged input (4E IN, or tagged IN), wild-type immunoprecipitation (WT IP) and eIF4E-TAP tagged immunoprecipitation (4E IP, or tagged IP). It was expected that both WT and tagged inputs, being whole cell lysates digested in parallel with the immunoprecipitation reactions, would have even coverage across the transcriptome. The WT IP sample was expected to contain little RNA and represent background binding with coverage profiles similar to the inputs. Finally, the tagged IP sample was expected to show enriched coverage at the 5' end, a depletion of reads in the middle, and potential coverage enrichment at the 3' end.

Short read RNA-sequencing, among its many applications, is a technique often used to determine the comparative presence of mRNAs within and between biological samples. A standard RNA-sequencing pipeline for differential gene analysis generally consists of RNA extraction, fragmentation of the RNA, reverse transcription into cDNA, short-read sequencing, mapping to the genome/transcriptome and finally comparison of reads mapping to each gene between samples. These studies have been widely performed, and various bioinformatic processes and pipelines have been produced to analyse such data. However, while the aim of this study was also to compare mapping of reads between genes, there are some crucial differences that render a standard differential gene analysis approach unsuitable. For example, while poly(A) selection of mRNAs, followed by fragmentation and random priming for cDNA synthesis is a widely adopted library preparation method, our samples already consist of fragmented mRNA from the use of RNase I digestion. Therefore, alternative methods for mRNA selection and cDNA synthesis were required. Furthermore, most differential gene analysis programs count and normalise the reads mapping to each gene and compare these numbers between samples, regardless of where the reads are located on the gene. However, as we are most interested in capturing and comparing reads from the ends of mRNAs to suggest closed-loop formation, we also needed to gather distribution information of where reads are mapping to along the length of the gene. Overall, it was

evident we needed to develop a custom RNA-seq methodology suitable for use with our aims and specific experimental method.

In this chapter, I detail the library preparation approaches trialled and the rationale for the method ultimately chosen. Next, I introduce the pipeline for the initial filtering and mapping of the libraries to the yeast genome. Mapping statistics are also presented. I also introduce a technical challenge we encountered, being the lack of 5' and 3' UTR annotations for many yeast genes. As our analysis focuses on the enrichments from the ends of mRNAs, the accurate annotation of yeast UTRs was particularly important. I therefore present a solution we devised to address this issue. Finally, reproducibility between the replicates is discussed.

Overall, in order to draw conclusions regarding the extent of closed-loop formation across the transcriptome and compare differences between mRNAs, the technical aspects of RNA-sequencing and its analysis first required careful consideration. This chapter explains these considerations and details the methods we have employed to address them.

4.1 Choice of library preparation and sequencing strategy

4.1.1 Initial sequencing strategy

Initially, the RNA isolated from the immunoprecipitation reaction described in chapter 3 was dried with vacuum centrifugation and sent for library preparation and sequencing by an Illumina short-read sequencing provider. Briefly, this strand-specific library preparation consisted of rRNA hybridisation and RNase H depletion, followed by DNase I depletion, cDNA synthesis, end repair, dA-tailing, adapter ligation and PCR enrichment. Libraries were selected for inserts of 150-200 bp using DNA clean-up beads. As no kits were available for depletion of yeast rRNA at the time, this kit included rRNA probes for human/mouse/rat. 150 bp paired-end sequencing was performed on the Illumina Hi-seq platform.

4.1.1.1 Overall sequencing quality

While RNA-sequencing is a high throughput method capable of generating large amounts of data regarding the composition of RNA species in a sample, this data must be tightly quality-

controlled for nucleotide sequencing accuracy, GC content or biases introduced during library preparation and sequencing. Therefore, quality control for each of the paired-end sequencing libraries was performed using FastQC and collated with MultiQC. All libraries had average read lengths of 150 bp, indicating that the sequencing platform used had consistently sequenced 150 bp fragments for all samples. All samples also had comparable GC content of between 47% - 51%. Although libraries were mixed and sequenced in equimolar proportions, the number of reads obtained for each sample ranged from 20.2 million to 190.8 million, highlighting the need for downstream library size normalisation (Figure 4.1a). Both mean quality scores and per sequence quality scores had Phred quality scores consistently above 30, indicating an error rate of less than 1 in 1000 nucleotides or 0.1% (Figure 4.1b,c). For all samples, per base N content was also low at an average of below 5 counts for every position along the read (Figure 4.1d). Overall, regarding the technical elements of the sequencing and library preparation, quality was consistently high.

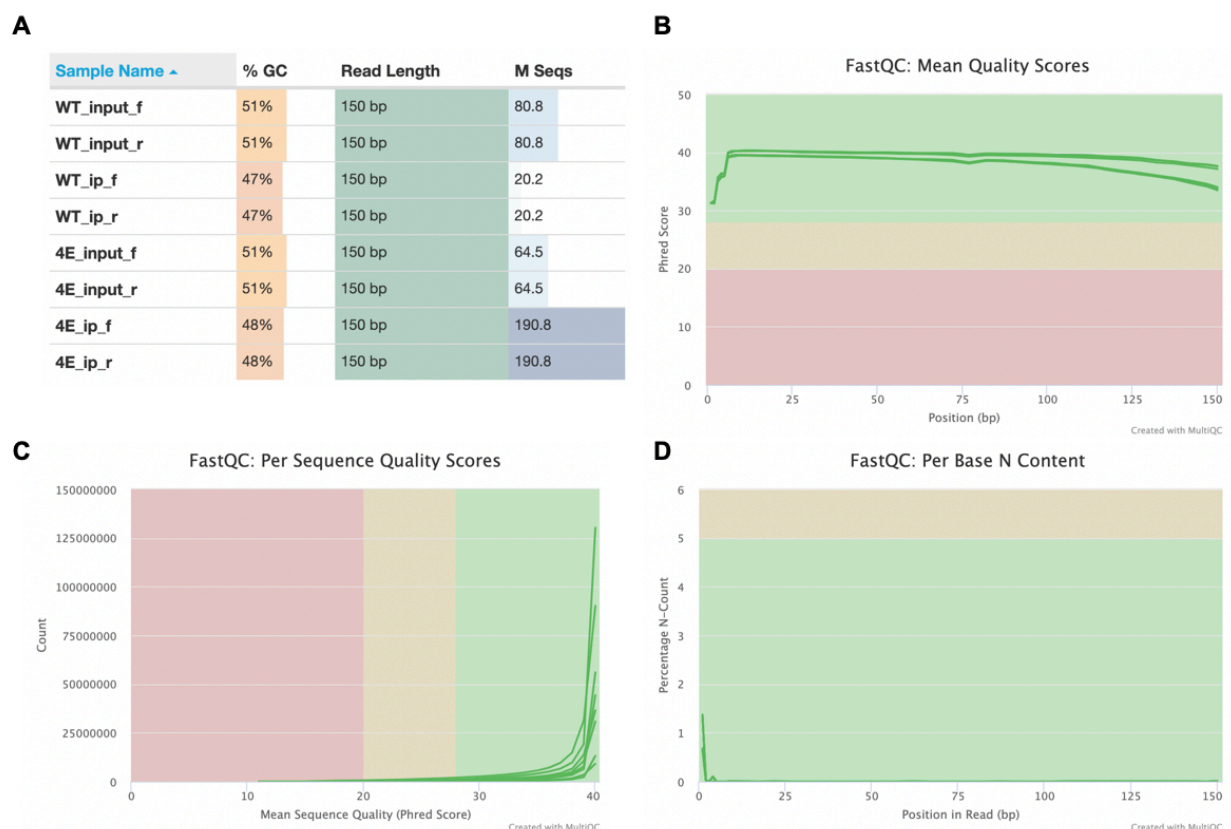


Figure 4.1. Initial quality control of sequencing libraries

Quality control statistics for raw sequencing reads received after strand-specific, 150-bp paired-end Illumina sequencing. FastQC was used to generate sequence quality reports and these were collated by MultiQC. *_f* and *_r* refer to R1 and R2 for each sample respectively. WT refers to the untagged wild-type sample, 4E refers to libraries made from the eIF4E-TAP, eIF4G1-3xFLAG strain. Input refers to whole cell lysate (WCL) digested in parallel with the immunoprecipitation reaction (ip). (A) Average %GC content, average read length in bp, and number of reads (in millions) for each sample (B) Collation of average Q scores (y-axis) for each position of the

read (x-axis) for all samples. A Q score above 30 represents a probability of an error at 1 in 1000 nucleotides. (C) Number of reads (y-axis) having a certain mean sequence Q score (x-axis) for each sample. (D) Number of Ns counted (y-axis) for each position along the read (x-axis) per sample.

4.1.1.2 Adapter trimming

Although sequencing quality was generally high, the initial FastQC analysis did indicate adapter contamination. All samples had at least 5% of their reads contain adapter sequence by positions 100bp – 125bp. Overall, all samples had between ~20-40% of reads with adapter sequences (Figure 4.2c). In order to improve subsequent mapping efficiency, I removed the adapter sequences using the CLIP step of Trimmomatic, a pre-processing tool for short-read sequencing. FastQC performed before and after trimming demonstrated efficient removal of adapter sequences, with an adapter content of <0.1% for all samples after trimming (Figure 4.2d). Average read lengths for each library also decreased from 150bp before trimming to between 139-142bp after trimming, consistent with the removal of adapter sequences. Total read numbers did not decrease significantly after trimming, suggesting overall few adapter-only read (Figure 4.2a,b)

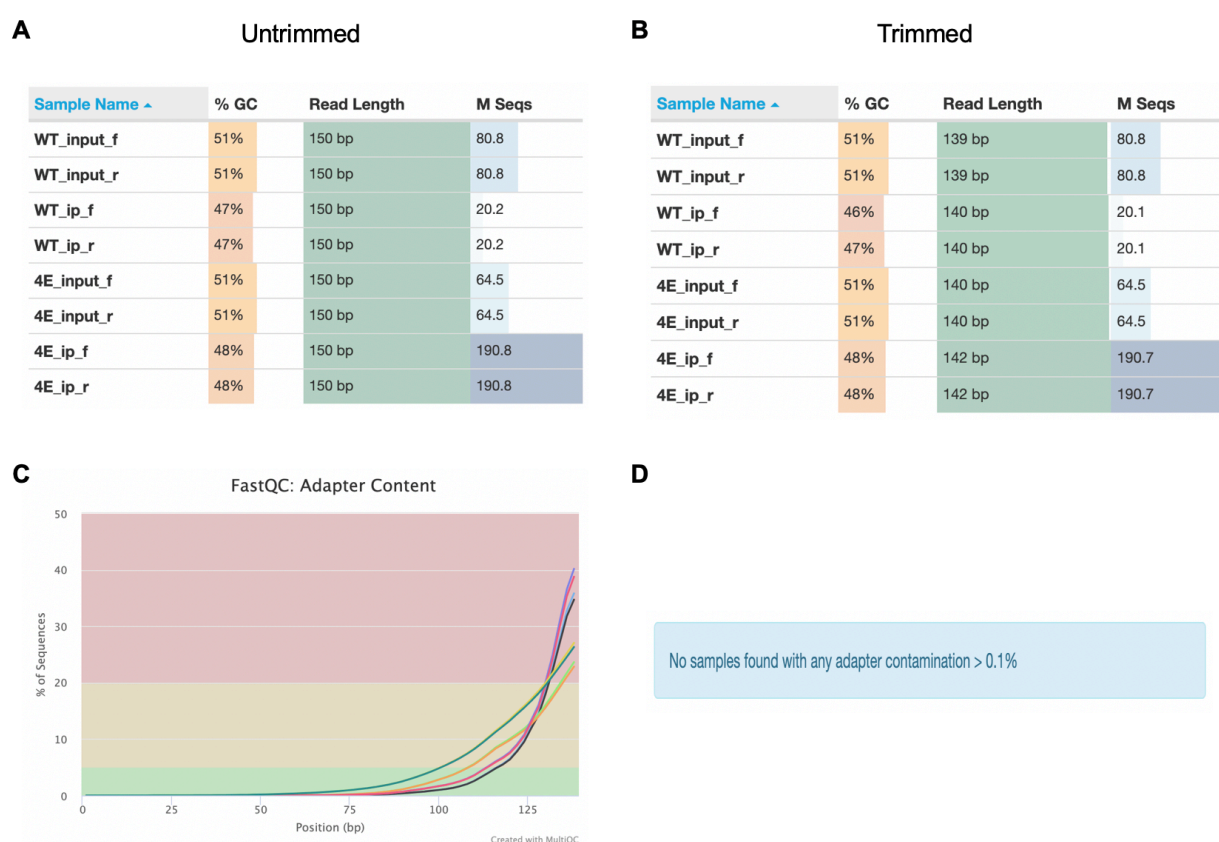


Figure 4.2. Sequencing statistics pre- and post- adapter trimming

*Comparison in adapter content, average read length and reads count before and after trimming with Trimmomatic. FastQC was used to generate sequence quality reports and these were collated by MultiQC. *_f* and *_r* refer to R1 and R2 for each sample respectively. WT refers to the untagged wild-type sample, 4E refers to libraries made from the eIF4E-TAP, eIF4G1-3xFLAG strain. Input refers to whole cell lysate (WCL) digested in parallel with the immunoprecipitation reaction (ip). (A-B) Average %GC content, average read length in bp, and number of reads (in millions) for each sample before (A), and after (B) trimming. (C-D) % of reads (y-axis) having adapter contamination by position along the read (x-axis) for each sample, before (C) and after (D) adapter trimming.*

4.1.1.3 rRNA content and overrepresented reads

As briefly mentioned previously, at the time of sequencing there were no yeast specific rRNA kits commercially available. Therefore, a library preparation kit with rRNA probes for human/mouse/rat was used instead. It was hoped that some sequence similarity would be shared with yeast, for even a partial depletion of rRNA. However, this was not the case. The rRNA content in each library was estimated to be 96.67% for WT IN, 99.3% for 4E IN, 31.8% for WT IP and 77.3% for 4E IP, as reported by the sequencing facility. FastQC analysis of the libraries also suggesting a significant proportion of overrepresented sequences, defined as sequences that occur in more than 0.1% of the total number of sequences. For input files, between ~78-88% of total sequences were overrepresented sequences, depending on library. This percentage was lower in the IP files, consistent with the selective enrichment of certain mRNAs in those samples. However, the 4E IP sample still contained ~68% and 53% overrepresented sequences for R1 and R2 respectively (Figure 4.3). A BLAST nucleotide search of the top 5 overrepresented sequences for all libraries consistently returned results corresponding to the yeast rRNA locus. It was therefore determined that rRNA was the main RNA species represented in this sequencing attempt.

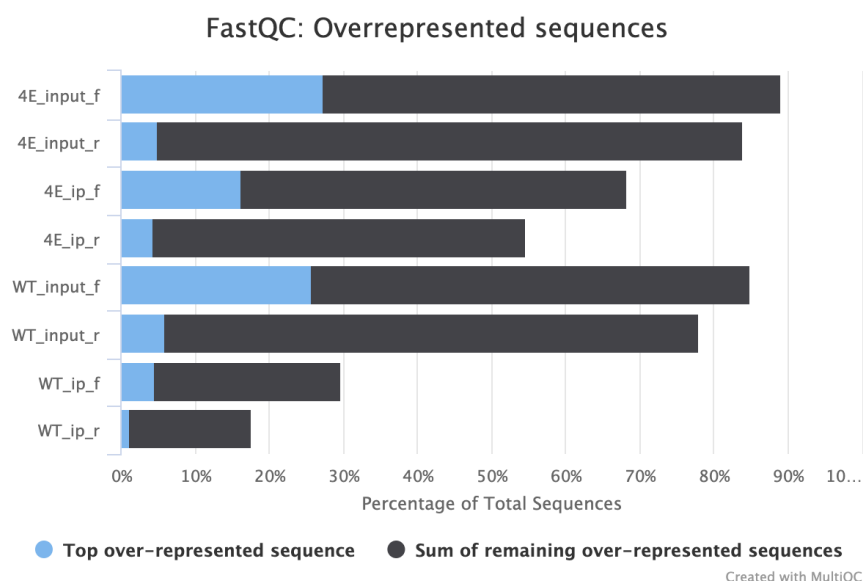


Figure 4.3. Overrepresented sequences as determined by FastQC

FastQC was used to determine the percentage of overrepresented sequences (over 0.1% of total reads) for all samples and statistics were collated with MultiQC. WT refers to the untagged wild-type sample, 4E refers to libraries made from the eIF4E-TAP, eIF4G1-3xFLAG strain. Input refers to whole cell lysate (WCL) digested in parallel with the immunoprecipitation reaction (IP). F and R refer to R1 and R2 respectively. Blue represents the % of total reads that contain the single top overrepresented sequence.

4.1.1.4 Alignment to genome

Despite the known high percentage of rRNA, it was hoped that with the deep sequencing performed, the number of useful mRNA reads obtained could still be sufficient for analysis. Therefore, in order to discern where the trimmed reads were originating from, the libraries were mapped back to the yeast S288C reference genome using STAR. Overall, only a small proportion of reads from each sample were uniquely mapped to the genome, ranging from 1.3% to 6.9%, corresponding to 2.6, 1.4, 0.8 and 9.7 million reads for WT input, WT IP, 4E input and 4E IP respectively (Figure 4.4a). A large proportion of reads were mapped to multiple loci, meaning they could not be definitively assigned to a genomic location. This is consistent with the high rRNA content of the sample, as the S288C reference genome has two annotated copies of the rDNA repeat locus, RDN37-1 and RDN37-2. Due to the high sequence similarity between the two, most rRNA species map to both. Finally, a smaller proportion were too short to map altogether (Figure 4.4). While a high number of reads were obtained for the 4E IP sample at 9.7 million, the input samples were lower at 2.6 and 0.8 million. While the number of reads were not insignificant, when spread out over the genome, the coverage, especially in input samples, was low. When coverage files were viewed in IGV, the coverages in the two input tracks were inconsistent and variable within genes. While this

may not be a significant issue for standard differential gene expression analysis, we required coverage tracks with enough coverage across the entire gene body in order to compare with the IP signal enrichments at the same locations. Therefore, it was concluded that due to the high rRNA content, the number of reads uniquely mapping to the genome was too low for meaningful transcriptome-wide interpretation of closed-loop formation. It was therefore decided a new library preparation approach was required to specifically deplete yeast rRNA.

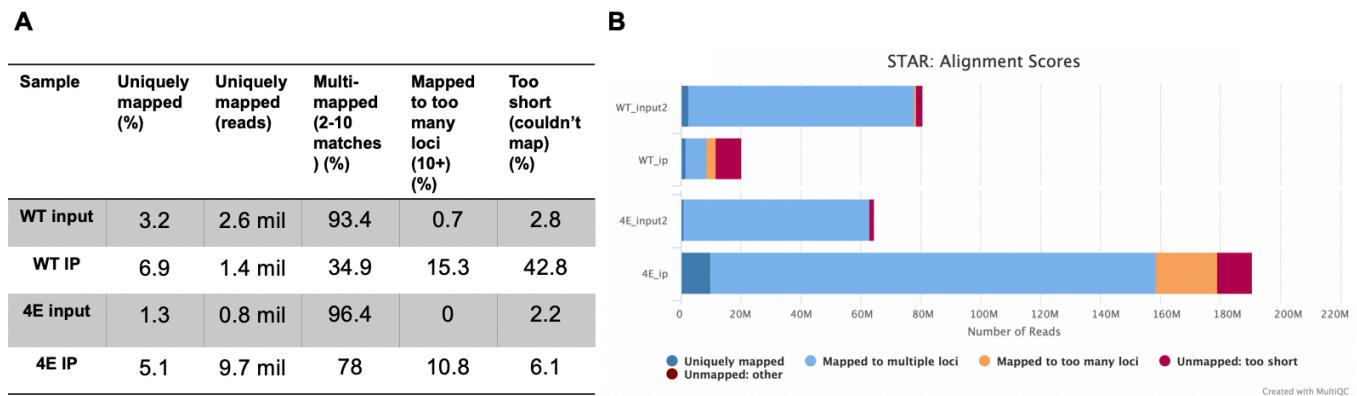


Figure 4.4. Mapping statistics

Overall mapping statistics for the samples sequenced. STAR was used to map reads to the genome and final mapping statistics were collated by MultiQC. WT refers to the untagged wild-type sample, 4E refers to libraries made from the eIF4E-TAP, eIF4G1-3xFLAG strain. Input refers to whole cell lysate (WCL) digested in parallel with the immunoprecipitation reaction (IP).

4.1.2 New library preparation approach for RNA-seq

In order to specifically deplete rRNA from the RNA samples collected, a library preparation method initially developed for translation complex profiling (TCP-seq) was utilised (Archer *et al.*, 2016). This method previously developed in the research group had the advantage of already incorporating an rRNA depletion step with a yeast rRNA probe set. It also had the advantage of being designed specifically for use with fragmented or digested RNA. However, this library preparation is much lengthier and involved than a standard Illumina library preparation protocol, with multiple overnight RNA or DNA precipitation steps and a total protocol length of 3-4 days compared to the one-day protocol typically required. It also requires much more starting material (preferably >1 µg per sample, Archer *et al.* (2016)) compared to the lower amounts possible with other library preparation methods. Although this was initially not the preferred library preparation method for these reasons, due to the specific requirements of this study, it was deemed the most suitable method available.

This method starts with PNK-mediated end-repair to convert the 2', 3'-cyclic phosphate at the end of digested RNAs to 3'-OH ends capable of subsequent poly(A) tailing by 3' EPAP polyadenylation enzyme. A split adapter containing the Illumina sequencing adapters, a carbon spacer and an anchored poly(T)₂₀ primer, binds to the poly(A)₂₀ stretch of the RNAs directed after the start of the original RNA sequence (due to the TVN anchor). The RNA is then reverse transcribed and the primer extended with the reverse-transcribed cDNA is circularised, resulting in single-stranded, circularised DNA molecules. Probes corresponding to yeast rRNA are hybridised to the circularised molecules containing rRNA sequences, and a double-strand nuclease is used to specifically digest these molecules. The remaining circular cDNA molecules are amplified with sequencing adapters containing hexanucleotide barcodes, then size-selected on an agarose gel to excise libraries of the desired size (Figure 4.5).

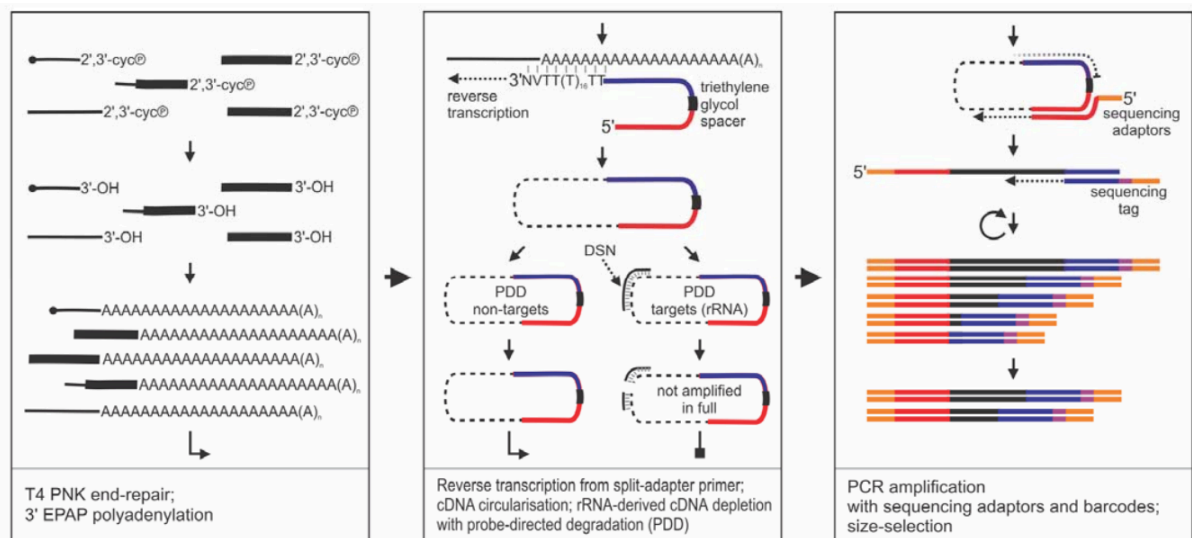


Figure 4.5. Overview of final library preparation approach taken

Figure was taken from Archer et al., 2016, supplementary materials. T4 PNK = T4 Polynucleotide Kinase; EPAP = E. coli Poly(A) Polymerase; DSN = Double Strand Nuclease; PDD = probe directed degradation

4.1.2.1 Final libraries for sequencing

Libraries were successfully obtained by following the established library preparation protocol (Figure 4.6), after adjustments were made for increased material required and changed availability of reagents. As anticipated, very little material was visible in the final library for WT IP samples as demonstrated by lack of a nucleic acid smear in the corresponding lane, consistent with this sample representing an untagged binding control (Figure 4.6a, lanes 4,8

and see also b). There was also a difference in library size distribution in samples pre- and post DSN-mediated rRNA depletion, with a shift to smaller library fragment sizes in the target range of 200 bp – 500 bp. With the empty adapter being 140bp (Figure 4.6a, last lane), this corresponds to inserts in the range of ~60-360 bp. For each sample, material in the range of 200 – 500bp was excised from the gel (Figure 4.6c), extracted and sent for sequencing by an Illumina NovaSeq600 platform sequencing provider.

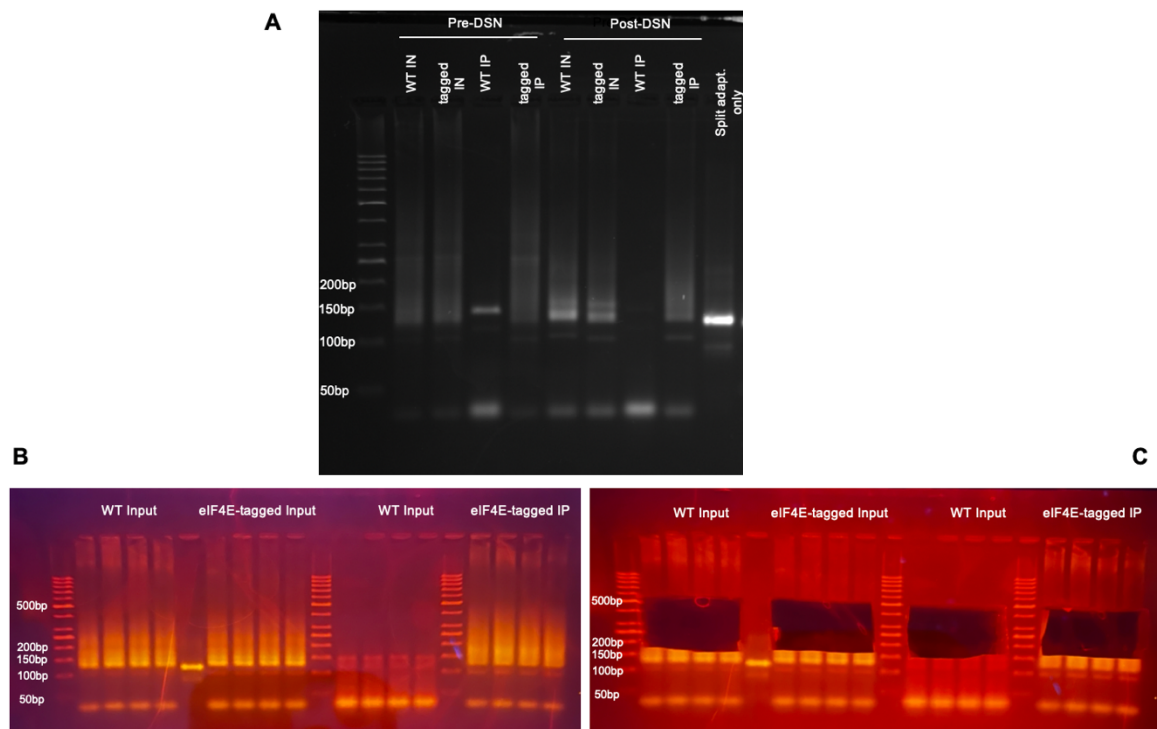


Figure 4.6. Final libraries obtained for sequencing

A) Final sequencing libraries pre- and post-double strand nuclease (DSN) mediated rRNA depletion. Final libraries were separated by size on a 2% agarose gel and imaged with the GelDoc system. B) and C) Final libraries (B) were separated by size on a 2% agarose gel and visualised with the SafelImager blue light transilluminator, then excised (C) in the region between 200 and 500 bp. This region was extracted and sent for sequencing. WT refers to the untagged wild-type sample, eIF4E-tagged or tagged refers to libraries made from the eIF4E-TAP, eIF4G1-3xFLAG strain. Input refers to whole cell lysate (WCL) digested in parallel with the immunoprecipitation reaction (IP).

4.1.2.2 Initial mapping and filtering workflow

150 bp paired-end short read sequencing was performed on the NovaSeq600 platform with a target of 100 million reads per sample. The initial mapping and filtering workflow broadly followed the same principles as previously described in section 4.1.1 of trimming low quality

sequences, removing adapters, mapping to the genome and visualisation. However, a few addition steps were introduced, namely trimming of the artificial poly(A) tail introduced during library preparation, filtering out of rRNA, tRNA and structural RNA sequences and normalisation for library size. Trimming of the poly(A) tail was introduced in order to improve mapping to the genome, as the string of poly(A)₂₀ was expected to be on all fragments due to the library preparation method. If not removed, this stretch of 20 As would prevent mapping of these sequences to the genome annotation. Filtering of the rRNA, tRNA and structural RNA step was performed by mapping the whole sequencing library to datasets of known yeast rRNA, tRNA or structural RNA sequences and filtering out matches. Normalisation for library size was performed for each sample by dividing the coverage at each genomic location by the total coverage for that sample, then multiplication by 10⁸ for visualisation. In this way, the signal at each position in the final coverage files represents the relative proportion of the library made up of reads from that location.

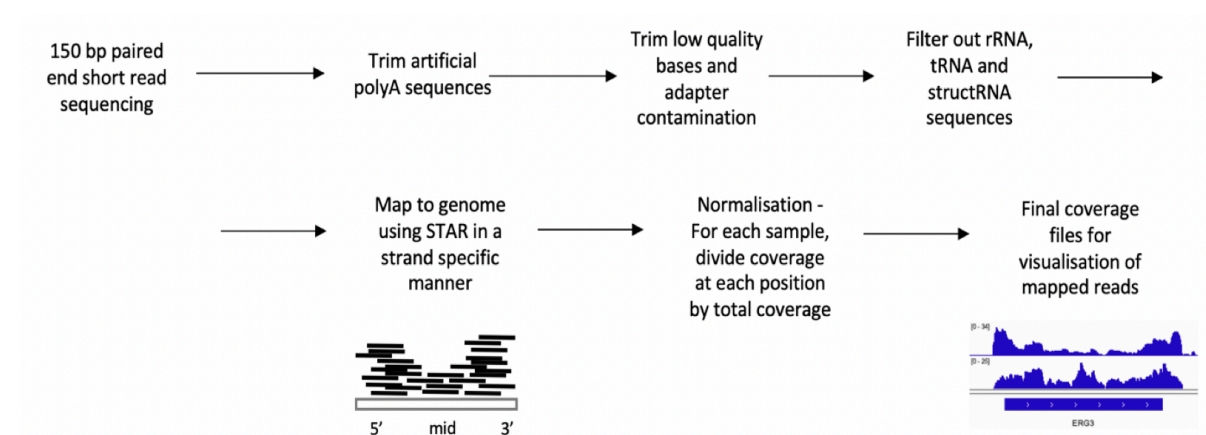


Figure 4.7. Overview of sequencing pre-processing steps

4.1.2.3 Mapping and filtering statistics

Overall, the number of reads mapping to the genome vastly improved using this library preparation method. Using the previous library preparation method, between 1.3 – 6.9% of reads were mapping to the genome after trimming (Figure 4.4). However, with the new library preparation method to deplete rRNA, the percentage of trimmed reads mapping to the genome increased to 13% for WT IP, 25% for both inputs, and 47% for the tagged IP (average of three replicates) (Figure 4.8d,e). This increase in proportion of ‘useful’ reads is even greater when considering the lack of rRNA, tRNA and structural RNA removal in the earlier pre-processing pipeline. The proportion of rRNA reads with this new library

preparation method also decreased overall, to between 20-60% of all trimmed reads for all replicates across all samples (Figure 4.8a). Percentages of reads aligning to each category was similar between the tagged and untagged inputs, suggesting an overall similarity in the composition of these samples. Compared to the inputs, the tagged IP also displayed a different distribution of reads, with a higher percentage mapping to the genome and a much smaller presence of tRNAs and structural RNA species, consistent with this sample's selective enrichment of mRNAs associating with the closed-loop (Figure 4.8a).

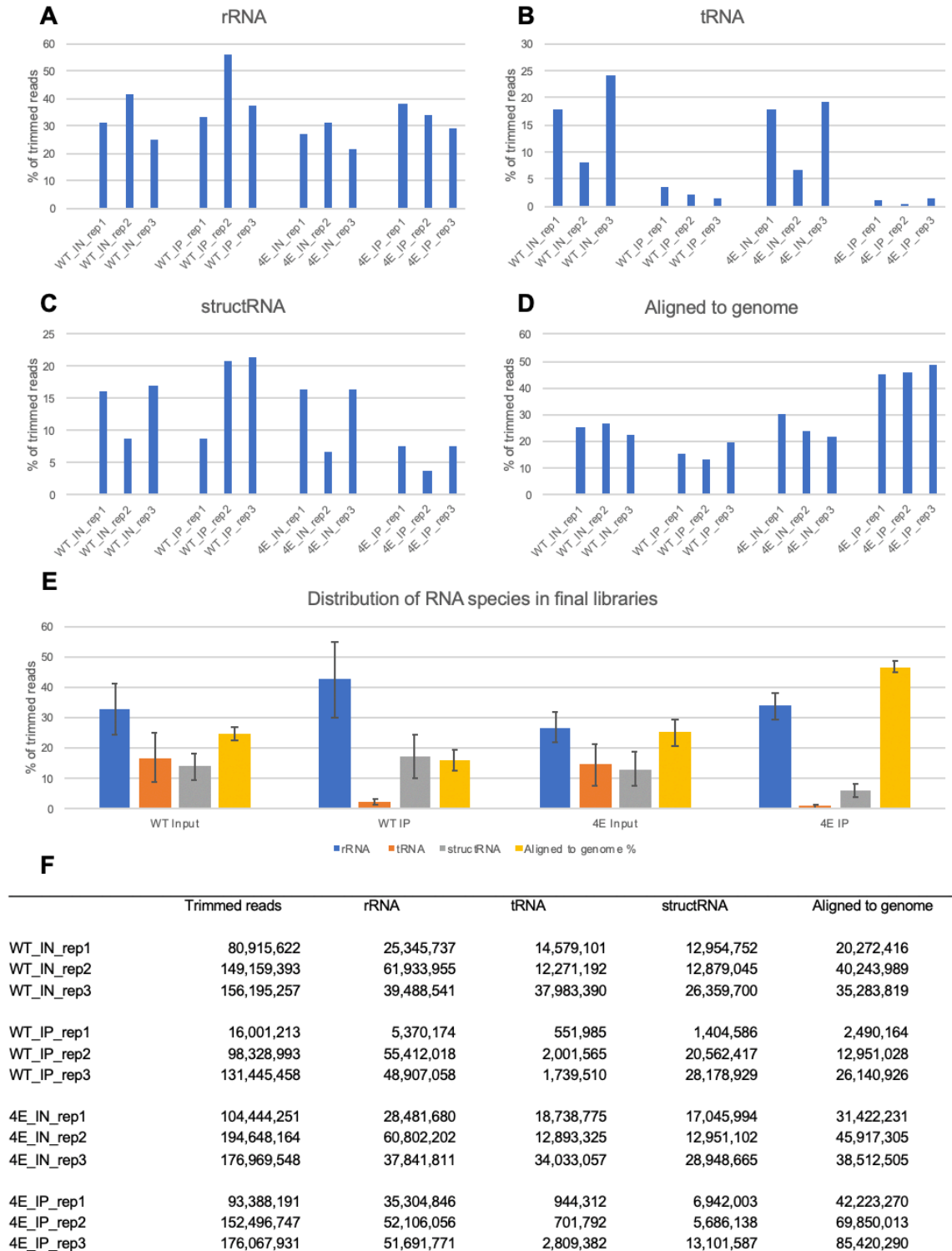


Figure 4.8. Number and distribution of reads mapping to different RNA species

Reads remaining after trimming were sequentially matched to a dataset of yeast rRNA, tRNA and structural RNA sequences, followed by mapping to the genome using STAR. The percentages of reads mapping to rRNA (A), tRNA (B), structural RNA (C) or the genome (D) for each sample was calculated as a percentage of total trimmed reads. (E) The average percentage of reads mapping to each category was calculated from three replicates each for the four samples WT input, 4E input, WT IP and 4E IP. Error bars denote mean $\pm$ SD. (F) Raw read counts for each category are provided as reference. WT refers to the untaged wild-type sample,

eIF4E-tagged or tagged refers to libraries made from the eIF4E-TAP, eIF4G1-3xFLAG strain. Input refers to whole cell lysate (WCL) digested in parallel with the immunoprecipitation reaction (IP).

4.2 Defining transcription units *de novo*

Given the nature of the closed-loop in contacting the mRNA cap and poly(A) tail, our interest was predominantly in the mRNA 5' and 3' extremities, where we expect there to be an enrichment of signal (Figure 4.9). Therefore, an accurate annotation of the mRNA 5' and 3' UTRs was of particular importance for our downstream analyses. The most comprehensive collection of yeast 5' and 3' UTR annotations as defined by RNA-seq has been compiled by Nagalakshmi *et al.* (2008). This study used a combination of RNA-seq and 5'-RACE to map the 5' and 3' UTRs of the majority of the yeast transcriptome. However, for some genes, only the 5' UTR annotation or only the 3' UTR annotation was determined. This annotation was also based on the BY4741 strain, while our study was conducted in the w303-1A strain. As the yeast transcriptome is known to display heterogeneity in transcript start site selection and have alternate polyadenylation based on a variety of signals (Rojas-Duran and Gilbert, 2012; Pelechano, Wei and Steinmetz, 2013; Kang *et al.*, 2015), it was decided that for the most accurate annotation applicable for our purposes, it would be best to define the transcription unit *de novo* for our conditions. In this way, we would be able to directly compare our IP end signals at the most applicable locations for each gene.

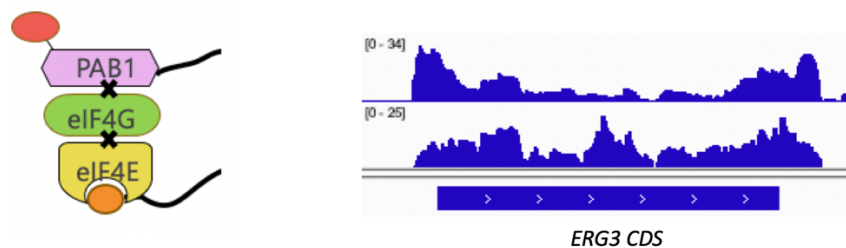


Figure 4.9. The focus of the study is on sequences at the starts and ends of mRNAs

As the experimental method is designed to deplete RNA-seq signal in the middle portion and enrich it at the 5' and 3' ends (left), the annotation of the mRNA extremities is particularly important. However, while the CDS annotation for yeast is well-established, much of the important signal to be analysed lies outside of the annotated CDS region (right).

4.2.1 Collating list of 'high confidence' examples

Before we began to define the transcription units, we first sought to establish a reference list of 'high confidence' examples that could be viewed in IGV to validate our *de novo*

predictions. For this purpose, I visually scanned the whole genome in IGV for examples of genes with a prominent high, medium or low 3' end enrichment compared to the 5' end. This resulted in a list of 72 genes with near equal end enrichments, a list of 94 genes with medium levels of 3' end enrichment compared to the 5' end and a list of 204 genes with very little 3' enrichment. Some examples are given in Figure 4.10. As the following *de novo* transcript unit definition strategy was developed, these examples were used to check appropriateness of thresholds and whether other factors needed to be taken into consideration.

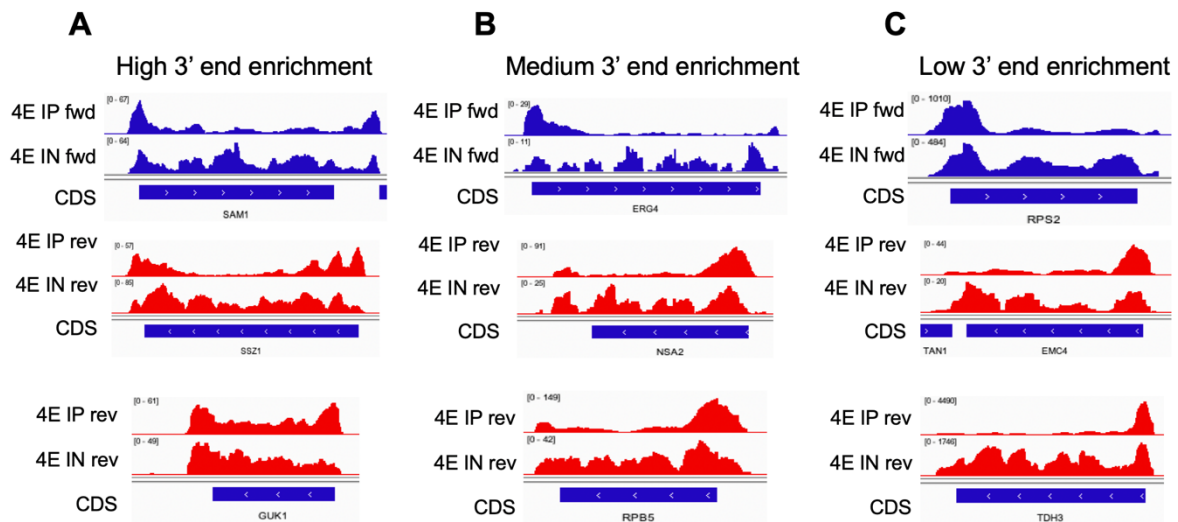


Figure 4.10. List of 'high confidence' individual examples

The yeast genome was visually scanned chromosome by chromosome in IGV to collect examples of high (A), medium (B) or low (C) 3' end enrichment (relative to 5' end) genes. Blue represents forward strand examples (read L-R) while red represents reverse strand examples (read R-L). Numbers represent the signal range of that track. For each example, the top lane represents the signal from the immunoprecipitation (IP) sample, the middle lane represents the signal from the input (IN) sample, and the thick blue bars represent coding sequence (CDS) annotations.

4.2.2 Development of workflow

The overall strategy was to define 'signal islands' in the IP and input samples – continuous stretches of signal above a certain threshold, then overlap these signal islands with the CDS annotation for a merged transcription unit prediction (Figure 4.11). Both IP and input tracks were used for this calculation, as the end boundaries in the IP tracks were often more sharply defined than in the input tracks (see Figure 4.9, Figure 4.10). While the inputs were expected to have even coverage from the start of the 5' UTR to the end of the 3' UTR, this was often not the case. In genes with lower coverage, the input signal was sometimes discontinuous and tapered off towards the ends, while the IP signal was the opposite – high at the ends but

patchy in the middle. Therefore, both were used, together with the CDS annotation. The actual decisions regarding thresholds were made iteratively, with different values set bioinformatically then tested against the high confidence examples for accuracy. This process was repeated until the final transcription unit prediction matched the coverage observed in IGV as accurately as possible for the high confidence examples. Some more in-depth examples of this iterative process are explained in the following sections.

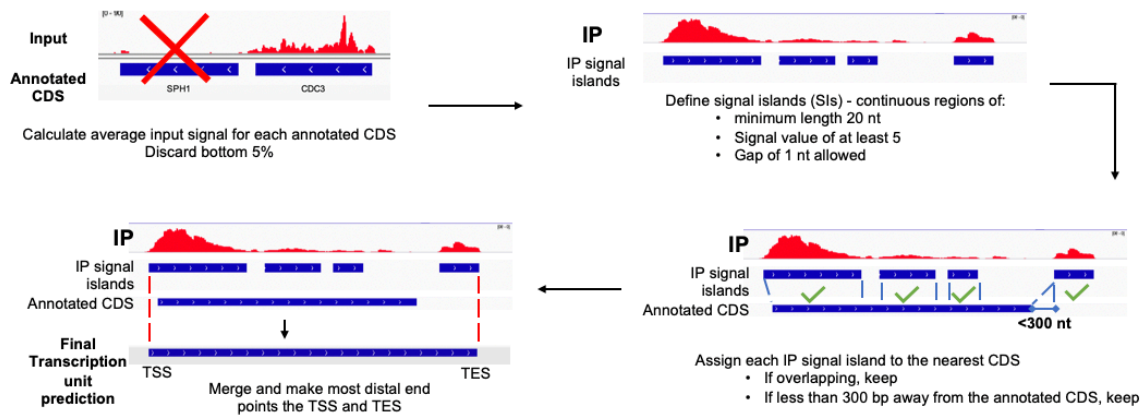


Figure 4.11. Overview of *de novo* transcription unit prediction approach

TSS = transcription start site; *TES* = transcription end site

4.2.2.1 Defining thresholds for signal islands

A signal island was defined as a continuous stretch of at least **X** length with a signal above **Y**. A gap of 1 nt being below the threshold was allowed at a time. At first, a minimum length (**X**) of 50 nt was tested with signal values of 1, 5 and 10 (**Y**). In other words, a signal island would be recognised in either the IP or input coverage profiles if there was a 50 nt+ stretch of signal with all values consistently above 1, 5, or 10. Comparing signal thresholds of 1, 5 and 10 with a minimum length of 50 nt, it can be observed that all three generally capture the whole transcription unit, with the signal islands ending where the reads are seen to drop off (Figure 4.12). While a signal threshold of 1 most accurately covers the entire region of reads in all 4 examples, it is also non-specific, resulting in extra signal islands outside of the likely transcription unit. For example, in HHF1 (bottom right), with a threshold of 1, multiple signal islands can be seen outside of where the HHF1 signal appears to begin and end. This was prominent particularly around highly expressed examples, where some level of background around the main signal body leads to these spurious signal islands being created. On the other end of the scale, a signal threshold of 10 results in very few ‘false positive’ signal islands, but

also results in some regions of low signals not being recognised (e.g. MET6). A threshold of 5 therefore appeared to be the best compromise between sensitivity and specificity (Figure 4.12).

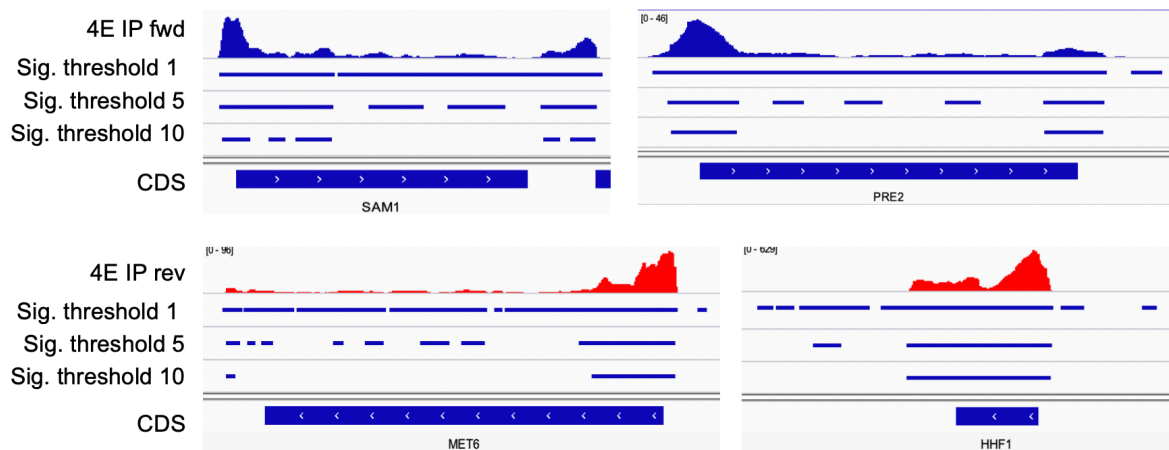


Figure 4.12. Comparing signal threshold values for defining signal islands

Continuous 50 nt stretches of signal of at least 1, 5 or 10 are compared for signal island calculation. The IP coverage track for example genes are shown in the first rows, with blue corresponding to the forward (fwd) strand and red to the reverse (rev) strand. Reverse strands should be read R-L. Thin blue lines refer to signal islands detected with that threshold value and a minimum length requirement of 50 nt. The CDS annotation is shown by the thick blue bar.

However, while a threshold of 5 was a good balance between sensitivity and specificity, there were still some regions that were not recognised as signal islands within the overall signal body. The minimum length required was therefore decreased from 50 nt of continuous signal above 5, to be a minimum length of 20 nt of continuous signal above 5. This resulted in better detection of shorter signals, particularly at the 3' ends (Figure 4.13).

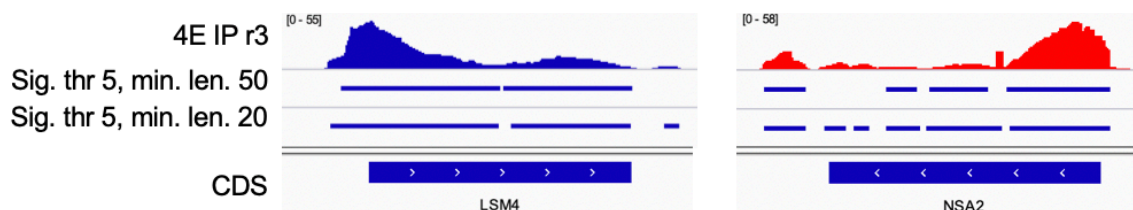


Figure 4.13. Comparison between a minimum length of 20 and 50 with a signal threshold of 5

Continuous 50 or 20 nt stretches of signal above 5 are compared for suitability as signal islands. The IP coverage track for example genes are shown in the first row, with blue corresponding to the forward strand and red to the reverse strand. Reverse strands should be read R-L. Thin blue lines refer to signal islands detected with that threshold value and a minimum length (min. len.) requirement of 50 or 20 nt. The CDS annotation is shown by the thick blue bar.

4.2.2.2 Merging of IP and input signal islands with CDS

Once all signal islands were determined, they were each assigned to their nearest CDS annotation. First, if a signal island from the input was overlapping fully or partially with the coordinates of a CDS annotation, the signal was automatically assigned to that CDS.

However, due to the sometimes-low signals in input samples, and due to often-seen tapering of input signal down towards the ends of mRNAs, these input signal islands often did not extend out far enough to overlap with the IP signals at the very ends, suggesting a need for including the signal islands from the IP as well into the final transcription unit calculation.

When assigning signal islands for the IP tracks, an allowance was made for signal islands further from the CDS. As the IP samples had been depleted of the majority of middle signals due to RNase I digestion and immunoprecipitation, the signal island corresponding to the ends of the signal often did not run into the CDS continuously enough to overlap with the CDS annotation (Figure 4.11). Therefore, a maximum distance value needed to be determined for a signal island to still be counted towards a nearby CDS. To determine this maximum value, the file of 5' UTR and 3' UTR lengths from Nagalakshmi *et al.* (2008) was used to determine the distribution of 5' UTR and 3' UTR lengths in yeast from the available annotations. Although the median lengths were fairly short at 50 bp and 104 bp for 5' UTRs and 3' UTRs respectively, a longer tail of lengths did exist past these medians, with the 95th percentile being 243 bp and 327 bp for 5' UTRs and 3' UTRs respectively (Figure 4.14). As the issue with the IP end signal not overlapping with the CDS was most prominent for genes with long UTRs, we chose to be more lenient in the maximum distance allowed cut off, so we tested values between 200 – 400 bp, taking into account the 95th percentile length values. After a few rounds of iterative testing of different lengths, it was determined that a maximum distance of 300 nt from the end of the CDS to the start of the signal island would be allowed, for that signal island to still be assigned to the CDS. Although this value is longer than the median 5' or 3' UTR, this generally was not an issue, as each signal island was first assigned to its overlapping or nearest CDS, so signal islands from nearby genes were often already assigned and did not get assigned twice to create UTRs that were artificially too long.

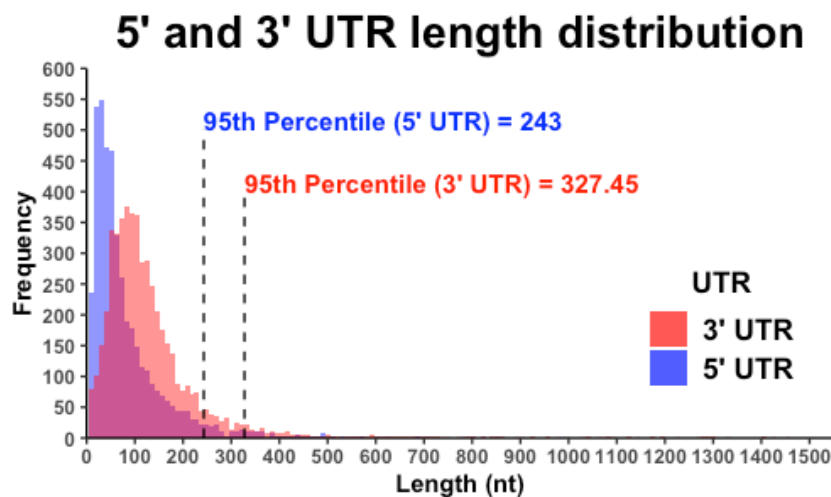


Figure 4.14. 5' and 3' UTR length distribution from Nagalakshmi et al. (2008)

5' and 3' UTR lengths from Nagalakshmi et al. (2008) were plotted as a histogram of counts at each length using R.

The genomic coordinates of these assigned IP signal islands were then merged with the genomic coordinates of the input signal islands, and finally with the genomic coordinates of the annotated CDS (Figure 4.11). The outermost coordinates were newly defined as the transcription start site and transcription end site for that gene. These transcription unit predictions were visually checked against the coverage profiles of the high confidence example gene set to ensure accurate mapping of 5' and 3' boundaries.

Finally, as we were working with genomic co-ordinates, it was important to remove intronic sequences. Therefore, the genomic coordinates of the introns were subtracted from the final transcription unit genomic coordinates, to result in transcription units without intronic stretches (Figure 4.15). This was to avoid calculations being made over intronic regions in downstream analyses.

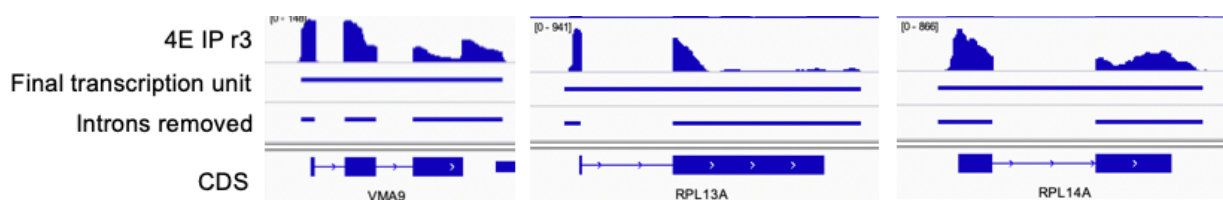


Figure 4.15. Removal of introns from transcription unit predictions

The IP coverage track for example genes are shown in the first row, with blue corresponding to the forward strand. The continuous thin blue lines in row two refer to the final transcription units predicted, before intron removal. The third row depicts the same transcription units with introns removed. The last row displays the original annotation, with introns shown by thin blue lines with arrows.

4.2.3 Validation of de novo transcription unit predictions

As a check for the validity of our transcription unit prediction method, I compared the length distribution of our predicted transcript annotations with the length distribution of known CDS annotations. A slight shift to longer lengths upon transcript unit prediction can be seen when comparing the annotated CDS (red) to the new transcript unit predictions (blue) (Figure 4.16). As the median lengths of 5' and 3' UTRs, as determined by Nagalakshmi *et al.* (2008), were 50 bp and 104 bp respectively, only a small shift would be expected given the scale of the x-axis. Therefore, it was concluded the predictions had extended the original CDS lengths within an appropriate and expected range.

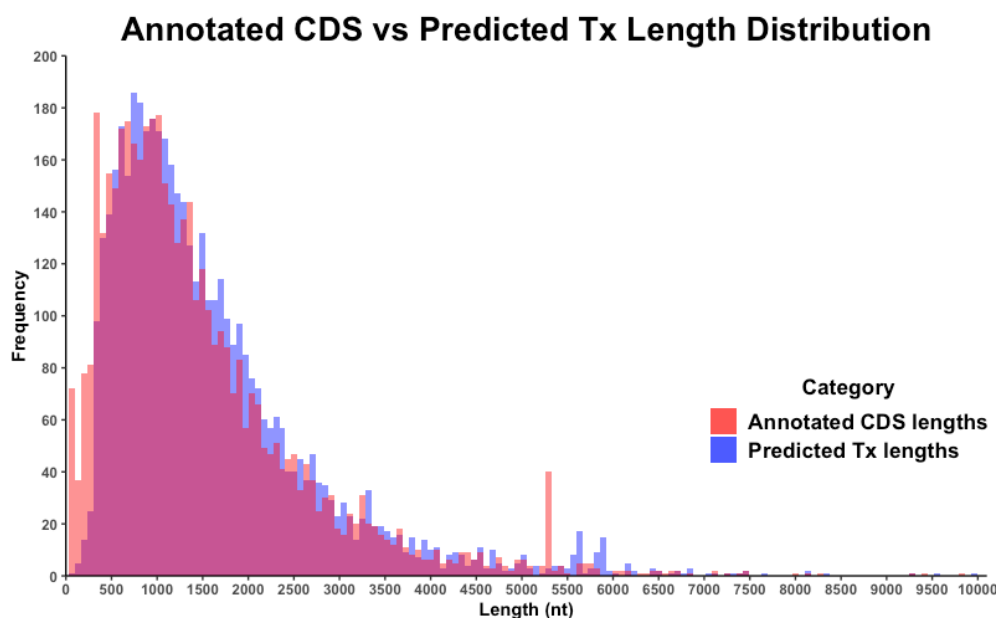


Figure 4.16. Lengths of predicted transcription units compared with CDS lengths

CDS lengths were calculated by taking the absolute value of the difference between CDS start and end coordinates for each gene, as downloaded from yeastgenome.org. Transcription units (Tx) were predicted *de novo* as described in previous sections. The number of genes (y-axis) of a certain length (x-axis) for either the CDS annotation (red) or predicted transcription unit (blue) are shown.

Our predicted transcription unit lengths were also compared to those calculated by Nagalakshmi *et al.* (2008). The genomic coordinates for the start of the 5' UTR and the end of the 3' UTR as mapped by Nagalakshmi *et al.* (2008) were subtracted, to obtain a list of transcription unit lengths for genes containing both annotations. These were compared to the transcription unit lengths obtained by us (Figure 4.17). It can be observed that from approximately 1000 bp and larger, our annotation and the Nagalakshmi *et al.* (2008) annotation are broadly in agreement and the overall distribution of lengths appears very

similar. Our annotation also predicted more transcription unit annotations in the range of 250-1000 bp. Taken together, this suggests that our transcription unit predictions are largely consistent with the existing literature, while also increasing the number of annotations for our purposes.

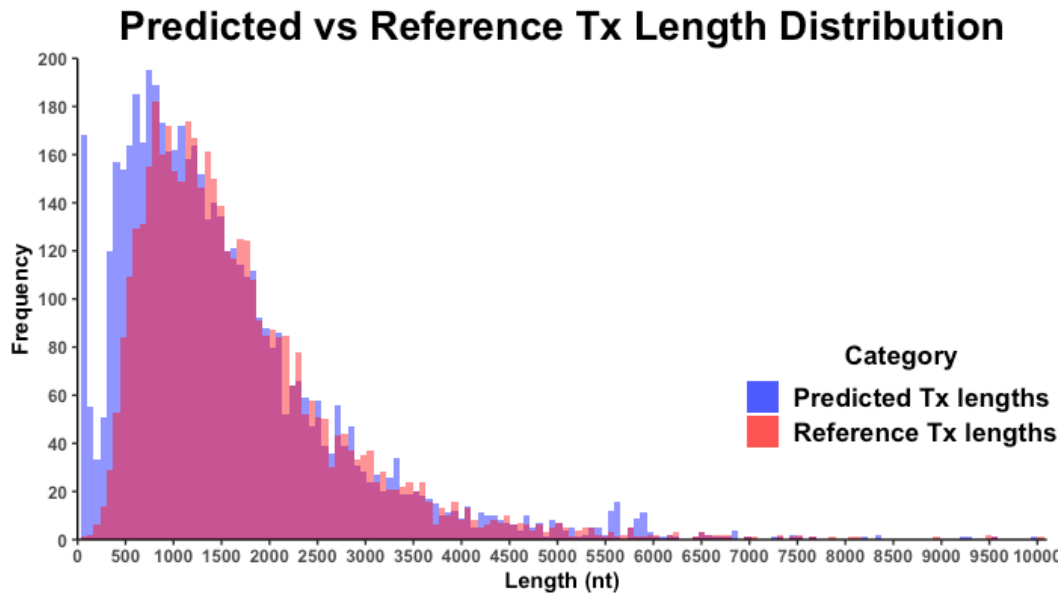


Figure 4.17. Comparison of predicted vs reference transcription unit lengths

Reference transcription unit (Tx) lengths (red) were calculated by taking the absolute value of the difference between 5' UTR start and 3' UTR end genomic coordinates for each gene, as reported by Nagalakshmi et al. (2008). Only genes with both 5' and 3' UTR annotations were considered. Predicted transcription units (blue) were determined de novo as described in previous sections. The number of genes (y-axis) of a certain length (x-axis) for either the reference annotation (red) or predicted transcription unit (blue) are shown.

4.3 Reproducibility of replicates

As the transcription unit predictions had only been derived from one replicate (replicate 1), we performed the whole workflow twice more. These two replicates consisted of the full set of samples – WT input, eIF4E input, WT IP and eIF4E IP. These two additional replicates, though starting from the same freezer yeast strain stocks, were refreshed on separate agar plates, grown to mid-logarithmic phase, crosslinked and harvested all on independently occasions from one another. Immunoprecipitation and library preparation were performed independently, while sequencing for replicate 2 and 3 were performed at the same time using unique de-multiplexing barcodes. The spearman's correlation between replicates was determined on the transcription units defined – i.e. for the genomic regions defined from

replicate 1 as being the transcription unit annotation, the signal distribution of all replicates was tested for correlation with each other.

Overall, the corresponding samples from all three replicates were similar to each other, with high spearman correlation values when compared to one another. Clustering of the samples by relative similarity also places the same condition from the different replicates close together (Figure 4.18). For example, the three tagged IP samples (4E IP r1-r3) are found adjacent to each other in this clustering and display very high similarity to each other.

Although replicate 1 of the tagged input (4E IP r1) is not clustered as closely to the other two replicates (4E IP r2, r3), it does cluster closely to two untagged inputs (WT IN r1, r2).

Similarly, while the third replicate of WT/untagged input does not cluster as closely to its replicate 1 and 2, it is very similar to two tagged inputs (4E IN r2,3). These two groups of three input samples also show high similarity to each other, as depicted by the dark orange – red. This suggests that broadly, there is a high level of similarity between all input samples, regardless of whether they are from the tagged or untagged strain. This is promising, as it indicates there are no strong inherent expression profile biases between the two strains. Two of the untagged IP samples (WT IP r1, r3) also cluster in between the 6 inputs, consistent with the expectation that the untagged IP is representative of background binding, so should be a reflection of input sequence distributions. The only sample that did not cluster strongly with any other sample was the second untagged IP sample (WT r2), showing very little correlation to any other sample, including its own replicates. Examination of this sample's coverage profile in IGV confirms this general difference in profile. This sample showed greater signs of overamplification of some regions and patchy signal otherwise. However, given the expectation for the WT IP sample to have very little binding, this was not unexpected, and likely represents a particularly low-background negative sample that has been overamplified when treated and sequenced in parallel with samples containing more material.

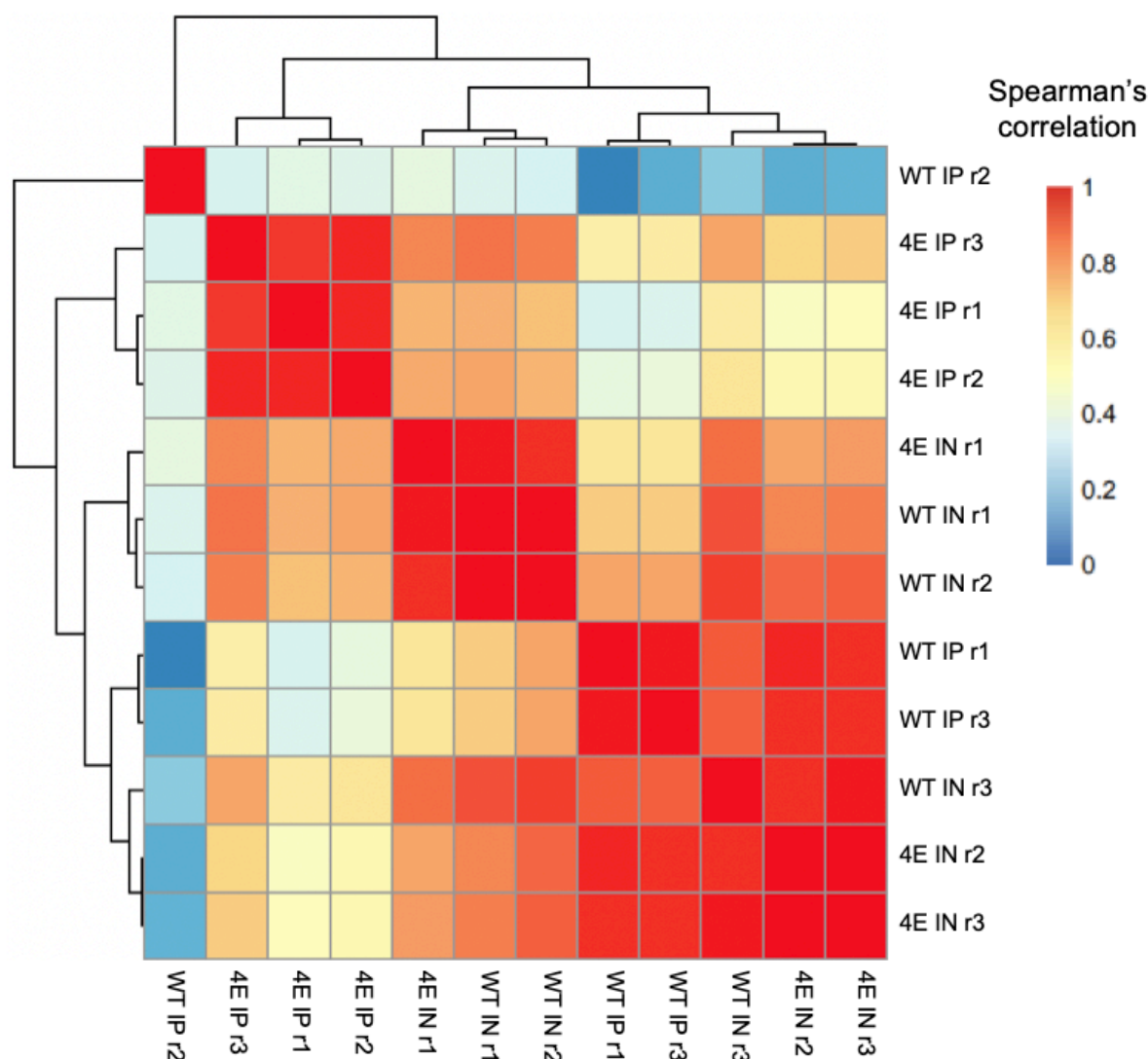


Figure 4.18. Clustered spearman's correlation matrix of samples between replicates

The signal distributions over the transcription units defined using replicate 1 were analysed for all replicates and correlated with each other. The spearman's correlation between all conditions and replicates were determined, and individual samples were clustered with each other in a matrix depending on similarity. Increasing red represents higher levels of similarity, while increasing blue represents lower levels of similarity. WT refers to the untagged strain, 4E to the eIF4E-TAP, eIF4G1-3xFLAG strain. IN refers to input, IP refers to the immunoprecipitated sample. r1, r2 and r3 refer to replicates 1, 2 and 3 respectively.

4.3.1 Reproducibility between replicates

Looking at the individual correlations more closely, the spread of points is also reflective of differences between the samples. Overall, all untagged inputs were similar, with spearman's correlation values between 0.87-0.94. Similarly, all tagged inputs were similar to each other, with values between 0.78-0.95. For both sets of inputs, replicate 1 and 3 were the most strongly correlated, while replicate 2 was slightly less similar to the others (Figure 4.19a).

The lack of a strong correlation between untagged IP replicates is also evident, with the larger spread of individual points in the correlation, and a spearman's correlation value of between 0.41-0.56 (Figure 4.19b). As mentioned in the previous section, this was not surprising, as we expected this condition to represent random background binding, and with less material overall, the similarity between these samples was not expected to be high. The tagged pulldown, on the other, showed a strong correlation between replicates, with a spearman's value between 0.93-0.96, again with a slightly higher similarity between replicate 1 and 3 compared to 2 (Figure 4.19b, row 2).

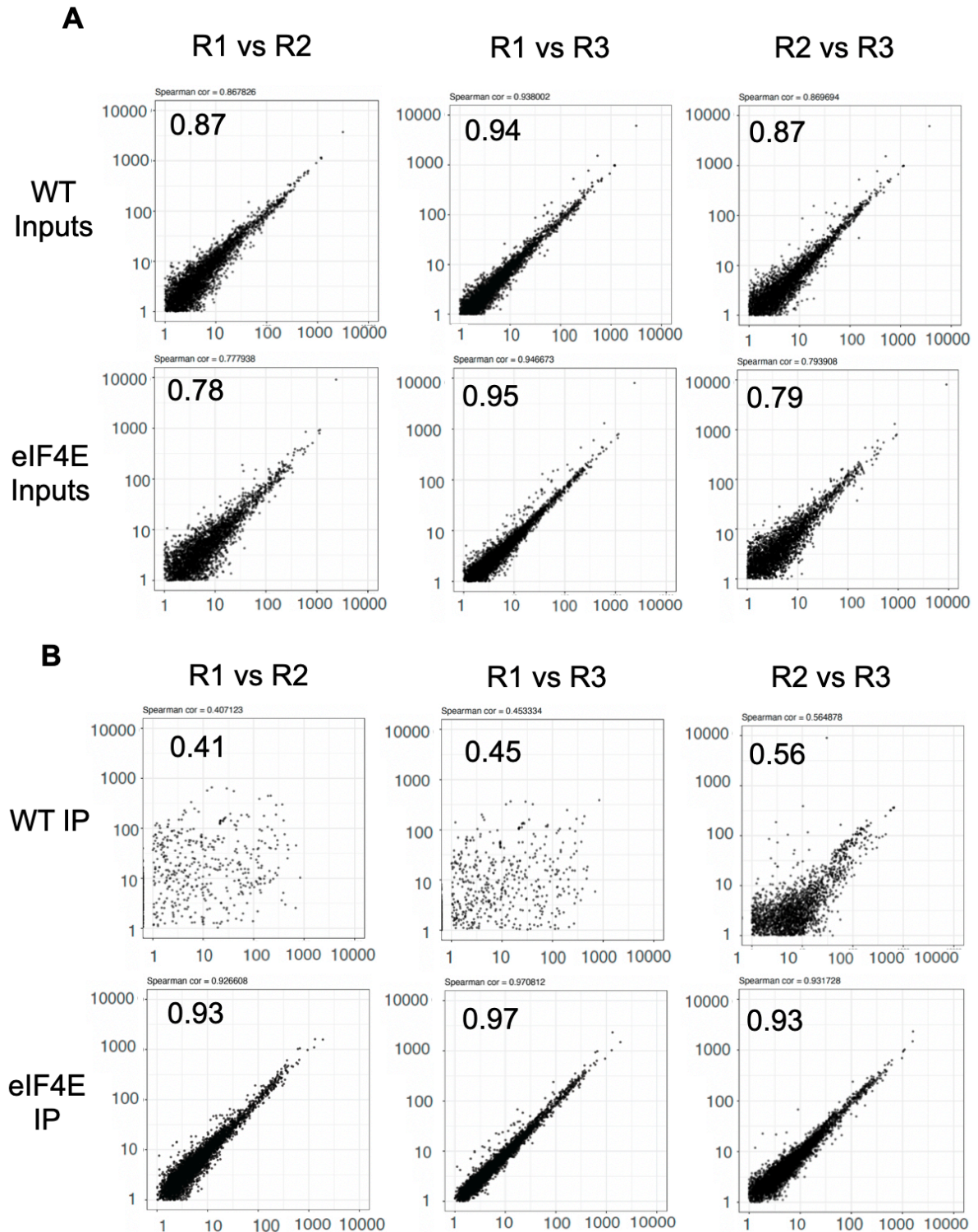


Figure 4.19. Individual correlations between replicates

The signal distributions over the transcription units defined using replicate 1 were analysed for all replicates and correlated with each other. The spearman's correlation between two replicates of the same condition was determined (top left corners), with the x axis representing the first replicate in the label and the y axis representing the second. WT refers to the untagged strain, 4E to the eIF4E-TAP, eIF4G1-3xFLAG strain. IN refers to input (A), IP refers to the immunoprecipitated sample (B). R1, R2 and R3 refer to replicates 1, 2 and 3 respectively.

4.3.2 Comparison within replicates

The individual correlations are also interesting to inspect in the context of comparing different conditions to each other within the same replicate. This serves as an indication that the experimental method resulted in sufficiently distinct or similar samples. For all replicates, the tagged and untagged inputs are similar to each other, with spearman correlation values of between 0.73 – 0.93 (Figure 4.20). This suggests that the two strains are comparable in terms of overall signal distribution and expression and there is little inherent bias between them. When the tagged and untagged IPs are compared, the similarity is much lower for all replicates, with a spearman correlation value of between 0.42-0.56. This is promising, as it is an indication that the enrichment profiles observed in the tagged IP are distinct from background binding. Overall, the similarities and differences between conditions presented are consistent with expectations for this experimental set up.

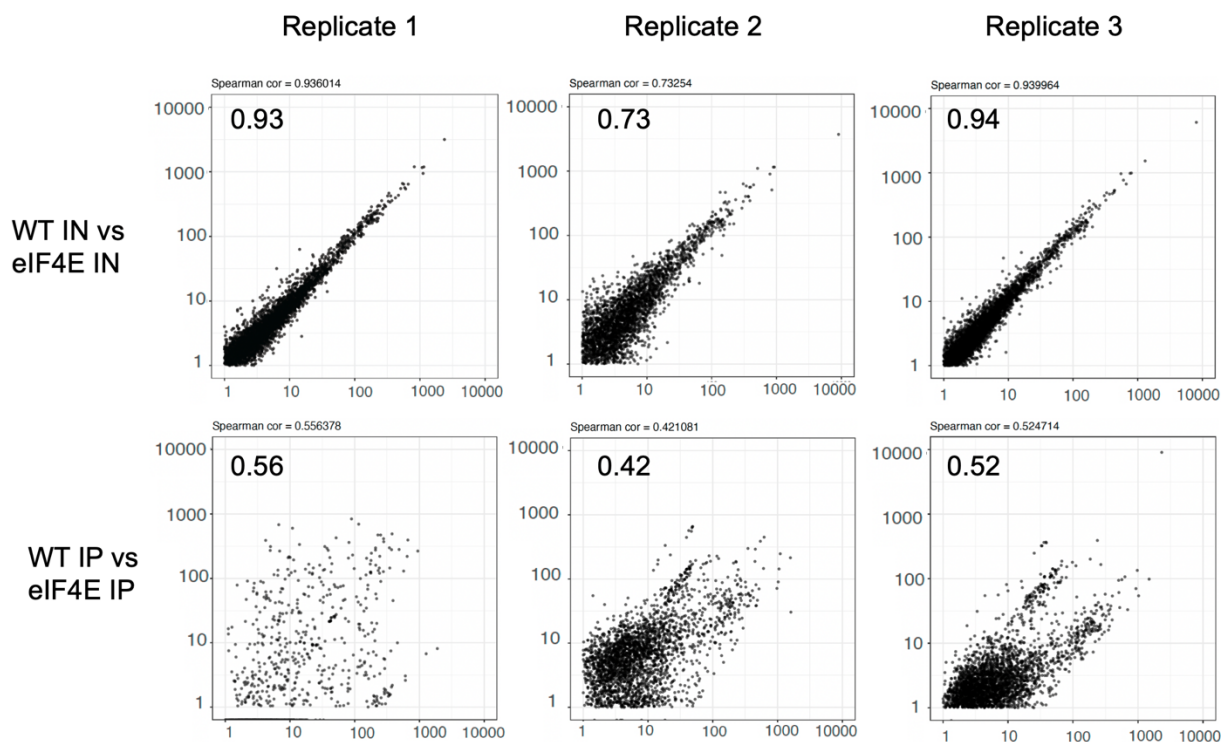


Figure 4.20. Comparison of samples within replicates

The signal distributions over the transcription units defined using replicate 1 were analysed for all samples and correlated with each other. The spearman's correlation between two conditions of the same replicate was determined (top left corners), with the x axis representing the tagged sample and the y axis representing the untagged sample. WT refers to the untagged strain, 4E to the eIF4E-TAP, eIF4G1-3xFLAG strain. IN refers to input, IP refers to the immunoprecipitated sample.

4.4 Chapter 4 conclusions

In this chapter, I presented the development of a library preparation and RNA-seq pre-processing pipeline to prepare the sequencing data for downstream interpretation.

I detail the two library preparation approaches, one with and one without yeast-specific rRNA depletion. While the first library preparation method – a standard strand-specific Illumina library protocol – required less starting material and resulted in reads of good sequencing quality, the lack of yeast-specific rRNA depletion resulted in a library containing majority rRNA sequences, with very few useful mRNA reads. The second preparation method – developed originally to prepare RNA-seq libraries from TCP-seq material (Archer *et al.*, 2016) – though more complex and requiring more starting material, resulted in much better depletion of yeast rRNA and a subsequent increase in useful mRNA reads.

I also present our approach to circumvent the lack of complete 5' and 3' UTR annotations in yeast. In order to obtain accurate start and end boundaries for downstream analysis of the mRNA ends, we defined signal islands from the IP and input coverage tracks and merged them with existing CDS annotations. These predictions were validated visually for a set of 'high confidence' genes with well-defined but varying 5' and 3' enrichments in the IP sample. These predicted lengths were compared to the original CDS annotation ensure a lengthening of signal, and also compared to the reference transcript unit annotations to ensure concordance.

Finally, the reproducibility of the replicates was also examined. Overall, replicates showed high levels of similarity to each other, demonstrating the high reproducibility of the experimental method, library preparation and sequencing performed. The different conditions within replicates were also similar or dissimilar to each other in line with expectation, reflecting the validity of the experimental set up in either adequately differentiating, or retaining consistency, between conditions.

Overall, in order to study closed-loop formation transcriptome-wide, deep and reproducible coverage of reads across well-defined mRNA boundaries was required. This chapter detailed the strategies we have employed to meet these requirements.

Chapter 5 – Investigating mRNA features affecting the closed-loop

Note on contribution

The approaches detailed in this chapter were decided upon by collaboration between myself and Dr Horvath. Coding and data processing were performed by Dr Horvath. Visual verification of bioinformatic outputs was performed by myself. Schematics were generated by myself, while plots were made by Dr Horvath.

5 *Investigating mRNA features affecting the closed-loop*

After development of an experimental method (chapter 3), and development of a sequencing and bioinformatic approach as a readout (chapter 4), we then aimed to examine the landscape of closed-loop occurrence across the transcriptome.

As presented in the introduction, one body of research detailed how interactions between the mRNA cap, poly(A) tail and the closed-loop components were required for productive translation both *in vitro* and with reporter mRNAs *in vivo* (Gallie, 1991; Tarun and Sachs, 1995; Preiss and Hentze, 1998; Wells *et al.*, 1998). Another body of research has suggested mRNAs do not always form the closed-loop while actively translating, and that the closed-loop is dispensable for translation to take place (Park, Zhang, *et al.*, 2011; Adivarahan *et al.*, 2018; Khong and Parker, 2018). However, this model has largely been extrapolated from study of example mRNAs (Adivarahan *et al.*, 2018; Khong and Parker, 2018), or in the case of Park *et al.* (2011), by analysis of translation upon eIF4G1 mutation. Therefore, a more global *in vivo* analysis of closed-loop formation at a transcriptome-wide level was missing. A study reporting differential association of mRNAs with the individual closed-loop components was also an indication of variable closed-loop formation, however, association with the closed-loop factors may not necessarily translate directly to closed-loop formation. A study to more directly probe closed-loop formation was previously performed by our laboratory (Archer *et al.*, 2015), but focused on a small panel of mRNAs. Therefore, in this study we aimed to investigate the prevalence of the closed-loop at a cellular, transcriptome-wide level using the method established in the first two results chapters of this thesis. We also aimed to investigate whether there are mRNA features which could correlate with differential closed-loop formation between mRNAs, if this was indeed observed.

More specifically, there were two main questions we aimed to investigate. The first question was how prevalent the closed-loop was in the global population of mRNAs at steady state. Specifically, we aimed to observe the extent of 3' end recovery observed when the closed-loop complex was immunoprecipitated via a 5' pulldown with eIF4E. We expect that due to the RNase I digestion, recovery at the 3' end would be largely mediated by the protein-mRNA and protein-protein contacts of the closed-loop. We aimed to use this relative end

enrichment as an indication of how prominently the closed-loop appears in cells actively translating mRNAs.

The second question we aimed to investigate was whether there were mRNA features that correlated with closed-loop formation. First, we aimed to elucidate whether there were physical mRNA features, such as transcript length or CDS length, that could potentially affect closed-loop formation. Secondly, regulatory and functional outputs were tested for closed-loop involvement. These included features such as association to the closed-loop proteins, mRNA half-life or translation efficiency. However, to perform these correlation analyses, we first required a metric to compare the extent of closed-loop between mRNAs. Our approach for defining this metric is detailed and explained in this chapter, before presenting the various correlation analyses performed.

Due to time constraints, we focused our strategy development and analysis on the first replicate generated. From the replicate correlation analyses presented in section 4.3, we expect a good concordance in signal distribution between the replicates, so we would expect replicate 2 and 3 to show similar results as those presented in this chapter. However, replicates 2 and 3 will continue to be examined and compared to the results obtained with replicate 1. There are also analyses still in progress at the time of writing. Nevertheless, the findings in this chapter present initial insights into closed loop occurrence under physiological conditions and begin to raise interesting questions regarding the role of the closed-loop in translation.

5.1 Snapshot of the closed-loop in living cells

In light of the two bodies of research suggesting either the closed-loop was important for translation or that it was not essential, the prevalence of the closed-loop across the transcriptome was probed. In order to study this, we aimed to understand the overall distribution of reads across mRNAs when the closed-loop was pulled down at the 5' end via eIF4E. We expected an even distribution of reads across mRNAs in the input, a very low level of binding in the untagged IP and for the 5' pulldown through eIF4E, we expected a high enrichment of reads at the 5' end, a depletion in the middle, and an enrichment at the 3' end if the closed-loop was detectable.

We first sought to observe the distribution of reads at a metagene level in order to obtain an overall picture of the closed-loop. To do this, we took the transcript annotations defined in the previous chapter and observed the enrichment at their 5' and 3' ends. We then scaled all transcripts to the same length and summed their coverage profiles, to produce a representation of the relative distribution of reads at different positions along the mRNAs.

5.1.1 Metagene plots – by read

We first summed the RNA seq-coverage at a per-read level, with all reads given equal weighting regardless of which mRNA they came from. A metagene plot was generated to summarise the combined read distribution across all mRNAs. In this way, the metagene plot is a representation of the end enrichments all the mRNA *molecules* in the library, with highly-abundant mRNAs contributing more to the read distribution profile than low-expressed mRNAs.

Both the untagged and tagged input profiles appeared similar, showing a fairly even coverage across the mRNAs apart from the presence of a few prominent spikes (Figure 5.1a,b). It appeared these spikes were from a few very highly expressing mRNA sequences, as they only appear in this by-read metagene. The untagged pulldown mirrored this profile (Figure 5.1c), suggesting this negative control accurately represents background binding by non-selectively associating with the same distribution of mRNA fragments found in the input.

Unlike the inputs and the untagged IP, in the case of the eIF4E 5' pulldown, a strong enrichment at the 5' end can be observed, followed by a gradual decrease to the 3' end, then a slight uptick at the end of the aggregated mRNA body (Figure 5.1d). If the closed-loop was a prominent feature of all mRNA molecules, a larger enrichment of reads increasing to the TES would be expected. Therefore, it was concluded that few mRNA molecules are found in the closed-loop state at steady state.

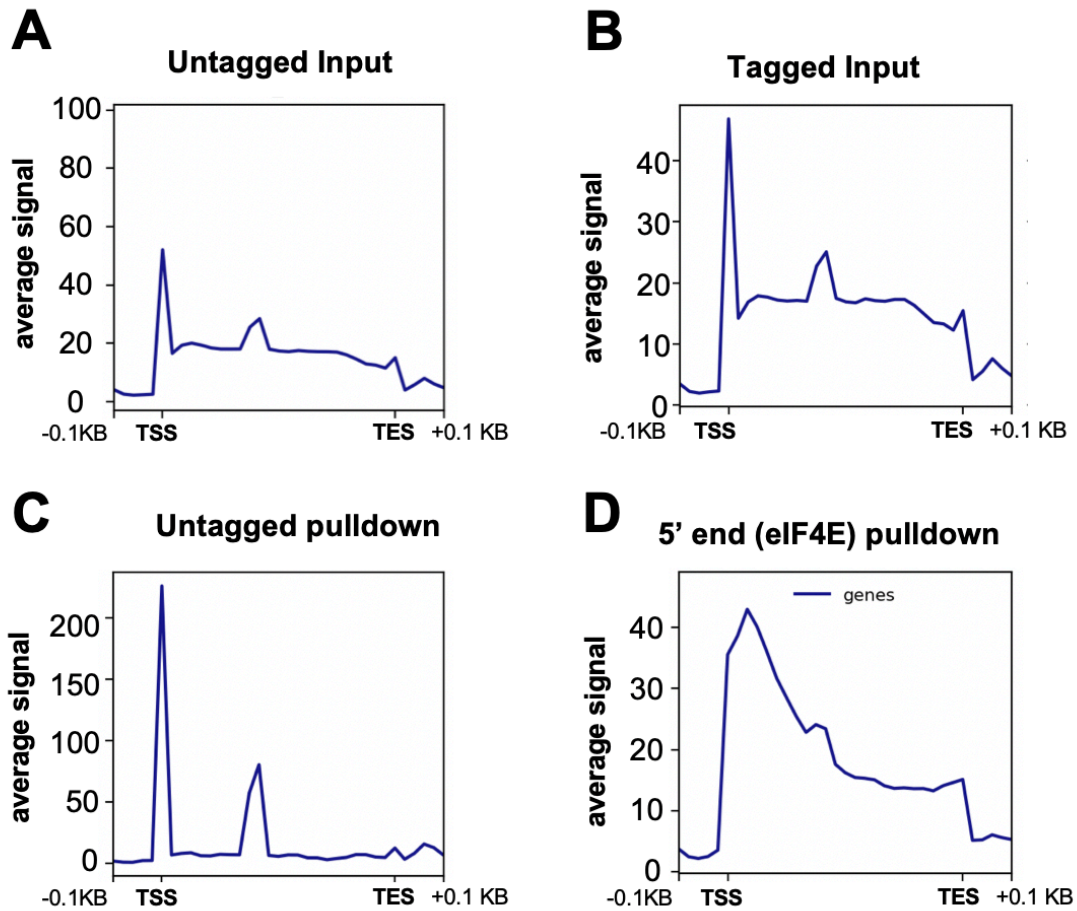


Figure 5.1. Summation of RNA-seq coverage at a per-read level, with all reads given equal weighting

The *de novo* transcription unit predictions from chapter 4 were scaled to a common length. The read coverage values from each mRNA were summed, then averaged for each position along the scaled representative mRNA body. TSS refers to the transcription start site, or start of the transcript, and TES refers to transcription end site, or end of the transcript. 0.1 kb before and after the mRNA body are shown as +0.1KB and -0.1KB respectively. The average signal (y-axis) refers to average coverage signal obtained from summing all reads at that scaled position and dividing by the number of transcript annotations. Untagged (A) and tagged (B) inputs refer to whole cell lysates (WCL) of the untagged WT strain and the eIF4E-TAP, eIF4G1-3xFLAG strain respectively. These were digested in parallel with their corresponding immunoprecipitation reactions, untagged pulldown (C) and eIF4E pulldown (D).

5.1.2 *Metagene plots, by gene*

A different profile appears when the metagene plots are normalised to a per-gene basis, where all genes are considered equally regardless of their expression levels. In this analysis, coverage profiles from all genes contribute equally to the final profile, regardless of whether they are a highly expressed or lowly expressed mRNA. Each mRNA had its highest signal within the defined transcription unit designated 1, with all other positions in the mRNA given a score relative to this highest signal. These normalised coverage values were then averaged across all mRNAs for each scaled position.

Similar to the metagene profiles summed at a per-read basis, the coverage profiles of both inputs show a relatively even coverage across the aggregated mRNA body, in line with our expectations for the input to equally represent all positions of the mRNAs. The spikes previously seen are also absent, suggesting the previous spikes seen in Figure 5.1 were mainly influenced by a small group of highly expressed fragments.

The untagged pulldown again mirrored the inputs, with mRNAs showing a low and flat normalised coverage signal across the whole mRNA body. The spikes which were prominent in the previous metagene plot also were no longer present, again indicating these spikes had most likely belonged to a few highly-expressed fragments, which were subsequently normalised out by the equal per-gene weighting in this metagene plot.

Regarding the tagged immunoprecipitation (IP) sample under these conditions, the same large enrichment at the 5' end, followed by a gradual decrease towards the middle was observed. However, there was a more pronounced 3' end enrichment compared to our earlier observations, suggesting that at the gene-normalised level, where all mRNAs are treated equally regardless of their expression levels, the presence of a closed-loop structure becomes more noticeable. This finding indicates that a subset of mRNAs is contributing to this 3' enrichment profile more than others, and that these mRNAs with more 3' enrichment may not be the most highly expressed ones. This group of mRNAs contribute to the 3' enrichment profile more prominently when given equal weighting with the more highly expressing mRNAs dominating the profile in the un-normalised metagene plots (section 5.1.1).

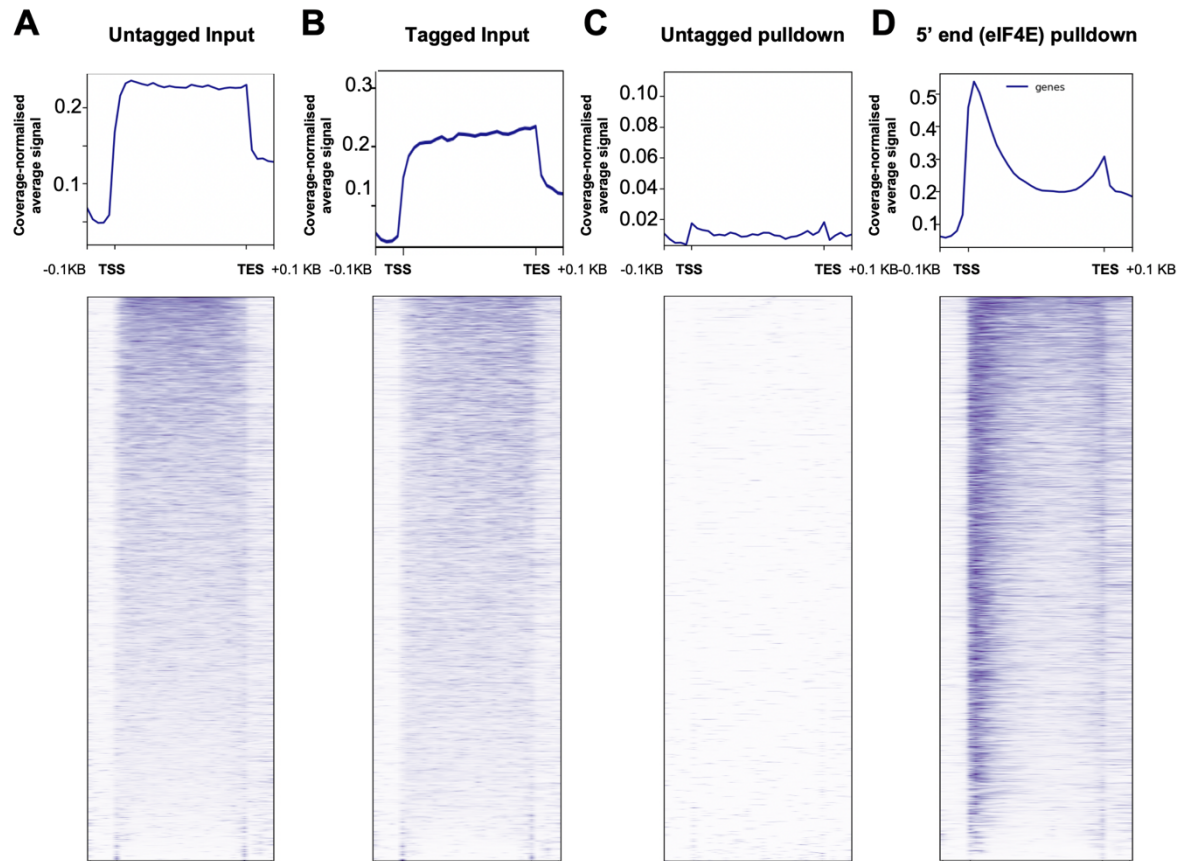


Figure 5.2. Summation of RNA-seq coverage at a per-gene level, where normalised coverage profiles are first determined for each gene individually before aggregation

The *de novo* transcription unit predictions from chapter 4 were scaled to a common length. The read coverage values from each transcript were internally normalised, with the highest signal within the transcription unit designated 1, and all other positions in the mRNA given a score relative to this highest signal. The result of these individual gene-wise normalisations are depicted by the heatmaps under each metagene, with each row representing an individual transcription unit. These normalised profiles were then aggregated and averaged across the scaled representative mRNA body, to give the coverage-normalised average signal (y-axis). TSS refers to the transcription start site, or start of the transcript, and TES refers to transcription end site, or end of the transcript. 0.1 kb before and after the mRNA body are shown as +0.1KB and -0.1KB respectively. Untagged (A) and tagged (B) inputs refer to whole cell lysates (WCL) of the untagged WT strain and the eIF4E-TAP, eIF4G1-3xFLAG strain respectively. These were digested in parallel with their corresponding immunoprecipitation reactions, untagged pulldown (C) and eIF4E pulldown (D).

5.2 Development of closed-loop score

Before correlation analyses could be performed between the extent of closed-loop formation and the various features introduced in section 5, we first required a method of sorting mRNAs relative to each other on the basis of their extent of closed-loop. Therefore, for each mRNA, we required an unbiased way to compare the RNA-seq signal at 5' and 3' ends as a proxy for closed-loop occurrence (Figure 5.3). While conceptually straightforward, this task required careful technical considerations. As seen in Figure 5.3, the RNA signals at the ends are not always uniform, and there is great variety in the length, shape and slope of the signals at the ends. Although it was unlikely any approach could accurately capture and compare all different variations of end signals, we aimed to devise a closed-loop measure to be as representative of the relative read distribution at the ends as possible. Our approach consisted of first defining start and end signals by finding the maximum signal at defined and consistent 5' and 3' regions of each mRNA. Then, an appropriate middle region was chosen as an internal point of comparison. Finally, the ratio between the start and end signal for each mRNA was calculated and transformed to a ranking measure for use in correlation analyses. For the development of this approach, the same 'high confidence' set of genes from chapter 4 were used for confirming the appropriateness and validity of thresholds and overall strategy.

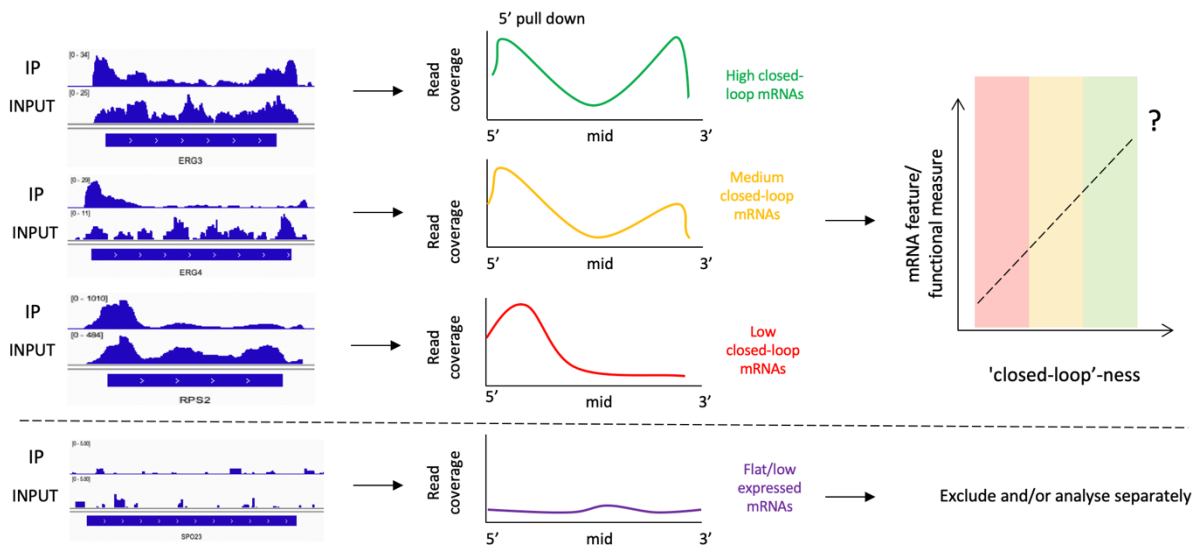


Figure 5.3. Approach for comparing mRNAs by their extent of relative end enrichment, as proxy for closed-loop formation.

Read coverage profiles differed greatly between mRNAs (left), but a method of consistently assessing 'closed-loop status' for each mRNA was required (middle), in order to perform downstream correlation analyses (right).

5.2.1 Start and end signals

The ratio between start and end signals was our predominant measure for an mRNA's closed loop propensity. However, as mentioned previously, there was a large variation observed in the coverage profile at the ends of mRNAs across the transcriptome. Therefore, we required an approach to defining end signals that could incorporate a level of leeway for the different presentations of these signals. mRNAs also vary greatly in their lengths, so this also needed to be taken into account when defining end signals. Taking these considerations into account, we first defined a radius from each end of the annotation as the segment to search in (Figure 5.4, blue dashed box). This was set at 30% of the length of the annotation from each end. Then within this blue section, we identified the top half containing the highest continuous stretch of signal (red) and averaged the signal in this region over the region's length. We believed this to be a compromise between capturing the full and highest end signal, while also keeping the signal relatively anchored to the ends. In terms of how these values were settled on specifically, multiple rounds of testing various thresholds were conducted and checked visually against the 'high confidence example set' for fit.

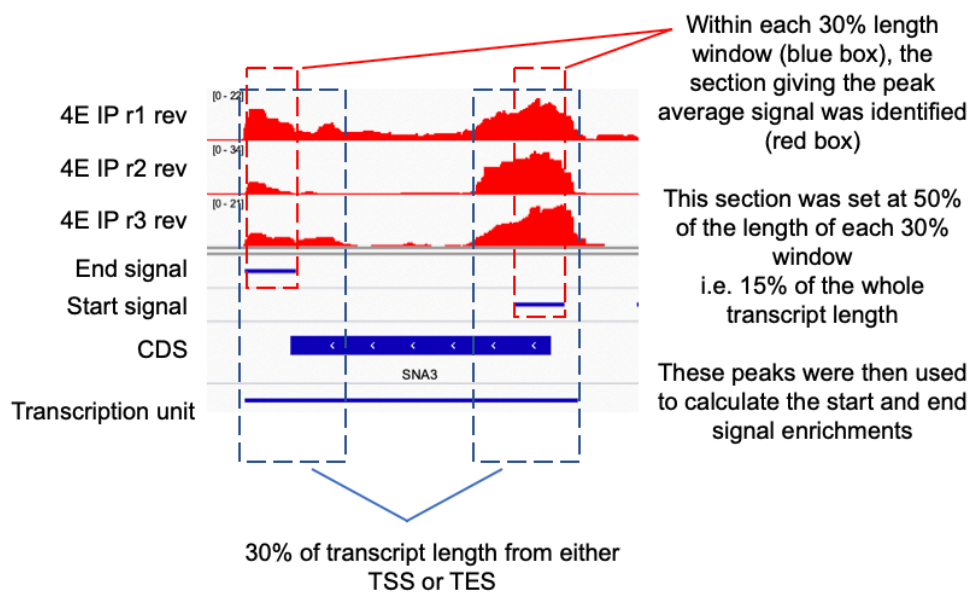


Figure 5.4. Methodology for defining start and end signals

Three replicates are shown, labelled r1-r3. 4E IP refers to the 5' pulldown through eIF4E-TAP. End signal and start signal refer to the final start and end signals as calculated by our approach. This example is on the reverse strand, so should be viewed R(5')-L(3'). TSS refers to the transcription start site, or start of the transcript, and TES refers to transcription end site, or end of the transcript.

Some examples of the final start and end predictions are shown below (Figure 5.5). As previously mentioned, different mRNAs displayed differently shaped enrichments at the ends. For example, in the case of SAM1, the coverage starts cleanly from the predicted transcription unit ends and decreases towards the middle, whereas RPA34 had more rounded peaks starting further in from the ends, and SSC1 had asymmetrically shaped, uneven peaks at the ends. As seen in the start and end signal annotation tracks, these different presentations of the end signals have been accounted for by our approach.

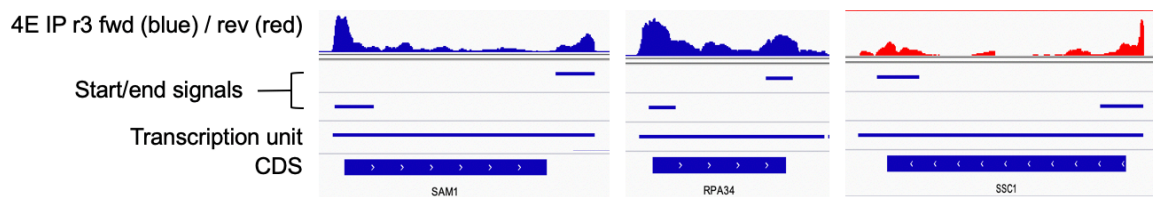


Figure 5.5. Examples of start and end signal predictions.

4E IP refers to the 5' pulldown through eIF4E-TAP, with blue profiles on the forward strand (L-R) and red profiles on the reverse strand (R-L). End signal and start signal refer to the final start and end signals as calculated by our approach. Transcription unit refers to our de novo predicted annotations.

5.2.2 Choice of middle region signal

Next, we required a middle section to compare the ends to, to ensure that there was a depletion of the middle signal. Given our experimental approach, we have the strongest indication of closed-loop formation if both 5' and 3' ends are both enriched upon 5' pulldown, and the middle section is also depleted. If the middle section is not depleted, this could indicate that the mRNA had not been sufficiently digested and was still intact. The 3' end would not need to be in contact with the closed-loop proteins for the 3' end signal to also be present, as the whole mRNA would be pulled down through its contact with eIF4E, regardless of its closed-loop status. In such a scenario, it would be difficult to discern whether the 3' end signal was also observed due to closed-loop formation, or due to insufficient digestion of the mRNA. Therefore, a representative middle section to compare the end signals to was required.

In order to define an appropriate middle signal for all mRNAs, we again tested multiple options for choosing the length and location of this middle signal and compared each option to our list of 'high confidence' examples for validity. Our first attempt was to take the

average signal of the intercept to the start and end signals, being the signal across the rest of the transcription unit excluding the start and end signals. However, we found this to be biased against shorter genes. We observed that as these mRNAs were shorter, they would have less RNase I digestion overall, and therefore more signal decreasing into the middle. When the average was taken over the whole intercept to calculate the middle signal, this resulted in a higher middle signal, contributed by the tails of the signal from each end, and division by a smaller value (Figure 5.6). Longer mRNAs would therefore appear to have a more prominent depletion in the middle due to the longer region the intercept signals were averaged over. We then trialled defining a 500 bp region from the centre. However, we faced a similar problem as before, with a number of predicted transcription units being under 500 bp in length. For the short genes in this case, the start and end signals often therefore contributed again to the calculation of the middle signal, while the longer genes did not have this effect. Finally, we set the middle section to be the average signal over the middle 50 bp of each transcription unit. This gave us a region over which we could calculate an average signal without length affecting the measurement substantially. In general, for all examples observed, a 50 bp region in the middle was sufficiently depleted for both long and short mRNAs.

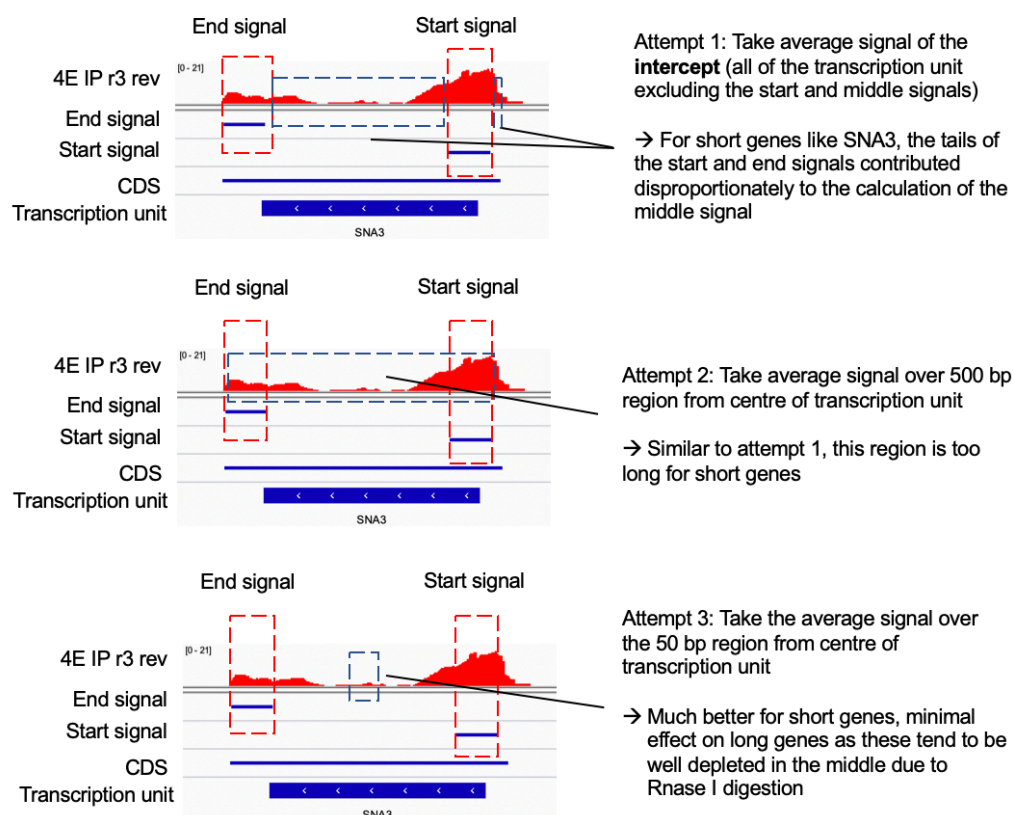


Figure 5.6. Options tested for defining the 'middle' signal.

Red dashed boxes denote the start and end signals, while the blue dashed box denotes the trialled middle signal. 4E IP refers to the 5' pulldown through eIF4E-TAP, with the red profiles on the reverse strand (R-L). End signal and start signal refer to the final start and end signal locations as calculated by our approach. Transcription unit refers to our de novo predicted annotations.

5.2.3 Exclude low expressed genes + flat genes

With the start, end and middle signals calculated for each mRNA, we first divided the average start and ends signals by the average middle signal to find the fold change of each end relative the middle for each mRNA. Genes with low expression levels in the input, defined as having an average signal in the input of below 2, were filtered out. Then, for the purposes of this analysis, we also excluded from the analysis 'flat genes', genes which we identified as not having start and end signals sufficiently enriched above the middle. Our criteria for this filter was that either the start or the end signal had to be at least 2-fold higher than the middle to pass. This step was incorporated because if just the starts and ends of 'flat' genes are compared, they would be even, indistinguishable from the more confidently 'closed-loop' examples where the ends are both enriched above the middle. In total, 359 genes were excluded for being too flat, and a larger set of 1497 were excluded for not meeting the input expression cut off.

5.2.4 Final score and split into groups (high 5' UTR, u-shape, high 3' UTR)

From the list of genes that had passed the flat and low expression filters, we then performed the final calculation, resulting in a score of $\log_2 \left(\frac{\text{start signal}}{\text{end signal}} \right)$, which we called the 'closed-loop measure'. This value represented the relative enrichment of the 5' signal over the 3' signal. The closer the average 5' and 3' signals were to each other, the closer to 0 this final calculated value was, and according to our interpretation, suggests a higher level of closed-loop formation for that gene.

Using this metric, we obtained a number for each mRNA remaining and sorted them into categories. First, we had the 'u-shaped' examples, which we believe are showing the most closed-loop, defined as genes having a closed-loop measure of between -1 and 1. This meant they had ends with signals within double of one another. Then, genes with scores above 1 were classed as 'high 5' UTR' genes, where the 5' end is more than double that of the 3' end. Finally, we also identified a smaller group we named 'high 3' UTR' genes, with a score < -1,

suggesting the 3' enrichment was at least double that of the 5' UTR signal, which was opposite to expectation given the 5' end pulldown. Again, the classification within the high confidence examples set were compared to the numerical rankings to ensure concordance.

5.3 *Variety in closed-loop between mRNAs*

To observe the range of scores obtained by our closed-loop metric, we normalised all average signals (start/middle/end) to the middle signal, then clustered genes based on the ratio between the starts and ends using the categories defined previously – u-shaped, high 5' UTR and high 3' UTR. As seen in Figure 5.7, different genes do appear in these three categories to different extents. Approximately 850 genes fell within the u-shaped, or high closed-loop category, ~1300 genes were found in the high 5' UTR category and a smaller group of approximately 60 genes were classified in the 3' UTR category.

As a whole, the heatmap of middle-normalised start and end signals does align with expectation, with the u-shaped category having more equal levels of red for the $\log_2(\frac{\text{start signal}}{\text{middle signal}})$ and $\log_2(\frac{\text{end signal}}{\text{middle signal}})$ columns, suggesting comparable enrichment of both ends above the middle. The high 5' UTR group was majority red in the $\log_2(\frac{\text{start signal}}{\text{middle signal}})$ column, indicating an positive enrichment of the 5' end signal above the middle. However, unlike the u-shaped group, the $\log_2(\frac{\text{end signal}}{\text{middle signal}})$ column of this high 5' UTR group was predominantly yellow or blue, suggesting either equal or lower levels of signal at the 3' end compared to the middle. Finally, the high 3' UTR category showed the opposite, with a predominantly red signal in the 3' end and majority yellow or blue signal at the 5' end. Furthermore, a general enrichment of the ends over the middle can be observed, as mirrored in the by-gene metagene plot in Figure 5.2 and as expected from the experimental set up.

Although the overall pattern of end enrichments in each category is consistent with expectation, the u-shaped group does contain some proportion of light orange/yellow in both $\log_2(\frac{\text{start signal}}{\text{middle signal}})$ and $\log_2(\frac{\text{end signal}}{\text{middle signal}})$ columns, indicating that a strong enrichment of either end above the middle was not consistently observed in this group. This might suggest the 'flatness' filter could be increased slightly, or that a stricter definition of the 'u-shaped'

category could be used. We recognise that such definitions inherently require an element of qualitative judgement, so would therefore be hesitant to interpret this categorisation of genes as the definitive proportional breakdown of closed-loop vs non-closed-loop genes in yeast. Hence, in subsequent correlation analyses, we have generally chosen to consider the genes as the one range of scores, rather than as three distinct categories. However, we do believe that overall, this categorisation for the purposes of clustering does aid in conceptualising and visualising the diversity in end enrichments present in our sample.

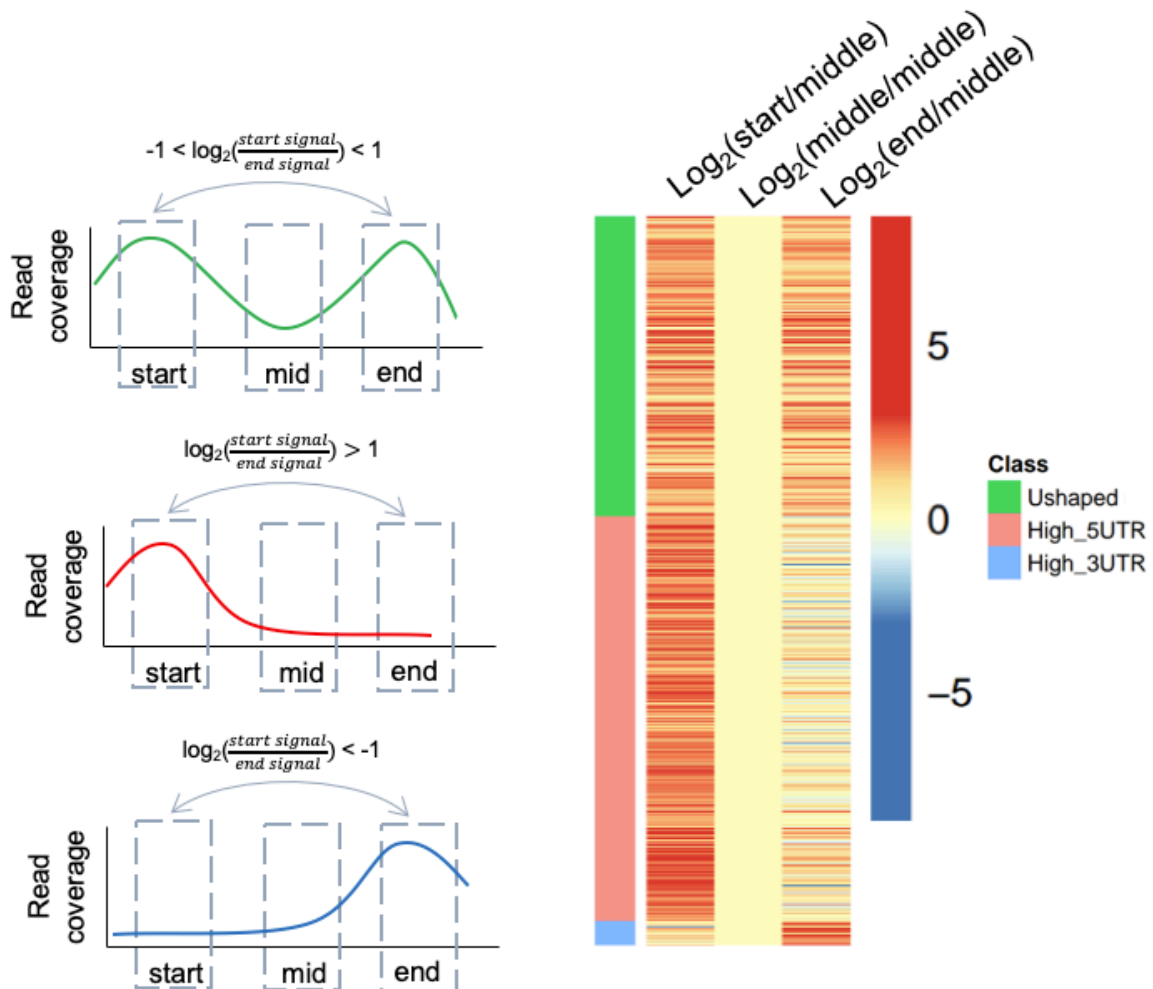


Figure 5.7. Distribution of closed-loop measurements across the transcriptome

All regions (start/middle/end) were individually normalised to the middle signal and log2 transformed. These three measurements were then displayed as three separate columns for each gene, with colour indicative of the level of enrichment above the middle, with yellow being no enrichment, increasing red as increasing enrichment and increasing blue as increasing level of depletion compared to the middle. Genes with a $\log_2\left(\frac{\text{start signal}}{\text{end signal}}\right)$ value between -1 and 1 were clustered as 'u-shaped' (green), genes with $\log_2\left(\frac{\text{start signal}}{\text{end signal}}\right)$ above 1 were classified as high 5' UTR (red) and those below -1 were classified as high 3' UTR (blue).

5.4 mRNA features affecting the closed-loop

With the manually-clustered heatmaps indicating a large diversity in closed-loop scores across the transcriptome, we next aimed to investigate whether any correlations could be found between the extent of closed-loop formation and mRNA features, either physical, or regulatory and functional. With the current closed-loop measure of $\log_2(\frac{\text{start signal}}{\text{end signal}})$, genes with near-equal enrichment of their ends would have a value converging on 0. For the sake of visualisation and correlation analyses, for some analyses we performed a technical transformation of our closed loop measure to the inverse of its absolute value (i.e.

$\log_2(\frac{\text{start signal}}{\text{end signal}}) \rightarrow \frac{1}{|\log_2(\frac{\text{start signal}}{\text{end signal}})|}$), in order to have a scale where the higher the value, the more equal the enrichment at the ends and we believe therefore, the more closed-loop.

5.4.1 Physical attributes of mRNA – CDS, UTR and overall transcript lengths

As formation of the closed-loop is a phenomenon occurring in 3D space, physical features such as the length of mRNA could play a role in determining closed-loop formation. On one hand, a shorter mRNA molecule may lead to more effective circularisation due to the decreased distance between the ends, leading to increased ease for the closed-loop proteins to come into contact with both ends. However, it could also be possible for longer mRNAs to have greater overall flexibility in the RNA molecule, allowing them to more easily overcome the steric challenges of bending into a circular structure.

In order to determine whether overall mRNA length was positively or negatively related to our closed-loop measure, we correlated our *de novo* transcription unit predictions to our transformed closed-loop measure, where increasing x-axis values were indicative of more equal enrichment at the ends. Unexpectedly, we did not observe a strong positive or negative correlation between mRNA length and our closed-loop measure (Figure 5.8a). However, when comparing the ‘non-closed-loop’ groups of high 5’ UTR and high 3’ UTR (red and blue, transformed closed-loop measure < 1) to the ‘u-shaped’ group (green, transformed closed-loop measure > 1), the distribution of individual datapoints between the groups is slightly different. The ‘non-closed-loop’ group appears to skew slightly towards shorter transcription unit lengths, while the ‘u-shaped’ or high closed loop group appears to skew

very slightly towards longer lengths. This distribution could indicate a weak correlation between increased mRNA length and increased closed-loop formation. However, this would require more detailed analysis as the potential correlation remains weak at the present moment.

Regarding CDS length and closed-loop formation, one study had reported a correlation between shorter CDS lengths and association to the closed-loop factors (Thompson and Gilbert, 2017). We therefore aimed to observe whether this correlation to the closed-loop factors translated to increased closed-loop formation on short mRNAs. However, correlating annotated CDS lengths to our transformed closed-loop measure, we did not see this relationship (Figure 5.8b). Similar to the transcript length correlation, there was instead a possible weak correlation between increasing transcript length and increasing closed-loop measure.

Finally, the UTR lengths derived from the *de novo* predictions were compared against the closed-loop measure, to determine whether there were any effects of UTR length on closed-loop formation. However, no correlation was observed between the length of either UTR and the close-loop measure (Figure 5.8c,d).

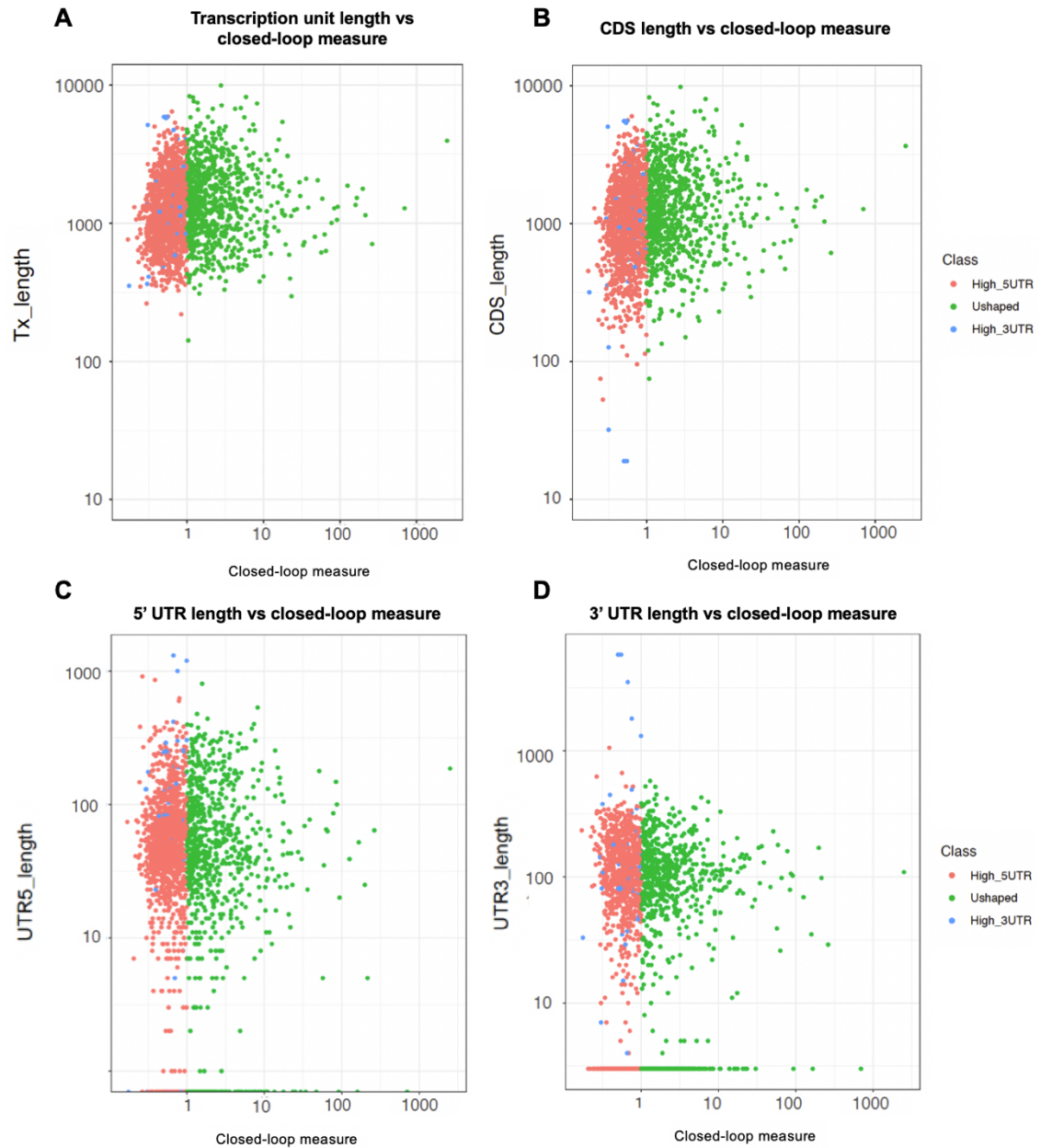


Figure 5.8. Correlation between closed-loop measure and mRNA features

The transformed closed-loop measure (x-axis) was compared to either the full transcript length as defined by our de novo predictions (A), annotated CDS length (B), 5' UTR length (C) or 3' UTR length (D) as defined by subtraction of the annotated CDS coordinates from the predicted de novo transcript unit predictions. Green refers to the original 'u-shaped' or high closed-loop mRNAs, red to the high 5' UTR category and blue to the high 3' UTR category as defined in section 5.2.4 and discussed in section 5.3.

5.4.2 Regulatory and functional features vs closed-loop formation

Regulatory and functional features were also investigated for potential closed-loop involvement. These included features such as gene expression, gene ontology, association to the closed-loop proteins, mRNA half-life and translation efficiency.

5.4.2.1 Gene expression

From observing the differences in the metagene plots shown in Figure 5.1 (by read) and Figure 5.2 (by mRNA), it appeared that a relatively small subset of mRNA species, which together constituted a large proportion of mRNA bulk, were responsible for driving the lack of closed-loop seen in Figure 5.1. Therefore, we compared the relative input signals for different mRNAs against their closed-loop scores, in order to determine whether there was a correlation between gene expression and closed-loop formation. Consistent with the metagene plots mentioned above, mRNAs exhibiting less closed-loop (high 5' UTR group) appeared to have higher average input signals.

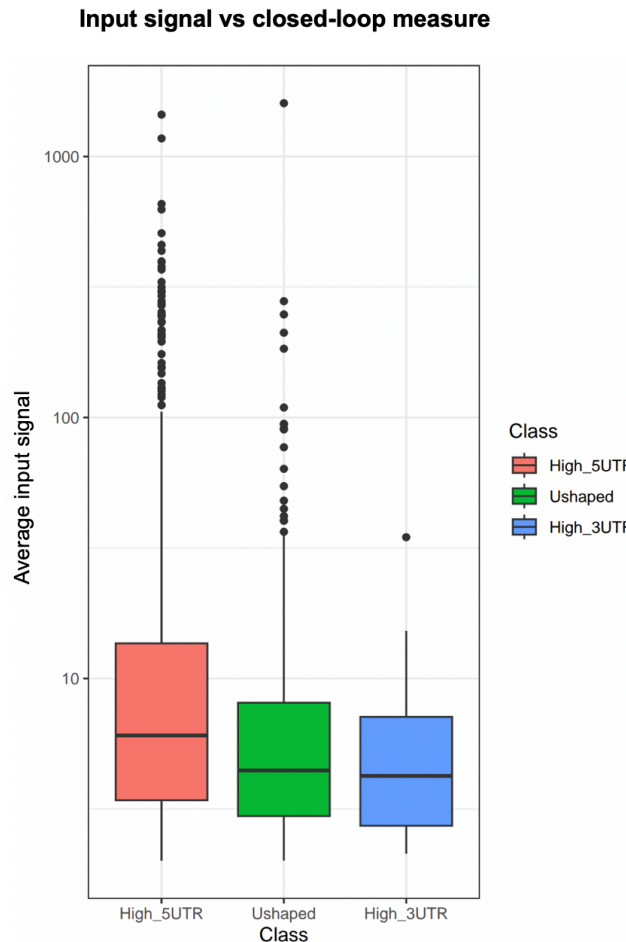


Figure 5.9. Correlation between gene expression and closed-loop measure

Relative gene expression was determined by comparing total length-normalised input signals of different mRNAs (y-axis). mRNAs were binned by their un-transformed closed-loop score (x-axis). Green refers to the original ‘u-shaped’ or high closed-loop mRNAs, red to the high 5’ UTR category and blue to the high 3’ UTR category as defined in section 5.2.4 and discussed in section 5.3.

5.4.2.2 Gene Ontology analysis

In order to determine whether the ‘u-shaped/high closed-loop’ category was enriched for any specific gene ontology terms, the list of genes from this category were compared to a reference set of the final genes belonging to any of the ‘u-shaped’, high 5’ UTR or high 3’ UTR groups. Using GO Term Finder (v 0.86) through Saccharomyces genome database (SGD), no significant GO terms were detected for either function ontology or component ontology using a p-value cut off of 0.01. Some terms were enriched for the biological process ontology with a p value cut off of 0.01, however, the top matches were for broad, general terms such as regulation of - cellular process, biological process and macromolecule

metabolic process. Therefore, it is difficult to definitively conclude that the closed-loop primarily forms on any one subset of mRNAs.

5.4.2.3 Association to closed-loop factors

Previous research has suggested mRNAs may associate with the closed-loop factors to different extents. Costello *et al.* (2015) individually tagged and immunoprecipitated the closed-loop factors and performed RNA-sequencing to identify mRNAs which were enriched in the IP above input levels. Differential association to the closed-loop factors was observed between mRNAs. Therefore, to observe whether this differential association to the closed-loop factors translated to differential closed-loop formation, we correlated our transformed closed-loop measure with enrichment data obtained from Costello *et al.* (2015) for each of the factors eIF4E, eIF4G1 and Pab1. Interestingly, for eIF4E and Pab1, there was no visible correlation between increased association to the closed-loop factor (y-axis, literature data) and increased closed-loop propensity (x-axis, our data) (Figure 5.10a,b). The mRNAs with low closed-loop scores (red/blue) had a large distribution in regard to association with either eIF4E or Pab1, while the mRNAs with the highest closed-loop measure scores appeared to congregate around average in the distribution of association to either eIF4E or Pab1. However, in the case of eIF4G1, there did appear to be a visually more prominent – though still weak – correlation between increasing association with eIF4G1 and increasing closed-loop score. In this case, the population deemed ‘low closed-loop forming’ (red/blue) appears to skew slightly lower in association with eIF4G1, while the high closed-loop or u-shaped group appears to skew slightly higher in association with eIF4G1 (Figure 5.10c). While this association is not particularly strong, the difference between the overall profiles of eIF4E or Pab1 and eIF4G1 is interesting, given the difference between these factors and their contacts with mRNA. Considering first the functions of eIF4E and Pab1, they have affinity for the cap and poly(A) tail respectively. Given the near-ubiquitous nature of these two features on mRNA, it would stand to reason that the binding of eIF4E and Pab1 might be more non-specific, and not a major determinant of closed-loop formation. However, as eIF4G1 is the scaffolding protein that is thought to bring eIF4E and Pab1 together into this loop, it would be logical for eIF4G1 to be a greater determinant of closed-loop formation.

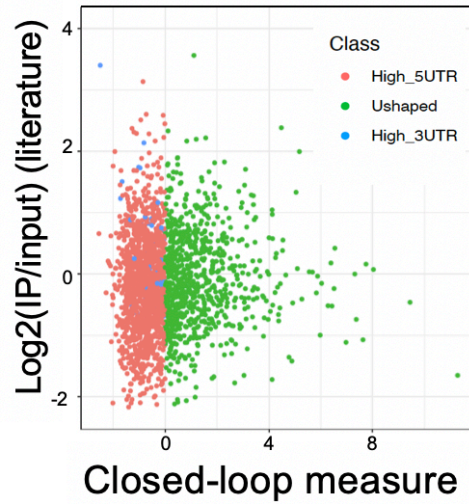
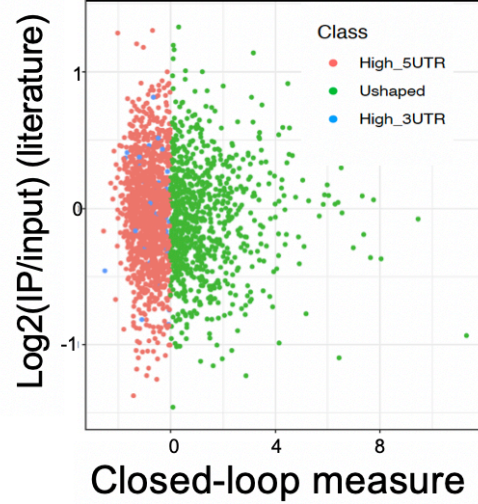
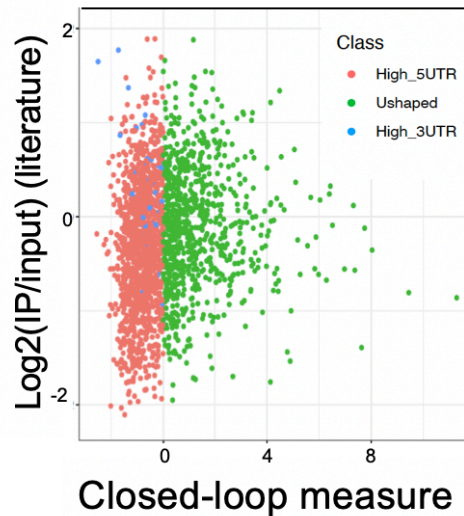
A**eIF4E association****B****PAB1 association****C****eIF4G1 association**

Figure 5.10. Correlation between closed-loop factor associations and closed-loop measure

The transformed closed-loop measure of each mRNA (x-axis) was compared to its level of association with either eIF4E (A), Pab1 (B) or eIF4G1 (C). Association of the mRNA to the factor in question (y-axis) was determined by Costello et al. (2015,) by measuring enrichment of the mRNA above input after co-immunoprecipitation with the tagged protein factor. Green refers to the original 'u-shaped' or high closed-loop mRNAs, red to the high 5' UTR category and blue to the high 3' UTR category as defined in section 5.2.4 and discussed in section 5.3.

5.4.2.4 mRNA half life

mRNA half-life was another feature we studied as potentially correlated to the extent of closed-loop formation. It has been hypothesised that contact between the poly(A) tail and the closed-loop factors may protect the mRNA from 3' → 5' degradation, resulting in longer half-lives for mRNAs that preferentially form the closed-loop. Therefore, we binned genes by closed-loop score (untransformed) in 0.5 increments and compared the average of each bin to mRNA half-life data reported by Geisberg *et al.* (2014). In this presentation of the closed-loop measure, the original measure of $\log_2(\frac{\text{start signal}}{\text{end signal}})$ was used. This translated to the high closed-loop genes (green) clustering around 0, high 5' UTR genes to the right (red) and high 3' UTR genes to the left (blue). We did not observe a relationship between our closed-loop measure and mRNA half-life data determined by Geisberg *et al.* (2014). Apart from the presence of a few outliers with particularly long half-lives, average half-lives were largely consistent across all groups (Figure 5.11). Therefore, we concluded that formation of the closed-loop was not significantly correlated with mRNA stability and half-life.

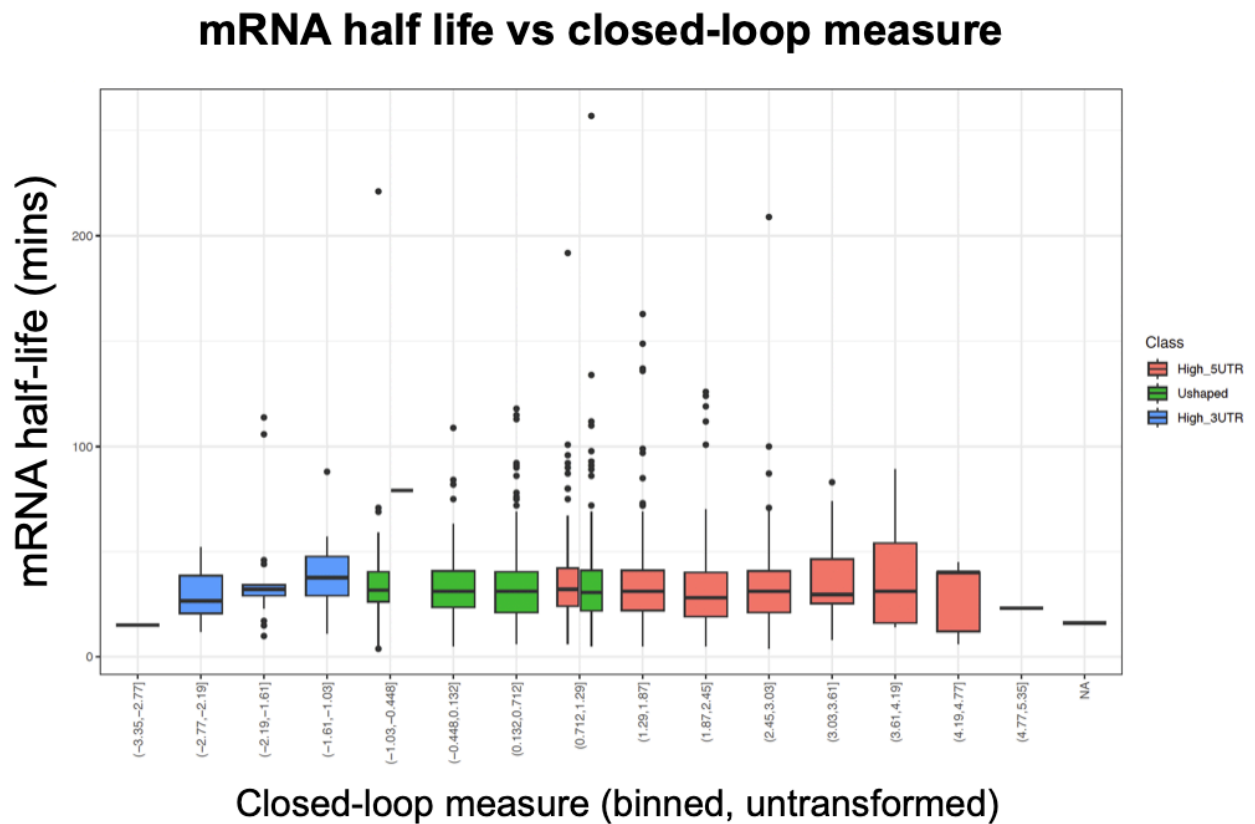


Figure 5.11. Comparison between mRNA half-life and closed-loop measure

mRNAs were binned in 0.5 increments by their un-transformed closed-loop score (x-axis). mRNA half-life measurements were obtained from Geisberg *et al.* (2014). Green refers to the original 'u-shaped' or high closed-loop mRNAs, red to the high 5' UTR category and blue to the high 3' UTR category as defined in section 5.2.4 and discussed in section 5.3.

5.4.2.5 Translation efficiency

Translation efficiency is perhaps the functional feature most expected to correlate with closed-loop formation. Since the conceptualisation of the closed-loop model, a major rationalisation for its existence has been that formation of the closed-loop allows for ribosome recycling after one round of translation, allowing for more efficient protein synthesis. Multiple studies have also detailed the stimulatory effects of the cap, tail and closed-loop factors on mRNA translation and protein output (Gallie, 1991; Salvador Z Tarun *et al.*, 1997; Preiss and Hentze, 1998). However, other studies have reported a negative correlation between closed-loop formation and active translation (Adivarahan *et al.*, 2018; Khong and Parker, 2018).

To investigate whether closed-loop formation and translation efficiency were correlated, either positively or negatively, we binned genes by closed-loop score (untransformed) in 0.5 increments and determined the average protein synthesis rate per bin, as measured by SILAC (Riba *et al.*, 2019). In this presentation of the closed-loop measure, the original measure of $\log_2(\frac{\text{start signal}}{\text{end signal}})$ was used. This translated to the high closed-loop genes (green) clustering around 0, high 5' UTR genes to the right (red) and high 3' UTR genes to the left (blue).

On the left-hand side of the spectrum, the high 3' UTR bins were variable in their protein synthesis rates. However, as this group was much smaller with only ~60 genes, it is likely that a few prominent examples have skewed the average protein synthesis rates of this group.

Unexpectedly, we did not observe a positive relationship between high levels of closed-loop (green) and increased protein synthesis rates. Across the 4 bins in the 'u-shaped' category, it appears that protein synthesis may instead be increasing slightly with increasing 5' enrichment over the 3' signal. We see this trend continue as the ratio of 5' signal to 3' signal increases further, and overall we observed that genes with less enrichment at their 3' end compared to the 5' end (red) were the ones with higher average protein synthesis rates (Figure 5.12). Interesting, the highly abundant ribosomal protein genes commonly appear in

this high 5' UTR category and visually, they consistently show low enrichment at the 3' ends compared to high levels of enrichment at the 5' end. It is possible that this abundant group drives this trend of increasing protein synthesis rate upon increasing 5' to 3' end signal ratio.

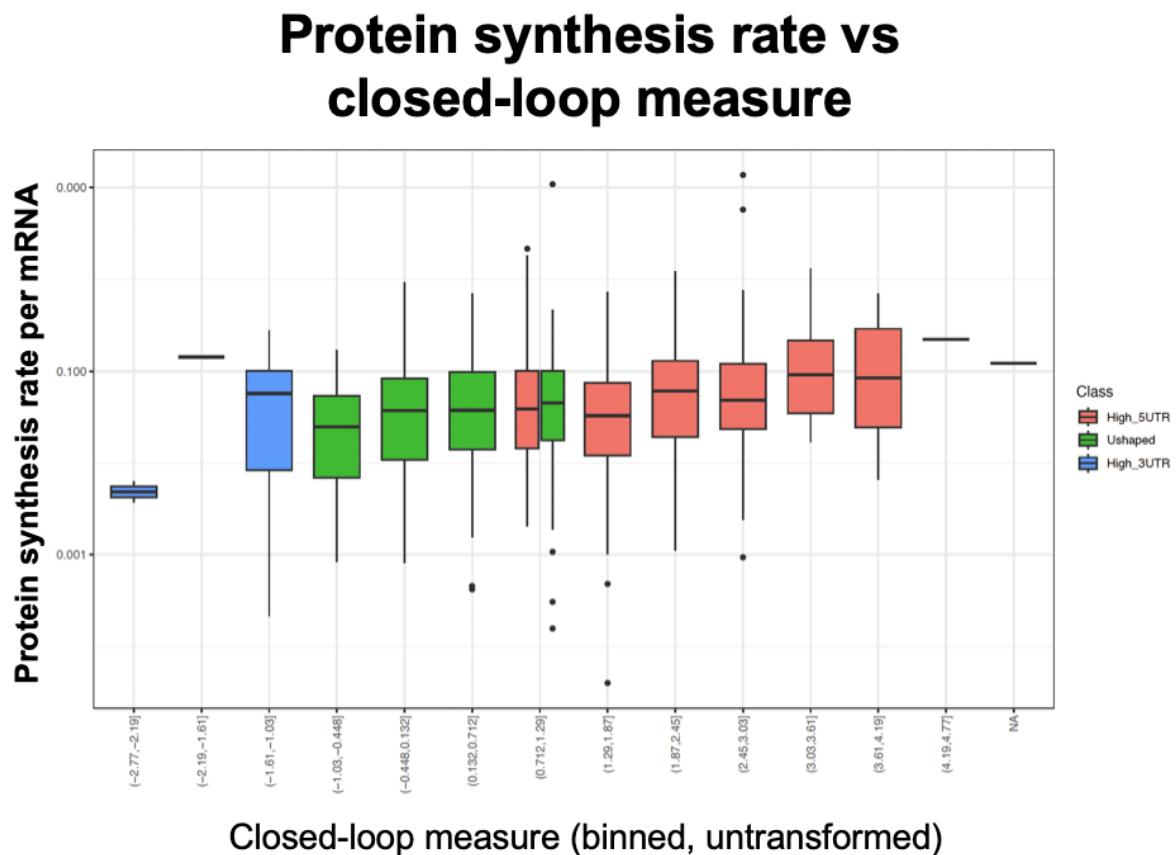


Figure 5.12. Comparison between protein synthesis rate and closed-loop measure

mRNAs were binned in 0.5 increments by their un-transformed closed-loop score (x-axis). Protein synthesis rates per mRNA were obtained by SILAC by Riba et al. (2019). Green refers to the original 'u-shaped' or high closed-loop mRNAs, red to the high 5' UTR category and blue to the high 3' UTR category as defined in section 5.2.4 and discussed in section 5.3.

5.5 Chapter 5 conclusions

This chapter presents our findings to a few major biological questions. First, we asked whether the closed-loop could be observed, and whether it was a prevalent feature at a cellular, transcriptome-wide level in actively growing cells. We then sought to investigate further whether diversity between mRNAs could be observed in the extent of their closed-loop formation, and whether this differential closed-loop formation could be correlated to various physical, regulatory and functional features.

Regarding the first question, from metagene analyses representing the read distribution over an aggregated and standardised gene body, it appears the majority of reads from a 5' pulldown are found at the 5' end, with only a small enrichment at the 3' end. This suggests that few 3' end fragments are remaining associated with the 5' end through the closed-loop. Therefore, it was concluded that few mRNA molecules are found in the closed-loop state at steady state. However, when the same metagene analysis was performed at a per-gene basis, with all genes given equal weighting regardless of expression level, the enrichment at the 3' end became more prominent. This suggests there are some mRNAs contributing to this 3' enrichment profile more than others, and that these mRNAs with more 3' enrichment may not be the most highly expressed ones.

To better understand this group of mRNAs which may be driving the 3' end enrichment seen in the per-gene metagene plots, we devised a method to compare the relative 5' and 3' end enrichment for each mRNA as proxy for the extent of closed-loop formation. By defining regions of high signal at the ends of mRNAs and normalising these regions to a middle signal, we observed a large variation in end enrichment profiles between mRNAs. Notably, we observed a group of mRNAs with close to equal enrichment at the ends (which we termed the 'u-shaped' group), a larger group with predominantly 5' enrichment only, and a much smaller group with predominantly 3' enrichment.

We then used this 'closed-loop measure' to rank mRNAs and investigate whether our proxy measure for closed-loop formation correlated to any physical or functional features. We observed a possible weak correlation between increasing CDS length and increasing closed-loop formation, but no correlation between our closed-loop measure and UTR lengths. No specific gene ontology terms were highlighted as correlating either. We observed little correlation between association to either eIF4E or Pab1 and our closed-loop measure. However, a weak correlation may be present between association with eIF4G1 and our closed-loop measure, possibly highlighting the scaffolding role of eIF4G1 in maintaining the closed-loop structure. No correlation between mRNA half-life and closed-loop measure was observed. Finally, protein synthesis rate appeared to be negatively correlated with closed-loop formation, with a higher average protein synthesis rate observed for genes with more enrichment at their 5' end compared to their 3' end.

The results presented in this chapter begin to challenge some of the initial explanations and justifications for the closed-loop model and have raised interesting questions, some of which will be discussed in the next chapter. Overall, these findings suggest that the closed-loop model is perhaps more complex than initially hypothesised, and therefore its role in translation warrants further investigation.

Chapter 6 – Discussion

6 Discussion

6.1 Summary

From the initial discovery of the synergistic effects of the cap and poly(A) tail in translation (Gallie, 1991; Tarun and Sachs, 1995; Preiss and Hentze, 1998), to the characterisation of the closed-loop factors and their binding to each other (Grifo *et al.*, 1983; Salvador Z Tarun *et al.*, 1997; Kessler and Sachs, 1998), and finally to the visualisation of mRNA circularisation in the presence of the closed-loop proteins (Wells *et al.*, 1998), the closed-loop model has been upheld as an explanation for how mRNA end proximity might enhance translation. Although early studies reported a strong dependence on the closed-loop for productive translation, this conclusion has been challenged by more recent findings. Some more recent studies have found formation of the closed-loop to be negatively correlated with active translation in *in vivo* conditions (Adivarahan *et al.*, 2018; Khong and Parker, 2018), or found that disruption of the closed-loop interactions did not lead to global translational impairment (Salvador Z Tarun *et al.*, 1997; Park, Zhang, *et al.*, 2011). Finally, arising from these two seemingly opposing bodies of research, was the suggestion that the closed-loop could be mRNA dependent or be a source of translational regulation (Archer *et al.*, 2015; Costello *et al.*, 2015)

Considering the lack of clarity around the function of the closed-loop in translation, and the many unanswered questions regarding the prevalence of its formation and a potential mRNA-dependent relationship, this thesis aimed to better understand the closed-loop in an *in vivo* context. We developed an experimental method to capture the closed-loop complex and associated mRNAs via an immunoprecipitation approach. Coupling the immunoprecipitation with RNase I digestion allowed us to observe relative end enrichments from the 5' and 3' ends as a proxy for closed-loop formation. We then adapted this experimental immunoprecipitation approach for transcriptome-wide analysis by short-read RNA-seq. During the process of transcriptome-wide analysis, we also developed a method to more accurately define mRNA transcript boundaries in a project-specific manner. Finally, we devised a method to identify the signals at the ends of transcripts after immunoprecipitation, and used these signals to compare closed-loop formation across the transcriptome. Our

method appears to be reproducible and specific, and has the potential therefore to serve as an avenue for further study into characteristics of the closed-loop.

From our transcriptome-wide analyses, we find that the closed-loop likely can form *in vivo*. We are able to observe strong examples of mRNAs with significant signal enrichment at their 3' ends upon 5' end pulldown with eIF4E. We also observe some mRNA recovery at the 3' ends when we take an aggregate view of gene-normalised read distribution across the transcriptome. However, despite some prominent examples of mRNAs which appear to favour formation of the closed-loop, the closed-loop does not appear to be a widespread or consistently observed phenomenon. The majority of mRNAs do not show significant signal enrichment at their 3' ends, and when taking into account the relative expression levels of different mRNAs, it appears that overall, the majority of mRNA molecules in a population are not in the closed-loop formation at steady state.

Finally, various mRNA features were studied for their potential association with closed-loop formation. Physical features such as CDS and UTR lengths were tested as possible determinants of closed-loop formation, with only CDS length and transcript length identified as potentially weakly correlated to increased closed-loop formation. Interestingly, we also failed to identify strong correlations between increased closed-loop formation and functional features such as enrichment of specific gene ontology terms, increased mRNA half-life or increased translation efficiency. However, we did observe a potential correlation between eIF4G1 association and increased closed-loop.

6.2 Study approach and method development

A key focus of this thesis was on developing an *in vivo* method for transcriptome-wide analysis of closed-loop formation. There were a few key differences between our approach and previous studies reported in the literature. First, the *in vivo* aspect of this method presents some differences compared to the translation systems predominantly used in early studies into closed-loop-mediated translation. Many early studies investigating the contributions of the cap and poly(A) tail were performed in cell-free systems, or with reporter mRNAs introduced into *in vivo* environments (Gallie, 1991; Tarun and Sachs, 1995; Preiss and Hentze, 1998). At the time, differing levels of synergy between the ends in influencing

translation output was observed depending on translation system used and experimental conditions. This suggests that while these experiments were crucial for mechanistic understanding of the effects of different mRNA and protein components, it is likely that the formation and function of the closed-loop may be context dependent, and these artificially perturbed systems may not fully represent *in vivo* conditions.

Given the complexity and heavily-regulated nature of translation initiation, there are therefore benefits to studying this process in conditions as close to unperturbed as possible, especially when the aim is to understand large-scale characteristics of the system. By rapidly fixing cellular complexes in actively growing cells using formaldehyde, our study has aimed to probe the closed-loop in conditions as close to physiological as possible. The speed of formaldehyde crosslinking was an advantage in our investigation of translation complexes, as formaldehyde swiftly permeates cell membranes and fixes these dynamic complexes, preserving their near-native state (Thavarajah *et al.*, 2012; Hoffman *et al.*, 2015).

Although the central immunoprecipitation method had previously been established for qPCR as a readout (Archer *et al.*, 2015), its adaption for use with short-read RNA sequencing was another key feature of this study. By employing RNA-seq as a readout, we were able to rapidly scale our analysis of the closed-loop to the transcriptome. Another advantage of RNA-seq analysis is the ability to gain information regarding read distribution over the entire mRNA, rather than only at small, directed amplicon regions. Therefore, a more comprehensive picture of end enrichments could be obtained, and the comparison of signals at the ends could be tailored to regions with real enrichments. The depth of the RNA-seq data also allowed us to map *de novo* the bounds of mRNA extremities, avoiding the need to rely on incomplete UTR annotations. This study-specific mapping of the UTRs was then able to guide us to identify end signals as close as possible to where mRNA contacts the closed-loop.

Finally, a key advantage of our crosslinking and RNase I digestion method was the ability to more directly probe the closed-loop beyond studying association to the individual closed-loop factors. A defining feature of our approach was the ability to retain and analyse positional information of mRNA fragments associated to the closed-loop complex. Through detection of end enrichments, we were able to detect the closed-loop by proxy, while also having the advantage of retaining *in vivo* conditions and allowing for larger-scale transcriptome-wide analysis.

6.3 *The closed-loop is likely able to form in vivo*

With the various closed-loop protein-protein interactions and protein-mRNA interactions identified, the mechanistic possibility for mRNA circularisation through the closed-loop model has been generally accepted. The strongest evidence for direct circularisation has been *in vitro*, where mRNA circularisation was observed by atomic force microscopy in the presence of the closed-loop proteins (Wells *et al.*, 1998). However, observations of this phenomenon *in vivo* have remained less prevalent. Both Adivarahan *et al.* (2018) and Khong and Parker (2018) have utilised single-molecule FISH to observe limited put possible co-localisation of mRNA ends *in vivo*, but this approach had the limitation of not directly addressing involvement of the closed-loop proteins in this circularisation. Findings were also limited to the example mRNAs selected.

By employing a closed-loop protein-based pulldown approach, we were able to involve the protein components of the closed-loop in this observation of end proximity *in vivo* as well. Although our approach is not a direct, microscopic visualisation as are the smFISH studies above, our findings do suggest that a circularisation of mRNA by the closed-loop factors is likely possible *in vivo*, irrespective of biological function. Given the well-established connections between the closed-loop factors and mRNA, and the incorporation of RNase I digestion to deplete the middle section of mRNAs, concurrently observed 5' and 3' enrichments most likely indicate presence of an intact closed-loop structure. The pulled down factor, eIF4E, also has not been reported to bind mRNA outside of contacts with the cap, so association of end fragments with the pulldown is more likely to be through contacts with another factor in the complex. This factor is most likely Pab1, given the bulk of the 3' end signal is found adjacent to the expected poly(A) tail, but this does not rule out other possibilities also co-existing, such alternative closed-loop conformations as have been proposed (Amrani *et al.*, 2008; Park, Walker, *et al.*, 2011; Wang, Xi, Toomey, Chiang, Hasek, Thomas M. Laue, *et al.*, 2016). However, ultimately, with the middle sections of the mRNA depleted by RNase I digestion, it is difficult to explain 3' end enrichment upon a 5' pulldown independent of some form of a bridging mechanism between the ends. Therefore, it is likely that the closed-loop structure observed *in vitro* can also persist *in vivo*.

6.4 *The pervasiveness and necessity of CL during translation*

An interesting finding from our study is that overall, while it seems formation of the closed-loop is possible, few mRNAs appear to be in this conformation at any one time. From the by-read metagene plots, few molecules appeared to be in the closed-loop formation at steady state. An examination of the individual coverage profiles from some of the most highly expressed mRNAs such as *TDH2*, *TDH3*, *ENO2* and *PDC1* (taken from Velculescu *et al.* (1997)) – which collectively make up a large proportion of mRNA molecules in each cell, confirms the lack of closed-loop formation on highly expressed mRNAs.

Interestingly, the occurrence of the closed-loop also did not appear to be positively correlated to increased translation efficiency/rate of protein synthesis per mRNA. The opposite was instead observed – that mRNAs with high translation efficiency generally had larger imbalances between their 5' and 3' ends. A visual examination of the mRNAs defined by Riba *et al.* (2019) as having the highest reported protein synthesis rates per mRNA, such as *MET6*, *ENO1*, *HSP12* and *SSA2*, confirms they do not show appreciable enrichment at their 3' ends. Therefore, unlike the findings from initial *in vitro* studies, our findings suggest that formation of a closed-loop structure is not essential for translation efficiency.

These conclusions appear to contradict the model of circularisation in its original form, in which the closed-loop was strongly implicated in promoting productive translation to take place. For example, synergistic effects of the cap and poly(A) tail on translation were originally observed when reporter constructs were electroporated into tobacco, Chinese hamster ovary cells (Gallie, 1991) and yeast spheroplasts (Preiss and Hentze, 1998). Removal of contact between eIF4G1 and Pab1 was also shown to abolish poly(A) dependent translation in yeast (Kessler and Sachs, 1998). However, these synergistic or stimulatory effects were often reported in the context of comparison with uncapped or un-polyadenylated reporter mRNA constructs, categories which the overwhelming majority of mRNAs do not belong to. When studying a population of majority capped and polyadenylated endogenous mRNAs in an *in vivo* context, there may be greater redundancy in the pathways available for productive translation to take place outside of the closed-loop structure formation.

Our results instead are more consistent with the smFISH studies (Adivarahan *et al.*, 2018; Khong and Parker, 2018) where actively translating mRNAs were observed to have less co-localisation of the ends. These studies found that for the example mRNAs tested, translating mRNAs rarely had close ends, and that the narrowing of end-co-localisation distances was instead observed upon stress or exit from active translation. It should be noted, however, that this observed lack of closed-loop formation on well translated mRNAs does not rule out a possible role of the closed-loop in enhancing the translation of a different subset of mRNAs. For example, a disruption of the closed-loop by eIF4G1 mutation was found to reducing the TE of a subset of mRNAs while increasing the TE of another group, while having little to no effect on the majority of mRNAs (Park, Zhang, *et al.*, 2011).

However, one caveat to this general conclusion should be considered. It should be noted that due to the inherent efficiency of formaldehyde fixation, and the need to balance capture of the closed-loop complex with over-fixation of non-specific interactions, it is possible that not all instances of the closed-loop were captured by our methodology. As paraformaldehyde was also used for fixation in the smFISH experiments, those findings also need to be considered with this caveat. Hence, our findings, and the smFISH experiment findings, may potentially be an underrepresentation of the level of closed-loop present in cells. However, in our case, as we do see prominent examples of genes with near-equal enrichment at their 3' ends compared to their 5' ends, this technical consideration is unlikely to be the only explanation for the low levels of closed-loop observed.

Overall, our research – consistent with the body of *in vivo* work suggesting a non-essential role of the closed-loop during translation – highlights the importance of further examining the potential biological role of the closed-loop.

6.5 mRNA-specific formation of the closed-loop

Our clustering of mRNA enrichment profiles demonstrates a large variation between mRNAs in the levels of their relative end enrichment. We observed that the majority of mRNAs did not show significant enrichment of their 3' end signal, but a sizeable number did have varying levels of 3' enrichment, ranging from a small amount to almost symmetrical enrichments compared to the 5' end. Although this was already reported previously by our

research group (Archer *et al.*, 2015), with the adaption for RNA-seq, we have now been able to show this on a larger scale. The increase in number of mRNAs for which end enrichment comparisons have been made has also allowed us to better examine potential relationships between closed-loop formation and physical or functional mRNA features.

Previous studies have also indicated that formation of the closed-loop could potentially be mRNA dependent. For example, immunoprecipitation of the individual closed-loop factors and analysis of co-immunoprecipitated mRNAs observed a difference in association to the closed-loop factors between mRNAs (Costello *et al.*, 2015). Crucially, they found that some mRNAs associated strongly with eIF4E and eIF4G1, while other mRNAs were more variable. However, little differential association with Pab1 was observed, consistent with the role of Pab1 as a non-specific poly(A) binding protein. Consistent with this, we did not find an association between Pab1 association and increased closed-loop formation. However, we also interestingly did not observe a correlation between eIF4E association and closed-loop formation, but we did observe a potential positive, albeit weak, correlation between eIF4G1 association and closed-loop formation. This could suggest that between the three closed-loop proteins, eIF4G1 is the major driver of closed-loop formation. As eIF4G1 is known to have RNA recognition domains (Berset *et al.*, 2003), is usually found in complex with eIF4E (Grifo *et al.*, 1983), and its binding to eIF4E is shown to increase eIF4E affinity for the mRNA cap (Haghighat and Sonenberg, 1997), it is possible that eIF4G1 is actually driving the differential association of eIF4E to mRNAs observed by Costello *et al.* (2015). This possibility is strengthened by the observation that the majority of individual mRNAs show similar eIF4E and eIF4G1 association profiles (i.e. mRNAs associating more with eIF4E pulldown were often also associating more with eIF4G1 pulldown) (Costello *et al.*, 2015).

However, despite the large observed differences in closed-loop formation between mRNAs, it is still unclear what could be driving the differential propensity for closed-loop formation. We did observe a weak correlation between CDS or transcript length with our closed-loop measure, but this small correlation is unlikely to fully explain the disparity. It is possible that closed-loop formation is dependent on multiple factors rather than one alone, and that individual effects of features such as length are too weak to observe alone. Along a similar vein, it is possible that cellular and environmental context plays a large role in closed-loop formation or lack thereof. There are also some factors that have not yet been examined, such as poly(A) tail length. Finally, although our gene ontology analysis did not result in any

specific enrichments, specific regulatory motifs could determine mRNA contact, especially with eIF4G1, given eIF4G1 has known RNA-binding domains. eIF4G1-mediated decay of select mRNAs during glucose starvation has been reported (Chang and Huh, 2018b), further strengthening the possibility of eIF4G1 binding as a potential determinant of closed-loop formation.

6.6 *Alternative roles for the closed-loop*

Aside from the original model of mRNA circularisation as a means of increasing translation efficiency, other biological functions or explanations for the closed-loop have been proposed. As our results do not point to a simple relationship between closed-loop and increased translation, it is worthwhile to consider our study in the broader context of these possibilities.

One possibility is that the closed-loop structure allows for storage of mRNA under cellular stress. mRNAs with the closed-loop components have been observed to localise to cytoplasmic granules termed EGP-bodies under glucose starvation in yeast, possibly in order for translation to resume after the stress has been removed (Hoyle *et al.*, 2007). It was suggested that this closed-loop conformation could protect mRNAs from degradation in these granules. This would advocate for a role of the closed-loop in protection and storage, rather than just in translation initiation.

Another possibility is that the closed-loop is involved in setting up the initial round of translation to bring all the translation initiation factors and pre-initiation complex in close proximity to the mRNA. This would be consistent with the mutually stabilising interactions reported between the individual factors and mRNA (Haghighat and Sonenberg, 1997; Wei *et al.*, 1998; Gross *et al.*, 2003), and with the role of eIF4G1 as a broader scaffolding protein for other initiation factors like eIF4A (Imataka and Sonenberg, 1997) and eIF3 (LeFebvre *et al.*, 2006). It could be possible that once the initial round of translation has been established, and the required translation factors and ribosomes are in close proximity, further rounds of translation can occur with the translation factors required remaining close by for re-initiation, without the need for the strict closed-loop structure.

Finally, a broader idea that encompasses some of these theories is that formation of a closed-loop structure is dynamic and transient, and more related to the spatial and temporal qualities of an mRNA rather than its function or physical presentation. For example, circular, linear and spiral polysome conformations have all been visualised on the rough endoplasmic reticulum (Christensen, Kahn and Bourne, 1987). smFISH experiments also report diversity in colocalisation between mRNA ends within a population of the same mRNA (Adivarahan *et al.*, 2018).

6.7 Future directions

Our results have raised some interesting questions. They do not appear to be consistent with the idea of the closed-loop model as being a major determinant of translational efficiency. The results in this study also have not yet identified a strong, definitive link between mRNA features and closed-loop formation. Therefore, taking into account our results already obtained, and the more abstract ideas presented in the preceding subsection, some future avenues of research are proposed below.

- The analyses described in chapter 5 should be repeated for the two additional replicates, to confirm the reproducibility of our findings.
- The IP experiment can be repeated with pull downs via tagged Pab1 and eIF4G1. Lists of 'high closed-loop genes' could be identified from pull downs of each factor, then these lists could be compared between pull downs to see the level of overlap.
- More correlations can be performed with this comparative metric we have devised for closed-loop formation. At the time of writing, correlation analyses with poly(A) tail length and sequence motifs are ongoing.
- This method can be extended to study closed-loop formation behaviour during glucose starvation in yeast. By comparing relative end enrichments in glucose-starved cells to end enrichments in non-starved cells, we may be able to observe mRNA-specific differences in closed-loop formation due to stress. Combining this new data with our research group's existing data on translational changes to glucose starvation,

we will be able to better understand if the closed-loop plays a role in cellular responses to stress

- A better understanding of the protein-protein and protein-mRNA contacts aiding circularisation of mRNA can be gained by repeating this study, but with mutation of the closed-loop factors to disrupt binding. Work is currently in progress to disrupt the eIF4G1-Pab1 interaction via a point mutation in the Pab1 binding domain of eIF4G1. A comparison of end enrichments between mutated and non-mutated strains could help to more directly confirm the role of the canonical closed-loop interaction. If the 3' end enrichment is still observed upon disruption of the eIF4G1-Pab1 interaction, this could suggest alternate contacts are enabling circularisation. Measuring the same correlations presented in this study for a closed-loop disrupted strain would also lend insights into what functional outputs the closed-loop is involved in.
- To study whether closed-loop formation is related to rounds of translation or polysomal composition, sucrose density gradient ultracentrifugation can be used to probe the closed-loop prevalence in different fractions (monosomes, disomes, etc).
- Prominent mRNAs showing either low or high levels of closed-loop formation can be visualised by the smFISH method available in the literature, to validate whether co-localisation of the ends is observed as would be expected from our results. Doing this could also allow us to better understand whether the closed-loop is dynamic or transient, by comparing molecules from the one mRNA, for examples where we expect high levels of closed-loop formation.

6.8 Conclusion and outlook

In eukaryotes, translation initiation is a complex and tightly controlled stage of protein synthesis, with its regulation and dysregulation heavily implicated in health and disease. Hence, a fundamental understanding of translation initiation characteristics is imperative for the evidence-based development of translation targeting therapeutics, and especially in recent years – RNA-based therapeutics. However, an aspect of translation initiation that has remained poorly understood is the formation and function of the closed-loop structure.

In this thesis, we have developed a method to study the closed-loop *in vivo* and transcriptome wide. We find that the closed-loop likely can form *in vivo*, but does not appear to be particularly prevalent. Some mRNAs appear to show greater closed-loop formation propensity, but we are yet to elucidate definitive reasons for this differential closed-loop formation. With the many unanswered questions that still exist regarding the closed-loop, we anticipate that the method and findings presented in this thesis can be capitalised upon to produce further insights into this elusive biological phenomenon.

7 References

- Adam, S.A. *et al.* (1986) 'mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence.', *Molecular and Cellular Biology*, 6(8), pp. 2932–2943.
- Adivarahan, S. *et al.* (2018) 'Spatial Organization of Single mRNPs at Different Stages of the Gene Expression Pathway Molecular Cell Spatial Organization of Single mRNPs at Different Stages of the Gene Expression Pathway', *Molecular Cell*, 72, pp. 727–738.e5. Available at: <https://doi.org/10.1016/j.molcel.2018.10.010>.
- Akiri, G. *et al.* (1998) 'Regulation of vascular endothelial growth factor (VEGF) expression is mediated by internal initiation of translation and alternative initiation of transcription', *Oncogene*, 17(2), pp. 227–236. Available at: <https://doi.org/10.1038/sj.onc.1202019>.
- Altmann, M. *et al.* (1985) 'Purification and characterization of protein synthesis initiation factor eIF-4E from the yeast *Saccharomyces cerevisiae*', *Biochemistry*, 24(22), pp. 6085–6089. Available at: <https://doi.org/10.1021/bi00343a009>.
- Altmann, M., Handschin, C. and Trachsel, H. (1987) 'mRNA cap-binding protein: cloning of the gene encoding protein synthesis initiation factor eIF-4E from *Saccharomyces cerevisiae*.', *Molecular and Cellular Biology*, 7(3), pp. 998–1003.
- Amrani, N. *et al.* (2008) 'Translation factors promote the formation of two states of the closed-loop mRNP', *Nature*, 453(7199), pp. 1276–1280. Available at: <https://doi.org/10.1038/nature06974>.
- Archer, S.K. *et al.* (2015) 'Probing the closed-loop model of mRNA translation in living cells', *RNA Biology*, 12(3), pp. 248–254. Available at: <https://doi.org/10.1080/15476286.2015.1017242>.
- Archer, S.K. *et al.* (2016) 'Dynamics of ribosome scanning and recycling revealed by translation complex profiling.', *Nature*, 535(7613), pp. 570–4. Available at: <https://doi.org/10.1038/nature18647>.
- Baer, B.W. and Kornberg, R.D. (1983) 'The protein responsible for the repeating structure of cytoplasmic poly(A)-ribonucleoprotein.', *Journal of Cell Biology*, 96(3), pp. 717–721. Available at: <https://doi.org/10.1083/jcb.96.3.717>.
- Banerjee, A.K. (1980) '5'-terminal cap structure in eucaryotic messenger ribonucleic acids', *Microbiological Reviews*, 44(2), pp. 175–205. Available at: <https://doi.org/10.1128/mr.44.2.175-205.1980>.
- Berset, C. *et al.* (2003) 'RNA-binding activity of translation initiation factor eIF4G1 from *Saccharomyces cerevisiae*', *RNA*, 9(7), pp. 871–880. Available at: <https://doi.org/10.1261/rna.5380903>.
- Blobel, G. (1971) 'Release, Identification, and Isolation of Messenger RNA from Mammalian Ribosomes', *Proceedings of the National Academy of Sciences*, 68(4), pp. 832–835. Available at: <https://doi.org/10.1073/pnas.68.4.832>.

Blobel, G. (1973) 'A Protein of Molecular Weight 78,000 Bound to the Polyadenylate Region of Eukaryotic Messenger RNAs', *Proceedings of the National Academy of Sciences*, 70(3), pp. 924–928. Available at: <https://doi.org/10.1073/pnas.70.3.924>.

Borman, A.M., Michel, Y.M. and Kean, K.M. (2000) 'Biochemical characterisation of cap–poly(A) synergy in rabbit reticulocyte lysates: the eIF4G–PABP interaction increases the functional affinity of eIF4E for the capped mRNA 5'-end', *Nucleic Acids Research*, 28(21), pp. 4068–4075. Available at: <https://doi.org/10.1093/nar/28.21.4068>.

Botstein, D., Chervitz, S.A. and Cherry, M. (1997) 'Yeast as a Model Organism', *Science*, 277(5330), pp. 1259–1260. Available at: <https://doi.org/10.1126/science.277.5330.1259>.

Botstein, D. and Fink, G.R. (2011) 'Yeast: An Experimental Organism for 21st Century Biology', *Genetics*, 189(3), pp. 695–704. Available at: <https://doi.org/10.1534/genetics.111.130765>.

Brambilla, M. *et al.* (2019) 'The *Saccharomyces cerevisiae* poly (A) binding protein (Pab1): Master regulator of mRNA metabolism and cell physiology', *Yeast*, 36(1), pp. 23–34. Available at: <https://doi.org/10.1002/yea.3347>.

Brenner, C. *et al.* (1988) 'CDC33 Encodes mRNA Cap-Binding Protein eIF-4E of *Saccharomyces cerevisiae*', *Molecular and Cellular Biology*, 8(8), pp. 3556–3559. Available at: <https://doi.org/10.1128/mcb.8.8.3556-3559.1988>.

de Breyne, S. *et al.* (2009) 'Direct functional interaction of initiation factor eIF4G with type 1 internal ribosomal entry sites', *Proceedings of the National Academy of Sciences*, 106(23), pp. 9197–9202. Available at: <https://doi.org/10.1073/pnas.0900153106>.

Brown, E.A. *et al.* (1992) 'Secondary structure of the 5' nontranslated regions of hepatitis C virus and pestivirus genomic RNAs', *Nucleic Acids Research*, 20(19), pp. 5041–5045. Available at: <https://doi.org/10.1093/nar/20.19.5041>.

Brune, C. *et al.* (2005) 'Yeast poly(A)-binding protein Pab1 shuttles between the nucleus and the cytoplasm and functions in mRNA export', *RNA*, 11(4), pp. 517–531. Available at: <https://doi.org/10.1261/rna.7291205>.

Chang, Y. and Huh, W.-K. (2018a) 'Ksp1-dependent phosphorylation of eIF4G modulates post-transcriptional regulation of specific mRNAs under glucose deprivation conditions', *Nucleic Acids Research*, 46(6), pp. 3047–3060. Available at: <https://doi.org/10.1093/nar/gky097>.

Chang, Y. and Huh, W.-K. (2018b) 'Ksp1-dependent phosphorylation of eIF4G modulates post-transcriptional regulation of specific mRNAs under glucose deprivation conditions', *Nucleic Acids Research*, 46(6), pp. 3047–3060. Available at: <https://doi.org/10.1093/nar/gky097>.

Christensen, A.K., Kahn, L.E. and Bourne, C.M. (1987) 'Circular polysomes predominate on the rough endoplasmic reticulum of somatotropes and mammatropes in the rat anterior pituitary', *The American Journal of Anatomy*, 178(1), pp. 1–10. Available at: <https://doi.org/10.1002/aja.1001780102>.

- Clarkson, B.K., Gilbert, W.V. and Doudna, J.A. (2010) 'Functional Overlap between eIF4G Isoforms in *Saccharomyces cerevisiae*', *PLOS ONE*, 5(2), p. e9114. Available at: <https://doi.org/10.1371/journal.pone.0009114>.
- Coldwell, M.J. *et al.* (2000) 'Initiation of Apaf-1 translation by internal ribosome entry', *Oncogene*, 19(7), pp. 899–905. Available at: <https://doi.org/10.1038/sj.onc.1203407>.
- Cooper, G.M. (2000) 'Translation of mRNA', in *The Cell: A Molecular Approach. 2nd edition*. Sinauer Associates. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK9849/> (Accessed: 25 August 2023).
- Costello, J. *et al.* (2015) 'Global mRNA selection mechanisms for translation initiation', *Genome Biology*, 16(1), p. 10. Available at: <https://doi.org/10.1186/s13059-014-0559-z>.
- Dalgarno, L. and Shine, J. (1973) 'Conserved Terminal Sequence in 18S rRNA May Represent Terminator Anticodons', *Nature New Biology*, 245(148), pp. 261–262. Available at: <https://doi.org/10.1038/newbio245261a0>.
- von Der Haar, T., Ball, P.D. and McCarthy, J.E. (2000) 'Stabilization of eukaryotic initiation factor 4E binding to the mRNA 5'-Cap by domains of eIF4G', *The Journal of Biological Chemistry*, 275(39), pp. 30551–30555. Available at: <https://doi.org/10.1074/jbc.M004565200>.
- Dickson, K.S. *et al.* (2001) 'Poly(A) Polymerase and the Regulation of Cytoplasmic Polyadenylation *', *Journal of Biological Chemistry*, 276(45), pp. 41810–41816. Available at: <https://doi.org/10.1074/jbc.M103030200>.
- Drazkowska, K. *et al.* (2022) '2'-O-Methylation of the second transcribed nucleotide within the mRNA 5' cap impacts the protein production level in a cell-specific manner and contributes to RNA immune evasion', *Nucleic Acids Research*, 50(16), pp. 9051–9071. Available at: <https://doi.org/10.1093/nar/gkac722>.
- Fekete, C.A. *et al.* (2007) 'N- and C-terminal residues of eIF1A have opposing effects on the fidelity of start codon selection', *The EMBO Journal*, 26(6), pp. 1602–1614. Available at: <https://doi.org/10.1038/sj.emboj.7601613>.
- Ford, L.P., Bagga, P.S. and Wilusz, J. (1997) 'The Poly(A) Tail Inhibits the Assembly of a 3'-to-5' Exonuclease in an In Vitro RNA Stability System', *Molecular and Cellular Biology*, 17(1), pp. 398–406. Available at: <https://doi.org/10.1128/MCB.17.1.398>.
- Fuke, H. and Ohno, M. (2008) 'Role of poly (A) tail as an identity element for mRNA nuclear export', *Nucleic Acids Research*, 36(3), pp. 1037–1049. Available at: <https://doi.org/10.1093/nar/gkm1120>.
- FURUICHI, Y. (2015) 'Discovery of m7G-cap in eukaryotic mRNAs', *Proceedings of the Japan Academy. Series B, Physical and Biological Sciences*, 91(8), pp. 394–409. Available at: <https://doi.org/10.2183/pjab.91.394>.
- Furuichi, Y., LaFiandra, A. and Shatkin, A.J. (1977) '5'-Terminal structure and mRNA stability', *Nature*, 266(5599), pp. 235–239. Available at: <https://doi.org/10.1038/266235a0>.

- Gallie, D.R. (1991) 'The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency', *Genes and Development*, 5(11), pp. 2108–2116. Available at: <https://doi.org/10.1101/gad.5.11.2108>.
- Galloway, A. and Cowling, V.H. (2019) 'mRNA cap regulation in mammalian cell function and fate', *Biochimica et Biophysica Acta. Gene Regulatory Mechanisms*, 1862(3), pp. 270–279. Available at: <https://doi.org/10.1016/j.bbagr.2018.09.011>.
- García-García, C. *et al.* (2015) 'Factor-dependent processivity in human eIF4A DEAD-box helicase', *Science*, 348(6242), pp. 1486–1488. Available at: <https://doi.org/10.1126/science.aaa5089>.
- Gasior, E. *et al.* (1979) 'The preparation and characterization of a cell-free system from *Saccharomyces cerevisiae* that translates natural messenger ribonucleic acid.', *Journal of Biological Chemistry*, 254(10), pp. 3965–3969. Available at: [https://doi.org/10.1016/S0021-9258\(18\)50681-5](https://doi.org/10.1016/S0021-9258(18)50681-5).
- Geisberg, J.V. *et al.* (2014) 'Global Analysis of mRNA Isoform Half-Lives Reveals Stabilizing and Destabilizing Elements in Yeast', *Cell*, 156(4), pp. 812–824. Available at: <https://doi.org/10.1016/j.cell.2013.12.026>.
- Gershon, P.D. *et al.* (1991) 'Poly(A) polymerase and a dissociable polyadenylation stimulatory factor encoded by vaccinia virus', *Cell*, 66(6), pp. 1269–1278. Available at: [https://doi.org/10.1016/0092-8674\(91\)90048-4](https://doi.org/10.1016/0092-8674(91)90048-4).
- Gingras, A.-C. *et al.* (1999) 'Regulation of 4E-BP1 phosphorylation: a novel two-step mechanism', *Genes & Development*, 13(11), pp. 1422–1437.
- Goffeau, A. *et al.* (1996) 'Life with 6000 Genes', *Science*, 274(5287), pp. 546–567. Available at: <https://doi.org/10.1126/science.274.5287.546>.
- Goldstein, A.L. and McCusker, J.H. (1999) 'Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*', *Yeast*, 15(14), pp. 1541–1553. Available at: [https://doi.org/10.1002/\(SICI\)1097-0061\(199910\)15:14<1541::AID-YEA476>3.0.CO;2-K](https://doi.org/10.1002/(SICI)1097-0061(199910)15:14<1541::AID-YEA476>3.0.CO;2-K).
- Goyer, C. *et al.* (1993) 'TIF4631 and TIF4632: two yeast genes encoding the high-molecular-weight subunits of the cap-binding protein complex (eukaryotic initiation factor 4F) contain an RNA recognition motif-like sequence and carry out an essential function', *Molecular and Cellular Biology*, 13(8), pp. 4860–4874. Available at: <https://doi.org/10.1128/mcb.13.8.4860-4874.1993>.
- Grange, T. *et al.* (1987) 'Human mRNA polyadenylate binding protein: evolutionary conservation of a nucleic acid binding motif', *Nucleic Acids Research*, 15(12), pp. 4771–4787. Available at: <https://doi.org/10.1093/nar/15.12.4771>.
- Grifo, J.A. *et al.* (1983) 'New initiation factor activity required for globin mRNA translation', *The Journal of Biological Chemistry*, 258(9), pp. 5804–5810.
- Gross, J.D. *et al.* (2003) 'Ribosome Loading onto the mRNA Cap Is Driven by Conformational Coupling between eIF4G and eIF4E', *Cell*, 115(6), pp. 739–750. Available at: [https://doi.org/10.1016/S0092-8674\(03\)00975-9](https://doi.org/10.1016/S0092-8674(03)00975-9).

- Grüner, S. *et al.* (2016) ‘The Structures of eIF4E-eIF4G Complexes Reveal an Extended Interface to Regulate Translation Initiation’, *Molecular Cell*, 64(3), pp. 467–479. Available at: <https://doi.org/10.1016/j.molcel.2016.09.020>.
- Haghighat, A. and Sonenberg, N. (1997) ‘eIF4G Dramatically Enhances the Binding of eIF4E to the mRNA 5'-Cap Structure *’, *Journal of Biological Chemistry*, 272(35), pp. 21677–21680. Available at: <https://doi.org/10.1074/jbc.272.35.21677>.
- Haizel, S.A. *et al.* (2020) ‘5'-UTR recruitment of the translation initiation factor eIF4GI or DAP5 drives cap-independent translation of a subset of human mRNAs’, *Journal of Biological Chemistry*, 295(33), pp. 11693–11706. Available at: <https://doi.org/10.1074/jbc.RA120.013678>.
- Hartley, A.D. *et al.* (1996) ‘Preparation and Use of Yeast Cell-Free Translation Lysate’, in I.H. Evans (ed.) *Yeast Protocols: Methods in Cell and Molecular Biology*. Totowa, NJ: Humana Press (Methods in Molecular Biology™), pp. 249–257. Available at: <https://doi.org/10.1385/0-89603-319-8:249>.
- Held, W.A., West, K. and Gallagher, J.F. (1977) ‘Importance of initiation factor preparations in the translation of reovirus and globin mRNAs lacking a 5'-terminal 7-methylguanosine’, *The Journal of Biological Chemistry*, 252(23), pp. 8489–8497.
- Hellen, C.U.T. (2018) ‘Translation Termination and Ribosome Recycling in Eukaryotes’, *Cold Spring Harbor Perspectives in Biology*, 10(10), p. a032656. Available at: <https://doi.org/10.1101/cshperspect.a032656>.
- Hellen, C.U.T. and Sarnow, P. (2001) ‘Internal ribosome entry sites in eukaryotic mRNA molecules’, *Genes & Development*, 15(13), pp. 1593–1612. Available at: <https://doi.org/10.1101/gad.891101>.
- Hentze, M.W. (1997) ‘eIF4G--A Multipurpose Ribosome Adapter?’, *Science*, 275(5299), pp. 500–501. Available at: <https://doi.org/10.1126/science.275.5299.500>.
- Hernan, R., Heuermann, K. and Brizzard, B. (2000) ‘Multiple Epitope Tagging of Expressed Proteins for Enhanced Detection’, *BioTechniques*, 28(4), pp. 789–793. Available at: <https://doi.org/10.2144/00284pf01>.
- Hershey, P.E.C. *et al.* (1999) ‘The Cap-binding Protein eIF4E Promotes Folding of a Functional Domain of Yeast Translation Initiation Factor eIF4G1*’, *Journal of Biological Chemistry*, 274(30), pp. 21297–21304. Available at: <https://doi.org/10.1074/jbc.274.30.21297>.
- Higgs, P.G. (2000) ‘RNA secondary structure: physical and computational aspects’, *Quarterly Reviews of Biophysics*, 33(3), pp. 199–253. Available at: <https://doi.org/10.1017/S0033583500003620>.
- Hinnebusch, A.G. (2014) ‘The scanning mechanism of eukaryotic translation initiation’, *Annual Review of Biochemistry*, 83, pp. 779–812. Available at: <https://doi.org/10.1146/annurev-biochem-060713-035802>.

- Hoffman, E.A. *et al.* (2015) 'Formaldehyde Crosslinking: A Tool for the Study of Chromatin Complexes', *The Journal of Biological Chemistry*, 290(44), pp. 26404–26411. Available at: <https://doi.org/10.1074/jbc.R115.651679>.
- Hoyle, N.P. *et al.* (2007) 'Stress-dependent relocalization of translationally primed mRNPs to cytoplasmic granules that are kinetically and spatially distinct from P-bodies', *Journal of Cell Biology*, 179(1), pp. 65–74. Available at: <https://doi.org/10.1083/jcb.200707010>.
- Hua, S. *et al.* (1997) 'Minimum Length of Sequence Homology Required for in Vivo Cloning by Homologous Recombination in Yeast', *Plasmid*, 38(2), pp. 91–96. Available at: <https://doi.org/10.1006/plas.1997.1305>.
- Humphreys, D.T. *et al.* (2005) 'MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function', *Proceedings of the National Academy of Sciences*, 102(47), pp. 16961–16966. Available at: <https://doi.org/10.1073/pnas.0506482102>.
- Imataka, H. and Sonenberg, N. (1997) 'Human Eukaryotic Translation Initiation Factor 4G (eIF4G) Possesses Two Separate and Independent Binding Sites for eIF4A', *Molecular and Cellular Biology*, 17(12), pp. 6940–6947. Available at: <https://doi.org/10.1128/MCB.17.12.6940>.
- Jackson, R.J., Hellen, C.U.T. and Pestova, T.V. (2010) 'The mechanism of eukaryotic translation initiation and principles of its regulation', *Nature Reviews Molecular Cell Biology*, 11(2), pp. 113–127. Available at: <https://doi.org/10.1038/nrm2838>.
- Jackson, R.J. and Standart, N. (1990) 'Do the poly(A) tail and 3' untranslated region control mRNA translation?', *Cell*, 62(1), pp. 15–24. Available at: [https://doi.org/10.1016/0092-8674\(90\)90235-7](https://doi.org/10.1016/0092-8674(90)90235-7).
- Jacobson, A. and Favreau, M. (1983) 'Possible involvement of poly(A) in protein synthesis.', *Nucleic Acids Research*, 11(18), pp. 6353–6368.
- Jang, S.K. and Wimmer, E. (1990) 'Cap-independent translation of encephalomyocarditis virus RNA: structural elements of the internal ribosomal entry site and involvement of a cellular 57-kD RNA-binding protein.', *Genes & Development*, 4(9), pp. 1560–1572. Available at: <https://doi.org/10.1101/gad.4.9.1560>.
- Jiang Q., Lin L. and Wang T. (2006) 'A New Model for Apoptosis Research: Yeast', *Progress in Biochemistry and Biophysics* [Preprint]. Available at: <http://dx.doi.org/> (Accessed: 27 August 2023).
- Johannes, G. and Sarnow, P. (1998) 'Cap-independent polysomal association of natural mRNAs encoding c-myc, BiP, and eIF4G conferred by internal ribosome entry sites', *RNA (New York, N.Y.)*, 4(12), pp. 1500–1513. Available at: <https://doi.org/10.1017/s1355838298981080>.
- Kang, Y.-N. *et al.* (2015) 'Genome-wide profiling of untranslated regions by paired-end ditag sequencing reveals unexpected transcriptome complexity in yeast', *Molecular Genetics and Genomics*, 290(1), pp. 217–224. Available at: <https://doi.org/10.1007/s00438-014-0913-6>.

- Kapp, L.D. and Lorsch, J.R. (2004) 'The Molecular Mechanics of Eukaryotic Translation', *Annual Review of Biochemistry*, 73(1), pp. 657–704. Available at: <https://doi.org/10.1146/annurev.biochem.73.030403.080419>.
- Kessler, S.H. and Sachs, A.B. (1998) 'RNA Recognition Motif 2 of Yeast Pab1p Is Required for Its Functional Interaction with Eukaryotic Translation Initiation Factor 4G', *Molecular and Cellular Biology*, 18(1), pp. 51–57. Available at: <https://doi.org/10.1128/mcb.18.1.51>.
- Khong, A. and Parker, R. (2018) 'mRNP architecture in translating and stress conditions reveals an ordered pathway of mRNP compaction', *The Journal of Cell Biology*, 217(12), pp. 4124–4140. Available at: <https://doi.org/10.1083/jcb.201806183>.
- Kozak, M. (1984) 'Selection of initiation sites by eucaryotic ribosomes: effect of inserting AUG triplets upstream from the coding sequence for preproinsulin.', *Nucleic Acids Research*, 12(9), pp. 3873–3893.
- Kozak, M. (1989) 'The scanning model for translation: an update.', *Journal of Cell Biology*, 108(2), pp. 229–241. Available at: <https://doi.org/10.1083/jcb.108.2.229>.
- Kozak, M. (1999) 'Initiation of translation in prokaryotes and eukaryotes', *Gene*, 234(2), pp. 187–208. Available at: [https://doi.org/10.1016/S0378-1119\(99\)00210-3](https://doi.org/10.1016/S0378-1119(99)00210-3).
- Lai, W.J.C. *et al.* (2018) 'mRNAs and lncRNAs intrinsically form secondary structures with short end-to-end distances', *Nature Communications*, 9(1), pp. 1–11. Available at: <https://doi.org/10.1038/s41467-018-06792-z>.
- Laurberg, M. *et al.* (2008) 'Structural basis for translation termination on the 70S ribosome', *Nature*, 454(7206), pp. 852–857. Available at: <https://doi.org/10.1038/nature07115>.
- Lee, J.H. *et al.* (2002) 'Initiation factor eIF5B catalyzes second GTP-dependent step in eukaryotic translation initiation', *Proceedings of the National Academy of Sciences*, 99(26), pp. 16689–16694. Available at: <https://doi.org/10.1073/pnas.262569399>.
- LeFebvre, A.K. *et al.* (2006) 'Translation Initiation Factor eIF4G-1 Binds to eIF3 through the eIF3e Subunit *', *Journal of Biological Chemistry*, 281(32), pp. 22917–22932. Available at: <https://doi.org/10.1074/jbc.M605418200>.
- Leija-Martínez, N., Casas-Flores, S., Cadena-Nava, Rubén D., *et al.* (2014) 'The separation between the 5'-3' ends in long RNA molecules is short and nearly constant', *Nucleic Acids Research*, 42(22), pp. 13963–13968. Available at: <https://doi.org/10.1093/nar/gku1249>.
- Leija-Martínez, N., Casas-Flores, S., Cadena-Nava, Rubén D., *et al.* (2014) 'The separation between the 5-3 ends in long RNA molecules is short and nearly constant', *Nucleic Acids Research*, 42, pp. 13963–13968. Available at: <https://doi.org/10.1093/nar/gku1249>.
- Linder, P. (1992) 'Molecular biology of translation in yeast', *Antonie van Leeuwenhoek*, 62(1), pp. 47–62. Available at: <https://doi.org/10.1007/BF00584462>.
- Lindmark, R., Thorén-Tolling, K. and Sjöquist, J. (1983) 'Binding of immunoglobulins to protein A and immunoglobulin levels in mammalian sera', *Journal of Immunological Methods*, 62(1), pp. 1–13. Available at: [https://doi.org/10.1016/0022-1759\(83\)90104-7](https://doi.org/10.1016/0022-1759(83)90104-7).

- Liu, W. *et al.* (2017) 'From *Saccharomyces cerevisiae* to human: The important gene co-expression modules', *Biomedical Reports*, 7(2), pp. 153–158. Available at: <https://doi.org/10.3892/br.2017.941>.
- Lomakin, I.B. and Steitz, T.A. (2013) 'The initiation of mammalian protein synthesis and mRNA scanning mechanism', *Nature*, 500(7462), pp. 307–311. Available at: <https://doi.org/10.1038/nature12355>.
- López de Quinto, S. and Martínez-Salas, E. (2000) 'Interaction of the eIF4G initiation factor with the aphthovirus IRES is essential for internal translation initiation in vivo.', *RNA*, 6(10), pp. 1380–1392.
- Lorsch, J.R. and Dever, T.E. (2010) 'Molecular View of 43 S Complex Formation and Start Site Selection in Eukaryotic Translation Initiation *', *Journal of Biological Chemistry*, 285(28), pp. 21203–21207. Available at: <https://doi.org/10.1074/jbc.R110.119743>.
- Mandel, C.R. *et al.* (2006) 'Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease', *Nature*, 444(7121), pp. 953–956. Available at: <https://doi.org/10.1038/nature05363>.
- Marcotrigiano, J. *et al.* (1997) 'Cocrystal Structure of the Messenger RNA 5' Cap-Binding Protein (eIF4E) Bound to 7-methyl-GDP', *Cell*, 89(6), pp. 951–961. Available at: [https://doi.org/10.1016/S0092-8674\(00\)80280-9](https://doi.org/10.1016/S0092-8674(00)80280-9).
- Marcotrigiano, J. *et al.* (1999) 'Cap-Dependent Translation Initiation in Eukaryotes Is Regulated by a Molecular Mimic of eIF4G', *Molecular Cell*, 3(6), pp. 707–716. Available at: [https://doi.org/10.1016/S1097-2765\(01\)80003-4](https://doi.org/10.1016/S1097-2765(01)80003-4).
- Marintchev, A. and Wagner, G. (2004) 'Translation initiation: structures, mechanisms and evolution', *Quarterly Reviews of Biophysics*, 37(3–4), pp. 197–284. Available at: <https://doi.org/10.1017/S0033583505004026>.
- Marzluff, W.F. (2005) 'Metazoan replication-dependent histone mRNAs: a distinct set of RNA polymerase II transcripts', *Current Opinion in Cell Biology*, 17(3), pp. 274–280. Available at: <https://doi.org/10.1016/j.ceb.2005.04.010>.
- Matsuo, H. *et al.* (1997) 'Structure of translation factor eIF4E bound to m7GDP and interaction with 4E-binding protein', *Nature Structural Biology*, 4(9), pp. 717–724. Available at: <https://doi.org/10.1038/nsb0997-717>.
- Melamed, D. *et al.* (2013) 'Deep mutational scanning of an RRM domain of the *Saccharomyces cerevisiae* poly(A)-binding protein', *RNA*, 19(11), pp. 1537–1551. Available at: <https://doi.org/10.1261/rna.040709.113>.
- Merrick, W.C. (1979) 'Evidence that a single GTP is used in the formation of 80 S initiation complexes', *The Journal of Biological Chemistry*, 254(10), pp. 3708–3711.
- Merrick, W.C. (2004) 'Cap-dependent and cap-independent translation in eukaryotic systems', *Gene*, 332, pp. 1–11. Available at: <https://doi.org/10.1016/j.gene.2004.02.051>.
- Merrick, W.C. (2015) 'eIF4F: A Retrospective', *The Journal of Biological Chemistry*, 290(40), pp. 24091–24099. Available at: <https://doi.org/10.1074/jbc.R115.675280>.

- Miras, M. *et al.* (2017) 'Structure of eIF4E in Complex with an eIF4G Peptide Supports a Universal Bipartite Binding Mode for Protein Translation[OPEN]', *Plant Physiology*, 174(3), pp. 1476–1491. Available at: <https://doi.org/10.1104/pp.17.00193>.
- Mohammadi, S. *et al.* (2015) 'Scope and limitations of yeast as a model organism for studying human tissue-specific pathways', *BMC Systems Biology*, 9(1), p. 96. Available at: <https://doi.org/10.1186/s12918-015-0253-0>.
- Munroe, D. and Jacobson, A. (1990) 'mRNA poly(A) tail, a 3' enhancer of translational initiation.', *Molecular and Cellular Biology*, 10(7), pp. 3441–3455.
- Murthy, K.G. and Manley, J.L. (1995) 'The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation.', *Genes & Development*, 9(21), pp. 2672–2683. Available at: <https://doi.org/10.1101/gad.9.21.2672>.
- Muthukrishnan, S. *et al.* (1975) '5'-Terminal 7-methylguanosine in eukaryotic mRNA is required for translation', *Nature*, 255(5503), pp. 33–37. Available at: <https://doi.org/10.1038/255033a0>.
- Muthukrishnan, S. *et al.* (1978) 'Influence of 5'-terminal cap structure on the initiation of translation of vaccinia virus mRNA', *The Journal of Biological Chemistry*, 253(5), pp. 1710–1715.
- Nanbru, C. *et al.* (1997) 'Alternative translation of the proto-oncogene c-myc by an internal ribosome entry site', *The Journal of Biological Chemistry*, 272(51), pp. 32061–32066. Available at: <https://doi.org/10.1074/jbc.272.51.32061>.
- Neve, J. *et al.* (2017) 'Cleavage and polyadenylation: Ending the message expands gene regulation', *RNA Biology*, 14(7), pp. 865–890. Available at: <https://doi.org/10.1080/15476286.2017.1306171>.
- O'Leary, S.E. *et al.* (2013) 'Dynamic Recognition of the mRNA Cap by *Saccharomyces cerevisiae* eIF4E', *Structure*, 21(12), pp. 2197–2207. Available at: <https://doi.org/10.1016/j.str.2013.09.016>.
- Orr-Weaver, T.L., Szostak, J.W. and Rothstein, R.J. (1981) 'Yeast transformation: a model system for the study of recombination', *Proceedings of the National Academy of Sciences of the United States of America*, 78(10), pp. 6354–6358. Available at: <https://doi.org/10.1073/pnas.78.10.6354>.
- Park, E.H., Zhang, F., *et al.* (2011) 'Depletion of eIF4G from yeast cells narrows the range of translational efficiencies genome-wide', *BMC Genomics*, 12(1), p. 68. Available at: <https://doi.org/10.1186/1471-2164-12-68>.
- Park, E.H., Walker, S.E., *et al.* (2011) 'Multiple elements in the eIF4G1 N-terminus promote assembly of eIF4G1•PABP mRNPs in vivo', *The EMBO Journal*, 30(2), p. 302. Available at: <https://doi.org/10.1038/EMBOJ.2010.312>.
- Passmore, L.A. *et al.* (2007) 'The Eukaryotic Translation Initiation Factors eIF1 and eIF1A Induce an Open Conformation of the 40S Ribosome', *Molecular Cell*, 26(1), pp. 41–50. Available at: <https://doi.org/10.1016/j.molcel.2007.03.018>.

- Passmore, L.A. and Collier, J. (2021) 'Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression', *Nature Reviews Molecular Cell Biology* 2021, pp. 1–14. Available at: <https://doi.org/10.1038/s41580-021-00417-y>.
- Paterson, B.M. and Rosenberg, M. (1979) 'Efficient translation of prokaryotic mRNAs in a eukaryotic cell-free system requires addition of a cap structure', *Nature*, 279(5715), pp. 692–696. Available at: <https://doi.org/10.1038/279692a0>.
- Pelechano, V., Wei, W. and Steinmetz, L.M. (2013) 'Extensive transcriptional heterogeneity revealed by isoform profiling', *Nature*, 497(7447), pp. 127–131. Available at: <https://doi.org/10.1038/nature12121>.
- Pelletier, J. and Sonenberg, N. (1988) 'Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA', *Nature*, 334(6180), pp. 320–325. Available at: <https://doi.org/10.1038/334320a0>.
- Pestova, T.V. and Kolupaeva, V.G. (2002) 'The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection', *Genes & Development*, 16(22), pp. 2906–2922. Available at: <https://doi.org/10.1101/gad.1020902>.
- Peterson, D.T., Safer, B. and Merrick, W.C. (1979) 'Role of eukaryotic initiation factor 5 in the formation of 80 S initiation complexes', *The Journal of Biological Chemistry*, 254(16), pp. 7730–7735.
- Preiss, T. and Hentze, M.W. (1998) 'Dual function of the messenger RNA cap structure in poly(A)-tail- promoted translation in yeast', *Nature*, 392(6675), pp. 516–520. Available at: <https://doi.org/10.1038/33192>.
- Prévôt, D., Darlix, J.-L. and Ohlmann, T. (2003) 'Conducting the initiation of protein synthesis: the role of eIF4G', *Biology of the Cell*, 95(3–4), pp. 141–156. Available at: [https://doi.org/10.1016/S0248-4900\(03\)00031-5](https://doi.org/10.1016/S0248-4900(03)00031-5).
- Ptushkina, M. *et al.* (1998) 'Cooperative modulation by eIF4G of eIF4E-binding to the mRNA 5' cap in yeast involves a site partially shared by p20', *The EMBO Journal*, 17(16), pp. 4798–4808. Available at: <https://doi.org/10.1093/emboj/17.16.4798>.
- Ramanathan, A., Robb, G.B. and Chan, S.-H. (2016) 'mRNA capping: biological functions and applications', *Nucleic Acids Research*, 44(16), pp. 7511–7526. Available at: <https://doi.org/10.1093/nar/gkw551>.
- Riba, A. *et al.* (2019) 'Protein synthesis rates and ribosome occupancies reveal determinants of translation elongation rates', *Proceedings of the National Academy of Sciences*, 116(30), pp. 15023–15032. Available at: <https://doi.org/10.1073/pnas.1817299116>.
- Richter, J.D. and Sonenberg, N. (2005) 'Regulation of cap-dependent translation by eIF4E inhibitory proteins'. Available at: <https://doi.org/10.1038/nature03205>.
- Rogers, G.W., Lima, W.F. and Merrick, W.C. (2001) 'Further Characterization of the Helicase Activity of eIF4A: SUBSTRATE SPECIFICITY *', *Journal of Biological Chemistry*, 276(16), pp. 12598–12608. Available at: <https://doi.org/10.1074/jbc.M007560200>.

Rojas-Duran, M.F. and Gilbert, W.V. (2012) 'Alternative transcription start site selection leads to large differences in translation activity in yeast', *RNA*, 18(12), pp. 2299–2305. Available at: <https://doi.org/10.1261/rna.035865.112>.

Ross-Kaschitzka, D. and Altmann, M. (2020) 'eIF4E and Interactors from Unicellular Eukaryotes', *International Journal of Molecular Sciences*, 21(6), p. 2170. Available at: <https://doi.org/10.3390/ijms21062170>.

Sachs, A. (1990) 'The role of poly(A) in the translation and stability of mRNA', *Current Opinion in Cell Biology*, 2(6), pp. 1092–1098. Available at: [https://doi.org/10.1016/0955-0674\(90\)90161-7](https://doi.org/10.1016/0955-0674(90)90161-7).

Sachs, A.B., Bond, M.W. and Kornberg, R.D. (1986) 'A single gene from yeast for both nuclear and cytoplasmic polyadenylate-binding proteins: Domain structure and expression', *Cell*, 45(6), pp. 827–835. Available at: [https://doi.org/10.1016/0092-8674\(86\)90557-X](https://doi.org/10.1016/0092-8674(86)90557-X).

Sachs, A.B., Davis, R.W. and Kornberg, R.D. (1987) 'A single domain of yeast poly(A)-binding protein is necessary and sufficient for RNA binding and cell viability.', *Molecular and cellular biology*, 7(9), pp. 3268–3276. Available at: <https://doi.org/10.1128/mcb.7.9.3268-3276.1987>.

Sachs, A.B., Sarnow, P. and Hentze, M.W. (1997) 'Starting at the Beginning, Middle, and End: Translation Initiation in Eukaryotes', *Cell*, 89(6), pp. 831–838. Available at: [https://doi.org/10.1016/S0092-8674\(00\)80268-8](https://doi.org/10.1016/S0092-8674(00)80268-8).

Sarnow, P. (1989) 'Translation of glucose-regulated protein 78/immunoglobulin heavy-chain binding protein mRNA is increased in poliovirus-infected cells at a time when cap-dependent translation of cellular mRNAs is inhibited.', *Proceedings of the National Academy of Sciences*, 86(15), pp. 5795–5799. Available at: <https://doi.org/10.1073/pnas.86.15.5795>.

Sawazaki, R. *et al.* (2018) 'Characterization of the multimeric structure of poly(A)-binding protein on a poly(A) tail', *Scientific Reports*, 8(1), p. 1455. Available at: <https://doi.org/10.1038/s41598-018-19659-6>.

Shirokikh, N.E. and Preiss, T. (2018) 'Translation initiation by cap-dependent ribosome recruitment: Recent insights and open questions', *WIREs RNA*, 9(4), p. e1473. Available at: <https://doi.org/10.1002/wrna.1473>.

Sonenberg, N. *et al.* (1978) 'A polypeptide in eukaryotic initiation factors that crosslinks specifically to the 5'-terminal cap in mRNA.', *Proceedings of the National Academy of Sciences of the United States of America*, 75(10), pp. 4843–4847.

Sonenberg, N. *et al.* (1979) 'Eukaryotic mRNA cap binding protein: purification by affinity chromatography on sepharose-coupled m7GDP.', *Proceedings of the National Academy of Sciences*, 76(9), pp. 4345–4349. Available at: <https://doi.org/10.1073/pnas.76.9.4345>.

Sonenberg, N. *et al.* (1980) 'Differential stimulation of capped mRNA translation in vitro by cap binding protein', *Nature*, 285(5763), pp. 331–333. Available at: <https://doi.org/10.1038/285331a0>.

- Sonenberg, N. (2008) 'eIF4E, the mRNA cap-binding protein: from basic discovery to translational research', *Biochemistry and Cell Biology*, 86(2), pp. 178–183. Available at: <https://doi.org/10.1139/O08-034>.
- Sonenberg, N. and Gingras, A.-C. (1998) 'The mRNA 5' cap-binding protein eIF4E and control of cell growth', *Current Opinion in Cell Biology*, 10(2), pp. 268–275. Available at: [https://doi.org/10.1016/S0955-0674\(98\)80150-6](https://doi.org/10.1016/S0955-0674(98)80150-6).
- Sonenberg, N. and Hinnebusch, A.G. (2009) 'Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets', *Cell*, 136(4), pp. 731–745. Available at: <https://doi.org/10.1016/j.cell.2009.01.042>.
- Steitz, J.A. and Jakes, K. (1975) 'How ribosomes select initiator regions in mRNA: base pair formation between the 3' terminus of 16S rRNA and the mRNA during initiation of protein synthesis in Escherichia coli.', *Proceedings of the National Academy of Sciences of the United States of America*, 72(12), pp. 4734–4738.
- Su, K.-H. and Dai, C. (2017) 'mTORC1 senses stresses: Coupling stress to proteostasis', *BioEssays : news and reviews in molecular, cellular and developmental biology*, 39(5), p. 10.1002/bies.201600268. Available at: <https://doi.org/10.1002/bies.201600268>.
- Tarun, Salvador Z *et al.* (1997) *Translation initiation factor eIF4G mediates in vitro poly(A) tail-dependent translation*, pp. 9046–9051. Available at: www.pnas.org. (Accessed: 3 April 2020).
- Tarun, Salvador Z. *et al.* (1997) 'Translation initiation factor eIF4G mediates in vitro poly(A) tail-dependent translation', *Proceedings of the National Academy of Sciences*, 94(17), pp. 9046–9051. Available at: <https://doi.org/10.1073/pnas.94.17.9046>.
- Tarun, S.Z. and Sachs, A.B. (1995) 'A common function for mRNA 5' and 3' ends in translation initiation in yeast', *Genes and Development*, 9(23), pp. 2997–3007. Available at: <https://doi.org/10.1101/gad.9.23.2997>.
- Tarun, S.Z. and Sachs, A.B. (1996) 'Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G.', *The EMBO Journal*, 15(24), pp. 7168–7177.
- Thavarajah, R. *et al.* (2012) 'Chemical and physical basics of routine formaldehyde fixation', *Journal of Oral and Maxillofacial Pathology : JOMFP*, 16(3), pp. 400–405. Available at: <https://doi.org/10.4103/0973-029X.102496>.
- Thompson, M.K. and Gilbert, W.V. (2017) 'mRNA length-sensing in eukaryotic translation: reconsidering the “closed loop” and its implications for translational control', *Current Genetics*, 63(4), pp. 613–620. Available at: <https://doi.org/10.1007/s00294-016-0674-3>.
- Tian, B. and Manley, J.L. (2017) 'Alternative polyadenylation of mRNA precursors', *Nature Reviews Molecular Cell Biology*, 18(1), pp. 18–30. Available at: <https://doi.org/10.1038/nrm.2016.116>.
- Tudek, A. *et al.* (2021) 'Global view on the metabolism of RNA poly(A) tails in yeast *Saccharomyces cerevisiae*', *Nature Communications*, 12(1), p. 4951. Available at: <https://doi.org/10.1038/s41467-021-25251-w>.

- Uchida, N. *et al.* (2002) 'A Novel Role of the Mammalian GSPT/eRF3 Associating with Poly(A)-binding Protein in Cap/Poly(A)-dependent Translation *', *Journal of Biological Chemistry*, 277(52), pp. 50286–50292. Available at: <https://doi.org/10.1074/jbc.M203029200>.
- Velculescu, V.E. *et al.* (1997) 'Characterization of the Yeast Transcriptome', *Cell*, 88(2), pp. 243–251. Available at: [https://doi.org/10.1016/S0092-8674\(00\)81845-0](https://doi.org/10.1016/S0092-8674(00)81845-0).
- Vicens, Q., Kieft, J.S. and Rissland, O.S. (2018) 'Revisiting the Closed-Loop Model and the Nature of mRNA 5'–3' Communication', *Molecular Cell*, 72(5), pp. 805–812. Available at: <https://doi.org/10.1016/j.molcel.2018.10.047>.
- Voorhees, R.M. and Ramakrishnan, V. (2013) 'Structural Basis of the Translational Elongation Cycle', *Annual Review of Biochemistry*, 82(1), pp. 203–236. Available at: <https://doi.org/10.1146/annurev-biochem-113009-092313>.
- Wang, X., Xi, W., Toomey, S., Chiang, Y.-C., Hasek, J., Laue, Thomas M., *et al.* (2016) 'Stoichiometry and Change of the mRNA Closed-Loop Factors as Translating Ribosomes Transit from Initiation to Elongation', *PLOS ONE*, 11(3), p. e0150616. Available at: <https://doi.org/10.1371/journal.pone.0150616>.
- Wei, C.-C. *et al.* (1998) 'Wheat Germ Poly(A) Binding Protein Enhances the Binding Affinity of Eukaryotic Initiation Factor 4F and (iso)4F for Cap Analogues', *Biochemistry*, 37(7), pp. 1910–1916. Available at: <https://doi.org/10.1021/bi9724570>.
- Wells, S.E. *et al.* (1998) 'Circularization of mRNA by eukaryotic translation initiation factors', *Molecular Cell*, 2(1), pp. 135–140. Available at: [https://doi.org/10.1016/S1097-2765\(00\)80122-7](https://doi.org/10.1016/S1097-2765(00)80122-7).
- Wodnar-Filipowicz, A. *et al.* (1978) '5'-Terminal 7-Methylguanosine and mRNA Function', *European Journal of Biochemistry*, 92(1), pp. 69–80. Available at: <https://doi.org/10.1111/j.1432-1033.1978.tb12724.x>.
- Yoffe, A.M. *et al.* (2011) 'The ends of a large RNA molecule are necessarily close', *Nucleic Acids Research*, 39(1), pp. 292–299. Available at: <https://doi.org/10.1093/nar/gkq642>.