

Deciphering Messenger Ribonucleoprotein Composition Using Single Molecule Resolution Microscopy

Mathew Kramar

Department of Experimental Medicine

McGill University

Montreal, QC

September 2018

A thesis submitted to McGill University in partial fulfillment of the requirements
of the degree of Master of Science under the Faculty of Medicine

© MATHEW KRAMAR, 2018

Abstract:

Messenger RNA (mRNA) molecules expressed in the nuclei of eukaryotic organisms associate with many proteins to facilitate processing and maturation of mRNA, as well as packaging into messenger ribonucleoprotein particle (mRNP) complexes to facilitate nuclear export of transcripts. Existing methods used to study mRNPs have succeeded in identifying a number of proteins involved in mRNP biogenesis, but have failed to provide information with regards to mRNP structure or the heterogeneous pool of mRNPs existing in the cell. Here we describe steps taken towards the development of a novel technique aimed at quantifying protein abundance, in terms of stoichiometry, of mRNP proteins on specifically purified mRNA transcripts from cellular extracts by affinity purification and single-molecule microscopy. In developing this method, we aim to investigate the influence of mRNA length on mRNP protein stoichiometry, and to determine how mRNAs of different lengths may use the pool of mRNP biogenesis factors differently to achieve effective nuclear export.

Résumé:

Les molécules d'ARN messager (ARNm) exprimées dans les noyaux des organismes eucaryotes s'associent à certaines protéines pour faciliter le traitement et la maturation de l'ARNm, ainsi que pour former des complexes de particules ribonucléoprotéines messagères (RNPm) pour faciliter l'exportation hors du noyau. Les méthodes existantes utilisées pour étudier les RNPm ont réussi à identifier un certain nombre de protéines impliquées dans la biogenèse de la RNPm, mais ne parviennent pas à fournir des informations sur la structure de la RNPm ou l'hétérogénéité des RNPm qui existent dans la cellule. Nous décrivons ici les mesures prises pour le développement d'une nouvelle technique visant à quantifier l'abondance des protéines, en termes de stœchiométrie, des RNPm sur des ARNm spécifiquement purifiés à partir d'extraits cellulaires par purification d'affinité et microscopie à molécule unique. En développant cette méthode, nous cherchons à étudier l'influence de la longueur de l'ARNm sur la stœchiométrie des RNPm, afin de voir comment les ARNm de différentes longueurs peuvent utiliser différemment les facteurs de biogenèse de la RNPm pour faciliter l'exportation hors du noyau.

Table of Contents

Abstract:	2
Résumé:.....	2
Acknowledgements.....	10
Author Contributions	11
Introduction.....	12
The basics of mRNA, RBPs and mRNPs:	13
The basics of transcription:	16
5' Cap Binding Complex.....	18
THO complex	22
THO function in Metazoa.....	24
TREX	25
Mex67-Mtr2/TAP-p15/NXF1-NXT1 recruitment	27
Yra1 and Aly/REF.....	27
Hpr1	28
Nab2.....	28
Npl3	29
TREX-2	33
TOR regulation of mRNP Methylation.....	36
3' End Processing.....	39
Nuclear Export	44
mRNP Heterogeneity	46
Introns	46
3' Processing.....	47
Localization	48
Research Project.....	49
Project justification	50
Single-molecule pull-down and imaging methodology	55
ssAP from Cryolysate	59
PP7 RNA Tagging System	60
SNAP and CLIP site specific fluorescent protein labelling	61
Single-molecule pull-down experiments	65

Materials and methods	67
Yeast strains:	68
Yeast Media:	69
PCR:	69
Yeast tagging:.....	69
Competent Yeast Cells	70
Yeast transformation protocol.....	70
Screening of yeast colonies	70
Yeast chromosomal extraction	71
Separation of DNA Fragments.....	71
Total yeast protein extraction.....	72
SDS-PAGE.....	72
Western blotting	72
Yeast spotting Assay	73
Cell harvest and cryo-lysis	74
Dynabead conjugation with rabbit IgG	76
Single step affinity Protein A purification	76
Silver staining of proteins in polyacrylamide gels	77
RT-PCR.....	77
Native Protein A Elution.....	78
Bead Saturation binding assay	79
Targeted Bradford Assay	79
SNAP and CLIP dye preparation	80
Relative time course labelling assay	81
Mass spec labelling efficiency assay.....	82
Plasmid cloning.....	83
PSACH recombinant protein purification.	84
NTA Coverslips.....	86
NTA PEG coverslips.....	87
IgG Coverslips.....	87
Gel Filtration of mRNP complexes.....	88
Chapter 1: Specific mRNP subunit counting	89

Purification of transcript specific mRNPs	90
Native elution of PrA bound complexes.....	93
Labelling of SNAP and CLIP tagged constructs	96
Optimization of slide wash buffer	97
Imaging of cap binding complex on PMA1 transcripts.....	100
Relative labelling efficiencies of SNAP and CLIP tags	106
Chapter 2: General mRNP subunit counting	111
Quantification of isolated bait protein	112
Absolution quantification of labelling efficiency by mass spectrometry	115
Single-molecule imaging of Pab1 on Cbp80 purified complexes	118
Imaging of Cbp80 purified complexes on PrA coverslips	122
Separation of Cbp80 purified mRNPs by SEC.....	125
Pab1-CLIP-AP binding assay	128
Discussion	130
Conclusion	137
References.....	139

List of Figures

Figure 1. Co-transcriptional formation of mRNPs.....	17
Figure 2. Assembly and rearrangement of mRNP export factors	
Figure 3. mRNP NPC docking and quality control.	45
Figure 4. Specific transcript quantification vs. general transcript quantification goals.....	54
Figure 5. Schematic representation of methodology used for transcript specific protein quantification.	57
Figure 6. Schematic representation of SNAP and CLIP labelling reactions.	64
Figure 7. Purification of PMA1 specific mRNPs.	92
Figure 8. Native elution of PrA bound complexes with PEGyLOX.	94
Figure 9. Functional labelling of SNAP and CLIP tagged complexes	95
Figure 10. Non-specific binding of PP7-SNP-PrA to coverslips in buffers of different stringency.	98
Figure 11. Imaging of CBC on single purified PMA1 mRNPs by TIRF.	101
Figure 12. Single-molecule imaging of PSACH construct purified from yeast.	103
Figure 13. Single-molecule imaging of labelled rPSACH.....	105
Figure 14. Relative time course labelling assay of rPSACH in solution or on IgG conjugated dynabeads.....	108
Figure 15. Single-molecule imaging of rPSACH labelled for 90 minutes	109
Figure 16. Relative time course labelling assay of Cbp80-CLIP-His10 and PP7-SNAP-PrA in solution or on IgG conjugated dynabeads.....	110
Figure 17. Quantification of purified protein by targeted Bradford assay.....	114
Figure 18. Absolute quantification of SNAP-tag labelling by mass spectrometry.....	117
Figure 19. Single-molecule imaging of Cbp80 and Pab1 bound mRNPs on Ni-NTA coverslips.	120
Figure 20. Pab1-CLIP-His10 binding assay to Ni ²⁺ beads.....	121
Figure 21. Single-molecule imaging of Cbp80 and Pab1 bound mRNPs on IgG coverslips	124
Figure 22. Pab1-CLIP-AP binding assay with monomeric avidin beads	129

List of abbreviations used

5': five prime

5'm7G: 5' 7-methylguanosine cap

Amp: ampicillin

AP: Acceptor peptide

AP-MS: Affinity purification mass spectrometry

bp: base pair

CBC: cap-binding complex

CFIA: cleavage factor I A

ChIP: chromatin immunoprecipitation

CPF: cleavage and polyadenylation factor

CTD :C-terminal domain

DNA: deoxyribonucleic acid

EJC: exon junction complex

EtOH: ethanol

G418: geneticin

His5: histidine gene

hnRNP: heterogeneous ribonucleoprotein particle

IgG: immunoglobulin gamma

IMAC: immobilized metal affinity chromatography

IPTG: isopropylthio- β -galactoside

KanMX: yeast kanamycin gene

Kb: kilobases

kDa: kilodalton

LB: Lysogeny broth

mg: milligram

mRNA: messenger ribonucleic acid
mRNP: messenger ribonucleoprotein
MS: mass spectrometry
nm: nanometer
NPC: Nuclear pore complex
ORF: *open reading frame*
PAR-CLIP: photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation
PAS: polyadenylation sequence
PBS: Phosphate buffer saline
PCR: Polymerase chain reaction
PEG: Polyethylene glycol
PIF: Progressive Idealization and Filtering
pmol: picomol
Pol I: RNA polymerase I
poly(A): polyadenosine
PP7CP: PP7 cap protein
PrA: ProteinA
PSACH: PP7-SNAP-PrA-CLIP-His₁₀
PVDF: polyvinylidene difluoride
RBD: RNA binding domain
RBP: RNA binding protein
RNA: ribonucleic acid
RNAPII: RNA polymerase II
RNP: ribonucleoprotein
rpm: rotation per minute
rPSACH: recombinant PSACH
rRNA: ribosomal ribonucleic acid
SDS: sodium dodecyl sulphate

SDS-PAGE: sodium dodecyl sulphate polyacrylamide gel electrophoresis

SEC: Size exclusion chromatography

SGD: *Saccharomyces genome database*

SLBP: stem-loop binding protein

SOC: super optimal broth with catabolic repression

ssAP: single step affinity purification

TIC: Total ion chromatogram

TIRF: total internal reflection fluorescence

TOR: Target of rapamycin

TORC1: target of rapamycin complex 1

TREX: transcription elongation complex

UTR: untranslated region

YPD: yeast peptone dextrose

α : anti

μ : micro

μ L: microliter

Acknowledgments

I would like to thank Dr. Oeffinger without whom this project would not be possible. I thank her for the opportunity to pursue scientific endeavours. I thank all members of the Oeffinger lab, past and present, for their help over the years, their contributions to my training and their entertaining and insightful discussions.

Thank you to Dr. Zenklusen for his help and thoughtful discussions over the years and the members of his lab for sharing their expertise and reagents. Thank you to Dr. Blunck for his expertise on single-molecule microscopy experiments and allowing the use of his TIRF microscope. A very special thanks to Lena Moeller for her great assistance in operating the TIRF microscope and making the coverslips necessary for this project.

Thank you to Vicky Kottis for allowing me to contribute to undergraduate training for several years.

Thank you to my family for their support and patience with my studies.

Merci à Geneviève pour ton soutien, ta compréhension et ta patience.

Author Contributions

Dr. Oeffinger and Dr. Zenklusen devised the basis and concept for this project. Lena Moeller generated Ni-NTA coverslips, operated the TIRF microscope and analyzed images with the PIF program created by her supervisor Dr. Blunck. Dr. Christian Trahan operated the FPLC for size exclusion chromatography. The IRCM proteomic platform was responsible for running tryptic digested peptides on the mass spectrometry instruments and performing preliminary processing of the raw data. Aside from that which is stated above, I performed all experiments, generated all strains and generated all plasmids used for this work. I conceived of generating IgG coverslips and was assisted by Lena Moeller in preparing them.

Introduction

The basics of mRNA, RBPs and mRNPs:

Eukaryotic cells have evolved to compartmentalize the sites of RNA transcriptions to the nucleus and translation away to the cytoplasm. In order for mRNA to transition out of the nucleus, a ribonucleoprotein complex must form about the mRNA to indicate its export competence[1]. Naked cellular RNA is a rapid target for degradation and is inherently unstable; therefore RNA binding proteins (RBPs) must quickly associate with nascent mRNA to not only protect it but also to guide it along the maturation path for exportation out of the nucleus. RBPs act as major effectors of RNA fate, regulating maturation, localization, translational status and degradation. Quality control mechanisms ensure that aberrant mRNA transcripts are degraded prior to their nuclear export, but also that only properly processed and matured mRNPs are exported. The complement of RBPs, and the structure adopted by the mRNP, influences the cellular function of mRNPs and highlights the relation between structure and function [2, 3].

mRNP structures are dynamic, they undergo frequent rearrangements as they grow, mature and develop through progression along the export pathway and within the cytoplasm of the cell. During transcription of mRNA by RNA polymerase II (RNAPII), the increase in the length of nascent mRNA offers the opportunity for the loading of more protein factors onto the mRNA through protein-nucleic acid interactions, as well as through protein-protein interactions between bound proteins[2, 3]. Some of the factors recruited are believed to be only transiently associated, while others remain bound for longer periods and possibly for most of the mRNAs lifetime within the cell. During export, a number of nuclear RBPs and mRNA associated proteins are released for exchange with cytoplasmic RBPs, while others remain associated until the initial round of translation or mRNA localization[2].

Here I will review and highlight key mRNP components, the regulation of mRNPs through post-translational modifications and the mechanisms of mRNP maturation which allow for structured mRNP formation allowing for nuclear export. As my work focused on the use of the model organism *S. cerevisiae*, much of the literature review will focus on research and examples taken from works pertaining to this species, with examples in other organisms mentioned where relevant and appropriate. Prior to delving into the details of this system, I will begin with quick review and summary of the basics of transcription in *S. cerevisiae*.

Table 1. mRNP Components and their mammalian homologs

	<u>Yeast</u>	<u>Human</u>
CBC	Cbp80/Cbc1/Sto1 Cbp20/Cbc2	Cbp80/NCBP1 Cbp20/NCBP2
THO	Hpr1 Tho2 Mtf1 Thp2 Tex1	Thoc1 Thoc2 Thoc7 Thoc3 Thoc 5, Thoc6
TREX	Yra1 Sub2	Aly/Ref/Thoc4 UAP56
TREX-2/AMEX	Sus1 Sac3 Sem1 Thp1 Cdc31	Eny2 Ganp Dss1 Q5JVF3 Cen2
Export Receptor	Mex67 Mtr2	Nxf1/Tap Nxt1/Ntf2/p15
N-Methyltransferase	Hmt1	Prmt1

The basics of transcription:

The eukaryotic fungal species *S. cerevisiae* is regarded as a simple organism, with a simple genomic organization in comparison to higher organisms such as mammalian cells. Of *S. cerevisiae*'s 6000 ORFs, only 5% (281) contain introns, while this situation is reversed in mammalian cells where only 4.5% (687) of genes lack introns[4]. The prevalence of introns in higher organisms adds complexity to the system of mRNA maturation and also causes some differences in mRNP structure and composition, most notably in the regulated recruitment of the spliceosome to intron-containing genes

In either case, mRNP biogenesis begins co-transcriptionally through the interaction of mRNP components with the RNA polymerase II transcription machinery[1]. The large subunit of RNA polymerase II in yeast, Rpb1, contains a heptapeptide repeat (YSPTSPS) which is repeated 52 times in humans and 26 times in yeast and is referred to as the C-terminal domain (CTD)[5, 6]. This repeat sequence undergoes differential phosphorylation as the RNAPII progresses through transcription of the gene, most notably is the decrease of Ser5 phosphorylation and the accumulation of Ser2 phosphorylation towards the 3' end of the gene. The CTD acts as a scaffold for the recruitment of RNP components by facilitating the transfer of factors from the transcription machinery to the nascent mRNA. Some proteins, such as Yra1, have been suggested to bind directly to the CTD and are then transferred to the mRNA, while other mRNP component proteins are transferred through protein-protein interactions with CTD bound factors, such as the THO complex [7]. The state of CTD phosphorylation acts as an important driving signal to begin the assembly of mRNP complexes [8].

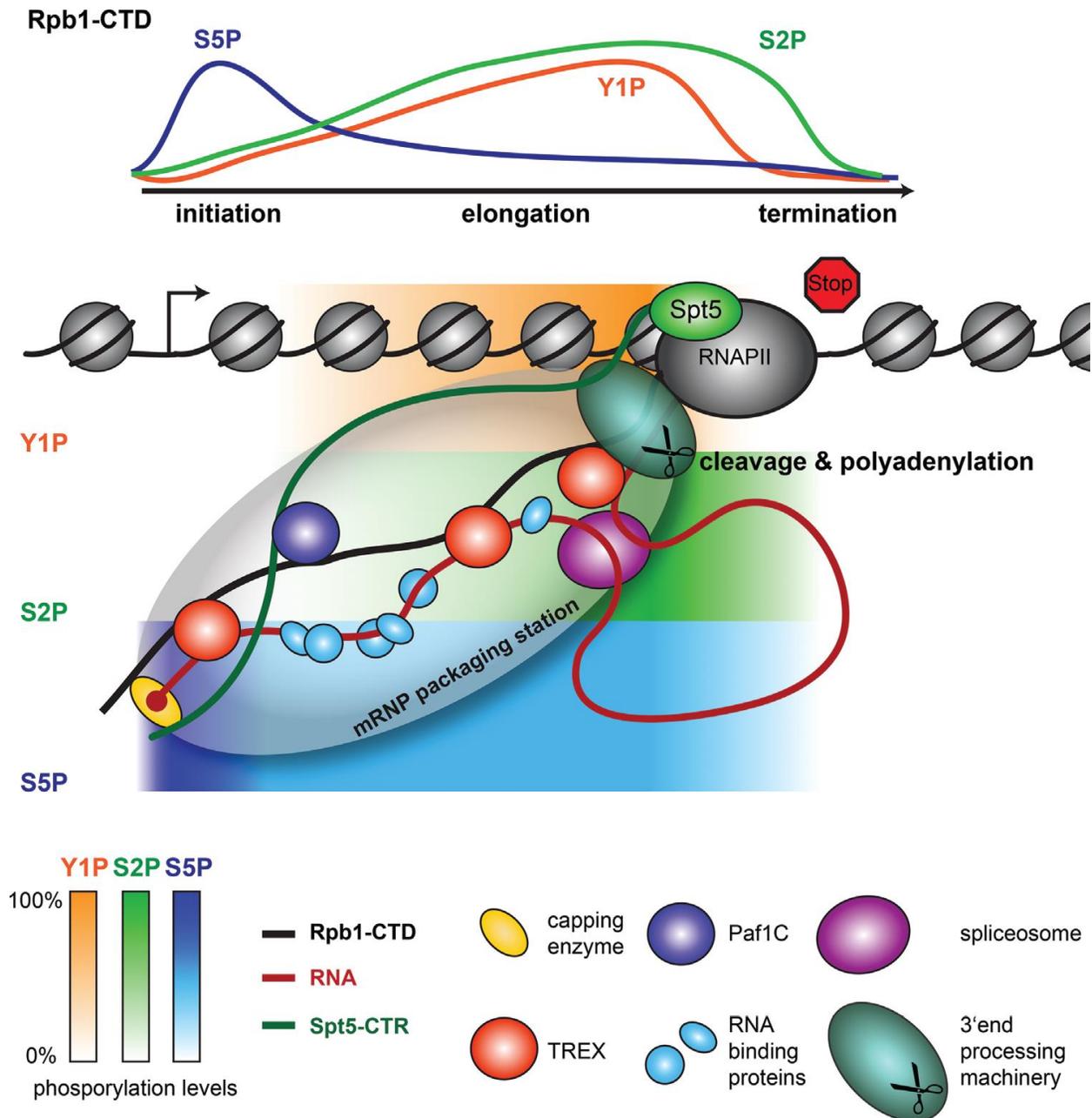


Figure 1. Co-transcriptional formation of mRNPs. Illustration demonstrating the co-transcriptional formation of a mRNP with factors interacting with RNAPII according to the phosphorylation pattern of the CTD of Rbp1. The figure is taken from [9]

5' Cap Binding Complex

The first processing event to occur in the life of mRNA occurs early with the addition of a specialized 7-methylguanosine cap structure to the 5' end of the mRNA (5'm7G). So-called capping occurs co-transcriptionally, following the synthesis of the first 25-30nt and the emergence of the 5' end of nascent mRNA from RNAPII, three enzymes, a triphosphatase, a guanyltransferase, and a methyltransferase, act to form the structure unique to RNAPII transcripts[10-14]. Cet1 acts to cleave phosphate from the 5' end of the first nucleotide which harbours a triphosphate, followed by the addition of a guanosine by Ceg1 which is then methylated by Abd1, resulting in the final formation of the cap linked to the first nucleotide by a 5'-5' phosphate bridge[15]. Cet1 and Ceg1 exist as a heterotetrameric complex, with two subunits of each protein, which is directly recruited to RNAPII through Ceg1 interaction with phosphorylated Ser5 of the CTD prevalent at the 5' end of genes[16-18].

The 5'm7G cap structure serves as a marker for RNAPII transcripts and promotes mRNA stability through protection from 5' exonucleases, such as nuclease Rat1 found in complex with pyrophosphohydrolase Rai1 which functions in 5' quality control of mRNA and RNA metabolism [19, 20]. Cap synthesis is essential, as null mutations for the synthesis enzyme Cet1, Ceg1 or Abd1 are non-viable in yeast [21, 22]. The cap also serves as the binding site for the nuclear Cap Binding Complex (CBC), composed of Cbp20 (Cbc2) and Cbp80 (Sto1/Cbc1), which are recruited shortly after 5'm7G synthesis[23]. CBC is only capable of cap binding as a dimeric complex, as both Cbp20 and Cbp80 individually exhibit low affinity for 5'm7G despite the localization of the cap-binding domain within Cbp20 [15, 23-26]. Cbp20 requires Cbp80 as the monomeric protein is unstable, yet Cbp80 does not require Cbp20, as it is stable alone [27, 28]. Surprising, these two proteins are expressed at different levels in yeast cells, with Cbp80 in

twofold excess over Cbp20 (11,000 vs 6,200 molecules/cell)[29]. While this proposed disparity in abundance would suggest that all Cbp20 is saturated and in complex with Cbp80 leaving a large free pool of Cbp80, no cellular role or function has been described for either protein independent of the other. [27].

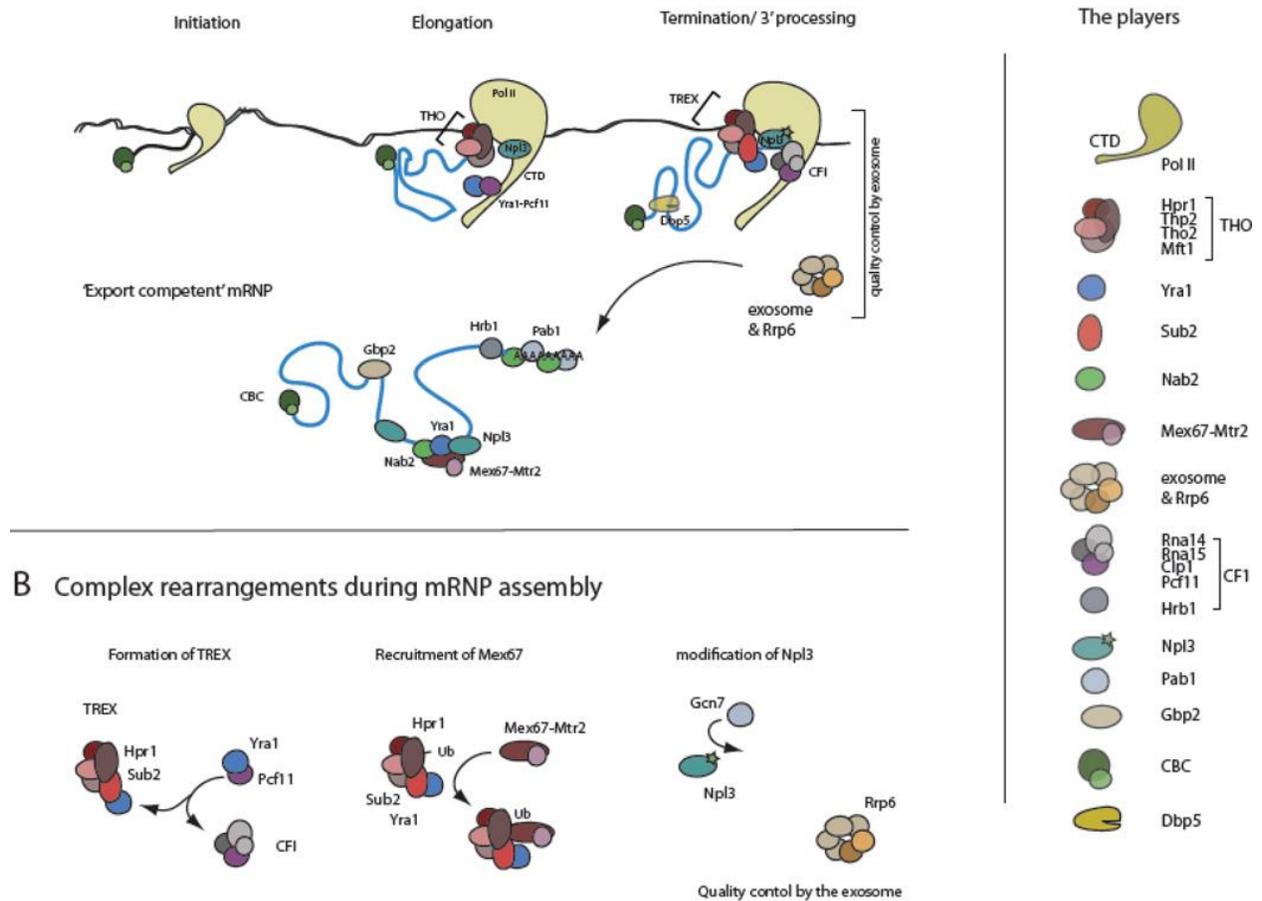
Aside from promoting mRNA stability, the CBC plays several roles in mRNP maturation. While CBC may only be recruited to mRNA once transcription is underway, it has also been shown to have functions upstream of active transcription. CBC interacts with transcription factor Mot1 to recruit the protein to the pre-initiation complex (PIC) at yeast promoters to enhance the formation of the PIC [30]. Furthermore, CBC promotes the recruitment of CTK and BUR kinase complexes to phosphorylate RNAPII CTD Ser2, as well as establish histone H3K36me3, a marker of active transcription [31]. CBC also serves to couple transcription with splicing, being required for efficient pre-mRNA splicing. CBC both physically and genetically interacts with spliceosome components, with both CBC components being synthetically lethal with spliceosome factors and depletion of CBC causing a reduction in spliceosome assembly. [23, 24, 27, 32, 33]. Likewise, for splicing, CBC is also involved in promoting efficient 3' processing and polyadenylation, however, these influences are more pronounced in mammalian cells than yeast [34, 35]. In *S. cerevisiae* the 5' m7G has little effect on 3' processing; however, CBC has been shown to interact with Npl3 to repress the recruitment of the termination complex CFIA to weak polyadenylation sites [28, 36]. Despite the involvement of the CBC in the many maturation events of mRNA, the complex is non-essential in yeast, where deletion strains suffer growth defects but remain viable.

CBC is only known to function in relation to the binding of the 5' m7G cap, yet both Cbp20 and Cbp80 contain RNA recognition motifs (RRM) and are capable of binding RNA[37-

39]. Chromatin immunoprecipitation experiments have not only detected enrichment of the CBC at the 5' end of genes but also well within the body of genes, suggesting that CBC may associate with elongating RNAPII like many co-transcriptionally recruited mRNP proteins[10, 30, 40, 41]. Consistent with these observations, CBC may transfer from RNAPII to the nascent mRNA before finally binding 5' m7G or multiple CBC proteins may bind to a single mRNA. The direct RNA binding sites of Cbp20 and Cbp80 were investigated with two different techniques, PAR-CLIP and CRAC respectively, which were able to detect enriched occupancy of both proteins within the 5' region of mRNAs and not only at the 5' cap, with Cbp20 specifically detected enriched up to 90nt from the transcription start site[42, 43]. Regardless, the primary site for CBC occupancy must be the 5' m7G cap, not only for reasons of mRNA stability but due to the complex's high affinity for the structure. Using a synthetic cap analog, the association constant for CBC was determined to be approximately $2 \times 10^8 \text{ M}^{-1}$, which is several folds greater than the association constant of the cytoplasmic cap-binding complex eIF4E[25].

Balabiani ring particles produced in *Drosophila* salivatory glands have historically served as a model for the study of mRNPs due to their large size[44]. These mRNPs have consistently been observed to begin translation, even as the rest of the mRNA is being translocated through the NPC. In addition, associated mRNAs have been observed to be exported with their 5' ends exiting towards the cytoplasm first, suggesting that the CBC may serve as the primary marker for orientation of these mRNPs at the NPC for export. [45].

A Cotranscriptional recruitment of mRNA export factors



B Complex rearrangements during mRNP assembly

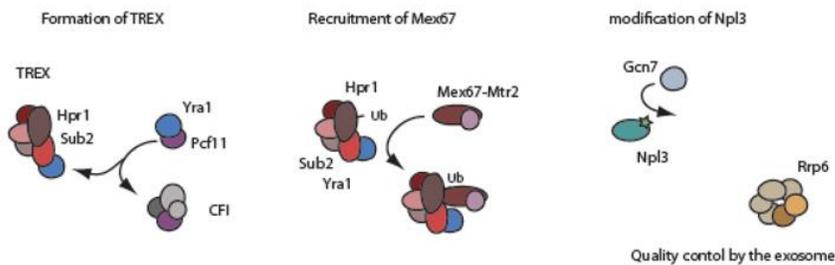


Figure 2. Assembly and rearrangement of mRNP export factors. Hypothesized order of assembly of mRNP associated proteins with their approximate localization on the mRNA shown. Associated protein undergo modification which may trigger rearrangement or disassociation of factors from the complex throughout the maturation of the mRNP. The figure is taken from [46]

THO complex

As RNAPII moves to the body of a gene, efficient transcription is dependent on the recruitment of the pentameric THO complex. Originally only described as the heterotetrameric complex composed of Tho2, Hpr1, Mft1 and Thp2, the complex was also later found to include the factor Tex1[47, 48]. The THO complex is evolutionary conserved and necessary to facilitate mRNP packaging, protein recruitment and nuclear export through interactions with splicing components, poly(A) binding proteins and export adaptors. The THO complex recruits Yra1 (Aly/Ref/Thoc4) and Sub2 (UAP56) which then forms the Transcription Elongation (TREX) complex coupling transcription with export. Sub2 is a DEAD box helicase first identified with a role in splicing but then discovered to have essential roles in intron lacking mRNA export.

The THO complex is a relatively stable complex which can be purified under high salt conditions using any component as bait [49]. Recently, a structure of the THO complex has been described using SAXS and docking to EM structures to give candidate positions of each protein within the complex which is informative in the absence of a high-resolution protein structure [50, 51]. However, it is unknown if this complex is constitutively formed or undergoes assembly during transcription. Interestingly, each complex component is present at varying levels, with Tho2 at the lower end estimated at only 521 molecules/cell and Mft1 at the upper bound with an estimated ten-fold greater prevalence at 5910 molecules/cell [29].

THO/TREX components are co-transcriptionally recruited through interaction with RNAPII transcription machinery. THO/TREX shows a characteristic ChIP profile of accumulation towards the 3' end of genes. For THO, optimal interaction with RNAPII CTD is achieved with Ser2/Ser5 phosphorylation which allows for direct binding, but also with Ser2 and/or Tyr1 phosphorylation which accumulates towards the 3' end of genes[7, 18]. While the

mechanism for Sub2 recruitment remains unknown, Yra1 has been shown to interact directly with CTD Ser2 phosphorylation and indirectly with CTD associating protein Pcf11, which is a component of the CF1A complex involved in 3' processing [52, 53].

In *S. cerevisiae*, the THO complex, specifically Hpr1, can be found genome-wide recruited to actively transcribed RNAPII genes [54]. THO null mutants are viable and not essential for mRNA maturation and nuclear export per se. Using Poly(A) FISH, deletion of any THO component has been observed to result in nuclear accumulation of transcripts; a phenotype enhanced when cells are subjected to temperatures of 37°C due to the induction of the heat shock response [47]. Similarly, Hpr1 and Tho2 deletions each affect global gene expression; however, they tend to affect genes which are longer, most expressed and with higher GC content to a greater extent [54]. While Tho2 and Hpr1 deletions tend to produce the same transcriptional phenotypes, with no synergistic effects from the double mutant, individual gene deletions seem to result in the differential expression of certain genes different genes [54, 55]. Increasing evidence shows that deletion of different THO subunits affects the expression of different genes producing different phenotypes. For example, deletion of Tho2 or Hpr1 was shown to decrease expression telomere interacting protein Rif1 resulting in telomere lengthening, yet no such effect was observed with the deletion of Mft1 or Thp2 [56, 57]

The major effects seen in THO null mutants is the stalling of transcription elongation, the accumulation of R-loops and DNA hyper-recombination resulting in genomic instability [55, 58-61]. THO mutants are also reported to undergo defects in 3' processing resulting in retention of mRNPs at the site of transcription and genomic locus [62]. 3' processing defects for genes which contact the NPC results in the accumulation of transient intermediates in the transcription and export pathways resulting in the formation of so-called "heavy chromatin", with the formation of

these dense structures limited to the gene boundaries of THO target genes [62, 63]. With the propensity for so many genome damaging events and alterations to gene expression, it remains surprising that THO mutant only suffers from mild growth defects and remain viable.

THO function in Metazoa

The THO complex is evolutionary conserved, however, the mammalian THO complex is composed of 6 protein factors, Thoc1 (Hpr1), Thoc2 (Tho2), Thoc5, Thoc6, Thoc7 (Mft1) and Tex1 [64]. Human THO does not associate with pre-spliced mRNA, but rather only spliced mRNA in a splicing coupled manner [49]. Despite these early observations, Thoc1/hHpr1 knockdown was recently shown to indeed influence RNAPII elongation and exhibit similar genome instabilities as seen in yeast [65].

The biological importance of Thoc1 has begun to be investigated in mice. Using mouse models, it was discovered that Thoc1 was essential for early development due to a strong dependence of inner cell mass cells for the protein resulting in embryonic lethality [66, 67]. Conditional knockout mutants later revealed the importance of Thoc1 in granulocyte progenitor cell viability, gametogenesis affecting spermatocyte viability and leading to male infertility and small intestinal stem cells [68-70]. More interesting still is the emerging role for Thoc1 in cancer, where Thoc1 overexpression has been found to increase with breast cancer tumour size and metastasis, prostate cancer cells and small cell lung cancer [71-73]. Recently, it was shown that loss of Thoc1 in transformed fibroblasts results in increased cell death by apoptosis not observed in untransformed cells, suggesting Thoc1 as a possible future cancer therapeutic target [74]. Of additional note for THO complex components in metazoan function is THOC5 and THOC2 which are required for mouse embryonic stem cell self-renewal. With Thoc5 mediating the interaction of Thoc2, these proteins were shown to preferentially associate with pluripotent

gene transcripts to mediate their export and expression, with reduced expression of Thoc5 leading to cell differentiation [75]. Additionally, Thoc5 is also important for hematopoiesis [76]. The THO complex is important not only at the cellular level but also for whole organisms to function properly.

TREX

Yra1 was first identified in yeast as an essential protein with potent RNA annealing activity, leading to the hypothesis for a potential role as a mRNA chaperone [77]. Yra1's involvement in RNA export was solidified with the discovery of its interactions with the heterodimeric nuclear export receptor Mex67-mtr2 and its necessary role in mRNA export [78].

Yra1 is one of the few intron-containing genes in *S. cerevisiae*. In the case of Yra1, its intron was discovered to aid in the regulation of Yra1 protein levels through an autoregulatory loop. Overexpression of Yra1 cDNA was found to be toxic and lead to the nuclear accumulation of poly(A) transcripts, yet overexpression of the intron-containing transcript did not affect Yra1 protein levels [79, 80]. It was determined that Yra1 is capable of inhibiting the splicing of its own mRNA, thus targeting it for degradation once exported to the nucleus [81]. Such regulatory loops have not been described for any other mRNP gene, however, CBP80 does also contain a similar intron with its gene.

Yeast also express a homolog of Yra1, Yra2, which shares high sequence and structural similarities with Yra1. Yra2 overexpression can compensate for the loss of Yra1, suggesting functional redundancy to some extent [82]. Yra1 and Yra2 have also found to interact, but this was only shown from purified Yra1 in an RNA dependent manner [83]. Additional high copy suppressor to compensate for the loss of Yra1 were identified, and these include SUB2 and

ULP1, an NPC associated SUMO protease, with Sub2 confirmed to also be a direct interactor with Yra1 [41, 83, 84].

Sub2 was identified as the yeast homolog of human UAP56 already known to be involved in splicing [85, 86]. While first only hypothesized to interact with mRNA in a splicing-dependent manner, Sub2 was quickly shown to be required for export of intronless mRNAs [87]. Both Sub2 and Yra1 associate with elongating RNAPII and show similar ChIP profiles as those for THO mutants, however Yra1 association is independent of Sub2 [52]. Yra1 and Sub2 are recruited to the THO complex through interaction with Hpr1 [41]. Yra1 and Sub2 interact genetically with the tetrameric THO complex (Hpr1, Mft1, Tho2, Thp2), with Sub2 overexpression also able to resolve defects arising from HPR1 depletion [47, 88]. However, deletion of any subunit of THO abolishes Sub2 association, which then raises questions on how Sub2 is then able to associate with mRNPs and how these cells can still be viable since Sub2 is an essential protein [47].

While TREX components show enrichment towards the 3' ends of genes, they are surprisingly found to bind preferentially towards the 5' end of mRNAs with little observed at the 3' end [42, 43]. In mammalian cells, TREX and Cbp80 have been shown to physically associate. Specifically, Aly/Ref and Thoc2 are capable of both mutual and independent association with Cbp80, however, both are required to mediate the association of UAP56 which is unable to interact with Cbp80 alone [64]. Given the preferential enrichment 5' mRNA for THO/TREX observed in yeast, these associations are likely to be conserved.

Mex67-Mtr2/TAP-p15/NXF1-NXT1 recruitment

Yra1 and Aly/REF

One of the principal reasons for the loading of Yra1 onto mRNA is to serve as an adaptor for the recruitment of the nuclear heterodimeric export receptor Mex67-Mtr2 (TAP-p15). Yra1 and Mex67-Mtr2 interact through the same protein surface interface as Sub2/Yra1, therefore Sub2 must first be released to allow Mex67-Mtr2 to subsequently interact with Yra1 [87, 89]. Sub2 has recently been shown to be ubiquitinated by Mdm30, resulting in the targeting of Sub2 for proteasomal degradation. Degradation of Sub2 was shown to enhance Yra1 recruitment and promote mRNA export, thus providing a possible mechanism for the exchange of Sub2 for Yra1 with Mex67 [90].

Mex67-Mtr2 and TAP-p15 are both poor RBPs with a weak innate affinity for RNA due to an intermolecular interaction which sequesters its RBD and limits the complexes' ability to associate with mRNA [91, 92]. The interaction between Yra1 and Mex67-Mtr2 triggers a conformational change in Mex67 which allows the RBD to become functional, with severalfold greater RNA binding affinity [89, 91]. In human cells, the handover of RNA from Aly/Ref to TAP/NXF1 is promoted through arginine methylation of Aly/REF by PMRT1/Hmt1 which reduces Aly/Ref RNA binding affinity and leads to Aly/Ref dissociation from mRNA [93]. In yeast Yra1 has been shown to be a substrate for Hmt1 in a yeast two-hybrid screen, in addition to other mRNP component proteins such as Npl3 and Gbp2, suggesting the yeast homolog of PMRT1 may encourage the same mRNA handover reaction to Mex67-Mtr2 [94]. In addition, Yra1 is ubiquitinated by Tom1 at its C-terminal region to promote its dissociation from mRNPs, as cells lacking Tom1 show increased Yra1 association with mRNPs and mRNA export defect [95].

Hpr1

Yra1 is not the only means for Mex67-Mtr2 recruitment, as several other mechanisms and contributing interactions have been proposed. During RNAPII elongation, Hpr1 undergoes polyubiquitination by Rsp5, resulting in targeting to the proteasome for degradation [96]. However, Mex67 is capable of interacting with Hpr1, with this interaction promoting a structural rearrangement which exposes the ubiquitin-associated (UBA) domain of Mex67 which then binds polyubiquitinated Hpr1, protecting Hpr1 from degradation and promoting Mex67 recruitment to the mRNP [96, 97]. Dissociation of Mex67 from the mRNP at the nuclear pore may cause Hpr1 degradation and THO complex separation from the mRNP.

Nab2

Nab2 (ZC3H14) is an essential RBP and poly(A) binding protein with a role in controlling mRNA poly(A) tail length[98, 99]. Nab2 associates with mRNA during 3' processing and polyadenylation through the use of seven CCCH zinc finger domains at its C-terminus, with the last three domains sufficient for potent RNA binding and polyadenylation control [99-102]. Through ChIP experiments, Nab2 has also been detected at RNAPII actively transcribed genes with enrichment at the 5' end of genes and with a decreasing signal progressing within the body of most genes [103]. Nab2 only allows for nuclear export of associated mRNAs if it has undergone methylation in its RGG domain by the action of Hmt1. Nab2 remains associated with mRNA following nuclear export through the NPC and is only removed on the cytoplasmic face of the NPC by the remodelling action of the DEAD-box helicase Dbp5 [98, 104]. Nab2 is also capable of acting as an adaptor for Mex67 recruitment in conjunction with Yra1, which together are capable of forming a trimeric complex. While not essential, Yra1 interaction with Nab2 is able to enhance the Nab2-Mex67 interaction, which is nonetheless possible in the absence of

Yra1. The dissociation of Yra1 from Nab2 and Mex67 is prompted by Yra1 ubiquitination by Tom1, thus leaving Nab2 associated only with Mex67 within the mRNP structure.

Furthermore, while Yra1 is normally an essential protein, it is completely dispensable in cells which overexpress Nab2, showing effective Mex67 recruitment and no observed mRNA export defects even with Yra1 deletion [95]. However, the means of Nab2, Yra1 and Mex67 interaction remains intriguing in light of recent conflicting RNA binding profiles. While Mex67 occupancy is approximately evenly across the body of a transcript by PAR-CLIP experiments, Nab2 is enriched at the 3' end at the Poly(A) site and Yra1 enriched at the 5' end with low occupancy at the poly(A) site [43]. Such occupancy may be possible if the mRNP adopts a compact structure which allows for the overlap and close association of the 5' and 3' ends, reminiscent of the circularization adopted by actively translating polysomes in the cytoplasm. However, analysis of Nab2 purified mRNPs by cryo-electron microscopy have only been able to reveal elongated compact structures, offering limited evidence of such a hypothesis [102].

Npl3

While Npl3 has also been shown to serve as a potential Mex67 adaptor, it also serves many essential roles in mRNP maturation. Npl3 is an SR protein, similar to Gbp2 and Hrb1, which contains many serine and arginine dipeptide residues and RGG motifs which serve as targets for post-translational modifications to vary the protein's function [105, 106]. Npl3 associates with mRNPs in the nucleus and remains complexed within mRNPs in the cytoplasm, serving to influence mRNA translation rates until phosphorylated by kinase Sky1 to trigger dissociation and eventual nuclear import via Mtr10 [106-109]. In the nucleus, Glc7 dephosphorylation of Npl3 is essential for its mRNA association and essential for Mex67 recruitment to mRNPs, as neither occur with Npl3 in a phosphorylated state. While

dephosphorylation of Npl3 is required for mRNA export, the significance of Mex67 interaction with Npl3 in the context of mRNA has not been well described [107]. Aside from mRNA, Npl3 and Mex67 have also been shown to be involved in the export of pre-ribosomes possibly confounding interactions pertaining to mRNPs [110-112].

Unphosphorylated Npl3 is co-transcriptionally recruited to RNAPII through interaction with phosphorylated Ser2 residues of the CTD and stimulates transcriptional elongation [113, 114]. Along with CBC, with which it directly interacts, Npl3 also functions to impede the recruitment of cleavage factor IA (CFIA) to weak termination sites [28, 113, 115]. Along with CTD binding, Npl3 binds RNA to antagonize the binding of Rna15, a component of CFIA, and negatively regulate transcription termination. Alleviation from Npl3 stimulation is achieved through phosphorylation by Casein Kinase 2 (CK2) which prompts Npl3 dissociation from both RNAPII and mRNA allowing for CFIA recruitment and 3' processing [113, 116]. Glc7 is found to associate with 3' processing proteins where it may then be in a position to dephosphorylate Npl3, allowing for its association with processed mRNA to serve as a Mex67 adaptor [107, 117, 118]. As Mex67 shows preferential association with transcripts having already undergone 3' processing, it is less likely that Mex67 and Npl3 would interact during RNAPII stimulated elongation [43]. Npl3 shows enrichment exclusively with the 5' end of mRNA and at the poly(A) site, consistent with an interaction with Cbp80 and 3' processing antagonization. This interaction has been hypothesized to be possible as a result of folding of the mRNP to bring Cbp80 in contact with Npl3 at the 3' end, or may be the result of different Npl3 molecules bound to the same mRNP, or even still different Npl3 interactions occurring on different mRNAs [31, 43]. However, there is a lack of structural and stoichiometric data of protein factors available which would be able to shed light on these possibilities.

In addition to phosphorylation, Npl3 is regulated through the differential methylation of its 17 RGG motifs by Hmt1[119]. The RGG motifs of Npl3 serve to inhibit several functions of the protein, but this inhibition is overcome through methylation or observed through R→K mutation, and allows for modulation of Npl3 activity through the action of Hmt1 [120, 121]. A defect in Npl3 methylation results in decreased occupancy at genes, decreased RNAPII elongation, increased termination and recruitment of CFIA to cryptic polyadenylation sites, consistent with the decreased anti-termination activity of Npl3 [113, 122]. Notably, RGG methylation does not affect Npl3 RNA binding activity [120]. Non-methylated Npl3 also interacts with Tho2 in an RNA dependent manner, reducing Tho2 occupancy at genes and contributing to decreased RNAPII elongation, with Hmt1 methylation of Npl3 disrupts the interaction and possibly serving as a means for initial recruitment of Tho2[119, 121, 122]. Methylation is also essential for nuclear export of Npl3 and presumably mRNP associated complexes, with the mutation of all methylation sites resulting in Hmt1 independent nuclear export. Regulation of Npl3 through methylation is essential, the progressive mutation of targeted RGG motifs associated with increased growth defects and would otherwise prevent the modulation of Npl3 function or signal for export-competent mRNPs[119-122].

Npl3 is one of the most prevalent mRNP proteins within yeast cells, likely due to the many cellular processes with which it interacts and modulates [29]. While Npl3 is recruited to the transcripts of most genes in yeast, it was discovered to be required for efficient splicing of a large set of intron-containing pre-mRNA, most notably the highly expressed ribosomal protein genes[43, 123]. Npl3 promotes the co-transcriptional binding of early splicing factor to chromatin and associates with U1 snRNP and the branch point binding protein to stimulate splicing [123]. Npl3 was also found to physically and genetically associate with the Rad6

complex for H2B mono-ubiquitination through interaction with Bre1 and Ubp8 of the SAGA chromatin remodelling complex. Deletion of BRE1 or UBP8 genes were found to exacerbate splicing defects of Npl3 deletion strains, revealing connections between Npl3, chromatin factors H2B ubiquitination and pre-mRNA splicing [124]. As well, due to a close relationship with elongating RNAPII, Npl3, similar to THO/TREX, has been described to have a role in genome integrity and the prevention of R-loop formation [125]. With the many interaction modes of Npl3, it is not surprising that its levels of expression must be well regulated within the cell with either overexpression and underexpression leading to cell toxicity [126, 127]. Taking advantage of its influence on RNAPII elongation and anti-termination effects at weak poly(A) sites, Npl3 has been described to autoregulate expression of its own mRNA through a negative feedback loop. The 3' end of the Npl3 gene contains alternative polyadenylation sites, with the use of a later site resulting in the production of an unstable transcript that is quickly degraded. In the presence of excess Npl3, read-through and production of unstable NPL3 transcripts dominate, while the use of an early polyadenylation site and production of a stable transcript dominates with less Npl3, thus achieving a means of regulating cellular protein levels [115].

TREX-2

Alongside the THO/TREX complex, *S. cerevisiae* also expressed a TREX-2/AMEX complex required for mRNA export. TREX-2 is a conserved protein complex that serves to mediate gene gating, the association of a number of actively transcribing genes, such as the GAL and ARG1, with the nuclear pore. Sac3 serves as the core scaffold of yeast TREX-2 for the recruitment of Thp1, Cdc31, Sus1, and more recently Sem1, a subunit of the 19S proteasome [128-131]. Sus1 is also a component of the transcriptional co-activator complex SAGA, a complex executing histone tail modifications to clear nucleosomes from transcription initiation sites [132-136]. The SAGA complex represents the minor pathway for transcriptional activation, acting on 10% of *S. cerevisiae* genes which are typically upregulated in response to environmental stress [137, 138]. While acting primarily at gene promoters, SAGA has also been localized in the coding region of genes influencing transcription elongation [139]. Similar to THO/TREX components, Sus1 associates with elongating RNAPII, but through Ser5 CTD phosphorylation rather than Ser2 [139]. As the common component between SAGA, elongating RNAPII and TREX-2, Sus1 is believed to mediate the recruitment of TREX-2 to nascent transcripts [140].

TREX-1 and TREX-2 do not operate independently, as Sus1 co-purifies with Yra1 and Sac3 with THO/TREX components, indicating the presence of both complexes on certain mRNPs [130, 141]. Sac3 was found to interact with Mex67 directly, suggesting that it too may act as a Mex67 adaptor for mRNP nuclear export [130]. Furthermore, while *thp1* deletion strains show transcriptional and mRNA export defects, these can be overcome through overexpression of Nab2 [130].

The docking of inducible genes at the NPC through TREX-2 is thought to allow for tight regulation of expression, allowing for rapid repression following gene inactivation but not necessarily allowing for rapid expression [142]. Under steady state conditions, TREX-2 is found associated with the NPC, an interaction mediated through Sac3 and Sus1 contact with two critical residues of Nup1 [143]. Localization of TREX2 appears to be critical for efficient gene expression, as mutations solely disrupting Nup1:TREX2 interaction and not complex assembly results in growth and mRNA export defects. Disruption of the TREX2 complex in yeast results in decreased RNAPII elongation, nuclear poly(A) accumulation and growth defects. These effects are not limited to NPC localized or inducible genes. Recent ChIP analysis showed Thp1 and Sac3 recruitment to approximately 2000 actively transcribed genes, showing a graded accumulation towards the 3' end of genes similar to that of the THO complex [144]. Given that these were not all SAGA-dependent genes, the mechanism of TREX-2 recruitment must be slightly different than the transfer of Sus1 from SAGA to RNAPII to recruit the complex. Consistent with the genes to which it was recruited, loss of Thp1 and Sac3 was found to preferentially affect the expression of long, GC rich and highly expressed genes [144].

Similar to THO/TREX, loss of TREX-2 leads to genome instability in both yeast and mammalian cells, however, unlike THO/TREX, this instability does not arise through R-loop formation. Instead, TREX-2 functions to prevent transcription-replication conflicts, as TREX-2 mutants show increased accumulation of replicative Rrm3 helicase, required for progression of replication forks through obstacles in DNA. [144-146]. Likewise, siRNA knockdown of mammalian TREX-2 components leads to increased accumulation DNA damage markers Rad51, 53BP1 and γ -H2AX, as well as alterations in DNA replication profiles but not R-loops. Mammalian TREX-2 is believed to promote genome stability through its interaction and

recruitment of DNA repair factor BRCA2, as loss of BRCA2 was recently shown to lead to an accumulation of R-loops [147]. Yeast do not express clear homologs of either of the BRCA proteins; therefore conservation of TREX-2 mediated recruitment of DNA repair factors is expected to be limited [148, 149]. However, both THO/TREX and TREX-2 have been implicated in transcription-coupled repair, which may represent some level of conserved mRNP biogenesis associated DNA repair and genome stability [61, 150].

TOR regulation of mRNP Methylation

As outlined above, Hmt1 represents an important regulator of mRNP dynamics through the methylation of protein targets, which can modify the substrate protein's properties, such as RNA binding ability and nuclear export. Hmt1 requires phosphorylation at its N-terminus to allow for dimerization and activation of the protein, as the monomeric protein shows little to no methylation activity [151]. While the dimeric form shows enzymatic activity, Hmt1 further assembles into a hexameric complex, a trimer of dimers, to act upon its targets, with the human homolog of Hmt1, Prmt1, believed to do the same [151]. Regulation of Hmt1 multimerization and activity is thus capable of regulating mRNP components and maturation.

Such regulation was recently demonstrated in the context of the Clb2 mRNA encoding for cyclin B, required for M phase progression and cell division [152]. Active Hmt1 and methylation of its targets were shown to be required to generate stabilized CLB2 mRNA transcripts, which would otherwise undergo rapid degradation in the cytoplasm by the CCR4-NOT complex. The two principle regulators of Hmt1 affecting CLB2 mRNP stability were found to be activated through the kinase Dbf2, triggering multimerization, and inactivation through the phosphatase Pph22, triggering monomerization [152].

Pph22 activity is regulated as a downstream component of the TORC1 (Target Of Rapamycin Complex 1) signal transduction pathway. TORC1 is a kinase complex serving as a nutrient and cellular state sensor, promoting cell growth through phosphorylation of its substrates in its active state when cells are in favourable conditions. However, TORC1 signalling is reduced in cells exposed to poor growth conditions, leading to activation of catabolic and stress responses through different signalling effectors, including activated Pph22 [153, 154]. Thus a model for CLB2 mRNA decay was presented, whereby Pph22, activated under poor conditions, inactivates

Hmt1 to prevent methylation of mRNPs, promoting degradation of CBL2 mRNA in the cytoplasm and therefore preventing transcript translation and mitotic entry in an environment not favourable for growth or replication. Interestingly, a protein complementation assay in this study revealed a number of altered protein-protein interactions between Hmt1 and a number of mRNP associated factors affected by inhibition of TORC1 signalling. While only investigated specifically for the CBL2 mRNP, this outlines the possibility of TORC1 signalling in regulating mRNPs through Hmt1 [152].

A recent study in mammalian cells aimed not only to investigate the possible influence of mTOR on mRNA export but also the interconnected upstream signalling pathways PI3K and AKT [155]. Using selective chemical inhibitor of pathway components in combination with fluorescent microscopy and fluorescent recovery after photobleaching (FRAP) experiments, a significant reduction in the associations of Aly/REF, UAP56, NXF1 and EJC (Exon Junction Complex) components with nuclear complexes was observed following PI3 / AKT signalling inhibition. While inhibition of both Akt and PI3K promoted general poly(A) mRNA export, further investigation revealed differential regulation of the signalling pathways on different transcripts. Inhibition of Akt, but not PI3K promoted nuclear export of specifically investigated transcripts, while neither influenced the export of ER or mitochondrial targeted transcripts. This study showed a minimal influence of mTOR signalling on mRNA export complexes, with the only significantly reduced association of the EJC component MAGOH. However, this does illustrate the ability of upstream signalling components to differentially affect mRNP proteins and nuclear export of specific transcripts, similar to what was observed in the case of Hmt1 and CBL2 mRNP stability.

The PI3K/AKT/TOR pathway is not well conserved between yeast and mammalian cells, rendering it difficult to compare or seek out similar mRNP influences between the two organisms [156, 157]. In mammalian cells, PI3K/AKT/TOR plays important roles in apoptotic control and cell survival, events less important to unicellular yeast [156, 158, 159]. *S. cerevisiae* does not produce any detectable PIP₃ necessary for PI3K signalling, and only produce a weakly similar version of the upstream components and are in fact used for different biological processes, such as cell integrity and endocytosis [160-162]. These signalling pathways may be condensed in yeast funnelling more towards conserved TORC1 regulation. The possibility of PI3K/AKT/mTOR regulation of mRNPs remains interesting, as AKT was shown to phosphorylate Aly/REF and S6K, activated downstream of mTORC1, was shown to phosphorylate mammalian Cbp80 as one of its selective substrates [163, 164]. Identification of the mediators and mRNP effects will yield a better understanding of this potential regulatory mechanism, as well as how it may be conserved in yeast.

3' End Processing

The addition of a poly(A) tail to the 3' end of mRNA serves as a marker to define transcripts as export-competent, in addition to the association of export proteins. Like mRNP biogenesis, 3' processing of mRNA is a dynamic process requiring the action of nearly 30 proteins assembled into complexes, which can further be divided into sub-complexes, that act in a coordinated manner. 3' processing is complex and remains an area of study unto itself and will only be described briefly.

Elements encoded within the pre-mRNA act to localize and direct the recruitment of processing factors, with these cis-elements referred to as the polyadenylation signals (PAS). PAS consists of a conserved A(A/U)UAAA motif 10-30 nucleotides upstream of the cleavage/polyadenylation site, with additional variable GU-rich elements ahead of the A(A/U)UAAA motif [165-167]. Proteins recognizing these elements associate with transcribing RNAPII initially, but are then transferred to the pre-mRNA once the cis-regulatory elements are transcribed and exposed. Yeast cleavage factor I (CFI), composed of two sub-complexes CFIA and CFIB, binds upstream of cleavage site to the A(A/U)UAAA motif, while cleavage and polyadenylation factor (CPF) binds to the cleavage site. Coordinate interactions between CFI, CPF and associate factors induce pausing of RNAPII, allowing for endonucleolytic cleavage at the processing site [168-171]. The free 3' end which results from nucleolytic cleavage, is taken as a substrate by poly(A) polymerase (Pap1), adding the 3' poly(A) tail recognized by factors such as the poly(A) binding protein (Pab1) [172-174]. The exposed 5' end of cleaved pre-mRNA is recognized by exonuclease complex Rat1-Rai1, which degrade the remaining RNA still associated with RNAPII and template DNA. The collision of Rat1-Rai1 is believed to induce dissociation of RNAPII from DNA, thus completing the transcription termination [175, 176].

Of the factors involved in 3' processing, of particular interest is the CFI complex. CFIA is composed of Rna14, Rna15, Clp1, and Pcf11, while CFIB is solely composed of Hrp1. Pcf11 directly associates with phosphorylated Ser2 of RNAPII CTD to recruit the CFIA complex to elongating RNAPII and is poised to allow recruitment of CFIA once the PAS is transcribed [177-179]. Likewise, the observed association of Pcf11 with Yra1 is believed to allow for the coupling of 3' processing and mRNA export [52, 53]. Similar to the THO/TREX components with which they genetically interact, Rna14 and Rna15 are required for effective RNAPII transcription elongation, with mutants showing increased transcription-dependent hyper-recombination [179]. In addition, both Rna15 and Hrp1 interact genetically with Npl3, with their abilities to suppress Npl3 temperature-sensitive mutations [178, 180]. Like Npl3, Hrp1 is a shuttling protein which remains associated with mRNPs through the NPC and is involved in nonsense-mediated decay of transcripts containing a premature stop codon in the cytoplasm [181]. Similar to what has been reported for THO/TREX, CFI has also been linked to genome stability through the promotion of DNA damage repair mechanisms [182]. Of interest is the recent observation regarding Hrp1 localization on mRNA transcripts. While as expected, Hrp1 is found at the 3' end of mRNA near the poly(A) site, enrichment has also been observed at the 5' end, similar to what is seen for Npl3 [42]. Occupancy of Hrp1 near the 5' end may mediate interactions between CFI and THO/TREX, accounting for similar interactions observed between the two complexes.

The CFI and CPF complexes are chiefly responsible for the transcription termination and processing of pre-mRNAs, while the yeast Nrd1 complex functions in the termination and processing of non-coding RNAs, such as snoRNAs and cryptic unstable transcripts (CUTs) [183, 184]. The *S. cerevisiae* Nrd1 complex is trimeric, composed of RBPs Nrd1, Nab3, and DNA/RNA helicase Sen1. Nrd1 and Nab3 function to bind to recognition sites transcribed in

nascent RNA and recruit the nuclear exosome-TRAMP complex for target processing or degradation, while Sen1 functions to disrupt RNAPII from template DNA and effectuate dissociation and transcription termination [185-187]. While Pcf11 binds RNAPII CTD Ser2 phosphorylation to mediate recruitment of CFI factors, Nrd1 differs by recognizing Ser5 phosphorylation to mediate the recruitment of associated proteins [185, 188]. Ser5 CTD phosphorylation drops considerably once RNAPII moves into the body of long protein genes. However, snoRNA genes are much shorter, usually less than 1kB, and show comparatively higher Ser5 phosphorylated signal at their 3' end, which still permits Nrd1 recruitment [189-191].

Greater study of the Nrd1 complex has begun to expand its role to include the processing of mRNAs. In the absence of Rat1-Rai1, Nrd1 has been described to operate as a redundant fail-safe for RNAPII termination. In order to prevent transcription interference, Nrd1 is capable of recruiting the nuclear exosome to read-through transcripts to trigger their degradation, along with an additional termination pathway mediated by ribonuclease Rnt1 [192]. The efficiency of Nrd1 recruitment is affected by the distance of RNAPII from the transcription starts site and the resulting drop of CTD Ser5 phosphorylation [189]. However, in the absence of Rat1-Rai1 mediated termination, Nrd1 inefficient termination became viable and was still observed [192].

Recently, cross-linking and immunoprecipitation data greatly expanded the repertoire of Nrd1-Nab3 associated mRNAs to 25% of yeast protein-coding transcripts. Preferentially associating with smaller genes less than 1Kb in length, Nrd1 and Nab3 were detected in the body of transcripts, along with greater enrichment in the 5' and 3' ends [193]. The high quantity of binding sites detected is possible due to the two alternative simple preferential binding motifs for both Nrd1 and Nab3, UGUA/GUAG and CUUG/UCUU respectively, which are highly prevalent

in the yeast transcriptome [194-196]. While the prevalence of Nrd1 and Nab3 at the 3' end of transcripts is believed to arise as a result of transcription termination, an association of the complex at the 5' end of transcripts is believed to mediate mRNA regulation. Indeed, the 5' end of NRD1 mRNA has been shown to contain a Nrd1 complex recognition site and is responsible for mediating the degradation of the transcript through an auto-regulatory feedback inhibition mechanism to regulate expression [184].

The enrichment of Nrd1 in 5' UTRs was found to occur primarily in transcripts coding for genes involved in biosynthetic or metabolic processes, ideal candidates for regulated expression, suggesting a likely role for Nrd1 in this process [193]. Furthermore, a high-throughput synthetic lethal screen for genes interacting with NRD1 identified a role for the complex in the regulation of cellular growth and response to nutrient availability [197]. Subsequent investigation further implicated Npr1 in a variety of cellular functions including regulation of cell size, regulation of mitochondria proliferation, negative regulation of Ras/PKA signalling, inducible regulation of metabolic transcripts and re-localization into nuclear speckle domains upon glucose deprivation [197]. Therefore it is clear the Nrd1 complex is an important factor with diverse cellular functions which merit further study to decipher fully.

The synthetic lethal screen made use of a temperature sensitive Npr1 mutant, which was quickly observed to cause cells to swell to more than three times their size at the restrictive temperature. As a follow up to genetic interactions identified in a screen using a temperature-sensitive mutant, Npr1 was implicated in the Ras signalling pathway, and its loss of function at the restrictive temperature resulted in an increased number of cellular mitochondria. The Ras pathway functions to regulate cellular growth in response to environmental nutrients in *S. cerevisiae* [198, 199]. Several negative regulators of the Ras signalling pathway were identified

in the synthetic lethal screen, which revealed overactive Ras signalling in Npr1 mutants and dysregulation of cell size, with mutant cells swelling to more than three times their size.

Likewise, Npr1 was found to be a downstream substrate of the Ras signalling

Nuclear Export

Loaded with the requisite proteins and having completed processing, mRNPs diffuse through the nucleoplasm and make their way to the NPC. The nuclear membrane functions to contain the contents of the nucleus from the cytoplasm. The only means of exchange between the two compartments is through the NPCs which stud the membrane the nuclear membrane. The NPC is a large 66MDa protein complexes composed of more than 30 proteins [200]. The nucleoplasmic and cytoplasmic surfaces adopt different structures, while the portion of the complex embedded in the membrane shows eight-fold symmetry with repeating structural units. The protruding nuclear portion of the NPC forms the nuclear basket composed of Mlp proteins Mlp1 and Mlp2, with which mRNPs first make contact and facilitate docking to the NPC. While associated with the nuclear basket, a final quality control check will retain and degrade aberrant mRNPs through recruitment of the nuclear TRAMP complex [200, 201].

Export competent mRNPs make contact with FG Nups lining the central passage of the NPC to allow transport to the cytoplasm. Directional export is established through the action of DEAD-box helicase Dbp5 and its cofactor Gle1, bound to inositol hexakisphosphate (InsP₆), which associates with protruding filaments on the cytoplasmic face of the NPC. Passage through the NPC is believed to trigger the release of several mRNP associated proteins, while others disassociate shortly after reaching the cytoplasm, while others still remain bound until the first round of translation [2, 200, 201].

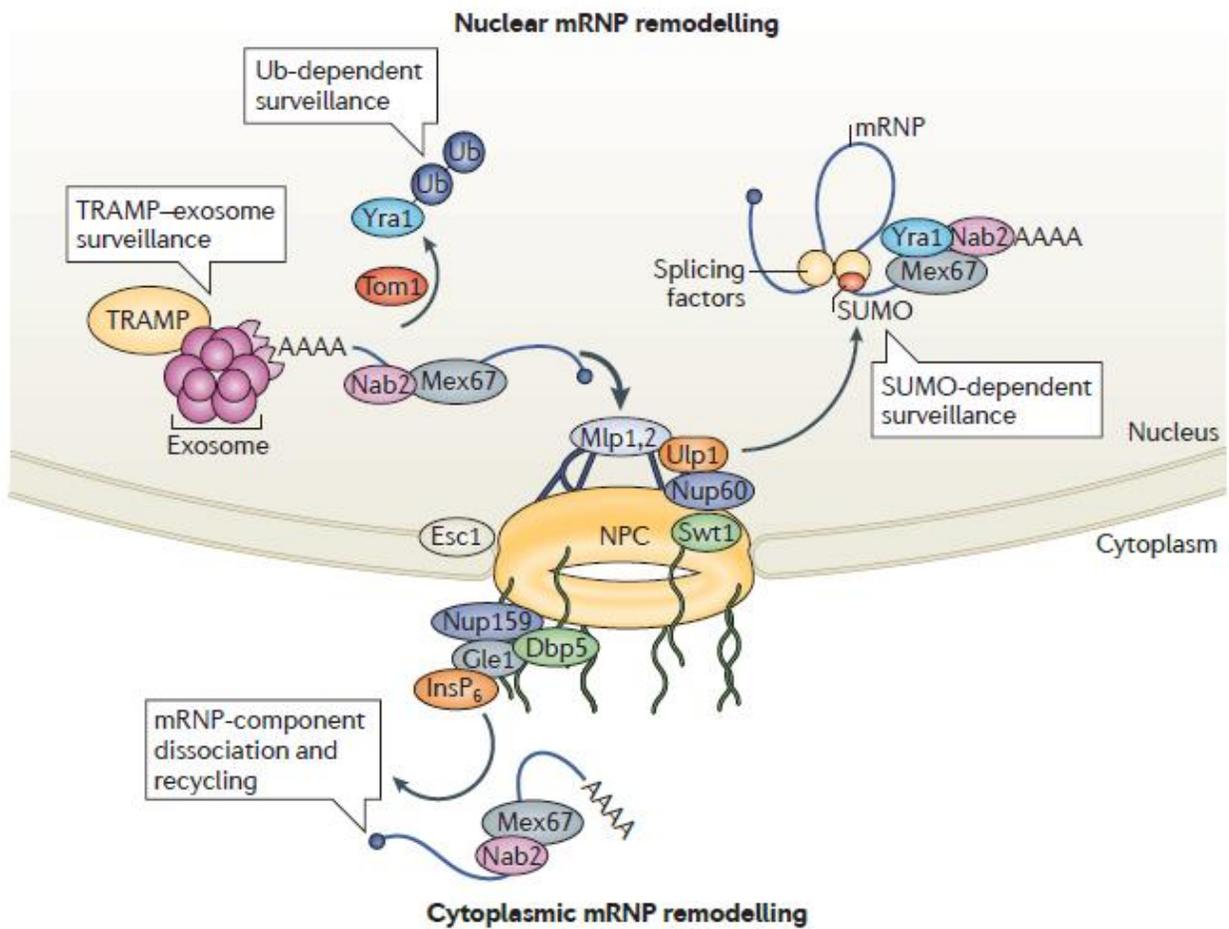


Figure 3. mRNP NPC docking and quality control. mRNPs undergo a final quality control checkpoint once docked to the nucleoplasmic face of the NPC. Final mRNP remodelling events take place prior to the eventual passage through the NPC which triggers the disassociation of several bound protein in the nucleus or shortly after passage to the cytoplasm. Several nuclear mRNP proteins remain associated with the mRNA up to the first round of translation. Figure taken from [201]

mRNP Heterogeneity

The process of mRNP assembly does not take place in a linear manner with clearly defined steps, but rather a number of events may occur concurrently on different regions of the mRNA and to the different copies of proteins presents, which presents a challenge in deciphering the mechanisms of the pathway. As mRNP assembly is dynamic, it is important to realize that there exists no archetypal mRNP complex, as the complement of associated proteins will vary depending on the stage of maturation of the complex. Otherwise said, even when considering only the mRNA molecules resulting from a single gene, the complement of associated proteins will change over the course of time, giving rise to a heterogeneous pool of mRNPs within the nucleus[202, 203]. Although many mRNP maturation factors are essential for viability or crucial for processing steps, not every mRNP factor is recruited to every transcript. An early study sought to identify transcript bound by Yra1 and Mex67 and revealed that each only bound to 20% and 36% of genomic mRNA respectively [204]. These results are surprising given that Mex67 is the only yeast mRNA export receptor and Yra1 plays a crucial role in its recruitment, while gene deletion of both leads to detrimental effects. Furthermore, work in recent years has provided evidence for the existence of sub-classes of mRNPs, mRNPs which consist of a unique composition of proteins which is not found in the majority of nuclear mRNPs.

Introns

One of the most prominent examples pertains to the 5% of yeast mRNAs which contain an intron. The presence of an intron can have a significant influence on the maturation process of the mRNA and the protein composition of the resulting mRNP. A key difference in the maturation of intron-containing mRNAs features the recruitment of the spliceosome [2]. The spliceosome is a large multimeric ribonucleoprotein complex composed of nearly 90 proteins

and snRNPs. Several RBPs factors have been implicated in the recruitment of the spliceosome in yeast, or have been shown to associate preferentially with intron-containing genes [2, 205, 206].

While the splicing reaction may be relatively quick, mRNPs will continue to bear a scar marking them as having been formed from a mRNA containing an intron through the deposition of exon-junction complexes (EJC) at exon-exon junctions [2, 207]. EJCs play a critical role in quality control of intron-containing mRNAs through nonsense-mediated decay (NMD) which occurs in the cytoplasm, targeting mis-spliced or non-spliced mRNAs for degradation. While EJCs are recruited in the nucleoplasm, they remain bound to the mRNA following intron removal and through the NPC to the cytoplasm [2, 207].

3' Processing

While polyadenylation is regarded as an important step in mRNA maturation, not all mRNAs in the cell are subject to this event. In mammalian cells and other metazoan, histone mRNAs are unique in that they are the only known cellular mRNAs which are not polyadenylated [208, 209]. Rather, histone mRNAs contain a unique sequence in their 3' UTR which forms a stem-loop structure, following endonuclease cleavage to release the mRNA from the site of transcription, which is quickly recognized by stem-loop binding protein (SLBP). SLBP remains associated with histone mRNAs in the nucleus and is loaded into the forming mRNP for nuclear export. SLBP is an important regulator of histone mRNA metabolism, influencing translation in the cytoplasm, as well as mRNA processing and decay [209, 210]. In yeast, histone mRNAs have been observed to be polyadenylated, yet the poly(A) tail length has been found to be shorter in comparison to most cellular mRNAs and undergo differential 3' processing according to the cell cycle [209].

Localization

Across a number of species, certain mRNAs have been found to localize to specific regions of the cell to permit localized production of protein, or prime budding daughter cells with transcripts necessary to facilitate growth [211]. In yeast, the prime example of such a transcript is the *ASH1* mRNA encoding for a component of Histone Deacetylase Complex (HDAC) and essential for establishing yeast mating-type, which is loaded into budding daughter cells [212-214]. An intricate series of interactions must take place in the nucleus to permit for faithful localization, which results in *ASH1* mRNA associating with certain protein factors not found to associate with other mRNPs [215]. This process begins with the co-transcriptional recruitment of She2 to nascent mRNA. She2 functions to recruit another protein, Loc1p, which contains a nucleolar targeting domain, which results in the mRNP passing through the nucleolus to recruit Puf6 which is necessary for proper mRNP localization in the cytoplasm. As not all mRNA is targeted for localization within the budding yeast daughter cell, few of the nuclear mRNPs will be loaded with these factors [215]. As such, mRNA localization can result in the loading of transcript specific protein factors into nuclear mRNPs.

Research Project

Project justification

While much work has been done to identify mRNP associated proteins and begin to assign some sort of cellular function to these protein factors, little is known about the structure adopted by a mRNP in the cell. The median gene length in *S. cerevisiae* is 1,500bp, which corresponds to a linear mRNA molecule which measures nearly 450nm in length [216]. Given that the yeast genome measures approximately 2 μ m in diameter, it is understandable that these molecules must be folded into a compact structure to allow for nucleoplasmic diffusion while avoiding entanglement about chromosomes and limiting of macromolecular collisions [217, 218]. While simple nitrogenous base pairing offers some structure to RNA molecules, associated RBPs are likely candidates for effectuating the majority of the structure adopted by the mRNP. While many proteins associated with a mRNPs in the nucleus, the structure must be organized and defined, as the indiscriminate binding of proteins could hinder the accessibility of important regulatory domains necessary for the proper processing and maturation of the mRNA.

While the structures of a number of individual RBPs and mRNP associated sub-complexes have been solved, little is known about the structure or structural composition of mRNPs. Ideally, the structures of macromolecular complexes are resolved by expressing and purifying constituent proteins by recombinant protein technologies and then assembling these complexes in vitro to derive a representative structure which exhibits some degree of functionality. However, given the dynamic process of mRNP assembly which involves the coordinated interplay between several large enzymatic complexes and is regulated by diverse post-translational modifications, it is not conceivable to recreate such a macromolecular complex in vitro which could be conceived to resemble the complex which exists in vivo accurately.

Much of our limited knowledge of mRNP structure has been inferred through experiments involving affinity purification mass spectrometry and UV crosslinking experiments coupled with deep sequencing [43]. Mass spectrometry has permitted the identification of mRNP associated proteins, while UV crosslinking and purification experiments have allowed for the identification of mRNA sequences occupied by RBPs over the transcriptome. The combination of these techniques applied to mRNPs has generated averaged global models of mRNPs where we can identify constituent proteins with an idea of where they bind mRNA. However, a flaw in both these techniques is their reliance on population measurements, which fails to resolve temporal and transcript specific features. For instance, affinity purification mass spectrometry identifies all possible protein interactions made by a target protein of interest, while UV crosslinking experiments identify possible mRNA sequences occupied by the target. However, both these techniques measure all possible interaction made by a target over its lifetime in the cell and are unable to resolve spatial and temporal differences between these interactions. These techniques are unable to distinguish a protein moving and binding to a different mRNA sequencing and forming new protein interactions, versus two copies of the protein present on the mRNA molecule forming these interactions simultaneously and independently at two distinct sites. As such, several major, yet simple, questions pertaining to mRNP biology remains, such as potential variance in the co-occupancy of protein factors on the same transcript over time or between transcripts, and how the abundance of these factors may vary between mRNA transcripts of different properties. As current techniques fail in their capacity to address such questions, there exists a need to develop new methodologies capable of shedding light on this aspect of this field of research.

Little is definitively known about mRNP structure. There are only two defining features common to nearly all nuclear mRNAs, that is the 5' Cap and the 3' poly(A) tail [2]. The 5' cap is bound by the nuclear CBC, while the 3' tail is bound by several poly(A) binding proteins [219]. The location and abundance of proteins in the body of the mRNA are unknown, as is the adopted structure of the mRNP. mRNPs could be closely densely packed with multiple structurally repeating proteins, or more loosely assembled with certain proteins only present in low copy numbers. The only real mRNP structure data consists of low-resolution Cryo-EM pictures of purified mRNPs which revealed an average structure of 30nm in length and 8nm in diameter [102]. However, how well these size constraints apply to all mRNAs and mRNPs is unknown.

The primary goal of this project was to study the heterogeneity of nuclear mRNP and glean insight into their varied structural composition. The long-term aims of this project included deciphering the abundance of protein factors loaded on individual mRNAs, highlighting factors influencing the possible variability of RBPs loaded into different mRNPs and clearly identifying the co-occupancy of protein factors on individual mRNAs. However, due to the limitations of current methods, it was first necessary to develop a new technique capable of providing the necessary information to begin to answer these questions.

With the exception of the CBC, the abundance in terms of absolute stoichiometry of any RBP loaded into a mRNP is unknown. Likewise, while heterogeneity is assumed to exist between mRNPs resulting from mRNA of different properties, such as length, the existence of a variation with general mRNP assembly factors has not been shown, aside from the protein and transcript specific examples described above. As such, this project could be explored through two different avenues. Firstly, it could be desired to purify total nuclear mRNAs and identify the stoichiometry of target factors to show that a distribution does indeed exist across all nuclear

mRNPs. Otherwise, a more targeted approach could be undertaken whereby nuclear mRNPs for a single mRNA transcript could be isolated, and the abundance of a specific RBP could be quantified to identify how many copies of that specific protein associate with a specific mRNA. As the proposed methodology could easily be adapted to address either of these situations, both were investigated and referred to as either the transcript specific or general transcript approach. What follows will be an overview of the proposed method, followed by a more in-depth explanation of elements in the system

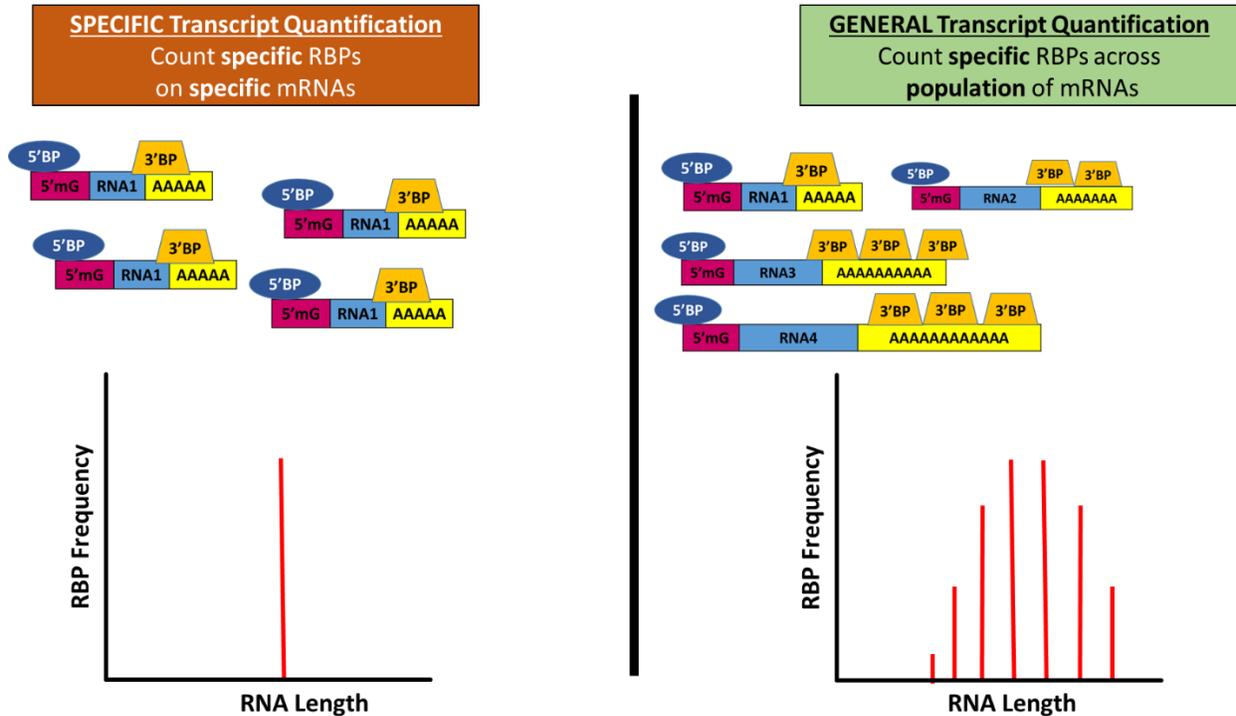


Figure 4. Specific transcript quantification vs. general transcript quantification goals. The example above illustrate the desire to quantify a specific 3' Binding Protein (3'BP) according to the two proposed methods. The goal of the specific transcript approach is to quantify the abundance of a specific 3'BP on a single specific mRNA molecule. The quantification of the specific approach will inform how many 3'BP's binding to a mRNA of a certain length. The general quantification approach will aim to show the distribution of a specific 3'BP over all transcripts in the cell. In the general transcript approach, a large pool of mRNA of different lengths will be purified and the abundance of the 3' BP will be determined across all of these transcripts.

Single-molecule pull-down and imaging methodology

The quantification of RBP stoichiometry on individual mRNAs is thought possible through the combination of existing methods with minor modifications and optimization. As the goal is to study a gene-specific mRNP, an important challenge lies in the ability to obtain a pure sample which maintains a native structure and is representative of the complex's organization in vivo. In order to achieve this purity, the proposed methodology will consist of a two-step purification against two differentially tagged proteins which will greatly reduce the complexity of the sample. The selected tagged proteins vary depending on whether global mRNPs or transcript specific mRNPs wish to be investigated, but the general proposed methodology is otherwise highly simple. In the transcript specific case, the first round of purification will be done to select for a specifically tagged mRNA transcript of interest, while the second round of selection will be for a mRNP associated protein factor of interest. By performing these two sequential purifications, we eliminate all cellular transcripts, save for the one we are interested in, and we then further select for only those transcript specific mRNPs which are bound by our RBP of interest. In this dual purified sample, we avoid purifying free RBP protein or other cellular complexes in which this RBP is loaded, as these are removed and avoided in the first step of the purification.

In order to isolate native protein complexes that are representative of the organization adopted in vivo, it is necessary to perform purification quickly, at cool temperatures and with buffer conditions which promote complex stability. In order to maximize this possibility, the first step of the purification is done using a cryo-lysate and a rapid single-step affinity tag which removes the protein complex from the lysate as quickly as possible and limits the possibility of degradation or protein rearrangement due to residual activity during the purification [141, 220].

Once a complex is isolated from a lysate, the risk of reorganization is reduced, which allows for sample manipulation and a slower second-step purification, if unavoidable. While the first step of the purification is performed in accordance with conventional methods, the second round of purification is performed in a novel manner, as it is performed on functionalized glass coverslips rather than functionalized beads. The purification on coverslips is necessary, as it allows for immobilization of complexes, such that they can be viewed and imaged by fluorescent microscopy, as this served as the means for the quantification of tagged proteins of interest. The quantification of proteins is achieved through exploitation of photo-bleaching of fluorescent tags engineered alongside and in frame with the affinity tags used for the purification, which allows for the inference of protein abundance in individual complexes [221-223].

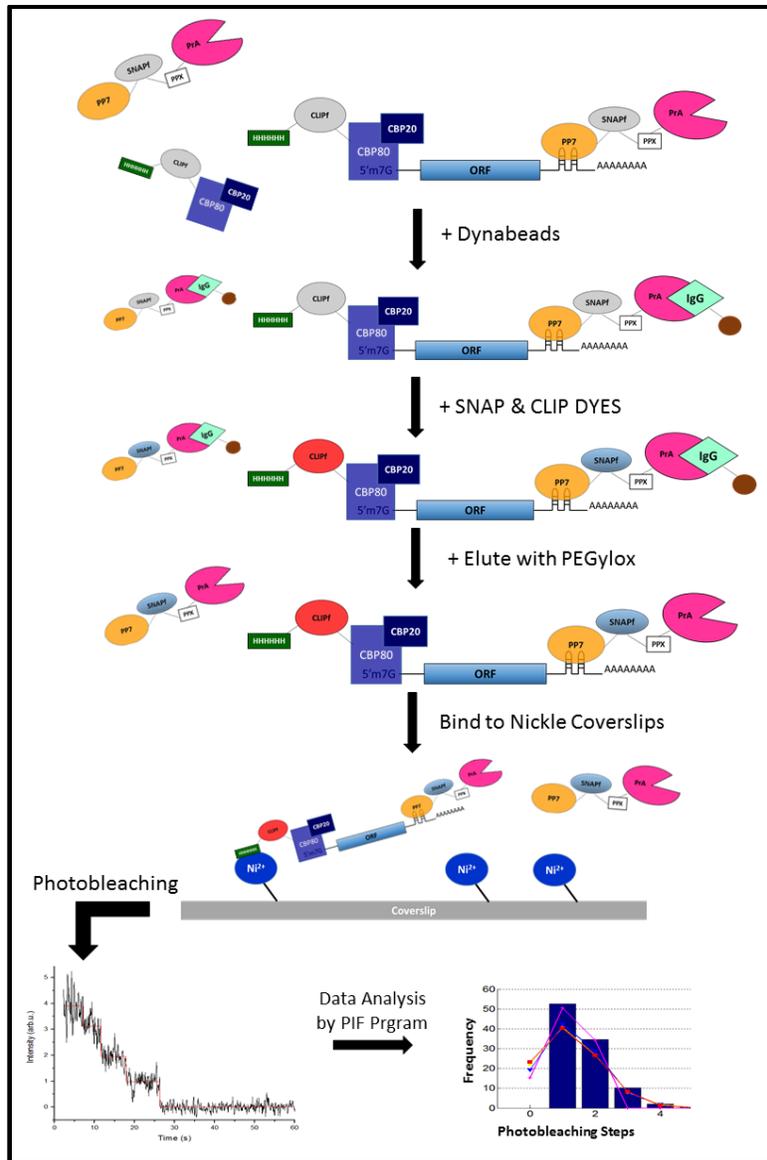


Figure 5. Schematic representation of methodology used for transcript specific protein quantification. Yeast strains are generated where a target gene is tagged with two PP7 stem-loops in 3' UTR. Yeast strains co-express PP7-SNAP-PrA and Cbp80-CLIP-His₆. In vivo, four copies of PP7-SNAP-PrA will bind to the two PP7 stem-loops of the target gene while Cbp80-CLIP-His₆ will bind to the 5' cap. Cryo-lysate is used for ssAP with IgG dynabeads to bind mRNPs from a specific gene. SNAP and CLIP dyes are added to label complexes while attached to beads and unincorporated dye is washed away. Complexes are natively eluted with PEGylox and bound to Ni-NTA coverslips which bind His₆ tagged protein. Free PP7-SNAP-PrA not loaded into mRNP complexes and not associated with Cbp80-CLIP-His₆ are washed away. Complexes are imaged by TIRF microscopy and undergo photobleaching which is measured as a fluorescence trace. Fluorescent traces from individual spots are compiled and analyzed by the program PIF to generate a histogram which indicates the number of fluorescent subunits present in each channel.

The general transcript approach features a very similar methodology, but in lieu of purifying specific mRNA transcripts two RBPs are used as bait instead. Two sequential rounds of purification are performed using two differently tagged proteins of interest. The result is largely expected to be a pool of all cellular mRNAs bound by these two proteins of interest or RNA independent complexes containing both proteins. Identical to the previous approach, the second round of purification is performed on functionalized coverslips and quantification is inferred through photo-bleaching. The result of two purifications is expected to be a pool of all cellular mRNPs which contain both proteins of interest, and not simply one specific mRNA transcript as with the transcript specific approach.

While no method currently exists to address the quantification of proteins loaded within specific mRNPs, it is believed to be possible through the rational combination of existing methods. Below are key elements, which if combined, are believed to generate a method capable of answering this biological question.

ssAP from Cryolysate

Single step affinity purification serves as a rapid means of isolating native protein complexes from cellular lysates of cells from which they are endogenously expressed [220, 224]. The method was devised as a combination of three elements: cryo-milling, use of magnetic dynabeads and use of a rapid and high-affinity tag [141]. Strains of interest are grown and harvested in a manner similar to existing methods, except that they are then rapidly frozen in liquid nitrogen. Cellular lysis is subsequently achieved through ball-mill grinding, with components cooled in liquid nitrogen, which generates a fine cellular powder with particles measuring less than a micron in size. Solid state lysis succeeds in preserving weak protein-protein interactions and reduces the time cellular components may exhibit residual activity, which diminishes the possibility of protein and RNA degradation [141, 220].

Use of magnetic dynabeads allows for high-density conjugation of ligands for affinity purification which aids to increase the kinetics of binding, which reduces the time needed for binding and allows for the capture of targets. As dynabeads are functionalized on their exposed surface, they also present no size exclusion limits for target capture compared to agarose beads, which are porous and require target complexes to be able to diffuse within the beads [141, 220].

Finally, effective purification is achieved by selecting an affinity tag which exhibits high affinity and fast kinetics of binding, even at 4°C at which the binding step of the protocol is performed. The most commonly used affinity tag in this method is the Protein A (PrA) tag which recognizes the Fc region of antibodies. The PrA tag, measuring 27kDa, is capable of binding to antibody conjugated dynabeads in as little as 15 minutes at 4°C and is stable under high salt conditions and mild detergent conditions [224, 225]. These properties allow for rapid purification

of protein associated complexes but come at the expense of difficulty at eluting the complex under native conditions, which has only become possible in recent years through the advent of a synthetic peptide capable of triggering disassociation of the bait protein from the purification matrix [225].

PP7 RNA Tagging System

Purification of specific mRNA molecules from cellular lysate is possible through the use of the PP7 stem-loop system [226]. During the life cycle of the RNA based virus PP7 bacteriophage, the PP7 coat protein (PP7CP) functions to bind a specific RNA hairpin structure in the genomic RNA as a dimer to function as a translational repressor of viral replicas. The PCPs recognition of a specific RNA hairpin structure, which measures less than 30nts, has been exploited in microscopy experiments to monitor single specific mRNA transcripts for years. In these experiments, the 3' UTR of a gene of interest is tagged with a variable number of these RNA hairpin structures, normally ranging from 12 to 24 repeats, and cells concurrently express PCP tagged with a fluorescent protein [227, 228]. PCP is capable of recognizing and binding to the PP7 RNA stem-loops quickly and with high affinity (1nM K_d) [226]. As a result, mRNA transcripts resulting from tagged genes will be bound by fluorescently labelled PCP in vivo, which allows for live cell imaging and monitoring of specific transcripts in the cell [227]. A variant of this system also exists, which makes use of MS2 bacteriophage components, a PP7 related RNA virus whose life cycle also features the recognition of a different RNA stem-loop structure by the MS2 cap protein (MS2CP) with similar high affinity and specificity [227].

These systems have been used to a limited extent to purify mRNAs and their associated proteins from cell lysates by using PP7CP or MS2CP carrying affinity purification tags, rather than fluorescent protein tags [229, 230]. One such recent study reported success in isolating transcript specific nuclear mRNP complexes by performing sequential purifications using tagged Nab2 and MS2CP as baits and subsequently performing mass spectrometry and proteomic analysis to identify associated proteins [231]. As such, it is possible to isolate transcript specific mRNA molecules by tagging the 3' UTR of a candidate gene with the appropriate stem-loop structure and co-expressing PP7CP with the desired affinity tag. While in vivo microscopy experiments require target mRNA molecules to be tagged with multiple stem-loop structures to improve the signal to noise ratio better and offer better resolution, such is not the case with affinity purification where a simple pair of stem-loop structure which will recruit two dimer pairs of tagged PP7 is expected to be sufficient to permit for the purification of the mRNP.

SNAP and CLIP site-specific fluorescent protein labelling

While GFP is the classical protein tag used for fluorescent microscopy experiments, it is non-ideal for single-molecule experiments. The major limitations of GFP relate to its maturation as an active fluorescent protein. Not only does GFP require nearly an hour maturation time to become photoactive following the protein's synthesis, but a significant proportion of molecules fail to mature at all [232, 233]. Single-molecules fluorescent microscopy experiments are dependent on a high degree of photo-activity of supposedly fluorescent molecules. As such, an improvement on the use of fluorescent proteins features the use of covalent labelling protein tags [234]. While many variants exist, the general principle is that these protein domain tags will specifically recognize substrates conjugated to fluorescent organic dyes and add these organic

dyes to site-specific regions of the protein tag. Two such tags are the SNAP and CLIP tags, developed and commercialized by NEB [234, 235].

The SNAP and CLIP tags are derivatives of one another, with the SNAP-tag developed first from human DNA repair enzyme O⁶-alkylguanine-DNA-alkyltransferase (hAGT). The SNAP-tag, measuring 20kDa, functions to specifically recognize O⁶-alkylguanine derivatives and, through an enzymatic reaction, will covalently link its substrate to a site-specific cysteine through a covalent bond. The CLIP-tag is a mutated version of the SNAP-tag and functions in the same way, save for its recognition of benzylcytosine derivatives. The SNAP and CLIP-tags were developed to permit for rapid, specific, simultaneous and orthogonal labelling in vivo or in cell lysates. The SNAP and CLIP-tags were developed to permit for rapid, specific, simultaneous and orthogonal labelling in vivo or in cell lysates under a variety of buffer conditions and temperatures ranging from 25°C to 37°C [234, 235].

Labelling reactions are reported to have a greater than 90% labelling efficiency with substrate concentrations of 10uM in 30-45min for both SNAP and CLIP tags. Nonetheless, mutant variants of both the SNAP and CLIP tags, called SNAPf and CLIPf, were developed which report similar labelling efficiencies in only 15-30min [236, 237]. The efficiency of labelling varies slightly between the SNAP and CLIP tags, with the SNAP-tag generally reported as having a greater efficiency, as well as variation based on the nature of the fluorescent conjugated substrate, used [234, 237] [238].

Organic fluorescent dyes tend to be brighter and offer greater photo-stability in comparison to fluorescent proteins. In combination with a high labelling efficiency yielding a greater pool of fluorescently labelled proteins, the use of SNAP and CLIP tags offer significant advantages over the use of fluorescent proteins and are more amenable to high-resolution

fluorescent microscopy experiments. To that effect, two groups have published works making use of SNAP and CLIP tags to investigate spliceosome assembly and reaction dynamics through high-resolution single-molecule microscopy pulldown experiments [223, 236].

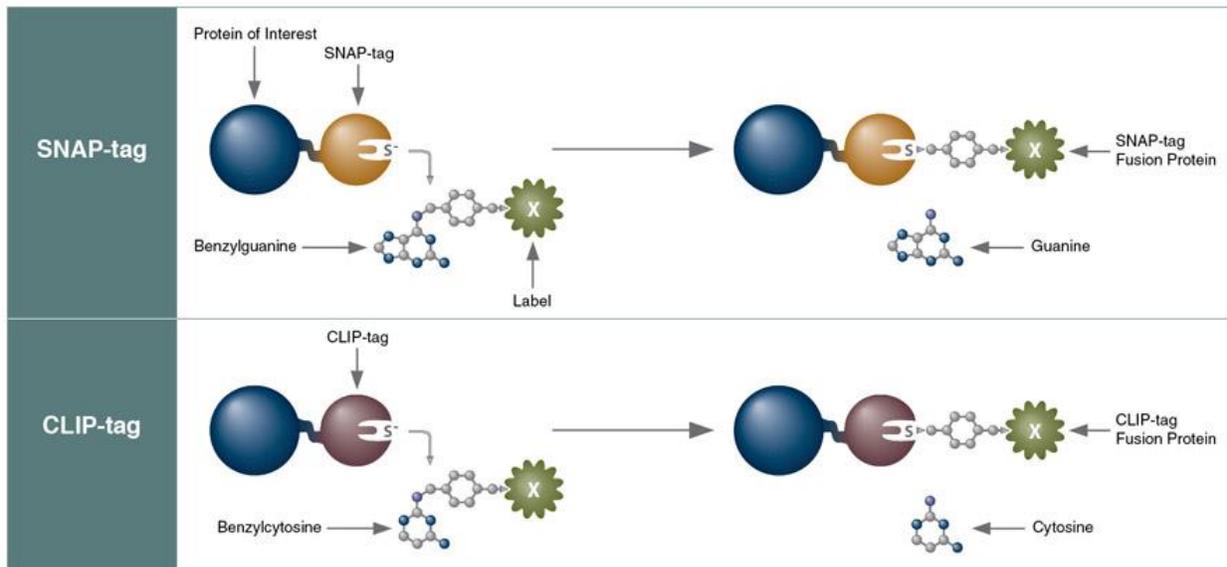


Figure 6. Schematic representation of SNAP and CLIP labelling reactions. SNAP and CLIP are 20kDa covalent protein labelling tags which covalently attach substrate conjugated dyes to a site-specific cysteine residue within each tag. SNAP and CLIP tags recognize different substrates which permit for simultaneous and orthogonal labelling of tagged proteins. (Image source: www.international.neb.com)

Single-molecule pull-down experiments

In recent years, several groups have published works featuring the high-resolution imaging of purified proteins by fluorescent microscopy. Common to many of these protocols is the immobilization of proteins or protein coverslips to allow for fluorescent imaging or event quantification [221, 223].

One of the earliest published works featured the purification of GFP tagged proteins from cell lysates on coverslips conjugated with anti-GFP antibodies in a functionalized layer of polyethylene glycol (PEG) [221]. The coating of coverslips in a layer of PEG functions to reduce non-specific binding of proteins to glass, while the conjugation of antibodies functions to select, enrich and immobilize tagged proteins of interest on coverslips. Following washing of coverslips to reduce non-specific binding, bound fluorescent proteins could be analyzed and protein abundance inferred through photobleaching. Photobleaching is the process by which a fluorophore permanently loses its ability fluoresce due to a chemical alternation sustained while in the excited [239]. Photobleaching can be induced through prolonged and intense imaging of fluorescent molecules, when observing a fluorescence trace over time, a photobleaching event is observed as a rapid and sudden drop in fluorescence, reminiscent of a step function. This phenomenon can be exploited to infer abundance and stoichiometry of a particular fluorescent protein loaded within a protein complex, as the number of photobleaching events, and therefore steps directly reflect the number of copies of fluorescent protein present [222, 239].

Biophysical studies aimed at investigating protein structural dynamics through the use of fluorescent microscopy have also developed similar techniques for immobilization and imaging of proteins on coverslips. As most of these studies deal with one or a small number of proteins, these target proteins may first be expressed and purified through recombinant protein

technologies, generating protein samples which are significantly more pure and homogenous than those purified from a cellular lysate [222]. As the samples are of greater purity, there is a limited concern of non-specific protein binding to coverslips, and therefore these coverslips are seldom functionalized with blocking agents like a layer of PEG. Instead, coverslips are simply functionalized to react with the affinity purification tag carried by the protein of interest. The most common case features the use of a polyhistidine (His₆) tag to recombinantly purify proteins; therefore coverslips are functionalized to immobilize nickel ions and permit His₆ tagged protein binding, in the same manner as with immobilized metal affinity chromatography (IMAC) purification with agarose beads [240, 241].

Materials and methods

Yeast strains:

Yeast strain W303a (MATa {ade2, his3, leu2, trp1, ura3}) was used as the background yeast strain for all yeast experiments unless otherwise specified [242]. Below is a list of yeast strains generated for this work:

Table 2. Yeast strains generated for use in this work

PMA1-2xPP7
CBP80-CLIPf-His6-HIS3
TIF4631-CLIPf-His6-HIS3
CBP80-CLIPf-His6-HIS3 PMA1-2xPP7
TIF4631-CLIPf-His6-HIS3 PMA1-2xPP7
CBP80-CLIPf-His6-HIS3 PMA1-2xPP7 pPP7-SNAPf-PPX-PrA::URA3
TIF4631-CLIPf-His6-HIS3 PMA1-2xPP7 pPP7-SNAPf-PPX-PrA::URA3
Cbp80-CLIPf-His6::HIS3 PP7-SNAPf-PrA
Tif4631-CLIPf-His6::HIS3 PP7-SNAPf-PrA
Pab1-CLIPf-TEV-His10::HIS3
Nab2-CLIPf-TEV-His10::HIS3
CDC33-SNAPf-PPX-PrA::KanMX, Nab2-CLIPf-TEV-His10::HIS3
CBP80-SNAPf-PPX-PrA::KanMX, NAB2-CLIPf-TEV-His10::HIS3
CBP80-(GSG)4-SNAPf-PPX-PrA::URA3, Nab2-CLIPf-TEV-His10::HIS3
CBP80-(GSG)4-SNAPf-PPX-PrA::URA3
CBP80-SNAPf-PPX-PrA::KanMX
Cbp80-(GSG)4-SNAP-PrA::KanMX
Cbp80-(GSG)4-SNAP-PrA::KanMX pHis6-BirA::URA3
Cbp80-(GSG)4-SNAP-PrA::KanMX pHis6-BirA::URA3 Pab1-(GSG)4-CLIP-AP::HIS3
Cbp80-(GSG)4-SNAP-PrA::KanMX Pab1-(GSG)4-CLIP-AP::HIS3

Yeast Media:

YPD (1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) glucose) was used as the base media to grow yeast cells. YPD media was supplemented with antibiotics when selection was required, namely G418 (gentamycin) was used at a concentration of 200ug/mL for a first-round selection, followed by 600ug/mL for a second-round selection.

Drop-out media was used when the selection was performed using auxotrophic markers, such as URA3, TRP1, LEU2 or HIS3. Drop-out media was composed of 1.5% (w/v) yeast nitrogenous bases (amino acid and ammonium sulfate free), 5% (w/v) ammonium sulfate, 1x amino acid mix (except for amino acid used for selection), 2% (w/v) glucose. [243]

PCR:

For yeast tagging and cloning, iProof (NEB) was the standard Taq polymerase used according to the manufacturer's instructions. For long or complicated sequences, LaTaq (Sigma) was used according to manufacturer's instructions. For control PCR to validate cassette integration when performing yeast tagging or validate vector constructs generated by cloning, Taq polymerase (NEB) was used according to manufacturer's instructions.

Yeast tagging:

Primers were designed to contain 45-60bp homology boxes to the regions where integration was desired, in addition to the sequences necessary to amplify tagging or deletion cassettes from plasmid vectors. For endogenous tagging of a yeast locus, 4x50uL PCR reactions were set up and performed, followed by phenol-chloroform extraction and DNA precipitation at 80C with sodium acetate and ethanol. Typically 1ug of PCR product was used per yeast transformation or 500ng of purified plasmid if plasmid integration was desired.

Competent Yeast Cells

Yeast transformation was performed by first generating competent yeast cells by the standard lithium chloride (LiCl) protocol. Briefly, a 10mL starter culture was first inoculated and left to grow overnight (12-16h) at 30°C. The following morning, the starter culture was used to inoculate 50mL of fresh media at 0.300 OD and left to grow at 30°C for approximately 3 hours until the O.D. reached 0.8 and cells re-entered log phase. 50mL cultures were then harvested by centrifugation, and pellets washed suspended and washed twice with sterile H₂O. Cells were then washed 4 times with TeLiAc buffer (10mM Tris pH 8.0, 1mM EDTA, 100mM LiAc) and competent cells were finally suspended in 1mL TeLiAc buffer. [244, 245]

Yeast transformation protocol

Yeast transformation reactions were set up as follows in accordance with previously described methods [243]. 50uL of competent yeast cells were mixed with 5uL carrier DNA (10 mg/mL salmon sperm DNA), 10uL of re-suspended DNA to be transformed and 300 uL of PEG 3500 50% w/v, 10mM Tris pH 8.0, 1mM EDTA, 100mM LiAc. Transformation reactions were mixed and incubated at 30°C for 15min followed by 42°C for 30min. Cells were pelleted in a microcentrifuge, suspended in 100uL TE buffer and plated on appropriate dishes containing agar media. In the case of a transformation with a drug-resistant marker, such as KanMX, transformed cells were suspended in 1mL of YPD media following incubation at 42°C and incubated overnight at room temperature in a microcentrifuge tube prior to plating on agar plates.

Screening of yeast colonies

Following yeast transformation, yeast cells were left to grow on agar plates for two days, or until colonies could be observed. Plates were replica plated onto agar plates with the same

selection and left to grow for two days, or until colonies could be observed. Colonies were picked from the replica plate and screened by PCR for cassette integration and western blot to validate the expression of the tagged protein with the added affinity or epitope tag

Yeast chromosomal extraction

YPD media was inoculated with yeast colonies and left to grow overnight at 30°C. The following day, 2mL of culture was collected by centrifugation in a micro-centrifuge tube. Cells were re-suspended in a mixture of 200uL of breaking buffer, 200uL phenol-chloroform and 200uL of glass beads. Cell pellets were vortexed for 5 minutes at room lyse cells. 200uL of TE buffer was added and samples were mixed. Samples were then centrifuged at 4°C for 10min at max speed in a microcentrifuge (16,700g). The supernatant was collected and transferred to a new tube. 1mL of 100% ethanol was subsequently added, followed by a mixing of the samples and a 20 minute incubation at -20°C. Precipitated DNA was then collected by centrifuging samples at 4°C for 10min at max speed. The DNA pellet was then washed twice with 70% ethanol, and the DNA pellet was left to air dry at room temperature for 10 minutes. DNA was then re-suspended in 50uL H₂O and the concentration determined by measuring OD₂₆₀ with a nano-drop. DNA concentrations were standardized to 300ng/uL and subsequently used in a control PCR reaction to validate the integration of the intended cassette at the desired chromosomal location. [243]

Separation of DNA Fragments

DNA fragments from PCR or molecular cloning reactions were separated by electrophoresis on TAE agarose gels and visualized through staining with GelRed dye.

Total yeast protein extraction

2mL of overnight yeast cultures were collected in a microcentrifuge tube. Pellets were suspended in 250uL lysis buffer (1.85M NaOH, 7.5% (v/v) β -mercaptoethanol) and incubated on ice for 10min. An equal volume, 250uL, of 50% TCA was added, and samples were mixed and incubated on ice for 15 minutes. Proteins were pelleted in a centrifuge by spinning samples at max speed for 10 minutes at 4°C. The supernatant was discarded and pellets were washed with 500uL acetone for 20 minutes at -20°C. Samples were spun again at max speed for 10 minutes at 4°C and the supernatant discarded. Pellets were suspended in a mix of SDS sample buffer, first 100uL of Solution A (0.5M Tris pH 6.8, 5% SDS) and then 100uL of Solution B (75% glycerol, 1.9% (w/v) DTT, 0.05% Bromophenol Blue). Samples were heated to 95°C for 10 minutes and spun at max speed in a centrifuge at room temperatures. Samples were loaded on an SDS-PAGE gel immediately or stored at -20°C. [243, 246]

SDS-PAGE

Protein samples were either loaded on a homemade polyacrylamide gel (Bio-Rad) or pre-cast 4-12% Bis-Tris NuPAGE gels. Homemade SDS-PAGE gels were run in Tris-Glycine running buffer at 100V through the stacking gel and then 120V through the separating gel until the bromophenol blue marker left the gel, nearly two hours in total. NuPAGE gels (Invitrogen) were run in 1X MES buffer at 200V for 25 minutes according to manufactures instructions [247]

Western blotting

Following gel electrophoresis, proteins were transferred to nitrocellulose membranes according to standard procedures [248]. Briefly, wet transfers were set up in transfer buffer

composed of 25mM Tris (pH 8.3), 192 mM glycine, 20% v/v methanol, 0.1% SDS and run in a cold room for 1 hour at 100V or 30V overnight. Transfers were verified by Ponceau S staining (0.1% Ponceau S, 5% (v/v) CH₃COOH). Membranes were blocked with 5% (m/v) skim milk in TBST (50mM Tris pH 7.4, 150mM NaCl, 0.1% Tween) for 1h at room temperature. Primary antibodies were incubated according to their specification, namely 1h at room temperature or overnight at 4°C. Membranes were then washed 4 times for 5 minutes with TBST and re-blocked again with 5% milk TBST for 20 minutes at room temperature. Membranes were then incubated with secondary antibodies for 1h at room temperature in TBST. Membranes were washed 3 times quickly with TBST and then 3 times for 5 minutes with TBST. Membranes probed with HRP conjugated antibodies were incubated with 1mL ECL solution for 1 minute and then placed in a cassette to be developed by film exposure or through scanning with a gel dock equipped with the appropriate filters. Membranes probed with fluorescently conjugated antibodies were directly imaged with a gel dock or Typhoon membrane scanner.

Yeast spotting Assay

Yeast cultures were left to grow overnight. The following day, the optical density of overgrown cultures were measured by spectrophotometry and subsequently diluted to 0.15 OD in sterile water. The diluted cultures were subsequently serially diluted 10 fold giving a final 1000 fold dilution. 5uL of each dilution was spotted on agar plates of the appropriate media and left to grow the desired temperature, for wild-type cells this was YPD at 30°C. Plates were incubated for several days to allow cells to grow and subsequently imaged to identify phenotypic growth defects [243]

Cell harvest and cryo-lysis

Yeast cells were grown and harvested as previously described [141, 220]. Briefly, 6L media in three 3L flasks were grown to an OD₆₀₀ of 0.8 and collected by centrifugation at 4000g for 5min at 4°C. Yeast cell pellets were transferred to falcon tubes by re-suspending in cold water and pelleted twice to wash the cells. Pellets were then re-suspended on ice in an equal volume of PVP buffer and spun twice at 2600g for 15 minutes at 4°C to remove the supernatant completely. The final pellet was passed through a syringe into a falcon tube filled with liquid nitrogen to generate frozen cell noodles, which were then stored at -80°C.

Cell noodles were lysed using a ball mill grinder in a metal jar cooled in liquid nitrogen, as previously described. Briefly, the metal jar fitted for the Retsch PM-100 ball mill grinder and associated ball bearings were cooled for at least 5 minutes in liquid nitrogen. Noodles were placed into the jar and cryo-lysed by operating eight cycles of 400 RPM, 1min rotation clockwise, 1 min rotation counterclockwise and a final 1min rotation clockwise, with jar cooled in liquid nitrogen for at least two minutes between each cycle. The final powder generated was collected and stored -80°C until needed.

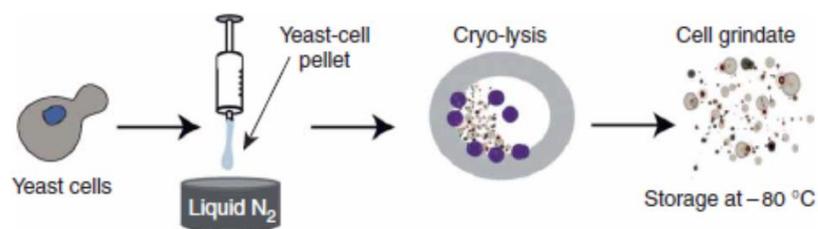


Figure 5. Cryo-lysis and storage of tagged yeast strains. Yeast cells are grown and harvested in log phase. Cells are collected by centrifugation, washed in PVP buffer and frozen in liquid nitrogen. Cells are lysed by ball mill grinding using a container and ball bearings cooled in liquid nitrogen. The grindate powder is stored at -80°C until use. [141]

Dynabead conjugation with rabbit IgG

Magnetic beads with epoxide-reactive groups were conjugated with rabbit IgG to a functional density of 160ug IgG per milligram of dynabeads, with a final concentration of 0.15ug per microliter of the solution as previously described [141, 220].

Briefly, a bottle of 300mg of M-270 epoxy Dynabeads (Thermo Fischer) was suspended and washed twice in 16 ml of 0.1M NaPO₄ pH 7.4 buffer for 10min at room temperature on a rocker. The supernatant was removed by first placing tubes against a magnetic holder to allow recovery of the beads, permitting the supernatant to be easily discarded. Beads were evenly suspended in 20ml of antibody conjugation buffer (2.45mg/mL rabbit IgG, 50mM NaPO₄ pH 7.4, 1M (NH₄)₃SO₄) and incubated overnight at 30°C for 18-24h with rotation. The following day, beads were washed quickly in succession with 100mM Glycine pH 2.5, 10mM Tris pH 8.8 and 100mM N(CH₂CH₃)₃. Beads were then washed four times for 5 minutes with PBS, once for 5 minutes with PBS with 0.5% Triton X-100 and once for 15 minutes with PBS with 0.5% Triton X-100, before finally being stored in 2mL PBS with 0.02% NaN₃ at 4°C.

Single step affinity Protein A purification

Single step affinity protein A purification was performed as previously described [141, 220]. Purifications were performed using TBT100 buffer (20mM HEPES pH 7.4, 100mM NaCl 110mM KOAc, 0.5% Triton X-100, 0.1% Tween-20, 1mM DTT, 1mM MgCl₂, Solution P 1:100, Antifoam B 1:5000). Purifications were performed using an optimized powder to bead ratio (see below in binding assay), typically using 250mg of powder re-suspended in a 1:4 ratio with 1ml of TBT100 buffer and incubated with 50-100ul of rabbit IgG dynabeads. Beads were washed by adding 1mL of buffer, shaking beads by inversion, then placing beads on a magnetic stand and

discarding the supernatant. A total of three washes with 1mL of buffer was performed to equilibrate the beads which were then kept on ice in a minimal amount of buffer. The cryo-lysate powder was weighed into a 2ml microcentrifuge tube and left to thaw for two minutes on ice. 1mL of chilled TBT100 buffer was added and the sample was vortexed for 30 seconds at room temperature and then spun in a 4°C microcentrifuge at 10,000g for 7 minutes. The supernatant was collected and transferred to the prewashed and equilibrated magnetic beads. Tubes were then placed on a rocker at 4°C for 30 minutes to bind Protein A tagged bait proteins and associated protein complexes to beads. Tubes were then placed on a magnetic rack with the supernatant discarded and the beads re-suspended and washed six times with 1mL of TBT100 buffer. For western blot and gel analysis, beads were mixed with an appropriate amount of a mixture of SDS sample buffers Solution A and Solution B and loaded on a gel along with samples from Input, supernatant, pellet and the flow through following the 30 minute binding.

Silver staining of proteins in polyacrylamide gels

Following separating of protein samples by SDS-PAGE, gels would be silver stained using the commercially available “SilverQuest™ Silver Staining Kit” from thermos fisher. The silver staining protocol would be performed as described by the manufacturer.

RT-PCR

Reverse transcription PCR was used to query for mRNAs associated with purified mRNP. ssAP was performed as described above up to the washing of the complexes bound to beads with TBT100 buffer. Beads were washed with 1mL of last washed buffer and subsequently suspended in 500uL last wash buffer to which 5uL of 10% SDS and 5uL of 2mg/mL Proteinase K were added. Samples were incubated at 30°C for 30min to degrade

proteins in the sample. 500µl of Phenol:Chloroform: IAA was added, and samples were vortexed for 30 seconds and subsequently spun at max speed for 2 minutes in a microcentrifuge. The supernatant was collected and transferred to a new tube. In order to precipitate RNA, 3 volumes of 100% EtOH and one-tenth volume of 3M NaOAc was added, samples were mixed and stored at -80°C for 1h. Samples were then spun at max speed for 10 minutes in a cold microcentrifuge, with the supernatant discarded and pellets washed twice with 500uL 70% EtOH. Pellets were allowed to air dry and subsequently re-suspended in 20uL DEPC treated water.

Reverse transcription reactions with Superscript II (Invitrogen) were set up according to manufacturer's instructions with either 500ng dT₁₈ or 100ng pdN₆ random primers or 1uL of resuspended RNA. Following the reverse transcription reaction, 1uL of the reaction mix was taken and directly used in PCR reactions with PMA1 or ACT1 gene-specific primers. The resultant PCR reaction was run on an agarose gel to visualize DNA bands.

Native Protein A Elution

Elution of Protein A bound complexes and associated proteins were eluted under native conditions using PEGylOx as previously described [225]. Briefly, the synthetic peptide PEGylOx was ordered as 1mg aliquots of lyophilized powder (21st Century Biochemicals) and re-suspended in PEGylOx Buffer (20mM HEPES pH 7.4, 100mM NaCl, 0.01% Tween-20, 1mM EDTA, 5% EtOH) at a concentration of 2-3mM. The suspended powder was mixed for 5 minutes at room temperature and then centrifuged at room temperature in a microcentrifuge at max speed (14,000g). The supernatant was collected and transferred to a new tube. The PEGylOx solution was measured on a nanodrop at 280nm to determine the concentration and then stored at -20°C until use.

In order to natively elute Protein A tagged proteins bound and associated complexes from IgG dynabeads, the single step affinity purification was performed as described above. Following washing of beads six times with 1ml of TBT100 buffer, beads were washed once with 1ml of PEGylOx Buffer and the supernatant was subsequently removed. 10ul of thawed PEGylOx solution was added per 15ul of IgG beads and incubated at room temperature for 15 minutes with rotation, with the supernatant containing eluted proteins collected by first placing the tube on a magnetic rack and transferring the supernatant containing eluted proteins transferred to a new tube.

Bead Saturation binding assay

In order to determine the optimum quantity of beads necessary to capture the tagged protein of interest, a bead saturation binding assay was performed. Approximately 250mg of cryo lysate powder was resuspended in 1ml of TBT100 buffer, vortexed and cleared by centrifugation. 100UL (1/10th) of the supernatant was then incubated with varying amounts of washed and equilibrium beads, ranging from 1ul to 20ul. The single step purification was then performed as previously described, with input, flow-through and elution samples retained and loaded on SDS-PAGE gel for subsequent western blot analysis. Developed western blots were analyzed for the bead concentration which left little to no tagged protein remaining in the flow-through after a 30 minute incubation. This bead concentration was subsequently proportionally scaled up and used as the bead amount needed for that particular protein purification in subsequent experiments.

Targeted Bradford Assay

Targeted Bradford assays were performed to quantify the amount of tagged protein binding to affinity capture IgG dynabeads. 250mg of cryo-lysate was suspended in a more

stringent variant of TBT buffer containing 350mM NaCl, 0.01% SDS, 100mM urea and 10ug/mL RNase A. The single step purification protocol was performed as described above with the optimal amount of beads previously determined by the beads saturation binding assay. A fraction of the eluted samples were loaded on an SDS-PAGE gel along with a series of standard dilution of BSA, ranging from 0.09ug to 0.9ug. Following separation of the proteins by electrophoresis, the gel was stained with Coomassie blue and destained with methanol and acetic acid solution to reveal protein bands according to standard protocols. Gels were then scanned and bands corresponding to the known size of the proteins of interest were quantified with ImageJ software and the amount of protein present was determined by generating a standard curve with the quantification of BSA standards.

SNAP and CLIP dye preparation

SNAP-Surface-Alexa488 (SNAP-488) and CLIP-Surface-647 (CLIP-647) dyes were purchased from NEB, with each 50pmol pack suspended in 50ul DMSO and stored as 5ul dry aliquotes at -80C following brief desiccation in a speed vac. When necessary for labelling reactions, a 5uL dry aliquot was resuspended in 5ul DMSO, further separated as five 1ul aliquots which were then brought to a volume of 5ul as working stock. All aliquots were stored at -80C when not in use.

Relative time course labelling assay

Time course labelling assays were performed to determine the optimal time necessary to achieve maximal labelling of SNAP and CLIP tagged proteins with the relative SNAP-488 and CLIP-647 dyes, similar to as previously described [249]. The on bead labelling assay was performed by first performing the single step affinity purification with IgG conjugated dynabeads as described above. Following binding of tagged proteins for 30 minutes at 4C and washing the beads 3 times with 1ml TBT100 buffer, beads were suspended in 110ul of TBT100 buffer. 10ul of was initially removed as the zero time point. Subsequently, 1ul (200 pmol) of SNAP-488 or CLIP-647 or both were added to the remaining 100ul of bead slurry and left to incubate at room temperature with rotation in the dark. At the desired time intervals of 15, 30, 45, 60, 90 and 120 minutes, 10uL aliquotes were removed and immediately SDS sample buffer was added and samples placed at 95C for 10 minutes. All samples were loaded on an SDS-PAGE gel, separated by electrophoresis and then scanned with a Typhoon scanner using the filter appropriate for each dye under investigation. Visualized bands were then quantified using ImageJ to determine the time necessary to achieve maximal labelling.

In solution labelling assays were performed in a similar manner. Cryo lysate was first suspended in 1ml TBT100 buffer and cleared by centrifugation. A 100uL sample of the supernatant was removed and placed in a new tube, to which 1ul (200pmol) of the appropriate dye was added and left to incubate and room temperature with rotation in the dark. At the desired time intervals of 15, 30, 45, 60, 90 and 120 minutes, 10uL aliquotes were removed and immediately SDS sample buffer was added and samples placed at 95C for 10 minutes. Samples were then loaded on an SDS-PAGE gel and analyzed as described above.

Mass spec labelling efficiency assay

In order to determine the absolute labelling efficiency, SNAP and CLIP peptides labelled with dye were quantified by mass spectrometry. 250mg of cryolysate was suspended in 1ml of TBT100 buffer and the single step purification protocol was carried out as described above. Following washing of the beads with 1ml of TBT100 buffer three times following the binding at 4°C for 30 minutes, beads were suspended in 100ul TBT100 buffer and 1uL (200pmol) of CLIP-647 and SNAP-488 dyes were added. Samples were incubated at room temperature for 90 minutes with rotation in the dark. The supernatant was removed and samples were washed three times with 1ml of TBT100 buffer, once with 1ml Last Wash Buffer (100mM NH₄OAc, 0.1mM MgCl₂) quickly, once with 1ml Last wash buffer for 5 minutes at room temperature with rotation and once with 1ml 20mM Tris pH 8.0. Beads were resuspended in 50ul 20mM Tris pH 8.0, 10mM DTT and incubated at 37°C for 30min. Iodoacetamide was added to a final concentration of 10mM and incubated for 30 minutes in the dark at room temperature. Additional DTT was added to a final concentration of 40mM and then one volume of 20mM Tris pH8.0, 20% acetonitrile. 500ng Trypsin (Sigma) was added and samples were incubated at 37°C with 1000rpm rotation in a microcentrifuge block overnight (16 hours). The following day tubes were placed on a magnetic rack and the supernatant was transferred to a new tube with formic acid added to a final concentration of 2%. Digested peptides samples were given to the IRCM proteomics facility, cleaned on MCX columns according to manufacturer's instructions and analyzed on an LTQ-Orbitrap mass spectrometer.

The spectrum corresponding to the various mass charge ratios of the tryptic peptides corresponding to the dye conjugated and unconjugated forms the SNAP and CLIP-tags were identified in the reconstituted total ion chromatogram (TIC). The area of elution peaks in the TIC

was quantified and used as a proxy for ion concentration to permit for comparative quantification and identification of labelling efficiency in samples.

Plasmid cloning

pFA6a-SNAPf-His₆-HIS3 was generated by cloning the SNAPf gene from pAAH0034, with primers encoding His₆, into pFA6a-yEGFP-His3 by removing yEGFP. pKAN-2xPP7-loxP-KanMX-loxP was generated by removing the 12xPP7 sequence of pKAN-12xPP7-loxP-KanMX-loxP and cloning overlapping primers encoding 2xPP7 RNA hairpin sequence as BamHI/BglIII restriction digested fragments. pURA-ADE3p-PP7-CLIPf-PPX-PrA-ADH1 was generated by first removing the 2xGFP-CYC sequence from pURA-ADE3p-PP7-2xGFP-CYC by restriction digest. The sequences of PPX-PrA-ADH1 and CLIPf were amplified independently from pFA6a-PPX-PrA-ADH1 and pAAH0026 respectively by PCR, and subsequently mixed for the second round of PCR to generate CLIPf-PPX-PrA-ADH1 fragments which were then cloned into pURA-ADE3p-PP7 to generate the final plasmid.

Table 3. Plasmids generated and used in this work

pFA6a-CLIPf-His6-HIS3 clone 1
pKAN-2xPP7-loxP-KanMX-loxP V4 Clone 1
pFA6a-CLIPf-His10-HIS3 clone 1
pURA3 PADE3-PP7-SNAPf-PrA-tADH1 Clone 1
pCLIPf-AP-HIS3
pPCYC1-PP7-SNAP-PPX-PrA-URA3
pPSTE5-PP7-SNAP-PPX-PrA-URA3
pET28-PP7-SNAPf-PPX-PrA-CLIPf-His10 (PSACH) clone 6
pFA6a-SNAPf-PPX-PrA-KanMX
pFA6a SNAPf-PPX-PrA URA3
pVTU260-His6-BirA

PSACH recombinant protein purification.

The sequences of PP7-SNAP-PrA and Clip-His₆ were amplified by PCR, such to retain their linker sequences, ligated through a 3' and 5' PacI restriction site and cloned into bacterial expression vector pET28-MHL. The validated plasmid was transformed into Arctic Expressed DE3 bacterial cell for protein expression of PP7-SNAP-PrA-CLIP-His₆ (PSACH). Protein expression was induced with 0.1mM IPTG for 20h at 16⁰C with media supplemented with 0.5% glucose and arginine to promote protein solubility. Collected bacterial pellets were first re-suspended in 2ml/g Lysozyme buffer (20mM Tris pH 8.0, 100mM NaCl, 1mM EDTA, 0.01% tween, 0.01% Triton, 10mg/mL Lysozyme, 1mM PMSF, 15ug/mL benzamidine, 2ug/mL Aprotinin, 5mh/mL 6-aminocaproic acid) and incubated at room temperature for 30 minutes with rotation. The bacterial suspension was then quickly frozen in liquid nitrogen and allowed to thaw in cold water, with this cycle repeated a total of four times. 3x resuspension buffer (90mM HEPES pH 7.4, 600mM NaCl, 0.01% Tween-20, 0.01% Triton X-100, 15mM β-mercaptoethanol, 30mM imidazole, 150mM arginine, 150mM glutamate) was added to bring the final concentration to 1X, prior to lysing cells by sonicating five times for 15 seconds on ice. The lysate was then cleared by centrifuging at 13,000g for 15 min at 4°C, and the supernatant was transferred to a new tube with the centrifugation repeated again at the same conditions.

Recombinant protein was purified in batch using Ni²⁺ agarose beads and gravity columns [250]. 5mL of 50% slurry of Ni²⁺ agarose beads were washed once with water and then equilibrated with 10 column volumes of 1x resuspension buffer (30mM HEPES pH 7.4, 300mM NaCl, 0.01% Tween-20, 0.01% Triton X-100, 5mM β-mercaptoethanol, 10mM imidazole, 50mM arginine, 50mM Glutamate). The cleared bacterial lysate was incubated with the equilibrated Ni²⁺ agarose beads for 1h at 4°C with rotation. Following binding, the column was

washed with a succession of wash buffers, with 25ml applied to the column and left to drain to just above the bead volume before applying the next buffer. The following table indicates the series in which wash buffers were used, the volume used and the buffer composition

Table 4. Buffers used for IMAC purification of recombinant PSACH		
	Volume	Buffer Composition
Resuspension buffer	1x25ml	30mM HEPES pH 7.4, 300mM NaCl, 0.01% tween-20, 0.01% Triton X-100, 5mM β -mercaptoethanol, 10mM imidazole, 50mM arginine, 50mM Glutamate
Wash buffer 1	4x25ml	30mM HEPES pH 7.4, 100mM NaCl, 0.01% tween-20, 0.01% Triton X-100, 20mM imidazole
Wash buffer 2	2x25ml	30mM HEPES pH 7.4, 100mM NaCl, 0.01% tween-20, 0.01% Triton X-100, 100mM urea
Wash buffer 3	2x25ml	30mM HEPES pH 7.4, 500mM NaCl, 0.01% tween-20, 0.01% Triton X-100, 100mM urea
Wash buffer 4	1x25ml	30mM HEPES pH 7.4, 500mM NaCl, 0.01% tween-20, 0.01% Triton X-100
Wash buffer 5	1x25ml	30mM HEPES pH 7.4, 500mM NaCl, 0.01% tween-20, 0.01% Triton X-100, 20mM imidazole
Wash buffer 6	1x25ml	30mM HEPES pH 7.4, 500mM NaCl, 0.01% tween-20, 0.01% Triton X-100, 5% EtOH

Recombinant protein was eluted from the column by applying elution buffer (30mM HEPES pH 7.4, 100mM NaCl, 0.01% Tween-20, 0.01% Triton X-100, 500mM imidazole) and collected as 4x5ml fractions. The two fractions containing the most protein, as tested by Bradford reagent, were pooled and placed in a 50K MWCO filter and spun at 4,000 RPM in a tabletop centrifuge until the volume was reduced by 75%. The flow-through of the spin filter was discarded, dialysis buffer (30mM HEPES pH 7.4, 100mM NaCl, 0.01% Tween-20, 0.01% Triton X-100) was added to fill the upper chamber of the spin filter, and the sample was spun again with this process repeated a total of four times. Following the final concentration, the concentrated protein sample was collected and glycerol added to a final concentration of 5% and samples stored as 100uL aliquots at -80°C.

NTA Coverslips

Nickle NTA coverslips were made by Lena Moeller in the Rickard Blunk Lab at l'Université de Montreal as previously described [240]. Briefly, 2cm circular glass coverslips were first washed with concentrated KOH to remove organic material and then overnight in Piranha solution (concentrated sulfuric acid and hydrogen peroxide) overnight at 70°C. GPTMS, a silicon-containing molecule, is used to react with exposed hydroxyl groups on the glass surface and polymerizes through hydrolysis and condensation reaction to leave an exposed epoxide reactive group. 10mM AB-NTA is then added to the coverslips to react with epoxide groups for at least three hours in a humidor at room temperature, followed by a glycine solution which is added to quench unreacted epoxide groups. Following a series of washes, a nickel chloride solution is added to generate coverslips containing functionalized bound nickel capable of binding polyhistidine-tagged proteins. Final functionalized coverslips were glued to 10cm Petri dishes, into which a 2cm hole was drilled, by using chloroform to melt the plastic and provide a waterproof seal. The waterproof seal was only stable for 36h, as chloroform would continue to degrade the petri dish; therefore new coverslips were made on a weekly basis fresh for each experiment.

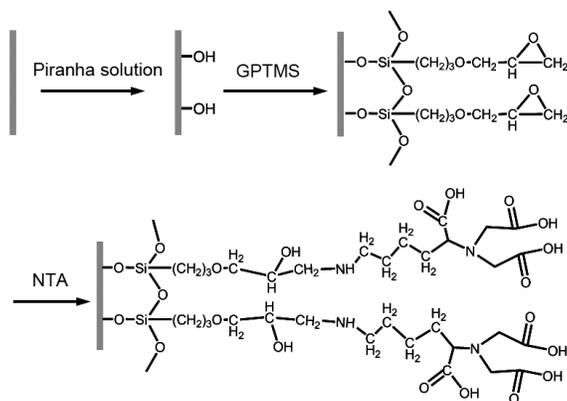


Figure 6. The synthetic route for generating Ni-NTA coverslips. [240]

NTA PEG coverslips

The protocol for generating PEGylated nickel NTA coverslips was performed in the manner identical to that for generating normal NTA coverslips, up to the addition of GPTM and the subsequent washes. Amine functionalized PEG of differing lengths, either a molecule weight of 2000 g/mol or 550 g/mol, were ordered and mixed in a 5:1 ratio with AB-NTA (50mM PEG-NH₂:10mM AB-NTA) and left to react with GPTM functionalized coverslip for at least three hours in a humidior at room temperature. The supernatant was subsequently removed and coverslips were washed according to the protocol for generating normal nickel NTA coverslips.

IgG Coverslips

The protocol for generating IgG coverslips was performed in the manner identical to that for generating NTA coverslips, up to the addition of GPTM and the subsequent washes. Instead of then using AB-NTA, 500uL of IgG conjugation buffer (2.47mg/ml rabbit IgG, 1M NH₄SO₄, 0.1M NaPO₄ pH 7.4) was applied to GPTM reacted coverslips and left to incubate overnight (16h) at room temperature in a sealed petri dish under a biological safety hood. The following day, the solution was removed and coverslips were washed by applying 500-1000ul of the following buffers to the coverslips and leaving them to incubate for 30sec before removing the solution and applying the next buffer. Coverslips were first washed with 100mM glycine pH 2.5, followed by 10mM Tris pH 8.8, then 100mM Triethylamine, twice with 1x PBS and once with 1x PBS left to incubate for 5 minutes, and then twice with 1x PBS with 0.01% TritonX-100 for 5 minutes, once again with 1x PBS quickly and finally stored with an overlayer of 1x PBS.

Gel Filtration of mRNP complexes

mRNP complexes were purified by ssAP as described above and eluted under native conditions with PEGyIOX in a volume of 30uL. Superdex 200 10/300 GL columns were equilibrated in PEGyIOX buffer installed on an FPLC and operated according to manufacturer's instructions at 4°C with UV output spectrophotometer measuring wavelengths of 260, 280 and 488nm. A volume of 120ul of Dextran Blue standard (Sigma) was injected and used for mark the void volume of the column. Subsequently, 100uL of 3mg/ml working stock of Aldolase standard (Sigma) was injected on the column. 30uL of mRNP purified sample was injected and run over the column. The results from the UV-VIS spectrometer from the sample run was combined and graphed.

Chapter 1: Specific mRNP subunit counting

As the fundamental individual components necessary for the proposed single-molecule pulldown and subunit counting protocol have been previously described, it was believed that optimizing their combined use would be a trivial matter. Therefore, the much more detailed and precise investigation of transcript specific RBP stoichiometry was undertaken first.

Purification of transcript specific mRNPs

The development and optimization of the proposed methodology first required selection of a candidate target mRNP. To that effect, the gene PMA1 was selected as it is a long gene of relatively high expression which has previously served as a reporter in genetic, biochemical and microscopy experiments. The essential Pma1 protein functions as a plasma membrane ATPase proton pump, while the mRNA transcript is expressed at an estimated 30 copies per cell and measures approximately 3,000nt. The 3' UTR of the PMA1 gene was tagged with two PP7 hairpin structures, with the KanMX selection marker flanked by LoxP sites subsequently removed by expression of Cre recombinase to restore the 3' UTR with minimal perturbations, save for a residual LoxP nucleotide sequence. Strains were also transformed with stable CEN plasmids expressing PP7CP tagged with SNAPf and PrA (PP7-SNAP-PrA) underexpression of the ADE3 promoter.

ssAP was performed was performed using IgG conjugated dynabeads against PP7-SNAP-PrA using cryo-lysate from strains in which the 3' UTR of PMA1 was also tagged. In order to determine the optimum salt concentration needed for the purification of this particular bait, the purification was performed with a varying concentration of NaCl, either 0, 50, 100 or 300mM NaCl added to the TBT buffer. Prior to elution, samples were split with a fraction eluted under denaturing conditions and loading on gel for silver staining, while the other fraction was used to

investigate the specificity of the purification for PMA1 mRNA transcripts through RT-PCR.

Western blots, the silver stained gel and the results of RT-PCR are shown in the following figure.

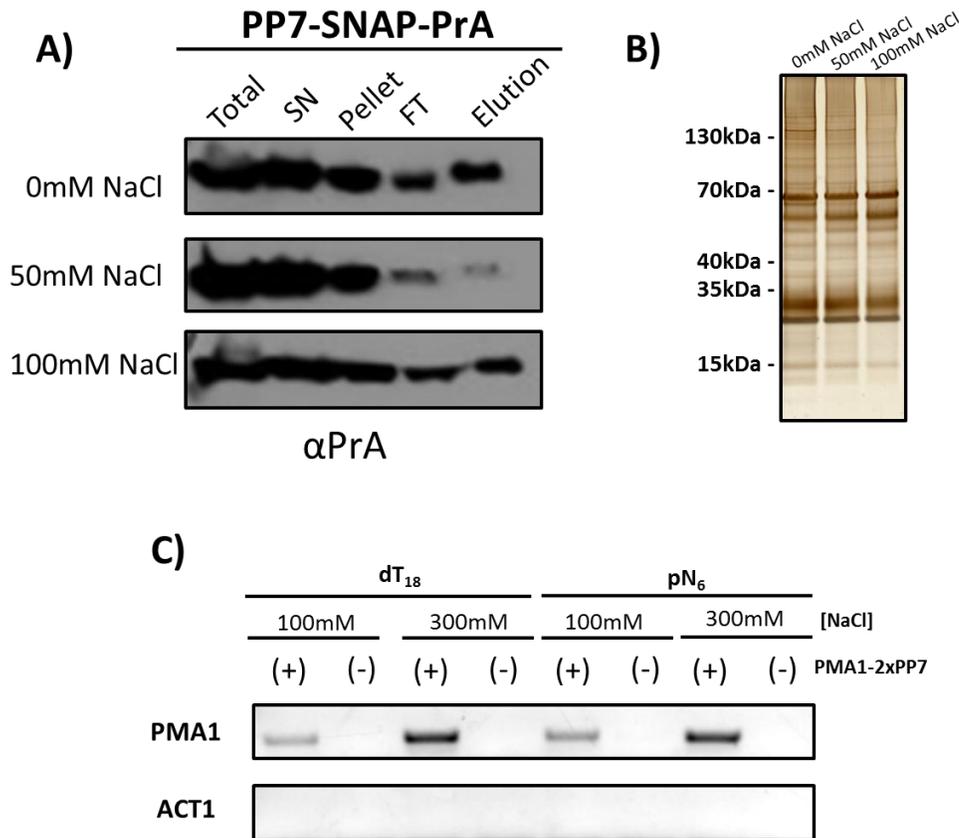


Figure 7. Purification of PMA1 specific mRNPs. Yeast strains expressing PMA1 tagged with PP7 stem-loop repeats and PP7-SNAP-PrA allow for the purification of gene-specific transcripts. A) Shows the effective purification of PP7-SNAP-PrA using IgG dynabeads under different salt concentrations and probing western blots to reveal PrA tagged proteins. B) Shows a silver stained gel of PP7-SNAP-PrA purified eluate under different salt concentrations demonstrated the purification of intact mRNPs. C) Demonstrates the result of RT-PCR using PMA1 or ACT1 specific primers on cDNA generated following ssAP for PP7-SNAP-PrA. Purification was performed in yeast strains where PMA1 was tagged with 2xPP7 stem-loops or no tag at all. The purification was performed in TBT100 or TBT300 buffer with 100mM or 300mM NaCl respectively. RT reactions were performed on isolated RNA using either dT₁₈ or pN₆ primers.

Western blots confirmed effective purification of target PP7-SNAP-PrA. Analysis of purified mRNA by RT-PCR confirmed the specific presence of PMA1 mRNA in purified samples, as compared to negative controls and ACT1 mRNA as a non-specific reporter. These results, coupled with multiple bands observed from silver staining of elution fractions suggest purification of intact PMA1 mRNA with a bound complement of proteins, suggesting the isolation of a gene-specific mRNP. Limited differences were observed with the different concentrations of NaCl tested. As such, 100mM NaCl TBT buffer was selected for use during microscopy experiments, as this was considered to be a mild salt concentration capable of striking a balance between sample purity, maintaining protein interactions and limiting possible interference during labelling reactions. However, 100mM and 300mM NaCl buffer salt concentrations were both repeatedly tested during phases of optimization.

Native elution of PrA bound complexes

While eluting PrA bound complexes from IgG conjugated dynabeads under denaturing conditions is quick, cheap and convenient for simple verifications, our method requires the capacity to elute these complexes in their native state. In order to validate the capacity to natively elute PP7-SNAP-PrA bound complexes from IgG dynabeads with PEGylOx, complexes were first bound to beads in TBT100 or TBT300 buffer. The protocol was performed by eluting complexes using 2.2mM PEGylOx and a 15 minute incubation at room temperature, with taken for western blot to validate the elution. Figure 8 shows an elution in the 70-80% range, in line with what is to be expected according to the authors who developed the synthetic PEGylOx peptide [225]. As such, it is shown that it is possible to elute PP7-SNAP-PrA bound complexes from IgG dynabeads natively.

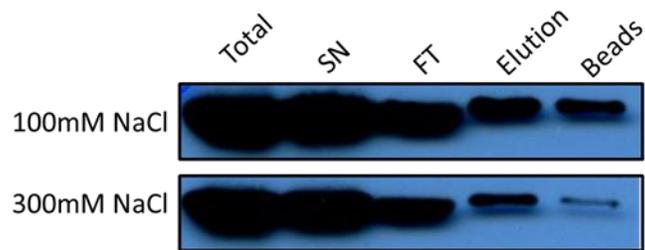


Figure 8. Native elution of PrA bound complexes with PEGyIOX. PP7-SNAP-PrA was eluted with 2.3mM PEGyIOX. Western blotting against using PAP antibody reveals effective elution in TBT100 or TBT300 buffer. Elution efficiency coincides with reported claims.



Figure 9. Functional labelling of SNAP and CLIP tagged complexes. Triple tagged yeast strain Tif4631-CLIP-His₆, PMA1-2xPP7, PP7-SNAP-PrA was labelled with SNAP-Alexa488 and CLIP-Surface-647 dyes on IgG or Ni²⁺ beads. Scanning samples ran on gel with Typhoon scanner at the wavelengths appropriate for each dye shows a band at the appropriate size for each protein.

Labelling of SNAP and CLIP tagged constructs

The nuclear cap binding protein, Cbp80, was selected as the initial RBP for which the stoichiometry would be determined by the proposed single-molecule pulldown and quantification method. This protein was selected as it is widely believed to be recruited to all nuclear mRNA transcripts and can be found at a known quantity of one copy per transcript. However, only a fraction of a mRNA's lifetime is spent in the nucleus of the cell, with a majority in fact spent in the cytoplasm [46, 219]. As such, the number of cellular transcripts associated with the nuclear cap-binding protein would be limited and subsequent experimentation might fail due to difficulties with the limits of detection. Therefore, in addition to generating strains where Cbp80 was tagged with CLIP and His₆ tag, strains were also generated where Tif4631, a component of the cytoplasmic cap-binding complex, was also tagged instead [251]. While the main focus of this project was to study the structure of nuclear mRNP complexes, performing optimization with a higher abundant target was believed to be a useful redundancy.

CBP80 or TIF4631 genes were tagged with tandem CLIP and His₆ tags (Cbp80-CLIP-His₆ and Tif4631-CLIP-His₆) in strains with PMA1 tagged with PP7 stem-loops and expressing PP7-SNAP-PrA (Cbp80-CLIP-His₆, PMA1-2xPP7, PP7-SNAP-PrA and Tif4631-CLIP-His₆, PMA1-2xPP7, PP7-SNAP-PrA). The functionality of these tags for effective labelling was investigated by performing on bead labelling reactions. Complexes were first purified by ssAP from cryo-lysate, however, following initial washing steps, substrate dyes were added to suspended beads, with bound complexes for 30 minutes at room temperature. Unincorporated dyes were removed by washing beads prior to eventually eluting complexes under denaturing conditions and loading samples on gel.

Analysis of scanned gels revealed fluorescent bands corresponding the expected size of target proteins. The observance of these bands indicates that the CLIP-tag is functional when fused to yeast proteins and labelling of these tags while complexes are on beads can be achieved.

Optimization of slide wash buffer

The eluate from the IgG purification of dual tagged strain Tif4631-CLIP-His₆, PP7-SNAP-PrA was used as a negative control to test for non-specific binding of complexes to coverslips. This strain expressed no PP7 stem-loop tagged RNA, and as such, PP7-SNAP-PrA would not be expected to bind to coverslips as it contains no His₆ and cannot be recruited to mRNPs with a tag inserted into the mRNA transcript. PP7-SNAP-PrA was labelled on beads with the addition of SNAP substrate dye, eluted under native conditions and placed on Ni-NTA coverslips. Coverslips were washed with buffers of increasing salt or imidazole concentration, with the number of fluorescent spots from PP7-SNAP-PrA measured in three fields with the PIF program and averaged after each wash (Figure 10, Table 5).

The coverslips used for this experiment suffered from contaminating background fluorescence, as fluorescent spots can be observed in the image labelled “BLANK” prior to the addition of sample protein. Nonetheless, these coverslips were used and appropriate wash buffer conditions were selected by determining the condition in which the number of observed fluorescent spots were reduced to below background levels, that is the average number of fluorescent spots was less than that observed on blank coverslips prior to addition of protein sample.

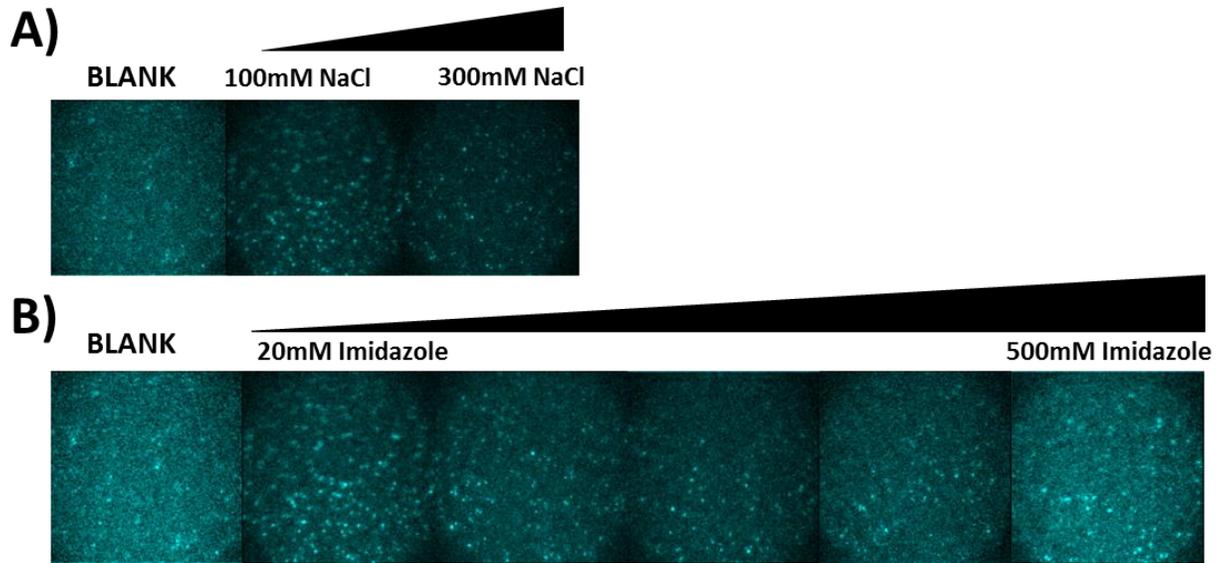


Figure 10. Non-specific binding of PP7-SNP-PrA to coverslips in buffers of different stringency. PP7-SNP-PrA protein purified by IgG conjugated dynabeads was placed on Ni-NTA coverslips and washed with buffers of increasing stringency. Yeast strain Tif4631-CLIP-His₆, PP7-SNP-PrA expressed no PP7 tagged mRNA and thus no PP7 was expected to bind to coverslips. BLANK represents image taken of coverslips before addition of antibodies. Coverslips were washed with PEGlyOX buffer with additional NaCl (A) or added imidazole (B). See Table 5 for average quantification of the number of spots remaining after each washing.

Wash Condition	Number of fluorescent spots
Blank	72
100mM NaCl, 20mM Imidazole	89
300mM NaCl, 20mM Imidazole	82.5
100mM NaCl, 50mM Imidazole	128
100mM NaCl, 75mM Imidazole	60.5
100mM NaCl, 100mM Imidazole	42
100mM NaCl, 500mM Imidazole	37

To that effect, PEGylOX buffer supplemented with 100mM NaCl and 100mM imidazole was selected as the optimal coverslip wash buffer. While fewer spots, 61, were first initially observed compared to the 72 spots observed prior to the addition of any proteins with 100mM NaCl and 75mM imidazole, the number of background PP7-SNAP-PrA spots binding was even further reduced to 42 with a wash of 100mM NaCl and 100mM Imidazole. Even with a very stringent wash with 100mM NaCl and 500mM imidazole, the number of spots was reduced only slightly to 37; therefore the 100mM NaCl and 100mM imidazole wash was deemed of sufficient stringency.

Imaging of cap binding complex on PMA1 transcripts

PMA1 tagged transcripts were purified in strains where PP7-SNAP-PrA was expressed, alongside either Cbp80-CLIP-His₆ or Tif4631-CLIP-His₆. Fluorescent tags were labelled on beads during purification and complexes were natively eluted and subsequently placed on Ni-NTA functionalized coverslips. Following the washing of coverslips with PEGylOX buffer supplemented with 100mM NaCl and 100mM imidazole, dyes were imaged at the appropriate wavelengths, with several fields of view taken (Figure 11). The results highlight several problems with this experiments. Firstly, it is observed that many more spots, corresponding to PP7-SNAP-PrA, are observed in cyan compared to red which reflects the cap-binding protein responsible for tethering the complex to the coverslip. Secondly, it is observed that of the few red spots observed, few co-localize with spots in cyan. Lastly, analysis of photobleaching traces from multiple fields for PP7-SNAP-PrA yielded a histogram showing a prevalence for monomers rather than the expected tetramers. As similar results were observed for both Cbp80 and Tif4631 tagged strains, and the observed problems could have arisen as a result of one of the multiple issues, it was decided to generate a positive control construct to eliminate several of these possibilities.

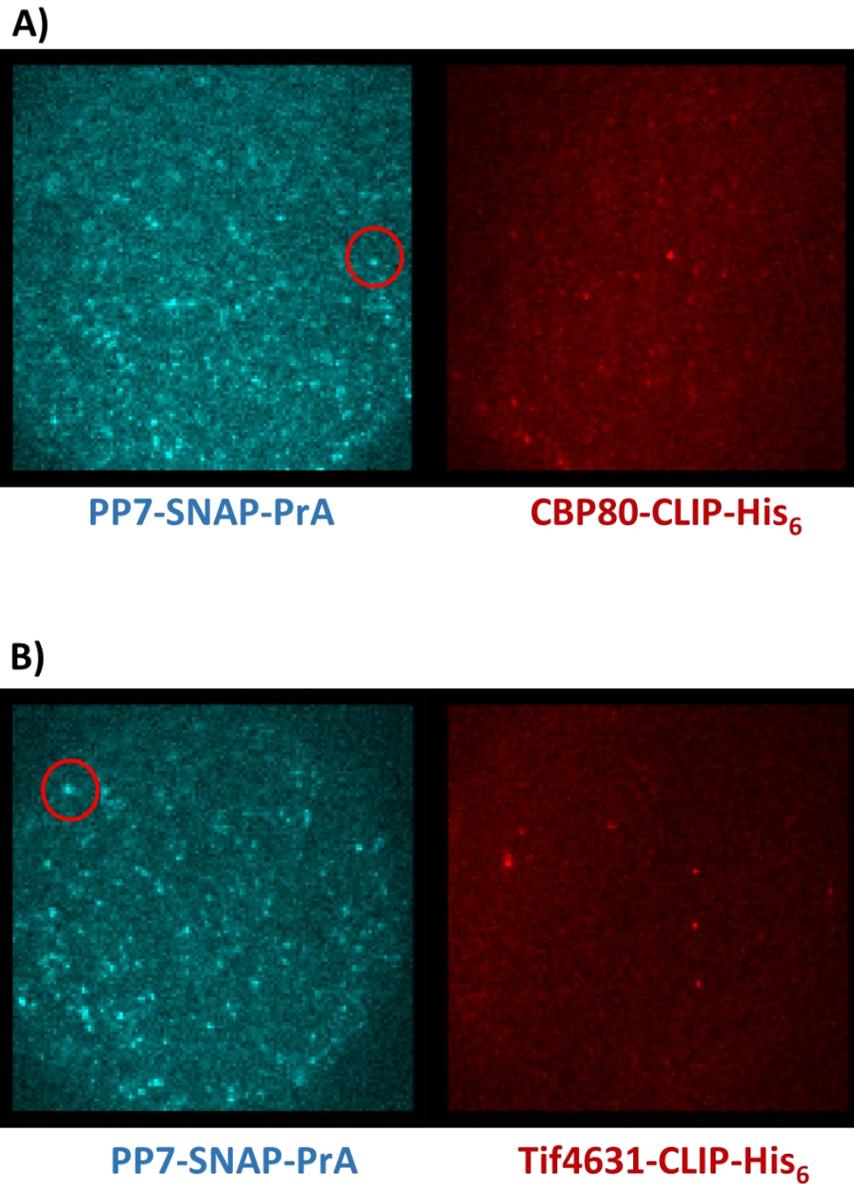


Figure 11. Imaging of CBC on single purified PMA1 mRNPs by TIRF. Fluorescent microscopy image of purified PMA1 mRNPs with PP7 labelled with SNAP substrate dye and Cbp80-CLIP-His6 (A) or Tif4631-CLIP-His₆ (B) labelled with CLIP substrate dye. The labelled eluate from IgG dynabead purification against PP7-SNAP-PrA in yeast strains where the PMA1 gene was tagged with 2xPP7 stem-loops was placed on Ni-NTA coverslips and imaged. The same field is shown in each channel. Red circle illustrates single mRNP complex. 1pixel = 0.4 μ m, Field = 51.2 μ m x 51.2 μ m.

Imaging of rPSACH labelled constructs

Devised as a positive control for co-localization, a fusion construct was devised which integrated all of the tags used for the method into one protein. The construct, PP7-SNAP-PrA-CLIP-His₁₀ (PSACH), was first expressed ectopically from a plasmid in yeast cells under the regulation of the constitutive ADE3 promoter. The PSACH construct was purified from cryo-lysate using the PrA handle and labelled on beads for 30 minutes, as was done for the normal mRNP purification, prior to natively eluting the complex and placing it on Ni-NTA coverslips for imaging (Figure 12). Limited co-localization was observed, as well as a discrepancy of more signal from the SNAP dye compared to the CLIP dye. Analysis of samples taken during the course of the purification by western blot with PAP, to reveal PrA tagged targets, revealed the presence of a prominent lower molecule weight band which is indicative of a degradation product. The presence of this abundant degradation product was believed to account for the discrepancy of abundance in the two channels, as well as the poor co-localization, and therefore did not function well as a positive control.

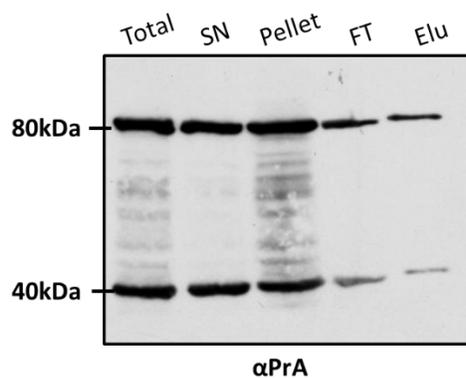
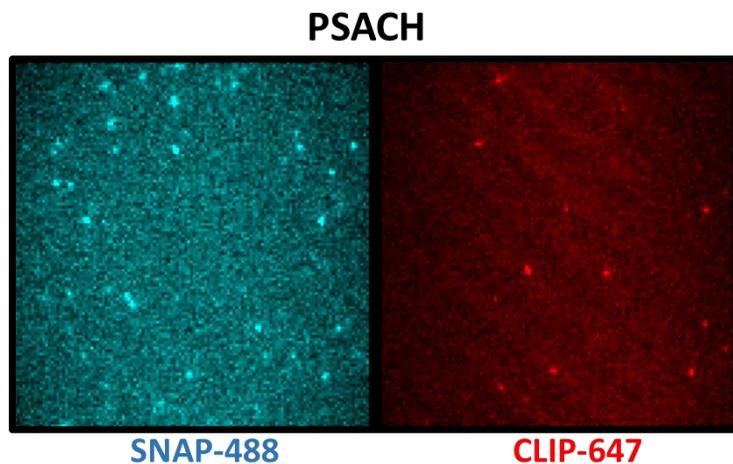


Figure 12. Single-molecule imaging of PSACH construct purified from yeast. The PSACH construct was expressed in yeast and purified from cryo-lysate on IgG conjugated dynabeads. Samples were labelled with SNAP and CLIP substrate dyes and eluted under native conditions and placed on Ni-NTA coverslips. The same field is shown in two different channels for the SNAP fluorophore and the CLIP fluorophore. Below is a western blot of samples taken during the course of the PSACH purification blotted using the PAP antibody to react with the PrA tag.

As such, the PSACH construct was instead expressed in bacterial cells and purified by the polyhistidine tag through recombinant protein technology. Following optimization of rPSACH expression and purification, a sample of significantly greater purity, and containing only minor degradation products, was obtained. The functionality of the construct was initially validated by observing specific labelling of the construct with the SNAP specific dye. Subsequently, 50 pmol of the construct was labelled for 30min at room temperature with both dyes in solution, instead of performing labelling on beads. The complexes were then bound to Ni²⁺ magnetic beads to wash away unincorporated dye, before eluting complexes and placing them in dilution buffer to bind to Ni-NTA coverslips and image by microscopy (Figure 13). While the number of spots observed in both channels was more comparable than previously seen with previously purified rPSACH constructs or mRNPs, a 2-3 fold excess from the SNAP dye was still observed.

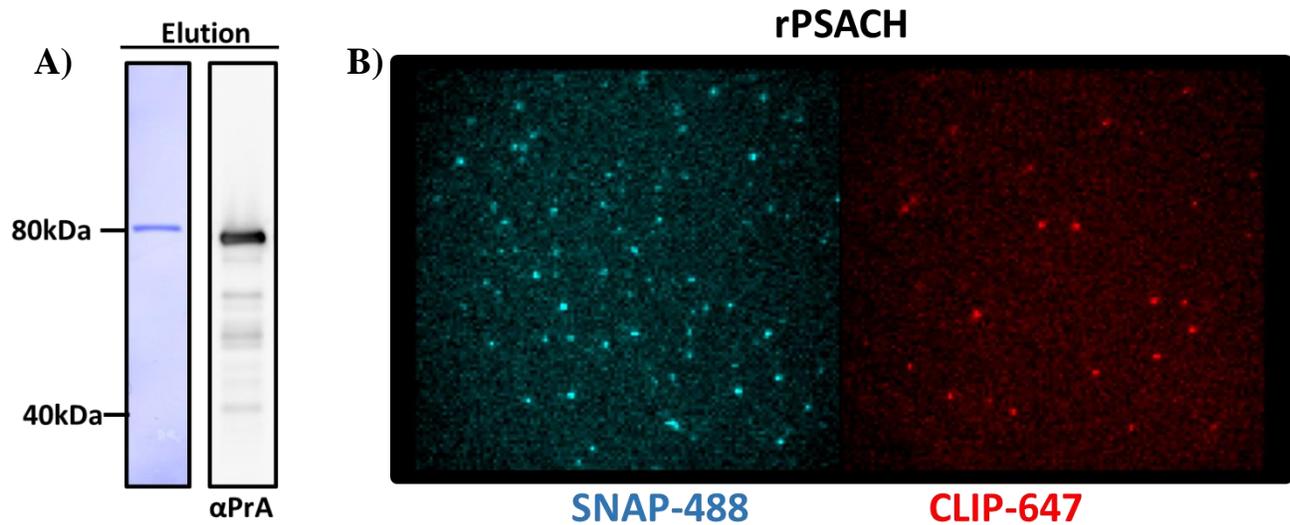


Figure 13. Single molecule imaging of labelled rPSACH. A) Shows coomassie stained gel (Left) and western blot (Right) of eluate of rPSACH expressed in bacterial cells and purified by Ni^{2+} agarose beads. B) shows 50pmol of rPSACH labelled with SNAP and CLIP substrate dyes for 30min on beads as viewed under TIRF microscopy. The same field is shown in either cyan or red channel.

Relative labelling efficiencies of SNAP and CLIP tags

While 30min is often reported as the time required for labelling reactions with SNAP and CLIP dyes in publications and the product manufacturer's, it is possible that certain tagged proteins and the use of certain SNAP or CLIP dyes exhibit variable kinetics of labelling and require additional reaction times. Furthermore, most labelling reactions are reported with complexes in solution rather than bound to an affinity purification matrix, and as such, the labelling on-bead method may be a source of interference for the reaction. A relative time course assay was performed using the recombinant PSACH construct, where the construct was labelled independently with either SNAP or CLIP specific dyes, and the labelling reaction was either performed in solution or with complexes bound to IgG conjugated dynabeads. While little difference was observed between performing the labelling reaction in solution or while the construct was bound to beads, it was noted that greater levels of CLIP-tag labelling were observed after 90min had elapsed, compared to the SNAP-tag which showed no increase in labelling after 15 minutes (Figured 14).

50 pmol of the PSACH construct was labelled with both dyes using newly determined optimized conditions of 90min and observed under the microscope on Ni-NTA coverslips. A comparable number of spots could be observed in both channels, with a slight excess from SNAP labelled spots, however most importantly a number of these spots were observed to co-localize (Figure 15).

The relative time course labelling assay was repeated using triple tagged strain, Cbp80-CLIP-His₁₀, PMA1-2xPP7, PP7-SNAP-PrA, where SNAP and CLIP tagged proteins were labelled independently either in solution or while bound to IgG dynabeads. Similar results were obtained for in solution labelling with Cbp80-CLIP-His₁₀ and PP7-SNAP-PrA as with PSACH,

with maximal SNAP labelling achieved within 15 to 30 minutes, whereas CLIP labelling required 90 minutes to achieve maximal levels. While the same results were observed for SNAP labelling when complexes were bound to IgG conjugated dynabeads, no conclusions could be drawn for CLIP labelling, as no protein could be detected, similar to what was observed in previous purifications with either cap-binding protein. With these results, it was determined that the time required for labelling reactions would be extended to 90 minutes, rather than the 15-30min used previously (Figure 16).

However, with doubts raised due to the lack of detectable signal observed from 5' cap-binding proteins when purifying PMA1 transcripts by tagged PP7, originally believed to be due to low abundance, it was decided to move away from looking at protein stoichiometry on specific mRNP transcripts and proceed to investigate the more general trend of heterogeneity between mRNPs.

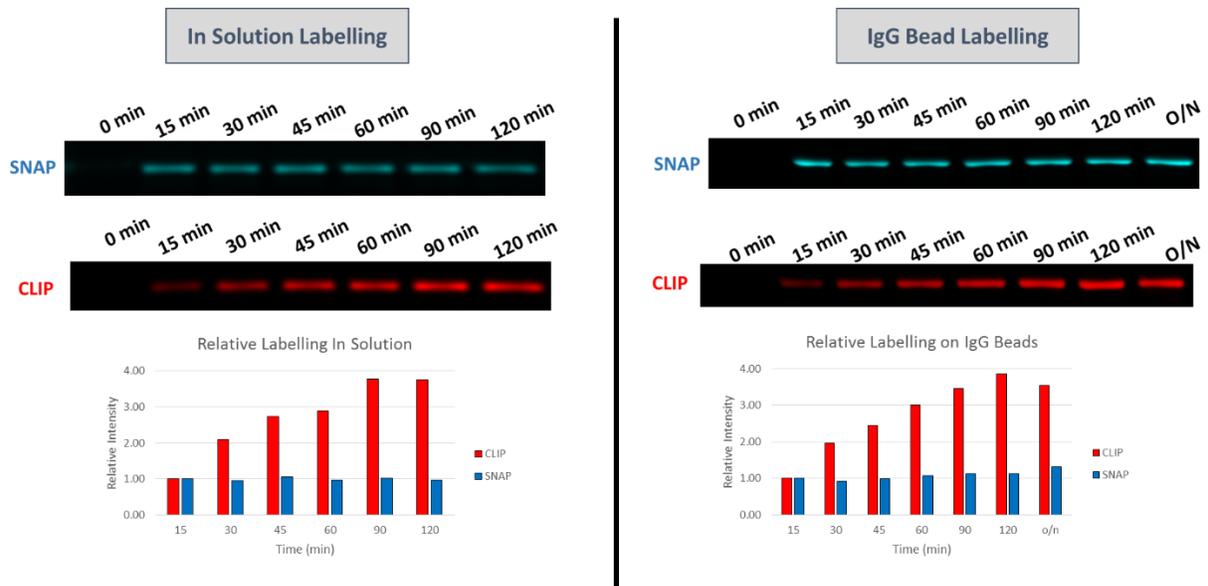


Figure 14. Relative time course labelling assay of rPSACH in solution or on IgG conjugated dynabeads. rPSACH was labelled with either SNAP or CLIP substrate dye in solution or by first binding complexes to IgG conjugated dynabeads. Samples were taken at indicated time points, denatured and subsequently ran on gel and imaged with Typhoon fluorescent scanner. The intensity of each band was determined using ImageJ software with intensities at each time point reported relative to the intensity measured at 15min.

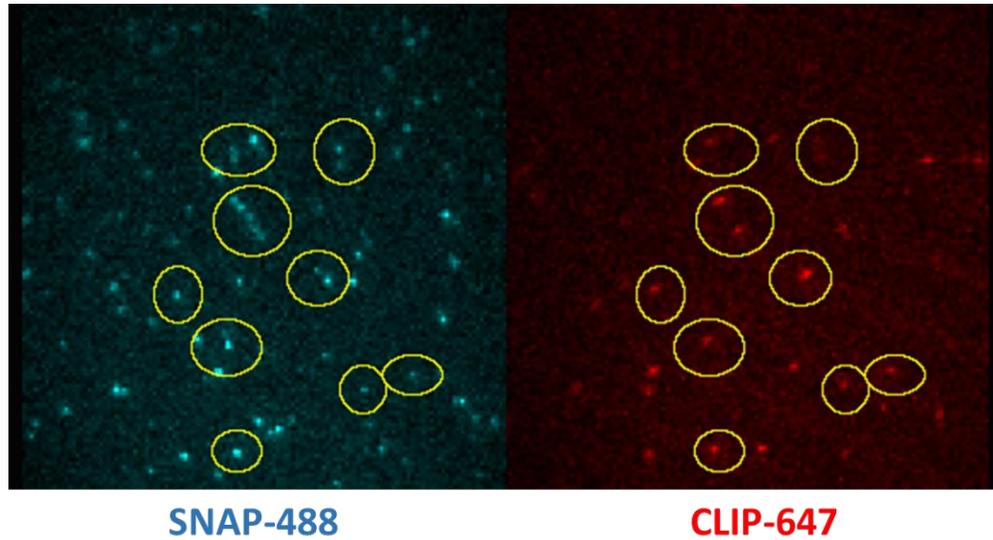


Figure 15. Single molecule imaging of rPSACH labelled for 90 minutes. 50pmol of rPSACH was labelled in solution for 90min, bound to Ni^{2+} agarose beads to removed unincorporated dye, natively eluted and bound to Ni-NTA coverslips. Yellow circles show the same region in both channels.

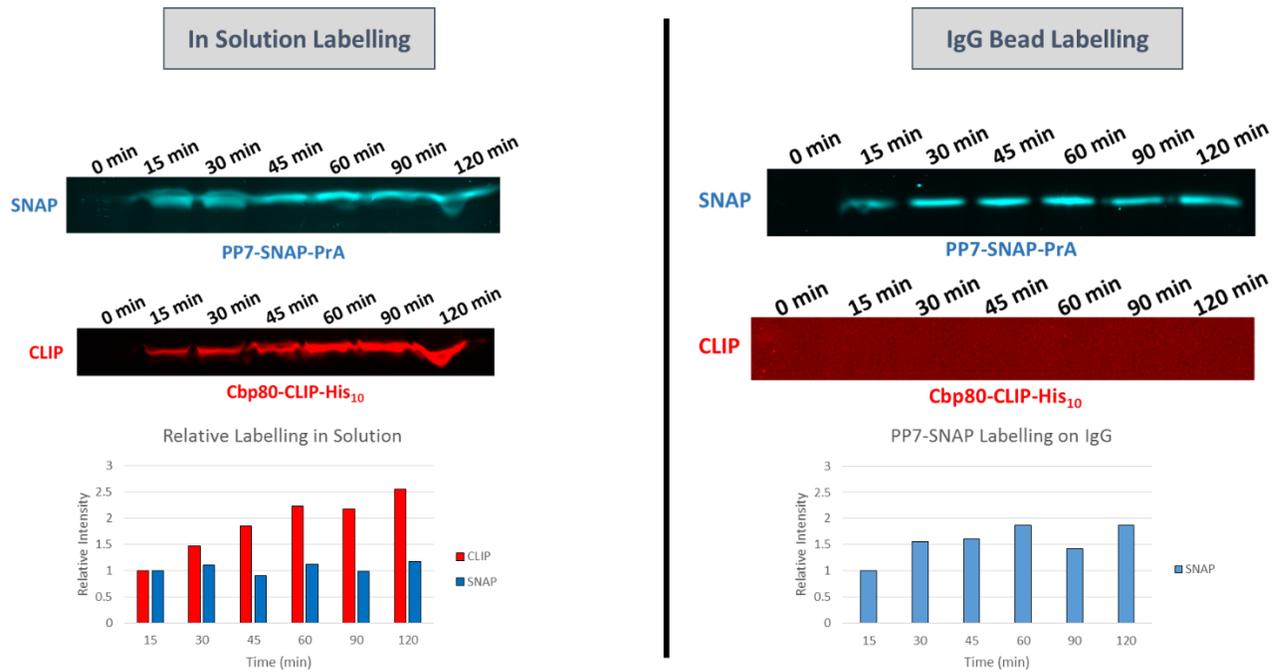


Figure 16. Relative time course labelling assay of Cbp80-CLIP-His10 and PP7-SNAP-PrA in solution or on IgG conjugated dynabeads. Cryo-lysate from yeast strain Cbp80-CLIP-His10, PMA1-2xPP7, PP7-SNAP-PrA was labelled in solution or bound to IgG conjugated dynabeads with SNAP or CLIP substrate dyes. Samples were taken at indicated time points, denatured and subsequently ran on gel and imaged with Typhoon fluorescent scanner. The intensity of each band was determined using ImageJ software with intensities at each time point reported relative to the intensity measured at 15min.

Chapter 2: General mRNP subunit counting

Prior to conducting further microscopy experiments, there was a realization that there was a need to further quantify and optimize certain elements pertaining to the experimentation procedure. Namely, there was a need to quantify the amount of bait protein purified to ensure that SNAP and CLIP dyes were added in great excess, as well as quantify the labelling efficiency in absolute terms rather than relative levels. Furthermore, it was decided to substitute the use of tagged Tif4631 as a control with Cdc33 tagged strains. While Tif4631 forms the scaffold for cytoplasmic cap-binding complex, it does not directly bind with the 5' cap, as does Cdc33. Additionally, yeast express a paralogue of Tif4631, Tif4632, which also associates with cellular mRNPs and may, therefore, exclude recruitment of Tif4631 and may cause the protein not to be found in all nuclear mRNPs.

Quantification of isolated bait protein

While Bradford assays are typically performed to quantify the amount of protein in a sample, this classic method would not be directly suitable, as the interested lied in quantify the abundance of two particular proteins in the sample, rather than all proteins present. To that effect, a more targeted protein quantification assay was devised, whereby all tagged protein in the sample would be purified by affinity purification from cellular cryo-lysate under moderately stringent conditions, and then separated on SDS-PAGE with a known amount of BSA standards for comparison.

For this quantification, it was desired that all of the tagged protein in the lysate be captured, in order to ensure maximal material for further microscopy experiments. To that effect, a bead saturation binding assay was performed where a fraction of re-suspended yeast cryo-lysate was incubated with an increasing amount of IgG dynabeads in order to determine the amount needed to capture all tagged protein in solution. This experiment was performed to

quantify the abundance of several PrA tagged proteins, namely Cbp80, Cdc33 and PP7-SNAP-PrA. The bead saturation binding assay revealed the amount of IgG conjugated dynabeads necessary to isolate the target bait protein by identifying the condition under which bait protein could no longer be detected in flow-through of the binding reaction.

Using the determined optimum quantity of beads, the purification was repeated with a more stringent buffer variant of TBT buffer in order to maximize the purification of only the target protein with few associated protein. A fraction of eluted samples were run on SDS-PAGE gels, along with varying amounts of BSA standards. Following separation, the proteins in the gel were stained with Coomassie blue, and bands corresponding to the size of the purified protein were quantified along with the BSA standards (Figure 17). The quantity of the specific protein was determined by generating a standard curve using the band intensities of the BSA standards, generating a normalized quantity of each bait protein for every gram of cryo-lysate powder measured.

Analysis validated that SNAP and CLIP dyes had indeed been added in suitable excess over the SNAP and CLIP tagged proteins, normally at least 3-4 times in excess. This analysis eliminated the possibility that insufficient dye substrate was added to labelling reactions with mRNP components, which could have given rise to the poor co-localization of labelled proteins between the two channels.

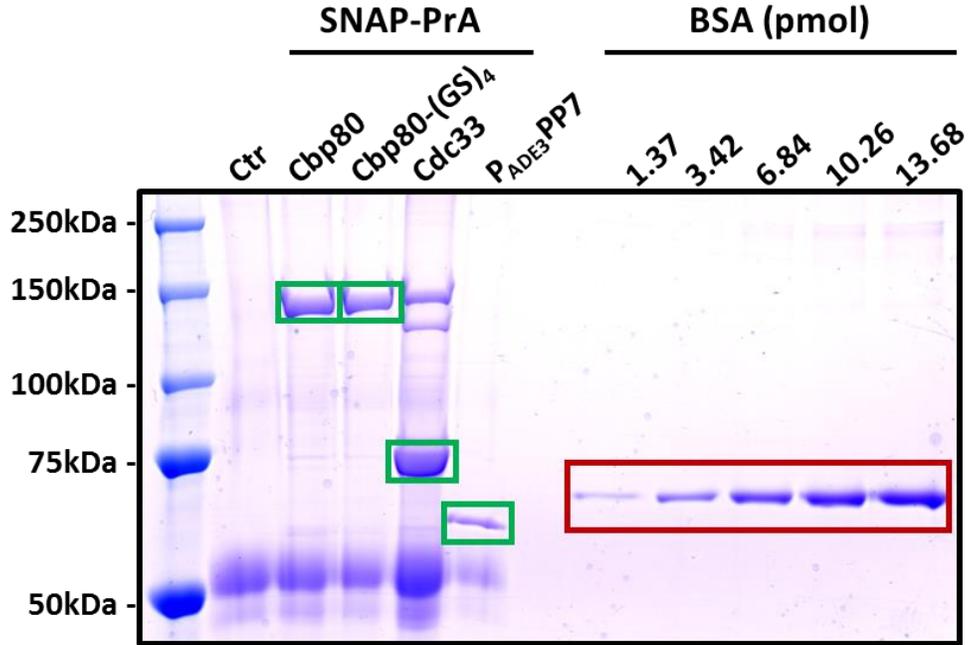


Table 6. Quantification of protein specific bands from targeted Bradford assay

Bait	Band Intensity	Control Intensity	Normalized Intensity	pmol Total	pmol/g
Cbp80	24444	4956	19488	25.8	93.8
Cbp80-(GS) ₄	19049	4956	14093	18.4	86.0
CDC33	39588	3299	36289	48.8	206.9
PP7	8516	1429	7086	8.8	44.6

Figure 17. Quantification of purified protein by targeted Bradford assay. The eluate from Cbp80, Cbp80-(GS)₄, Cdc33 or PP7 purified by ssAP on IgG conjugated dynabeads were run on SDS-PAGE gel along with BSA standards and proteins stained with Coomassie. Bands corresponding to the bait proteins were quantified using ImageJ along with the BSA standards. A standard curve was generated to determine the quantity of bait protein purified and used to determine the pmol of protein present per gram of cryo-lysate as reported in Table 6. Green boxes indicate SNAP-PrA tagged bait proteins and red boxes indicate BSA standards.

Absolution quantification of labelling efficiency by mass spectrometry

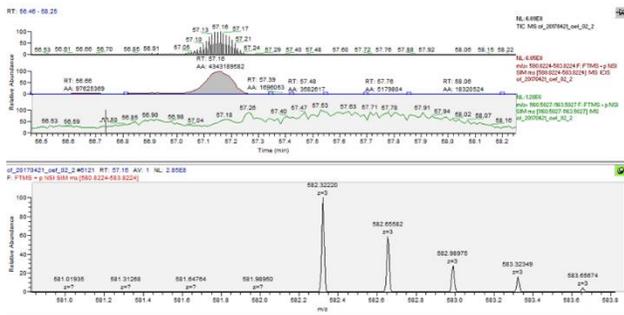
While previously, it was determined that an extended labelling time of 90 minutes resulted in three-fold more labelling of the CLIP-tag compared to 15 minutes, it does not indicate the actual quantity of CLIP tagged protein which is labelled. While maximum labelling is observed after 90 minutes, it is unknown if this translates to 100% of CLIP-tags being labelled, or an upper limit of only 75% of molecules are labelled with the other 25% remaining non-functional and unlabelled for an unknown reason. As such, the absolute labelling efficiency of SNAP and CLIP tagged proteins was investigated through mass spectrometry.

New yeast strains were generated, both for this experiment and for use in the general mRNP quantification experiment, where Cbp80 was tagged with SNAP-PrA and the 3' poly(A) binding protein Nab2 was tagged with CLIP-His₁₀. Nab2 was selected as it an essential mRNP biogenesis and export protein believed to be recruited to all nuclear mRNAs and was believed to be found in conjunction with Cbp80 on nuclear mRNAs. Variations of these strains were also generated with two different linkers were used to separate the SNAP-PrA tags from the tagged protein. The sequence of these two linkers was either GDGAGLIN, found in the pFA6a plasmid series and used in all previous constructs or (GSG)₄, which was a linker used to previous publications to separate the SNAP domain from the tagged protein. Also included, was a comparison between if the labelling were done in solution or on IgG conjugated dynabeads, with the recombinant PSACH construct also included for comparison.

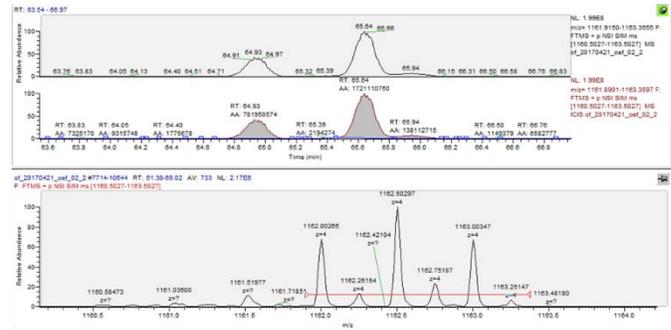
The purification was performed using cryo-lysate from all strains, with the exception of the recombinant PSACH construct, and labelling reactions were either performed in solution

prior to affinity purification or while complexes were bound to beads, and for a time of either 15 minutes or 90 minutes. Samples were not eluted from beads, but rather digested on bead with trypsin and subsequently analyzed by mass spectrometry. The mass spectrometry runs focused on identifying SNAP and CLIP tryptic peptides, with and without the conjugated dye. The quantity of each peptide was assessed in two different ways making use of the reconstituted TIC profile. Peptides were either quantified by the most intense peak intensity or by calculating the area under the peaks corresponding to the particular ions. Of note was the observation that SNAP peptide conjugated to Alexa488 dye eluted as two or three distinct peaks, as observed in the TIC, depending on the sample. In this case, the sum of the area of each peak was calculated and combined for determining abundance (Figure 18, Table 7).

This method succeeded in quantifying the SNAP-tag labelling efficiencies and identified on bead labelling for 90 minutes with a (GSG)₄ linker, as the optimal labelling conditions yielding more than 80%, which is suitable for single-molecule microscopy experiments. However, no data could be obtained for the CLIP-tag, as the Nab2 protein was found not to associate with these purified complexes.



SNAP



SNAP-Alexa488

Figure 18. Absolute quantification of SNAP-tag labelling by mass spectrometry. Yeast strain Cbp80-SNAP-PrA, Nab2-CLIP-His₁₀ was used to assess the labelling efficiency of the SNAP and CLIP tags under various condition. Depicted are screenshots taken from XCalibur analysis and visualization software of mass spectrometry data. Left depicts the analysis of unlabelled SNAP tryptic peptide, while right depicts SNAP peptide labelled with Alexa-488 SNAP substrate dye. The TIC chromatogram for the ion under investigation is shown above while the MS1 spectrum is shown below. SNAP-Alexa488 labelled peptides eluted as multiple peaks in the MS run, and the area of each was calculated and combined.

Table 7. Quantification of Alexa488 labelled and unlabelled SNAP tryptic peptides by Mass Spectrometry

Samples	Intensity			Area Under Curve				% Labeled
	SNAP	SNAP-Dye	% Labeled	SNAP	SNAP-DYE Peak1	SNAP-DYE Peak 2	SNAP-DYE Peak 3	
CBP80-SNAP-PrA, Nab2-CLIP-His10 Bead, 15min, pFA	6.82E+08	5.63E+07	7.63	6.63E+09	6.15E+08	1.06E+09	4.03E+08	23.88
CBP80-SNAP-PrA, Nab2-CLIP-His10 Bead, 15min, GS4	2.98E+08	1.95E+08	39.55	4.00E+09	6.13E+08	9.21E+08	6.29E+08	35.09
CBP80-SNAP-PrA, Nab2-CLIP-His10 Bead, 90min, pFA	3.58E+08	3.07E+08	46.17	3.56E+09	8.59E+08	8.71E+08	1.11E+09	44.42
CBP80-SNAP-PrA, Nab2-CLIP-His10 Bead, 90min, GS4	4.40E+07	2.27E+08	83.76	3.60E+08	6.75E+08	6.83E+08	8.38E+08	85.93
CBP80-SNAP-PrA, Nab2-CLIP-His10 Solution, 15min, pFA	6.36E+08	4.06E+08	38.96	9.83E+09	1.13E+09	1.93E+09	7.63E+08	27.97
CBP80-SNAP-PrA, Nab2-CLIP-His10 Solution, 15min, GS4	2.68E+08	9.38E+07	25.93	5.02E+08	3.27E+08	2.57E+08	0.00E+00	53.77
CBP80-SNAP-PrA, Nab2-CLIP-His10 Solution, 90min, pFA	9.42E+07	1.18E+08	55.61	1.54E+09	6.18E+08	8.76E+08	0.00E+00	49.22
CBP80-SNAP-PrA, Nab2-CLIP-His10 Solution, 90min, GS4	7.93E+07	8.79E+07	52.57	8.87E+08	4.73E+08	4.53E+08	0.00E+00	51.08
rPSACH, Solution, 90min	2.77E+07	1.31E+08	82.55	1.41E+09	6.90E+08	6.40E+08	0.00E+00	48.52

Single molecule imaging of Pab1 on Cbp80 purified complexes

Following re-evaluation of published interactions between Cbp80 and mRNP associated proteins, Pab1 was selected as a suitable candidate RBP target for stoichiometric quantification on global mRNPs. Strains were generated where Cbp80 was tagged with SNAP-PrA along with Pab1 tagged with CLIP-His₁₀. The interaction between these proteins was first validated by generating a cryo-lysate from this strain and performing a purification using IgG conjugated dynabeads. The purification was done in TBT100 buffer with or without RNase A, and complexes were labelled on beads with SNAP and CLIP substrate dyes for 90 minutes before eluting under denaturing conditions. Running samples on SDS-PAGE gel confirmed a co-IP signal from Pab1, which was RNase A sensitive, suggesting that intact mRNPs bound by both proteins were affinity purified.

Subsequently, the experiment was repeated without RNase A treatment and complexes were eluted under native conditions and placed on Ni-NTA coverslips for imaging by TIRF microscopy. However, imaging revealed few spots from Cbp80-SNAP-PrA labelled constructs and essentially none from Pab1-CLIP-His₁₀ constructs which would be expected to anchor the complex to the coverslips. It was observed that much of the fluorescence remained in solution above the coverslip, suggesting that much of the protein in the sample was incapable of binding to the Ni-NTA coverslips.

In order to validate the functionality of the His₁₀ tag added to the Pab1 protein, a bead binding assay was performed where re-suspended cryo-lysate from this strain was added to varying amounts of Ni²⁺ magnetic beads to assess the tagged protein's capacity for binding (Figure 20). The results show that very little Pab1 was capable of binding, and no Cbp80 could be detected from this small-scale purification. The results suggest that the His₁₀ tag on Pab1 is

non-functional or inaccessible, and as such is not suitable for purification or immobilization on Ni-NTA coverslips.

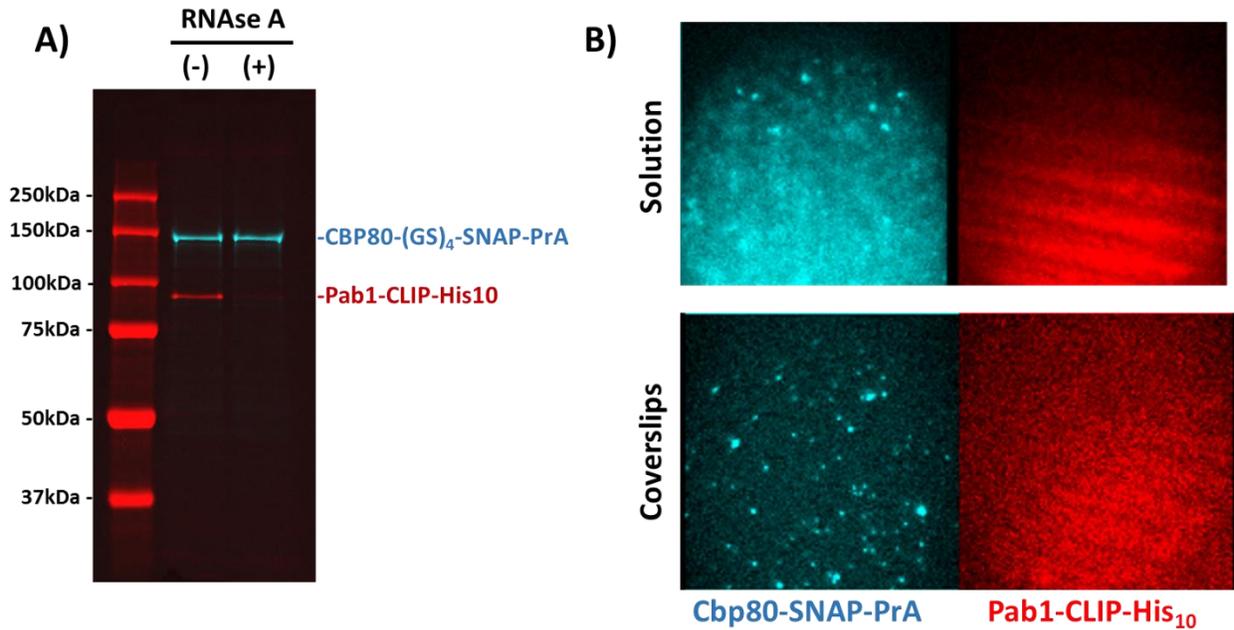


Figure 19. Single-molecule imaging of Cbp80 and Pab1 bound mRNPs on Ni-NTA coverslips. **A)** Shows the result of a co-IP experiment performed using yeast strain Cbp80-SNAP-PrA, Pab1-CLIP-His₁₀, where mRNPs were purified using IgG conjugated dynabeads, labelled using SNAP and CLIP substrate specific dyes and eluted under denaturing conditions. Purifications were performed with or without the addition of RNase A to degrade RNA in the samples. The gel was imaged using Typhoon scanner at dye specific wavelengths. **B)** Shows the result of imaging mRNPs by IgG conjugated dynabeads from yeast strain Cbp80-SNAP-PrA, Pab1-CLIP-His₁₀ where complexes were labelled with SNAP and CLIP-tag specific dyes and eluted under native conditions and placed on Ni-NTA coverslips. The same field of view is shown two different imaging channels. The solution shows images taken above the plane of the coverslips to demonstrate fluorescent proteins remaining in solution and failing to bind to coverslips. Coverslips indicates images taken on the planar surface of the coverslips where no fluorescence can be detected from Pab1-CLIP-His₁₀

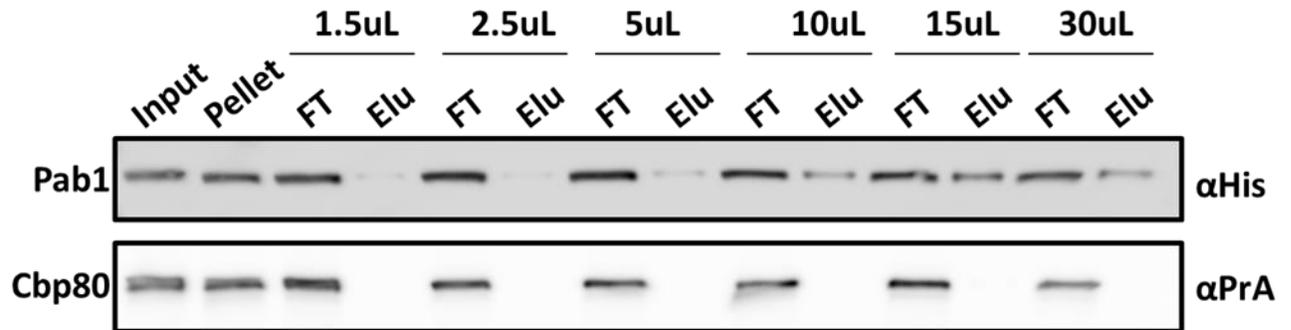


Figure 20. Pab1-CLIP-His10 binding assay to Ni²⁺ beads. Cryo-lysate from yeast strain Cbp80-SNAP-PrA, Pab1-CLIP-His₁₀ was suspended, and 1/10th fraction was incubated with varying amounts of Ni²⁺ magnetic beads. Samples were bound, washed and eluted under denaturing conditions and run on gel to perform Western blots to detect binding of Pab1-CLIP-His₁₀ and co-IP of Cbp80-SNAP-PrA. Limited effective binding of Pab1 was detected and no Cbp80 could be detected in eluate samples suggestion inefficiency of the His₁₀ tag on Pab1

Imaging of Cbp80 purified complexes on PrA coverslips

Given that the His₁₀ tag added to Pab1 was determined to be non-functional, it was hypothesized that it might be possible to perform a single step purification using only Cbp80-SNAP-PrA, and subsequently anchor labelled and eluted complexes on coverslips with the PrA tag if coverslips were instead functionalized with IgG rather than Ni-NTA. While this methodology would be expected to purify significantly more Cbp80 protein rather than Pab1 and therefore generates much more fluorescent spots from SNAP rather than CLIP, it was believed that this could be corrected for during data analysis if the abundance between the two factors was of a comparable magnitude.

The purification of Cbp80-SNAP-PrA bound mRNPs by IgG conjugated dynabeads from yeast cryo-lysate of the yeast strain where Pab1 was also tagged with CLIP-His₁₀, was prepared where SNAP and CLIP-tags were labelled on beads during the course of the purification for 90 minutes. Samples were eluted under native conditions; however, it was subsequently necessary to remove the PEGyLOX as it would interfere with Cbp80-SNAP-PrA binding to IgG functionalized coverslips. Samples were either diluted 20 fold, such that PEGyLOX could no longer interfere with the PrA and IgG interaction, or run on a 7K MWCO micro-spin size exclusion column to remove PEGyLOX, as previously described. Regardless, the results were similar for both samples, where significantly more signal was observed from Cbp80-SNAP-PrA than Pab1-CLIP-His₁₀, such that post-experimental data correction could not be possible. Undiluted samples revealed clear spots for Pab1-CLIP-His₁₀, yet observing the same region in the channel appropriate for the SNAP dye revealed that the field of view was sutured with fluorescence. While samples could be diluted to permit individual fluorescent spots to be resolved, the dilution factor was too great and resulted in the loss of fluorescent signal from the

CLIP dye and Pab1 protein. While the use of IgG conjugated coverslips did appear to function better than Ni-NTA coverslips by binding more target protein, the results indicate that a single purification and selection step cannot appropriately reduce the complexity of the sample to levels suitable for single-molecule imaging.

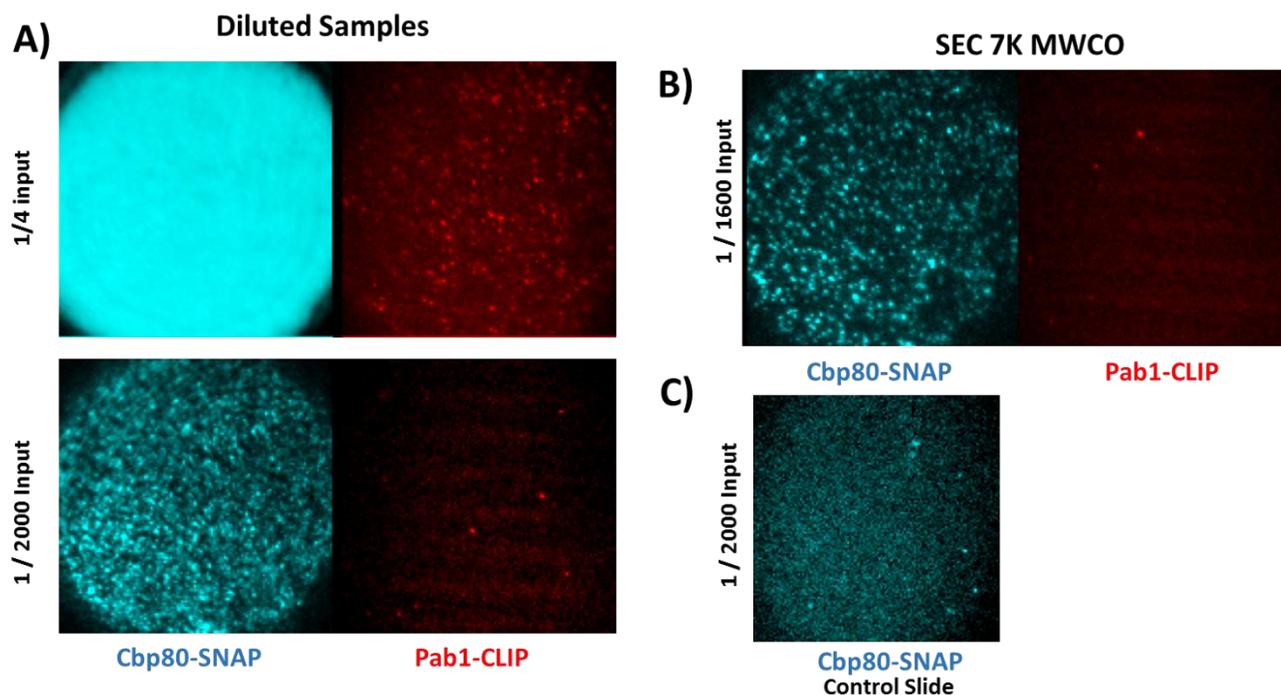


Figure 21. Single-molecule imaging of Cbp80 and Pab1 bound mRNPs on IgG coverslips. Yeast strain Cbp80-SNAP-PrA, Pab1-CLIP-His₁₀ was used for purification of mRNPs with IgG conjugated dynabeads. Proteins were labelled on beads with substrate specific dyes for 90min and eluted under native conditions with PEGylOx. Elute was split in two and either diluted in PEGylOx buffer or run over a 7K MWCO filter prior to placing on IgG conjugated coverslips to prevent PEGylOx. **A)** Shows results of image samples diluted prior to placing on coverslips. Diluted samples show single Cbp80-SNAP-PrA spots, but insufficient spots from Pab1-CLIP-His₁₀ in the same field with CLIP dye specific filter. Concentrated sample added to coverslips show the presence of Pab1-CLIP-His₁₀ signal, but the signal from Cbp80-SNAP-PrA saturates the imaging field and does not allow for single spot resolution. **B)** Shows the result of samples placed on a 7K MWCO spin filter prior to addition on coverslips. Similar results are observed where single spots can be resolved from Cbp80-SNAP-PrA, but insufficient fluorescent spots from Pab1-CLIP-His₁₀ are observed. **C)** Shows elute from dilution sample placed on control IgG coverslips which are not functionalized with IgG to show that there is minimal non-specific binding and cyan spots observed in **A** and **B** are the result of Cbp80-SNAP-PrA binding to coverslips.

Separation of Cbp80 purified mRNPs by SEC

As the excess signal from Cbp80-SNAP-PrA observed under the microscope was believed to be due to the prevalence of free Cbp80, protein molecules associated with any mRNPs, efforts were made to use classical chromatographic separation techniques to separate large mRNP complexes from smaller single free protein molecules. To that effect, cryo-lysate from the strain Cbp80-SNAP-PrA, Pab1-CLIP-His₁₀ was used for IgG dynabead purification, where complexes were labelled with SNAP and CLIP substrate dyes and eluted under native conditions with PEGylOX. The eluate was then passed over a gel filtration column along with several protein standards, and eluting fractions were measured by UV spectrometry at wavelengths appropriate for nucleic acid, proteins and the SNAP dyes. The protein standard Aldolase, measuring 158kDa, was used to approximate the expected elution time of tagged Cbp80, as the Cbp80-SNAP-PrA construct measured 150kDa. In a standard run, Aldolase was observed to elute at the 13ml mark, while the Dextran Blue standard, used to mark the void column volume of the column, was observed to begin eluting at the 2.75ml mark. The average size of a mRNP containing only one copy of Cbp80-SNAP-PrA, one copy of Pab1-CLIP-His₁₀ and a mRNA molecule of 1,500nt in length were estimated as having a molecular weight of 740 kDa. Given this conservative estimate, and the fact that the Superdex 200 10/300 GL gel filtration column used has an exclusion limit of 1.3×10^6 Da, it was believed that mRNPs would readily elute early in the void volume of the column, while free Cbp80-SNAP-PrA molecules would elute later, at approximately the same time as was observed with the aldolase standard.

However, analysis of the chromatogram generated by measuring eluting fractions by UV spectrophotometer revealed a failure to separate free Cbp80 from associated mRNPs. While four peaks are observable on the chromatogram, all contain protein co-eluting with nucleic acids. In

addition, the minor peak at 8mL elution volume does not correspond to the expected elution peak at 13mL for Cbp80-SNAP-PrA, and nor was mRNPs observed to elute in the void volume early from the column. As such, it was determined that SEC could not be used to separate free Cbp80-SNAP-PrA from cellular mRNPs containing Cbp80-SNAP-PrA (Figure 22).

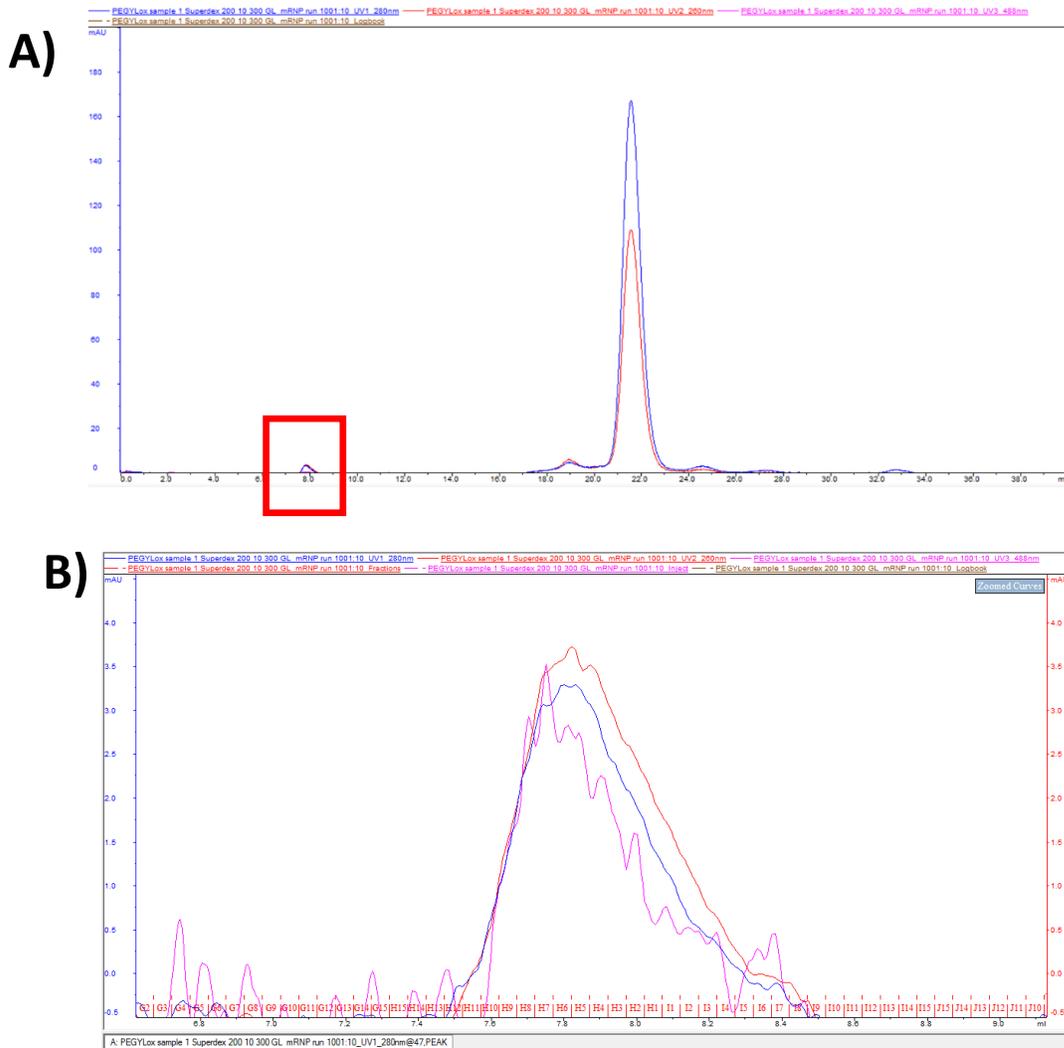


Figure 21. Separation of Cbp80-SNAP-PrA purified complexes by SEC. Cryo-lysate from yeast strain Cbp80-SNAP-PrA, Pab1-CLIP-His₁₀ purified using IgG conjugated dynabeads, labelled with SNAP and CLIP substrate specific dyes and natively eluted using PEGyLOx. Samples were run on Superdex 200 10/300 GL column with an FPLC using PEGyLOx buffer as the running buffer. Elutions were monitored with a UV-Vis spectrophotometer measuring at 260nm (Red), 280nm (Blue) and 488nm (Purple). **A)** shows the full elution profile while **B)** shows a highlight of the red box in **A)**. Free Cbp80-SNAP-PrA was expected to elute at 13mL, while large intact mRNPs were expected to elute at 3mL.

Pab1-CLIP-AP binding assay

Realizing the need to substitute the His₁₀ affinity tag placed on Pab1 with one which is effective and would allow for two selection and affinity purification, a new yeast strain was generated where Pab1 was tagged with CLIP-AP. The Acceptor Peptide (AP) tag is a small peptide which is specifically recognized by BirA biotin ligase co-transfected into cells, which biotinylates the peptide in vivo and allows for affinity purification of associated proteins. The affinity purification can be performed with streptavidin-conjugated beads; however, the interaction between streptavidin and biotin is so strong that proteins can only be eluted under denaturing conditions. Instead, the use of affinity purification beads conjugated with monomeric avidin allows for elution under native condition, if biotin is added to elution buffers to trigger competitive binding and trigger the release of bound complexes from beads.

As such, yeast strains were where Pab1 was tagged with CLIP-AP and BirA was expressed from a co-transformed plasmid were generated. Cryo-lysate of this strain was generated and used to assess AP tag accessibility by performing a binding assay with commercially available monomeric avidin beads. While the manufacturer's protocol recommends performing purification with monomeric avidin beads in PBS buffer at room temperature, our protocol requires that binding is performed at 4°C in TBT100 buffer in order to minimize potential protein degradation and promote purification of intact mRNPs. In order to evaluate the potential for varying the purification conditions, purification and binding assays were performed in TBT100 buffer or PBS and at room temperature or at 4°C. Samples taken during the course of the purification were analyzed by western blot against biotinylated proteins, where it was revealed that Pab1-CLIP-AP exhibited both poor binding and poor elution from monomeric avidin under all conditions tested.

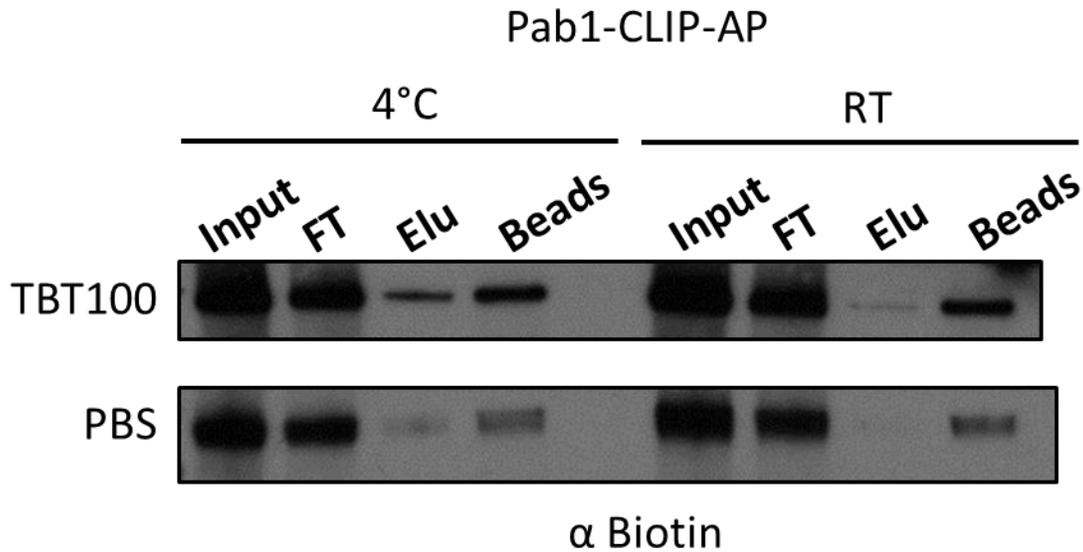


Figure 22. Pab1-CLIP-AP binding assay with monomeric avidin beads. A binding assay was performed with Cyro-lysate from yeast strain Cbp80-SNAP-PrA, Pab1-CLIP-AP, pBirA. The lysate was suspended in either TBT100 or PBS buffer, with 250ul of cleared suspended cryo-lysate incubated with 100ul of monomeric avidin conjugated magnetic beads. 30min binding reactions were performed at 4°C or room temperature (RT) and beads were subsequently washed six times with 1ml of the buffer. Complexes were first eluted under native conditions with 2mM biotin for 10min at room temperature and the supernatant subsequently collected. Remaining beads were subsequently suspended in SDS sample buffer boiled at 95°C to elute under denaturing conditions. Samples were run on SDS-PAGE and western blots were probed with Streptavidin-HRP to detected biotinylated proteins.

Discussion

This work describes steps taken towards optimization of a method capable of quantifying RBPs loaded into mRNP complexes using a combination of single-step affinity purification and high-resolution fluorescence microscopy. The proposed methodology features a novel combination of existing techniques to allow for the eventual investigation of the heterogeneous structure of mRNPs. While these techniques have previously been used individually, the optimization of their use in combination has proven not to be a trivial matter.

The original experimental design sought to quantify the abundance of specific RBPs, or a mRNP associated proteins, on individual specific mRNPs. In order to optimize the system and allow for a proof of principle, it was desired to select proteins with a known abundance on mRNAs in order to demonstrate that the method could perform as expected. However, a limitation resides in the fact that the abundance of few mRNP proteins are known, save for Cbp80 and Cbp20 which bind to the one 5' cap structure of mRNA and are expected to be found in only one copy per transcript. The PP7 stem-loop system used for the purification of transcript specific mRNPs also allowed for the quantification of a protein present in a known quantity, which could further serve as an internal control. As the PP7CP protein binds to PP7 stem-loops as a dimer, four copies of the protein are expected to be found on PMA1 transcripts, as the PMA1 gene was tagged with two PP7 stem-loops in the 3' UTR of the gene.

Initial transcript specific experiments were plagued by several issues which were necessary to overcome. Observations under the microscope of the eluate from IgG purified complexes revealed many PP7-SNAP-PrA spots and few CLIP-His₆ spots, which was unexpected given that complexes were to be anchored to the Ni-NTA coverslips by the His₆ tagged protein. In addition, of the observable spots in both channels, few co-localized suggesting a possible issue with poor labelling efficiency, non-specific protein binding or complex

disassociation. While initial experiments performed with only 15min or 30min labelling reactions with SNAP and CLIP substrate dyes were sub-optimal, longer labelling reactions performed with later purified complexes were only able to slightly improve co-localization suggesting another contributing factor.

The PSACH construct was devised to serve as a positive control for labelling and co-localization. With both SNAP and CLIP tags integrated into the same construct, perfect co-localization would be expected when imaging the construct in both channels. The initial failure to observe co-localization with the PSACH construct hinted at the need to investigate the optimization of the labelling reactions, as optimum conditions were found to differ slightly from previously published reports and manufacturer's recommendations [223, 235, 249, 252-254]. However, even when performing labelling reactions for the optimally determined 90 minutes with both dye substrates, rPSACH co-localization was greatly improved but did not exhibit 100% co-localization. There are two dominant possibilities to explain the persisting lack of co-localization. The first is the presence of a lower molecular weight degradation product which contains the PrA tag, and possibly a residual functional portion of the SNAP-tag. While the abundance of this degradation product was reduced through optimization of the recombinant protein's expression and purification, these low abundance products could remain visible in single-molecule experiments and interfere with the analysis. The second possibility relates to the functionality of the construct given its structure and the fact that it was expressed and purified as a large recombinant protein construct, where ascertaining the absolute functionality of a novel recombinant protein remains difficult. The rPSACH construct is dense with several functional protein domains, as such, it is possible that adjacent domains interfere with one another and do not present an optimal labelling environment, resulting in some rPSACH molecules remaining

unlabelled with one of the dyes. In addition, despite the use of solubilizing agents and buffer conditions which dissuade aggregation, it is possible that a small number of soluble aggregates are formed and retained and the end of the purification, which exhibit poor functionality for labelling. Indeed, quantification of the absolute labelling efficiency of rPSACH was found to lie in the range of 48-82% (Table 7) illustrating that not all molecules were in fact labelled with the SNAP dye.

Experimentation with control coverslips lacking functionalized Ni-NTA or lacking tagged protein complexes capable of binding to functionalized coverslips revealed that there was indeed non-specific protein binding, mainly due to PP7-SNAP-PrA. The stringency of the slide wash buffer was optimized, but it is believed that residual non-specific protein binding to the glass coverslips remains. While experiments were performed where Ni-NTA coverslips were functionalized to also contain a layer of PEG of varying lengths, no significant improvement in terms of non-specific binding of PP7-SNAP-PrA was observed. The PP7-SNAP-PrA construct is believed to be expressed in excess over the number of PMA1 PP7 stem-loop tagged mRNA molecules in the cell, such there exists free PP7-SNAP-PrA protein which is not loaded into mRNPs but is nonetheless still isolated during the first step of ssAP with IgG conjugated dynabeads. Given that modulating buffer conditions were insufficient to eliminate excess PP7-SNAP-PrA from binding to Ni-NTA coverslips, an alternative approach would be to modify the promoter used to drive expression of the construct, such that protein expression could better match mRNA expression levels and thus diminish the amount of free PP7-SNAP-PrA isolated during the first purification step.

One possibility which was not extensively investigated was the disassociation of purified complexes over the course of the experiment. While great effort was made to perform

purifications as quickly as possible, under temperature and buffer conditions which would promote complex stability, protein complexes can nonetheless disassemble over time. In total, from the time cryo-lysate powder is removed from -80°C to the time the first image is taken under the microscope, at least 90 minutes has expired with complexes spending more than half this time at room temperature. However, while mRNP complexes may be transient and dynamic in vivo, once removed from a complex cellular lysate, their structure may be stable as proteins may no longer be targeted by post-translational modifications, and mRNA covered in RBPs is better protected from spontaneous degradation compared to naked RNA. Even if mRNP complexes are stable for a time after ssAP, their disassociation may be triggered by the extensive dilution factor which is required to observe discrete fluorescent foci by TIRF. Protein complexes are held together by a series of protein interactions, which are governed by association and disassociation equilibrium. While PP7 has a disassociation constant of 1nM , extensive dilution may drive the equilibrium to favour complex disassociation, which would otherwise not occur in the crowded environment of the cell or in a concentrated, purified sample [226].

Conversely, transcript specific experimentation may have also failed due to a commonly held assumption present in the majority of the mRNP literature. When investigating the absolute labelling efficiencies by mass spectrometry, strains were generated where Cbp80 and Nab2 were tagged, as these proteins are believed to be recruited to all nuclear mRNAs. Nab2 is recruited late during mRNP maturation as it is recruited to the 3' end mRNAs. Nab2 is believed to serve as a marker for nuclear mRNPs as it is believed to dissociate from mRNPs either immediately before, or shortly after mRNA translocation through the NPC [102, 255, 256]. On the other hand, Cbp80 is recruited early to mRNP, following the synthesis of the 5' m⁷G, and is believed to follow mRNP through the NPC to the cytoplasm, where it is exchanged for the cytoplasmic CBC

containing Tif4631 and Cdc33. Thus, a portion of nuclear mRNPs is expected to contain both Cbp80 and Nab2 at the later stages of mRNP biogenesis, with a number of reviews illustrating such a model. However, in conducting the mass spectrometry experiment for quantification of the SNAP and CLIP-tag labelling efficiency, where complexes were isolated by purifying Cbp80-SNAP-PrA and looking for labelled CLIP-tag on Nab2-CLIP-His₆, no Nab2 protein was identified. While the interaction between Nab2 and Cbp80 has been reported in a handful of studies, analysis of mass spectrometry data from these studies indicates that the evidence is only superficial [102, 203, 257]. Further analysis of published mass spectrometry data for Cbp80 purified complexes, as well as unpublished data from the Oeffinger Lab, revealed a previously unrealized trend, where a majority of 3' processing or 3' mRNA binding proteins tended to be absent from Cbp80 purified complexes. Furthermore, AP-MS performed using 3' processing or 3' mRNA binding proteins tended to find cytoplasmic CBC components Tif4631/4632 and Cdc33. This data suggests that the nuclear CBC may be exchanged for the cytoplasmic CBC much earlier than previously thought, while mRNPs are still within the nucleus and prior to 3' processing of the mRNA. As such, by the time the PP7 stem-loops in the 3' UTR of the PMA1 gene are transcribed, it is possible that Cbp80 has already dissociated and as such could not be found identified when purifying PMA1 mRNA by PP7-SNAP-PrA (Figure 9). However, interestingly enough, 3' binding protein Pab1 was found to associate in a number of these analyzed and published studies, which triggered its selection as a candidate RBP for quantification in place of Nab2 for the general transcript quantification approach.

The RNA dependent interaction between Cbp80 and Pab1 was confirmed through affinity purification and visualization of tagged protein on gel, similar to a co-IP experiment. However, visualization and quantification of Pab1 complexes on Ni-NTA coverslips proved difficult due to

accessibility and functionality issues of the His₆ tag placed on Pab1. In an attempt to overcome the limitations imposed by the inaccessible tag on Pab1, visualization of IgG dynabead purified Cbp80-SNAP-PrA complexes was attempted using only a single purification step. To that effect, the first ever IgG conjugated coverslips were generated through a novel synthetic route used for single-molecule microscopy, which appears to function better than with Ni-NTA coverslips and His₆ tagged proteins. While this experiment failed due to the great discrepancy in the abundance of the two tagged proteins, Cbp80 and Pab1, the improved complex binding to coverslips lead to the desire to use IgG conjugated coverslips for the future single-molecule microscopy experiments and alter the first step purification by changing the tag on Pab1. The AP tag was selected as it was previously reported to function with high affinity, high selectivity and permit for native protein elution, tag functionality and accessibility in mRNP purification conditions remained an important limiting issue which requires extensive time and works to investigate and optimize [249].

Conclusion

New techniques are required to investigate the structure and heterogeneity of nuclear mRNP complexes to better understand how the cell can manage to organize and regulate the export of mRNA molecules of greatly varying properties. While no technique has currently been described to investigate the stoichiometry and protein abundance of individual RBPs or mRNP associated proteins on single specific mRNP complexes by single-molecule resolution, presented here are steps taken to optimize and develop such a method. Additional work is needed to optimize the means of reducing the complexity of the sample obtained by affinity purification, such that a discrete subset of complexes can be analyzed with single-molecule resolution. While single-molecule microscopy experiments offer great resolution allowing for interrogation of individual complexes required to address the questions posed in this study, it is important to note the labour intensive and low throughput nature of such experiments. As a cell expresses many different mRNA transcripts loaded with a variety of proteins, an ideal method would allow for greater throughput to allow for a more rapid deciphering of mRNP heterogeneity. However, as the questions of mRNP composition and heterogeneity remain unanswered with limited active research in the field, the hope is that this work may serve as a primer for the development of a preliminary method capable of providing new insight into the study of mRNPs.

References

1. Rando, O.J. and F. Winston, *Chromatin and transcription in yeast*. Genetics, 2012. **190**(2): p. 351-87.
2. Singh, G., et al., *The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion*. Annu Rev Biochem, 2015. **84**: p. 325-54.
3. Mitchell, S.F. and R. Parker, *Principles and properties of eukaryotic mRNPs*. Mol Cell, 2014. **54**(4): p. 547-58.
4. Lang, O.W., et al., *An Introduction to the Saccharomyces Genome Database (SGD)*. Methods Mol Biol, 2018. **1757**: p. 21-30.
5. Napolitano, G., L. Lania, and B. Majello, *RNA polymerase II CTD modifications: how many tales from a single tail*. J Cell Physiol, 2014. **229**(5): p. 538-44.
6. Eick, D. and M. Geyer, *The RNA polymerase II carboxy-terminal domain (CTD) code*. Chem Rev, 2013. **113**(11): p. 8456-90.
7. Meinel, D.M., et al., *Recruitment of TREX to the transcription machinery by its direct binding to the phospho-CTD of RNA polymerase II*. PLoS Genet, 2013. **9**(11): p. e1003914.
8. Corden, J.L., *RNA polymerase II C-terminal domain: Tethering transcription to transcript and template*. Chem Rev, 2013. **113**(11): p. 8423-55.
9. Meinel, D.M. and K. Strasser, *Co-transcriptional mRNP formation is coordinated within a molecular mRNP packaging station in S. cerevisiae*. Bioessays, 2015. **37**(6): p. 666-77.
10. Perales, R. and D. Bentley, *"Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions*. Mol Cell, 2009. **36**(2): p. 178-91.
11. Shibagaki, Y., et al., *mRNA capping enzyme. Isolation and characterization of the gene encoding mRNA guanylyltransferase subunit from Saccharomyces cerevisiae*. J Biol Chem, 1992. **267**(14): p. 9521-8.
12. Itoh, N., K. Mizumoto, and Y. Kaziro, *Messenger RNA guanylyltransferase from Saccharomyces cerevisiae. II. Catalytic properties*. J Biol Chem, 1984. **259**(22): p. 13930-6.
13. Ho, C.K., B. Schwer, and S. Shuman, *Genetic, physical, and functional interactions between the triphosphatase and guanylyltransferase components of the yeast mRNA capping apparatus*. Mol Cell Biol, 1998. **18**(9): p. 5189-98.
14. Wang, S.P. and S. Shuman, *Structure-function analysis of the mRNA cap methyltransferase of Saccharomyces cerevisiae*. J Biol Chem, 1997. **272**(23): p. 14683-9.
15. Shuman, S., *What messenger RNA capping tells us about eukaryotic evolution*. Nat Rev Mol Cell Biol, 2002. **3**(8): p. 619-25.
16. Fabrega, C., et al., *Structure of an mRNA Capping Enzyme Bound to the Phosphorylated Carboxy-Terminal Domain of RNA Polymerase II*. Molecular Cell, 2003. **11**(6): p. 1549-1561.
17. Gu, M., K.R. Rajashankar, and C.D. Lima, *Structure of the Saccharomyces cerevisiae Cet1-Ceg1 mRNA capping apparatus*. Structure, 2010. **18**(2): p. 216-27.
18. Heidemann, M., et al., *Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription*. Biochim Biophys Acta, 2013. **1829**(1): p. 55-62.
19. Jiao, X., et al., *Identification of a quality-control mechanism for mRNA 5'-end capping*. Nature, 2010. **467**(7315): p. 608-11.
20. Xiang, S., et al., *Structure and function of the 5'-->3' exoribonuclease Rat1 and its activating partner Rai1*. Nature, 2009. **458**(7239): p. 784-8.
21. Tsukamoto, T., et al., *Isolation and characterization of the yeast mRNA capping enzyme beta subunit gene encoding RNA 5'-triphosphatase, which is essential for cell viability*. Biochem Biophys Res Commun, 1997. **239**(1): p. 116-22.

22. Giaever, G., et al., *Functional profiling of the Saccharomyces cerevisiae genome*. Nature, 2002. **418**(6896): p. 387-91.
23. Lewis, J.D., D. Gorlich, and I.W. Mattaj, *A yeast cap binding protein complex (yCBC) acts at an early step in pre-mRNA splicing*. Nucleic Acids Res, 1996. **24**(17): p. 3332-6.
24. Izaurralde, E., et al., *A nuclear cap binding protein complex involved in pre-mRNA splicing*. Cell, 1994. **78**(4): p. 657-68.
25. Worch, R., et al., *Specificity of recognition of mRNA 5' cap by human nuclear cap-binding complex*. RNA, 2005. **11**(9): p. 1355-63.
26. Mazza, C., et al., *Crystal structure of the human nuclear cap binding complex*. Mol Cell, 2001. **8**(2): p. 383-96.
27. Fortes, P., et al., *Genetic and physical interactions involving the yeast nuclear cap-binding complex*. Mol Cell Biol, 1999. **19**(10): p. 6543-53.
28. Wong, C.M., et al., *Yeast cap binding complex impedes recruitment of cleavage factor IA to weak termination sites*. Mol Cell Biol, 2007. **27**(18): p. 6520-31.
29. Ghaemmaghami, S., et al., *Global analysis of protein expression in yeast*. Nature, 2003. **425**(6959): p. 737-41.
30. Lahudkar, S., et al., *The mRNA cap-binding complex stimulates the formation of pre-initiation complex at the promoter via its interaction with Mot1p in vivo*. Nucleic Acids Res, 2011. **39**(6): p. 2188-209.
31. Hossain, M.A., et al., *The yeast cap binding complex modulates transcription factor recruitment and establishes proper histone H3K36 trimethylation during active transcription*. Mol Cell Biol, 2013. **33**(4): p. 785-99.
32. Gornemann, J., et al., *Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex*. Mol Cell, 2005. **19**(1): p. 53-63.
33. Colot, H.V., F. Stutz, and M. Rosbash, *The yeast splicing factor Mud13p is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex*. Genes Dev, 1996. **10**(13): p. 1699-708.
34. Flaherty, S.M., et al., *Participation of the nuclear cap binding complex in pre-mRNA 3' processing*. Proc Natl Acad Sci U S A, 1997. **94**(22): p. 11893-8.
35. Lewis, J.D. and E. Izaurralde, *The role of the cap structure in RNA processing and nuclear export*. Eur J Biochem, 1997. **247**(2): p. 461-9.
36. Fresco, L.D. and S. Buratowski, *Conditional mutants of the yeast mRNA capping enzyme show that the cap enhances, but is not required for, mRNA splicing*. RNA, 1996. **2**(6): p. 584-96.
37. Baltz, A.G., et al., *The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts*. Mol Cell, 2012. **46**(5): p. 674-90.
38. Castello, A., et al., *Insights into RNA biology from an atlas of mammalian mRNA-binding proteins*. Cell, 2012. **149**(6): p. 1393-406.
39. Pabis, M., et al., *Binding properties and dynamic localization of an alternative isoform of the cap-binding complex subunit CBP20*. Nucleus, 2010. **1**(5): p. 412-21.
40. Lidschreiber, M., K. Leike, and P. Cramer, *Cap completion and C-terminal repeat domain kinase recruitment underlie the initiation-elongation transition of RNA polymerase II*. Mol Cell Biol, 2013. **33**(19): p. 3805-16.
41. Zenklusen, D., et al., *Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p*. Mol Cell Biol, 2002. **22**(23): p. 8241-53.
42. Tuck, A.C. and D. Tollervy, *A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs*. Cell, 2013. **154**(5): p. 996-1009.
43. Baejen, C., et al., *Transcriptome maps of mRNP biogenesis factors define pre-mRNA recognition*. Mol Cell, 2014. **55**(5): p. 745-57.

44. Mehlin, H. and B. Daneholt, *The Balbiani ring particle: a model for the assembly and export of RNPs from the nucleus?* Trends Cell Biol, 1993. **3**(12): p. 443-7.
45. Visa, N., et al., *A nuclear cap-binding complex binds Balbiani ring pre-mRNA cotranscriptionally and accompanies the ribonucleoprotein particle during nuclear export.* J Cell Biol, 1996. **133**(1): p. 5-14.
46. Oeffinger, M. and D. Zenklusen, *To the pore and through the pore: a story of mRNA export kinetics.* Biochim Biophys Acta, 2012. **1819**(6): p. 494-506.
47. Strasser, K., et al., *TREX is a conserved complex coupling transcription with messenger RNA export.* Nature, 2002. **417**(6886): p. 304-8.
48. Chavez, S., et al., *A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae.* EMBO J, 2000. **19**(21): p. 5824-34.
49. Masuda, S., et al., *Recruitment of the human TREX complex to mRNA during splicing.* Genes Dev, 2005. **19**(13): p. 1512-7.
50. Poulsen, J.B., et al., *Structural characterization of the Saccharomyces cerevisiae THO complex by small-angle X-ray scattering.* PLoS One, 2014. **9**(7): p. e103470.
51. Pena, A., et al., *Architecture and nucleic acids recognition mechanism of the THO complex, an mRNP assembly factor.* EMBO J, 2012. **31**(6): p. 1605-16.
52. Johnson, S.A., G. Cubberley, and D.L. Bentley, *Cotranscriptional recruitment of the mRNA export factor Yra1 by direct interaction with the 3' end processing factor Pcf11.* Mol Cell, 2009. **33**(2): p. 215-26.
53. MacKellar, A.L. and A.L. Greenleaf, *Cotranscriptional association of mRNA export factor Yra1 with C-terminal domain of RNA polymerase II.* J Biol Chem, 2011. **286**(42): p. 36385-95.
54. Gomez-Gonzalez, B., et al., *Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles.* EMBO J, 2011. **30**(15): p. 3106-19.
55. Piruat, J.I. and A. Aguilera, *A novel yeast gene, THO2, is involved in RNA pol II transcription and provides new evidence for transcriptional elongation-associated recombination.* EMBO J, 1998. **17**(16): p. 4859-72.
56. Yu, T.Y., C.Y. Wang, and J.J. Lin, *Depleting components of the THO complex causes increased telomere length by reducing the expression of the telomere-associated protein Rif1p.* PLoS One, 2012. **7**(3): p. e33498.
57. Pfeiffer, V., et al., *The THO complex component Thp2 counteracts telomeric R-loops and telomere shortening.* EMBO J, 2013. **32**(21): p. 2861-71.
58. Huertas, P. and A. Aguilera, *Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination.* Mol Cell, 2003. **12**(3): p. 711-21.
59. Libri, D., et al., *Interactions between mRNA export commitment, 3'-end quality control, and nuclear degradation.* Mol Cell Biol, 2002. **22**(23): p. 8254-66.
60. Aguilera, A. and H.L. Klein, *HPR1, a novel yeast gene that prevents intrachromosomal excision recombination, shows carboxy-terminal homology to the Saccharomyces cerevisiae TOP1 gene.* Mol Cell Biol, 1990. **10**(4): p. 1439-51.
61. Gonzalez-Barrera, S., et al., *Defective nucleotide excision repair in yeast hpr1 and tho2 mutants.* Nucleic Acids Res, 2002. **30**(10): p. 2193-201.
62. Rougemaille, M., et al., *THO/Sub2p functions to coordinate 3'-end processing with gene-nuclear pore association.* Cell, 2008. **135**(2): p. 308-21.
63. Mouaikel, J., et al., *High-frequency promoter firing links THO complex function to heavy chromatin formation.* Cell Rep, 2013. **5**(4): p. 1082-94.

64. Chi, B., et al., *Aly and THO are required for assembly of the human TREX complex and association of TREX components with the spliced mRNA*. Nucleic Acids Res, 2013. **41**(2): p. 1294-306.
65. Dominguez-Sanchez, M.S., et al., *Genome instability and transcription elongation impairment in human cells depleted of THO/TREX*. PLoS Genet, 2011. **7**(12): p. e1002386.
66. Wang, X., et al., *An allelic series for studying the mouse Thoc1 gene*. Genesis, 2007. **45**(1): p. 32-7.
67. Wang, X., et al., *Thoc1/Hpr1/p84 is essential for early embryonic development in the mouse*. Mol Cell Biol, 2006. **26**(11): p. 4362-7.
68. Wang, X., et al., *Thoc1 deficiency compromises gene expression necessary for normal testis development in the mouse*. Mol Cell Biol, 2009. **29**(10): p. 2794-803.
69. Pitzonka, L., et al., *The Thoc1 encoded ribonucleoprotein is required for myeloid progenitor cell homeostasis in the adult mouse*. PLoS One, 2014. **9**(5): p. e97628.
70. Pitzonka, L., et al., *The THO ribonucleoprotein complex is required for stem cell homeostasis in the adult mouse small intestine*. Mol Cell Biol, 2013. **33**(17): p. 3505-14.
71. Guo, S., et al., *Linking transcriptional elongation and messenger RNA export to metastatic breast cancers*. Cancer Res, 2005. **65**(8): p. 3011-6.
72. Chinnam, M., et al., *The Thoc1 ribonucleoprotein and prostate cancer progression*. J Natl Cancer Inst, 2014. **106**(11).
73. Yang, J., et al., *Relationships of hHpr1/p84/Thoc1 expression to clinicopathologic characteristics and prognosis in non-small cell lung cancer*. Ann Clin Lab Sci, 2008. **38**(2): p. 105-12.
74. Li, Y., et al., *Cancer cells and normal cells differ in their requirements for Thoc1*. Cancer Res, 2007. **67**(14): p. 6657-64.
75. Wang, L., et al., *The THO complex regulates pluripotency gene mRNA export and controls embryonic stem cell self-renewal and somatic cell reprogramming*. Cell Stem Cell, 2013. **13**(6): p. 676-90.
76. Mancini, A., et al., *THOC5/FMIP, an mRNA export TREX complex protein, is essential for hematopoietic primitive cell survival in vivo*. BMC Biol, 2010. **8**: p. 1.
77. Portman, D.S., J.P. O'Connor, and G. Dreyfuss, *YRA1, an essential Saccharomyces cerevisiae gene, encodes a novel nuclear protein with RNA annealing activity*. RNA, 1997. **3**(5): p. 527-37.
78. Strasser, K. and E. Hurt, *Yra1p, a conserved nuclear RNA-binding protein, interacts directly with Mex67p and is required for mRNA export*. EMBO J, 2000. **19**(3): p. 410-20.
79. Preker, P.J., K.S. Kim, and C. Guthrie, *Expression of the essential mRNA export factor Yra1p is autoregulated by a splicing-dependent mechanism*. RNA, 2002. **8**(8): p. 969-80.
80. Preker, P.J. and C. Guthrie, *Autoregulation of the mRNA export factor Yra1p requires inefficient splicing of its pre-mRNA*. RNA, 2006. **12**(6): p. 994-1006.
81. Dong, S., et al., *YRA1 autoregulation requires nuclear export and cytoplasmic Edc3p-mediated degradation of its pre-mRNA*. Mol Cell, 2007. **25**(4): p. 559-73.
82. Zenklusen, D., et al., *The yeast hnRNP-Like proteins Yra1p and Yra2p participate in mRNA export through interaction with Mex67p*. Mol Cell Biol, 2001. **21**(13): p. 4219-32.
83. Kashyap, A.K., et al., *Biochemical and genetic characterization of Yra1p in budding yeast*. Yeast, 2005. **22**(1): p. 43-56.
84. Takahashi, Y., et al., *Yeast Ulp1, an Smt3-specific protease, associates with nucleoporins*. J Biochem, 2000. **128**(5): p. 723-5.
85. Libri, D., et al., *Multiple roles for the yeast SUB2/yUAP56 gene in splicing*. Genes Dev, 2001. **15**(1): p. 36-41.
86. Zhang, M. and M.R. Green, *Identification and characterization of yUAP/Sub2p, a yeast homolog of the essential human pre-mRNA splicing factor hUAP56*. Genes Dev, 2001. **15**(1): p. 30-5.

87. Strasser, K. and E. Hurt, *Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p*. Nature, 2001. **413**(6856): p. 648-52.
88. Fan, H.Y., R.J. Merker, and H.L. Klein, *High-copy-number expression of Sub2p, a member of the RNA helicase superfamily, suppresses hpr1-mediated genomic instability*. Mol Cell Biol, 2001. **21**(16): p. 5459-70.
89. Hautbergue, G.M., et al., *Mutually exclusive interactions drive handover of mRNA from export adaptors to TAP*. Proc Natl Acad Sci U S A, 2008. **105**(13): p. 5154-9.
90. Durairaj, G., S. Lahudkar, and S.R. Bhaumik, *A new regulatory pathway of mRNA export by an F-box protein, Mdm30*. RNA, 2014. **20**(2): p. 133-42.
91. Viphakone, N., et al., *TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export*. Nat Commun, 2012. **3**: p. 1006.
92. Segref, A., et al., *Mex67p, a novel factor for nuclear mRNA export, binds to both poly(A)+ RNA and nuclear pores*. EMBO J, 1997. **16**(11): p. 3256-71.
93. Hung, M.L., et al., *Arginine methylation of REF/ALY promotes efficient handover of mRNA to TAP/NXF1*. Nucleic Acids Res, 2010. **38**(10): p. 3351-61.
94. Erce, M.A., et al., *Interactions affected by arginine methylation in the yeast protein-protein interaction network*. Mol Cell Proteomics, 2013. **12**(11): p. 3184-98.
95. Iglesias, N., et al., *Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export*. Genes Dev, 2010. **24**(17): p. 1927-38.
96. Gwizdek, C., et al., *Ubiquitin-associated domain of Mex67 synchronizes recruitment of the mRNA export machinery with transcription*. Proc Natl Acad Sci U S A, 2006. **103**(44): p. 16376-81.
97. Hobeika, M., et al., *Coordination of Hpr1 and ubiquitin binding by the UBA domain of the mRNA export factor Mex67*. Mol Biol Cell, 2007. **18**(7): p. 2561-8.
98. Green, D.M., et al., *Nab2p is required for poly(A) RNA export in Saccharomyces cerevisiae and is regulated by arginine methylation via Hmt1p*. J Biol Chem, 2002. **277**(10): p. 7752-60.
99. Viphakone, N., F. Voisinet-Hakil, and L. Minvielle-Sebastia, *Molecular dissection of mRNA poly(A) tail length control in yeast*. Nucleic Acids Res, 2008. **36**(7): p. 2418-33.
100. Brockmann, C., et al., *Structural basis for polyadenosine-RNA binding by Nab2 Zn fingers and its function in mRNA nuclear export*. Structure, 2012. **20**(6): p. 1007-18.
101. Kelly, S.M., et al., *Recognition of polyadenosine RNA by the zinc finger domain of nuclear poly(A) RNA-binding protein 2 (Nab2) is required for correct mRNA 3'-end formation*. J Biol Chem, 2010. **285**(34): p. 26022-32.
102. Batisse, J., et al., *Purification of nuclear poly(A)-binding protein Nab2 reveals association with the yeast transcriptome and a messenger ribonucleoprotein core structure*. J Biol Chem, 2009. **284**(50): p. 34911-7.
103. Gonzalez-Aguilera, C., et al., *Nab2 functions in the metabolism of RNA driven by polymerases II and III*. Mol Biol Cell, 2011. **22**(15): p. 2729-40.
104. Tran, E.J., et al., *The DEAD-box protein Dbp5 controls mRNA export by triggering specific RNA:protein remodeling events*. Mol Cell, 2007. **28**(5): p. 850-9.
105. Graveley, B.R., *Sorting out the complexity of SR protein functions*. RNA, 2000. **6**(9): p. 1197-211.
106. Windgassen, M., et al., *Yeast shuttling SR proteins Npl3p, Gbp2p, and Hrb1p are part of the translating mRNPs, and Npl3p can function as a translational repressor*. Molecular and Cellular Biology, 2004. **24**(23): p. 10479-10491.
107. Gilbert, W. and C. Guthrie, *The Glc7p nuclear phosphatase promotes mRNA export by facilitating association of Mex67p with mRNA*. Mol Cell, 2004. **13**(2): p. 201-12.
108. Krebber, H., et al., *Uncoupling of the hnRNP Npl3p from mRNAs during the stress-induced block in mRNA export*. Genes Dev, 1999. **13**(15): p. 1994-2004.

109. Senger, B., et al., *Mtr10p functions as a nuclear import receptor for the mRNA-binding protein Npl3p*. EMBO J, 1998. **17**(8): p. 2196-207.
110. Hackmann, A., et al., *The mRNA export factor Npl3 mediates the nuclear export of large ribosomal subunits*. EMBO Rep, 2011. **12**(10): p. 1024-31.
111. Faza, M.B., et al., *Role of Mex67-Mtr2 in the nuclear export of 40S pre-ribosomes*. PLoS Genet, 2012. **8**(8): p. e1002915.
112. Yao, W., et al., *Nuclear export of ribosomal 60S subunits by the general mRNA export receptor Mex67-Mtr2*. Mol Cell, 2007. **26**(1): p. 51-62.
113. Dermody, J.L., et al., *Unphosphorylated SR-like protein Npl3 stimulates RNA polymerase II elongation*. PLoS One, 2008. **3**(9): p. e3273.
114. Lei, E.P., H. Krebber, and P.A. Silver, *Messenger RNAs are recruited for nuclear export during transcription*. Genes Dev, 2001. **15**(14): p. 1771-82.
115. Lund, M.K., T.L. Kress, and C. Guthrie, *Autoregulation of Npl3, a yeast SR protein, requires a novel downstream region and serine phosphorylation*. Mol Cell Biol, 2008. **28**(11): p. 3873-81.
116. Bucheli, M.E., et al., *Polyadenylation site choice in yeast is affected by competition between Npl3 and polyadenylation factor CFI*. RNA, 2007. **13**(10): p. 1756-64.
117. He, X. and C. Moore, *Regulation of yeast mRNA 3' end processing by phosphorylation*. Mol Cell, 2005. **19**(5): p. 619-29.
118. Ryan, K., *Pre-mRNA 3' cleavage is reversibly inhibited in vitro by cleavage factor dephosphorylation*. RNA Biol, 2007. **4**(1): p. 26-33.
119. McBride, A.E., et al., *Arginine methylation of yeast mRNA-binding protein Npl3 directly affects its function, nuclear export, and intranuclear protein interactions*. J Biol Chem, 2005. **280**(35): p. 30888-98.
120. Xu, C. and M.F. Henry, *Nuclear export of hnRNP Hrp1p and nuclear export of hnRNP Npl3p are linked and influenced by the methylation state of Npl3p*. Mol Cell Biol, 2004. **24**(24): p. 10742-56.
121. Yu, M.C., et al., *Arginine methyltransferase affects interactions and recruitment of mRNA processing and export factors*. Genes Dev, 2004. **18**(16): p. 2024-35.
122. Wong, C.M., et al., *Yeast arginine methyltransferase Hmt1p regulates transcription elongation and termination by methylating Npl3p*. Nucleic Acids Res, 2010. **38**(7): p. 2217-28.
123. Kress, T.L., N.J. Krogan, and C. Guthrie, *A single SR-like protein, Npl3, promotes pre-mRNA splicing in budding yeast*. Mol Cell, 2008. **32**(5): p. 727-34.
124. Moehle, E.A., et al., *The yeast SR-like protein Npl3 links chromatin modification to mRNA processing*. PLoS Genet, 2012. **8**(11): p. e1003101.
125. Santos-Pereira, J.M., et al., *Npl3, a new link between RNA-binding proteins and the maintenance of genome integrity*. Cell Cycle, 2014. **13**(10): p. 1524-9.
126. Flach, J., et al., *A yeast RNA-binding protein shuttles between the nucleus and the cytoplasm*. Mol Cell Biol, 1994. **14**(12): p. 8399-407.
127. Sopko, R., et al., *Mapping pathways and phenotypes by systematic gene overexpression*. Mol Cell, 2006. **21**(3): p. 319-30.
128. Jani, D., et al., *Sus1, Cdc31, and the Sac3 CID region form a conserved interaction platform that promotes nuclear pore association and mRNA export*. Mol Cell, 2009. **33**(6): p. 727-37.
129. Gonzalez-Aguilera, C., et al., *The THP1-SAC3-SUS1-CDC31 complex works in transcription elongation-mRNA export preventing RNA-mediated genome instability*. Mol Biol Cell, 2008. **19**(10): p. 4310-8.
130. Fischer, T., et al., *The mRNA export machinery requires the novel Sac3p-Thp1p complex to dock at the nucleoplasmic entrance of the nuclear pores*. EMBO J, 2002. **21**(21): p. 5843-52.
131. Fischer, T., et al., *Yeast centrin Cdc31 is linked to the nuclear mRNA export machinery*. Nat Cell Biol, 2004. **6**(9): p. 840-8.

132. Daniel, J.A., et al., *Deubiquitination of histone H2B by a yeast acetyltransferase complex regulates transcription*. J Biol Chem, 2004. **279**(3): p. 1867-71.
133. Kuo, M.H., et al., *Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines*. Nature, 1996. **383**(6597): p. 269-72.
134. Galan, A. and S. Rodriguez-Navarro, *Sus1/ENY2: a multitasking protein in eukaryotic gene expression*. Crit Rev Biochem Mol Biol, 2012. **47**(6): p. 556-68.
135. Ellisdon, A.M., et al., *Structural basis for the interaction between yeast Spt-Ada-Gcn5 acetyltransferase (SAGA) complex components Sgf11 and Sus1*. J Biol Chem, 2010. **285**(6): p. 3850-6.
136. Kohler, A., et al., *The mRNA export factor Sus1 is involved in Spt/Ada/Gcn5 acetyltransferase-mediated H2B deubiquitinylation through its interaction with Ubp8 and Sgf11*. Mol Biol Cell, 2006. **17**(10): p. 4228-36.
137. Huisinga, K.L. and B.F. Pugh, *A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae*. Mol Cell, 2004. **13**(4): p. 573-85.
138. Lee, T.I., et al., *Redundant roles for the TFIID and SAGA complexes in global transcription*. Nature, 2000. **405**(6787): p. 701-4.
139. Pascual-Garcia, P., et al., *Sus1 is recruited to coding regions and functions during transcription elongation in association with SAGA and TREX2*. Genes Dev, 2008. **22**(20): p. 2811-22.
140. Rodriguez-Navarro, S., et al., *Sus1, a functional component of the SAGA histone acetylase complex and the nuclear pore-associated mRNA export machinery*. Cell, 2004. **116**(1): p. 75-86.
141. Oeffinger, M., et al., *Comprehensive analysis of diverse ribonucleoprotein complexes*. Nat Methods, 2007. **4**(11): p. 951-6.
142. Green, E.M., et al., *A negative feedback loop at the nuclear periphery regulates GAL gene expression*. Mol Biol Cell, 2012. **23**(7): p. 1367-75.
143. Jani, D., E. Valkov, and M. Stewart, *Structural basis for binding the TREX2 complex to nuclear pores, GAL1 localisation and mRNA export*. Nucleic Acids Res, 2014. **42**(10): p. 6686-97.
144. Santos-Pereira, J.M., et al., *A genome-wide function of THSC/TREX-2 at active genes prevents transcription-replication collisions*. Nucleic Acids Res, 2014. **42**(19): p. 12000-14.
145. Ivessa, A.S., et al., *The Saccharomyces cerevisiae helicase Rrm3p facilitates replication past nonhistone protein-DNA complexes*. Mol Cell, 2003. **12**(6): p. 1525-36.
146. Azvolinsky, A., et al., *Highly transcribed RNA polymerase II genes are impediments to replication fork progression in Saccharomyces cerevisiae*. Mol Cell, 2009. **34**(6): p. 722-34.
147. Bhatia, V., et al., *BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2*. Nature, 2014. **511**(7509): p. 362-5.
148. Liu, J. and W.D. Heyer, *Who's who in human recombination: BRCA2 and RAD52*. Proc Natl Acad Sci U S A, 2011. **108**(2): p. 441-2.
149. Jasin, M., *Homologous repair of DNA damage and tumorigenesis: the BRCA connection*. Oncogene, 2002. **21**(58): p. 8981-93.
150. Gaillard, H., R.E. Wellinger, and A. Aguilera, *A new connection of mRNP biogenesis and export with transcription-coupled repair*. Nucleic Acids Res, 2007. **35**(12): p. 3893-906.
151. Weiss, V.H., et al., *The structure and oligomerization of the yeast arginine methyltransferase, Hmt1*. Nat Struct Biol, 2000. **7**(12): p. 1165-71.
152. Messier, V., D. Zenklusen, and S.W. Michnick, *A nutrient-responsive pathway that determines M phase timing through control of B-cyclin mRNA stability*. Cell, 2013. **153**(5): p. 1080-93.
153. Crespo, J.L. and M.N. Hall, *Elucidating TOR Signaling and Rapamycin Action: Lessons from Saccharomyces cerevisiae*. Microbiology and Molecular Biology Reviews, 2002. **66**(4): p. 579-591.

154. Loewith, R. and M.N. Hall, *Target of rapamycin (TOR) in nutrient signaling and growth control*. Genetics, 2011. **189**(4): p. 1177-201.
155. Quaresma, A.J., R. Sievert, and J.A. Nickerson, *Regulation of mRNA export by the PI3 kinase/AKT signal transduction pathway*. Mol Biol Cell, 2013. **24**(8): p. 1208-21.
156. Vivanco, I. and C.L. Sawyers, *The phosphatidylinositol 3-Kinase AKT pathway in human cancer*. Nat Rev Cancer, 2002. **2**(7): p. 489-501.
157. Rodriguez-Escudero, I., et al., *Reconstitution of the mammalian PI3K/PTEN/Akt pathway in yeast*. Biochem J, 2005. **390**(Pt 2): p. 613-23.
158. Wilkinson, D. and M. Ramsdale, *Proteases and caspase-like activity in the yeast Saccharomyces cerevisiae*. Biochem Soc Trans, 2011. **39**(5): p. 1502-8.
159. Koren, A., *Is Saccharomyces cerevisiae apoptotic cell death associated with gene transfer?* IUBMB Life, 2006. **58**(4): p. 203-7.
160. Wera, S., J.C. Bergsma, and J.M. Thevelein, *Phosphoinositides in yeast: genetically tractable signalling*. FEMS Yeast Res, 2001. **1**(1): p. 9-13.
161. Heymont, J., et al., *TEP1, the yeast homolog of the human tumor suppressor gene PTEN/MMAC1/TEP1, is linked to the phosphatidylinositol pathway and plays a role in the developmental process of sporulation*. Proc Natl Acad Sci U S A, 2000. **97**(23): p. 12672-7.
162. Casamayor, A., et al., *Functional counterparts of mammalian protein kinases PDK1 and SGK in budding yeast*. Curr Biol, 1999. **9**(4): p. 186-97.
163. Wilson, K.F., W.J. Wu, and R.A. Cerione, *Cdc42 stimulates RNA splicing via the S6 kinase and a novel S6 kinase target, the nuclear cap-binding complex*. J Biol Chem, 2000. **275**(48): p. 37307-10.
164. Okada, M., S.W. Jang, and K. Ye, *Akt phosphorylation and nuclear phosphoinositide association mediate mRNA export and cell proliferation activities by ALY*. Proc Natl Acad Sci U S A, 2008. **105**(25): p. 8649-54.
165. Ozsolak, F., et al., *Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation*. Cell, 2010. **143**(6): p. 1018-29.
166. Chen, F., C.C. MacDonald, and J. Wilusz, *Cleavage site determinants in the mammalian polyadenylation signal*. Nucleic Acids Res, 1995. **23**(14): p. 2614-20.
167. Irniger, S., C.M. Egli, and G.H. Braus, *Different classes of polyadenylation sites in the yeast Saccharomyces cerevisiae*. Mol Cell Biol, 1991. **11**(6): p. 3060-9.
168. Mandel, C.R., Y. Bai, and L. Tong, *Protein factors in pre-mRNA 3'-end processing*. Cell Mol Life Sci, 2008. **65**(7-8): p. 1099-122.
169. Park, N.J., D.C. Tsao, and H.G. Martinson, *The two steps of poly(A)-dependent termination, pausing and release, can be uncoupled by truncation of the RNA polymerase II carboxyl-terminal repeat domain*. Mol Cell Biol, 2004. **24**(10): p. 4092-103.
170. Nag, A., K. Narsinh, and H.G. Martinson, *The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase*. Nat Struct Mol Biol, 2007. **14**(7): p. 662-9.
171. Kazerouninia, A., B. Ngo, and H.G. Martinson, *Poly(A) signal-dependent degradation of unprocessed nascent transcripts accompanies poly(A) signal-dependent transcriptional pausing in vitro*. RNA, 2010. **16**(1): p. 197-210.
172. Preker, P.J., et al., *A multisubunit 3' end processing factor from yeast containing poly(A) polymerase and homologues of the subunits of mammalian cleavage and polyadenylation specificity factor*. EMBO J, 1997. **16**(15): p. 4727-37.
173. Zhao, J., M.M. Kessler, and C.L. Moore, *Cleavage factor II of Saccharomyces cerevisiae contains homologues to subunits of the mammalian Cleavage/ polyadenylation specificity factor and*

- exhibits sequence-specific, ATP-dependent interaction with precursor RNA.* J Biol Chem, 1997. **272**(16): p. 10831-8.
174. Sachs, A.B., R.W. Davis, and R.D. Kornberg, *A single domain of yeast poly(A)-binding protein is necessary and sufficient for RNA binding and cell viability.* Mol Cell Biol, 1987. **7**(9): p. 3268-76.
175. Luo, W., A.W. Johnson, and D.L. Bentley, *The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model.* Genes Dev, 2006. **20**(8): p. 954-65.
176. Luo, W. and D. Bentley, *A ribonucleolytic rat torpedoed RNA polymerase II.* Cell, 2004. **119**(7): p. 911-4.
177. Lunde, B.M., et al., *Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain.* Nat Struct Mol Biol, 2010. **17**(10): p. 1195-201.
178. Kessler, M.M., et al., *Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3'-end formation in yeast.* Genes Dev, 1997. **11**(19): p. 2545-56.
179. Sadowski, M., et al., *Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination.* EMBO J, 2003. **22**(9): p. 2167-77.
180. Henry, M., et al., *Potential RNA binding proteins in Saccharomyces cerevisiae identified as suppressors of temperature-sensitive mutations in NPL3.* Genetics, 1996. **142**(1): p. 103-15.
181. Gonzalez, C.I., et al., *The yeast hnRNP-like protein Hrp1/Nab4 marks a transcript for nonsense-mediated mRNA decay.* Mol Cell, 2000. **5**(3): p. 489-99.
182. Gaillard, H. and A. Aguilera, *Cleavage factor I links transcription termination to DNA damage response and genome integrity maintenance in Saccharomyces cerevisiae.* PLoS Genet, 2014. **10**(3): p. e1004203.
183. Arigo, J.T., et al., *Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3.* Mol Cell, 2006. **23**(6): p. 841-51.
184. Steinmetz, E.J., et al., *RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts.* Nature, 2001. **413**(6853): p. 327-31.
185. Vasiljeva, L., et al., *The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain.* Nat Struct Mol Biol, 2008. **15**(8): p. 795-804.
186. Ursic, D., et al., *The yeast SEN1 gene is required for the processing of diverse RNA classes.* Nucleic Acids Res, 1997. **25**(23): p. 4778-85.
187. Vasiljeva, L. and S. Buratowski, *Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts.* Mol Cell, 2006. **21**(2): p. 239-48.
188. Steinmetz, E.J. and D.A. Brow, *Control of pre-mRNA accumulation by the essential yeast protein Nrd1 requires high-affinity transcript binding and a domain implicated in RNA polymerase II association.* Proc Natl Acad Sci U S A, 1998. **95**(12): p. 6699-704.
189. Gudipati, R.K., et al., *Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice.* Nat Struct Mol Biol, 2008. **15**(8): p. 786-94.
190. Guo, J., et al., *Poly(A) signals located near the 5' end of genes are silenced by a general mechanism that prevents premature 3'-end processing.* Mol Cell Biol, 2011. **31**(4): p. 639-51.
191. Kim, H., et al., *Gene-specific RNA polymerase II phosphorylation and the CTD code.* Nat Struct Mol Biol, 2010. **17**(10): p. 1279-86.
192. Rondon, A.G., et al., *Fail-safe transcriptional termination for protein-coding genes in S. cerevisiae.* Mol Cell, 2009. **36**(1): p. 88-98.
193. Webb, S., et al., *PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast.* Genome Biol, 2014. **15**(1): p. R8.

194. Creamer, T.J., et al., *Transcriptome-wide binding sites for components of the Saccharomyces cerevisiae non-poly(A) termination pathway: Nrd1, Nab3, and Sen1*. PLoS Genet, 2011. **7**(10): p. e1002329.
195. Carroll, K.L., et al., *Identification of cis elements directing termination of yeast nonpolyadenylated snoRNA transcripts*. Mol Cell Biol, 2004. **24**(14): p. 6241-52.
196. Carroll, K.L., et al., *Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements*. RNA, 2007. **13**(3): p. 361-73.
197. Darby, M.M., et al., *The Saccharomyces cerevisiae Nrd1-Nab3 transcription termination pathway acts in opposition to Ras signaling and mediates response to nutrient depletion*. Mol Cell Biol, 2012. **32**(10): p. 1762-75.
198. Zaman, S., et al., *How Saccharomyces responds to nutrients*. Annu Rev Genet, 2008. **42**: p. 27-81.
199. Tamanoi, F., *Ras signaling in yeast*. Genes Cancer, 2011. **2**(3): p. 210-5.
200. Aitchison, J.D. and M.P. Rout, *The yeast nuclear pore complex and transport through it*. Genetics, 2012. **190**(3): p. 855-83.
201. Tutucci, E. and F. Stutz, *Keeping mRNPs in check during assembly and nuclear export*. Nat Rev Mol Cell Biol, 2011. **12**(6): p. 377-84.
202. Oeffinger, M. and B. Montpetit, *Emerging properties of nuclear RNP biogenesis and export*. Curr Opin Cell Biol, 2015. **34**: p. 46-53.
203. Bretes, H., et al., *Sumoylation of the THO complex regulates the biogenesis of a subset of mRNPs*. Nucleic Acids Res, 2014. **42**(8): p. 5043-58.
204. Hieronymus, H. and P.A. Silver, *Genome-wide analysis of RNA-protein interactions illustrates specificity of the mRNA export machinery*. Nat Genet, 2003. **33**(2): p. 155-61.
205. Muddukrishna, B., C.A. Jackson, and M.C. Yu, *Protein arginine methylation of Npl3 promotes splicing of the SUS1 intron harboring non-consensus 5' splice site and branch site*. Biochim Biophys Acta, 2017. **1860**(6): p. 730-739.
206. Lei, E.P. and P.A. Silver, *Intron status and 3'-end formation control cotranscriptional export of mRNA*. Genes Dev, 2002. **16**(21): p. 2761-6.
207. Meaux, S., A. van Hoof, and K.E. Baker, *Nonsense-mediated mRNA decay in yeast does not require PAB1 or a poly(A) tail*. Mol Cell, 2008. **29**(1): p. 134-40.
208. Davila Lopez, M. and T. Samuelsson, *Early evolution of histone mRNA 3' end processing*. RNA, 2008. **14**(1): p. 1-10.
209. Marzluff, W.F., E.J. Wagner, and R.J. Duronio, *Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail*. Nat Rev Genet, 2008. **9**(11): p. 843-54.
210. Beggs, S., T.C. James, and U. Bond, *The PolyA tail length of yeast histone mRNAs varies during the cell cycle and is influenced by Sen1p and Rrp6p*. Nucleic Acids Res, 2012. **40**(6): p. 2700-11.
211. Singer-Kruger, B. and R.P. Jansen, *Here, there, everywhere. mRNA localization in budding yeast*. RNA Biol, 2014. **11**(8): p. 1031-9.
212. Shen, Z., et al., *Nuclear shuttling of She2p couples ASH1 mRNA localization to its translational repression by recruiting Loc1p and Puf6p*. Mol Biol Cell, 2009. **20**(8): p. 2265-75.
213. Sil, A. and I. Herskowitz, *Identification of asymmetrically localized determinant, Ash1p, required for lineage-specific transcription of the yeast HO gene*. Cell, 1996. **84**(5): p. 711-22.
214. Bobola, N., et al., *Asymmetric accumulation of Ash1p in postanaphase nuclei depends on a myosin and restricts yeast mating-type switching to mother cells*. Cell, 1996. **84**(5): p. 699-709.
215. Niedner, A., et al., *Role of Loc1p in assembly and reorganization of nuclear ASH1 messenger ribonucleoprotein particles in yeast*. Proc Natl Acad Sci U S A, 2013. **110**(52): p. E5049-58.
216. Hurowitz, E.H. and P.O. Brown, *Genome-wide analysis of mRNA lengths in Saccharomyces cerevisiae*. Genome Biol, 2003. **5**(1): p. R2.

217. Taddei, A., H. Schober, and S.M. Gasser, *The budding yeast nucleus*. Cold Spring Harb Perspect Biol, 2010. **2**(8): p. a000612.
218. Taddei, A. and S.M. Gasser, *Structure and function in the budding yeast nucleus*. Genetics, 2012. **192**(1): p. 107-29.
219. Gonatopoulos-Pournatzis, T. and V.H. Cowling, *Cap-binding complex (CBC)*. Biochem J, 2014. **457**(2): p. 231-42.
220. Trahan, C., L.C. Aguilar, and M. Oeffinger, *Single-Step Affinity Purification (ssAP) and Mass Spectrometry of Macromolecular Complexes in the Yeast *S. cerevisiae**. Methods Mol Biol, 2016. **1361**: p. 265-87.
221. Jain, A., et al., *Probing cellular protein complexes using single-molecule pull-down*. Nature, 2011. **473**(7348): p. 484-8.
222. McGuire, H., et al., *Automating single subunit counting of membrane proteins in mammalian cells*. J Biol Chem, 2012. **287**(43): p. 35912-21.
223. Hansen, S.R., M.L. Rodgers, and A.A. Hoskins, *Fluorescent Labeling of Proteins in Whole Cell Extracts for Single-Molecule Imaging*. Methods Enzymol, 2016. **581**: p. 83-104.
224. Oeffinger, M., *Two steps forward--one step back: advances in affinity purification mass spectrometry of macromolecular complexes*. Proteomics, 2012. **12**(10): p. 1591-608.
225. LaCava, J., et al., *Improved native isolation of endogenous Protein A-tagged protein complexes*. Biotechniques, 2013. **54**(4): p. 213-6.
226. Lim, F. and D.S. Peabody, *RNA recognition site of PP7 coat protein*. Nucleic Acids Res, 2002. **30**(19): p. 4138-44.
227. Larson, D.R., R.H. Singer, and D. Zenklusen, *A single molecule view of gene expression*. Trends Cell Biol, 2009. **19**(11): p. 630-7.
228. Hocine, S., et al., *Single-molecule analysis of gene expression using two-color RNA labeling in live yeast*. Nat Methods, 2013. **10**(2): p. 119-21.
229. Yoon, J.H., S. Srikantan, and M. Gorospe, *MS2-TRAP (MS2-tagged RNA affinity purification): tagging RNA to identify associated miRNAs*. Methods, 2012. **58**(2): p. 81-7.
230. Yoon, J.H. and M. Gorospe, *Identification of mRNA-Interacting Factors by MS2-TRAP (MS2-Tagged RNA Affinity Purification)*. Methods Mol Biol, 2016. **1421**: p. 15-22.
231. Fridy, P.C., *Nanobody-Based Interactomic Studies of Single Transcripts During mRNA Maturation*. 2016.
232. Craggs, T.D., *Green fluorescent protein: structure, folding and chromophore maturation*. Chem Soc Rev, 2009. **38**(10): p. 2865-75.
233. Shashkova, S., et al., *Characterising Maturation of GFP and mCherry of Genomically Integrated Fusions in *Saccharomyces cerevisiae**. Bio Protoc, 2018. **8**(2): p. e2710.
234. Dean, K.M. and A.E. Palmer, *Advances in fluorescence labeling strategies for dynamic cellular imaging*. Nat Chem Biol, 2014. **10**(7): p. 512-23.
235. Kolberg, K., et al., *SNAP-tag technology: a general introduction*. Curr Pharm Des, 2013. **19**(30): p. 5406-13.
236. Shcherbakova, I., et al., *Alternative spliceosome assembly pathways revealed by single-molecule fluorescence microscopy*. Cell Rep, 2013. **5**(1): p. 151-65.
237. Gautier, A., et al., *An engineered protein tag for multiprotein labeling in living cells*. Chem Biol, 2008. **15**(2): p. 128-36.
238. Bosch, P.J., et al., *Evaluation of fluorophores to label SNAP-tag fused proteins for multicolor single-molecule tracking microscopy in live cells*. Biophys J, 2014. **107**(4): p. 803-14.
239. Song, L., et al., *Photobleaching kinetics of fluorescein in quantitative fluorescence microscopy*. Biophys J, 1995. **68**(6): p. 2588-600.

240. Cheng, W., et al., *Ultrasensitive scanometric strategy for detection of matrix metalloproteinases using a histidine tagged peptide-Au nanoparticle probe*. Chem Commun (Camb), 2011. **47**(10): p. 2877-9.
241. Kang, E., et al., *Specific adsorption of histidine-tagged proteins on silica surfaces modified with Ni²⁺/NTA-derivatized poly(ethylene glycol)*. Langmuir, 2007. **23**(11): p. 6281-8.
242. Ralser, M., et al., *The Saccharomyces cerevisiae W303-K6001 cross-platform genome sequence: insights into ancestry and physiology of a laboratory mutt*. Open Biol, 2012. **2**(8): p. 120093.
243. Evans, I.H., *Yeast protocols: methods in cell and molecular biology*. 1996: Springer.
244. Janke, C., et al., *A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes*. Yeast, 2004. **21**(11): p. 947-62.
245. Wach, A., et al., *New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae*. Yeast, 1994. **10**(13): p. 1793-808.
246. Kushnirov, V.V., *Rapid and reliable protein extraction from yeast*. Yeast, 2000. **16**(9): p. 857-60.
247. Schägger, H., *Tricine-sds-page*. Nature protocols, 2006. **1**(1): p. 16.
248. Towbin, H., T. Staehelin, and J. Gordon, *Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications*. Proc Natl Acad Sci U S A, 1979. **76**(9): p. 4350-4.
249. Hoskins, A.A., et al., *Ordered and dynamic assembly of single spliceosomes*. Science, 2011. **331**(6022): p. 1289-95.
250. Bornhorst, J.A. and J.J. Falke, *Purification of proteins using polyhistidine affinity tags*. Methods Enzymol, 2000. **326**: p. 245-54.
251. Goyer, C., et al., *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*. Mol Cell Biol, 1993. **13**(8): p. 4860-74.
252. Correa, I.R., Jr., et al., *Substrates for improved live-cell fluorescence labeling of SNAP-tag*. Curr Pharm Des, 2013. **19**(30): p. 5414-20.
253. Hoehnel, S. and M.P. Lutolf, *Capturing Cell-Cell Interactions via SNAP-tag and CLIP-tag Technology*. Bioconjug Chem, 2015. **26**(8): p. 1678-86.
254. Barth, S., *The SNAP-tag technology - a versatile tool with many applications*. Curr Pharm Des, 2013. **19**(30): p. 5404-5.
255. Anderson, J.T., et al., *NAB2: a yeast nuclear polyadenylated RNA-binding protein essential for cell viability*. Mol Cell Biol, 1993. **13**(5): p. 2730-41.
256. Suntharalingam, M., A.R. Alcazar-Roman, and S.R. Wenthe, *Nuclear export of the yeast mRNA-binding protein Nab2 is linked to a direct interaction with Gfd1 and to Gle1 function*. J Biol Chem, 2004. **279**(34): p. 35384-91.
257. Wilmes, G.M., et al., *A genetic interaction map of RNA-processing factors reveals links between Sem1/Dss1-containing complexes and mRNA export and splicing*. Mol Cell, 2008. **32**(5): p. 735-46.