



***Transcriptome maps of general eukaryotic RNA
degradation factors and identification and
functional characterization of the novel mRNA
modification N³-methylcytidine***

Dissertation

for the award of the degree
“Doctor rerum naturalium”
of the Georg-August-Universität Göttingen

within the doctoral program
“Molecular Biology of Cells”
of the Georg-August University School of Science (GAUSS)

submitted by

Katharina Bettina Hofmann

from Würzburg

Göttingen 2019

Thesis Committee

Prof. Dr. Patrick Cramer	Max-Planck-Institute for Biophysical Chemistry, Göttingen
Prof. Dr. Reinhard Lührmann	Max-Planck-Institute for Biophysical Chemistry, Göttingen
Prof. Dr. Claudia Höbartner	Julius-Maximilians-University, Würzburg

Members of the Examination Board

Referee:	Prof. Dr. Patrick Cramer, Max-Planck-Institute for Biophysical Chemistry, Göttingen
2 nd Referee:	Prof. Dr. Reinhard Lührmann, Max-Planck-Institute for Biophysical Chemistry, Göttingen

Further members of the Examination Board

Prof. Dr. Markus Bohnsack	Department of Molecular Biology, University Medical Center, Göttingen
Prof. Dr. Lutz Walter	German Primate Center, Göttingen
Prof. Dr. Henning Urlaub	Max-Planck-Institute for Biophysical Chemistry, Göttingen

Date of oral examination: 06.05.2019

Promovierenden-Erklärung

Georg-August-Universität Göttingen

Name: Hofmann, Katharina Bettina

Anschrift: Otto-Lauffer-Str. 34, 37077 Göttingen

Ich beabsichtige, eine Dissertation zum Thema "Transcriptome maps of general eukaryotic RNA degradation factors and identification and functional characterization of the novel mRNA modification N^3 -methylcytidine" an der Georg-August-Universität Göttingen und dem Max-Planck-Institut für Biophysikalische Chemie anzufertigen. Dabei werde ich von Herrn Prof. Dr. Patrick Cramer betreut.

Ich gebe folgende Erklärung ab:

1. Die Gelegenheit zum vorliegenden Promotionsvorhaben ist mir nicht kommerziell vermittelt worden. Insbesondere habe ich keine Organisation eingeschaltet, die gegen Entgelt Betreuerinnen und Betreuer für die Anfertigung von Dissertationen sucht oder die mir obliegenden Pflichten hinsichtlich der Prüfungsleistungen für mich ganz oder teilweise erledigt.
2. Hilfe Dritter wurde bis jetzt und wird auch künftig nur in wissenschaftlich vertretbarem und prüfungsrechtlich zulässigem Ausmaß in Anspruch genommen. Insbesondere werden alle Teile der Dissertation selbst angefertigt; unzulässige fremde Hilfe habe ich dazu weder unentgeltlich noch entgeltlich entgegengenommen und werde dies auch zukünftig so halten.
3. Die Richtlinien zur Sicherung der guten wissenschaftlichen Praxis an der Universität Göttingen werden von mir beachtet.
4. Eine entsprechende Promotion wurde an keiner anderen Hochschule im In- oder Ausland beantragt; die eingereichte Dissertation oder Teile von ihr wurden nicht für ein anderes Promotionsvorhaben verwendet.

Mir ist bekannt, dass unrichtige Angaben die Zulassung zur Promotion ausschließen bzw. später zum Verfahrensabbruch oder zur Rücknahme des erlangten Grades führen.

Göttingen, den 31.03.2019

Katharina B. Hofmann

Acknowledgements

First, I would like to thank my supervisor Prof. Dr. Patrick Cramer for the opportunity to work on such interesting and state-of-the-art projects. Furthermore, I would like to express my gratitude for his supervision, guidance and all the discussions throughout my PhD time.

Moreover, I am very thankful for the support and discussions to the members of my thesis committee, Prof. Dr. Reinhard Lührmann and Prof. Dr. Claudia Höbartner. I would also like to thank Prof. Dr. Markus Bohnsack, Prof. Dr. Lutz Walter and Prof. Dr. Henning Urlaub for being part of my examination board and their time to evaluate this dissertation.

A special thanks goes to Salma Sohrabi-Jahromi from the group of Dr. Johannes Söding. She performed all the bioinformatic analysis on the degradation project and thereby provided the basis for data interpretation leading to many results for this dissertation. I am very happy to see that all our efforts finally turned into a nice manuscript.

I would like to thank the GGNB Office for their support and for providing a fantastic basis for graduation. I would also like to thank the members of the GGNB Program “Molecular Biology of Cells” for the great retreats, good scientific discussions and leisure activities.

I would like to thank all current and former members of the Cramer group for scientific discussions and a nice working atmosphere. I am particularly grateful to Svetlana for sharing the lab with me and the great time together. A special thanks goes to Janine, Kirsten, Kerstin, Petra, Thomas, Sigurd, Andrea, Angelika, and Manuela for lab infrastructure, organization and technical support.

I am very happy that I can count many of my current and former colleagues to my friends, especially Carina, Ania, Livia, Kristina, Svetlana, Kerstin, Jinmi, and Saskia. I want to thank them for a great time in Göttingen and especially for our fantastic trips to Krakow, Bremen, Barcelona, Berlin, Hamburg, Dublin, Goslar, Lübeck, Brocken, and of course New Zealand. I am very thankful to all my friends, who supported me during my PhD and visited me in Göttingen, particularly Susi, Ina, Carina, Lisa, Sandra, Janine, and Dolores.

I would like to thank Kerstin, Ania, Svetlana, and Jannik for critical reading of this dissertation and their valuable feedback.

I would like to express my deepest gratitude to my family for their constant support and understanding. Last but not least, I am most grateful to Jannik, who cheered me up in difficult times and for his support, patience, understanding and love.

Publications associated with this dissertation

Sohrabi-Jahromi S*, Hofmann KB*, Boltendahl A, Roth C, Gressel S, Baejen C, Soeding J, Cramer P. 2019. **Transcriptome maps of general eukaryotic RNA degradation factors.** *Elife* **8**. doi: 10.7554/eLife.47040

*These authors contributed equally to this work.

Author contributions: SSJ designed and performed bioinformatics analysis. KH optimized and carried out PAR-CLIP experiments. AB, CB, SG carried out PAR-CLIP experiments. CR contributed to bioinformatics analysis. JS and PC designed and supervised research. SSJ, KH, JS and PC prepared the manuscript, with input from other authors.

Hofmann KB, Villamil GJ, Schwalb B, Sohrabi-Jahromi S, Söding J, Cramer P. **Mapping and functional characterization of the N^3 -methylcytidine mRNA modification.**

Current author list.

Current author contributions: KH generated all experimental data. GV did bioinformatic analysis on 4tU-seq data under supervision of BS. SSJ did co-localization analysis under supervision of JS. KH and PC wrote the manuscript with input from all authors.

(Manuscript in preparation)

Other publications derived during the time of this dissertation

Jones JL, Hofmann KB, Cowan AT, Temiakov D, Cramer P, Anikin M. 2019.

Yeast mitochondrial protein Pet111p binds directly to two distinct targets in COX2 mRNA, suggesting a mechanism of translational activation. *J Biol Chem* **294(18)**:1-20. doi:10.1074/jbc.RA118.005355

Author contributions: JJ: validation, investigation, visualization, methodology. KH: data curation, formal analysis, investigation, visualization, methodology, writing-original draft. AC: data curation, investigation. DT: investigation; methodology, writing-original draft, writing-review and editing. PC: conceptualization, supervision, funding acquisition, validation, methodology, writing-original draft. MA: conceptualization, data curation, supervision, funding acquisition, validation, investigation, methodology, writing-original draft, writing-review and editing

Vos SM*, Pöllmann D*, Caizzi L, Hofmann KB, Rombaut P, Zimniak T, Herzog F, Cramer P. 2016. **Architecture and RNA binding of the human negative elongation factor.** *Elife* **5**:1–27. doi:10.7554/eLife.14981

*These authors contributed equally to this work.

Author contributions: SMV: conception and design, acquisition of data, analysis and interpretation of data, drafting or revising the article. DP: acquisition of data, analysis and interpretation of data, drafting or revising the article. LC: acquisition of data, analysis and interpretation of data, drafting or revising the article. KH: acquisition of data, analysis and interpretation of data, drafting or revising the article. PR: acquisition of data, analysis and interpretation of data. TZ: acquisition of data, analysis and interpretation of data. FH: acquisition of data, analysis and interpretation of data. PC: conception and design, analysis and interpretation of data, drafting or revising the article

The following chapters and sections were taken from Sohrabi *et al.*:

3.3 Data analysis

3.3.1 PAR-CLIP data pre-processing

3.3.2 Transcript class enrichment

3.3.3 Metagene analysis

3.3.4 Co-occupancy

3.3.5 Co-localization

3.3.6 Codon-enrichment analysis

3.3.7 Relating occupancies to various transcript features

3.3.8 Motif enrichment analysis

4.1 Results

4.1.1 Degradation factors exhibit transcript class specificity

4.1.1.1 RNA end- processing complexes differ in their targets

4.1.1.2 The exosome and surveillance factors

4.1.2 Distinct factor distribution along mRNA

4.1.3 Surveillance of aberrant nuclear non-coding RNA

4.1.4 Interactions between RNA processing machineries

4.1.5 5' degradation machinery senses translation efficiency

4.1.6 Decapping factors are enriched upon RNA degradation

The following elements have been adapted to the format of this dissertation and may deviate from Sohrabi *et al.*:

- Numbering of figures and tables
- References to figures and tables

The following figures, figure legends, and table were taken from Sohrabi *et al.* (manuscript table and figure number shown in parenthesis):

- Figure 7: Number of PAR-CLIP cross-link sites and replicate correlation. (adapted from Figure 1 and Figure 1 - figure supplement 1)
- Figure 8: Distribution of degradation factor cross-link sites over the yeast transcriptome. (Figure 2)
- Figure 9: Metagene analysis of degradation factor binding on protein-coding mRNAs. (Figure 3)
- Figure 10: Surveillance of aberrant nuclear non-coding RNA by components of the exosome and the TRAMP4 complex. (Figure 4)
- Figure 11: Global co-occupancy and co-localization analysis reveals unexpected cooperation between factors from different complexes and pathways. (Figure 5)
- Figure 12: Binding preferences reveal a link between decapping-mediated degradation and translation. (Figure 6)
- Figure 13: Location and recruitment of the decapping complex Dcp1/Dcp2 and decapping enhancers Edc3, Dhh1, and Edc2. (Figure 7)
- Figure 28: Different transcript classes have comparable U-content. (Figure 1 – figure supplement 2)
- Figure 29: Metagene profiles of yeast RNA degradation factors centered on translation start and stop sites in comparison to TIF-annotated TSS and pA sites. (Figure 3 – figure supplement 1)
- Figure 30: Distributions of transcript length, half-life, expression level and transcript optimality for yeast mRNAs. (Figure 6 – figure supplement 1)
- Figure 31: Occupancies of deadenylation factors (Ccr4, Pop2, Not1, Caf40, Pan2, and Pan3) compared to transcript length, optimality, expression level, and half-life. (Figure 6 – figure supplement 2)
- Figure 32: Occupancies of decapping factors (Dcp2, Dcp1, Edc2, Edc3, and Dhh1) compared to transcript length, optimality, expression level, and half-life. (Figure 6 – figure supplement 3)

- Figure 33: Occupancy of Xrn1 compared to transcript length, optimality, expression level, and half-life. (Figure 6 – figure supplement 4)
- Figure 34: Occupancies of exosome components (Rrp6, Csl4, Rrp40, Rrp4, and Rrp44) compared to transcript length, optimality, expression level, and half-life. (Figure 6 – figure supplement 5)
- Figure 35: Occupancies for components of the TRAMP complex (Air1, Trf5, Mtr4, Air2, and Trf4) compared to transcript length, optimality, expression level, and half-life. (Figure 6 – figure supplement 6)
- Figure 36: Occupancies for components of the Ski complex (Ski2, Ski3, Ski7, and Ski8) compared to transcript length, optimality, expression level, and half-life. (Figure 6 – figure supplement 7)
- Figure 37: Occupancies for components of the NMD pathway (Upf1, Upf2, Upf3, and Nmd4) compared to transcript length, optimality, expression level, and half-life. (Figure 6 – figure supplement 8)
- Figure 38: Correlation between binding to degradation factors and transcript length, codon-optimality, expression, and half-life. (Figure 6 – figure supplement 9)
- Figure 39: Metagene profiles for subunits of the TRAMP complex on snoRNA genes. (Figure 2 – figure supplement 2)
- Figure 40: Comparison of binding profiles on genes containing annotated upstream sense NUTs with all mRNAs. (Figure 3 – figure supplement 2)

Table 13: Overview of RNA processing factors and their respective published PAR-CLIP experiments. (Table supplement 1)

Table of Contents

<i>Promovierenden-Erklärung</i>	III
<i>Acknowledgements</i>	V
<i>Publications associated with this dissertation</i>	VII
1 Summary	1 -
2 Introduction	3 -
2.1 The central dogma of molecular biology	3 -
2.1.1 <i>Transcription</i>	3 -
2.1.2 <i>Translation</i>	5 -
2.2 The RNA degradation pathways	8 -
2.2.1 <i>Nuclear surveillance of aberrant non-coding RNAs</i>	8 -
2.2.2 <i>Cytoplasmic RNA degradation</i>	9 -
2.2.2.1 <i>Deadenylation</i>	9 -
2.2.2.2 <i>Decapping and exonucleolytic 5' → 3' RNA degradation</i>	10 -
2.2.2.3 <i>Exosomal 3' → 5' RNA degradation</i>	11 -
2.2.2.4 <i>Quality control mechanism for defective mRNA</i>	13 -
2.3 The emerging field of epitranscriptomics	14 -
2.3.1 <i>Biogenesis of eukaryotic tRNAs and their function in translation</i>	14 -
2.3.2 <i>tRNA modifications and their function</i>	14 -
2.3.3 <i>mRNA modifications</i>	16 -
2.3.3.1 <i>N⁶-methyladenosine modification</i>	18 -
2.3.3.2 <i>5-methylcytidine modification</i>	20 -
2.3.3.3 <i>N¹-methyladenosine modification</i>	20 -
2.3.3.4 <i>Pseudouridine modification</i>	21 -
2.3.3.5 <i>N⁴-acetylcytidine modification</i>	22 -
2.3.4 <i>N³-methylcytidine modification</i>	23 -
2.3.4.1 <i>N³-methylcytidine methyltransferase in <i>S. cerevisiae</i></i>	23 -
2.3.4.2 <i>N³-methylcytidine methyltransferases in higher eukaryotes</i>	24 -
2.4 Aims of this thesis	25 -
2.4.1 <i>Transcriptome maps of general eukaryotic RNA degradation factors</i> -	25 -
2.4.2 <i>Identification and functional characterization of the novel mRNA modification N³-methylcytidine</i>	25 -
3 Materials and Methodology	27 -
3.1 Materials	27 -
3.1.1 <i>Bacterial strains</i>	27 -

3.1.2	<i>Yeast strains</i>	- 27 -
3.1.3	<i>Human cell lines</i>	- 28 -
3.1.4	<i>Media and supplement</i>	- 29 -
3.1.5	<i>Spike-ins</i>	- 30 -
3.1.6	<i>Primers and oligonucleotides</i>	- 30 -
3.1.7	<i>Thermal cycler programs</i>	- 31 -
3.1.8	<i>Plasmids</i>	- 32 -
3.1.9	<i>Buffers and solutions</i>	- 32 -
3.1.10	<i>Antibodies</i>	- 33 -
3.2	Experimental methodology	- 34 -
3.2.1	<i>Yeast culture</i>	- 34 -
3.2.2	<i>S. cerevisiae strain validation by SDS-PAGE and Western Blot</i>	- 34 -
3.2.3	<i>Generation of the METTL8-3xFLAG Flp-In™ T-REx™ 293 Cell Line</i> -	34 -
3.2.4	<i>PAR-CLIP</i>	- 35 -
3.2.4.1	<i>S. cerevisiae</i>	- 35 -
3.2.4.2	<i>Human</i>	- 38 -
3.2.5	<i>m3C-CLIP</i>	- 40 -
3.2.5.1	<i>S. cerevisiae</i>	- 40 -
3.2.5.2	<i>Human</i>	- 40 -
3.2.6	<i>4tU-seq in S. cerevisiae</i>	- 41 -
3.2.7	<i>Protein purification of AlkB and Trm140</i>	- 41 -
3.2.8	<i>RNA demethylation using AlkB</i>	- 42 -
3.2.9	<i>RNA re-methylation using Trm140 and SAM</i>	- 42 -
3.2.10	<i>Reverse transcription assay and denaturing PAGE</i>	- 42 -
3.3	Data analysis	- 43 -
3.3.1	<i>PAR-CLIP data pre-processing</i>	- 43 -
3.3.2	<i>Transcript class enrichment</i>	- 44 -
3.3.3	<i>Metagene analysis</i>	- 45 -
3.3.4	<i>Co-occupancy</i>	- 46 -
3.3.5	<i>Co-localization</i>	- 47 -
3.3.6	<i>Codon-enrichment analysis</i>	- 47 -
3.3.7	<i>Relating occupancies to various transcript features</i>	- 48 -
3.3.8	<i>Motif enrichment analysis</i>	- 49 -
3.3.9	<i>Trm140 PAR-CLIP and m3C-CLIP analysis</i>	- 50 -
3.3.10	<i>Calling of m3C modification sites</i>	- 51 -
3.3.11	<i>Calculation of ribosome P site occupancy</i>	- 51 -
3.3.12	<i>Calculation of translational efficiency on m3C containing codons</i>	- 51 -
3.3.13	<i>4tU-seq data pre-processing and normalization</i>	- 52 -

3.3.14	<i>Estimation of RNA synthesis and degradation rates</i>	- 53 -
4	Transcriptome maps of general eukaryotic RNA degradation factors	- 55 -
4.1	Results	- 55 -
4.1.1	<i>Degradation factors exhibit transcript class specificity</i>	- 57 -
4.1.1.1	<i>RNA end- processing complexes differ in their targets</i>	- 57 -
4.1.1.2	<i>The exosome and surveillance factors</i>	- 58 -
4.1.2	<i>Distinct factor distribution along mRNA</i>	- 61 -
4.1.3	<i>Surveillance of aberrant nuclear non-coding RNA</i>	- 64 -
4.1.4	<i>Interactions between RNA processing machineries</i>	- 66 -
4.1.5	<i>5' degradation machinery senses translation efficiency</i>	- 69 -
4.1.6	<i>Decapping factors are enriched upon RNA degradation</i>	- 71 -
4.2	Discussion and Outlook	- 74 -
4.2.1	<i>Occupancy profiles of general eukaryotic degradation factors give new insights into RNA degradation processes</i>	- 74 -
4.2.2	<i>Initial step of RNA decay - mRNA deadenylation</i>	- 74 -
4.2.3	<i>Decapping as first step for RNA degradation from the 5' end</i>	- 76 -
4.2.4	<i>Nuclear RNA surveillance mechanism</i>	- 77 -
4.2.5	<i>RNA degradation by the exosome complex and auxiliary factors</i>	- 78 -
4.2.6	<i>Cytoplasmic RNA surveillance by NMD factors</i>	- 80 -
4.2.7	<i>Perspectives of mechanistic insights into the RNA degradation process by complementation of the transcriptome maps with functional studies</i>	- 81 -
4.2.8	<i>Protein-RNA binding studies of eukaryotic degradation factors in response to stress conditions will allow for better functional characterization of the RNA degradation machinery</i>	- 81 -
4.2.9	<i>Conservation of general RNA degradation factors in human cells</i>	- 82 -
5	Identification and functional characterization of the novel mRNA modification N³-methylcytidine (m3C)	- 85 -
5.1	Results	- 85 -
5.1.1	<i>Experimental set-up for Trm140 PAR-CLIP and m3C-CLIP</i>	- 85 -
5.1.2	<i>Sequencing coverage of Trm140 PAR-CLIP and m3C-CLIP at established m3C modification sites</i>	- 87 -
5.1.3	<i>Reverse transcription signature of m3C modification</i>	- 89 -
5.1.4	<i>Identification of m3C modification in tRNA Arginine (CCU)</i>	- 90 -
5.1.5	<i>Mapping of m3C modification sites in mRNA</i>	- 93 -
5.1.6	<i>Translation efficiency change upon m3C depletion</i>	- 94 -
5.1.7	<i>Co-localization of Trm140 with RNA degradation factors</i>	- 97 -

5.1.8	<i>Depletion of m3C modification alters RNA synthesis rate and half-life</i>	- 98 -
5.1.9	<i>Experimental set-up for PAR-CLIP of METTL8</i>	- 100 -
5.1.10	<i>Mapping of the m3C modification in human mRNA</i>	- 102 -
5.2	Discussion and Outlook	- 104 -
5.2.1	<i>Identification of the m3C modification in tRNA Arginine</i>	- 104 -
5.2.2	<i>Effect of m3C loss on other tRNA modifications</i>	- 105 -
5.2.3	<i>Functional relevance of the m3C modification</i>	- 105 -
5.2.4	<i>Future directions for studies on the identification of m3C binding proteins in different eukaryotic cells</i>	- 108 -
5.2.5	<i>Functional implication of the m3C modification in the context of the epitranscriptome</i>	- 109 -
5.2.6	<i>Dissecting functional implications of the m3C modification in tRNA and mRNA in the future</i>	- 110 -
5.2.7	<i>Functional characterization of the m3C modification in human cells in the future</i>	- 111 -
5.2.8	<i>Technical challenges in the field of epitranscriptomics and future perspectives</i>	- 112 -
6	Bibliography	- 115 -
7	Appendix	- 141 -
7.1	Supplementary Information	- 141 -
7.1.1	<i>Figures</i>	- 141 -
7.1.2	<i>Tables</i>	- 152 -
7.1.3	<i>Spike-in sequences</i>	- 155 -
7.1.3.1	<i>Spike-in 2</i>	- 155 -
7.1.3.2	<i>Spike-in 4</i>	- 155 -
7.1.3.3	<i>Spike-in 5</i>	- 156 -
7.1.3.4	<i>Spike-in 8</i>	- 156 -
7.1.3.5	<i>Spike-in 9</i>	- 157 -
7.1.3.6	<i>Spike-in 12</i>	- 157 -
7.2	Abbreviations	- 159 -
7.3	List of Figures	- 165 -
7.4	List of Tables	- 167 -
8	Curriculum Vitae	- 169 -

1 Summary

The first part of this dissertation focused on the different RNA degradation pathways in the yeast *Saccharomyces cerevisiae* (*S. cerevisiae*). We used the method photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) to systematically generate transcriptome-wide protein binding profiles for 30 general RNA degradation factors. In-depth bioinformatic analysis and comparison with previously reported PAR-CLIP data provided factor enrichment on different RNA classes and the binding behavior for mRNAs and their associated antisense transcripts. The results also gave insights into how the various degradation complexes, and different subunits in these complexes, may be involved in the processing and degradation of different RNA species. Several conclusions were drawn with respect to degradation pathway selection, new functions for known factors were proposed, and several hypotheses have emerged that may be tested in the future. The generated datasets provide a rich resource for future studies of eukaryotic RNA degradation pathways, mechanisms, and the integration of mRNA metabolism.

The second part of this dissertation addressed the finding of a novel mRNA modification within the emerging field of epitranscriptomics. We performed PAR-CLIP experiments of the *N*³-methylcytidine (m3C) tRNA methyltransferase Trm140 to investigate its mRNA binding in the yeast *S. cerevisiae*. Using this approach in combination with a newly developed method applying anti-m3C antibodies for RNA immunoprecipitation (m3C-CLIP), we were able to call novel m3C modification sites. Next generation sequencing methods and biochemical verification assays showed the m3C modification in tRNA Arginine with anticodon CCU in addition to the established modification sites in tRNA Serine and Threonine in yeast. Our genome wide datasets also provided the first evidence that the m3C modification is present in yeast mRNAs and we mapped the modification on nucleotide resolution. Functional analyses suggested an effect of m3C mRNA modification on RNA degradation mediated by translation efficiency changes. Mapping of the m3C modification sites in human RNA revealed a similar localization compared to yeast over the entire

mRNA with higher occupancy towards the end of protein coding transcripts and within CG rich sequences.

2 Introduction

2.1 The central dogma of molecular biology

Genetic information in eukaryotic cells is stored in the nucleus as deoxyribonucleic acid (DNA). The replication and decoding of DNA, one of the most important processes in living cells, is also known as the “central dogma of molecular biology” (Crick, 1970). The conversion of genetic information into proteins is crucial for nearly all cellular functions. This conversion is sub-divided into two steps: transcription and translation. Transcription is the generation of ribonucleic acid (RNA) complementary to a given DNA sequence in the nucleus. During translation, the RNA serves as a template for protein synthesis in the cytoplasm.

2.1.1 Transcription

Transcription is performed by large multi-subunit complexes, which are called DNA dependent RNA polymerases (Pol). In eukaryotes, three distinct multi-subunit enzymes are necessary to transcribe all genes into various RNA molecules. Pol I transcribes the 18S, 5.8S and 23S ribosomal RNA (rRNA). Pol II transcribes many protein-coding messenger RNAs (mRNA) as well as a variety of non-coding RNAs (ncRNA), such as small nuclear RNAs (snRNA), long intergenic non-coding RNAs (lincRNA), and cryptic unstable transcripts (CUT). Pol III transcribes mostly transfer RNAs (tRNA), 5S rRNA and other small RNAs.

The synthesis of mRNAs by Pol II, which are later translated into proteins is completed in a cyclic process. This can be divided into three major steps: initiation, elongation, and termination (reviewed in Hantsche and Cramer, 2016).

The formation of the transcription competent pre-initiation complex (PIC) on the promoter sequence is the first crucial step of Pol II transcription initiation. The PIC contains, in addition to Pol II, the general transcription factors (GTFs) (Sainsbury et al., 2015). Conformational changes of Pol II regulate the

transition from initiation to elongation (Proudfoot et al., 2002). Rpb1, the largest subunit of Pol II, contains a C-terminal domain (CTD) consisting of multiple tyrosine-serine-proline-threonine-serine-proline-serine (YSPTSPS) heptapeptide repeats (26 in yeast, 52 in mammals) (Corden et al., 1985). These provide a scaffold for co-transcriptional RNA processing factors (Hirose and Manley, 2000; McCracken et al., 1997b). All amino acids of the CTD can be phosphorylated, except proline. The phosphorylation status of the CTD changes dynamically during the transcription cycle ('CTD code') (Buratowski, 2003; Hahn, 2004; Nechaev and Adelman, 2011; Proudfoot et al., 2002), which influences the maturation steps of pre-mRNAs (Fong and Bentley, 2001). A 7-methyl-guanosine cap is added to the 5' end of the nascent RNA when it reaches a length of 20-30 nucleotides (nt) (Rasmussen and Lis, 1993). This process is catalyzed by capping enzymes, which are recruited to the CTD by phosphorylation of the serine 5 (S5) residues (Cho et al., 1997; McCracken et al., 1997a). The 5' cap is important for RNA stability, nuclear export, and translation initiation (Proudfoot et al., 2002).

Transcription termination is induced by the recruitment of 3' end processing factors to Pol II via hyper-phosphorylation of the CTD at serine 2 (S2) residues (Ahn et al., 2004; McCracken et al., 1997b). 3' end processing and transcription termination are tightly coupled for eukaryotic protein-coding genes (Whitelaw and Proudfoot, 1986). After the polyadenylation (pA) signal has been transcribed, the nascent transcript gets cleaved directly after the pA site in order to release it from Pol II (Gilmartin and Nevins, 1989). A polyadenylated (polyA) tail is added to the newly generated 3' end of the mRNA by the poly-adenosine polymerase (PAP) (Colgan and Manley, 1997; Moore and Sharp, 1985). The polyA tail is required to export the mRNA to the cytoplasm for subsequent translation (Huang and Carmichael, 1996). The released Pol II becomes dephosphorylated and may enter a new transcription cycle. Efficient recycling and reinitiation of Pol II on the same template are facilitated by gene looping and the promoter binding GTFs forming a 'reinitiation platform' (Ansari and Hampsey, 2005; Dieci and Sentenac, 1996; Yudkovsky et al., 2000).

Thus, a full transcription cycle of Pol II produces an mRNA with a 5' cap and 3' polyA tail, which among other criteria renders the mRNA competent for subsequent translation.

2.1.2 Translation

During translation, the mRNA directs the synthesis of a polypeptide chain, which folds into a protein (reviewed in Green and Noller, 1997). This process is catalyzed by a large ribonucleoprotein particle, called the ribosome. The ribosome consists of two subunits, the small 40S and large 60S, together making up the 80S ribosome in eukaryotes. Both subunits harbor three binding sites for tRNA as shown in Figure 1: the acceptor (A') site for the incoming aminoacylated tRNA, the peptidyl (P) site for peptide bond formation, and the exit (E) site for holding the deacylated tRNA before leaving the ribosome. The small ribosomal subunit binds the mRNA and the complementary tRNA anticodon stem loop. The open reading frame of mRNA consists of codons, which are made up of three consecutive nucleotides. One codon of the mRNA matches the nucleotide triplet of the tRNA's anticodon loop. Translation is like transcription a cyclic process consisting of initiation, elongation, and termination (reviewed in Green and Noller, 1997).

For translation initiation, the scanning ribosome first has to localize a start codon AUG within the mRNA strand, from which translation can begin. For scanning, the small ribosomal subunit and many auxiliary initiation factors bind to the mRNA in the 5' untranslated region (5' UTR) between the 5' cap and the start codon. After start codon recognition, the large ribosomal subunit joins the small ribosomal subunit and the mRNA to form a translation competent 80S ribosome (Pestova et al., 2001).

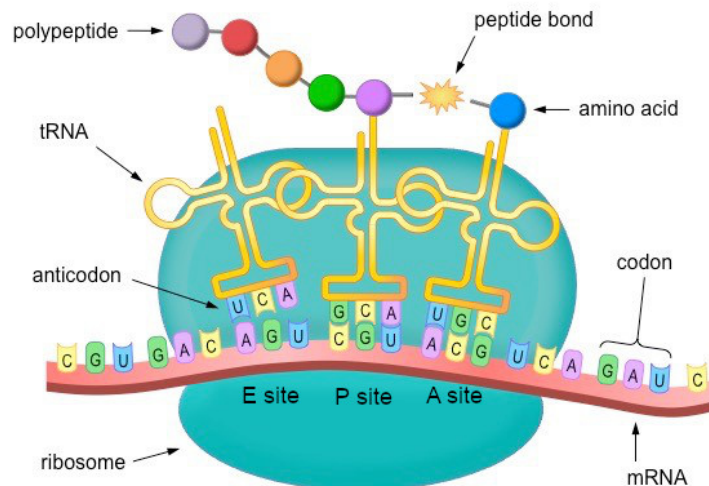


Figure 1: Translation by the ribosome.

The mRNA sequence shown in red serves as a template for protein production via the ribosome shown in green. The codons of the mRNA are complementary to the anticodon of the tRNA loaded with one specific amino acid. During translation the peptide bond formation between the amino acids takes place between the A' site and P site of the ribosome thereby forming the polypeptide sequence. (modified from <http://ib.bioninja.com.au/standard-level/topic-2-molecular-biology/27-dna-replication-transcri/translation.html>, 18.02.2019)

For translation elongation, the small ribosomal subunit, together with the mRNA identifies the correct anticodon stem loop of the tRNA (Schmeing et al., 2011, 2009). Thereby, it contributes to the fidelity of translation via monitoring the decoding process of highly regulated RNA-RNA interactions of codon and anticodon base pairing (Demeshkina et al., 2010). The large ribosomal subunit can bind the acceptor arm of the tRNA and then peptide bond formation between the nascent polypeptide chain in the P site and the aminoacylated tRNA in the A' site is catalyzed in the peptidyl transferase center. Both ribosomal subunits and auxiliary elongation factors contribute to translocation, which is the precise movement of tRNA and mRNA through the ribosome one codon at a time (Gao et al., 2009).

Translation termination is the process after stop codon (UAA, UGA, and UAG) recognition in the A' site and separation of the polypeptide chain from the ribosome. Release of the protein requires two release factors binding to the A' site in complex with GTP and subsequent GTP hydrolysis to trigger hydrolysis of the polypeptidyl-tRNA. This results in the free protein product with the

ribosome still bound to mRNA, the deacylated tRNA in the P site, and the release factors in the A' site. For further rounds of translation, the ribosome needs to be recycled by disassembling of termination factors and mRNA release (reviewed in Hellen, 2018; Jackson et al., 2012).

In order to get a genome-wide picture of translating ribosomes and an estimation of translation rates and translation efficiency, a method called ribosome-profiling has been developed (Ingolia et al., 2009). By sequencing ribosome protected mRNA fragments, this method allowed monitoring of the ribosome position on the mRNA with single-codon resolution. Comparing translation efficiency with mRNA levels showed its substantial contribution to the dynamic nature of gene expression, which was invisible to measurements of mRNA levels alone. Thus, regulation of transcription and translation contributes to variable production of mRNAs and polypeptide chains.

2.2 The RNA degradation pathways

An active RNA degradation mechanism is implied by comparing RNA production with RNA steady-state levels, which showed more RNA transcription than RNA accumulation in the cell (reviewed in Houseley and Tollervey, 2009). In general, in order to maintain RNA homeostasis each RNA molecule is degraded after a certain lifespan, which can be long for rRNAs or very short for introns or regulatory elements and is precisely controlled for most mRNA transcripts (Miller et al., 2011; Turowski and Tollervey, 2015). Defective RNA molecules in regard to processing, folding, or assembling with proteins are identified and rapidly degraded by the surveillance machinery in order to prevent aberrant protein production (reviewed in Houseley and Tollervey, 2009). Since RNA decay is omnipresent in all cells, it implies a highly accurate controlling mechanism to recognize target RNAs. In addition, upon environmental stimuli such as temperature change or nutrient deprivation, the cell adjusts gene expression by regulating RNA abundance either by changing transcription output or degradation rate (Gasch et al., 2000; Gasch and Werner-Washburne, 2002; Jelinsky and Samson, 1999). In principal, RNA degradation can occur from both ends of the transcripts (Parker, 2012), however, the characteristics of a transcript that determine the selection of one or the other decay pathway is a highly studied question in the field.

2.2.1 Nuclear surveillance of aberrant non-coding RNAs

Transcription initiation by Pol II exhibits a poor directionality, which leads to pervasive transcription (Core et al., 2008; Neil et al., 2009; Seila et al., 2008; Xu et al., 2009). In yeast, the phenomenon of bi-directional transcription has been observed by the formation of two adjacent PICs within nucleosome-depleted regions (NDRs) (Murray et al., 2012; Rhee and Pugh, 2012). Two mechanisms have been shown to restrict the amount of pervasive transcription in eukaryotes: First, gene looping and preferred formation of PICs in one orientation can influence transcription directionality and thereby restrict initiation of divergent ncRNA transcription (Rhee and Pugh, 2012; Tan-Wong

et al., 2012). Second, nuclear 3'→5' RNA degradation by the exosome and its catalytically active subunit Rrp6 removes excessive transcription products instantly after targeting of the ncRNAs by the Nrd1-Nab3-Sen1 complex (Neil et al., 2009; Schulz et al., 2013; Steinmetz et al., 2001; Xu et al., 2009). The Nrd1-Nab3-Sen1 complex also interacts with the cap binding complex as well as the TRAMP complex (see section 2.2.2.3) (Vasiljeva and Buratowski, 2006). In addition to degradation from the 3' end, the exonuclease Xrn1 degrades ncRNAs from the 5' end (van Dijk et al., 2011).

2.2.2 Cytoplasmic RNA degradation

2.2.2.1 Deadenylation

RNA degradation is initiated by shortening of the polyA tail by two major deadenylation machineries: the multi-subunit Ccr4-Not complex (Ccr4, Not1, Pop2, Caf40) and the Pan2-Pan3 complex as shown in Figure 2A (Wolf and Passmore, 2014). Deadenylation is one of the rate-limiting steps for mRNA decay, thus, the enzymatic activities of the two mRNA deadenylase complexes make up an important target for the control of mRNA decay (Wolf and Passmore, 2014). Previous investigations regarding decay mechanisms of specific mRNAs showed that selective recruitment of deadenylating enzymes to the substrate mRNA is involved (Finoux and Seraphin, 2006; Goldstrohm et al., 2006; Semotok et al., 2005). Nevertheless, the exact mechanism leading to mRNA deadenylation has not been fully elucidated. A direct link between translation termination and mRNA decay has been revealed by several studies (reviewed in Huch and Nissan, 2014), in particular deadenylation, which is dependent on the poly-adenylate binding protein 1 (Pab1) and Ccr4 (Webster et al., 2018). It has also been shown that translation elongation and mRNA decay are coupled, which depends on the net effect of different cognate tRNA expression termed “codon optimality” (Presnyak et al., 2015). mRNA degradation is enhanced in the presence of rare codons, also non-optimal codons, which are encoded by tRNAs of relatively low abundance (Caponigro et al., 1993).

Upon translational termination, a stepwise model was proposed for deadenylation by first reducing the average polyA tail length of 90 nt to 50 nt via the Pan2-Pan3 complex, which is then further shortened by the Ccr4-Not complex (Beilharz and Preiss, 2007; Brown and Sachs, 1998; Tucker et al., 2001). The mRNA with an oligo A tail of 10-12 nt can then become a substrate for decapping (Chowdhury et al., 2007; Tharun and Parker, 2001), or it is subjected to exosome catalyzed decay (Bonneau et al., 2009).

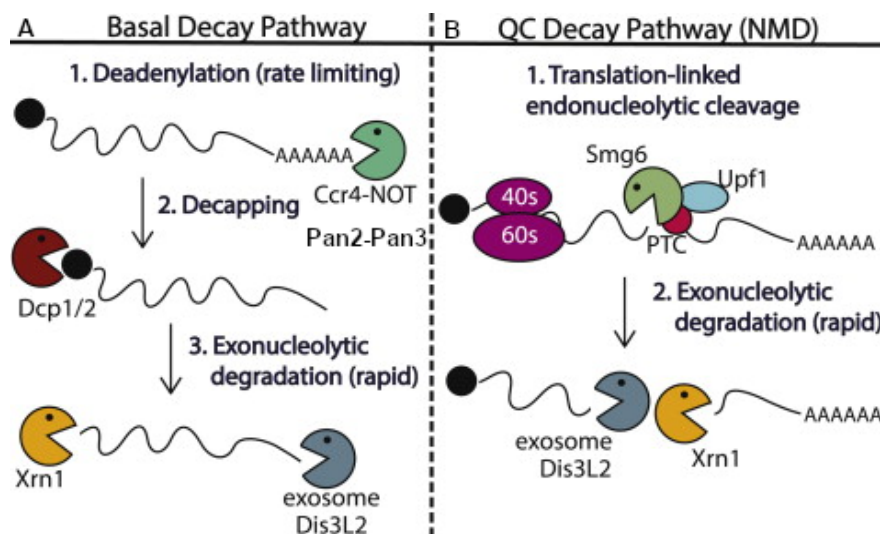


Figure 2: Schematic overview of RNA degradation

A) RNA decay starts with the rate-limiting step of deadenylation via Pan2-Pan3 and Ccr4-Not complexes, followed by decapping via Dcp1 and Dcp2 and exonucleolytic degradation of the transcript by Xrn1 from the 5' end and the exosome from the 3' end. B) Quality control decay pathways in particular nonsense mediated decay (NMD) factors such as the Upf1 and Smg6 recognize aberrant mRNAs during translation, mostly in the presence of premature termination codons (PTC), and induce endonucleolytic cleavage, whereupon the fragments are degraded by exonucleases from the 5' end (Xrn1) or 3' end (exosome) (adapted from Abernathy and Glaunsinger, 2015).

2.2.2.2 Decapping and exonucleolytic 5'→3' RNA degradation

The second step in mRNA degradation after deadenylation is removal of the 5' cap, which protects the RNA from 5'→3' degradation by the exonuclease Xrn1 (Stevens and Poole, 1995). In order to remove the cap, a three-step model is proposed: First, the 5' cap structure needs to be unprotected by loss of the

cap-binding complex. Second, recruitment of the decapping enzyme to the mRNA needs to take place, which seems to be regulated by the assembly of a larger decapping complex. Third, catalysis by Dcp2 occurs together with Dcp1, inducing rapid 5'→3' decay of the mRNA by the exonuclease Xrn1 (Hsu et al., 2017; Muhlrud and Parker, 1994). Decapping is highly regulated by decapping enhancers like Dhh1, Edc2 and Edc3. Different mechanisms may trigger decapping by binding RNA and providing a platform for assembly of the decapping machinery, by interfering with translation initiation factors or by stimulating Dcp2 catalytic function. Assembly of the decapping machinery occurs mostly after shortening of the polyA tail, which prevents binding of the protective Pab1, opens the mRNA closed-loop structure, and allows decapping complex formation on the 3' end of deadenylated mRNA (Caponigro and Parker, 1995; Gallie, 1991; Morrissey et al., 1999). Opening of the mRNA closed-loop structure disrupts the proximity of the 5' and 3' ends of the mRNA. This prevents complex formation of translation initiation factors binding to the 5' cap and Pab1 at the 3' end, thereby contributing to mRNA expression regulation (Gallie, 1991; Wells et al., 1998).

2.2.2.3 Exosomal 3'→5' RNA degradation

Upon deadenylation, 3'→5' degradation of the mRNA is performed by the exosome and many auxiliary factors (Anderson and Parker, 1998). The exosome is a multi-subunit complex consisting of 10 main subunits. Three small RNA-binding proteins and six members of the RNase PH protein family form a ring structure similar to bacterial PNPase as shown in Figure 3 (Liu et al., 2006). The Rrp44/Dis3 protein is the catalytically active subunit possessing both an exonuclease and endonuclease domain (Lebreton and Seraphin, 2008; Schaeffer et al., 2009). The exosome complex takes part in many nuclear RNA processing and degradation processes in addition to its functions within the cytoplasm (reviewed in Lykke-Andersen et al., 2009). In order to fulfill its functions in the nucleus, the exosome is additionally bound by the 3'→5' exonuclease Rrp6 as well as RNA binding proteins Rrp47 and Mpp6 (Milligan et al., 2008; Mitchell et al., 2003; Synowsky et al., 2009).

One of the various co-factors for nuclear exosomal RNA maturation and degradation in eukaryotes is the TRAMP polyadenylation complex (reviewed in Houseley and Tollervey, 2009). This complex consists of two sub-complexes: TRAMP4 (Trf4, Air2 and Mtr4) and TRAMP5 (Trf5, Air1 and Mtr4). The TRAMP complexes harbor a poly-adenosine polymerase (Trf4 or Trf5), a zinc-knuckle putative RNA-binding protein (Air1 or Air2), and an RNA helicase (Mtr4) (Falk et al., 2014). Defective nuclear RNAs can be tagged with a short polyA tail and bound proteins can be detached by the TRAMP complex. This renders the defective RNA a more suitable target for exosomal degradation by the sterically restricted exosome core (Schmidt and Butler, 2013).

The Ski complex is required for cytoplasmic 3'→5' exosomal degradation to feed the RNA into the exosome. The Ski7 protein is stably associated with the cytoplasmic exosome by the Csl4 adapter protein (van Hoof et al., 2002). The Ski2, Ski3, and Ski8 proteins assemble in a discrete protein complex (Brown et al., 2000; Wang et al., 2005) interacting with the Ski7 subunit (Araki et al., 2001; Wang et al., 2005). This interaction is required for 3'→5' degradation of mRNAs by the exosome. The Ski2 subunit is an ATPase, which belongs to the RNA helicase family. It provides energy by ATP hydrolysis to unwind RNA secondary structures and dissociate attached proteins to pass the RNA to the exosome (Johnson and Jackson, 2013).

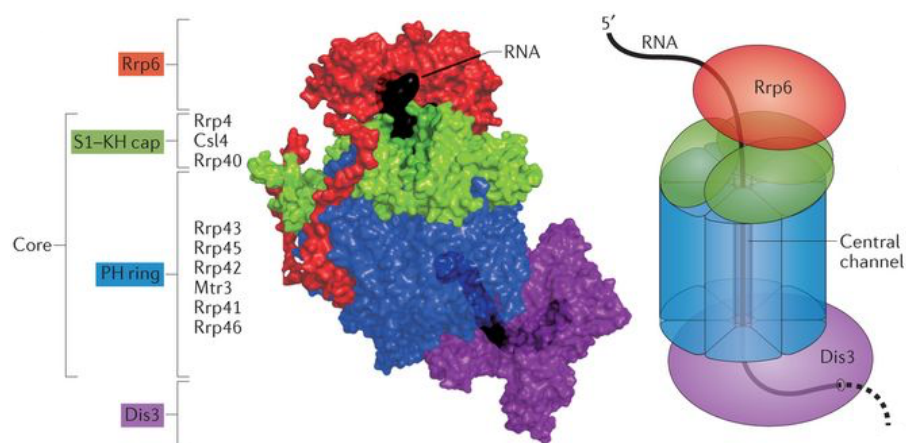


Figure 3: Model of the exosome complex.

Schematic overview of the exosome complex assembly including the nuclear Rrp6 in red and cytoplasmic Dis3 (Rrp44) subunit in violet. The exosome core consists of the cap (Rrp4, Csl4 and Rrp40) depicted in green and the ring domain (Rrp43, Rrp45, Rrp42, Mtr3, Rrp41 and Rrp46) in blue forming a central channel (adapted from Kilchert et al., 2016).

2.2.2.4 Quality control mechanism for defective mRNA

Aberrant mRNAs need to be determined from the pool of normal mRNAs by adapter proteins. This can be accomplished by interactions of specific adapter proteins with the translation machinery. The aberrant mRNA can then be subjected to one of the RNA decay pathways (Doma and Parker, 2007). mRNAs with aberrant translation termination due to premature translation termination codons are degraded by an mRNA quality control system called nonsense mediated decay (NMD) as depicted in Figure 2B (Losson and Lacroute, 1979). Substrates for NMD are distinguished by the Upf1 protein interacting with the translation termination complex. Subsequently the proteins Upf2 and Upf3 bind to Upf1 thereby enhancing its helicase activity (reviewed in Baker and Parker, 2004; Chakrabarti et al., 2011). Binding of the NMD machinery to an aberrant mRNA has several effects on the fate of the transcript: the polyA tail of the mRNA is subjected to enhanced deadenylation (Cao and Parker, 2003; Mitchell and Tollervey, 2003; Muhlrud and Parker, 1994), rapid deadenylation-independent decapping of the 5' end (Muhlrud and Parker, 1994), slightly accelerated rates of 3'→5' degradation by the exosome complex after deadenylation (Cao and Parker, 2003; Mitchell and Tollervey, 2003), and inhibition of translation (Muhlrud and Parker, 1999). In addition to its role in degrading aberrant mRNAs, the NMD machinery is involved in the regulation of many biological processes such as stress responses, cell survival and differentiation by controlling the steady state level of normal mRNAs (Goetz and Wilkinson, 2017; Karam et al., 2015; T. Li et al., 2015; Lou et al., 2016; Lykke-Andersen and Jensen, 2015; Nelson et al., 2016). It has been shown that the expression regulation of normal mRNAs is achieved by fine tuning of the NMD activity (Huang and Wilkinson, 2012; Karam et al., 2013).

2.3 The emerging field of epitranscriptomics

2.3.1 Biogenesis of eukaryotic tRNAs and their function in translation

In *S. cerevisiae*, Pol III transcribes a total of 275 tRNA genes (Chan and Lowe, 2009). The pre-tRNA transcripts contain a 5' leader sequence, a 3' trailer sequence and can harbor an intronic sequence, which need to be removed during biogenesis. After removal of the 5' leader sequence by RNase P, RNase Z can remove the 3' trailer sequence at nucleotide 73 and a nucleotidyl-transferase is able to add a CCA sequence to the 3' end of the tRNA (Aebi et al., 1990). Correctly folded and end-processed tRNAs are exported from the nucleus into the cytoplasm using the export factor Los1 (Arts et al., 1998; Lund and Dahlberg, 1998; Sarkar and Hopper, 1998). The conserved family of Sen proteins removes intronic sequences from 61 tRNAs in a process called tRNA splicing in the cytoplasm (Melton et al., 1980; Yoshihisa et al., 2003). The final step of tRNA maturation is aminoacylation by aminoacyl-synthetases (Grosshans et al., 2000; Lund and Dahlberg, 1998; Steiner-Mosonyi and Mangroo, 2004). The correctly loaded tRNAs can then serve as adapter molecules during translation of the mRNA into the amino acid sequence of the protein catalyzed by the ribosome.

2.3.2 tRNA modifications and their function

In ncRNAs like rRNA, tRNA and snRNA, more than 100 chemical modifications have been characterized. tRNAs can be highly modified on numerous nucleotides as shown in Figure 4, which is crucial for their specific function in translation (Phizicky and Hopper, 2010). The correct folding of tRNAs into the clover leaf structure (Figure 4) including the acceptor stem, the D-loop, the TΨC-loop, the anticodon-loop, and the variable loop is ensured by the different RNA modifications installed in the nucleus and cytoplasm depending on the specific modification (reviewed in Hopper, 2013). The tRNA's secondary structure is folded into an inverted L-shaped tertiary structure containing the CCA-acceptor stem at the top and the anticodon-loop at the bottom (Shi and Moore, 2000). The most variably and extensively modified

nucleotides are in the anticodon loop (ACL) at the wobble position (34) and in the extended anticodon (37) of tRNAs. These modifications may play important roles in decoding precision, efficiency and reading frame maintenance during translation (Björk et al., 2001; Chen C, Huang B, Eliasson M, Ryde'n P, 2011; Esberg et al., 2006; Gerber et al., 1998; Johansson et al., 2008; Murphy and Ramakrishnan, 2004; Urbonavičius et al., 2001; Waas et al., 2007; Weixlbaumer et al., 2007; Yacoubi et al., 2011). Effects on tRNA structure stability are mainly influenced by modifications in the body of the tRNA leading to more rigid or loose parts (reviewed in Lorenz et al., 2017). Incorrectly modified tRNAs or the loss of tRNA modifying enzymes have been implicated in a number of human diseases such as neurological and metabolic diseases as well as several kinds of cancer (reviewed in Torres et al., 2014). A perturbed protein synthesis in response to incorrect codon recognition is considered to be the molecular reason for those diseases (Torres et al., 2014).

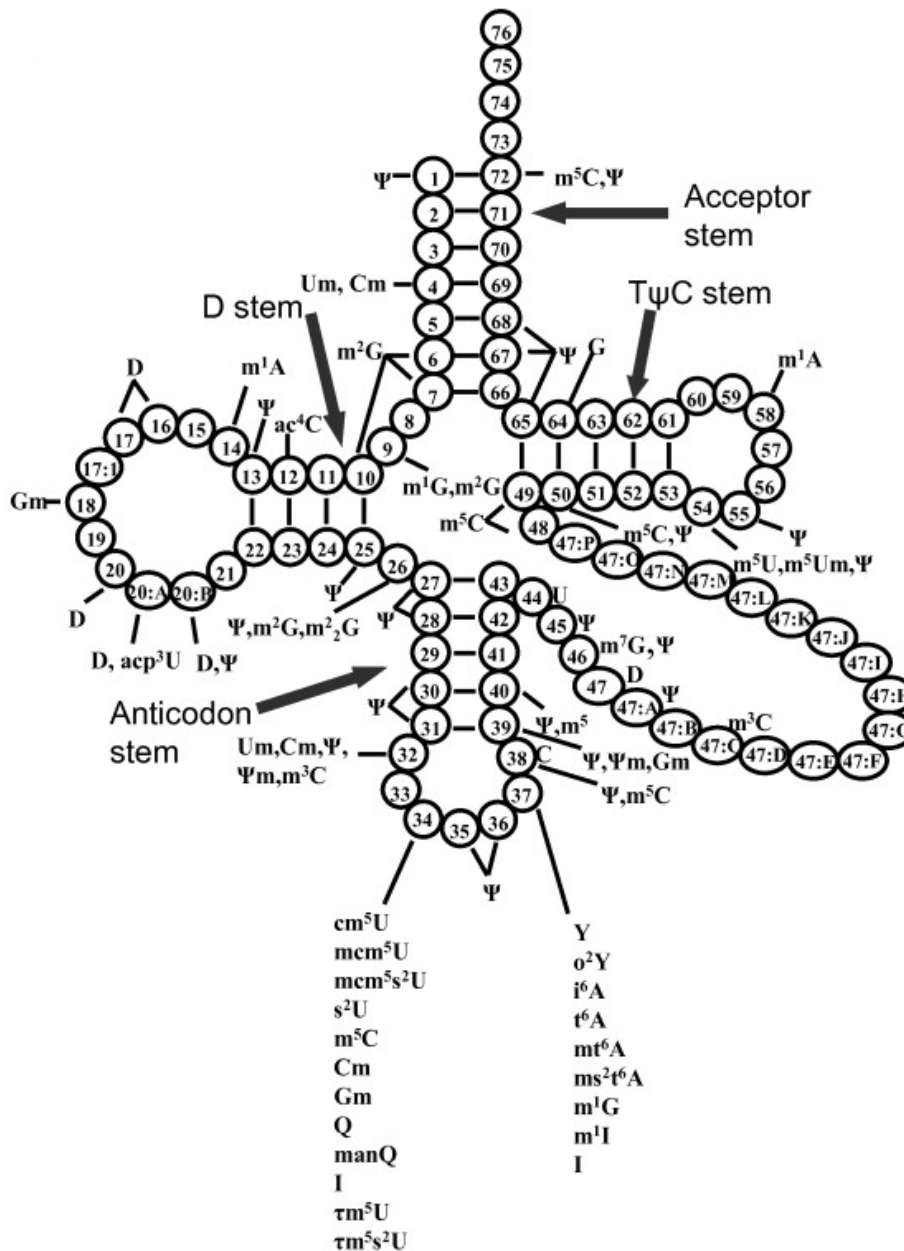


Figure 4: Modified nucleosides in the clover-leaf structure of the eukaryotic tRNA. Each circle represents a nucleotide, numbered from 5'- to 3'- end. The clover-leaf structure is divided into the acceptor stem, the D stem, the anticodon stem, the variable stem and the TΨC stem loop. Modified nucleosides found at different positions are shown (adapted from Chen et al., 2010).

2.3.3 mRNA modifications

In the last ten years a lot of effort has been made to detect RNA modifications in mRNA leading to the emerging field of epitranscriptomics

(Saletore et al., 2012). The epitranscriptome is defined as a functional change in the mRNA that does not involve changes to the ribonucleotide sequence. The epitranscriptome is analogous to the epigenome, which describes the chemical modifications on DNA and histones. The first modifications described in mRNA were *N*⁶-methyladenosine (m6A) and 5-methylcytidine (m5C) in the 1970s (Dubin and Taylor, 1975; Schibler et al., 1977). The currently known base modifications in protein-coding transcripts are m6A, m5C, inosine (I), pseudouridine (Ψ), *N*¹-methyladenosine (m1A), 5-hydroxymethylcytidine (hm5C), and *N*⁴-acetylcytidine (ac4C) (see Figure 5). Different RNA editing events within the mRNA can change the encoded protein sequence and thereby eventually generate premature stop codons or alter the differential expression levels of micro RNAs (Chawla and Sokol, 2014; Powell et al., 1987). The localization of the different mRNA modifications in eukaryotic protein-coding transcripts is shown in Figure 5. Epitranscriptomic modifications can also influence mRNA stability, structure, splicing, translation and degradation (reviewed in Frye et al., 2016).

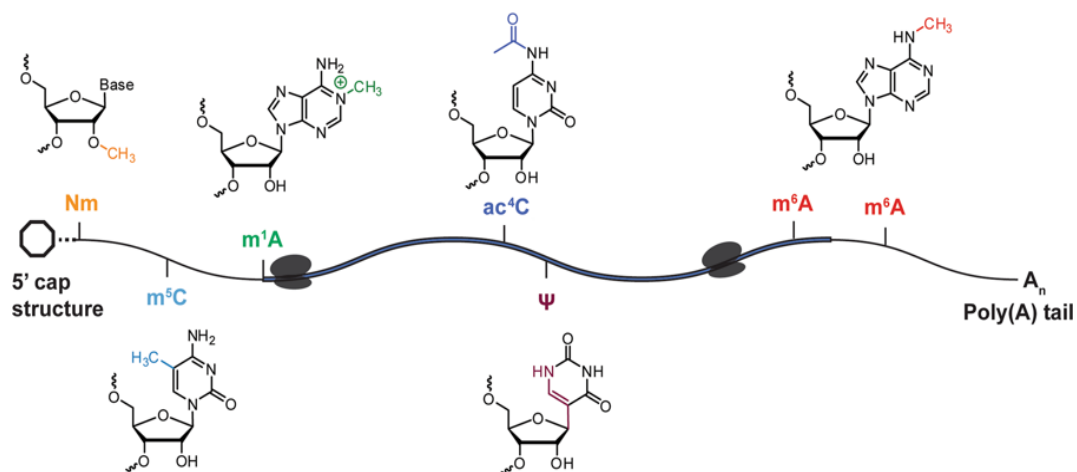


Figure 5: Schematic representation of known chemical modifications mapped in eukaryotic mRNA transcripts.

The protein coding region is highlighted in bold. The to date known base modifications within the mRNA are *N*⁶-methyladenosine (m6A), 5-methylcytidine (m5C), inosine (I), pseudouridine (Ψ), *N*¹-methyladenosine (m1A), and *N*⁴-acetylcytidine (ac4C). Structure of the nucleoside including the chemical modification in color is shown at the position within the transcript, where the modification has been localized. In addition to the base modifications, also the 2'-O-methylation of ribose sugars (Nm) is also shown close to the 5' cap (adapted from He and He, 2019).

2.3.3.1 N^6 -methyladenosine modification

Development of next-generation sequencing (NGS) techniques enabled a more accurate analysis of the m6A modification and revealed its extent over the transcriptome after its identification in the 1970s (Dubin and Taylor, 1975). The m6A modification is generally found in mRNAs of nearly all eukaryotes from yeast to fruit fly and mouse up to human (Dominissini et al., 2012; Hongay and Orr-weaver, 2011; Schwartz et al., 2013). Antibodies were used to specifically detect the m6A modification in cellular mRNAs (Chen et al., 2015; Dominissini et al., 2012) and by using cross-linking methods coupled to specific mutational signatures, the modification was mapped at single-nucleotide resolution (Linder et al., 2015). These studies revealed over 12,000 possible m6A modification sites in mRNAs and ncRNAs of more than 7,000 human transcripts. However, in yeast the presence of the m6A modifications is reduced to about 1,000 mRNAs and limited to meiosis suggesting a function in translation of certain mRNAs (Schwartz et al., 2013). Mapping of the modification on the mRNA transcript has revealed its localization in humans around stop-codons, long internal exons and in the 3' UTRs (Chen et al., 2015; Dominissini et al., 2012; Saletore et al., 2012). The consensus motif RRACH, which was already proposed in the 1970s, was confirmed by the highly enriched GGACU motif within the sequencing data (Chen et al., 2015; Dominissini et al., 2012; Saletore et al., 2012; Schibler et al., 1977).

The m6A modification is installed by a methylation complex consisting of the methyltransferase METTL3, the putative methyltransferase METTL14, and the regulatory proteins WTAP and KIAA1429 (Liu and Jia, 2014; Ping et al., 2014; Schwartz et al., 2014b). Other m6A methyltransferases seem to be involved in introducing such modifications, as knockdown of the METTL3/METTL14 complex does not completely abolish m6A modification. Furthermore, the known m6A modification sites just overlap partially with the binding sites of METTL3/METTL14 on cellular RNAs indicating the presence of other m6A methyltransferases (Chen et al., 2015; Liu and Jia, 2014; Schwartz et al., 2014b). Indeed, METTL16 was identified to modify pre-mRNAs and various ncRNAs with m6A (Pendleton et al., 2017; Warda et al., 2017). Interestingly, METTL16-dependent m6A modification sites do not have the

RRACH sequence motif and they occur within introns and at intron-exon boundaries (Pendleton et al., 2017; Warda et al., 2017).

Functional analyses of the m6A modification have revealed that the modification may change the secondary structure of the RNA by acting as a molecular switch. This can lead to the display of RNA binding structures or motifs for specific RNA binding proteins (Liu et al., 2015). On a cellular level the m6A modification is suggested to be involved in stress response, cap-independent translation and regulation of the circadian rhythm of cells (Engel et al., 2018; Fustin et al., 2013; Meyer et al., 2015). In mouse, the m6A modification was observed to keep the omnipotence of the cells thereby regulating embryonic stem cells and was also shown to be involved in the cell cycle regulation of human cells (Dominissini et al., 2012; Y. Wang et al., 2014). Recently, a study revealed a crosstalk between histone modifications and m6A modifications to regulate gene expression (Huang et al., 2019). Hereby, METTL14 recognizes the trimethylation of histone H3 at Lys 36, which is an established marker for transcription elongation, and facilitates binding of the methyltransferase complex in the vicinity of Pol II to deposit the m6A modification co-transcriptionally (Huang et al., 2019; Kizer et al., 2005).

The m6A modification has been identified to be reversible via the ALKBH5/FTO protein, which erases the mark, thereby allowing for the dynamic nature of the m6A modification associated with different functions in mice and humans (Jia et al., 2011; Zheng et al., 2013). However, not only writer and eraser proteins for m6A modification have been identified, but also reader proteins. Those reader proteins specifically bind to the m6A mark thereby influencing the fate of the transcript. The first identified m6A binding proteins share a specific domain, the YT521-B homology (YTH) domain (Zhang et al., 2010). In humans, five YTH domain containing proteins are known, namely YTHDF1, YTHDF2, YTHDF3, YTHDC1, and YTHDC2. These proteins are associated with different functions in cellular processes upon m6A recognition in mRNA such as splicing, stability and translation (Kretschmer et al., 2018; Theler et al., 2014; Wang et al., 2015; X. Wang et al., 2014; Xiao et al., 2016). Increased translation of specific m6A containing transcripts has been observed to control anti-tumor immunity in dendritic cells (Han et al., 2019). In *Drosophila*,

the m6A reader protein YT521B has been shown to control neurogenesis and sex determination (Lence et al., 2016).

2.3.3.2 5-methylcytidine modification

In addition to its occurrence in tRNA, rRNA, and the well characterized function in transcription regulation on DNA, the m5C modification was also observed in human and archaeal mRNA (Edelheit et al., 2013; Squires et al., 2012). Bisulfite treatment coupled with NGS enabled mapping of m5C and its oxidation products 5-hydroxymethylcytidine (hm5C) and 5-formylcytidine (f5C) on a transcriptome-wide level (Booth et al., 2014; Edelheit et al., 2013; Lee and Kim, 2016). In ncRNAs and mRNAs of HeLa cells over 10,000 m5C modification sites were discovered (Squires et al., 2012). Global positional analysis of the m5C modification revealed an overall distribution in mRNAs with an increased occupancy in 5' and 3' UTRs (Squires et al., 2012). This localization of the modifications implies a function in protein translation regulation (Squires et al., 2012). Two additional experiments in fruit flies support this hypothesis (Delatte et al., 2016). The m5C content can also be regulated in human cells by oxidation to hm5C and f5C, indicating a similar mechanism being present in mammals (Huber et al., 2015).

2.3.3.3 *N*¹-methyladenosine modification

Specific chemical properties of the m1A modification in mRNA allowed its discovery by using an antibody-based approach to enrich transcripts containing an m1A modification. These were later on sequenced to map the modified nucleoside within the transcripts (Dominissini et al., 2016; Li et al., 2016a; Li and Xiong, 2017; Safra et al., 2017b). The number of reported m1A marks in mRNAs varies between each study, ranging from very few (Safra et al., 2017b), to around 800 (Li et al., 2016a), and up to over 4,000 (Dominissini et al., 2016). m1A is reported to occur at low stoichiometry and in invariable tRNA TΨC-loop structures (Safra et al., 2017b). Whereas another study

reported the average methylation level of an m1A containing transcript of approximately 20% within one mRNA (Dominissini et al., 2016). This study also observed a positional distribution of the m1A marks in the 5' UTR and the 5' end of mRNAs (Dominissini et al., 2016). In addition to its presence in the 5' UTR, others reported occurrence of the m1A modification also in the 3' UTR (Li et al., 2016a). The last two studies indicated a functional role of the m1A mark in translation regulation as they observed changed methylation patterns in response to different physiological conditions and external stress (Dominissini et al., 2016; Li et al., 2016a). Furthermore, the alpha-ketoglutarate dependent dioxygenase ALKBH3 was identified to erase the m1A modification, thus, making it a reversible modification with a dynamic function (Li et al., 2016a).

2.3.3.4 Pseudouridine modification

Pseudouridine modifications have also been mapped by transcriptome-wide sequencing. A chemical labeling approach by using the reactivity of the molecule N^3 -[*N*-cyclohexyl-*N'*- β -(4-methylmorpholinium)ethylcarbodiimide (CMC) and pseudouridine introduces a bulky modification (CMC- Ψ), which causes a reverse transcription stop and therefore allows mapping of the modified nucleotide (Bakin and Ofengand, 1993; Zaringhalam and Papavasiliou, 2016). By using this method, different studies reported the modification to be present in yeast and human mRNAs and ncRNAs (Carlile et al., 2014; X. Li et al., 2015; Lovejoy et al., 2014; Safra et al., 2017a; Schwartz et al., 2014a). Conserved pseudouridine synthases of the Pus family and TRUB1 catalyze isomerization from uridine to pseudouridine with a conversion rate of 0.2 – 0.6% identified via mass spectrometry analysis (Carlile et al., 2014; Lovejoy et al., 2014; Safra et al., 2017a; Schwartz et al., 2014a). The high abundance of the pseudouridine mark enables a high regulatory potential, which was observed by altered modification patterns and abundance caused by starvation induced stress response in yeast and human cells (Carlile et al., 2014; X. Li et al., 2015; Schwartz et al., 2014a). Pseudouridylation of nonsense stop codons UAA, UAG or UGA can change their effect and makes it impossible for the ribosome to recognize the stop codon, hence translation continues with

incorporation of the corresponding tRNA (Hoernes et al., 2016; Karijolich and Yu, 2011). An alternative way to regulate translation was shown by *in vitro* and *in vivo* assays in which the pseudouridylated mRNA had an increased translation capacity as well as higher biological stability in mice and humans (Kariko et al., 2008). Altered pseudouridylation levels in humans have been shown to cause intellectual disability and microcephaly (Shaheen et al., 2019).

2.3.3.5 *N*⁴-acetylcytidine modification

Very recently, the ac4C modification has been identified in yeast and has also been mapped in human mRNAs (Arango et al., 2018; Tardu et al., 2018). The ac4C modification is installed by NAT10 in a broad range of human mRNAs in addition to 18S rRNA, Serine and Leucine tRNAs (Arango et al., 2018; Chimnarongk et al., 2009; Ito et al., 2014; Sharma et al., 2015). Over 4,000 ac4C marks were identified within human mRNAs with usually one or two ac4C modified nucleosides per modified transcript. Positional analysis of ac4C revealed a uniform distribution all over the mRNAs with a general bias towards the 5' UTR close to translation start sites (Arango et al., 2018). Depletion of NAT10 caused downregulated expression levels of ac4C containing mRNA indicating a positive effect of the ac4C modification on translation efficiency (Arango et al., 2018). Additionally, the mRNA stability of ac4C containing transcripts was significantly prolonged compared to mRNAs in NAT10 deficient cells (Arango et al., 2018). Moreover, mRNA decay and translation are intricately linked. Thus, decreased mRNA stability results in a reduced translation, what decreases mRNA stability reciprocally (Hanson and Coller, 2018). In addition to the elevated translation efficiency of ac4C containing mRNAs, ac4C in the wobble position of tRNAs also stimulates translation (Arango et al., 2018).

2.3.4 N^3 -methylcytidine modification

The presence of a N^3 -methylcytidine (m3C) (Figure 6) modification was observed in tRNAs Serine and Threonine at position C32 in the model organism *S. cerevisiae* (D'Silva et al., 2011; Noma et al., 2011) and humans (Clark et al., 2016). The variable arm of tRNA Serine and C32 in tRNA Arginine with anticodon CCU and UCU also contain m3C in higher eukaryotes (Arimbasseri et al., 2015; Capone et al., 1985; Cribbs et al., 1987; Keith, 1984; Partial et al., 1971).

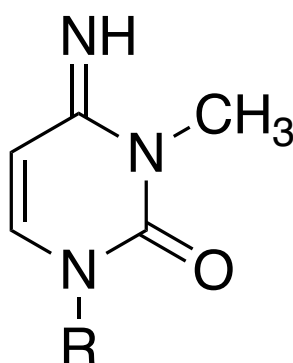


Figure 6: Structure of the N^3 -methylcytosine (m3C) modification. The additional methyl group is added to the N-3 of cytidine by the methyltransferase Trm140 in yeast and METTL2, 6 or 8 in humans and S-adenosylmethionine (SAM).

2.3.4.1 N^3 -methylcytidine methyltransferase in *S. cerevisiae*

The methyltransferase for m3C is Trm140 in *S. cerevisiae* (D'Silva et al., 2011; Noma et al., 2011). Trm140 was first identified as an actin binding protein (Abp140) (Asakura et al., 1998) with its N-terminal part binding to actin filaments and thereby localizing to actin patches and cables. The C-terminal part is responsible for m3C modification and contains the S-adenosylmethionine (SAM) binding motif. Trm140 is encoded on two separate open reading frames (ORFs) that are translated into one protein by a +1 frame shift (Farabaugh et al., 2006). It seems likely that the m3C modification on position C32 of tRNA is important for cells, as residue 32 interacts with residue 38 of the anticodon loop to maintain its structure (Auffinger and Westhof, 2001,

1999). A Trm140 knockout (KO) does not have a significant phenotype, whereas the Trm140 and Trm1 double knockout strain, which lacks m³C and N²-N²-dimethylguanine (m²₂G), is sensitive to cycloheximide, indicating impaired translocation on the ribosome (D'Silva et al., 2011). N⁶-isopentenyl adenosine (i6A37), N⁶-threonylcarbamoyl adenosine (t6A37) or their derivatives at position 37 are necessary for m³C modification at position 32 (Han et al., 2017). However, the t6A37 is not a sufficient mark for m³C modification as shown for several tRNAs (Met, Ile, Asn Lys) (Thiaville et al., 2016). It has been shown that Trm140 recognizes its tRNA substrates in two different ways. It can either recognize the sequence G35-U36-t6A37 of the anticodon loop of tRNA Threonine or t6A37 and i6A37 of tRNA Serine, since the anticodons of tRNA Serine do not have G35-U36 and no other nucleotides in common. Additionally, seryl-tRNA synthetase and the distinctive tRNA Serine variable loop are important to achieve Trm140 specificity on tRNA Serine (Han et al., 2017).

2.3.4.2 N³-methylcytidine methyltransferases in higher eukaryotes

In mammals, several methyltransferase-like (METTL) proteins have been well characterized and a high sequence similarity of the human and mouse METTL2, METTL6 and METTL8 to the yeast Trm140 was shown (Xu et al., 2017). METTL2 and METTL6 contribute to specific tRNA modification, whereas METTL8 forms m³C in mRNAs (Xu et al., 2017). A knockout of METTL8 in two different human cell lines drastically reduced the m³C level in mRNAs, suggesting that METTL8 is the only methyltransferase for m³C in mRNAs (Xu et al., 2017). The overall presence of m³C in mRNAs has been shown by HPLC-coupled triple quadrupole mass spectrometry (LC-MS/MS). Thereby, a similar abundance of m³C (5 per 10⁵ C) as m¹A (7 per 10⁵ A) and m⁵C (9 per 10⁵ C) and lower levels than m⁶A (1-2 per 10³ A) have been found (Xu et al., 2017). However, NGS sequencing and thus genome wide information about its localization on transcripts is still missing. Localization on mRNA transcripts in the 5' UTR, the coding region or the 3' UTR may shed light on the functional implication of the m³C modification.

2.4 Aims of this thesis

2.4.1 Transcriptome maps of general eukaryotic RNA degradation factors

A large variety of different RNA degradation factors poses the question how RNA degradation pathways are selected and how the RNA sequence encodes or influences this selection. Answering this question requires a systematic analysis of the RNA-binding profiles of the involved protein factors. Although several transcriptome profiles of the RNA degradation factors 5'→3' exonuclease Xrn1, exosome subunits (Rrp44, Csl4, Rrp41, and Rrp6), TRAMP subunits (Mtr4, Trf4, and Air2) and Ski complex subunit Ski2 have been reported (Delan-Forino et al., 2017; Milligan et al., 2016; Schneider et al., 2012; Tuck and Tollervey, 2013), there is a lack of transcriptome-wide binding profiles for components of the deadenylation, decapping, and NMD machineries, as well as subunits of the exosome complex and exosome associated factors. Thus, the task of systematically analyzing the binding of subunits from many known factors involved in RNA degradation to a eukaryotic transcriptome ('transcriptome mapping') has not been accomplished yet. I will use the established method photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) for 30 general RNA degradation factors to systematically generate transcriptome-wide protein binding profiles in the yeast *S. cerevisiae*. In combination with thorough bioinformatic analyses we will draw conclusions regarding protein complex interactions within degradation pathways.

2.4.2 Identification and functional characterization of the novel mRNA modification *N*³-methylcytidine

The m³C modification has so far only been observed in tRNAs of yeast and human cells and in human mRNAs (D'Silva et al., 2011; Noma et al., 2011; Xu et al., 2017). However, the presence in yeast mRNA and the positional information regarding the localization within human mRNA transcripts is still lacking. To investigate and localize the m³C modification in mRNAs of *S. cerevisiae*, I will perform PAR-CLIP experiments of Trm140 to analyze its

mRNA binding ability. Protein binding to mRNA suggests a modification in proximity to the binding site. I will use this approach in combination with a newly developed method using anti-m³C antibodies for RNA immunoprecipitation, called m³C-CLIP. Functional analyses by metabolic labeling and sequencing and analyses of ribosome profiling data in yeast will be used to elucidate a potential function of the m³C modification regarding RNA stability and translational fidelity. In addition to the work in yeast, I will perform PAR-CLIP experiments of METTL8 and m³C-CLIP in human HEK293 cells to elucidate the localization of the m³C modification on protein coding transcripts.

3 Materials and Methodology

3.1 Materials

3.1.1 Bacterial strains

Table 1: Bacterial strains used in this study.

Strain	Genotype	Source
BL21-Codon Plus(DE3)-RIL	<i>recA1 endA1 gyrA96 thi-1 hsdR17</i> Stratagene <i>supE44 relA1 lac</i> [F' <i>proAB lacI^q ZΔM15 Tn10 (Tet^r)</i>]	Stratagene
XL1-Blue	<i>recA1 endA1 gyrA96 thi-1 hsdR17</i> Stratagene <i>supE44 relA1 lac</i> [F' <i>proAB lacI^q ZΔM15 Tn10 (Tet^r)</i>]	Stratagene

3.1.2 Yeast strains

Table 2: Yeast strains used in this study.

Strain	Genotype	Source
BY4741 (Wildtype)	MATa; <i>his3Δ1; leu2Δ0; met15Δ0;</i> <i>ura3Δ0</i>	Euroscarf
Air1-TAP	BY4741; <i>AIR1-TAP::HIS3MX6</i>	Euroscarf
Air2-TAP	BY4741; <i>AIR2-TAP::HIS3MX6</i>	Euroscarf
Caf40-TAP	BY4741; <i>CAF40-TAP::HIS3MX6</i>	Euroscarf
Ccr4-TAP	BY4741; <i>CCR4-TAP::HIS3MX6</i>	Euroscarf
Csl4-TAP	BY4741; <i>CSL4-TAP::HIS3MX6</i>	Euroscarf
Dcp1-TAP	BY4741; <i>DCP1-TAP::HIS3MX6</i>	Euroscarf
Dcp2-TAP	BY4741; <i>DCP2-TAP::HIS3MX6</i>	Euroscarf
Dhh1-TAP	BY4741; <i>DHH1-TAP::HIS3MX6</i>	Euroscarf
Edc2-TAP	BY4741; <i>EDC2-TAP::HIS3MX6</i>	Euroscarf
Edc3-TAP	BY4741; <i>EDC3-TAP::HIS3MX6</i>	Euroscarf
Mtr4-TAP	BY4741; <i>MTR4-TAP::HIS3MX6</i>	Euroscarf
Nmd4-TAP	BY4741; <i>NMD4-TAP::HIS3MX6</i>	Euroscarf
Not1-TAP	BY4741; <i>NOT1-TAP::HIS3MX6</i>	Euroscarf

Pan2-TAP	BY4741; <i>PAN2-TAP::HIS3MX6</i>	Euroscarf
Pan3-TAP	BY4741; <i>PAN3-TAP::HIS3MX6</i>	Euroscarf
Pop2-TAP	BY4741; <i>POP2-TAP::HIS3MX6</i>	Euroscarf
Rrp4-TAP	BY4741; <i>RRP4-TAP::HIS3MX6</i>	Euroscarf
Rrp6-TAP	BY4741; <i>RRP6-TAP::HIS3MX6</i>	Euroscarf
Rrp40-TAP	BY4741; <i>RRP40-TAP::HIS3MX6</i>	Euroscarf
Rrp44-TAP	BY4741; <i>RRP44-TAP::HIS3MX6</i>	Euroscarf
Ski2-TAP	BY4741; <i>SKI2-TAP::HIS3MX6</i>	Euroscarf
Ski3-TAP	BY4741; <i>SKI3-TAP::HIS3MX6</i>	Euroscarf
Ski7-TAP	BY4741; <i>SKI7-TAP::HIS3MX6</i>	Euroscarf
Ski8-TAP	BY4741; <i>SKI8-TAP::HIS3MX6</i>	Euroscarf
Trf4-TAp	BY4741; <i>TRF4-TAP::HIS3MX6</i>	Euroscarf
Trf5-TAP	BY4741; <i>TRF5-TAP::HIS3MX6</i>	Euroscarf
Trm140-TAP	BY4741; <i>BUR1-TAP::HIS3MX6</i>	Euroscarf
Trm140 knockout	BY4741; MATa; <i>ura3Δ0</i> ; <i>leu2Δ0</i> ; <i>his3Δ1</i> ; <i>met15Δ0</i> ; <i>YOR239w::kanMX4</i>	Euroscarf
Upf1-TAP	BY4741; <i>UPF1-TAP::HIS3MX6</i>	Euroscarf
Upf2-TAP	BY4741; <i>UPF2-TAP::HIS3MX6</i>	Euroscarf
Upf3-TAP	BY4741; <i>UPF3-TAP::HIS3MX6</i>	Euroscarf
Xrn1-TAP	BY4741; <i>XRN1-TAP::HIS3MX6</i>	Euroscarf

3.1.3 Human cell lines

Table 3: Human cell lines used in this study.

Cell line	Description
Flp-In™ T-REx™ 293 Cell Line (Thermo Fisher Scientific)	Modified HEK293 cell line used for generation of a stable cell line of METTL8-3xFLAG from a Flp-In™ expression vector by homologous recombination.

3.1.4 Media and supplement

Table 4: Growth media used in this study.

Name	Description/ source	Species
<i>Lysogeny Broth</i> (LB)	1% (w/v) tryptone; 0.5% (w/v) yeast extract; 0.5% (w/v) NaCl; (+1.5% (w/v) agar for solid media plates)	<i>E. coli</i>
Yeast extract peptone dextrose (YPD)	2% (w/v) peptone; 2% (w/v) glucose; 1.5% (w/v) yeast extract (+1.8% (w/v) agar for solid media plates)	<i>S. cerevisiae</i>
Synthetic complete medium (SCM)	0.69% (w/v) yeast nitrogen base without amino acids; 0.04% (w/v) complete supplement mixture; 0.04% (w/v) complete supplement mixture without uracil (Formedium)	<i>S. cerevisiae</i>
Dulbecco's Modified Eagle Medium (DMEM)	Growth and maintenance medium from Thermo Fisher Scientific	<i>H. sapiens</i>

Table 5: Media supplements used in this study.

Name	Description/ source	Working concentration
4-thiouracil	Labeling of nascent RNA (<i>S. cerevisiae</i>) from Sigma	1 mM
4-thiouridine	Labeling of nascent RNA (human) from Carbosynth	500 μ M
Blasticidin HCl	Antibiotic from Thermo Fisher Scientific	15 μ g/mL
FBS	Fetal bovine serum as supplement for DMEM from Thermo Fisher Scientific	1:10
GlutaMAX (200 mM)	Supplement for DMEM from Thermo Fisher Scientific	1:100
Hygromycin	Antibiotic from Invitrogen	50 μ g/mL

Isopropyl-β- D-thiogalactopyranosid (IPTG)	Gene expression induction in <i>E. coli</i> from Carl Roth	1 mM
Kanamycin	Antibiotic from Carl Roth	50 µg/mL
Tetracycline	Gene of interest expression induction in the Flp-In™ T-REx™ 293 Cell Line from Thermo Fisher Scientific.	1 µM
Zeocin	Antibiotic from Invitrogen	100 µg/mL

3.1.5 Spike-ins

Table 6: Spike-ins used in this study provided by Dr. Kristina Zumer.

Spike-in	ERCC-ID	length	Number of U	GC content (%)	4sU labeled
Spike 2	Derived from ERCC-00043	985	297	33.9	yes
Spike 4	Derived from ERCC-00136	1014	268	42.6	yes
Spike 5	Derived from ERCC-00145	1015	264	42.8	no
Spike 8	Derived from ERCC-00092	1079	287	51.8	yes
Spike 9	Derived from ERCC-00002	1037	263	52.8	no
Spike 12	Derived from ERCC-00170	949	296	34.8	no

3.1.6 Primers and oligonucleotides

Table 7: Primers and oligonucleotides used in this study were purchased from IDT.

Name	Sequence
3' adapter	/5rApp/TGGAA TTCTCGGGTGCCAAGG/3ddC/
5' adapter	/5InvddT/rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrCrNrNrNrNrN
RT primer	CCTTGGCACCCGAGAATTCCA

NEXTflex barcode primer	CAAGCAGAAGACGGCATAACGAGA-Barcode-GTGACTGGAGTTC CTTGGCACCCGAGAA TTCCA
Barcode 1	ATCACG
Barcode 2	CGATGT
Barcode 3	TTAGGC
Barcode 4	TGACCA
Barcode 5	ACAGTG
Barcode 6	GCCAAT
Barcode 7	CAGATC
Barcode 8	ACTTGA
Barcode 9	GATCAG
Barcode 10	TAGCTT
Universal primer	AATGATACGGCGACCACCGAGATCTACACGTTC AGAGTTCTACAGTCCGA
Nextera primer 1	AATGATACGGCGACCACCGA
Nextera primer 2	CAAGCAGAAGACGGCATAACGA
RT primer Serine	FAM-CGACACCAGCAGGATTTGAA
RT primer Arginine	FAM-CGTTCCGTACGGGACT

3.1.7 Thermal cycler programs

Table 8: List of thermal cycler programs used in this study.

Program	Step	Temperature /°C	Time /sec	Repeat
Fusion PCR	1	98	120	
	2	98	20	
	3	60	30	
	4	72	30	Repeat steps 2-4 29 times
	5	72	300	
One-step PCR	1	95	300	
	2	98	80	
	3	55	30	
	4	72	60	
	5	72	180	

3.1.8 Plasmids

Table 9: List of plasmids used in this study.

Vector	Description	Source
438-C	N-terminal His 6× tag, MBP tag, N10 linker and a TEV protease cleavage site	UC Berkeley
pcDNA ^{TM5} /FRT	Expression vector into which the gene of interest was cloned	Thermo Fisher Scientific
pOG44	Flp recombinase expression plasmid under the control of the human CMV promoter.	Thermo Fisher Scientific

3.1.9 Buffers and solutions

Table 10: List of buffers and solutions used in this study.

Name	Ingredients
Lysis buffer (yeast)	50 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.1% SDS, 0.5% NP-40, 0.5% Na-deoxycholate
Lysis buffer (human)	50 mM HEPES-KOH pH 7.5, 150 mM KCl, 2 mM EDTA-NaOH, pH 8.0, 1 mM NaF, 0.5% (v/v) NP40 substitute, 0.5 mM DTT, complete EDTA-free protease inhibitor cocktail
Wash buffer (yeast)	50 mM Tris-HCl pH 7.5, 1 M NaCl, 0.1% SDS, 0.5% NP-40, 0.5% Na-deoxycholate
T1 buffer	50 mM Tris-HCl pH 7.5, 2 mM EDTA
High salt wash buffer	50 mM HEPES-KOH pH 7.5, 500 mM KCl, 0.05% (v/v), NP40 substitute, 0.5 mM DTT, complete EDTA-free protease inhibitor cocktail
Phosphatase buffer	50 mM Tris-HCl pH 7.0, 1 mM MgCl ₂ , 0.1 mM ZnCl ₂
Phosphatase wash buffer	50 mM Tris-HCl pH 7.5, 20 mM EGTA, 0.5% (v/v) NP40 substitute
PNK buffer	50 mM Tris-HCl pH 7.5, 50 mM NaCl, 10 mM MgCl ₂
Proteinase K buffer	50 mM Tris-HCl pH 7.5, 6.25 mM EDTA, 75 mM NaCl, 1% (v/v) SDS
Buffer A	300 mM NaCl, 20 mM Na-HEPES pH 7.4, 30 mM imidazole, 1 mM DTT, 10% glycerol, 0.284 mg/mL leupeptin, 1.37 mg/mL pepstatin A, 0.17 mg/mL PMSF, 0.33 mg/mL benzamidine

Buffer B	1 M NaCl, 20 mM Na-HEPES pH 7.4, 30 mM imidazole, 1 mM DTT, 10% (v/v) glycerol, 0.284 mg/mL leupeptin, 1.37 mg/mL pepstatin A, 0.17 mg/mL PMSF, 0.33 mg/ mL benzamidine
Buffer C	300 mM NaCl, 20 mM Na-HEPES pH 7.4, 500 mM imidazole, 1 mM DTT, 10% (v/v) glycerol, 0.284 mg/mL leupeptin, 1.37 mg/mL pepstatin A, 0.17 mg/mL PMSF, 0.33 mg/mL benzamidine
Buffer D	300 mM NaCl, 20 mM Na-HEPES pH 7.4, 30 mM imidazole, 1 mM DTT, 10% (v/v) glycerol
Buffer E	300 mM NaCl, 20 mM Na-HEPES pH 7.4, 1 mM DTT, 10% (v/v) glycerol

3.1.10 Antibodies

Table 11: List of antibodies used in this study.

Name	Dilution	Application	Source
Anti-m3C	30 μ L per IP	m3C-CLIP	Active Motif
Anti-FLAG M2 magnetic beads	100 μ L per IP	PAR-CLIP	Sigma Aldrich
Anti-FLAG M2	1:1000	WB	Sigma Aldrich
Anti-METTTL8	1:1000	WB	Thermo Fisher Scientific
Anti-mouse-HRP	1:3000	WB	Abcam
Anti-rabbit-HRP	1:5000	WB	GE Healthcare
Anti-rat-HRP	1:5000	WB	Sigma Aldrich
IgG from rabbit serum	100 μ g per IP	PAR-CLIP	Sigma

3.2 Experimental methodology

3.2.1 Yeast culture

S. cerevisiae cells expressing the TAP-tagged protein were cultured from OD₆₀₀ ~0.1 to OD₆₀₀ ~0.5 in SCM minimal medium (Formedium) supplemented with 89 µM uracil, 100 µM 4-thiouracil (4tU), and 2% glucose at 30 °C. After reaching OD₆₀₀ ~0.5, 4tU was added to a final concentration of 1 mM, and cells were cultured further for 4 h (final OD₆₀₀ ~1.3–1.6).

3.2.2 *S. cerevisiae* strain validation by SDS-PAGE and Western Blot

S. cerevisiae BY4741 strains containing tagged genes were tested for expression of the specific C-terminally tandem affinity purification (TAP)-tagged (Euroscarf) protein by Western blotting. Cells were lysed and the diluted lysate was subjected to a precast 4–12% NuPAGE Bis-Tris gel (Invitrogen) for gel electrophoresis. Following SDS-PAGE, samples were transferred onto a PVDF membrane (Bio-Rad). Subsequently, the membrane was incubated with a primary antibody coupled to horseradish peroxidase (HRP) against the TAP tag (PAP; Sigma-Aldrich). The antibody was detected using Pierce enhanced chemiluminescence (ECL) Western blotting substrate (Thermo Fisher Scientific) and a ChemoCam imager (Intas).

3.2.3 Generation of the METTL8-3xFLAG Flp-In™ T-REx™ 293 Cell Line

For the generation of a stable cell line expression METTL8 C-terminally 3xFLAG tagged, the coding sequence was cloned into a pcDNA5 vector including the 3xFLAG sequence (DYKDDDDK). HEK293 stable cell lines expressing METTL8-3xFlag were generated using the Flp-In™ T-REx™ system (Thermo Fisher Scientific) according to the manufacturer's instructions. HEK293 Flp-In™ T-REx™ cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1X GlutaMAX (Thermo Fisher Scientific) at 37 °C with 5% CO₂.

3.2.4 PAR-CLIP

3.2.4.1 *S. cerevisiae*

PAR-CLIP and data acquisition were performed as described (Baejen et al., 2017, 2014; Battaglia et al., 2017; Schulz et al., 2013) with minor modifications. Yeast cells harboring the TAP-tagged protein were grown in minimal medium (CSM mixture, Formedium) containing 89 μM uracil, 100 μM 4-thiouracil (4tU) and 2% glucose from OD₆₀₀ ~0.1 to ~0.5 at 30 °C. After addition of 4tU to a final concentration of 1 mM, cells were grown further for 4 h. 4tU-labeled yeast cells were collected and resuspended in 10 mL PBS for UV-irradiation with an energy dose of 12 J/cm² at 365 nm on ice and constant shaking. Harvested cells were resuspended in lysis buffer and lysed by bead beating (FastPrep-24 Instrument, MP Biomedicals) with 1 mL silica-zirconium beads (Roth) with 8 x 40 sec repeats at 4 m/s and with 1 min incubation on ice between each step. Immunoprecipitation was performed on a rotating wheel over night at 4 °C with rabbit IgG-conjugated protein G magnetic Dynabeads (Invitrogen) using 330 μL beads and 100 μg antibody. Beads were washed twice in 1 mL wash buffer and twice in 1 mL T1 buffer. A partial digest of the crosslinked RNA was performed with 20 U RNase T1 in 400 μL T1 buffer for 25 min at 25°C. To stop the RNase reaction, the sample was immediately cooled on ice for 5 min. To remove spare RNase T1, beads were washed twice in T1 buffer and phosphatase buffer. Dephosphorylation was performed with antarctic phosphatase (NEB) and 1 U/ μL RNase OUT (Invitrogen) for 30 min at 37 °C. Beads were washed once in phosphatase wash buffer and twice in polynucleotide kinase (PNK) buffer. Phosphorylation was performed in T4 PNK reaction buffer A (Fermentas) with final concentration of 1 U/ μL T4 PNK, 1 U/ μL RNase OUT and 1 mM ATP (cold labeling) or 0.1 μCi gamma-32-P-ATP (Perkin Elmer) (radioactive labeling) for Trm140 PAR-CLIP. The reaction was incubated at 37 °C for 1 h. For the radioactive labeling reaction mix, the sample was subsequently spiked with 1 mM ATP final concentration for 5 min at 37 °C. Beads were washed 5 times with PNK buffer and prepared for cDNA library generation. For 3' adapter ligation to the RNA, beads were incubated in T4 RNA ligase buffer (NEB) containing 10 U/ μL T4 RNA ligase 2 (KQ) (NEB, M0373), 10 μM 3' adapter (5rApp-TGGAATTCTCGGGTGCCAAGG-3ddC

(IDT)), 1 U/ μ L RNase OUT, and 15% (w/v) PEG 8000. The reaction was incubated for 18 – 20 h at 16 °C. Beads were washed 5 times with PNK buffer to remove not ligated adapter. For 5' adapter ligation to the RNA, beads were resuspended in T4 RNA ligase buffer (NEB) final concentration of 6 U/ μ L T4 RNA ligase 1 (NEB), 10 μ M 5' adapter (5InvddT-GUUCAGAGUUCUACAGUCCGACGAUCN NNNN, IDT), 1 mM ATP, 1 U/ μ L RNase OUT, 5% (v/v) DMSO, and 10% (w/v) PEG 8000. The reaction was incubated for 4 h at 24 °C and 1 h at 37 °C. Beads were washed 5 times in PNK buffer. For radioactive labeling, the sample was resuspended in 25 μ L 2X NuPAGE LDS Sample Buffer (Thermo Fisher Scientific) and released from the beads by incubation at 95 °C for 5 min. The sample was subjected to SDS-PAGE on NuPAGE 4 – 12% Bis-Tris Polyacrylamide gel electrophoresis (Thermo Fisher Scientific) for 1 h at 160 V in 1X MOPS buffer (Thermo Fisher Scientific). Radioactive RNA-protein bands were detected with the Typhoon FLA 9500 instrument (GE Healthcare Life Sciences). Electro-elution of the protein-RNA complex of interest was performed with D-Tube™ Dialyzer Midi (Merck) according to manufacturer instructions with 800 μ L 1X MOPS buffer and 100 V for 2 h. For cold labeling samples, the beads were incubated twice at 95 °C for 5 min in proteinase K buffer. RNA-protein complexes were subjected to protein digest using 1.5 mg/mL proteinase K (NEB) for 2 h at 55 °C. Acidic phenol/chloroform extraction followed by ethanol precipitation was performed in presence of 1 μ L GlycoBlue (Invitrogen) to recover the RNA. Reverse transcription was done for 1 h at 44 °C and 1 h at 55 °C using SuperScript III RTase (Invitrogen). Fusion PCR amplification was performed using the NEXTflex barcode primer kit (Bio Scientific), universal primer and Phusion HF master mix (NEB). Generated cDNA was size selected using 4% E-Gel Agarose Gel (Invitrogen) and purified by MinElute gel extraction (Qiagen). Purified cDNA was amplified in a one-step PCR using the Nextera primers and the KAPA Library preparation kit (Roche). The library was purified using AMPure XP beads (Beckman Coulter) in a ratio of 1:1.8, subsequently quantified using TapeStation 2200 (Agilent Technologies) and Qubit (Qiagen). Samples were sequenced on an Illumina machine (HiSeq 2500 or 4000, NextSeq550). The number of independent biological replicates used for analysis is shown in Table 12. I performed 49 out of the 75 PAR-CLIP

experiments for the degradation project. In addition, Andrea Boltendahl and I generated further 20 PAR-CLIP experiments together. Single PAR-CLIP experiments of Mtr4 and Xrn1 were generated by Saskia Gressel and PAR-CLIP experiments of Air2, Dcp1, Dcp2, and Edc3 were generated by Dr. Carlo Baejen, who originally started this project.

Table 12: Biological replicates for yeast PAR-CLIP experiments.

Experiment	Number of replicates in the analysis
Air1-TAP	4
Air2-TAP	4
Caf40-TAP	2
Ccr4-TAP	3
Csl4-TAP	2
Dcp1-TAP	2
Dcp2-TAP	3
Dhh1-TAP	2
Edc2-TAP	2
Edc3-TAP	2
Mtr4-TAP	3
Nmd4-TAP	2
Not1-TAP	2
Pan2-TAP	3
Pan3-TAP	3
Pop2-TAP	2
Rrp4-TAP	2
Rrp6-TAP	2
Rrp40-TAP	2
Rrp44-TAP	4
Ski2-TAP	2
Ski3-TAP	3
Ski7-TAP	3
Ski8-TAP	2
Trf4-TAp	3
Trf5-TAP	3
Trm140-TAP	3
Upf1-TAP	2
Upf2-TAP	2
Upf3-TAP	2
Xrn1-TAP	2

3.2.4.2 Human

PAR-CLIP of METTL8 was performed with the Flp-In T-Rex cell line expressing METTL8 with a C-terminal 3xFLAG tag. Cells were cultured according to the manufacturer's instruction. Gene expression was induced with 1 µg/ml tetracycline for 24 h. After 23 h, cells were labeled with 500 µM 4-thiouridine (Carbosynth) for 1 h at 37 °C. For UV-crosslinking, medium was removed and cells were washed with PBS. UV-crosslinking was performed with the culture plate on ice and irradiated with 0.9 J/cm² at a wavelength of 365 nm. Cells were resuspended in PBS and collected by centrifugation. Cell lysis was performed in 3 volumes (of the cell pellet) lysis buffer on ice for 10 min. For efficient cell lysis, the lysate was passed through a 27H needle syringe for 10 times. For RNase digest, RNase I (Ambion 100 U/µL) was added to a final concentration of 200 U and incubated at 37 °C for 5 min. Immediately after incubation, lysate was incubated on ice for 5 min. RNase digest was finally stopped by adding 20 µL SUPERase-In RNase inhibitor (20 U/µL, Thermo Fisher Scientific). Lysate was cleared by centrifugation at 13.000 g for 10 min at 4 °C. The supernatant was transferred to 5 mL Eppendorf tube. Protein concentration was quantified by Bradford assay and input was taken to test IP efficiency. For immunoprecipitation, 100 µL Anti-FLAG® M2 Magnetic Beads (Sigma) were washed twice in PBS and lysis buffer per sample. Immunoprecipitation was performed with 25 mg lysate and 100 µL beads over night at 4 °C on a rotating wheel. After IP, beads were collected and washed 3 times with 1 mL high salt buffer and 1 time with 1 mL phosphatase buffer. Dephosphorylation of the 5' RNA end was performed with Antarctic Phosphatase (NEB, M0289S) for 30 min at 37 °C, shaking at 800 rpm. Beads were washed once with Phosphatase wash buffer and twice with PNK buffer. 10% of IP samples were taken as a control and compared to the input sample by Western Blot. Phosphorylation was performed with 1x T4 PNK reaction Mix with 1 U/µL T4 PNK (Thermo Fisher Scientific) and 0.1 µCi gamma-32P-ATP (Perkin Elmer) for 1 h at 37 °C, shaking at 800 rpm. To ensure all RNAs are fully phosphorylated, ATP (Fermentas) was added to a final concentration of 100 mM for 5 min at 37 °C, 800 rpm. Subsequently, beads were washed 5 times with 1 mL PNK buffer. 3' adapter ligation was performed in 40 µL of 3' adapter

ligation mix (4 μ L of 100 mM 3' adapter, 20 U/ μ L T4 RNA Ligase 2 (KQ) (NEB, M0373) over night at 16 °C and 1 h at 25 °C, shaking at 600 rpm. Beads were then washed 5 times with 1 mL PNK buffer. For 5' adapter ligation, beads were resuspended in 40 μ L 5' adapter mix including 4 μ L 100 mM 5' adapter and 6 U/ μ L RNA ligase I (NEB, M0437M) and incubated for 4 h at 25 °C and 1 h at 37 °C, shaking at 600 rpm. Beads were washed 5 times with 1 mL PNK buffer and resuspend in 25 μ L 2X NuPAGE LDS Sample Buffer (Thermo Fisher Scientific). Protein-RNA complexes were released from the beads by boiling for 10 min at 95°C. Supernatant was run on a NuPAGE 4 – 12% Bis-Tris Polyacrylamide gel (Thermo Fisher Scientific) in MOPS buffer (Thermo Fisher Scientific) in order to separate protein complexes. The phosphor-screen was incubated with the gel and the gel image was visualized using a Typhoon FLA9500 (GE Healthcare Life Sciences). Printout was aligned to the gel and the band corresponding to the protein of interested bound to RNA was excised. Protein bound RNA was eluted from the gel by electro-elution using D-Tube Dialyzer Midi MWCO 35 kDa (EMD Millipore) according to the manufacturer's instructions with 100 V for 2 h. For proteinase K digest, the electro-eluate was incubated with 40 μ L proteinase K (800 U/ μ L, NEB) for 1 h at 55 °C. RNA was purified by Phenol-Chloroform extraction and ethanol precipitation. Purified RNA was reverse transcribed using Superscript III (Invitrogen, 18080-044). Library generation was performed by PCR amplification using the NEXTflex barcode primer kit (Bio Scientific). 4% E-Gel Agarose Gel (Invitrogen) was used to size select the cDNA, which was then purified by MinElute gel extraction (Qiagen). cDNA was amplified in a one-step PCR using the Nextera primers and the KAPA Library preparation kit (Roche, Germany). AMPure XP beads (Beckman Coulter) in a ratio of 1:1.8 were used to purify the cDNA library, which was subsequently quantified using TapeStation 2200 (Agilent Technologies) and Qubit (Qiagen). An Illumina machine (HiSeq2500, HiSeq4000 or NextSeq550) was used for sequencing. The experiment was performed in three independent biological experiments.

3.2.5 m3C-CLIP

3.2.5.1 *S. cerevisiae*

S. cerevisiae wildtype and Trm140 knockout cells were cultured to OD₆₀₀ ~1.5 in SCM minimal medium (Formedium) supplemented with 10 mg/L uracil, 100 mM 4-thiouracil (4tU), and 2% glucose. Cells were lysed by bead beating and RNA was recovered by acidic phenol/chloroform extraction.

500 µg of RNA were immunoprecipitated with 30 µL anti-m3C antibody (Active Motif) for 2 h at 4 °C. Antibody-RNA complexes were UV-irradiated with an energy dose of 0.45 J/cm² at a wavelength of 365 nm. Antibody-RNA complexes were captured using protein G magnetic Dynabeads (Invitrogen) for 2 h at 4 °C. Crosslinked RNA was partially digested with RNase T1 and used for cDNA library preparation. RNA was recovered by Proteinase K digestion for 2 h at 55 °C and subsequent acidic phenol/chloroform extraction and ethanol precipitation. Reverse transcription was performed using SuperScript III RTase (Invitrogen). PCR amplification was done using the NEXTflex barcode primer kit (Bio Scientific). Generated cDNA was purified, size-selected, and quantified using TapeStation (Agilent Technologies). Samples were sequenced on an Illumina machine (HiSeq2500, HiSeq4000 or NextSeq550).

3.2.5.2 Human

m3C-CLIP of human cells was performed with the Flp-In T-Rex cell line expressing METTL8 with a C-terminal 3xFLAG tag. Cells were cultured as described. Gene expression was induced with 1 µg/mL tetracycline for 24 h. After 23 h, cells were labeled with 500 µM 4-thiouridine (Carbosynth) for 1 h at 37 °C. Cells were washed in PBS, collected by centrifugation, and resuspended in QIAzol (Qiagen) according to the manufacturer's instruction. RNA was purified using ethanol precipitation and a fraction of it was demethylated with AlkB (see section 3.2.8) as control. The subsequent protocol starting with RNA immunoprecipitation was performed as described in yeast in section 3.2.5.1.

3.2.6 4tU-seq in *S. cerevisiae*

WT and Trm140 KO strains were grown in yeast extract peptone dextrose (YPD) overnight at 30 °C. Cultures were diluted to OD₆₀₀ ~0.1 and cultured to OD₆₀₀ ~0.6. 4tU labeling was subsequently performed as described (Sun et al., 2012). Labeled RNA was purified from extracted total RNA. The Ovation Universal RNA-Seq System Library Kit (NuGen) was used according to the manufacturer's instructions with an input of 100 ng of labeled RNA. The Fragment Analyzer and Qubit (Invitrogen) were used to check the quality and quantity of the libraries. Libraries were pooled and sequenced on an Illumina Sequencer (NextSeq550).

3.2.7 Protein purification of AlkB and Trm140

Geneblocks coding for the respective protein were purchased from IDT and cloned into His6-TEV-tag containing LIC expression vectors in XL1 Blue cells supplemented with Kanamycin. Positive plasmids were transformed into BL21DE3RIL cells. Cells were grown in LB medium at 37 °C until OD₆₀₀ ~0.6 and protein expression was induced by the addition of IPTG with final concentration of 1 mM for 3 h. All purification steps were performed at 4 °C. Cells were resuspended and lysed in buffer A using a sonicator for 5 min at 60% output. The lysate was applied to a HisTrap column and washed with buffer B. Protein was eluted with buffer C. Fractions containing His6-TEV-protein were collected, TEV was added and dialyzed against buffer D. Sample was applied to a HisTrap column and flow through was collected. Gel filtration was performed on a HiLoad 16/600 Superdex 75 pg using buffer E. For Trm140 an additional Q column after His6-tag removal was applied and protein was eluted with buffer B.

3.2.8 RNA demethylation using AlkB

RNA demethylation was performed as described (D'Silva et al., 2011). Briefly, purified RNA (50-1000 µg) isolated from yeast was incubated with purified AlkB from *E. coli* (see Methods section 2.2.7) in buffer containing 50 mM Hepes KOH (pH 8), 75 mM Fe(NH₄)₂(SO₄)₂·6H₂O, 1 mM α-ketoglutarate, 2 mM Na-ascorbate, and 50 mg/mL BSA for 1 h at 37 °C. The reaction was stopped using a final concentration of 11 mM EDTA (Trewick et al., 2002), followed by phenol/chloroform extraction and ethanol precipitation to recover the RNA.

3.2.9 RNA re-methylation using Trm140 and SAM

Re-methylation assay was performed as described (D'Silva et al., 2011). Briefly, demethylated or Trm140 KO RNA was treated with 50 mM Tris-HCl pH 8.0, 1 mM DTT, 0.1 mM EDTA, 1 mM spermidine, 0.5 mM SAM, and purified Trm140 for 1 h at 30 °C, followed by phenol extraction and ethanol precipitation for recovery of the RNA.

3.2.10 Reverse transcription assay and denaturing PAGE

S. cerevisiae wildtype and Trm140 knockout cells were grown to OD₆₀₀ ~1.5 in YPD. Cells were lysed by bead beating and RNA was recovered by acidic phenol/chloroform extraction. 20 to 80 µg of RNA were used for reverse transcription for 1 h at 44 °C and 1 h at 55 °C using SuperScript III RTase (Invitrogen). Products were separated on a 15% 8 M Urea gel at 300 V. Products were visualized on a Typhoon FLA 8500 using 5' FAM labeled RT primers for tRNA Serine (5'-CGACACCAGCAGGATTTGAA-3') and (Arginine (5'-CGTTCCGTACGGGACT-3')).

3.3 Data analysis

3.3.1 PAR-CLIP data pre-processing

This analysis was performed by Salma Sohrabi-Jahromi. “Reads from PAR-CLIP experiments with replicates were merged after making sure that all samples showed high Spearman correlation values comparing binding occupancies of replicates on different genes (Figure 7B). Mapping and statistical evaluation of PAR-CLIP experiments was performed using our in-house software mockinbird (Roth and Torkler, 2018). In summary, the UMI is removed from the 5′ end with UMI-tools (T. Smith et al., 2017), and the 3′ adapter is trimmed with Skewer (Jiang et al., 2014). Reads with traces of the 5′ adapter are discarded. The preprocessed reads are then mapped to the *S. cerevisiae* genome (sacCer3, version 64.2.1). After mapping PCR duplicates are removed with UMI-tools.

We used two alternative approaches for mapping reads using Bowtie (Langmead et al., 2009): For all analyses except the ‘transcript class enrichment analysis’ in Figure 8, reads are uniquely mapped with up to one mismatch. We discard alignments shorter than 20 nt. This stringent mapping ensures that our high confidence PAR-CLIP cross-link sites are originating from correctly mapped reads on the reference genome. For Figure 8, unique mapping would cause the loss of most reads that fall into rRNAs and tRNAs because of duplicated rRNA genes and tRNA isodecoders. For Figure 8, we therefore allowed Bowtie multi-mapping in two regions with `-best`, `-starra` options and discarded reads shorter than 30 nt.

T→C transitions directly at the edge of the reads or with a Phred quality score lower than 20 are not considered as signature of protein binding as they suffer from higher technical noise. To obtain high confidence cross-link sites, we set a stringent cutoff of 0.005 for the p-value of cross-link sites and require a minimum coverage of 2 per site. Moreover, if we see the same transition in at least 75% of reads in the input library control (SRA: SRX532381) (Baejen et al., 2014), we annotate it as a single nucleotide polymorphism of our lab strain with respect to the genomic reference and remove such sites from our analysis. Finally, the occupancy of a factor on a verified cross-link site is defined as the

number of transitions obtained from our PAR-CLIP experiments divided by the concentration of RNAs covering the cross-linked site according to the input library control. This control coverage is measured under comparable conditions to PAR-CLIP experiments (Baejen et al., 2014). Occupancy values are capped at the 95th percentile. Subsequent analyses were performed using in-house python scripts. Mockinbird configuration files as well as the analysis scripts can be found at https://github.com/soedinglab/Degradation_scripts.”

3.3.2 Transcript class enrichment

The following analysis was performed by Salma Sohrabi-Jahromi. “We analyzed the distribution of reads from high-confidence cross-link sites over the genome (Figure 8A). We presented the sum of reads from 5′ and 3′ UTRs, coding sequences, and introns as the value for mRNAs. Reads that fall within genomic regions not annotated as categories analyzed here are shown with grey. These annotated transcript classes have comparable U-content, making the comparison between fractions of cross-link sites in each category possible (Appendix Figure 28).

For each factor studied here, we defined enrichment scores that represent their preferences for binding to various transcript classes *c*, in comparison to all other factors. We use annotations for rRNA, tRNA, snoRNA, snRNA, coding sequences (CDS), from *S. cerevisiae* genome sacCer3, version 64.2.1. Untranslated regions around coding boundaries (5′ and 3′ UTRs) were annotated based on TIF-seq experiment (Pelechano et al., 2013). We selected the most strongly expressed isoform for each gene. We then assigned boundaries to 3′ and 5′ UTRs based on annotated CDS of the same gene. We furthermore used annotations for stable, unannotated transcripts (SUTs), cryptic unstable transcripts (CUTs), and Nrd1- unterminated transcripts (NUTs) (Neil et al., 2009; Pelechano et al., 2013; Schulz et al., 2013). We removed overlapping annotations with the following priority list: rRNA, tRNA, snRNA, snoRNA, intron, CDS, UTR, SUT, CUT, NUT. For each factor, we counted the number of high-confidence reads falling in each transcript class. We then used the log₂-transformed matrix and normalized it in the following way for both rows

and columns to get log enrichment values that sum to zero in both rows and columns. The row- and sum-normalized enrichment score is defined as follows, where $X_{f,c}$ is the number of high-confidence reads for factor f that fall into transcript class c , and $X'_{f,c} = \log_2 X_{f,c}$ (Figure 8B):

$$\tilde{X}'_{f,c} = X'_{f,c} - \frac{X'_{f,o} X'_{o,c}}{X'_{o,o}} \quad (1)$$

We defined the row and sum averages of $X_{f,c}$,

$$X'_{f,o} = \frac{1}{C} \sum_{c=1}^C X'_{f,c} \quad (2)$$

$$X'_{o,c} = \frac{1}{F} \sum_{f=1}^F X'_{f,c} \quad (3)$$

$$X'_{o,o} = \frac{1}{FC} \sum_{f=1}^F \sum_{c=1}^C X'_{f,c} \quad (4)$$

F is the number of factors and C is the number of transcript classes (Figure 8B). The normalization can be interpreted as subtracting from the log enrichment matrix X' the first singular component of its singular-value decomposition.”

3.3.3 Metagene analysis

This analysis was performed by Salma Sohrabi-Jahromi. “We used the most abundant TIF-annotated isoform for mRNAs (Pelechano et al., 2013) as a reference. Transcripts longer than 1500 bases are chosen and aligned at their TSS or pA sites. The average occupancy per nucleotide is then calculated based on high-confidence cross-link sites of each PAR-CLIP experiment. The profiles are smoothed by a moving average in a 41 nt window and the 95%

confidence interval is estimated by 1500 bootstrap sampling iterations over the transcripts. To further denoise the profiles, the cross-link sites falling in snRNAs, rRNAs, and tRNAs are removed. Furthermore, to avoid ambiguous results, we made sure that the profile comes solely from the central gene. To do so, we performed the metagene analysis around the TSS on the sense strand on TIF-annotated mRNAs that have no other mRNA up to 700 bp upstream of their TSS (3193 transcripts in total). Analogously, for sense-strand pA site profiles we used mRNAs that have no nearby genes downstream of their pA site up to 700 bases on the same strand (3193 transcripts in total). For the antisense strand profiles, we applied the same criteria on the opposite strand which left us with 3076 and 3193 transcripts filtered around TSS and pA sites respectively. This ensures that the observed antisense binding does not originate from neighboring or overlapping transcripts on the antisense strand. In both cases we looked at the average occupancy in a window of $[\pm 700 \text{ nt}]$ around TSS and around pA sites. Occupancies were normalized to the maximum value, which is the background binding level for antisense profiles with no significant cross-linking to the antisense strand (Figure 9 and Figure 10). The same procedure was followed to plot metagene occupancies centered around protein-coding regions and snoRNAs from *S. cerevisiae* genome sacCer3, version 64.2.1 (Appendix Figure 29 and Figure 39)."

3.3.4 Co-occupancy

Salma Sohrabi-Jahromi performed the following analysis. "Co-occupancy measures the tendency of two factors to bind to the same transcripts. Occupancy of a factor on a transcript is defined as the sum of occupancies for all high-confidence cross-link sites falling within this transcript. Co-occupancy of two factors is defined as the Pearson correlation over all transcripts between the occupancies of these factors (Figure 11A). We used these correlation values between all pairs of RNA processing factors to assign distances to each pair and used tSNE (Van Der Maaten and Hinton, 2008) to visualize the two-dimensional nonlinear embedding of co-occupancies for all RNA-binding proteins in our dataset (Figure 11C)."

3.3.5 Co-localization

This analysis was performed by Salma Sohrabi-Jahromi. “Co-localization measures how likely two factors are to bind near each other in the transcriptome. More precisely, we first calculate the occupancy of a factor $f \in \{1, \dots, F\}$ around the cross-link sites of another factor f' ([- 40 nt, + 40 nt] excluding the centered T). We then normalize according to the total occupancy values,

$$z_{ff'} = \sum_{i=1}^{n_{f'}} \left(\sum_{j=-40}^{-1} Occ_{ff',i,j} + \sum_{j=1}^{40} Occ_{ff',i,j} \right) \quad (5)$$

$$\text{co-localization}(f, f') = \frac{z_{ff'}}{\sum_f z_{ff'} \sum_{f'} z_{ff'}} \quad (6)$$

Where, n_f is the number of cross-link sites for factor f , and $Occ_{ff',i,j}$ is the occupancy of f at position j around the i^{th} cross-link site from factor f' ($Occ_{ff',i,j} = 0$ if no verified cross-link sites exist). To improve signal-to-noise, we compute from the resulting matrix of co-localizations between all RNA-processing factors $C_{f,f'}$, the matrix of Pearson correlations between the rows of $C_{f,f'}$, (Figure 11B, Figure 21, Appendix Figure 41).“

3.3.6 Codon-enrichment analysis

The following analysis was performed by Salma Sohrabi-Jahromi. “To search for possible links between translation efficiency and RNA degradation, we checked if some degradation factors preferentially bind to translationally efficient/non-efficient transcripts. To do so we adapted the proposed normalized translation efficiency scale (Pechmann and Frydman, 2013). The authors generate a normalized optimality score for codons that incorporates the competition between supply and demand of tRNAs. The coding region for each transcript was extracted according to ORFs annotated by SGD. The codon

optimality score was averaged over the whole reading frame (Figure 12A, more detailed explanation in the next section).

We then checked whether mRNAs that bind to each factor are enriched or depleted in some codons compared to all mRNAs. To achieve this, we defined the following score for codon enrichment that represents deviations from average frequencies in all mRNAs,

$$\text{codon enrichment} = \frac{\sum_{t=1}^T \left(\frac{\text{occ}(t)}{\sum_{t'=1}^T \text{occ}(t')} \times F_{c,t} \right)}{\frac{1}{T} \sum_{t=1}^T F_{c,t}} \quad (7)$$

Here T is the number of mRNA transcripts, $F_{c,t}$ is the fraction of the codon c in transcript t , and $\text{occ}(t)$ is the *total* occupancy of the factor on transcript t . 90% confidence intervals were generated by bootstrapping: we sampled *with replacement* 1000 times the same number of mRNAs from the total set as in total, and for each set we recalculated the codon enrichment score. We colored the bars based on the previously ranked optimality of codons (Pechmann and Frydman, 2013) (Figure 12B, Appendix Figure 31-Figure 37)."

3.3.7 Relating occupancies to various transcript features

This analysis was performed by Salma Sohrabi-Jahromi. "We analyzed the correlation of the occupancy of all factors with transcript length, codon enrichment of the transcript, expression level, transcript stability, and polyA tail length. For expression, we used an RNA-seq experiment of wild-type yeast (SRA: SRX532381) (Baejen et al., 2017) and mapped the reads to mRNAs. We present the average number of reads per base as an estimate for gene expression. For half-life calculations, we used published yeast 4tU-seq (GEO: GSM2199309) and RNA-seq experiments (SRA: SRX532381) (Baejen et al., 2017). Transcript half-life is estimated with an optimized method that will be published elsewhere (Hofmann et al., unpublished).

Since there are only few transcripts with very low or very high half-life, codon optimality, and expression (Appendix Figure 30), we performed the analysis on a subset of mRNAs where the transcript property lies between the 5% and 95% quantiles. We then compared the *total* occupancy of degradation factors on each mRNA relative to such transcript features (Figure 12A, Figure 13B, and Appendix Figure 31-Figure 37). We show 95% confidence intervals generated by bootstrapping mRNAs in grey shade.

We checked whether such correlations originate from the feature of interest or merely shows up due to correlations between this feature and others (Appendix Figure 30). We used a multivariate linear regression to model total occupancy as a linear function of these four features:

$$occupancy'(t) \sim length + optimality + expression + halflife \quad (8)$$

In cases where the correlation is a direct effect from our feature of interest, we expect to lose significantly on our prediction when this variable is taken out of the equation. Therefore, we use p-values representing the importance of each feature in this linear regression as a score representing the significance of its contribution in explaining the final occupancies. Occupancy correlated strongly with transcript length, which dominated as explanatory variable in this regression, trivially because most factors bind along the entire transcript. To eliminate this trivial dependency, we used occupancy per nucleotide, denoted *occupancy'*, as the target variable in our regression (Figure 12C)."

3.3.8 Motif enrichment analysis

Salma Sohrabi-Jahromi performed the following analysis. "To find sequence preferences for binding events of degradation factors, we counted 4-mers in a window of $[\pm 5 \text{ nt}]$ intervals around high-confidence cross-link sites of PAR-CLIP experiments. Based on this count table, the enrichment score for each 4-mer was calculated using the following formula,

$$4\text{-mer enrichment} = \frac{n_{4\text{-mer},i} + 1}{N \times \prod_{j=1}^4 P_{4\text{-mer}[j]}} \quad (9)$$

Here N is the number of cross-link sites below the cut-off p-value (we used a maximum of 5000 cross link sites), $n_{4\text{-mer},i}$ is the number of observed 4-mers at position i in the set of binding sequences aligned at their cross-link site $i=0$, $4\text{-mer}[j]$ is the base at the j 'th position of the 4-mer, and P_b is the probability of observing base b . We used the probabilities: $P_A = P_T = 0.31$ and $P_C = P_G = 0.19$ based on frequencies in yeast genome and corrected for the T bias at the cross-link site (Figure 13C)."

3.3.9 Trm140 PAR-CLIP and m3C-CLIP analysis

This analysis was performed by Gabriel J. Villamil. Sequencing reads of independent biological replicates from Trm140 PAR-CLIP and m3C-CLIP sequencing experiments were merged. Single-end 50 nt and 75 nt reads were mapped to the R64-2-1 *S. cerevisiae* reference genome (Engel et al., 2013) using STAR 2.5.3a (Dobin and Gingeras, 2015) with a maximum of two mismatches allowed per alignment. In cases of multi-mapped reads, the single best alignment was taken. Trm140 and anti-m3C antibody binding was determined from chemical cross-links indicated by T→C transitions called using a statistical model as described (Baejen et al., 2017, 2014) with a maximum p-value threshold of 0.005 and a minimum coverage of 2. The Phred score of sequencing quality at the transition base was set to 20. RNA-seq data obtained under similar conditions to PAR-CLIP experiments was used to remove potential transition miscalls caused by SNPs. Briefly, transitions in PAR-CLIP reads that are also observed in at least 75% of RNA-seq reads covering the same base were considered as SNPs and removed from the analysis. Accumulation of binding sites are depicted metagene-wide aligned at the transcription start site (TSS) and polyadenylation site (pA) of designated loci.

3.3.10 Calling of m3C modification sites

Sites of m3C modification were determined from C→T transitions in PAR-CLIP reads. C→T transitions were called with a transition rate of a least 4%, a minimum coverage of 40, and a minimum sequencing quality of 20 at the transition base resulting in the high confidence list of m3C sites shown in Appendix Table 14.

3.3.11 Calculation of ribosome P site occupancy

This analysis was performed by Gabriel J. Villamil. We calculated the ribosome P site occupancy based on the data from WT and Trm140 KO ribosome profiling experiments (Chou et al., 2017). Briefly, global codon occupancy analysis was calculated (Nedialkova and Leidel, 2015) with minor modifications. Examination of the cumulative distribution of 28-31 nt reads aligned at the start codons using Plastid (online package) revealed the P site offset. The respective offset was applied to reads of each size and off frame sequencing reads were discarded. We removed the first 15 and last 5 codons of each transcript from the reference. The quotient of frequency of each codon in ribosomal A, P and E site to the average frequency of the same codon in the three downstream codons from the A site was used for normalization.

3.3.12 Calculation of translational efficiency on m3C containing codons

Translation efficiency (TE) was calculated by using the averaged sequencing reads of ribosome protected fragments from two independent biological replicates of WT and Trm140 KO ribosome profiling experiments at m3C containing codons and divided by the number of RNA-seq reads from two independent biological replicates of WT and Trm140 KO cells at the same codon (Chou et al., 2017). The following m3C containing transcripts were selected based on the mutational signature in the Trm140 PAR-CLIP experiments: YHR099W (at position chrVIII:303,572), YDR341C (at position

chrIV:1,153,577), YPR148C (at position chrXVI:827,581), YEL060C (at position chrV:41,712), YML058W (at position chrVIII:159,468), and YOR361C (at position chrXV:1,017,134). The change in TE was plotted using Microsoft® Excel (Microsoft Office 2017) for each codon in Figure 20B.

3.3.13 4tU-seq data pre-processing and normalization

Gabriel J. Villamil performed the following analysis. Paired-end 50 nt sequencing reads were mapped to the R64-2-1 *S. cerevisiae* reference genome assembly (Engel et al., 2013) using STAR 2.5.2b (Dobin and Gingeras, 2015) with a maximum of two mismatches allowed per alignment and maximum intron length set to 500,000. In cases of multi-mapped reads, the single best alignment was taken. Sequences of labelled and unlabelled RNA spike-ins were included as additional chromosomes in the reference sequence to map reads from spike-ins. SAMtools 1.3.1 (Li et al., 2009) was used to remove alignments with MAPQ scores lower than 7 and to sort and index BAM files. Piled-up reads on transcription units were counted using HTSeq 0.6.1p1 (Anders et al., 2015). Antisense bias correction was performed as described (Gressel et al., 2017). Briefly, counts of reads mapping antisense to spike-in sequences were used to measure the extent of antisense bias and correct counts for all reads.

Real read counts in labelled RNA samples L or in total cellular RNA samples T for transcription unit i in sample j were calculated as:

$$L_{ij} = \frac{k_{ij}^L}{l_i \sigma_j^L} \quad (10)$$

$$T_{ij} = \frac{k_{ij}^T}{l_i \sigma_j^T} \quad (11)$$

Here, k_{ij} is the raw read count of a transcription unit in a labelled or total sample, l_i is the length of the transcription unit, and σ_j is a scaling factor that accounts for variations in sequencing depth determined from spike-in counts.

3.3.14 Estimation of RNA synthesis and degradation rates

The following analysis was performed by Gabriel J. Villamil. Estimates of gene-specific amounts of labelled RNA α_{ij} and unlabelled RNA β_{ij} in a tuple of labelled and total sample were modelled as:

$$\alpha_{ij} = \frac{L_{ij} - \epsilon_j^L T_{ij}}{1 - \epsilon_j^L} \quad (12)$$

$$\beta_{ij} = \frac{T_{ij} - L_{ij}}{1 - \epsilon_j^L} \quad (13)$$

Here, ϵ_j is a cross-contamination rate that models the proportion of unlabelled reads purified in the labelled sample.

We extend the statistical model to estimate gene-specific synthesis rates μ_{ij} and degradation rates λ_{ij} , assuming first-order kinetics using the following equations:

$$\alpha_{ij} = \frac{\mu_{ij}}{\gamma_j + \lambda_{ij}} \cdot (1 - e^{-t(\gamma_j + \lambda_{ij})}) \quad (14)$$

$$\alpha_{ij} + \beta_{ij} = \frac{\mu_{ij}}{\gamma_j + \lambda_{ij}} \quad (15)$$

Here, t is the labelling duration of 6 minutes, and therefore:

$$\lambda_{ij}(t) = -\gamma_j - \frac{1}{t} \cdot \log\left(\frac{\beta_{ij}}{\alpha_{ij} + \beta_{ij}}\right) \quad (16)$$

$$\mu_{ij}(t) = (\alpha_{ij} + \beta_{ij}) \cdot (\gamma_j + \lambda_{ij}(t)) \quad (17)$$

γ_j is the dilution rate of RNA due to cell growth, it is calculated as:

$$\gamma_j = \frac{\log(2)}{CCL_j} \quad (18)$$

where CCL_j is the length of one cell cycle in minutes.

In order to maximize the number of successfully calculated decay rate values, labelled read counts L_{ij} were optimized across all samples with a common scaling factor that minimizes the occurrence of logarithms of negative values and negative decay rate values, both of which were excluded from further analysis (Figure 23). The inclusion of this factor does not affect the relative observations made in this study.

4 Transcriptome maps of general eukaryotic RNA degradation factors

4.1 Results

“In order to get a better understanding of RNA processing and degradation in a eukaryotic cell, we measured transcriptome-wide binding locations of 30 RNA degradation factors involved in mRNA deadenylation, decapping, exosome-mediated degradation, and in RNA surveillance pathways including nuclear RNA surveillance and cytoplasmic nonsense-mediated decay (NMD) (Table 12). We performed PAR-CLIP in *S. cerevisiae* using our published protocol (Battaglia et al., 2017), with minor modifications (Methods). The high reproducibility of these PAR-CLIP experiments is revealed by a comparison of two independent biological replicates that we collected for all 30 degradation factors (Figure 7B), with Spearman correlations between 0.87 and 1.00 (mean: 0.94). We typically obtained tens of thousands of verified factor-RNA cross-link sites with p-values ≤ 0.005 (Figure 7A). These transcriptome maps represent an extensive, high-confidence dataset of *in vivo* RNA-binding sites for factors involved in RNA degradation.”

Transcriptome maps of general eukaryotic RNA degradation factors

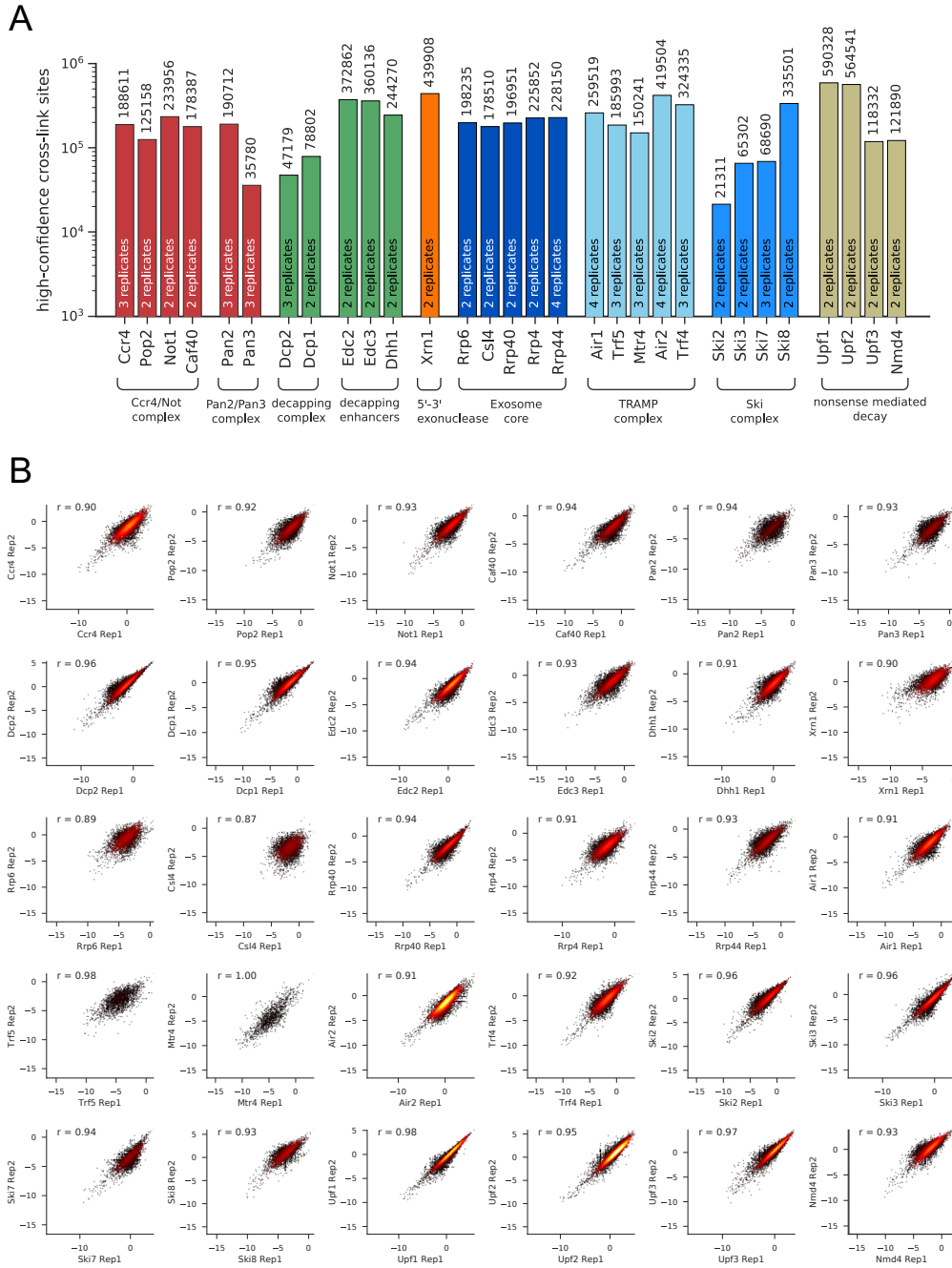


Figure 7: Number of PAR-CLIP cross-link sites and replicate correlation.

A) The number of high-confidence PAR-CLIP cross-link sites for each factor after merging of sequencing files of independent biological replicates is shown as a bar plot. The factors are sorted according to the complex they are residing in and colored in red (deadenylation), green (decapping), orange (exonuclease), dark blue (exosome), middle blue (TRAMP), light blue (Ski), and ochreous (NMD). B) Total transcript occupancy of all 30 factors in replicate experiments are plotted in log₂ space and Spearman correlation values are shown for each pair. Perfectly correlating pairs are located on the diagonal of the plotting area.

4.1.1 Degradation factors exhibit transcript class specificity

“We first compared degradation factor binding over different RNA classes. These included protein-coding messenger RNA (mRNA), where we distinguished the 5′ untranslated region (5′ UTR), the coding sequence (CDS), introns, and the 3′ untranslated region (3′ UTR). We also included several classes of ncRNAs: ribosomal (r), transfer (t), small nucleolar (sno), and small nuclear (sn) RNAs, as well as stable unannotated transcripts (SUTs), cryptic unstable transcripts (CUTs), and Nrd1- untruncated transcripts (NUTs) (Neil et al., 2009; Pelechano et al., 2013; Schulz et al., 2013) (Figure 8).

A first analysis revealed that most PAR-CLIP sequencing reads fall into the mRNA transcript class, although many of the factors also show a considerable number of sequencing reads in ncRNAs, in particular rRNAs (Figure 8A). To obtain a more quantitative comparison, we defined log enrichment scores that reflect the preferences of factors in binding to a specific transcript class in comparison to other factors and classes. To correct for the different sizes of classes and different numbers of measured factor binding sites, we normalized the log enrichment scores by subtracting class- and factor-specific offsets, such that the mean for each class and each factor vanishes (Figure 8B, Methods). This analysis highlights differences between degradation factors with respect to binding to various transcript classes, as will be discussed in detail below.“

4.1.1.1 RNA end- processing complexes differ in their targets

“The catalytic subunit Pop2 and the core subunits Not1 and Caf40 of the deadenylase complex Ccr4/Not have similar binding preferences for the 5′ UTR, the CDS and 3′ UTR of mRNAs, for rRNAs, tRNAs, snoRNAs, and snRNAs (Figure 8B, highlighted in red). Compared to other deadenylation factors of the Ccr4/Not complex, the catalytically active subunit Ccr4 has different binding preferences, and is strongly enriched at mRNA introns. The second deadenylation complex, Pan2/Pan3, shows a similar binding preference as the Ccr4/Not complex (except for the Ccr4 subunit), consistent

with its dominant role in yeast mRNA deadenylation (Boeck et al., 1996). Pan3 shows a strong binding preference for rRNAs and tRNAs.

For all decapping-related factors we observed similar binding preferences among each other (Figure 8B, highlighted in green). They show the strongest enrichment at SUTs and at mRNAs compared to the other transcript classes. Decapping factors bind preferentially to CDS and 3' UTR, as well as to SUTs. This is consistent with previous findings that SUTs are degraded via Dcp2-dependent pathways in the cytoplasm (Marquardt et al., 2011; Smith et al., 2014; Thompson and Parker, 2007). Dcp2, which harbors the hydrolase activity that removes the 5' cap, and the decapping activator Edc3, additionally bind to NUTs. The 5' exonuclease Xrn1 shows a similar binding preference as the decapping factors (Figure 8B, highlighted in orange). Taken together, complexes and enzymes that are known to target mRNA ends for 3' deadenylation and 5' decapping and degradation show remarkably distinct binding specificities to different transcript classes.”

4.1.1.2 The exosome and surveillance factors

“For the exosome we also observed binding to different RNA classes (Figure 8B, highlighted in royal blue). The core exosome subunits Csl4 and Rrp40 showed similar cross-linking to rRNAs, tRNAs, snoRNAs, and snRNAs. The catalytic exosome subunit Rrp44 and the core subunit Rrp4 binds to introns of mRNAs, but preferentially to the short-lived, nuclear CUTs and NUTs. Rrp6, a subunit that is exclusively present in the nuclear exosome complex, shows binding to rRNAs, snoRNAs, snRNAs, CUTs and NUTs. This is consistent with the suggestion that the factor is needed for nuclear processing of such non-coding transcripts and degradation of short-lived nuclear transcripts (Heo et al., 2013; Vasiljeva and Buratowski, 2006). This complex distribution of cross-links for different exosome subunits to different RNA classes reflects the distinct functions of the exosome in nuclear RNA surveillance, processing of stable ncRNAs, and cytoplasmic mRNA degradation (Zinder and Lima, 2017).

The two TRAMP complexes TRAMP4 and TRAMP5 show clearly distinct cross-linking patterns (Figure 8B, highlighted in light blue). TRAMP4 subunits

(Mtr4, Air2, Trf4) are enriched in introns, consistent with a function on mRNAs, and on SUTs, CUTs, and NUTs. The TRAMP5 complex (Mtr4, Air1, Trf5) shows binding enrichment for introns, rRNAs, tRNAs, snRNAs, and snoRNAs. This is in agreement with previous data, which showed rRNA binding for Mtr4 and exosome subunits (Delan-Forino et al., 2017; Schneider and Tollervey, 2013). Moreover, the TRAMP complex cooperates with the Nrd1/Nab3 complex and the nuclear exosome complex during the maturation and 3' pre-processing of snoRNAs (Grzechnik and Kufel, 2008). To distinguish binding upon degradation and binding in order to pre-process snoRNAs, we investigated metagene profiles of TRAMP subunits along snoRNA genes (Appendix Figure 39). Air1/Trf5 bind almost exclusively to the gene body whereas Air2/Trf4 bind downstream of the 3' end. This suggests that TRAMP5 is mainly involved in snoRNA degradation, whereas TRAMP4 may work together with the Nrd1/Nab3 machinery to pre-process snoRNAs (Appendix Figure 39) and to target NUTs, SUTs, and CUTs for degradation (Figure 8B).

The cross-linking preferences of subunits of the Ski complex differ only slightly from each other (Figure 8B, highlighted in cyan). All Ski complex subunits bind the 5' UTR, CDS, and 3' UTR of mRNAs, rRNAs, tRNAs, snoRNAs, and snRNAs. The Ski2 subunit preferentially binds to the CDS of mRNAs, consistent with its function as a helicase to detach bound proteins from the mRNAs (Houseley and Tollervey, 2009; Lebreton and Seraphin, 2008). The exosome adaptor subunit Ski7 preferentially binds rRNAs and tRNAs. These patterns are consistent with the model that the exosome cooperates with distinct accessory complexes and factors to target different transcript classes. Finally, we observed similar cross-linking patterns for all NMD factors with strong binding to SUTs and NUTs (Figure 8B, highlighted in yellow). Upf2 shows a binding preference to introns and CUTs. Upf3 also binds to the 5' UTR, CDS, and 3' UTR of mRNAs, and Nmd4 binds to introns and 3' UTRs of mRNAs."

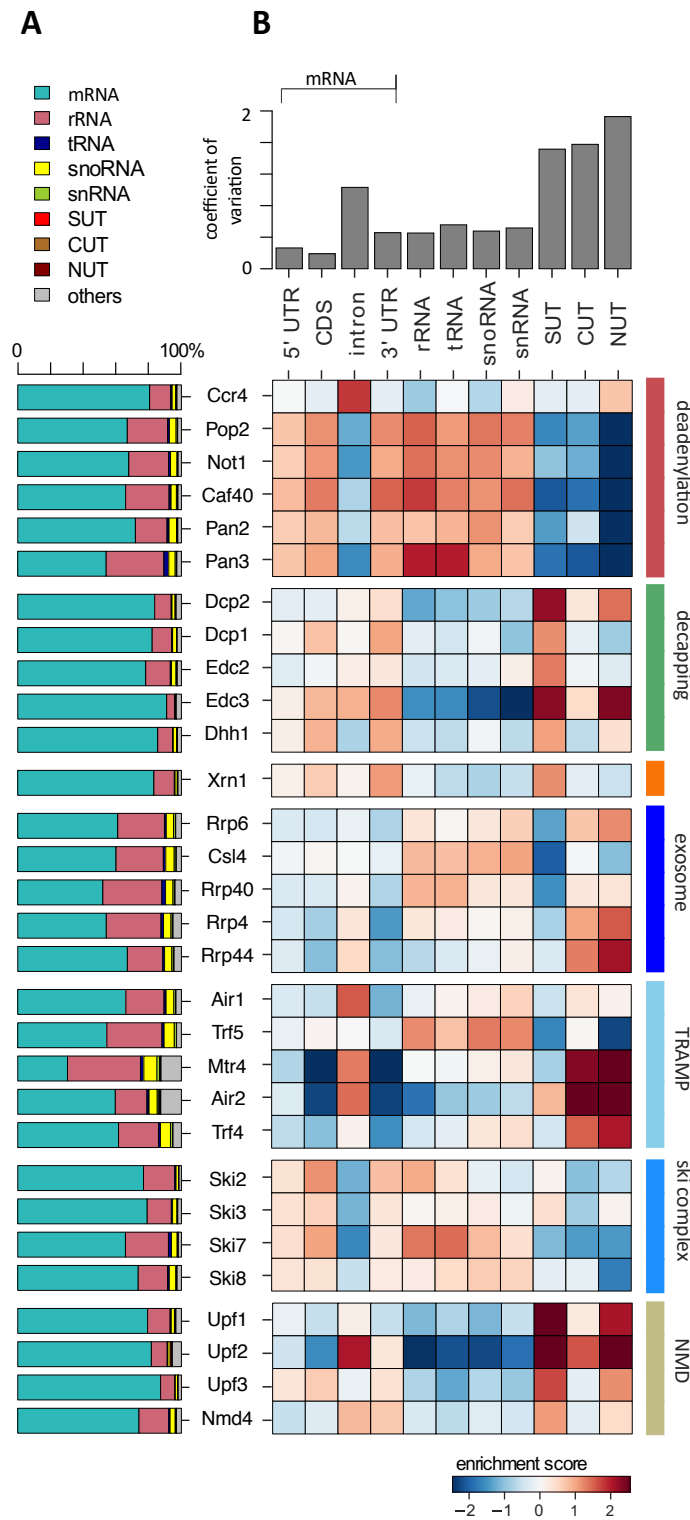


Figure 8: Distribution of degradation factor cross-link sites over the yeast transcriptome.

A) Fractions of high-confidence PAR-CLIP sequencing reads of 30 yeast degradation factors fall into various transcript classes. Depicted classes are the following: messenger RNA (mRNA) in turquoise (n=4,928), ribosomal RNA (rRNA) in antique pink (n=24), transfer RNA (tRNA) in dark blue (n=299), small nucleolar RNA (snoRNA)

in yellow (n=77), small nuclear RNA (snRNA) in green (n=6), stable unannotated transcripts (SUTs) in red (n=357), cryptic unstable transcripts (CUTs) in light brown (n=750), Nrd1-dependent unterminated transcripts (NUTs) in dark brown (n=317).

B) Enrichment z-scores of high-confidence PAR-CLIP cross-link sites of 30 yeast degradation factors (rows) in various segments of mRNA transcripts (left columns; UTR: untranslated region; intron; CDS: coding sequence), or other transcript classes as in A (other columns). The color-coded z-score is the fraction of binding sites of the factor x in the transcript class c minus the mean fraction for that transcript class c divided by the standard deviation of fractions for class c (color encoded, -0.8 in blue to 2.4 in red). The coefficient of variation on top is the standard deviation divided by the mean for each transcript class. Factors are grouped according to their functional role; from top to bottom: deadenylation machinery (Ccr4, Pop2, Not1, Caf40, Pan2, Pan3), decapping (Dcp2, Dcp1, Edc2, Edc3, Dhh1), Xrn1, exosome (Rrp6, Csl4, Rrp40, Rrp4, Rrp44), TRAMP polyadenylation complex (Trf5, Air1, Mtr4, Air2, Trf4), Ski complex (Ski2, Ski3, Ski7, Ski8), nonsense mediated decay (NMD) (Upf1, Upf2, Upf3, Nmd4).

4.1.2 Distinct factor distribution along mRNA

“We next focused on degradation factor distribution on mRNAs. We prepared metagene profiles showing the average occupancy of each factor around the mRNA transcription start sites (TSS) and poly-adenylation (pA) sites, respectively (Figure 9). The Pan2/Pan3 deadenylase complex and the Ccr4/Not subunits Pop2, Not1, and Caf40 all cross-link upstream of the 3' end of mRNA with the highest enrichment at the pA site, as expected from their function in shortening the polyA tail. The catalytic subunit Ccr4 binds strongly in the 5' region of mRNAs. All 5' decapping factors bind upstream of the pA site, and all but the catalytically active subunit Dcp2 show increasing occupancy towards the 3' end of mRNAs. These patterns can be explained if decapping factors are pre-bound to mRNAs that form a closed loop that holds the RNA ends in proximity. In contrast, Dcp2 binds almost exclusively at the pA site, suggesting that it might be recruited only upon active mRNA degradation. The cytoplasmic 5' exonuclease Xrn1 has the highest occupancy towards the 3' end, similar to the previously published crosslinking and cDNA analysis (CRAC) data (Tuck and Tollervey, 2013), thereby resembling the binding profiles of the decapping factors. Comparison of the binding profiles aligned at the pA site or alternatively with profiles aligned at the translation stop codon shows that the binding preference indeed lies at the end of the 3' UTR independent of the stop codon position (Appendix Figure 29).

The exosome core subunits (Csl4, Rrp40, and Rrp4) and the catalytically active subunits (nuclear: Rrp6, cytoplasmic: Rrp44) cross-link to the 5' end of the transcript (Figure 9), possibly because the exosome binds to the 5' end while digesting the 3' end, or more likely because the exosome slows down towards the remaining 5' end of mRNAs after rapid degradation from the 3' end. Both TRAMP complexes bind mainly in the 5' region of mRNAs near the TSS, as previously observed for Mtr4 and Trf4 (Tuck and Tollervey, 2013).

The Ski complex components Ski7 and Ski8 occupy the entire mRNA with increasing occupancy towards the pA site, whereas Ski2 and Ski3 show more discrete binding towards the polyA tail (Figure 9). The NMD factors Upf1 and Upf3 show binding over the entire mRNA with highest occupancy at the pA site, consistent with their role in scanning for premature stop codons in mRNAs and remodeling of the 3' end of protein-RNA complexes and completion of mRNA decay (Franks et al., 2010). In addition, Upf2 and Nmd4 show strongest binding near the 3' ends of mRNAs. Taken together, the distribution of cross-links along mRNA transcripts differs between degradation complexes and in some cases also between their subunits.”

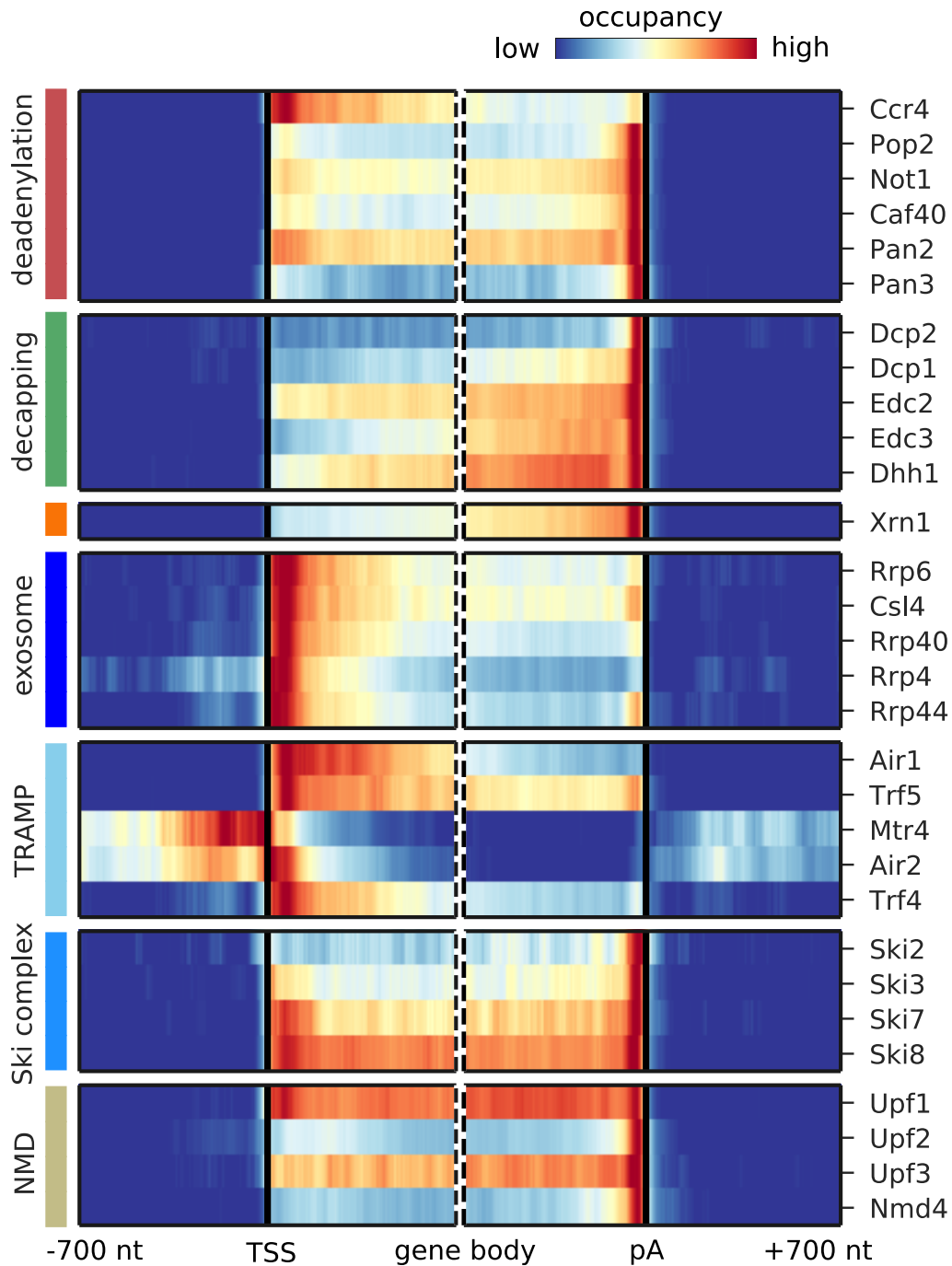


Figure 9: Metagene analysis of degradation factor binding on protein-coding mRNAs. Averaged occupancy profiles of degradation factors over mRNAs aligned around their transcription start site (TSS) (n=3,193, left) and around their poly(A) (pA) site (n=3,193, right) in a window of $[\pm 700\text{nt}]$. Regions that have neighboring transcripts on the same strand were removed to avoid contaminating profiles (Methods). Factors are grouped according to their functional role; from top to bottom: deadenylation, decapping, Xrn1, exosome, TRAMP complex, Ski complex, and NMD. The color code shows the average occupancy normalized between the minimum and maximum values per profile (with high occupancy shown as dark red, and low occupancy shown in dark blue).

4.1.3 Surveillance of aberrant nuclear ncRNA

“Pervasive transcription of the genome leads to many short-lived aberrant RNAs that must be rapidly detected and degraded in the nucleus. We previously reported that the RNA surveillance factors Nrd1 and Nab3 strongly cross-link to aberrant upstream antisense RNA that stems from bidirectional transcription (Schulz et al., 2013). In order to find factors cross-linking to aberrant ncRNAs, we plotted the occupancy of all 30 investigated factors on the antisense strand of known mRNAs (Figure 10). For comparison, we plotted the published Nrd1 and Nab3 profiles in the first two lanes of Figure 10. The factors involved in processing and degradation of Nrd1-terminated transcripts, or NUTs (Schulz et al., 2013) are expected to show similar binding to upstream antisense RNA as Nrd1 and Nab3. Indeed, we observed a similar binding pattern for all exosome subunits (Rrp6, Csl4, Rrp40, Rrp4, Rrp44) and subunits of the TRAMP4 complex (Mtr4, Air2, Trf4). Consistent with this, these factors also bind strongly to previously annotated NUTs and CUTs (Figure 8). It has been shown that Nrd1 is involved in terminating transcripts upstream of the TSS. We also observe a strong signal for binding upstream of the TSS on the sense strand for Air2 and Mtr4 (Figure 9). This suggests that the TRAMP4 complex is involved in degradation of those Nrd1-regulated upstream sense transcripts. To investigate this hypothesis, we compared the binding profiles around the TSS of 459 protein coding genes, previously annotated as having upstream Nrd1-terminated transcripts, or NUTs (Schulz et al., 2013) with the profiles obtained for all mRNAs (Appendix Figure 40). TRAMP4 and the exosome subunits show a strong preference for binding to the upstream promoter region of the genes that are controlled by the Nrd1/Nab3 complex (Appendix Figure 40).”

“These results are consistent with the idea that the nuclear RNA surveillance machinery involves, in addition to Nrd1 and Nab3, the TRAMP4 complex and the nuclear exosome. Indeed, it was reported that TRAMP4 can add a short polyA tail on aberrant RNAs (Wyers et al., 2005), which may trigger degradation by the nuclear exosome. It was also recently shown that Nrd1 and Trf4 interact, providing a basis for coupling surveillance-mediated termination to RNA degradation (Tudek et al., 2014).”

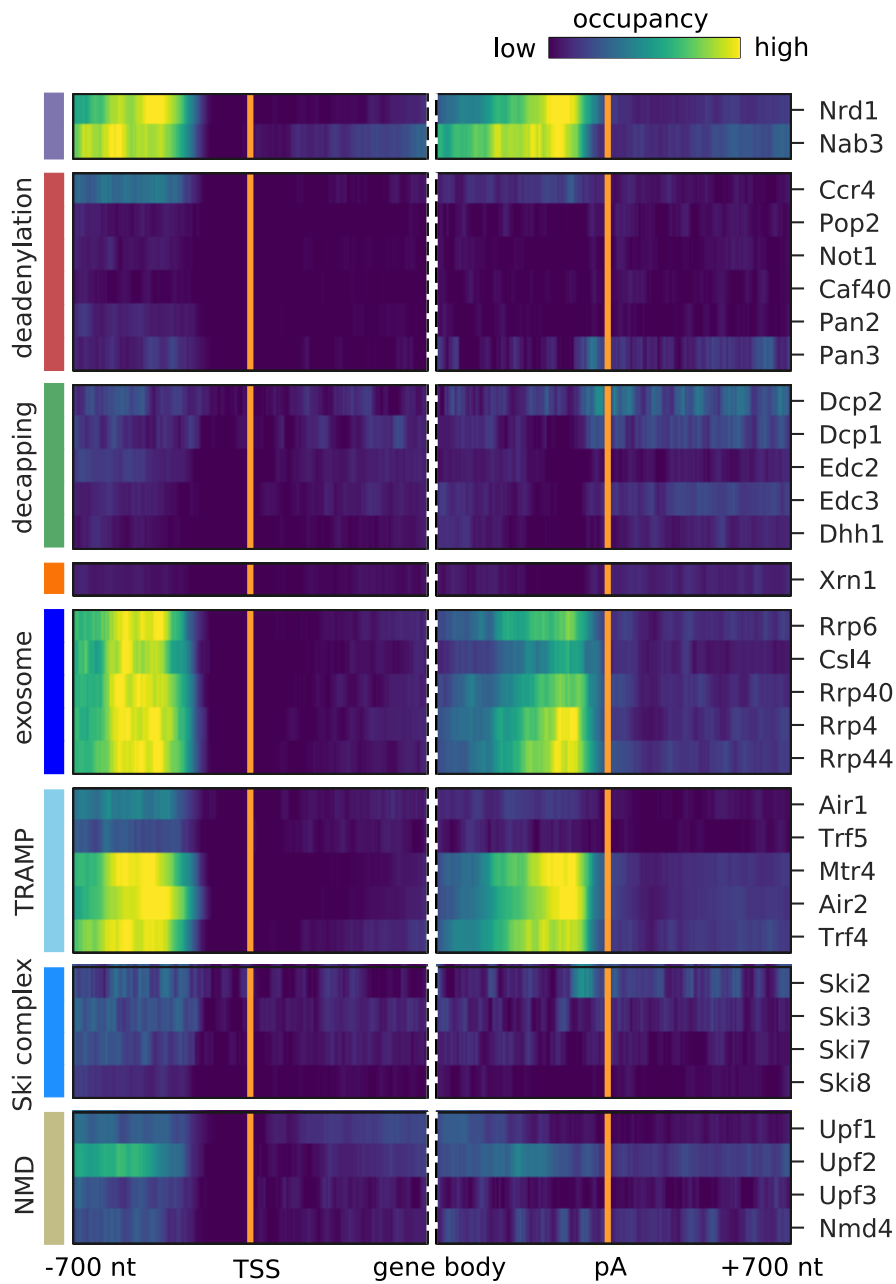


Figure 10: Surveillance of aberrant nuclear antisense RNAs by the exosome and the TRAMP4 complex.

Averaged occupancy profiles of degradation factors binding to transcripts antisense of mRNAs aligned around transcription start site (TSS) ($n=3,076$, left) and around their polyadenylation (pA) site ($n=2,705$, right) in a window of $[\pm 700\text{nt}]$. Regions with annotated genes on the antisense strand are removed to avoid contaminating the profiles (Methods). The color code shows the average occupancy normalized between the minimum and maximum values per profile (with high occupancy shown with yellow, and low occupancy in dark blue). On top, previously published PAR-CLIP profiles for Nrd1 and Nab3 are included for comparison (Schulz et al., 2013).

4.1.4 Interactions between RNA processing machineries

“To find out which groups of factors can work together in degrading transcripts, we analyzed their tendency to co-occupy the same transcripts by calculating the Pearson correlation of their occupancy across all transcripts (Figure 11A). We also analyzed their co-localization, that is, the tendency of a factor to bind near to another factor’s binding sites, using a range of ± 40 nt from each cross-link site (Figure 11B). To relate these profiles to those of other factors, we included previously published PAR-CLIP profiles from our lab (Appendix Table 13). Profiles were available for factors that function in nuclear RNA surveillance (Nrd1, Nab3), cap binding (Cbc2), mRNA transcript elongation (Bur1, Bur2, Ctk1, Ctk2, Cdc73, Ctr9, Leo1, Paf1, Rtf1, Set1, Set2, Dot1, Spt5, Spt6, Rpb1), pre-mRNA splicing (Ist3, Nam8, Mud1, Snp1, Luc7, Mud2, Msl5), pre-mRNA 3’ processing (Pab1, Pub1, Rna15, Mpe1, Cft2; Yth1), transcription termination (Rat1, Rai1, Rtt103, Pcf11), and mRNA export (Hrp1, Tho2, Gbp2, Hrb1, Mex67, Sub2, Yra1, Nab2, Npl3) (Baejen et al., 2017, 2014; Battaglia et al., 2017; Schulz et al., 2013).

A two-dimensional embedding of co-occupancy profiles between all these processing factors is shown in Figure 11C. It represents the degree of similarities between co-occupancy of transcripts (Figure 11A) in terms of the distance in two dimensions. The two-dimensional embedding of the co-localization matrix in Figure 11B shows a similar clustering. This extensive global analysis suggests which factors reside in functional complexes and which functional complexes may interact during RNA processing and degradation. The analysis recovers several established interactions between subunits of known complexes and between different complexes, providing a positive control. For example, all factors of the decapping complex show very high co-occupancy and co-localization, as do Air2 and Mtr4, which reside in the TRAMP4 complex.

The analysis contains a lot of new information, forcing us to focus here on a few interesting, novel findings (Figure 11C). First, the largest cluster is formed by the previously analyzed factors involved in transcription elongation by RNA polymerase II (cluster 1) and in co-transcriptional pre-mRNA processing, including cap-binding complex (Cbc2), 3’ processing, transcription

termination, and RNA export. The degradation factors Ccr4 and Air1 also reside in this cluster, maybe reflecting the role of Ccr4 in transcription elongation (Kruk et al., 2011). A second cluster is formed by splicing factors (cluster 2). Factors involved in nuclear and cytoplasmic exosomal degradation (Rrp6, Csl4, Rrp4, Rrp40 and Rrp44) form a third cluster (cluster 3). Close to cluster 3, we find the TRAMP4 complex subunit Trf4, the elongation factors Dot1, Paf1, Leo1, and the termination factors Pcf11 and Rai1. Rai1 has been shown to detect and remove incomplete 5' cap structures, to subject aberrant pre-mRNAs to nuclear degradation (Jiao et al., 2010).

A fourth cluster is formed by mRNA deadenylation factors together with pA tail binding proteins (Pab1 and Pub1), Ski7, Ski8, Trf5, and the export factor Yra1 (cluster 4). This is consistent with coupled mRNA deadenylation and subsequent degradation from its 3' end by the exosome with the Ski or TRAMP complex as adaptors. The fifth cluster is formed by mRNA decapping factors, which cluster together with Xrn1, suggesting a coupling of mRNA decapping with degradation from the 5' end by Xrn1 (cluster 5). The NMD-involved factors Upf1, Upf2, Upf3 and Nmd4, and Ski2 and Ski3 are also found in cluster 5. The high correlation between Xrn1 and Ski2 has been reported in a CRAC experiment (Tuck and Tollervey, 2013). The elongation factor Ctr9, the 3' processing factor Mpe1 and the export factors Tho2, Mex67 and Nab2 are also found in cluster 5. A last cluster (cluster 6) is formed by factors involved in nuclear RNA surveillance, including Air2, Mtr4 and the Nrd1/Nab3 complex. Taken together, these findings are consistent with known functional associations and physical interactions between factors and suggest intriguing new associations to be investigated in future work."

Transcriptome maps of general eukaryotic RNA degradation factors

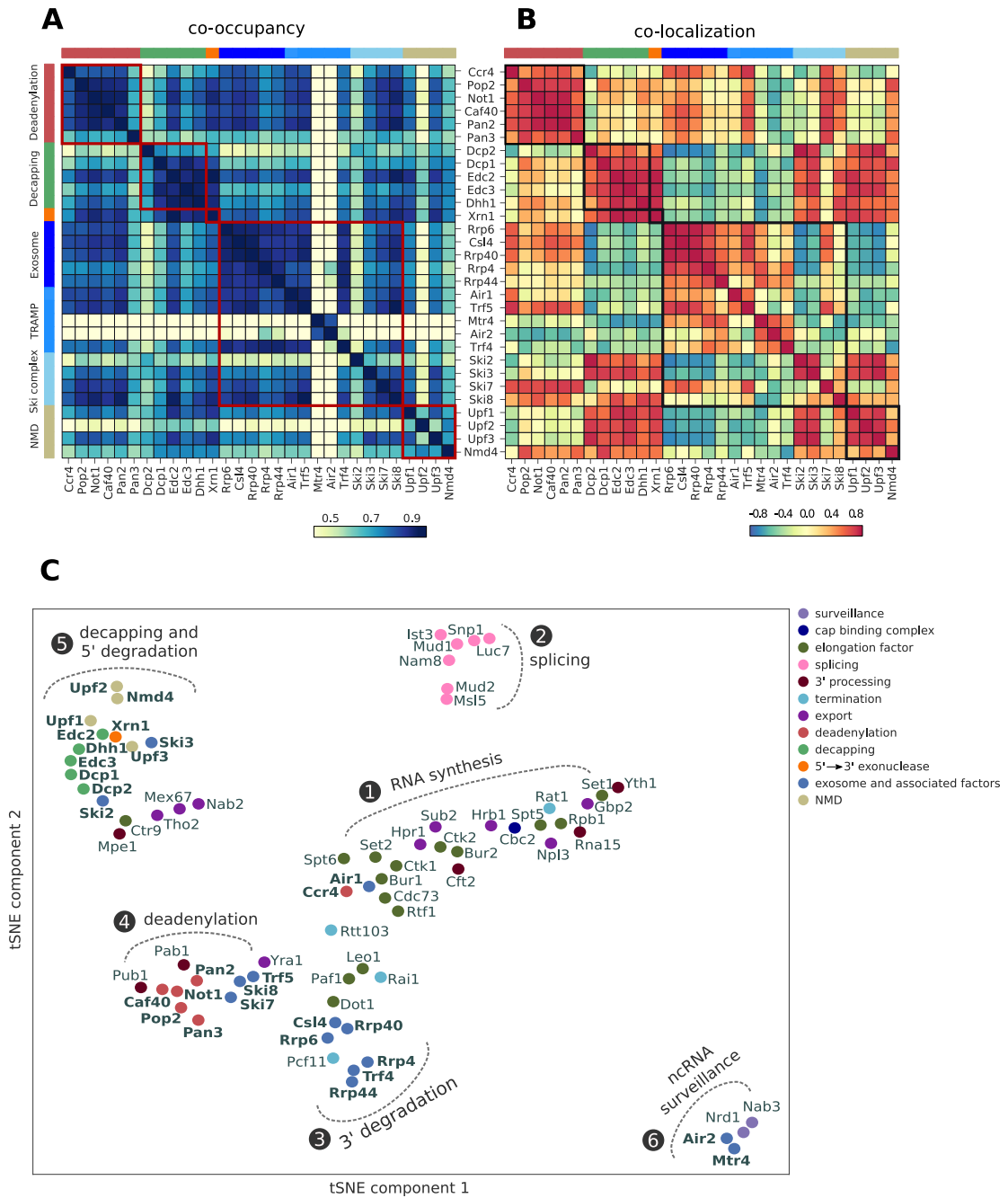


Figure 11: Global co-occupancy and co-localization analysis reveals unexpected cooperation between factors from different complexes and pathways.

A) Matrix of pairwise correlation coefficients of factor occupancies evaluated over all transcripts. B) Matrix of co-localization based on the enrichment of factor x binding within 40 nt upstream and downstream of the cross-link site of factor x' . C) Two-dimensional embedding of the co-occupancies in A) analyzed for 74 RNA processing factors with tSNE, including 30 factors from this study (highlighted in bold), and 44 factors from previous studies (Baejen et al., 2014, 2017; Battaglia et al., 2017; Schulz et al., 2013). Factors that are plotted in close proximity show a preference for binding to the same transcripts. Clusters present factors involved in RNA synthesis (1), splicing (2), 3' processing (3), deadenylation (4), decapping (5), and surveillance (6).

4.1.5 5' degradation machinery senses translation efficiency

“To study the link between cytosolic mRNA translation and degradation, we compared the occupancy of degradation factors on mRNAs to their average codon-optimality score (‘transcript optimality’) (Figure 12A, Appendix Figure 31-Figure 37). We found that the 5' decapping machinery and Xrn1 preferentially bind transcripts with low transcript optimality. In contrast, the 3' deadenylation machinery and the exosome bind more strongly to optimal transcripts. We asked whether this correlation with codon optimality is introduced by only a few differentially bound codons or by global enrichment/depletion of optimal codons. For this purpose, we introduced a ‘codon enrichment score’, which measures a codon’s enrichment in the set of transcripts bound by the factor relative to the yeast mRNA pool. For Dcp2 this enrichment score is high on non-optimal codons, and low on optimal codons, whereas the opposite trend is observed for Ccr4 and most degradation factors (Figure 12B, Appendix Figure 31-Figure 37). This is consistent with a model that ribosome stalling on translationally inefficient codons can lead to recruitment of Dcp2 and Xrn1 and subsequent 5' degradation of the transcript (Heck and Wilusz, 2018).

To investigate the significance of the correlation between transcript optimality and binding of the 5' degradation machinery, we compared the contribution of several mRNA features in explaining the occupancy patterns retrieved from PAR-CLIP experiments. Since mRNA expression, half-life, and translation optimality are inter-correlated (Appendix Figure 30), a causative effect of one of these features on binding strength may lead to correlations with all three features. To better distinguish correlation from causation, we used linear regression analysis to explore whether correlations between factor binding and optimality are better explained with other mRNA features (Appendix Figure 38). We assessed the significance of features via the likelihood ratio test on the multi-variate linear regression model for occupancy. The likelihood ratio test calculates the significance of a feature from the change of the likelihood (quantifying the prediction quality) upon removal of that feature from the regression model. For decapping enhancers (Edc2, Edc3, and Dhh1) and Xrn1, low codon optimality is the most determining feature for binding (Figure 12C). The same is true for NMD factors Upf1 and Upf3, which are

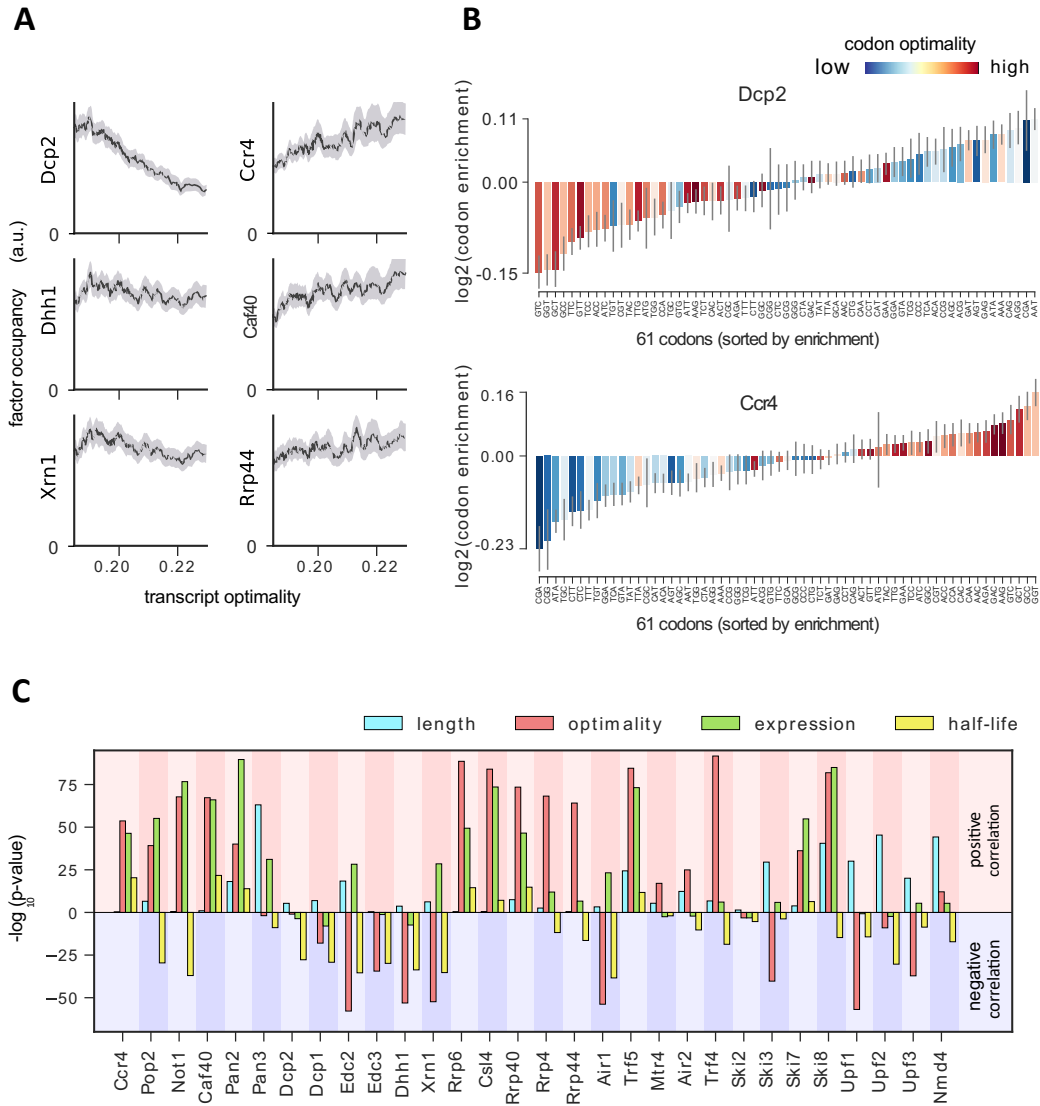


Figure 12: Binding preferences reveal a link between decapping-mediated degradation and translation.

A) Total occupancy per mRNA (according to TIF-seq annotation) for six factors as a function of the average mRNA codon optimality (transcript optimality). The occupancy of factors from the 5'→3' degradation machinery (decapping and Xrn1, left) decreases with increasing transcript optimality, whereas the occupancy of factors from the 3'→5' degradation machinery (Ccr4, Caf40 and exosome subunit Rrp44) increases with increasing average codon optimality. (Grey shading: 95% confidence intervals generated by bootstrapping mRNAs). B) Codon enrichment in transcripts bound by Dcp2 and Ccr4 compared to the average frequency over all mRNAs. The bar colors represent codon optimality, with highly optimal codons shown in dark red. (Thin grey lines: 90% confidence intervals generated by bootstrapping coding sequences.) C) Significance of correlations between the binding strength of degradation factors and transcript length, transcript optimality (Pechmann and Frydman, 2013), expression level (Baejen et al., 2017), and half-life derived by multivariate linear regression analysis (Methods). Bars are separated according to the direction of correlation with positive correlation marked by a red background and negative correlation marked by a blue background.

known to bind non-optimal transcripts (Celik et al., 2017). This result confirms the importance of the translation efficiency for the stability of cytosolic mRNAs and strengthens our finding that transcripts with low average codon optimality are preferentially targeted by the decapping machinery and degraded from the 5' end.”

4.1.6 Decapping factors are enriched upon RNA degradation

“Although decapping occurs at the 5' end of mRNAs, decapping factors show a strong occupancy near the 3' end (Figure 9). To investigate this further, we compared metagene profiles of decapping factors between stable (top 25%) and unstable (bottom 25%) transcripts, using mRNA half-life estimates (Figure 13A, Methods). On both stable and unstable mRNAs, Dcp1, Edc2, Edc3, and Dhh1 show increased binding near the 3' end, but unstable RNAs show a higher occupancy in the transcript body. The catalytically active subunit Dcp2 binds almost exclusively at the 3' end and has a higher occupancy on unstable transcripts. Moreover, A-rich 4-mers are abundant around the proximity (8 nt) of Dcp2-cross-link sites (Figure 13C), indicating a binding preference of Dcp2 for A-rich RNA sequences. Overall, these binding patterns suggest that decapping factors are bound in transcript bodies and near the 3' end of transcripts, and that through closed-loop formation of the mRNA they are in close proximity to the 5' end. Decapping factors might also travel with the 5'→3' exonuclease Xrn1 upon RNA degradation.

Decapping factors may bind to complete mRNAs or to transcripts that are in the process of being degraded. To quantify these two behaviors, we combined our PAR-CLIP occupancy data with RNA half-life estimates (Methods). We modeled the occupancy of factors on mRNA as the sum of binding to all transcripts (b) and surplus binding to transcripts that are in the process of degradation ($\frac{a}{t_{1/2}}$). Therefore, we can model occupancy as a function of half-life with a linear equation (occupancy = $\frac{a}{t_{1/2}} + b$). In cases where there

distribution, respectively). B) Dependence of total occupancy of factors on the transcripts half-life. The fitting function is plotted in red and the fitted value for b is marked with a dashed gray line. (Grey shade: 95% confidence intervals generated by bootstrapping transcripts). C) Sequence binding preference for the catalytically active subunit of decapping complex (Dcp2), illustrated with the 5 most enriched and the 3 most depleted 4-mers. The color code shows the \log_2 enrichment factor of 4-mers around PAR-CLIP cross-link sites [± 5 nt]. Dark red represents strong enrichment and dark blue shows strong depletion of a 4-mer. Infeasible combinations are shown with grey. The most highly enriched field is binding AAAAU with the cross-link at the U, which is enriched over random expectation approximately $2^3 = 8$ -fold.

is no surplus binding upon active degradation, i.e., the occupancy is the same as in intact RNAs, 'a' will be zero. For 5' decapping factors, this model closely fits the occupancy patterns retrieved from our experiments (Figure 13B), other degradation factors also follow this pattern to varying degrees (Appendix Figure 31-Figure 37). In particular, Dcp2 shows a very high a/b ratio, revealing that it cross-links preferentially to transcripts that are being degraded. This analysis strongly suggests that the 5' decapping machinery, although present to some extent on complete mRNAs, is enriched when mRNAs are degraded.”

4.2 Discussion and Outlook

4.2.1 Occupancy profiles of general eukaryotic degradation factors give new insights into RNA degradation processes

In this work we generated transcriptome-wide binding maps for 30 RNA degradation factors in the yeast *S. cerevisiae*. A detailed bioinformatics analysis of the NGS data revealed how degradation factors vary in their binding specificities for different classes of RNAs (Figure 8) and with respect to their preferred locations on RNA transcripts (Figure 9 and Figure 10). Global comparisons of the degradation factor profiles with previously published transcriptome-wide binding maps of other RNA-binding factors (Baejen et al., 2017, 2014; Battaglia et al., 2017; Schulz et al., 2013) revealed factors that co-occupy RNAs or co-localize on RNAs thereby forming clusters (Figure 11). Our data are consistent with published literature as discussed below and extend these findings on mostly single genes to a genome-wide scale. In addition, our analysis revealed several unexpected, novel insights, which we also discuss here. Although our data reflect factor binding signal and measure occupancy on transcripts, and do not directly reveal the function of factors, the correlations of occupancies between factors and with transcript properties indicate functional aspects and suggest functional associations between factors. We provide speculations on possible functional implications of our findings that can guide future studies.

4.2.2 Initial step of RNA decay - mRNA deadenylation

The first step of cytoplasmic mRNA decay is deadenylation of the polyA tail (Cao and Parker, 2003; Parker and Song, 2004). We have analyzed binding of Ccr4-Not and Pan2-Pan3 deadenylation complexes to RNAs by PAR-CLIP and observed similar binding patterns regarding RNA specificity (Figure 8). For the Pan2 subunit we observed cross-link sites throughout the mRNA (Figure 9) suggesting a function of Pan2 over the entire transcript such as scanning for degradation marks. For both subunits of the Pan2-Pan3 complex the occupancy is at the maximum close to the pA site (Figure 9). For the Ccr4-Not

complex we observed RNA binding more towards the 3' end of transcripts, except for Ccr4, which shows stronger cross-linking to the 5' end (Figure 9). It has been proposed that the Pan2-Pan3 complex initiates deadenylation of the polyA tail followed by the Ccr4-Not complex, which acts on the shortened tail (Beilharz and Preiss, 2007; Brown and Sachs, 1998; Tucker et al., 2001). In agreement with this hypothesis, we observed binding of deadenylation complexes at the pA site (Figure 9). One limitation of the PAR-CLIP technique is that factor binding to the polyA tail cannot be resolved. This is due to the inability to map the polyA tail as it is not encoded in the reference genome. Therefore, we are not able to detect binding of the deadenylation factors within the polyA tail directly. Our method only allows for mapping of factors located close to the pA site. With this information we cannot elucidate the order of events during deadenylation.

In addition to the pA site binding of deadenylation factors, we observed binding of the deadenylase Ccr4 to the 5' end of mRNAs (Figure 9), which indicates a different or additional function of this protein as previously suggested (Miller and Reese, 2012). This specific binding pattern and its colocalization with transcription elongation factors (Figure 11) can also reflect its function in transcription elongation by RNA Pol II (Kruk et al., 2011). During deadenylation, it has been shown that Pab1 interacts with the Ccr4-Not complex, thereby stimulating deadenylation activity and differentiating the roles of the nuclease enzymes Ccr4 and Pop2 (Webster et al., 2018). By using a fully reconstituted biochemical system with proteins from the fission yeast *Schizosaccharomyces pombe*, the authors showed that Pab1 release depends on Ccr4 activity. *In vivo* experiments in *S. cerevisiae* showed that Ccr4 acts on all mRNAs and thus, is a general deadenylase. Our PAR-CLIP data revealed Ccr4 occupancy at the 5' end of mRNAs (Figure 9) indicating that its binding to mRNAs is independent of the polyA tail and Pab1. In addition, we observed enriched binding of Ccr4 within introns (Figure 8). This leads us to the speculation that Ccr4 could be recruited to the mRNA 5' end or intronic sequences to release Pab1 from the polyA tail and further its deadenylation. In contrast to Ccr4, Pop2 has been found to just trim the polyA tail, which is not protected by Pab1 (Webster et al., 2018). In agreement with this, we were able

to detect Pop2 binding at the pA site (Figure 9) probably due to the absence of Pab1.

The authors concluded based on their data that Pop2 is a specialized enzyme deadenylating selected transcripts characterized by reduced Pab1 occupancy and lower translation elongation rates (Webster et al., 2018). Lower translation elongation rate indicates Pop2 binding to transcripts with non-optimal codons. In contrast, we observed a minor binding preference of Pop2 on mRNA with optimal codons (Appendix Figure 31). This inconsistency of data may be explained by the usage of two different experimental settings: a whole cell approach for our PAR-CLIP experiments, while the other experiment was performed in an isolated and biochemically reconstituted system. In contrast to Pop2, Ccr4 is bound preferentially to codons with high optimality (Figure 12) suggesting that deadenylation and degradation initiation of the mRNA are independent of a high translational efficiency.

4.2.3 Decapping as first step for RNA degradation from the 5' end

Following deadenylation, decapping is the next step in cytosolic mRNA degradation (Franks and Lykke-Andersen, 2008; Tharun and Parker, 2001). In our study, we observed binding of decapping factors primarily towards the polyA tail (Figure 9), and enriched cross-link sites on mRNAs and non-coding SUTs (Figure 8). The preferred localization of 5' decapping factors near the 3' end and 3' degradation factors close to the 5' end seems counterintuitive. Gallie *et al.* explained this phenomenon by the model of an mRNA closed-loop structure due to messenger ribonucleoprotein (mRNP) formation, in which the 5' cap is in close proximity to the polyA tail (Gallie, 1991). The decapping factors can be deposited near the 3' end of transcripts and upon polyA tail shortening, the decapping complex is activated. This leads to decapping and subsequent rapid degradation of the transcript by the 5' → 3' exonuclease Xrn1. In this model, decapping opens the RNA closed-loop structure, allowing access for Xrn1 and thereby enabling RNA degradation. Supporting this model, we determined that the catalytically active subunit Dcp2 shows the highest activity

on unstable transcripts compared to decapping activators (Figure 13). Our data suggest that decapping enhancers can be pre-bound to stable mRNAs waiting for the recruitment of the catalytic subunit (Figure 13).

Furthermore, we observed higher enrichment of decapping factors on translationally inefficient codons (Figure 12). This agrees with previous findings that suggested a link between RNA decay and translation (reviewed in Huch and Nissan, 2014). This link is thought to be established by the decapping enhancer Dhh1 sensing ribosome velocity. Ribosomes are slowed down on non-optimal codons, which reduces the translation rate and this can lead to activation of deadenylation and decapping (Radhakrishnan et al. 2016).

We observed further cross-link sites for Xrn1 throughout the transcript with higher occupancy towards the 3' end (Figure 9). Binding at the pA site has been observed for Xrn1 before using a similar approach (Tuck and Tollervey, 2013). This binding pattern may be explained by slower decay towards the transcript 3' end allowing for higher detection.

4.2.4 Nuclear RNA surveillance mechanism

Genomes of eukaryotic cells are pervasively transcribed with about 85% of the yeast genome giving rise to RNA transcripts, although only a few percent correspond to protein-coding mRNAs (David et al. 2006). In yeast, pervasive transcription stems from bidirectional Pol II transcription initiation of two opposing pre-initiation complexes within a nucleosome free region over the promoter region (Murray et al., 2012; Rhee and Pugh, 2012). The work from our laboratory has previously shown that the Nrd1-Nab3 surveillance machinery selectively terminates ncRNA synthesis, including transcripts synthesized in antisense direction to annotated genes and divergent transcription from bidirectional promoters (Schulz et al. 2013). With global mapping of degradation factors and in comparison to the metagene profiles of Nrd1-Nab3 (Schulz et al., 2013), we are able to propose a mechanism on how the surveillance machinery degrades antisense RNA in order to protect the cell from aberrant ncRNAs after early transcription termination. Several lines of evidence such as enrichment of these complexes on NUTs (Figure 8), similar

binding profiles (Figure 10), as well as transcript co-occupancy analysis (Figure 11) suggested that after recognition through the sequence specific Nrd1-Nab3 complex, the ncRNA is polyadenylated by the TRAMP4 complex. The interaction of the Nrd1-Nab3 complex and Trf4 subunit of the TRAMP4 complex has previously been shown by our group and colleagues (Tudek et al., 2014). The short polyA tail can then mark the RNA for degradation by the nuclear exosome. This mechanism is likely to be the same for regulation of introns and ncRNAs upstream of mRNAs on the same strand, which were annotated as NUTs and CUTs, for which we see similar occupancy of the same degradation factors (Figure 8 and Figure 40). Our results indicate that the degradation machinery for all short-lived ncRNAs in the nucleus requires the same degradation factors such as Nrd1-Nab3, TRAMP4 and the nuclear exosome.

4.2.5 RNA degradation by the exosome complex and auxiliary factors

The second major cytoplasmic mRNA degradation pathway in the 3' → 5' direction of transcripts involves the exosome complex (Zinder and Lima, 2017). In addition to mRNA decay, the exosome also targets long-lived transcripts such as tRNAs, rRNAs, snoRNAs, and snRNAs for processing and decay (Zinder and Lima, 2017). Our observations regarding the different binding patterns of the exosomal subunits and co-factors (Figure 8 and Figure 9) support the hypothesis that these factors are needed for RNA processing and degradation specificity (Delan-Forino et al., 2017). It has been proposed that exosome subunit localization regulates RNA degradation targets (Vanacova and Stefl, 2007). The RNA transcript can be channeled through the exosome core (Rrp4, Rrp40 and Csl4) towards the catalytic subunit Rrp44 (Kowalinski et al., 2016). This suggests that all factors have direct RNA contacts, which we were able to support by observing extensive crosslinking to mRNAs and many other non-coding transcripts (Figure 8). All exosome subunits cross-link towards the 5' end of mRNAs (Figure 9). This unexpected observation might be caused by the exosome moving rapidly from the 3' to the 5' end and then

residing at the 5' end for a longer time, enabling for extensive cross-linking and thus detection.

The exosome co-factor Ski2 and the other subunits of the Ski complex show strong cross-linking at the pA site of mRNA (Figure 9). This indicates that the Ski complex is necessary during initial steps of RNA degradation. This can be explained by the helicase activity of Ski2, which dissolves RNA secondary structures and detaches proteins from the RNA allowing the exosome to degrade the transcript from the 3' end (Schneider and Tollervey, 2013).

The TRAMP complexes are additional exosome co-factors and show binding to mRNAs towards the 5' end of transcripts similar to the core and catalytic subunits of the exosome (Figure 9). This indicates that TRAMP complex mediated targeting of defective nuclear mRNAs and aborted transcription products for exosomal degradation might be similar to the Ski complex in the cytoplasm. The helicase Mtr4 may have a similar function as the Ski2 helicase in dissolving the secondary structure of the RNA and disassembling protein complexes bound to RNA (Falk et al., 2014). Additionally, TRAMP4 binding extended upstream of the TSS (Figure 9). This suggests targeting of non-coding transcripts emerging upstream of promoters for degradation. We have shown that this signal stems from previously annotated transcripts under attenuation control by Nrd1-Nab3 (Schulz et al., 2013), the TRAMP complex and the nuclear exosome (Appendix Figure 40). Moreover, it has been shown for the TRAMP complexes that they function in 3' end processing and degradation of snoRNAs (Grzechnik and Kufel, 2008; Losh et al., 2015). However, it has not been fully elucidated whether the complexes share the functions or distinguish between them. Using metagene profiles on snoRNAs, we were able to distinguish the functions of the TRAMP complexes (Appendix Figure 39). We propose that the TRAMP4 complex due to its binding downstream of the 3' end of snoRNAs functions mainly in 3' end processing. Binding of the TRAMP5 complex within the gene body of snoRNAs suggest a function mainly in its degradation.

4.2.6 Cytoplasmic RNA surveillance by NMD factors

The NMD factors have a function in scanning for premature stop codons in mRNAs, remodeling of the 3' end of protein-RNA complexes and completion of mRNA decay (Franks et al., 2010). This is consistent with our observation that Upf1 and Upf3 bind to the entire length of the mRNAs and with highest occupancy at the pA site (Figure 9). Moreover, Upf2 and Nmd4 showed the strongest binding near the 3' ends of mRNAs close to endogenous stop codons. This indicates an additional role for NMD factors in recognition of the translation termination site.

The NMD-involved factors Upf1, Upf2, Upf3 and Nmd4 reside in a cluster together with mRNA decapping factors and the 5'→3' exonuclease Xrn1 (Figure 11C). This indicates that recognition of premature stop codons by the NMD machinery is coupled with initiation of decapping and degradation from the 5' end by Xrn1. It has been shown that the NMD machinery is directly linked to the decapping complex via an adapter protein (Cho et al., 2009; Lai et al., 2012). Furthermore, we showed binding for the NMD factors Upf1 and Upf3 to non-optimal codons (Figure 12C). This is consistent with previous findings that NMD factors bind to non-optimal transcripts (Celik et al., 2017). This result confirms the importance of NMD factor binding for the stability of cytosolic mRNAs.

Furthermore, unannotated transcripts, which were predicted to lack protein-coding capacity in yeast, have been shown to be translated and targeted for degradation by the NMD pathway (Smith et al., 2014). Consistently with this observation, we observed binding of the NMD factors on SUTs (Figure 8). In addition to cross-linking sites in SUTs, we detected binding of the NMD machinery to unstable ncRNAs such as NUTs (Figure 8). This unexpected finding indicates that some NUTs escape degradation in the nucleus by the Nrd1-Nab3, TRAMP and exosome complex using an unknown mechanism. After export of the NUTs to the cytoplasm they might be targeted for translation-dependent degradation via the NMD pathway.

4.2.7 Perspectives of mechanistic insights into the RNA degradation process by complementation of the transcriptome maps with functional studies

We provide binding profiles of 30 RNA processing and degradation factors as an extensive resource that gives novel insights into eukaryotic RNA metabolism. The findings of this study provide a better understanding of nuclear degradation of short-lived ncRNAs. In addition to regulation of mRNA degradation by decapping and Xrn1-mediated 5'→3' decay mainly on translation inefficient codons. Several questions remain to be answered. Further biochemical experiments are required to study and prove the indications of functional interactions of short-lived ncRNAs and the involved degradation machinery. The proposed slowing down of the different degradation machineries (exosome complex and Xrn1) towards either end of the transcript needs to be investigated with kinetic measurements *in vivo* using single molecule techniques or *in vitro* with a fully reconstituted system to confirm our hypotheses. Further research needs to be undertaken to get a more detailed understanding of how the different long-lived ncRNAs such as tRNAs, rRNAs, snoRNAs, and snRNAs can be processed and finally degraded by the same exosome complex with different nuclear and cytoplasmic accessory factors. Structural analysis of the different exosome components and accessory proteins with their target RNAs might help to answer this question as the structures of the exosome and accessory factors alone cannot answer all questions (Falk et al., 2014; Halbach et al., 2013; Schmidt et al., 2016). In conclusion, the findings of this study provide several important insights and establish a basis for future research.

4.2.8 Protein-RNA binding studies of eukaryotic degradation factors in response to stress conditions will allow for better functional characterization of the RNA degradation machinery

Yeast cells undergo major metabolic adaption processes when they encounter stress situations, such as changes of the nutrient environment. To

conserve energy, cells shut down gene expression of growth-related genes rapidly. At the same time, they increase expression levels of stress response and other metabolic genes. This leads to genome-wide changes in transcription and translation. Transcription levels are drastically reduced and translation is almost abolished in starved yeast cells (Ashe et al., 2000; Jona et al., 2000). At the same time, cytoplasmic degradation levels are also reduced and mRNAs are stabilized at a stage previous to polyA tail shortening (Jona et al., 2000). Down-regulation of growth-related genes in the nucleus is controlled by the Nrd1-Nab3-Sen1 and the TRAMP complexes targeting these genes for exosomal degradation (Bresson et al., 2017). Other genes like stress response genes escape degradation by decreased binding of Nrd1-Nab3 and TRAMP complexes, thereby resulting in upregulation of transcription. These findings have been obtained by genome-wide binding profiles of the Nab3 and Mtr4 proteins (Bresson et al., 2017). In addition, it has been observed that deadenylation by Ccr4 and Pan2 is inhibited in order to retain the majority of the cytoplasmic mRNA pool for later reuse after stress recovery (Hilgers et al., 2006). Nevertheless, how cytoplasmic degradation factors respond to nutrient changes has not been shown genome-wide. RNA binding studies using PAR-CLIP, in response to different stress conditions and in combination with RNA-seq measurements are required to study the feedback of various cytoplasmic degradation factors. Although inhibited, the degradation factors might show differential RNA binding behavior during stress response (Jona et al., 2000). The activity of Dcp2 is particularly interesting to analyze after stress induction as it showed the strongest binding to unstable transcripts in our study (Figure 13).

4.2.9 Conservation of general RNA degradation factors in human cells

Many RNA degradation proteins like the decapping and deadenylation factors, the TRAMP complex, the core exosome complex and associated factors are conserved amongst eukaryotes (Houseley and Tollervey, 2009). However, the RNA degradation mechanism in human is far less understood than in yeast. In order to compare different eukaryotic RNA degradation

pathways, it would be useful to perform PAR-CLIP of key players in RNA decay like the 5'→3' exonuclease Xrn1, the exosome catalytic subunits Rrp6 and Rrp44, as well as the NMD factor Upf1 in human cells. This would allow for a similar analysis as in yeast, regarding binding preference of specific RNA classes, preferred binding sites on mRNAs and antisense transcripts using metagene profiles. Additionally, the comparison with other published RNA binding profiles of human protein complexes could be helpful to further understand the different RNA decay pathways in human cells. In human cells, many more factors are involved in RNA degradation, which makes the understanding and experimental set-up to study RNA decay much more difficult when compared to yeast. In addition, some factors with homologous function have not even been discovered in human cells. For example the factors, which harbor homologous activities like Nrd1-Nab3 for nuclear surveillance in yeast, are assumed for human, but have not been uncovered yet, despite much effort (Arigo et al., 2006; Thiebaut et al., 2006). Systematic analysis of RNA binding factors and their mapping on transcripts might help to identify proteins with Nrd1-Nab3 homologous activities by studying the binding profiles on antisense transcripts. Improved methods such as enhanced crosslinking and immunoprecipitation (eCLIP) may provide a suitable set-up for large-scale and robust profiling of RNA binding proteins (Van Nostrand et al., 2016).

5 Identification and functional characterization of the novel mRNA modification *N*³-methylcytidine (m3C)

5.1 Results

5.1.1 Experimental set-up for Trm140 PAR-CLIP and m3C-CLIP

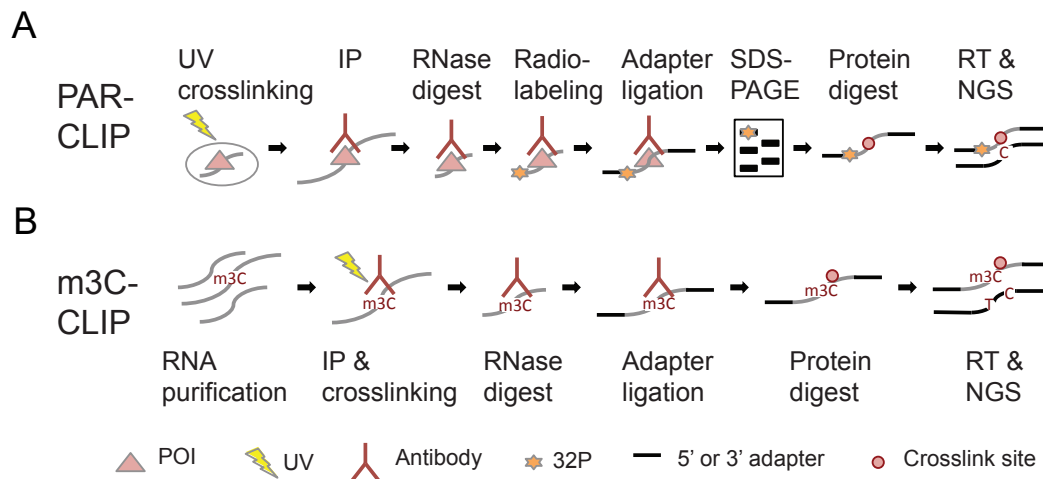


Figure 14: Schematic overview of PAR-CLIP and m3C-CLIP protocol.

A) PAR-CLIP was performed using *S. cerevisiae* or human cells cultured in the presence of 4tU or 4sU, respectively. The proteins were cross-linked to RNA by UV light at a wavelength of 365 nm. After cell lysis, the protein of interest was immunoprecipitated with antibody-coupled beads. The crosslinked RNA was fragmented and labeled radioactively using ³²P-ATP. Adapter ligation was performed on the beads. The protein bound to RNA was purified via SDS-PAGE and the protein-RNA complex corresponding to the predicted size was electro-eluted from the gel. The protein was digested and RNA was converted into cDNA for NGS. B) The protocol for m3C-CLIP was performed with purified and 4tU/4sU-labelled RNA from yeast or human cells. The antibody used for immunoprecipitation against the m3C modification was crosslinked to the RNA by UV light at the wavelength of 365 nm. The RNA not protected by the antibody was digested and sequencing adapter was ligated to the RNA on the beads. The antibody was digested and recovered RNA was subjected to NGS library generation.

I performed PAR-CLIP of Trm140 and m3C-CLIP as depicted in Figure 14 in the yeast *S. cerevisiae*. The PAR-CLIP experiment was performed as previously published with small adjustments (Baejen et al., 2017, 2014; Battaglia et al., 2017; Creamer et al., 2011). I have adapted the published protocol of PA-m6A-seq to perform m3C-CLIP (Chen et al., 2015). PAR-CLIP and m3C-CLIP protocols overlap in the following steps: by having the medium

supplemented with 4-thiouracil (4tU), UV crosslinking at 365 nm, and the same library preparation and sequencing strategy. Differences in the mentioned techniques are that immunoprecipitation (IP) was performed against the TAP-tag of a protein crosslinked to RNA in the PAR-CLIP experiment compared to m3C-CLIP, where previously purified RNA was directly immunoprecipitated with an anti-m3C antibody and afterwards the antibody-RNA complex was UV-crosslinked. For PAR-CLIP, I performed an additional radioactive RNA labeling and purification step via SDS-PAGE as visualized in Figure 15, which is not needed for the m3C-CLIP protocol. The radioactive signal of the RNA bound to the Trm140 protein appears at the predicted size as verified by Western Blot. Additional radioactive signals were observed for the antibody heavy and light chains. The gel fragment (1 cm x 0.5 cm) with the highest radioactive intensity including the RNA bound to Trm140 was used for further experimental procedure.

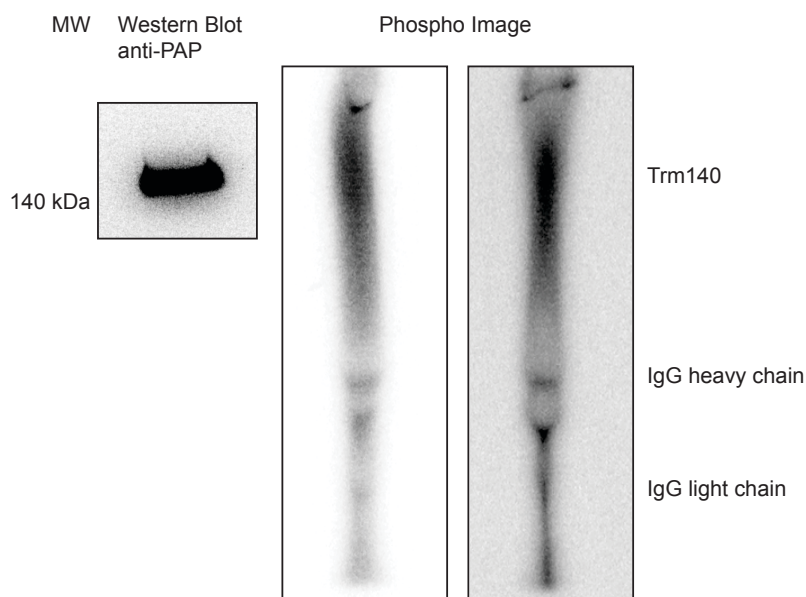
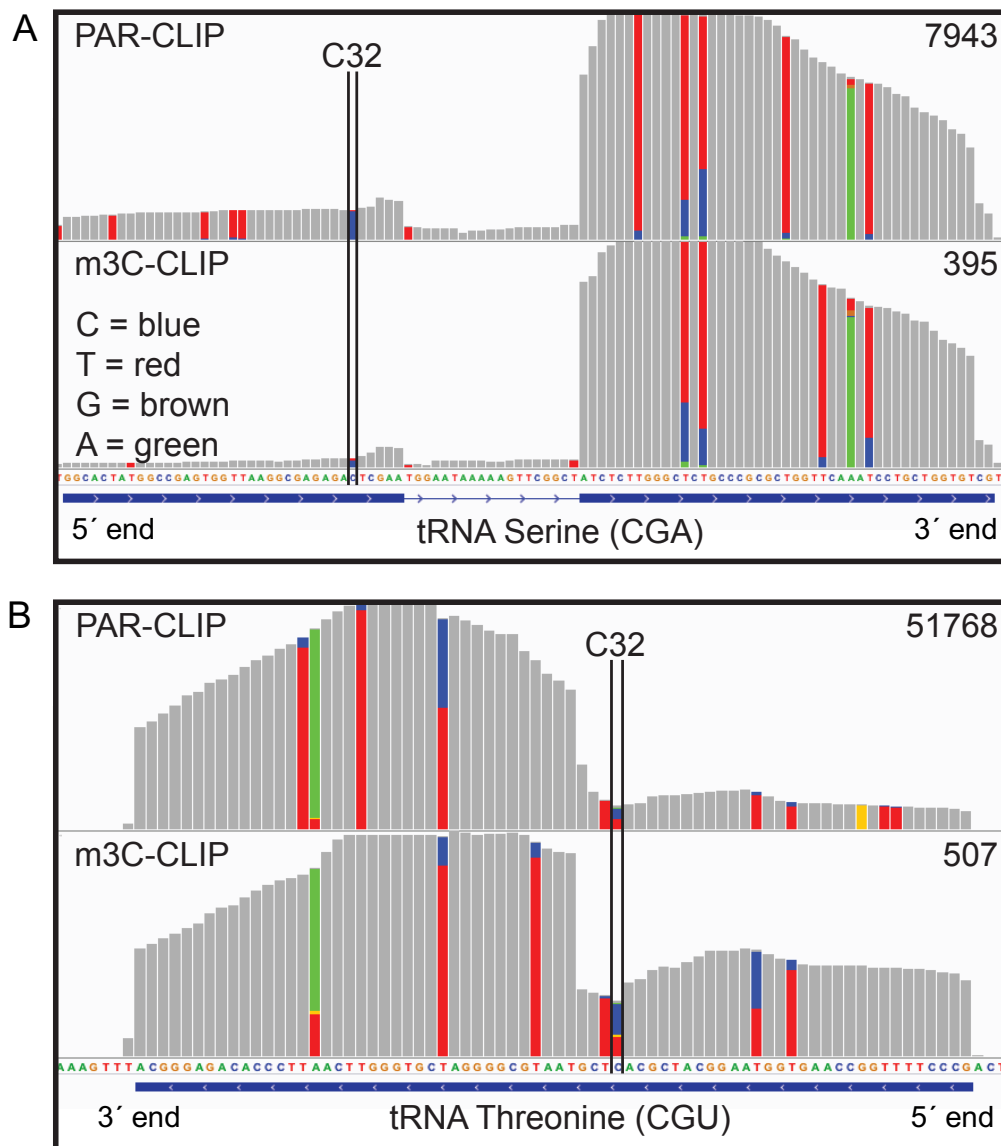


Figure 15: PAR-CLIP of Trm140 analyzed by Western Blot and phosphor imaging. PAR-CLIP of Trm140-TAP was verified by Western Blot as a size marker. SDS-PAGE was performed with NuPAGE 4-12% Bis-Tris gels in MOPS buffer. The size separated proteins were transferred onto a PVDF membrane. The protein of interest was visualized using anti-TAP (PAP) antibody coupled to HRP and chemiluminescence reaction. The radioactively labeled RNA bound to the protein of interest was visualized by phosphor imaging after gel electrophoresis (shown for two independent biological replicates).

5.1.2 Sequencing coverage of Trm140 PAR-CLIP and m3C-CLIP at established m3C modification sites

In order to cross validate the Trm140 PAR-CLIP and m3C-CLIP experiments, I compared the sequencing read coverage at the established modification site in tRNA Serine and Threonine by using the integrated genomics viewer (IGV) tool (Broad Institute) in Figure 16. I detected typical UV cross-linking induced thymine (T) to cytosine (C) transitions in the variable stem loop for Serine and in the T Ψ C-loop for both analyzed tRNAs in our experiments. The T \rightarrow C transition is shown in the IGV tool as blue bar for C compared to the encoded T (red) in the reference genome. The height of the colored bar represents the ratio of the mismatch compared to the encoded base. The UV induced cross-linking sites observed in the Trm140 PAR-CLIP experiment indicate binding of Trm140 to both tRNAs. The UV induced cross-linking sites observed in the m3C-CLIP experiment are caused by m3C antibody binding to the tRNAs.

Interestingly, at position 32 (black line), where tRNA Serine and Threonine carry the m3C modification, a high C \rightarrow T conversion of 4% and 24% was observed for tRNA Serine and 46% and 36% for tRNA Threonine in Trm140 PAR-CLIP and m3C-CLIP experiments, respectively. The m3C modification either induces C \rightarrow T mutational signature or reverse transcription (RT) block as previously published (Arimbasseri et al., 2015; D'Silva et al., 2011). These observations showed enrichment of m3C containing transcripts by Trm140 PAR-CLIP and ability of the m3C antibody to immunoprecipitate transcripts harboring the m3C modification.



*Figure 16: Identification of mutational signature of *N*³-methylcytidine (m3C) modification.*

A) Comparison of Trm140 PAR-CLIP and m3C-CLIP sequencing reads visualized using the IGV tool (Broad Institute) on position C32 of tRNA Serine (CGA) known to be m3C modified by Trm140. C(blue)→T(red) conversion at position 32 indicates m3C presence. Additional conversions (T→C) show PAR-CLIP typical UV induced cross-linking transitions. The DNA coding for the tRNA is located on the Watson strand and highlighted in blue, with the 5' end at the left and the 3' end on the right. The A(green)→T transition is caused by m1A modification on the respective position shown before (Arimbasseri et al., 2015) B) Trm140 PAR-CLIP and m3C sequencing reads over tRNA Threonine (CGU) (located on the Crick strand: 5' end on the right, 3' end on the left) show C→T conversion induced by m3C on C32. The T→C mutational signatures are caused by UV induced cross-linking. tRNA Threonine also harbors the m1A modification at A58 represented by the typical A→T transition (Arimbasseri et al., 2016).

5.1.3 Reverse transcription signature of m3C modification

The m3C mutational signature has been shown previously as a C→T transition or reverse transcription (RT) block (Arimbasseri et al., 2015; D'Silva et al., 2011). An RT block can be explained as a drop of the reverse transcriptase activity during cDNA synthesis due to a bulky RNA structure or the presence of a modified nucleoside. The drop of read coverage next to the modification site in our sequencing data for Trm140 PAR-CLIP and m3C-CLIP (Figure 16) also suggests RT block at the site of the m3C modification. In order to confirm that the RT block is caused by the m3C modification, I performed primer extension assay using a gene specific FAM-labeled primer for tRNA Serine (CGA) and different RNA samples. The products were visualized by denaturing PAGE and fluorescent imaging (Figure 17A). An RT block due to m3C modification results in a shorter fragment, which was observed with *S. cerevisiae* WT RNA. The shorter fragment was absent in the sample using RNA from a Trm140 KO strain lacking the m3C modification. The same RT products as with the Trm140 KO RNA were observed for the *in vitro* demethylated WT sample by AlkB. The AlkB enzyme is a demethylase specific for m1A and m3C modification derived and purified from *E. coli* (Aas et al., 2003). The shorter fragment caused by m3C RT block was again present in the *in vitro* AlkB demethylated and subsequently *in vitro* Trm140 re-methylated sample. As a negative control, no RT product was present in the primer alone condition.

The alignment of the RT primer on the sequence and structure of tRNA Serine (CGA) is shown in Figure 17B. The RT block can be explained by the additional methyl group on N-3 of the cytidine as shown in Figure 6 interfering with Watson-Crick base pairing. Additionally, I observed an RT block upstream of the m3C modification site (Figure 17A). This second RT block is most likely due to the N-2-N-2-dimethylguanosine (m²₂G) modification at G26 (Arimbasseri et al., 2016; Chan and Lowe, 2009). The full-length cDNA product is present above the m²₂G RT block for all conditions, except the negative control.

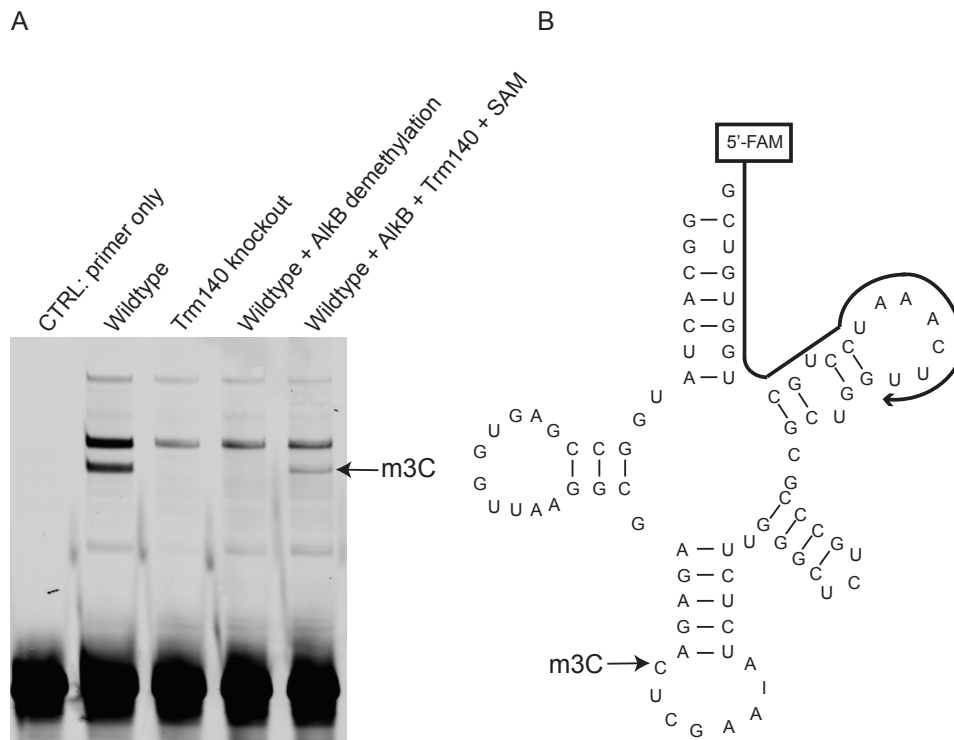


Figure 17: m3C modification causes reverse transcription block shown by primer extension assay on tRNA Serine.

A) Reverse transcription assay was performed with a FAM labeled RT primer complementary to the tRNA Serine (CGA) with different templates: without RNA as negative control (CTRL), WT RNA, Trm140 KO RNA, demethylated and re-methylated RNA. The RT product for each sample was separated on a denaturing 15% Urea gel for 1 h at 300 V. At the expected size, I observed a band in the WT and remethylated sample showing m3C dependent RT block. The RT block of the established modification site of N-2-N-2 dimethylguanosine (m²₂G) at G26 (above m3C band) does not change among tested samples. B) Schematic overview of tRNA Serine and RT primer binding (FAM-labeled) and location of the m3C modification at position C32 in the anticodon loop close to the anticodon (34-36).

5.1.4 Identification of m3C modification in tRNA Arginine (CCU)

In addition to the established modification sites, I observed many UV induced cross-link transitions on all three uridine residues in the TΨC-loop of tRNA Arginine with anticodon CCU in the Trm140 PAR-CLIP experiment (Figure 18). This was also shown on established m3C modification sites in tRNA Serine and Threonine. A high number of UV induced cross-link transitions at the same position suggest strong binding of Trm140 to tRNA Arginine.

Additionally, I observed m³C-CLIP sequencing reads on tRNA Arginine with anticodon CCU in both wildtype (WT) replicates. Interestingly, C at position 32 had a high transition rate to T of about 12% in the Trm140 PAR-CLIP experiment. This mutational signature has been described for the m³C modification (Arimbasseri et al., 2015). Furthermore, I observed a high C→T transition rate in the m³C-CLIP experiment with WT RNA of 20% and 42% for the two replicates. The control m³C-CLIP experiments performed in two replicates with RNA from a Trm140 KO strain showed only 1% and 3% of C→T transition frequencies. This demonstrates that the mutational signature on C32 of tRNA Arginine with anticodon CCU was significantly reduced in the Trm140 KO condition. In the PAR-CLIP experiments of Trm140 we also observe that the read coverage drops next to the m³C modification site due to a potential RT block. These findings indicate that tRNA Arginine is bound by Trm140 and therefore m³C modified on position C32 in *S. cerevisiae*, in addition to the established m³C modified tRNA Arginine in higher eukaryotes (Arimbasseri et al., 2015). In order to verify the m³C modification on tRNA Arginine with anticodon CCU *in vitro*, I performed a primer extension assay using a FAM-labeled RT primer for this particular tRNA. Alignment of the RT primer to the structure and sequence of tRNA Arginine is shown in Figure 18C. The RT assay products were subjected to denaturing PAGE and revealed an RT product corresponding to a shorter fragment in the WT sample in Figure 18B, which was absent in the Trm140 KO condition. Furthermore, it was also absent in the *in vitro* AlkB demethylated sample and present in the *in vitro* Trm140 re-methylated sample (previously demethylated by AlkB). No RT product was present in the primer alone and no RNA control conditions. The fragment caused by RT block due to m²₂G26 modification did not change among other tested samples. Taken together, these results show that the RT block is most likely due to the presence of the m³C modification at position C32 in tRNA Arginine (CCU).

Identification and functional characterization of the novel mRNA modification *N*³-methylcytinde (m3C)

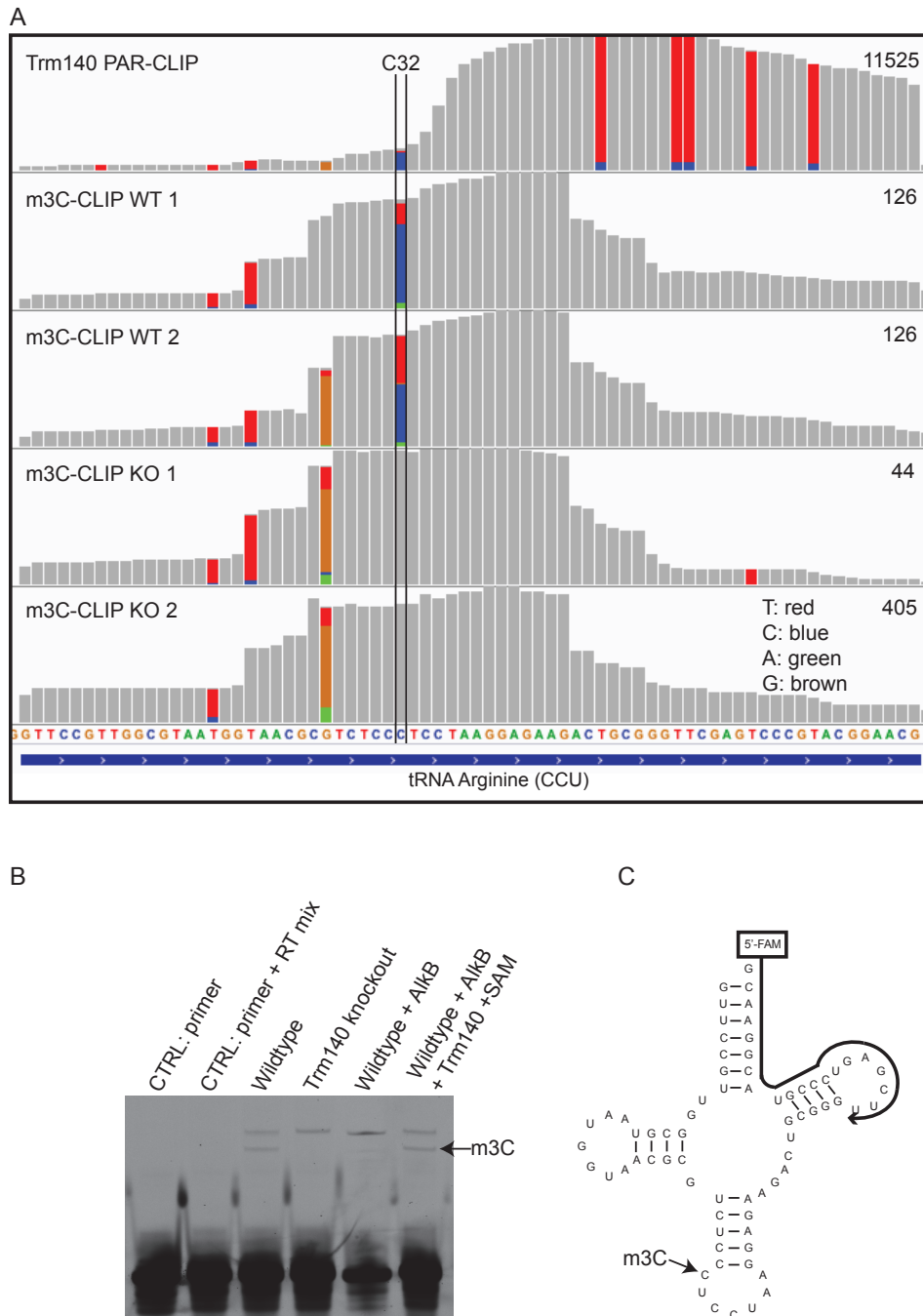


Figure 18: tRNA Arginine (CCU) contains m3C at C32 in *S. cerevisiae*.

A) Trm140 PAR-CLIP showed a high C→T conversion rate on C32 of tRNA Arginine (CCU). Binding of Trm140 to the T rich TΨC-loop of tRNA Arginine is represented as UV-induced T→C conversions was visualized using the IGV tool (Broad Institute). The replicates for m3C-CLIP performed with WT RNA showed high C→T conversion, whereas the experiment with RNA from the Trm140 KO strain had lower levels of conversion. B) Reverse transcription assay with FAM-labeled primer specific for tRNA Arginine showed a fragment caused by RT block with WT RNA and remethylated RNA, which was absent in Trm140 knockout RNA and demethylated RNA indicating Trm140 dependent m3C modification on tRNA Arginine. C) Structure and sequence of tRNA Arginine. m3C modification is indicated with black arrow. FAM-labeled RT primer binding is shown with black arrow starting from the first nucleotide of the 3' end.

5.1.5 Mapping of m3C modification sites in mRNA

To evaluate Trm40 binding on mRNA and m3C modification enriched sequencing reads in mRNA, we performed metagene analysis on mRNA aligned at the TSS and pA site for Trm140 PAR-CLIP and m3C-CLIP. UV induced cross-linking sites in mRNAs indicate binding of Trm140 within mRNAs and potential m3C sites. Metagene analysis of Trm140 PAR-CLIP and m3C-CLIP (Figure 19A) showed binding of the methyltransferase to mRNA and m3C-CLIP signal over the protein coding region. Comparing m3C-CLIP to Trm140 PAR-CLIP, we could see binding to the gene body for the first 250 nt of genes and at the end of genes the m3C-CLIP and the Trm140 PAR-CLIP signal is at its maximum.

We analyzed the mutational signatures of C→T transitions caused by m3C on a genome wide level. This approach identified potential modification sites within mRNA. A list of high-confidence m3C modification sites is shown in Appendix Table 14. Figure 19B shows an example of C→T transition at the same position present in Trm140 PAR-CLIP and m3C-CLIP experiments on mRNA Tra1 (genomic position chrVIII:303572). The transition rate was 22% in PAR-CLIP and 33% for m3C-CLIP. Another example of high C→T transition at the same position for both experimental set-ups is shown in Figure 19C for the mRNA Rrs1 (genomic position chrIV:1153577) close to the translation start codon. In this case the conversion rate for Trm140 PAR-CLIP and m3C-CLIP is 69% and 50%, respectively. In Figure 19D the third example of typical m3C mutational signature is shown on mRNA YCR024C-B (genomic position chrIII:162694) in both experiments close to the stop codon of the transcript. The transition rate is 86% for the Trm140 PAR-CLIP and 80% for the m3C-CLIP experiment. In summary, these results indicate the presence of m3C modification all over the protein coding transcripts. However, metagene analysis of Trm140 PAR-CLIP and m3C-CLIP suggest higher occupancy of the m3C modification at the end of mRNA transcripts.

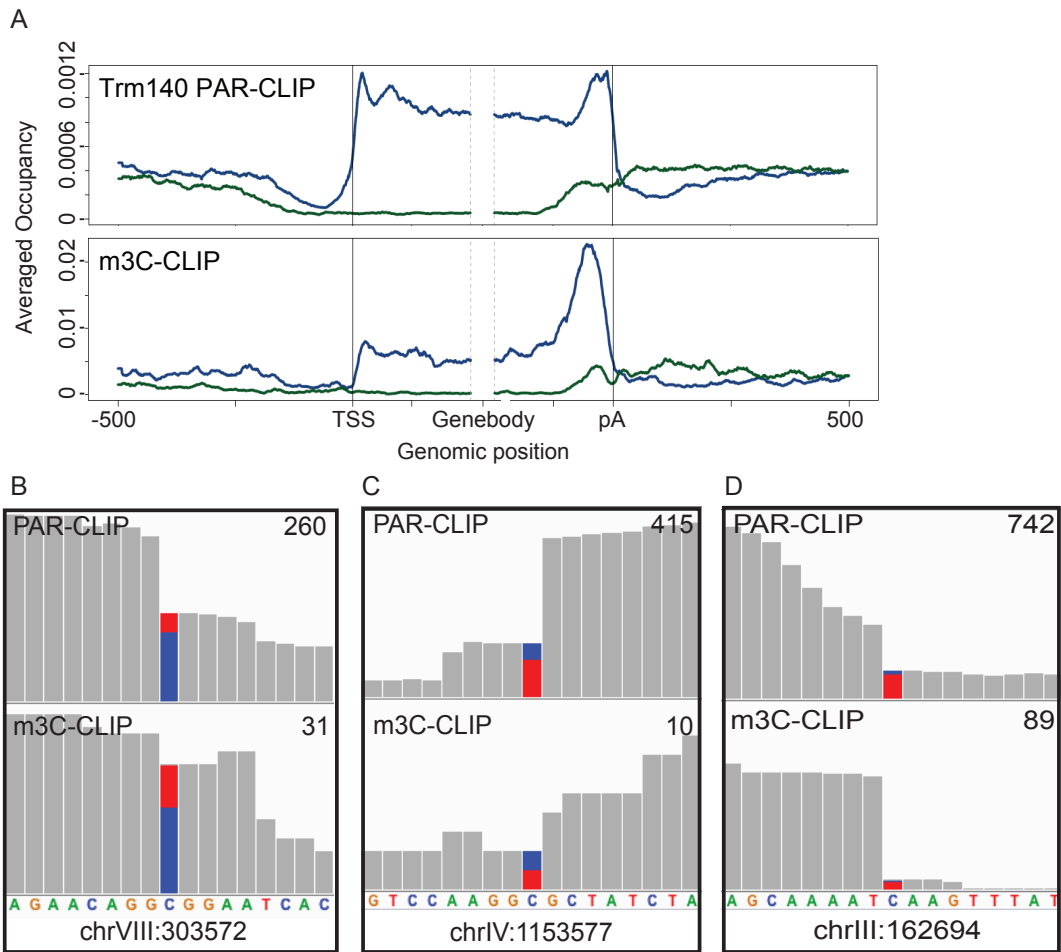


Figure 19: Presence of N^3 -methylcytidine modification in mRNA

A) Metagenome analysis of Trm140 PAR-CLIP and m3C-CLIP on mRNAs. PAR-CLIP signal over protein coding transcripts with peaks close to TSS and pA site. m3C-CLIP showed signal over the protein coding genes with its maximum at ~75 bp upstream of the pA site similar to the Trm140 PAR-CLIP. Mutational signature of m3C of Trm140 PAR-CLIP and m3C-CLIP in mRNAs at position chrVIII:303572 in B), chrIV:1153577 in C) and chrIII:162694 in D) visualized using the IGV tool (Broad Institute).

5.1.6 Translation efficiency change upon m3C depletion

Chemical modifications on tRNAs and mRNAs, as well as some reader proteins of RNA modifications have been shown to affect translation (Arango et al., 2018; Coats et al., 2017; Li et al., 2017; Nedialkova and Leidel, 2015). In order to test an effect of the m3C modification on translation via codon occupancy, we used published ribosome profiling data of WT and Trm140 KO cells (Chou et al., 2017). This revealed a change in codon occupancy on tRNA

Serine, Threonine and Arginine after Trm140 KO. The P-site of the ribosome shows higher occupancy compared to the WT strain on the mentioned tRNAs (Figure 20). This indicates a slower movement of the tRNA from P to E site within the ribosome and therefore slower translation on those codons in general. This suggests an effect on translation due to the absence of the m³C modification on tRNAs Serine, Threonine, and Arginine.

In addition to the effect of m³C modification on tRNAs, I analyzed the effect on mRNAs in response to the depletion of Trm140. For this, I used the translational efficiency, which is calculated as the ratio between translation (derived from counts of footprints per mRNA in ribosome profiling experiments) over transcription (derived from RNA-seq mRNA levels) of a particular mRNA (Chou et al., 2017). I calculated the translational efficiency for six m³C modified mRNA transcripts on the m³C containing codons of two independent biological replicates for Trm140 KO and WT cells (Figure 20B). For this, I used the ratio of read counts of ribosome footprints per codon and counts of RNA per codon (Chou et al., 2017). I selected six candidate transcripts based on their m³C modification status showing C→T transition in Trm140 PAR-CLIP and m³C-CLIP experiments (Methods Section 3.3.12). The translational efficiency at the selected codons is increased in the case of Trm140 KO compared to WT condition for all six tested transcripts. This suggests more efficient translation of the respective mRNA codon in the absence of the m³C modification. This leads to the hypothesis that the m³C modification in mRNA can be a roadblock for the ribosome and stalls translation thereby decreasing the translational efficiency.

Identification and functional characterization of the novel mRNA modification *N*³-methylcytinde (m3C)

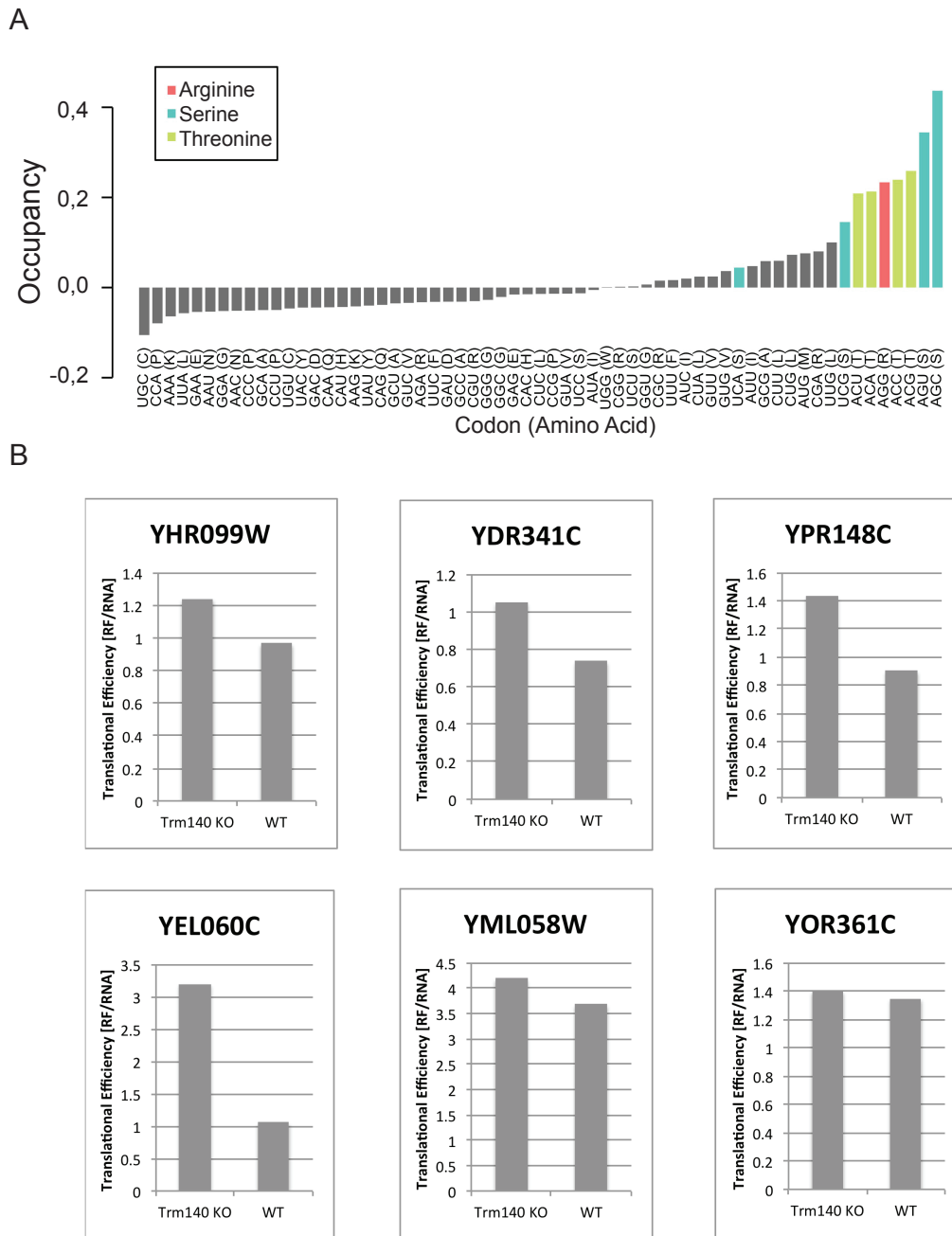


Figure 20: m3C modification loss changes ribosome dynamics.

A) Ribosome profiling of Trm140 KO compared to WT strain showed increased P-site occupancy for the established m3C containing tRNAs. This indicates prolonged residing time of the respective tRNAs inside the ribosome due to missing m3C modification on the tRNAs. B) Translational efficiency changes of m3C containing codons shown for six different mRNA transcripts. Translational efficiency is increased in Trm140 KO cells compared to the WT cells for the m3C containing codon.

5.1.7 Co-localization of Trm140 with RNA degradation factors

To evaluate a function of the m3C mRNA and tRNA modification in various processes, we compared Trm140 PAR-CLIP with experiments from general RNA processing and degradation factors comprising nuclear and cytoplasmic processes. Briefly, these factors function in nuclear RNA surveillance (Nrd1, Nab3), 5' cap binding (Cbc2), mRNA transcript elongation (Bur1, Bur2, Ctk1, Ctk2, Cdc73, Ctr9, Leo1, Paf1, Rtf1, Set1, Set2, Dot1, Spt5, Spt6, Rpb1), pre-mRNA splicing (Ist3, Nam8, Mud1, Snp1, Luc7, Mud2, Msl5), pre-mRNA 3' processing (Pab1, Pub1, Rna15, Mpe1, Cft2; Yth1), transcription termination (Rat1, Rai1, Rtt103, Pcf11), mRNA export (Hrp1, Tho2, Gbp2, Hrb1, Mex67, Sub2, Yra1, Nab2, Npl3) (Baejen et al., 2017, 2014; Battaglia et al., 2017; Schulz et al., 2013), deadenylation (Ccr4, Pop2, Not1, Caf40), decapping (Dcp1, Dcp2, Edc2, Edc3, Dhh1), 5'→3' exonuclease (Xrn1) exosome (Rrp6, Rrp40, Csl4, Rrp4, Rrp44), TRAMP complex (Trf4/5, Air1/2, Mtr4), Ski complex (Ski2, Ski3, Ski7, Ski8), and NMD (Upf1, Upf2, Upf3, Nmd4). Clustered co-localization analysis in a range of 40 nt around the crosslink site of each factor is shown in Appendix Figure 41. We observed co-localization of Trm140 mostly with cytoplasmic RNA degradation factors. Detailed co-localization analysis with degradation factors in Figure 21 revealed binding sites of Trm140 close to cross-link sites of the deadenylation machinery (Ccr4, Pop2, Not1, Caf40, Pan2, and Pan3), the exosome (Rrp6, Csl4, and Rrp40), as well as the exosome associated complexes TRAMP (via Trf5) and Ski (via Ski7) complex. Assuming Trm140 binding leads to deposition of a m3C modification, this observation suggests a cytoplasmic function of the m3C modification on RNA degradation.

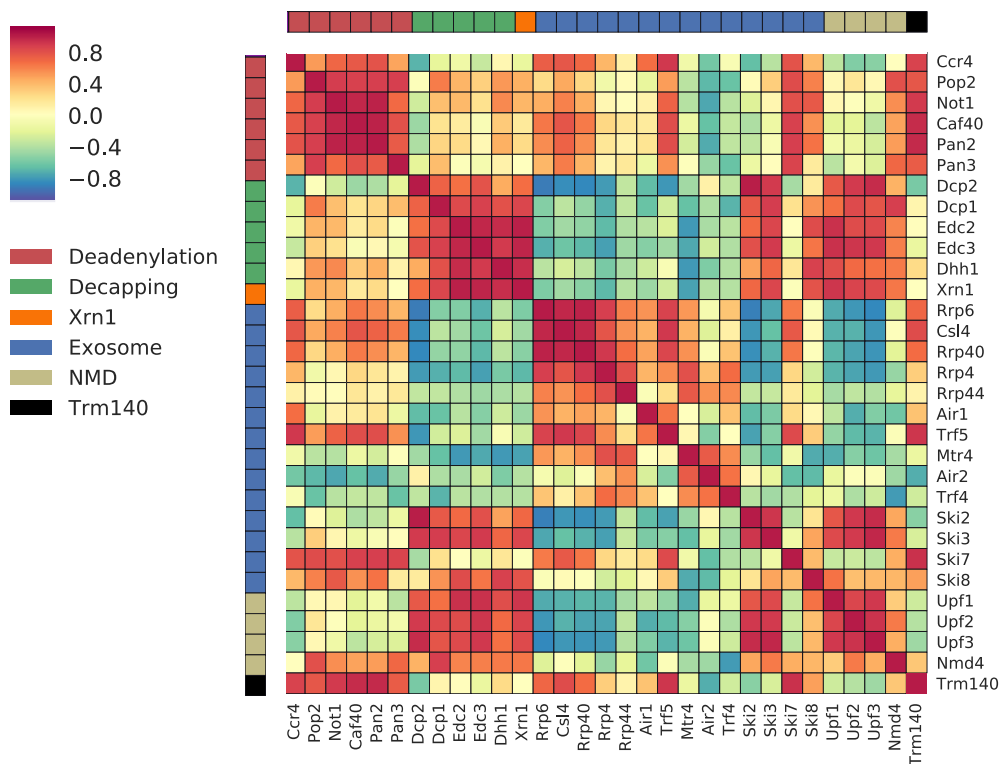


Figure 21: *Trm140* co-localizes with various RNA degradation factors.

A) Co-localization analysis 40 nt around the crosslink site of each factor performed with PAR-CLIP data from the deadenylation machinery, decapping factors, 5'→3' exonuclease, exosome and exosome associated complexes TRAMP and Ski as well as NMD compared to *Trm140*. This revealed high correlation (>0.6) of *Trm140* with RNA degradation factors of the deadenylation and exosome machinery, in particular *Ccr4*, *Pop2*, *Not1*, *Caf40*, *Pan2*, *Pan3*, *Rrp6*, *Csl4*, *Rrp40*, *Trf5*, and *Ski7*.

5.1.8 Depletion of m3C modification alters RNA synthesis rate and half-life

Metabolic labeling and sequencing of newly synthesized RNA (4tU-seq) can be used to measure RNA synthesis and degradation rates in cells (Schulz et al., 2013; Sun et al., 2012). To study the effect of *Trm140* KO and therefore lack of m3C modification on tRNAs and mRNAs, we performed 4tU-seq of *Trm140* KO and WT cells. The two biological replicates showed a high correlation for labeled (nascent) and total (steady state) RNA replicates (Figure 22).

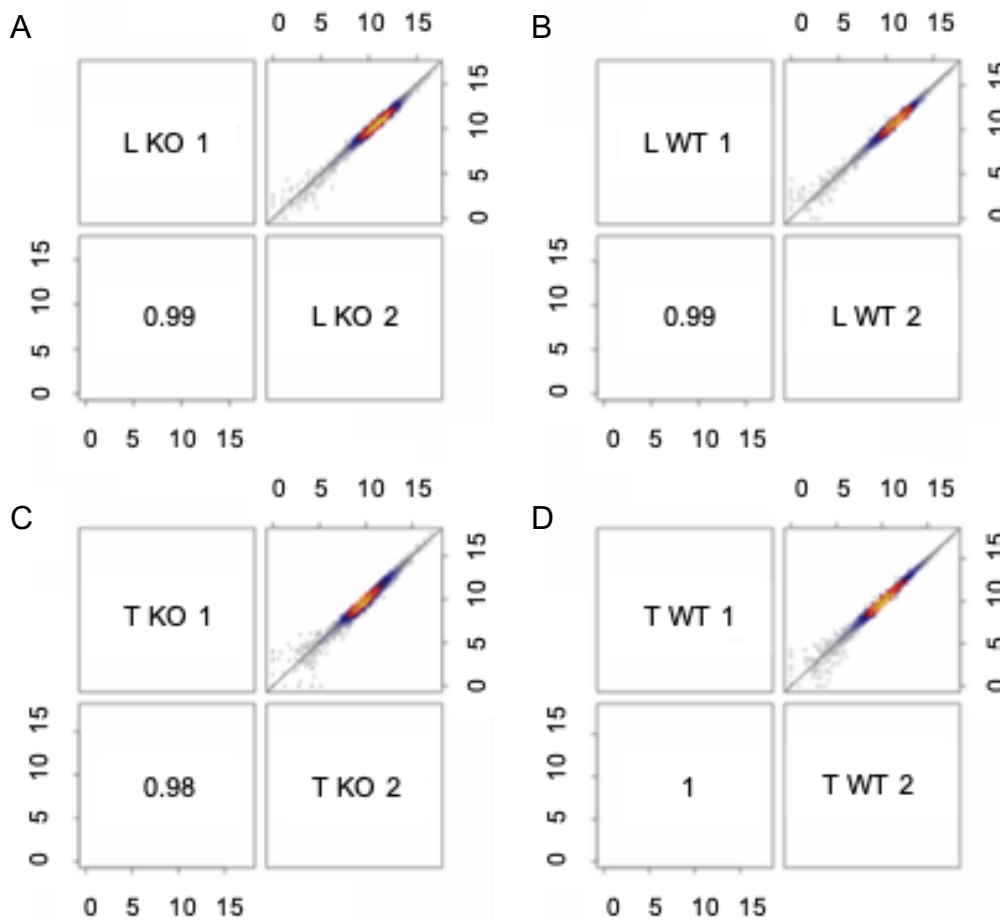


Figure 22: Spearman correlation of 4tU-seq data for *Trm140* KO and WT cells. A) The 4tU-seq replicates of *Trm140* KO labeled (L) RNA replicate 1 and replicate 2 showed a Spearman correlation of 0.99. B) The replicate correlation of WT L RNA was 0.99. C) The *Trm140* KO total (T) RNA replicates showed a correlation of 0.98. D) The correlation of WT T replicates showed a Spearman correlation of 1.

The 4tU-seq datasets (WT and *Trm140*KO) revealed a globally decreased synthesis rate (Wilcoxon, $p = 1.83e^{-60}$) (Figure 23A) and a slightly increased half-life (Wilcoxon, $p = 0.407$) of all protein coding transcripts (Figure 23B) in the *Trm140* KO condition compared to WT. Overall decreased translational efficiency revealed by ribosome profiling (Chou et al., 2017) in the context of *Trm140* KO cells can result in a reduced protein synthesis. This might explain the reduced RNA synthesis rate due to a feedback loop of less available RNA synthesis machinery.

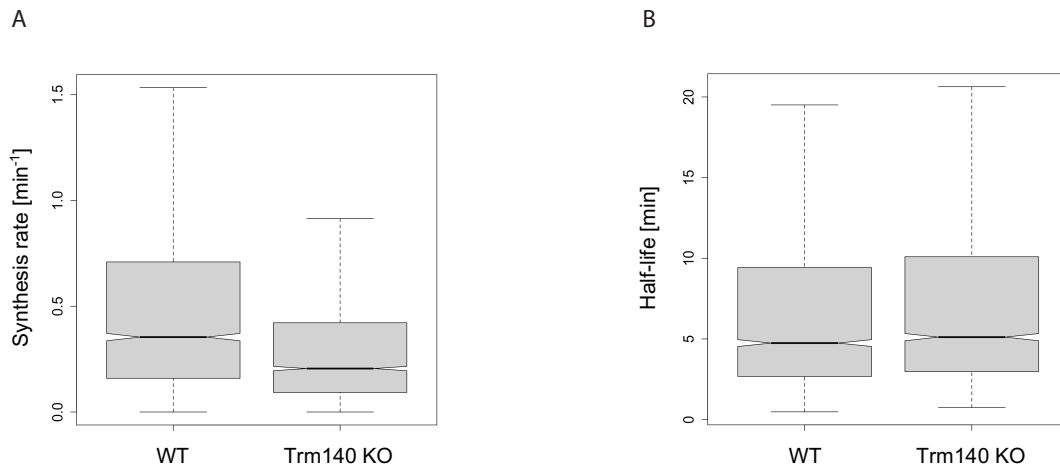


Figure 23: Functional characterization of m3C modification loss by using metabolic RNA labeling and sequencing in *S. cerevisiae*.

4tU-seq of WT versus Trm140 KO cells revealed significantly decreased synthesis rate (in min⁻¹) for mRNA in A) and slightly increased half-life (in min) of mRNA in B) in the absence of m3C modification.

5.1.9 Experimental set-up for PAR-CLIP of METTL8

In order to elucidate the localization of the m3C modification in human cells, I used the inducible Hek293 Flp-In™ T-REx™ Cell line (Invitrogen) to generate a C-terminally 3xFLAG tagged METTL8. In this system, protein expression is under the control of a tetracycline inducible promoter. Using this cell line, I performed PAR-CLIP of METTL8 to map its binding sites over the whole transcriptome. For the experiment, different conditions regarding tetracycline induction and crosslinking at the wavelength of 365 nm in the input and IP samples were tested and examined by Western Blot analysis as shown in Figure 24A. The non-induced sample only showed endogenous level of METTL8 and no signal with the FLAG antibody in the input. Thus, gene expression of METTL8-3xFLAG in the Flp-In™ T-REx™ 293 cells is only very weak if at all present in the cultured cells under standard conditions. After tetracycline induction, the METTL8 expression level stays constant as shown with the anti-METTL8 antibody and a signal for the FLAG-tagged METTL8 was observed with the anti-FLAG antibody. The Western Blot for the IP samples using the anti-FLAG M2 magnetic beads showed signal only for the tetracycline

induced samples in both blotting conditions (anti-FLAG and anti-METTL8). The double band in the non-induced IP sample blotted with the anti-FLAG antibody is caused by the heavy chain of the antibody and cross reactivity of the antibodies used for the Western Blot. Radioactive labeling (³²P) of the RNA crosslinked to the enriched protein of interest showed lower signal in the phosphor image for the control samples without tetracycline induction and without crosslinking compared to the two replicates with induction and crosslinking as depicted in Figure 24B. The area of the gel highlighted with black rectangles was used for further experimental procedure.

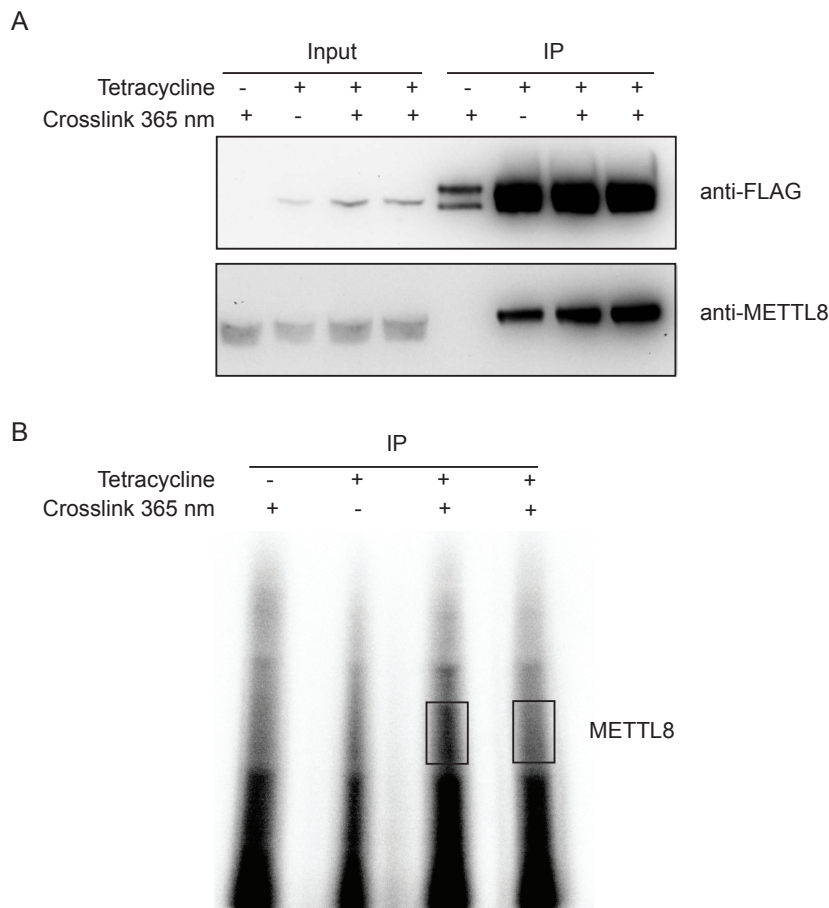


Figure 24: PAR-CLIP of METTL8-3xFLAG in *Flp-In*TM *T-REx*TM 293 cells.

A) Western Blot of different conditions varying Tetracycline induction and UV-crosslinking at 365 nm of Input and IP samples blotted with anti-FLAG and anti-METTL8 antibodies. B) Phosphor image of ³²P-labeled RNA immuno-precipitated with METTL8-3xFLAG using the same conditions as in the Western Blot. The area cut for further experimental steps is marked on the gel with black rectangles.

5.1.10 Mapping of the m3C modification in human mRNA

The sequencing data of three independent biological replicates of the 3xFLAG-METTL8 PAR-CLIP experiments were merged for analysis and resulted in 148,781 UV-induced crosslink sites. The UV induced crosslink sites were aligned to protein coding genes at the TSS and pA site and resulted in a metagene plot depicted in Figure 25A. The metagene plot shows binding of METTL8 over the entire length of the mRNAs with higher signal towards the end of transcripts. This indicates a distribution of m3C modifications over the protein coding transcripts with higher occurrence towards the 3' end. The search for m3C mutational signatures (C→T transition) in the METTL8 PAR-CLIP experiments resulted in 6144 sites (coverage > 2, pval > 0.005). XXmotif analysis (Luehr et al., 2012) of those sites revealed binding preference of METTL8 to CG rich elements (Figure 25B). The best two hits showed an occurrence of 48.13% and 20.4% with E-Values of 3.3e-24 and 1.08e-9, respectively.

Identification and functional characterization of the novel mRNA modification N^3 -methylcytosine (m3C)

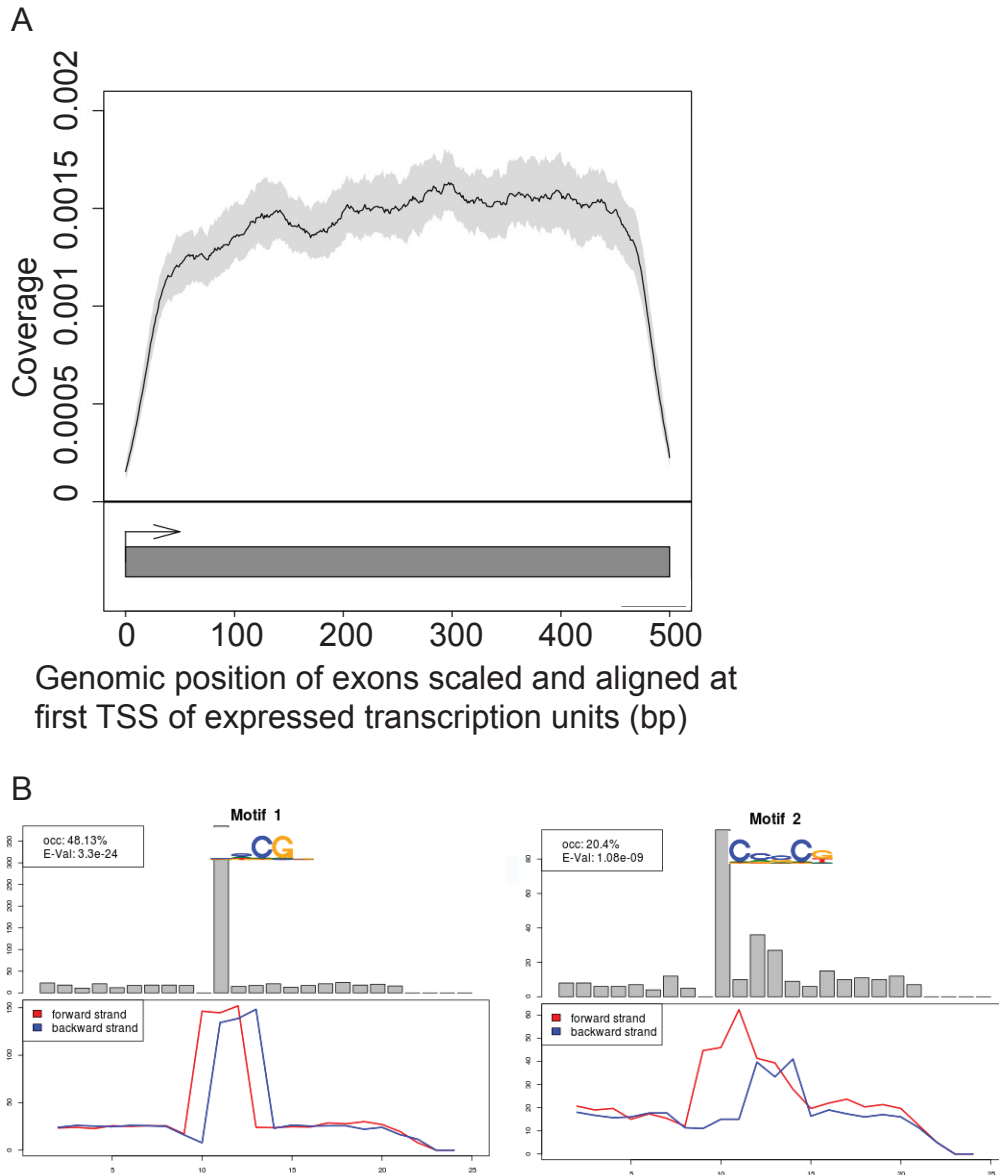


Figure 25: METTL8 PAR-CLIP reveals localization of m3C modification in mRNA. A) Metagene analysis of METTL8 PAR-CLIP replicates on exons scaled and aligned at the first annotated TSS. METTL8 cross-link sites are distributed over the entire mRNA. B) XXmotif analysis of the C→T mutational signature of m3C modification reveals a CG rich binding element for METTL8.

5.2 Discussion and Outlook

5.2.1 Identification of the m3C modification in tRNA Arginine

To study the m3C modification in the yeast *S. cerevisiae*, I performed PAR-CLIP of the m3C methyltransferase Trm140. The currently established targets of Trm140 are tRNA Serine and Threonine (D'Silva et al., 2011; Noma et al., 2011). I also developed a method called m3C-CLIP, using an anti-m3C antibody for RNA immunoprecipitation analogous to PA-m6A-seq (Chen et al., 2015). I showed that both methods enriched for sequencing reads containing mutational signatures typical for m3C modification on established targets tRNA Serine and Threonine (Figure 16). I observed that the m3C modification changes the base pairing properties of C during RT. The mutational signature appears mainly as C→T conversions as shown in Figure 16 and described before (Arimbasseri et al., 2015). Additionally, the m3C modification induced a RT block close to the modified nucleotide, which has been also observed by others (Arimbasseri et al., 2015; D'Silva et al., 2011). Using these mutational signatures, we were able to map m3C at single nucleotide resolution in the yeast transcriptome and identify tRNA Arginine with anticodon CCU as a new tRNA target (Figure 18A). The modification was detected at position C32, which is established for tRNA Serine and Threonine (D'Silva et al., 2011; Noma et al., 2011). I also applied a RT assay for tRNA Arginine to confirm m3C presence by RT block (Figure 18B). In order to demonstrate Trm140 dependency of the m3C modification as shown for tRNA Serine and Threonine (D'Silva et al., 2011; Noma et al., 2011), I also demethylated RNA with AlkB (Aas et al., 2003), derived and purified from *E. coli*. The demethylated RNA was subsequently re-methylated with purified yeast Trm140, which showed Trm140 methylation dependent RT block in the RT assay (Figure 18B). Using the combined approach of PAR-CLIP, m3C-CLIP and RT assay, I demonstrated the presence of Trm140-dependent m3C modification on tRNA Arginine (CCU) in the anticodon loop at C32. Trm140 has been shown to recognize the sequence G35-U36-t6A37 of the anticodon loop of tRNA Threonine or t6A37 and i6A37 of tRNA Serine (Han et al., 2017). The anticodon loop of tRNA Arginine (CCU) also contains U36 and the i6A37 modification (Chan and Lowe, 2009), which can be the recognition signal for Trm140 to modify this tRNA.

5.2.2 Effect of m3C loss on other tRNA modifications

The m3C-CLIP experiments using RNA from a Trm140 KO strain, as shown in Figure 18A, showed an increased mutational signature for G26 with G→A and G→T transitions on tRNA Arginine with anticodon CCU using the IGV tool (Broad Institute). This suggests a higher modification level of the nucleotide, which is known to be modified with m²₂G (Chan and Lowe, 2009). A similar mutational signature was observed before for this modification on G26 of tRNA Serine (Arimbasseri et al., 2016). Higher levels of m²₂G in the Trm140 KO strain compared to WT at tRNA Arginine could indicate a compensatory mechanism of the cells in response to the loss of m3C32. The need for a compensatory mechanism could be caused by a destabilized tRNA Arginine structure in the absence of the m3C modification. In line with this hypothesis, it has been shown before that tRNA modifications at position 32 are important to stabilize the structure of the anticodon loop (Auffinger and Westhof, 2001, 1999). Increased m²₂G modification levels might stabilize the tRNA structure (Steinberg and Cedergren, 1995), in particular the anticodon loop in the context of the Trm140 KO cells. An increased m²₂G level has only been observed for tRNA Arginine and not for tRNA Serine and Threonine in our experimental set-up. This suggests a more important function of the m3C modification on tRNA Arginine in the metabolism of the cell. This is also reflected in the P site occupancy analysis of the ribosome (Figure 20), where tRNA Arginine showed a strong increase compared to most other tRNAs.

5.2.3 Functional relevance of the m3C modification

For functional analysis, it has previously been shown in experiments with human cells that the ratio of ribosomes associated in polysomes to monosomes is slightly decreased in METTL8 KO compared to WT conditions (Xu et al., 2017). This suggests increased ribosome stalling when METTL8 and thus the m3C modification was absent. In addition to that, ribosome profiling analysis in yeast WT and Trm140 KO strains (Chou et al., 2017) showed increased P site ribosome occupancy on those codons, which are complementary to the tRNAs

lacking the m3C modification in the Trm140 KO strain (Figure 20A). The slower translation rate in absence of the m3C modification could be explained by incorrect modified tRNAs (Chou et al., 2017). However, our analysis revealed enhanced translational efficiency on m3C harboring codons (Figure 20B). Furthermore, 4tU-seq of the Trm140 KO in comparison to WT strain revealed an impact of the m3C modification on mRNA stability (Figure 23). Our analysis showed a decreased synthesis rate and a slightly increased half-life for mRNAs in the Trm140 KO compared to WT strain (Figure 23).

In the light of existing literature, our findings suggest that there are two contradictory effects: the role of m3C modification in tRNA versus in mRNA. The lack of m3C modification in tRNAs might destabilize the structure of the anticodon loop (Auffinger and Westhof, 2001, 1999). This may lead to an inefficient accommodation in the active center of the ribosome and therefore ribosome stalling. On the other hand, the m3C modification in mRNA can hinder Watson Crick base pairing (Figure 26) of the mRNA codon with the tRNA anticodon within the actively translating ribosome. The impaired Watson Crick base pairing is explained by H-H repulsion of the *N*³-methyl group of m3C with the *N*¹-H proton of G. This impeded Watson Crick base pairing of m3C with G has been observed also in our RT assay by RT block (Figure 17A and Figure 18B). The hindered Watson Crick base pairing between m3C and G suggests that the m3C modification in mRNAs stalls the ribosome like a roadblock and can thereby lead to recruitment of degradation factors (Radhakrishnan et al., 2016). Indeed, comparison of Trm140 PAR-CLIP data with genome wide RNA binding data of general processing factors in yeast revealed strong co-localization coefficients of Trm140 binding sites on transcripts with crosslink-site of degradation factors, particularly deadenylation factors (Ccr4, Pop2, Pan2 and Pan3), several exosome components (Rrp6, Csl4, and Rrp40) and an exosome associated factor (Ski7) (Figure 21). Co-localization with general cytoplasmic RNA degradation factors indicates a function of the m3C modification in mRNAs in the cytoplasm. Localization of Trm140 to the actin cytoskeleton has been shown before, supporting further a cytoplasmic deposition of the m3C modification (Kilchert and Spang, 2011). Our data analysis on translational efficiency suggests ribosome stalling upon translation of codons harboring the m3C modification (Figure 20) and this can lead to

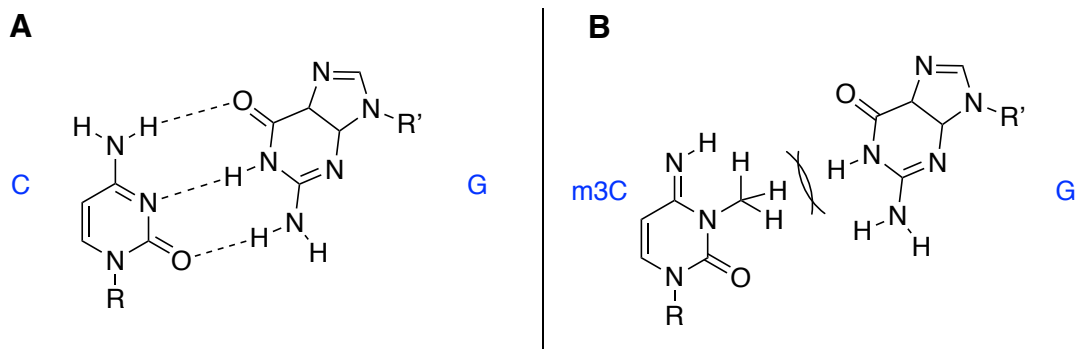


Figure 26: Schematic drawing of Watson Crick base pairing between unmodified C and G (A) as well as H-H repulsion of m3C and G (B).

indirect recruitment of general degradation factors in the cytoplasm. This hypothesis is supported by previous findings regarding the response of ribosome stalling on premature termination codons during NMD (Celik et al., 2017). During this process, the ribosome is blocked on premature stop codons and factors that initiate mRNA degradation by endonucleolytic cleavage, deadenylation (by the Ccr4-Not and Pan2/Pan3 deadenylation complexes) and/or decapping (by the decapping complex) are recruited (Lykke-Andersen and Jensen, 2015). It is also known that the Ski complex interacts with stalled ribosomes in case of NMD and links mRNAs to deadenylation and exosomal degradation (Mitchell and Tollervey, 2003). From the work of Pelechano et al., it is known that ribosome stalling increases degradation of the transcript from the 5' end (Pelechano et al., 2015). This is also in agreement with studies from Radhakrishnan et al. that the degradation machinery is recruited upon ribosome stalling. Ribosome stalling can be monitored by Dhh1, a decapping enhancer, which travels with the ribosome and senses translational efficiency (Radhakrishnan et al., 2016). In our co-localization analysis we observed also medium enrichment of Dhh1 crosslink-sites in proximity to Trm140 binding sites (Figure 21). Thus, our observations regarding co-localization of Trm140 with RNA degradation factors can be placed into context with existing literature mentioned above. Moreover, the lack of the m3C mRNA modification acting as a roadblock or “degradation signal” might explain the slightly prolonged half-life of mRNAs in the Trm140 KO cells as shown in Figure 23. However, the exact

mechanism how the m3C modification in mRNA can stall the ribosome or recruit degradation factors directly needs to be further analyzed.

5.2.4 Future directions for studies on the identification of m3C binding proteins in different eukaryotic cells

Epigenomic marks such as 5-methylcytosine in DNA can be recognized by specific binding proteins, thereby mediating its repressive effects (Klose and Bird, 2006). In addition to selective DNA modification binding proteins, methylation of RNA at specific positions can also affect interacting proteins (Dominissini et al., 2012). Novel m6A binding proteins were identified using methylated and negative control RNA baits in an RNA affinity chromatography approach accompanied with mass spectrometry analysis (Dominissini et al., 2012). Identification of m6A reader proteins was additionally performed by *in vitro* assays and binding experiments with a recombinantly expressed YTH domain. These studies showed RNA binding of the YTH domain in an m6A-dependent manner, although with different binding affinities, some in the nM range (Li et al., 2014; Luo and Tong, 2014; Zhu et al., 2014) and others with μ M affinity (Theler et al., 2014; Xu et al., 2015). Identification of different m6A binding proteins revealed additional implications of the m6A modification in various processes during the life-time of an RNA molecule (Han et al., 2019; Kretschmer et al., 2018; Lence et al., 2016; Theler et al., 2014; Wang et al., 2015; Y. Wang et al., 2014; Xiao et al., 2016). Identification of m3C binding proteins is required in the future to study the role of the modification in detail. This may be pursued in a similar way to the identification of m6A binding proteins (Dominissini et al., 2012): Using RNA affinity chromatography of m3C modified and control RNA with yeast and human cellular extract in combination with mass spectrometry analysis. These experiments can reveal the m3C interactome of yeast and human cells. Identification of specific binding proteins may uncover novel functions of the m3C mRNA modification. Identification of potential eraser proteins with homologous function to AlkB from *E. coli* (Trewick et al., 2002) can additionally reveal a dynamic nature of the m3C modification. A potential candidate to reverse the m3C modification in human cells is the

ALKBH3 protein, which has been shown to repair methylation damage in DNA and RNA (Aas et al., 2003; Duncan et al., 2002; Lee et al., 2005; Zhao et al., 2015).

5.2.5 Functional implication of the m3C modification in the context of the epitranscriptome

In this dissertation, the localization of the m3C modification was analyzed by metagene binding profiles on mRNA for the methyltransferases Trm140 and METTL8 in yeast and human cells. The analyses revealed a binding preference for both enzymes over the protein coding sequence with higher occupancy at the pA site (Figure 19A, Figure 25A). A similar localization of the m6A modification (see Figure 5) has been shown before in mouse and human mRNAs (Dominissini et al., 2012). Due to a similar localization of the m6A and m3C modifications, they may have similar functional implications. Epitranscriptomic marks may decorate the RNA and thereby influence *cis*-regulatory elements within the protein coding transcript. Various *cis*-regulatory elements within the RNA are highly enriched in the 5' and 3' UTRs regulating RNA localization, stability, and translation (Mignone et al., 2002). Reader proteins of the m6A modification are associated with mRNA stability and translation (Theler et al., 2014; Wang et al., 2015; X. Wang et al., 2014; Xiao et al., 2016). According to the analysis of a reduced translational efficiency for m3C harboring codons in yeast (Figure 19B), the m3C modification might also have specific binding proteins to influence ribosome stalling. Our co-localization analysis (Figure 21) revealed several potential binding partners or interacting partners of these complexes not included in the analysis here. In order to get a better understanding of the m3C modification, its function and to compare it with other epitranscriptomic marks, one needs to identify and study the m3C interactome.

It has been shown for the m6A modification to act as a switch for the mRNA secondary structure to expose *cis*-regulatory elements and to regulate protein-RNA interactions as shown in Figure 27 (Liu et al., 2015). Due to its ability to hinder Watson Crick base pairing of the m3C modification with G

(Figure 26), this epitranscriptomic mark harbors the potential to change RNA secondary structures. This might in turn allow access for RNA binding proteins to the single stranded RNA analogous to the m6A mark as shown in Figure 27.

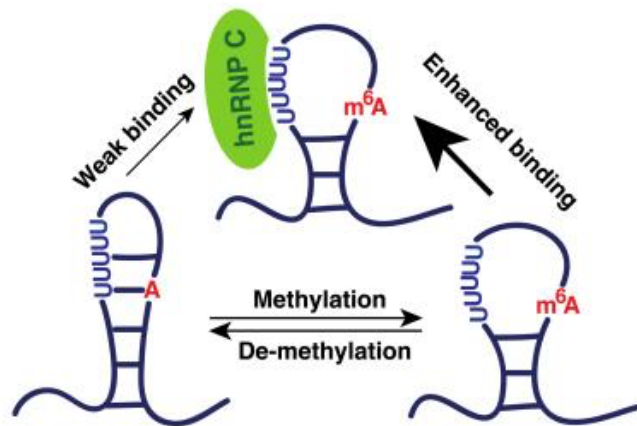


Figure 27: Illustration of the m6A switch model.

Reversible m6A methylation changes the secondary structure of mRNA to regulate RNA-protein interactions (from Liu et al., 2015).

5.2.6 Dissecting functional implications of the m3C modification in tRNA and mRNA in the future

In the context of Trm140 KO strain, we observed increased P site occupancy for m3C modified tRNAs (Figure 20A) implying reduced translational efficiency. In contrast, we observed higher translational efficiency in the absence of m3C on mRNA codons (Figure 20B). Global ribosome profiling (Chou et al., 2017) and 4tU-seq were not able to reveal the contradictory effects of the m3C modification in tRNAs and mRNAs independently. In order to distinguish the effect of the m3C modification on the level of tRNA and mRNA separately, different experimental set-ups need to be used.

One idea is to use the Trm140 KO strain but complementing the cells with purified tRNAs from WT strains to ensure proper tRNA modifications. Using this approach, the measured effects by ribosome profiling and 4tU-seq should just be caused by the missing m3C modification in mRNAs. On the other hand, this approach has various disadvantages. For one, uptake of the tRNAs into

the cytoplasm cannot be guaranteed. For two, the concentration of tRNA pool would be changed, which would influence the availability of all tRNAs during translation and thereby artificially perturb translational efficiency. It has been shown that yeast cells use differential expression of tRNAs to selectively adjust protein synthesis under stress conditions (Torrent et al., 2018). Inducing a stress situation by altered tRNA availability might additionally perturb gene expression.

An alternative approach could be the application of a fast degradation system for the methyltransferase Trm140. A fast degradation system was developed recently, using the dTAG mechanism (Nabet et al., 2018). This enables the immediate and selective control of single protein abundance by using the novel degrader FKBP12^{F36V} and in frame expression of FKBP12^{F36V}, fused to the protein of interest. This technology provides kinetic resolution to biological analyses such as RNA modifications (Nabet et al., 2018). Rapid degradation of Trm140 would lead to loss of the m3C modification on short-lived mRNAs, while the m3C modification would in theory still be present on long-lived tRNAs. This may enable us to distinguish the effect of the m3C modification on mRNAs and tRNAs. Changes in translational efficiency and RNA metabolism should now be only caused by the m3C mRNA modification.

5.2.7 Functional characterization of the m3C modification in human cells in the future

METTL8 PAR-CLIP has revealed the presence of the m3C modification in mRNAs and its localization within mRNAs (Figure 25A). However, functional studies such as ribosome profiling and RNA-seq are complementarily required to gain a better understanding of the function of the m3C modification in human cells. To study the effect of the m3C modification on mRNA, depletion of METTL8 is necessary to compare the effect of the m3C modification loss to wildtype cells. It has been shown that cells having a KO of METTL8 are viable (Xu et al., 2017). However, permanent KO of METTL8 could lead to secondary effects on long-lived RNAs. Depletion of METTL8 in a short regime of time, can help to study the functional implications of the m3C modification in mRNAs. For

rapid depletion of METTL8, the application of the dTAG system in human cells is possible as well (Nabet et al., 2018). Subsequent to METTL8 depletion, transient-transcriptome sequencing (TT-seq) similar to 4tU-seq is applicable to study the effect of m3C loss on RNA metabolism (Schwalb et al., 2016). It is now possible to calculate RNA synthesis and degradation rates from the sequencing data analogous to the analysis in yeast (Figure 23, Methods). In addition to the RNA metabolism, the effect of m3C loss in mRNAs on translation efficiency may be studied by ribosome profiling (Ingolia et al., 2009). Using both methods, TT-seq and ribosome profiling, in combination with the dTAG system, may give detailed insights on potential effects of the m3C modification in mRNAs of human cells.

5.2.8 Technical challenges in the field of epitranscriptomics and future perspectives

Our understanding of the localization, function and regulation of the currently known dynamic RNA modifications relies on sequencing technologies (Li et al., 2016b). However, RNA modifications cannot be detected directly so far on a high throughput level, since sequencing technologies depend on reverse transcription of RNA into cDNA. This process erases in many cases the information about RNA modifications, which do not change the base pairing properties of the modified bases. In addition, mRNA is of low abundance compared to tRNA and rRNA and this increases the challenge to detect specific base modifications. Thus, new experimental approaches are required to identify novel mRNA modifications and map them with nucleotide resolution. Potential techniques to study mRNA modifications in more detail may involve specific antibodies against the modification of interest or direct RNA sequencing approaches (Arango et al., 2018; Chen et al., 2015; Dominissini et al., 2016, 2012; Liu et al., 2019; A. M. Smith et al., 2017). A very promising technique to study RNA modifications using direct RNA sequencing was developed by Oxford Nanopore Technologies to overcome current limitations of the available techniques (Hussain, 2018; Liu et al., 2019; A. M. Smith et al., 2017). Hereby, a synthetic, electrically resistant membrane is interspersed with large

nanopores. A voltage is applied across this membrane providing an ionic current across the pores. RNA molecules fused to a tether protein are directed to nanopores and a motor protein pulls individual RNA strands through the biological nanopores at a controlled rate. Each RNA base shows a specific current profile across the nanopores and this allows calling of the nucleotide composition of the RNA sequence. However, this method possesses a sequencing accuracy of only around 85% (experience from our laboratory and personal communication) compared to 99% standard Illumina DNA sequencing (Fox et al., 2014). Thus, the accuracy of Oxford Nanopore Technologies direct RNA sequencing does currently not allow to distinguish between mutational signatures caused by an RNA modification or pure sequencing error. With an improved sequencing accuracy, this direct RNA sequencing technique may be able to distinguish modified nucleosides within an RNA stretch. A very recent study proposed that the presence of RNA modifications causes current intensity changes leading to decreased qualities and increased “errors” in the output of base-calling algorithms. Using this model and a machine learning approach, the authors showed that base-calling “errors” mapped m6A modifications in native RNA sequences with an overall accuracy of ~90% (Liu et al., 2019). In order to identify the m3C modification in native RNAs, the current intensity change due to the modification needs to be determined upon comparison with the raw current intensities.

It is important to note that not only experimental and technical improvements need to be introduced to improve the quality of the identification and calling of modifications, but also data analyses. It has been observed before that next-generation sequencing data has to be handled carefully to not draw false conclusions from it (Schwartz, 2018). Sequencing reads have to be mapped to the reference genome with high accuracy in order to call mutational signatures and reverse transcription blocks caused by modified nucleosides (Schwartz, 2018). Reverse transcription blocks occur only at the end of sequencing reads and thus, soft clipping of their ends during mapping may change the resulting read end and its localization (Schwartz, 2018). In addition, a sufficient sequencing depth is required to be able to call low abundant mRNA modification sites (Xiong et al., 2018). Thus, experimental and technical improvements to overcome current limitations are the most promising approach

Identification and functional characterization of the
novel mRNA modification *N*³-methylcytidine (m³C)

to advance studies in the field of epitranscriptomics in the future (Jantsch and Schaefer, 2019).

6 Bibliography

- Aas PA, Otterlei M, Falnes PO, Vagbo CB, Skorpen F, Akbari M, Sundheim O, Bjoras M, Slupphaug G, Seeberg E, Krokan HE. 2003. Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature* **421**:859–863. doi:10.1038/nature01388.1.
- Abernathy E, Glaunsinger B. 2015. Emerging roles for RNA degradation in viral replication and antiviral defense. *Virology* **479–480**:600–608. doi:https://doi.org/10.1016/j.virol.2015.02.007
- Aebi M, Kirchner G, Chen JY, Vijayraghavan U, Jacobson A, Martin NC, Abelson J. 1990. Isolation of a temperature-sensitive mutant with an altered tRNA nucleotidyltransferase and cloning of the gene encoding tRNA nucleotidyltransferase in the yeast *Saccharomyces cerevisiae*. *J Biol Chem* **265**:16216–16220.
- Ahn SH, Kim M, Buratowski S. 2004. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol Cell* **13**:67–76. doi:https://doi.org/10.1016/S1097-2765(03)00492-1
- Anders S, Pyl PT, Huber W. 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**:166–169. doi:10.1093/bioinformatics/btu638
- Anderson JS, Parker RP. 1998. The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. *EMBO J* **17**:1497–1506. doi:10.1093/emboj/17.5.1497
- Ansari A, Hampsey M. 2005. A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. *Genes Dev* **19**:2969–2978. doi:10.1101/gad.1362305
- Araki Y, Takahashi S, Kobayashi T, Kajihio H, Hoshino SI, Katada T. 2001. Ski7p G protein interacts with the exosome and the ski complex for 3'-to-5' mRNA decay in yeast. *EMBO J* **20**:4684–4693. doi:10.1093/emboj/20.17.4684
- Arango D, Sturgill D, Alhusaini N, Dillman AA, Sweet TJ, Hanson G, Hosogane M, Sinclair WR, Nanan KK, Mandler MD, Fox SD, Zenggeya TT, Andresson T, Meier JL, Coller J, Oberdoerffer S. 2018. Acetylation of Cytidine in mRNA Promotes Translation Efficiency. *Cell* **175**:1872-1886.e24. doi:10.1016/j.cell.2018.10.030
- Arigo JT, Eyster DE, Carroll KL, Corden JL. 2006. Termination of Cryptic Unstable

- Transcripts Is Directed by Yeast RNA-Binding Proteins Nrd1 and Nab3. *Mol Cell* **23**:841–851. doi:10.1016/j.molcel.2006.07.024
- Arimbasseri AG, Blewett NH, Iben JR, Lamichhane TN, Cherkasova V, Hafner M, Maraia RJ. 2015. RNA Polymerase III Output Is Functionally Linked to tRNA Dimethyl-G26 Modification. *PLoS Genet* **11**:e1005671. doi:10.1371/journal.pgen.1005671
- Arimbasseri AG, Iben J, Wei F, Rijal K, Tomizawa K, Hafner M, Maraia RJ. 2016. Evolving specificity of tRNA 3-methyl-cytidine-32 (m³ C32) modification : a subset of tRNAs Ser requires N⁶-isopentenylolation of A37. *RNA* **32**:1–11. doi:10.1261/ma.056259.116.5
- Arts GJ, Kuersten S, Romby P, Ehresmann B, Mattaj IW. 1998. The role of exportin-t in selective nuclear export of mature tRNAs. *EMBO J* **17**:7430–7441. doi:10.1093/emboj/17.24.7430
- Asakura T, Sasaki T, Nagano F, Satoh a, Obaishi H, Nishioka H, Imamura H, Hotta K, Tanaka K, Nakanishi H, Takai Y. 1998. Isolation and characterization of a novel actin filament-binding protein from *Saccharomyces cerevisiae*. *Oncogene* **16**:121–130. doi:10.1038/sj.onc.1201487
- Ashe MP, Long SK De, Sachs AB. 2000. Glucose Depletion Rapidly Inhibits Translation Initiation in Yeast. *Mol Biol Cell* **11**:833–848.
- Auffinger P, Westhof E. 2001. An extended structural signature for the tRNA anticodon loop. *RNA* **7**:334–341. doi:10.1017/S1355838201002382
- Auffinger P, Westhof E. 1999. Singly and bifurcated hydrogen-bonded base-pairs in tRNA anticodon hairpins and ribozymes. *J Mol Biol* **292**:467–483. doi:10.1006/jmbi.1999.3080
- Baejen C, Andreani J, Torkler P, Battaglia S, Schwalb B, Lidschreiber M, Maier KC, Boltendahl A, Rus P, Esslinger S, Söding J, Cramer P. 2017. Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes. *Mol Cell* **38**–49. doi:10.1016/j.molcel.2017.02.009
- Baejen C, Torkler P, Gressel S, Essig K, Söding J, Cramer P. 2014. Transcriptome maps of mRNP biogenesis factors define pre-mRNA recognition. *Mol Cell* **745**–757. doi:10.1016/j.molcel.2014.08.005
- Baker KE, Parker R. 2004. Nonsense-mediated mRNA decay: terminating erroneous gene expression. *Curr Opin Cell Biol* **16**:293–299. doi:10.1016/j.ceb.2004.03.003
- Bakin A, Ofengand J. 1993. Four newly located pseudouridylate residues in *Escherichia coli* 23S ribosomal RNA are all at the peptidyltransferase center: Analysis by the application of a new sequencing technique. *Biochemistry*

- 32:9754–9762. doi:10.1021/bi00088a030
- Battaglia S, Lidschreiber M, Baejen C, Torkler P, Vos SM, Cramer P. 2017. RNA-dependent chromatin association of transcription elongation factors and Pol II CTD kinases. *Elife*. doi:10.7554/eLife.25637
- Beilharz TH, Preiss T. 2007. Widespread use of poly (A) tail length control to accentuate expression of the yeast transcriptome. *RNA* **13**:982–997. doi:10.1261/rna.569407
- Björk GR, Jacobsson K, Nilsson K, Johansson MJO, Byström AS, Persson OP. 2001. A primordial tRNA modification required for the evolution of life? *EMBO J* **20**:231–239. doi:10.1093/emboj/20.1.231
- Boeck R, Tarun SJ, Rieger M, Deardorff JA, Muller-Auer S, Sachs AB. 1996. The yeast Pan2 protein is required for poly(A)-binding protein-stimulated poly(A)-nuclease activity. *J Biol Chem* **271**:432–438.
- Bonneau F, Basquin J, Ebert J, Lorentzen E, Conti E. 2009. The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation. *Cell* **139**:547–559. doi:10.1016/j.cell.2009.08.042
- Booth MJ, Marsico G, Bachman M, Beraldi D. 2014. Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. *Nat Chem* **6**:435–440. doi:10.1038/nchem.1893
- Bresson S, Tuck A, Staneva D, Tollervey D, Bresson S, Tuck A, Staneva D, Tollervey D. 2017. Nuclear RNA Decay Pathways Aid Rapid Remodeling of Gene Expression in Yeast Article Nuclear RNA Decay Pathways Aid Rapid Remodeling of Gene Expression in Yeast. *Mol Cell* **65**:787-800.e5. doi:10.1016/j.molcel.2017.01.005
- Brown CE, Sachs AB. 1998. Poly(A) tail length control in *Saccharomyces cerevisiae* occurs by message-specific deadenylation. *Mol Cell Biol* **18**:6548–6559.
- Brown JT, Bai X, Johnson AW. 2000. The yeast antiviral proteins Ski2p, Ski3p, and Ski8p exist as a complex in vivo. *RNA* **6**:449–457. doi:10.1017/S1355838200991787
- Buratowski S. 2003. The CTD code. *Nat Struct Biol* **10**:679–680. doi:10.1038/nsb0903-679
- Cao D, Parker R. 2003. Computational modeling and experimental analysis of nonsense-mediated decay in yeast. *Cell* **113**:533–545.
- Capone JP, Sharp PA, RajBhandary UL. 1985. Amber, ochre and opal suppressor tRNA genes derived from a human serine tRNA gene. *EMBO J* **4**:213–21.
- Caponigro G, Muhlrud D, Parker R. 1993. A small segment of the MAT alpha 1 transcript promotes mRNA decay in *Saccharomyces cerevisiae*: a stimulatory

- role for rare codons. *Mol Cell Biol* **13**:5141–5148.
- Caponigro G, Parker R. 1995. Multiple functions for the poly (A)-binding protein in mRNA decapping and deadenylation in yeast. *Genes Dev* **9**:2421–2432. doi:10.1101/gad.9.19.2421
- Carlile TM, Rojas-duran MF, Zinshteyn B, Shin H, Bartoli KM, Gilbert W V. 2014. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **515**:143–146. doi:10.1038/nature13802
- Celik A, Baker R, He F, Jacobson A. 2017. High-resolution profiling of NMD targets in yeast reveals translational fidelity as a basis for substrate selection. *RNA* **23**:735–748. doi:10.1261/rna.060541.116
- Chakrabarti S, Jayachandran U, Bonneau F, Fiorini F, Basquin C, Domcke S, Le Hir H, Conti E. 2011. Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. *Mol Cell* **41**:693–703. doi:10.1016/j.molcel.2011.02.010
- Chan PP, Lowe TM. 2009. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res* **37**:D93-7. doi:10.1093/nar/gkn787
- Chawla G, Sokol NS. 2014. ADAR mediates differential expression of polycistronic microRNAs. *Nucleic Acids Res* **42**:5245–5255. doi:10.1093/nar/gku145
- Chen C, Huang B, Eliasson M, Ryde'n P BA. 2011. Elongator Complex Influences Telomeric Gene Silencing and DNA Damage Response by Its Role in Wobble Uridine tRNA Modification. *PLoS One* **7**:1–11. doi:10.1371/journal.pgen.1002258
- Chen K, Lu Z, Wang X, Fu Y, Luo G-Z, Liu N, Han D, Dominissini D, Dai Q, Pan T, He C. 2015. High-resolution N(6) -methyl adenosine (m(6) A) map using photo-crosslinking-assisted m(6) A sequencing. *Angew Chem Int Ed Engl* **54**:1587–90. doi:10.1002/anie.201410647
- Chen P, Jager G, Zheng B. 2010. Transfer RNA modifications and genes for modifying enzymes in *Arabidopsis thaliana*. *BMC Plant Biol* **10**:201. doi:10.1186/1471-2229-10-201
- Chimnarong S, Suzuki Tateki, Manita T, Ikeuchi Y, Yao M, Suzuki Tsutomu, Tanaka I. 2009. RNA helicase module in an acetyltransferase that modifies a specific tRNA anticodon. *EMBO J* **28**:1–12. doi:10.1038/emboj.2009.69
- Cho EJ, Takagi T, Moore CR, Buratowski S. 1997. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev* **11**:3319–3326.
- Cho H, Kim KM, Kim YK. 2009. Human proline-rich nuclear receptor coregulatory protein 2 mediates an interaction between mRNA surveillance machinery and

- decapping complex. *Mol Cell* **33**:75–86. doi:10.1016/j.molcel.2008.11.022
- Chou H, Donnard E, Gustafsson HT, Garber M, Rando OJ, Chou H, Donnard E, Gustafsson HT, Garber M, Rando OJ. 2017. Transcriptome-wide Analysis of Roles for tRNA Modifications in Translational Regulation. *Mol Cell* **68**:978-992.e4. doi:10.1016/j.molcel.2017.11.002
- Chowdhury A, Mukhopadhyay J, Tharun S. 2007. The decapping activator Lsm1p-7p-Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. *RNA* **13**:998–1016. doi:10.1261/rna.502507
- Clark WC, Evans ME, Dominissini D, Zheng G, Pan T. 2016. tRNA base methylation identification and quantification via high-throughput sequencing. *RNA* **22**:1771–1784. doi:10.1261/rna.056531.116
- Colgan DF, Manley JL. 1997. Mechanism and regulation of mRNA polyadenylation. *Genes Dev* **11**:2755–2766.
- Coots RA, Liu X, Coots RA, Liu X, Mao Y, Dong L, Zhou J, Wan J, Zhang X. 2017. m6A Facilitates eIF4F-Independent mRNA Translation. *Mol Cell* **68**:504-514.e7. doi:10.1016/j.molcel.2017.10.002
- Corden JL, Cadena DL, Ahearn JMJ, Dahmus ME. 1985. A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proc Natl Acad Sci U S A* **82**:7934–7938.
- Core LJ, Waterfall JJ, Lis JT. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**:1845–1848. doi:10.1126/science.1162228
- Creamer TJ, Darby MM, Jamonnak N, Schaughency P, Hao H, Wheelan SJ, Corden JL. 2011. Transcriptome-wide binding sites for components of the *Saccharomyces cerevisiae* non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. *PLoS Genet* **7**. doi:10.1371/journal.pgen.1002329
- Cribbs DL, Gillam IC, Tener GM. 1987. Nucleotide sequences of three tRNA^{Ser} from *Drosophila melanogaster* reading the six serine codons. *J Mol Biol* **197**:389–395. doi:10.1016/0022-2836(87)90552-3
- Crick F. 1970. Central Dogma of Molecular Biology. *Nature* **227**:561–563. doi:10.1038/227561a0
- D’Silva SD, Haider SJ, Phizicky EM. 2011. A domain of the actin binding protein Abp140 is the yeast methyltransferase responsible for 3-methylcytidine modification in the tRNA anti-codon loop. *RNA* **17**:1100–1110. doi:10.1261/rna.2652611.reading
- Delan-Forino C, Schneider C, Tollervey D. 2017. Transcriptome-wide analysis of

- alternative routes for RNA substrates into the exosome complex. *PLoS Genet* **13**:e1006699. doi:10.1371/journal.pgen.1006699
- Delatte B, Wang F, Ngoc LV, Collignon E, Bonvin E, Deplus R, Calonne E, Hassabi B, Putmans P, Awe S, Wetzel C, Kreher J, Soin R, Creppe C, Limbach PA, Gueydan C, Kruys VV, Brehm A, Minakhina S, Defrance M, Steward R, Fuks FF. 2016. RNA biochemistry. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Science* **351**:282–285. doi:10.1126/science.aac5253
- Demeshkina N, Jenner L, Yusupova G, Yusupov M. 2010. Interactions of the ribosome with mRNA and tRNA. *Curr Opin Struct Biol* **20**:325–332. doi:10.1016/j.sbi.2010.03.002
- Dieci G, Sentenac A. 1996. Facilitated recycling pathway for RNA polymerase III. *Cell* **84**:245–252.
- Dobin A, Gingeras TR. 2015. Mapping RNA-seq Reads with STAR. *Curr Protoc Bioinforma* **51**:11.14.1-19. doi:10.1002/0471250953.bi1114s51
- Doma MK, Parker R. 2007. RNA quality control in eukaryotes. *Cell* **131**:660–668. doi:10.1016/j.cell.2007.10.041
- Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, Cesarkas K, Jacob-Hirsch J, Amariglio N, Kupiec M, Sorek R, Rechavi G. 2012. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**:201–206. doi:10.1038/nature11112
- Dominissini D, Nachtergaele S, Moshitch-Moshkovitz S, Peer E, Kol N, Ben-Haim MS, Dai Q, Di Segni A, Salmon-Divon M, Clark WC, Zheng G, Pan T, Solomon O, Eyal E, Hershkovitz V, Han D, Dore LC, Amariglio N, Rechavi G, He C. 2016. The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **530**:441–446. doi:10.1038/nature16998
- Dubin DT, Taylor RH. 1975. The methylation state of poly A-containing messenger RNA from cultured hamster cells. *Nucleic Acids Res* **2**:1653–1668.
- Duncan T, Trewick SC, Koivisto P, Bates PA, Lindahl T, Sedgwick B. 2002. Reversal of DNA alkylation damage by two human dioxygenases. *Proc Natl Acad Sci U S A* **99**:16660–16665. doi:10.1073/pnas.262589799
- Edelheit S, Schwartz S, Mumbach MR, Wurtzel O, Sorek R. 2013. Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. *PLoS Genet* **9**:e1003602. doi:10.1371/journal.pgen.1003602
- Engel M, Eggert C, Kaplick PM, Eder M, Röh S, Tietze L, Namendorf C, Arloth J, Weber P, Rex-Haffner M, Geula S, Jakovcevski M, Hanna JH, Leshkowitz D,

- Uhr M, Wotjak CT, Schmidt M V, Deussing JM, Binder EB, Chen A. 2018. The Role of m(6)A/m-RNA Methylation in Stress Response Regulation. *Neuron* **99**:389-403.e9. doi:10.1016/j.neuron.2018.07.009
- Engel SR, Dietrich FS, Fisk DG, Binkley G, Balakrishnan R, Costanzo MC, Dwight SS, Hitz BC, Karra K, Nash RS, Weng S, Wong ED, Lloyd P, Skrzypek MS, Miyasato SR, Simison M, Cherry JM. 2013. The reference genome sequence of *Saccharomyces cerevisiae*: then and now. *G3 (Bethesda)* **4**:389–398. doi:10.1534/g3.113.008995
- Esberg A, Huang B, Johansson MJO, Bystro AS. 2006. Elevated Levels of Two tRNA Species Bypass the Requirement for Elongator Complex in Transcription and Exocytosis 139–148. doi:10.1016/j.molcel.2006.07.031
- Falk S, Weir JR, Hentschel J, Reichelt P, Bonneau F, Conti E. 2014. The Molecular Architecture of the TRAMP Complex Reveals the Organization and Interplay of Its Two Catalytic Activities. *Mol Cell* **55**:856–867. doi:10.1016/j.molcel.2014.07.020
- Farabaugh PJ, Kramer E, Vallabhaneni H, Raman A. 2006. Evolution of +1 programmed frameshifting signals and frameshift-regulating tRNAs in the order saccharomycetales. *J Mol Evol* **63**:545–561. doi:10.1007/s00239-005-0311-0
- Finoux A-L, Seraphin B. 2006. In vivo targeting of the yeast Pop2 deadenylase subunit to reporter transcripts induces their rapid degradation and generates new decay intermediates. *J Biol Chem* **281**:25940–25947. doi:10.1074/jbc.M600132200
- Fong N, Bentley DL. 2001. Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. *Genes Dev* **15**:1783–1795. doi:10.1101/gad.889101
- Fox EJ, Reid-Bayliss KS, Emond MJ, Loeb LA. 2014. Accuracy of Next Generation Sequencing Platforms. *Next Gener Seq Appl* **1**:1000106. doi:10.4172/jngsa.1000106
- Franks TM, Lykke-Andersen J. 2008. The Control of mRNA Decapping and P-Body Formation. *Mol Cell* **32**:605–615. doi:10.1016/j.molcel.2008.11.001
- Franks TM, Singh G, Lykke-Andersen J. 2010. Upf1 ATPase-dependent mRNP disassembly is required for completion of nonsense- mediated mRNA decay. *Cell* **143**:938–950. doi:10.1016/j.cell.2010.11.043
- Frye M, Jaffrey SR, Pan T, Rechavi G, Suzuki T. 2016. RNA modifications: what have we learned and where are we headed? *Nat Rev Genet* **17**:365.
- Fustin J-M, Doi M, Yamaguchi Y, Hida H, Nishimura S, Yoshida M, Isagawa T, Morioka MS, Kakeya H, Manabe I, Okamura H. 2013. RNA-methylation-

- dependent RNA processing controls the speed of the circadian clock. *Cell* **155**:793–806. doi:10.1016/j.cell.2013.10.026
- Gallie DR. 1991. The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. *Genes Dev* **5**:2108–2116.
- Gao Y, Selmer M, Dunham CM, Weixlbaumer A, Kelley AC, Ramakrishnan V. 2009. The Structure of the Ribosome with Elongation Factor G Trapped in the Posttranslocational State. *Science* **326**:694 LP – 699. doi:10.1126/science.1179709
- Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. 2000. Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* **11**:4241–4257. doi:10.1091/mbc.11.12.4241
- Gasch AP, Werner-Washburne M. 2002. The genomics of yeast responses to environmental stress and starvation. *Funct Integr Genomics* **2**:181–192. doi:10.1007/s10142-002-0058-2
- Gerber A, Grosjean H, Melcher T, Keller W. 1998. Tad1p, a yeast tRNA-specific adenosine deaminase, is related to the mammalian pre-mRNA editing enzymes ADAR1 and ADAR2. *EMBO J* **17**:4780–4789. doi:10.1093/emboj/17.16.4780
- Gilmartin GM, Nevins JR. 1989. An ordered pathway of assembly of components required for polyadenylation site recognition and processing. *Genes Dev* **3**:2180–2190.
- Goetz AE, Wilkinson M. 2017. Stress and the nonsense-mediated RNA decay pathway. *Cell Mol Life Sci* **74**:3509–3531. doi:10.1007/s00018-017-2537-6
- Goldstrohm AC, Hook BA, Seay DJ, Wickens M. 2006. PUF proteins bind Pop2p to regulate messenger RNAs. *Nat Struct Mol Biol* **13**:533–539. doi:10.1038/nsmb1100
- Green R, Noller HF. 1997. Ribosomes and translation. *Annu Rev Biochem* **66**:679–716. doi:10.1146/annurev.biochem.66.1.679
- Gressel S, Schwalb B, Decker TM, Qin W, Leonhardt H, Eick D, Cramer P. 2017. CDK9-dependent RNA polymerase II pausing controls transcription initiation. *Elife* **6**. doi:10.7554/eLife.29736
- Grosshans H, Simos G, Hurt E. 2000. Review: transport of tRNA out of the nucleus—direct channeling to the ribosome? *J Struct Biol* **129**:288–294. doi:10.1006/jsbi.2000.4226
- Grzechnik P, Kufel J. 2008. Polyadenylation Linked to Transcription Termination Directs the Processing of snoRNA Precursors in Yeast. *Mol Cell* **32**:247–258. doi:10.1016/j.molcel.2008.10.003

- Hahn S. 2004. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol* **11**:394–403. doi:10.1038/nsmb763
- Halbach F, Reichelt P, Rode M, Conti E. 2013. The yeast ski complex: Crystal structure and rna channeling to the exosome complex. *Cell* **154**:814–826. doi:10.1016/j.cell.2013.07.017
- Han D, Liu J, Chen C, Dong L, Liu Yi, Chang R, Huang X, Liu Yuanyuan, Wang J, Dougherty U, Bissonnette MB, Shen B, Weichselbaum RR, Xu MM, He C. 2019. Anti-tumour immunity controlled through mRNA m6A methylation and YTHDF1 in dendritic cells. *Nature* **566**:270–274. doi:10.1038/s41586-019-0916-x
- Han LU, Marcus E, D'Silva S, Phizicky EM, Silva SD, Phizicky EM, D'Silva S, Phizicky EM. 2017. *S. cerevisiae* Trm140 has two recognition modes for 3-methylcytidine modification of the anticodon loop of tRNA substrates. *RNA* **23**:406–419. doi:10.1261/rna.059667.116
- Hanson G, Collier J. 2018. Translation and Protein Quality Control: Codon optimality, bias and usage in translation and mRNA decay. *Nat Rev Mol Cell Biol* **19**:20–30. doi:10.1038/nrm.2017.91
- Hantsche M, Cramer P. 2016. The Structural Basis of Transcription: 10 Years After the Nobel Prize in Chemistry. *Angew Chemie Int Ed* **55**:15972–15981. doi:10.1002/anie.201608066
- He PC, He C. 2019. mRNA acetylation: a new addition to the epitranscriptome. *Cell Res* **29**:91–92. doi:10.1038/s41422-018-0135-2
- Heck AM, Wilusz J. 2018. The Interplay between the RNA Decay and Translation Machinery in Eukaryotes. *Cold Spring Harb Perspect Biol* a032839. doi:10.1101/cshperspect.a032839
- Hellen CUT. 2018. Translation Termination and Ribosome Recycling in Eukaryotes. *Cold Spring Harb Perspect Biol* **10**. doi:10.1101/cshperspect.a032656
- Heo D, Yoo I, Kong J, Lidschreiber M, Mayer A, Choi B-Y, Hahn Y, Cramer P, Buratowski S, Kim M. 2013. The RNA polymerase II C-terminal domain-interacting domain of yeast Nrd1 contributes to the choice of termination pathway and couples to RNA processing by the nuclear exosome. *J Biol Chem* **288**:36676–36690. doi:10.1074/jbc.M113.508267
- Hilgers V, Teixeira D, Parker R. 2006. Translation-independent inhibition of mRNA deadenylation during stress in *Saccharomyces cerevisiae*. *RNA* **12**:1835–1845. doi:10.1261/rna.241006
- Hirose Y, Manley JL. 2000. RNA polymerase II and the integration of nuclear events. *Genes Dev* **14**:1415–1429. doi:10.1101/gad.14.12.1415
- Hoernes TP, Clementi N, Faserl K, Glasner H, Erlacher MD, Breuker K, Lindner H,

- Alexander H. 2016. Nucleotide modifications within bacterial messenger RNAs regulate their translation and are able to rewire the genetic code. *Nucleic Acids Res* **44**:852–862. doi:10.1093/nar/gkv1182
- Hongay CF, Orr-weaver TL. 2011. Drosophila Inducer of MEiosis 4 (IME4) is required for Notch signaling during oogenesis. *Proc Natl Acad Sci* **108**:14855–14860. doi:10.1073/pnas.1111577108
- Hopper AK. 2013. Transfer RNA post-transcriptional processing, turnover, and subcellular dynamics in the yeast *Saccharomyces cerevisiae*. *Genetics* **194**:43–67. doi:10.1534/genetics.112.147470
- Houseley J, Tollervey D. 2009. The Many Pathways of RNA Degradation. *Cell* **136**:763–776. doi:10.1016/j.cell.2009.01.019
- Hsu PJ, Shi H, He C. 2017. Epitranscriptomic influences on development and disease. *Genome Biol* **18**:1–9. doi:10.1186/s13059-017-1336-6
- Huang H, Weng H, Zhou K, Wu T, Zhao BS, Sun Mingli, Chen Z, Deng X, Xiao G, Auer F, Klemm L, Wu H, Zuo Z, Qin X, Dong Y, Zhou Y, Qin H, Tao S, Du J, Liu J, Lu Z, Yin H, Mesquita A, Yuan CL, Hu Y-C, Sun W, Su R, Dong L, Shen C, Li C, Qing Y, Jiang X, Wu X, Sun Miao, Guan J-L, Qu L, Wei M, Müschen M, Huang G, He C, Yang J, Chen J. 2019. Histone H3 trimethylation at lysine 36 guides m6A RNA modification co-transcriptionally. *Nature*. doi:10.1038/s41586-019-1016-7
- Huang L, Wilkinson MF. 2012. Regulation of nonsense-mediated mRNA decay. *Wiley Interdiscip Rev RNA* **3**:807–828. doi:10.1002/wrna.1137
- Huang Y, Carmichael GG. 1996. Role of polyadenylation in nucleocytoplasmic transport of mRNA. *Mol Cell Biol* **16**:1534–1542.
- Huber SM, Delft P Van, Mendil L, Bachman M, Smollett K, Werner F, Miska EA, Balasubramanian S. 2015. Formation and Abundance of 5-Hydroxymethylcytosine in RNA. *Chembiochem* **16**:752–755. doi:10.1002/cbic.201500013
- Huch S, Nissan T. 2014. Interrelations between translation and general mRNA degradation in yeast. *Wiley Interdiscip Rev RNA* **5**:747–763. doi:10.1002/wrna.1244
- Hussain S. 2018. Native RNA-Sequencing Throws its Hat into the Transcriptomics Ring. *Trends Biochem Sci* **43**:225–227. doi:10.1016/j.tibs.2018.02.007
- Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. 2009. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science* **324**:218–224.
- Ito S, Horikawa S, Suzuki TTT, Kawauchi H, Tanaka Y, Suzuki TTT, Suzuki TTT.

2014. Human NAT10 is an ATP-dependent RNA acetyltransferase responsible for N4-acetylcytidine formation in 18 S ribosomal RNA (rRNA). *J Biol Chem* **289**:35724–35730. doi:10.1074/jbc.C114.602698
- Jackson RJ, Hellen CUT, Pestova T V. 2012. Termination and Post-Termination Events in Eukaryotic Translation. *Adv Protein Chem Struct Biol* **86**:45-93ab.
- Jantsch MF, Schaefer MR. 2019. Mining the Epitranscriptome: Detection of RNA editing and RNA modifications. *Methods* **156**:1–4.
doi:https://doi.org/10.1016/j.ymeth.2019.02.016
- Jelinsky SA, Samson LD. 1999. Global response of *Saccharomyces cerevisiae* to an alkylating agent. *Proc Natl Acad Sci U S A* **96**:1486–1491.
- Jia G, Fu Y, Zhao X, Dai Q, Zheng G, Yang Y, Yi C, Lindahl T, Pan T, Yang Y-G, He C. 2011. N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol* **7**:885–887. doi:10.1038/nchembio.687
- Jiang H, Lei R, Ding S-W, Zhu S. 2014. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* **15**:182.
doi:10.1186/1471-2105-15-182
- Jiao X, Xiang S, Oh C, Martin CE, Tong L, Kiledjian M. 2010. Identification of a quality-control mechanism for mRNA 5'-end capping. *Nature* **467**:608–611.
- Johansson MJO, Esberg A, Huang B, Björk GR, Byström AS. 2008. Eukaryotic Wobble Uridine Modifications Promote a Functionally Redundant Decoding System. *Mol Cell Biol* **28**:3301–3312. doi:10.1128/MCB.01542-07
- Johnson SJ, Jackson RN. 2013. Ski2-like RNA helicase structures: common themes and complex assemblies. *RNA Biol* **10**:33–43. doi:10.4161/ma.22101
- Jona G, Choder M, Gileadi O. 2000. Glucose starvation induces a drastic reduction in the rates of both transcription and degradation of mRNA in yeast. *Biochim Biophys Acta - Gene Struct Expr* **1491**:37–48.
doi:https://doi.org/10.1016/S0167-4781(00)00016-6
- Karam R, Lou C-H, Kroeger H, Huang L, Lin JH, Wilkinson MF. 2015. The unfolded protein response is shaped by the NMD pathway. *EMBO Rep* **16**:599–609.
doi:10.15252/embr.201439696
- Karam R, Wengrod J, Gardner LB, Wilkinson MF. 2013. Regulation of nonsense-mediated mRNA decay: implications for physiology and disease. *Biochim Biophys Acta* **1829**:624–633. doi:10.1016/j.bbagr.2013.03.002
- Karijolich J, Yu Y-T. 2011. Converting nonsense codons into sense codons by targeted pseudouridylation. *Nature* **474**:395–398. doi:10.1038/nature10165
- Kariko K, Muramatsu H, Welsh FA, Ludwig J, Kato H, Akira S, Weissman D. 2008. Incorporation of pseudouridine into mRNA yields superior nonimmunogenic

- vector with increased translational capacity and biological stability. *Mol Ther* **16**:1833–1840. doi:10.1038/mt.2008.200
- Keith G. 1984. The primary structures of two arginine tRNAs (anticodons C-C-U and mcm5a2U-C-psi) and of glutamine tRNA (anticodon C-U-G) from bovine liver. *Nucleic Acids Res* **12**:2543–2548.
- Kilchert C, Spang A. 2011. Cotranslational transport of ABP140 mRNA to the distal pole of *S. cerevisiae*. *EMBO J* **30**:3567–3580. doi:10.1038/emboj.2011.247
- Kilchert C, Wittmann S, Vasiljeva L. 2016. The regulation and functions of the nuclear RNA exosome complex. *Nat Rev Mol Cell Biol* **17**:227.
- Kizer KO, Phatnani HP, Shibata Y, Hall H, Greenleaf AL, Strahl BD, Iol MOLCELLB. 2005. A Novel Domain in Set2 Mediates RNA Polymerase II Interaction and Couples Histone H3 K36 Methylation with Transcript Elongation. *Mol Cell Biol* **25**:3305–3316. doi:10.1128/MCB.25.8.3305
- Klose RJ, Bird AP. 2006. Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* **31**:89–97. doi:10.1016/j.tibs.2005.12.008
- Kowalinski E, Kögel A, Ebert J, Reichelt P, Stegmann E, Habermann B, Conti E. 2016. Structure of a Cytoplasmic 11-Subunit RNA Exosome Complex. *Mol Cell* **63**:125–134. doi:10.1016/j.molcel.2016.05.028
- Kretschmer J, Rao H, Hackert P, Sloan KE, Höbartner C, Bohnsack MT. 2018. The m6A reader protein YTHDC2 interacts with the small ribosomal subunit and the 5' -3' exoribonuclease XRN1. *RNA* **24**:1339–1350. doi:10.1261/ma.064238.117
- Kruk JA, Dutta A, Fu J, Gilmour DS, Reese JC. 2011. The multifunctional Ccr4-Not complex directly promotes transcription elongation. *Genes Dev* **25**:581–593. doi:10.1101/gad.2020911
- Lai T, Cho H, Liu Z, Bowler MW, Piao S, Parker R, Kim YK, Song H. 2012. Structural basis of the PNRC2-mediated link between mRNA surveillance and decapping. *Structure* **20**:2025–2037. doi:10.1016/j.str.2012.09.009
- Lebreton A, Seraphin B. 2008. Exosome-mediated quality control: substrate recruitment and molecular activity. *Biochim Biophys Acta* **1779**:558–565. doi:10.1016/j.bbagrm.2008.02.003
- Lee D-H, Jin S-G, Cai S, Chen Y, Pfeifer GP, O'Connor TR. 2005. Repair of methylation damage in DNA and RNA by mammalian AlkB homologues. *J Biol Chem* **280**:39448–39459. doi:10.1074/jbc.M509881200
- Lee S, Kim J. 2016. NGS-based deep bisulfite sequencing. *MethodsX* **3**:1–7. doi:10.1016/j.mex.2015.11.008
- Lence T, Akhtar J, Bayer M, Schmid K, Spindler L, Ho CH, Kreim N, Andrade-

- Navarro MA, Poeck B, Helm M, Roignant J-Y. 2016. m(6)A modulates neuronal functions and sex determination in *Drosophila*. *Nature* **540**:242–247. doi:10.1038/nature20568
- Li A, Chen Y-S, Ping X-L, Yang X, Xiao W, Yang Y, Sun H-Y, Zhu Q, Baidya P, Wang X, Bhattarai DP, Zhao Y-L, Sun B-F, Yang Y-G. 2017. Cytoplasmic m6A reader YTHDF3 promotes mRNA translation. *Cell Res* **27**:444.
- Li F, Zhao D, Wu J, Shi Y. 2014. Structure of the YTH domain of human YTHDF2 in complex with an m(6)A mononucleotide reveals an aromatic cage for m(6)A recognition. *Cell Res*. doi:10.1038/cr.2014.153
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079. doi:10.1093/bioinformatics/btp352
- Li T, Shi Y, Wang P, Guachalla LM, Sun B, Joerss T, Chen Y-S, Groth M, Krueger A, Platzer M, Yang Y-G, Rudolph KL, Wang Z-Q. 2015. Smg6/Est1 licenses embryonic stem cell differentiation via nonsense-mediated mRNA decay. *EMBO J* **34**:1630–1647. doi:10.15252/embj.201489947
- Li X, Xiong X. 2017. Base-Resolution Mapping Reveals Distinct m¹A Methylome in Nuclear- and Mitochondrial-Encoded Technology Base-Resolution Mapping Reveals Distinct m¹A Methylome in Nuclear- and Mitochondrial-Encoded Transcripts. *Mol Cell* **68**:993-1005.e9. doi:10.1016/j.molcel.2017.10.019
- Li X, Xiong X, Wang K, Wang L, Shu X, Ma S, Yi C. 2016a. Transcriptome-wide mapping reveals reversible and dynamic N¹-methyladenosine methylome. *Nat Chem Biol* **12**:311–316. doi:10.1038/nchembio.2040
- Li X, Xiong X, Yi C. 2016b. Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat Methods* **14**:23–31. doi:10.1038/nmeth.4110
- Li X, Zhu P, Ma S, Song J, Bai J, Sun F, Yi C. 2015. Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat Chem Biol* **11**:592–597. doi:10.1038/nchembio.1836
- Linder B, Grozhik A V, Olarerin-george AO, Meydan C, Mason CE, Jaffrey SR. 2015. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat Meth* **12**:767–772. doi:10.1038/nmeth.3453
- Liu H, Begik O, Lucas MC, Christopher E, Schwartz S, Mattick JS, Smith MA, Novoa M. 2019. Accurate detection of m6A RNA modifications in native RNA sequences. *bioRxiv*. doi:https://doi.org/10.1101/525741
- Liu J, Jia G. 2014. Methylation Modifications in Eukaryotic Messenger RNA. *J Genet Genomics* **41**:21–33. doi:10.1016/j.jgg.2013.10.002
- Liu N, Dai Q, Zheng G, He C, Parisien M, Pan T. 2015. N⁶-methyladenosine-

- dependent RNA structural switches regulate RNA-protein interactions. *Nature* **518**:560–564. doi:10.1038/nature14234
- Liu Q, Greimann JC, Lima CD. 2006. Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell* **127**:1223–1237. doi:10.1016/j.cell.2006.10.037
- Lorenz C, Lünse CE, Mörl M. 2017. tRNA Modifications: Impact on Structure and Thermal Adaptation. *Biomolecules* **7**:35. doi:10.3390/biom7020035
- Losh JS, King AK, Bakelar J, Taylor L, Loomis J, Rosenzweig JA, Johnson SJ, Hoof A Van. 2015. Interaction between the RNA-dependent ATPase and poly(A) polymerase subunits of the TRAMP complex is mediated by short peptides and important for snoRNA processing. *Nucleic Acids Res* **43**:1848–1858. doi:10.1093/nar/gkv005
- Losson R, Lacroute F. 1979. Interference of nonsense mutations with eukaryotic messenger RNA stability. *Proc Natl Acad Sci U S A* **76**:5134–5137.
- Lou C-H, Dumdie J, Goetz A, Shum EY, Brafman D, Liao X, Mora-Castilla S, Ramaiah M, Cook-Andersen H, Laurent L, Wilkinson MF. 2016. Nonsense-Mediated RNA Decay Influences Human Embryonic Stem Cell Fate. *Stem cell reports* **6**:844–857. doi:10.1016/j.stemcr.2016.05.008
- Lovejoy AF, Riordan DP, Brown PO. 2014. Transcriptome-Wide Mapping of Pseudouridines: Pseudouridine Synthases Modify Specific mRNAs in *S. cerevisiae*. *PLoS One* **9**:e110799. doi:10.1371/journal.pone.0110799
- Luehr S, Hartmann H, Söding J. 2012. The XXmotif web server for eXhaustive, weight matrix-based motif discovery in nucleotide sequences. *Nucleic Acids Res* **40**:W104–W109. doi:10.1093/nar/gks602
- Lund E, Dahlberg JE. 1998. Proofreading and aminoacylation of tRNAs before export from the nucleus. *Science* **282**:2082–2085.
- Luo S, Tong L. 2014. Molecular basis for the recognition of methylated adenines in RNA by the eukaryotic YTH domain. *Proc Natl Acad Sci U S A* **111**:13834–13839. doi:10.1073/pnas.1412742111
- Lykke-Andersen S, Brodersen DE, Jensen TH. 2009. Origins and activities of the eukaryotic exosome. *J Cell Sci* **122**:1487–1494. doi:10.1242/jcs.047399
- Lykke-Andersen S, Jensen TH. 2015. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nat Rev Mol Cell Biol* **16**:665–677. doi:10.1038/nrm4063
- Marquardt S, Hazelbaker DZ, Buratowski S. 2011. Distinct RNA degradation pathways and 3' extensions of yeast non-coding RNA species. *Transcription* **2**:145–154. doi:10.4161/trns.2.3.16298
- McCracken S, Fong N, Rosonina E, Yankulov K, Brothers G, Siderovski D, Hessel A,

- Foster S, Shuman S, Bentley DL. 1997a. 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev* **11**:3306–3318.
- McCracken S, Fong N, Yankulov K, Ballantyne S, Pan G, Greenblatt J, Patterson SD, Wickens M, Bentley DL. 1997b. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**:357–361.
doi:10.1038/385357a0
- Melton DA, De Robertis EM, Cortese R. 1980. Order and intracellular location of the events involved in the maturation of a spliced tRNA. *Nature* **284**:143–148.
- Meyer KD, Patil DP, Zhou J, Zinoviev A, Skabkin MA, Elemento O, Pestova T V, Qian S-B, Jaffrey SR. 2015. 5' UTR m(6)A Promotes Cap-Independent Translation. *Cell* **163**:999–1010. doi:10.1016/j.cell.2015.10.012
- Mignone F, Gissi C, Liuni S, Pesole G. 2002. Untranslated regions of mRNAs. *Genome Biol* **3**.
- Miller C, Schwalb B, Maier K, Schulz D, Dümcke S, Zacher B, Mayer A, Sydow J, Marciniowski L, Dölken L, Martin DE, Tresch A, Cramer P. 2011. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol Syst Biol* **7**. doi:10.1038/msb.2010.112
- Miller JE, Reese JC. 2012. Ccr4-Not complex: The control freak of eukaryotic cells. *Crit Rev Biochem Mol Biol* **47**:315–333. doi:10.3109/10409238.2012.667214
- Milligan L, Decourty L, Saveanu C, Rappsilber J, Ceulemans H, Jacquier A, Tollervey D. 2008. A Yeast Exosome Cofactor, Mpp6, Functions in RNA Surveillance and in the Degradation of Noncoding RNA Transcripts. *Mol Cell Biol* **28**:5446–5457. doi:10.1128/MCB.00463-08
- Milligan L, Huynh-Thu VA, Delan-Forino C, Tuck A, Petfalski E, Lombraña R, Sanguinetti G, Kudla G, Tollervey D. 2016. Strand-specific, high-resolution mapping of modified RNA polymerase II. *Mol Syst Biol* **12**:874.
doi:10.15252/msb.20166869
- Mitchell P, Petfalski E, Houalla R, Podtelejnikov A, Mann M, Tollervey D. 2003. Rrp47p is an exosome-associated protein required for the 3' processing of stable RNAs. *Mol Cell Biol* **23**:6982–6992.
- Mitchell P, Tollervey D. 2003. An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3'→5' degradation. *Mol Cell* **11**:1405–1413.
- Moore CL, Sharp PA. 1985. Accurate cleavage and polyadenylation of exogenous RNA substrate. *Cell* **41**:845–855.
- Morrissey JP, Deardorff JA, Hebron C, Sachs AB. 1999. Decapping of stabilized,

- polyadenylated mRNA in yeast pab1 mutants. *Yeast* **15**:687–702.
- Muhrad D, Parker R. 1999. Aberrant mRNAs with extended 3' UTRs are substrates for rapid degradation by mRNA surveillance. *RNA* **5**:1299–1307.
- Muhrad D, Parker R. 1994. Premature translational termination triggers mRNA decapping. *Nature* **370**:578–581. doi:10.1038/370578a0
- Murphy F V, Ramakrishnan V. 2004. Structure of a purine-purine wobble base pair in the decoding center of the ribosome. *Nat Struct Mol Biol* **11**:1251–1252.
- Murray SC, Serra Barros A, Brown DA, Dudek P, Ayling J, Mellor J. 2012. A pre-initiation complex at the 3'-end of genes drives antisense transcription independent of divergent sense transcription. *Nucleic Acids Res* **40**:2432–2444. doi:10.1093/nar/gkr1121
- Nabet B, Roberts JM, Buckley DL, Paulk J, Dastjerdi S, Yang A, Leggett AL, Erb MA, Lawlor MA, Souza A, Scott TG, Vittori S, Perry JA, Qi J, Winter GE, Wong K-K, Gray NS, Bradner JE. 2018. The dTAG system for immediate and target-specific protein degradation. *Nat Chem Biol* **14**:431–441. doi:10.1038/s41589-018-0021-8
- Nechaev S, Adelman K. 2011. Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. *Biochim Biophys Acta* **1809**:34–45. doi:10.1016/j.bbagr.2010.11.001
- Nedialkova DD, Leidel SA. 2015. Optimization of Codon Translation Rates via tRNA Modifications Maintains Proteome Integrity. *Cell* **161**:1606–1618. doi:10.1016/j.cell.2015.05.022
- Neil H, Malabat C, d'Aubenton-Carafa Y, Xu Z, Steinmetz LM, Jacquier A. 2009. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* **457**:1038–1042. doi:10.1038/nature07747
- Nelson JO, Moore KA, Chapin A, Hollien J, Metzstein MM. 2016. Degradation of Gadd45 mRNA by nonsense-mediated decay is essential for viability. *Elife* **5**. doi:10.7554/eLife.12876
- Noma A, Yi S, Kato T, Takai Y, Suzuki T. 2011. Actin-binding protein ABP140 is a methyltransferase for 3-methylcytidine at position 32 of tRNAs in *Saccharomyces cerevisiae*. *RNA* **17**:1111–1119. doi:10.1261/rna.2653411.1968
- Parker R. 2012. RNA degradation in *Saccharomyces cerevisiae*. *Genetics* **191**:671–702. doi:10.1534/genetics.111.137265
- Parker R, Song H. 2004. The enzymes and control of eukaryotic mRNA turnover. *Nat Struct Mol Biol* **11**:121–127. doi:10.1038/nsmb724
- Partial T, Digestion E, Ginsberg T, Rogg H. 1971. Nucleotide Sequences of Rat Liver Serine- tRNA and Derivation of its Total Primary Structure. *Eur J Biochem*

- 21:249–257.
- Pechmann S, Frydman J. 2013. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat Struct Mol Biol* **20**:237–243. doi:10.1038/nsmb.2466
- Pelechano V, Wei W, Steinmetz LM. 2013. Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* **497**:127–131. doi:10.1038/nature12121
- Pelechano V, Wei W, Steinmetz LMM. 2015. Widespread co-translational RNA decay reveals ribosome dynamics. *Cell* **161**:1400–1412. doi:10.1016/j.cell.2015.05.008
- Pendleton KE, Chen B, Liu K, Hunter O V, Xie Y, Tu BP, Conrad NK. 2017. The U6 snRNA m(6)A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. *Cell* **169**:824–835.e14. doi:10.1016/j.cell.2017.05.003
- Pestova T V, Kolupaeva VG, Lomakin IB, Pilipenko E V, Shatsky IN, Agol VI, Hellen CUT. 2001. Molecular mechanisms of translation initiation in eukaryotes. *Proc Natl Acad Sci* **98**:7029 LP – 7036. doi:10.1073/pnas.111145798
- Phizicky EM, Hopper AK. 2010. tRNA biology charges to the front. *Genes Dev* **24**:1832–1860. doi:10.1101/gad.1956510
- Ping X-L, Sun B-F, Wang L, Xiao W, Yang X, Wang W-J, Adhikari S, Shi Y, Lv Y, Chen Y-S, Zhao X, Li A, Yang Y, Dahal U, Lou X-M, Liu X, Huang J, Yuan W-P, Zhu X-F, Cheng T, Zhao Y-L, Wang X, Rendtlew Danielsen JM, Liu F, Yang Y-G. 2014. Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res* **24**:177–189. doi:10.1038/cr.2014.3
- Powell LM, Wallis SC, Pease RJ, Edwards YH, Knott TJ, Scott J. 1987. A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell* **50**:831–840.
- Presnyak V, Alhusaini N, Chen Y-H, Martin S, Morris N, Kline N, Olson S, Weinberg D, Baker KE, Graveley BR, Collier J. 2015. Codon Optimality Is a Major Determinant of mRNA Stability. *Cell* **160**:1111–1124. doi:10.1016/J.CELL.2015.02.029
- Proudfoot NJ, Furger A, Dye MJ. 2002. Integrating mRNA processing with transcription. *Cell* **108**:501–512.
- Radhakrishnan A, Chen YH, Martin S, Alhusaini N, Green R, Collier J. 2016. The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation by Monitoring Codon Optimality. *Cell* **167**:122–132.e9. doi:10.1016/j.cell.2016.08.053
- Rasmussen EB, Lis JT. 1993. In vivo transcriptional pausing and cap formation on three *Drosophila* heat shock genes. *Proc Natl Acad Sci U S A* **90**:7923–7927.

- Rhee HS, Pugh BF. 2012. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**:295–301. doi:10.1038/nature10799
- Roth C, Torkler P. 2018. soedinglab/mockinbird: Degradation PAR-CLIP. doi:10.5281/ZENODO.1342555
- Safra M, Nir R, Farouq D, Slutzkin IV. 2017a. TRUB1 is the predominant pseudouridine synthase acting on mammalian mRNA via a predictable and conserved code. *Genome Res* **27**:393–406. doi:10.1101/gr.207613.116.5
- Safra M, Sas-Chen A, Nir R, Winkler R, Nachshon A, Bar-Yaacov D, Erlacher M, Rossmanith W, Stern-Ginossar N, Schwartz S. 2017b. The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature* **551**:251–255. doi:10.1038/nature24456
- Sainsbury S, Bernecky C, Cramer P. 2015. Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol* **16**:129–143. doi:10.1038/nrm3952
- Saletore Y, Meyer K, Korlach J, Vilfan ID, Jaffrey S, Mason CE. 2012. The birth of the Epitranscriptome: deciphering the function of RNA modifications. *Genome Biol* **13**:175. doi:10.1186/gb-2012-13-10-175
- Sarkar S, Hopper AK. 1998. tRNA nuclear export in *saccharomyces cerevisiae*: in situ hybridization analysis. *Mol Biol Cell* **9**:3041–3055.
- Schaeffer D, Tsanova B, Barbas A, Reis FP, Dastidar EG, Sanchez-Rotunno M, Arraiano CM, van Hoof A. 2009. The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. *Nat Struct Mol Biol* **16**:56–62. doi:10.1038/nsmb.1528
- Schibler U, Kelley DE, Perry RP. 1977. Comparison of methylated sequences in messenger RNA and heterogeneous nuclear RNA from mouse L cells. *J Mol Biol* **115**:695–714.
- Schmeing TM, Voorhees RM, Kelley AC, Gao Y-G, Murphy F V, Weir JR, Ramakrishnan V, Iv FVM, Weir JR, Ramakrishnan V. 2009. The Crystal Structure of the Ribosome Bound to EF-Tu and Aminoacyl-tRNA. *Science* **326**:688 LP – 694. doi:10.1126/science.1179700
- Schmeing TM, Voorhees RM, Kelley AC, Ramakrishnan V. 2011. How mutations in tRNA distant from the anticodon affect the fidelity of decoding. *Nat Struct & Mol Biol* **18**:432.
- Schmidt C, Kowalinski E, Shanmuganathan V, Defenouillère Q, Braunger K, Heuer A, Pech M, Namane A, Berninghausen O, Fromont-Racine M, Jacquier A, Conti E, Becker T, Beckmann R. 2016. The cryo-EM structure of a ribosome-Ski2-Ski3-Ski8 helicase complex. *Science* **354**:1431–1433. doi:10.1126/science.aaf7520

- Schmidt K, Butler JS. 2013. Nuclear RNA surveillance: Role of TRAMP in controlling exosome specificity. *Wiley Interdiscip Rev RNA* **4**:217–231. doi:10.1002/wrna.1155
- Schneider C, Kudla G, Wlotzka W, Tuck A, Tollervey D. 2012. Transcriptome-wide Analysis of Exosome Targets. *Mol Cell* **48**:422–433. doi:10.1016/j.molcel.2012.08.013
- Schneider C, Tollervey D. 2013. Threading the barrel of the RNA exosome. *Trends Biochem Sci* **38**:485–493. doi:10.1016/j.tibs.2013.06.013
- Schulz D, Schwalb B, Kiesel A, Baejen C, Torkler P, Gagneur J, Soeding J, Cramer P. 2013. Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell* **155**:1075–1087. doi:10.1016/j.cell.2013.10.024
- Schwalb B, Michel M, Zacher B, Fruhauf K, Demel C, Tresch A, Gagneur J, Cramer P. 2016. TT-seq maps the human transient transcriptome. *Science* **352**:1225–1228. doi:10.1126/science.aad9841
- Schwartz S. 2018. M1A within cytoplasmic mRNAs at single nucleotide resolution: A reconciled transcriptome-wide map. *RNA*.
- Schwartz S, Agarwala SD, Mumbach MR, Jovanovic M, Mertins P, Shishkin A, Tabach Y, Mikkelsen TS, Satija R, Ruvkun G, Carr SA, Lander ES, Fink GR, Regev A. 2013. Resource High-Resolution Mapping Reveals a Conserved , Widespread , Dynamic mRNA Methylation Program in Yeast Meiosis. *Cell* **155**:1409–1421. doi:10.1016/j.cell.2013.10.047
- Schwartz S, Bernstein D a, Mumbach MR, Jovanovic M, Herbst RH, Leon-Ricardo BX, Engreitz JM, Guttman M, Satija R, Lander ES, Fink G, Regev A. 2014a. Transcriptome-wide Mapping Reveals Widespread Dynamic-Regulated Pseudouridylation of ncRNA and mRNA. *Cell* **159**:148–162. doi:10.1016/j.cell.2014.08.028
- Schwartz S, Mumbach MR, Jovanovic M, Wang T, Maciag K, Bushkin GG, Mertins P, Ter-Ovanesyan D, Habib N, Cacchiarelli D, Sanjana NE, Freinkman E, Pacold ME, Satija R, Mikkelsen TS, Hacohen N, Zhang F, Carr SA, Lander ES, Regev A. 2014b. Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. *Cell Rep* **8**:284–296. doi:10.1016/j.celrep.2014.05.048
- Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA. 2008. Divergent transcription from active promoters. *Science* **322**:1849–1851. doi:10.1126/science.1162253
- Semotok JL, Cooperstock RL, Pinder BD, Vari HK, Lipshitz HD, Smibert CA. 2005. Smaug recruits the CCR4/POP2/NOT deadenylase complex to trigger maternal

- transcript localization in the early *Drosophila* embryo. *Curr Biol* **15**:284–294.
doi:10.1016/j.cub.2005.01.048
- Shaheen R, Tasak M, Maddirevula S, Abdel-Salam GMH, Sayed ISM, Alazami AM, Al-Sheddi T, Alobeid E, Phizicky EM, Alkuraya FS. 2019. PUS7 mutations impair pseudouridylation in humans and cause intellectual disability and microcephaly. *Hum Genet*. doi:10.1007/s00439-019-01980-3
- Sharma S, Langhendries J, Watzinger P, Peter K, Entian K, Lafontaine DLJ. 2015. Yeast Kre33 and human NAT10 are conserved 18S rRNA cytosine acetyltransferases that modify tRNAs assisted by the adaptor Tan1 / THUMPD1. *Nucleic Acids Res* **43**:2242–2258. doi:10.1093/nar/gkv075
- Shi H, Moore PB. 2000. The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: a classic structure revisited. *RNA* **6**:1091–1105.
- Smith AM, Jain M, Mulroney L, Geralde DR, Akeson M. 2017. Reading canonical and modified nucleotides in 16S ribosomal RNA using nanopore direct RNA sequencing. *bioRxiv*.
- Smith JE, Alvarez-Dominguez JR, Kline N, Huynh NJ, Geisler S, Hu W, Collier J, Baker KE. 2014. Translation of small open reading frames within unannotated RNA transcripts in *Saccharomyces cerevisiae*. *Cell Rep* **7**:1858–1866.
doi:10.1016/j.celrep.2014.05.023
- Smith T, Heger A, Sudbery I. 2017. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* **27**:491–499. doi:10.1101/gr.209601.116
- Squires JE, Patel HR, Nusch M, Sibbritt T, Humphreys DT, Parker BJ, Suter CM, Preiss T. 2012. Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res* **40**:5023–5033.
doi:10.1093/nar/gks144
- Steinberg S, Cedergren R. 1995. A correlation between N²-dimethylguanosine presence and alternate tRNA conformers. *RNA* **1**:886–891.
- Steiner-Mosonyi M, Mangroo D. 2004. The nuclear tRNA aminoacylation-dependent pathway may be the principal route used to export tRNA from the nucleus in *Saccharomyces cerevisiae*. *Biochem J* **378**:809–816. doi:10.1042/BJ20031306
- Steinmetz EJ, Conrad NK, Brow DA, Corden JL. 2001. RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature* **413**:327–331. doi:10.1038/35095090
- Stevens A, Poole TL. 1995. 5'-exonuclease-2 of *Saccharomyces cerevisiae*. Purification and features of ribonuclease activity with comparison to 5'-exonuclease-1. *J Biol Chem* **270**:16063–16069.

- Sun M, Schwalb B, Schulz D, Pirkl N, Etzold S, Larivière L, Maier KC, Seizl M, Tresch A, Cramer P. 2012. Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome Res* **22**:1350–1359. doi:10.1101/gr.130161.111
- Synowsky SA, van Wijk M, Raijmakers R, Heck AJR. 2009. Comparative multiplexed mass spectrometric analyses of endogenously expressed yeast nuclear and cytoplasmic exosomes. *J Mol Biol* **385**:1300–1313. doi:10.1016/j.jmb.2008.11.011
- Tan-Wong SM, Zaugg JB, Camblong J, Xu Z, Zhang DW, Mischo HE, Ansari AZ, Luscombe NM, Steinmetz LM, Proudfoot NJ. 2012. Gene loops enhance transcriptional directionality. *Science* **338**:671–675. doi:10.1126/science.1224350
- Tardu M, Lin Q, Koutmou KS. 2018. N4-acetylcytidine and 5-formylcytidine are present in *Saccharomyces cerevisiae* mRNAs. *bioRxiv*. doi:10.1101/327585
- Tharun S, Parker R. 2001. Targeting an mRNA for decapping: displacement of translation factors and association of the Lsm1p-7p complex on deadenylated yeast mRNAs. *Mol Cell* **8**:1075–1083.
- Theler D, Dominguez C, Blatter M, Boudet J, Allain H, Allain FH-T. 2014. Solution structure of the YTH domain in complex with N6-methyladenosine RNA: a reader of methylated RNA. *Nucleic Acids Res* **42**:13911–13919. doi:10.1093/nar/gku1116
- Thiaville PC, Legendre R, Rojas-Benítez D, Baudin-Baillieu A, Hatin I, Chalancon G, Glavic A, Namy O, de Crécy-Lagard V. 2016. Global translational impacts of the loss of the tRNA modification t(6)A in yeast. *Microb cell* **3**:29–45. doi:10.15698/mic2016.01.473
- Thiebaut M, Kisseleva-Romanova E, Rougemaille M, Boulay J, Libri D. 2006. Transcription Termination and Nuclear Degradation of Cryptic Unstable Transcripts: A Role for the Nrd1-Nab3 Pathway in Genome Surveillance. *Mol Cell* **23**:853–864. doi:10.1016/j.molcel.2006.07.029
- Thompson DM, Parker R. 2007. Cytoplasmic decay of intergenic transcripts in *Saccharomyces cerevisiae*. *Mol Cell Biol* **27**:92–101. doi:10.1128/MCB.01023-06
- Torrent M, Chalancon G, Groot NS De, Wuster A, Babu MM. 2018. Cells alter their tRNA abundance to selectively regulate protein synthesis during stress conditions. *Sci Signal* **6409**:1–10.
- Torres AG, Batlle E, Ribas de Pouplana L. 2014. Role of tRNA modifications in human diseases. *Trends Mol Med* **20**:306–314.

- doi:10.1016/j.molmed.2014.01.008
- Trewick SC, Henshaw TF, Hausinger RP, Lindahl T, Sedgwick B. 2002. Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature*. doi:10.1038/nature00908
- Tuck AC, Tollervey D. 2013. A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs. *Cell* **154**:996–1009. doi:10.1016/j.cell.2013.07.047
- Tucker M, Valencia-Sanchez MA, Staples RR, Chen J, Denis CL, Parker R. 2001. The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in *Saccharomyces cerevisiae*. *Cell* **104**:377–386. doi:10.1016/S0092-8674(01)00225-2
- Tudek A, Porrua O, Kabzinski T, Lidschreiber M, Kubicek K, Fortova A, Lacroute F, Vanacova S, Cramer P, Stefl R, Libri D. 2014. Molecular basis for coordinating transcription termination with noncoding RNA degradation. *Mol Cell* **55**:467–481. doi:10.1016/j.molcel.2014.05.031
- Turowski TW, Tollervey D. 2015. Cotranscriptional events in eukaryotic ribosome synthesis. *Wiley Interdiscip Rev RNA* **6**:129–139. doi:10.1002/wrna.1263
- Urbanavičius J, Qian Q, Durand JMB, Hagervall TG, Björk GR. 2001. Improvement of reading frame maintenance is a common function for several tRNA modifications. *EMBO J* **20**:4863–4873. doi:10.1093/emboj/20.17.4863
- Van Der Maaten LJP, Hinton GE. 2008. Visualizing high-dimensional data using t-sne. *J Mach Learn Res* **9**:2579–2605. doi:10.1007/s10479-011-0841-3
- van Dijk EL, Chen CL, d'Aubenton-Carafa Y, Gourvenec S, Kwapisz M, Roche V, Bertrand C, Silvain M, Legoix-Ne P, Loeillet S, Nicolas A, Thermes C, Morillon A. 2011. XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* **475**:114–117. doi:10.1038/nature10118
- van Hoof A, Frischmeyer PA, Dietz HC, Parker R. 2002. Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science* **295**:2262–2264. doi:10.1126/science.1067272
- Van Nostrand EL, Pratt GA, Shishkin AA, Gelboin-Burkhart C, Fang MY, Sundararaman B, Blue SM, Nguyen TB, Surka C, Elkins K, Stanton R, Rigo F, Guttman M, Yeo GW. 2016. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* **13**:508–514. doi:10.1038/nmeth.3810
- Vanacova S, Stefl R. 2007. The exosome and RNA quality control in the nucleus. *EMBO Rep* **8**:651–657. doi:10.1038/sj.embor.7401005
- Vasiljeva L, Buratowski S. 2006. Nrd1 interacts with the nuclear exosome for 3'

- processing of RNA polymerase II transcripts. *Mol Cell* **21**:239–248.
doi:10.1016/j.molcel.2005.11.028
- Waas WF, Druzina Z, Hanan M, Schimmel P. 2007. Role of a tRNA Base Modification and Its Precursors in Frameshifting in Eukaryotes. *J Biol Chem* **282**:26026–26034. doi:10.1074/jbc.M703391200
- Wang L, Lewis MS, Johnson AW. 2005. Domain interactions within the Ski2/3/8 complex and between the Ski complex and Ski7p. *RNA* **11**:1291–1302.
doi:10.1261/rna.2060405
- Wang X, Lu Z, Gomez A, Hon GC, Yue Y, Han D, Fu Y, Parisien M, Dai Q, Jia G, Ren B, Pan T, He C. 2014. N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**:117–120. doi:10.1038/nature12730
- Wang X, Zhao BS, Roundtree IA, Lu Z, Han D, Ma H, Weng X, Chen K, Shi H, He C. 2015. N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **161**:1388–1399. doi:10.1016/j.cell.2015.05.014
- Wang Y, Li Y, Toth JI, Petroski MD, Zhang Z, Zhao JC. 2014. N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat Cell Biol* **16**:191–198. doi:10.1038/ncb2902
- Warda AS, Kretschmer J, Hackert P, Lenz C, Urlaub H, Höbartner C, Sloan KE, Bohnsack MT. 2017. Human METTL16 is a N6-methyladenosine (m6A) methyltransferase that targets pre-mRNAs and various non-coding RNAs **18**:2004–2014. doi:10.15252/embr.201744940
- Webster MW, Chen Y, Stowell JAW, Graveley BR, Collier J, Passmore LA, Webster MW, Chen Y, Stowell JAW, Alhusaini N, Sweet T, Graveley BR. 2018. mRNA Deadenylation Is Coupled to Translation Rates by the Differential Activities of Ccr4-Not Nucleases. *Mol Cell* **70**:1089-1100.e8.
doi:10.1016/j.molcel.2018.05.033
- Weixlbaumer A, Murphy F V, Dziergowska A, Malkiewicz A, Vendeix FAP, Agris PF, Ramakrishnan V. 2007. Mechanism of expanding the decoding capacity of tRNAs by modification of uridines. *Nat Struct Mol Biol* **14**:498–502.
doi:10.1038/nsmb1242
- Wells SE, Hillner PE, Vale RD, Sachs AB. 1998. Circularization of mRNA by eukaryotic translation initiation factors. *Mol Cell* **2**:135–140. doi:10.1016/S1097-2765(00)80122-7
- Whitelaw E, Proudfoot N. 1986. Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene. *EMBO J* **5**:2915–2922.
- Wolf J, Passmore LA. 2014. Deadenylation by Pan2 / Pan3. *Biochem Soc Trans*

- 42:184–187.** doi:10.1042/BST20130211.mRNA
- Wyers F, Rougemaille M, Badis G, Rousselle JC, Dufour ME, Boulay J, Régnault B, Devaux F, Namane A, Séraphin B, Libri D, Jacquier A. 2005. Cryptic Pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **121**:725–737. doi:10.1016/j.cell.2005.04.030
- Xiao W, Adhikari S, Xiao W, Adhikari S, Dahal U, Chen Y, Hao Y, Sun B, Sun H. 2016. Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. *Mol Cell* **61**:507–519. doi:10.1016/j.molcel.2016.01.012
- Xiong X, Li X, Wang KUN, Yi C. 2018. Perspectives on topology of the human m¹A methylome at single nucleotide resolution. *RNA* **24**:1437–1442. doi:10.1261/rna.067694.118.5
- Xu C, Liu K, Ahmed H, Loppnau P, Schapira M, Min J. 2015. Structural Basis for the Discriminative Recognition of N6-Methyladenosine RNA by the Human YT521-B Homology Domain Family of Proteins. *J Biol Chem* **290**:24902–24913. doi:10.1074/jbc.M115.680389
- Xu L, Liu X, Sheng N, Oo KS, Liang J, Chionh YH, Xu J, Ye F. 2017. Three distinct 3-methylcytidine (m3C) methyltransferases modify tRNA and mRNA in mice and humans. *J Biol Chem* **292**:14695–14703. doi:10.1074/jbc.M117.798298
- Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Munster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM. 2009. Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**:1033–1037. doi:10.1038/nature07728
- Yacoubi B El, Hatin I, Deutsch C, Kahveci T, Rousset J, Iwata-reuyl D. 2011. A role for the universal Kae1/Qri7/YgjD (COG0533) family in tRNA modification **30**:882–893. doi:10.1038/emboj.2010.363
- Yoshihisa T, Yunoki-Esaki K, Ohshima C, Tanaka N, Endo T. 2003. Possibility of cytoplasmic pre-tRNA splicing: the yeast tRNA splicing endonuclease mainly localizes on the mitochondria. *Mol Biol Cell* **14**:3266–3279. doi:10.1091/mbc.e02-11-0757
- Yudkovsky N, Ranish JA, Hahn S. 2000. A transcription reinitiation intermediate that is stabilized by activator. *Nature* **408**:225–229. doi:10.1038/35041603
- Zaringhalam M, Papavasiliou FN. 2016. Pseudouridylation meets next-generation sequencing. *Methods* **107**:63–72. doi:10.1016/j.ymeth.2016.03.001
- Zhang Z, Theler D, Kaminska KH, Hiller M, de la Grange P, Pudimat R, Rafalska I, Heinrich B, Bujnicki JM, Allain FH-T, Stamm S. 2010. The YTH domain is a novel RNA binding domain. *J Biol Chem* **285**:14701–14710. doi:10.1074/jbc.M110.104711

- Zhao Y, Majid MC, Soll JM, Brickner JR, Dango S, Mosammaparast N. 2015. Noncanonical regulation of alkylation damage resistance by the OTUD4 deubiquitinase. *EMBO J* **34**:1687–1703. doi:10.15252/embj.201490497
- Zheng G, Dahl JA, Niu Y, Fedorcsak P, Huang C-M, Li CJ, Vagbo CB, Shi Y, Wang W-L, Song S-H, Lu Z, Bosmans RPG, Dai Q, Hao Y-J, Yang X, Zhao W-M, Tong W-M, Wang X-J, Bogdan F, Furu K, Fu Y, Jia G, Zhao X, Liu J, Krokan HE, Klungland A, Yang Y-G, He C. 2013. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol Cell* **49**:18–29. doi:10.1016/j.molcel.2012.10.015
- Zhu T, Roundtree IA, Wang P, Wang X, Wang L, Sun C, Tian Y, Li J, He C, Xu Y. 2014. Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N6-methyladenosine. *Cell Res*. doi:10.1038/cr.2014.152
- Zinder JC, Lima CD. 2017. Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes Dev* **31**:88–100. doi:10.1101/gad.294769.116

7 Appendix

7.1 Supplementary Information

7.1.1 Figures

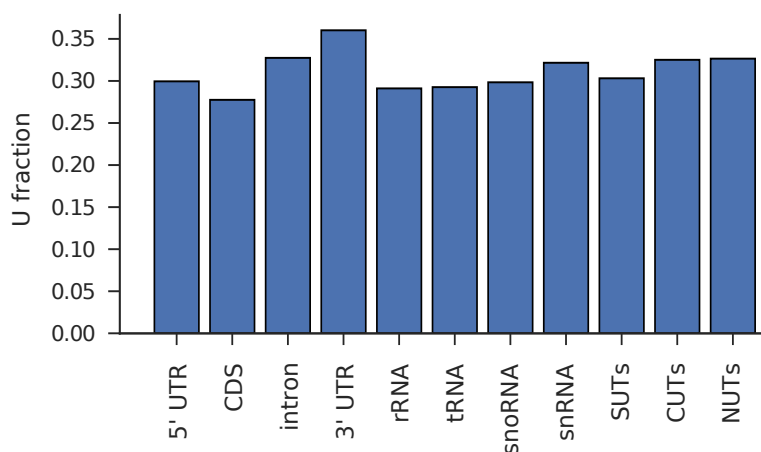


Figure 28: Different transcript classes have comparable U-content.

Fraction of U over all bases in transcript classes studied in Figure 8 (untranslated region (UTR); intron; coding sequence (CDS), ribosomal RNA (rRNA), transfer RNA (tRNA), small nucleolar RNA (snoRNA), small nuclear RNA (snRNA), stable unannotated transcripts (SUTs), cryptic unstable transcripts (CUTs), Nrd1-terminated transcripts (NUTs)).

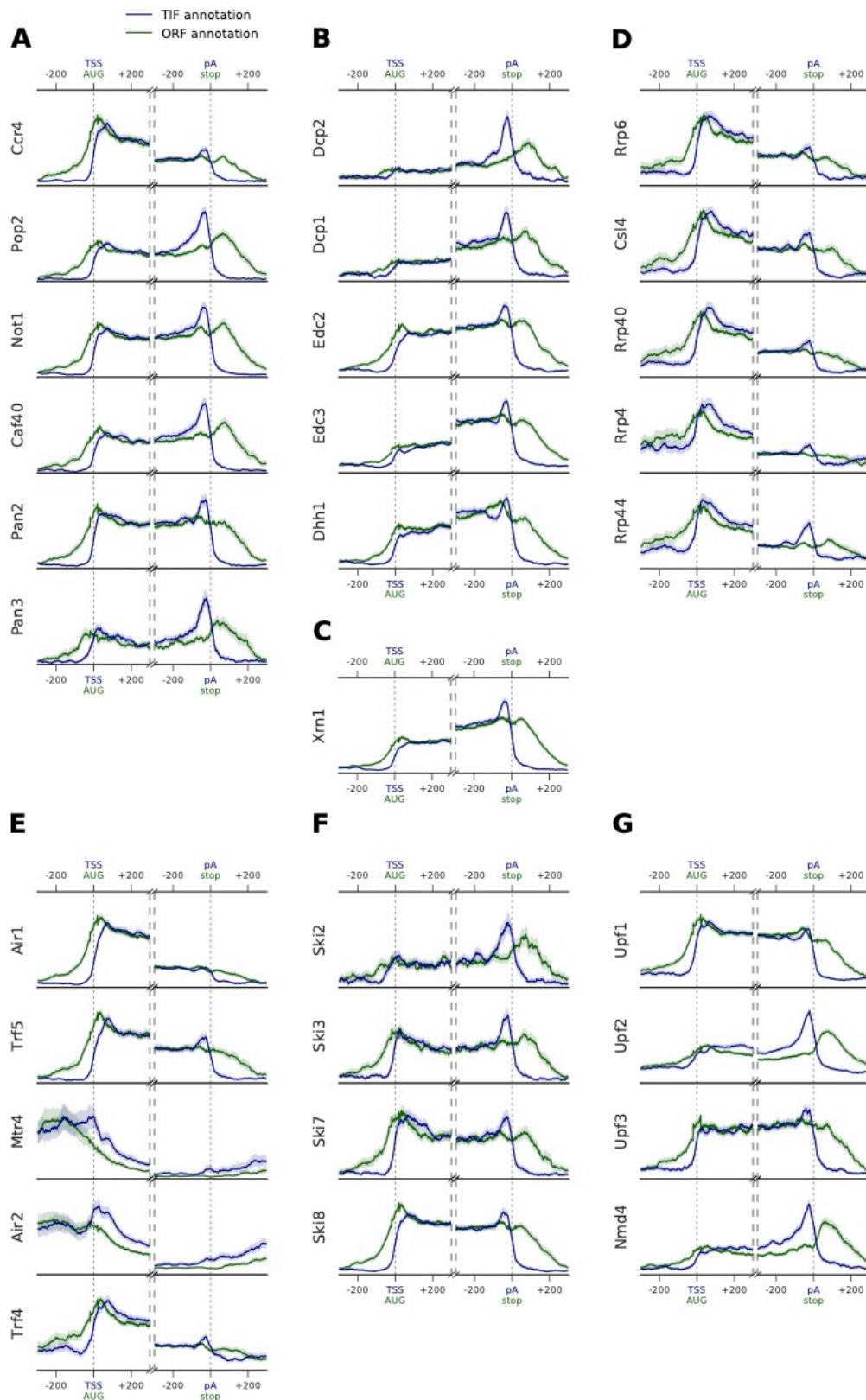


Figure 29: Metagene profiles of yeast RNA degradation factors centered on translation start and stop sites in comparison to TIF-annotated TSS and pA sites. Transcript-averaged PAR-CLIP occupancy profiles is shown for RNA degradation factors involved in A) deadenylation, B) decapping, C) 5'→3' exonuclease Xrn1, D)

exosome, E) TRAMP, F) Ski, and G) NMD. Transcripts are aligned either at transcript start site (TSS) and poly-adenylation (pA) site (marked with blue) or at their start and stop codons (marked with green). TIF-seq based annotation is shown in blue ($n=3,193$ for TSS and pA site profiles) (Pelechano et al., 2013). Open reading frames (ORF) annotated in the SGD (version 64.2.1) are shown in green ($n=4,012$ for TSS, and $n=3,965$ for pA site selected transcripts). To avoid contaminating signals from neighboring genes, we filtered out regions that had annotations upstream and downstream of the centered gene (up to 700 nt) (Methods). Shaded areas (in blue TIF-seq annotation, or in green for ORF annotation) depict 95% confidence intervals derived from bootstrapping genes. Comparison between these two profiles highlights preferences for end binding degradation factors in binding to untranslated regions at the two sides of the transcript.

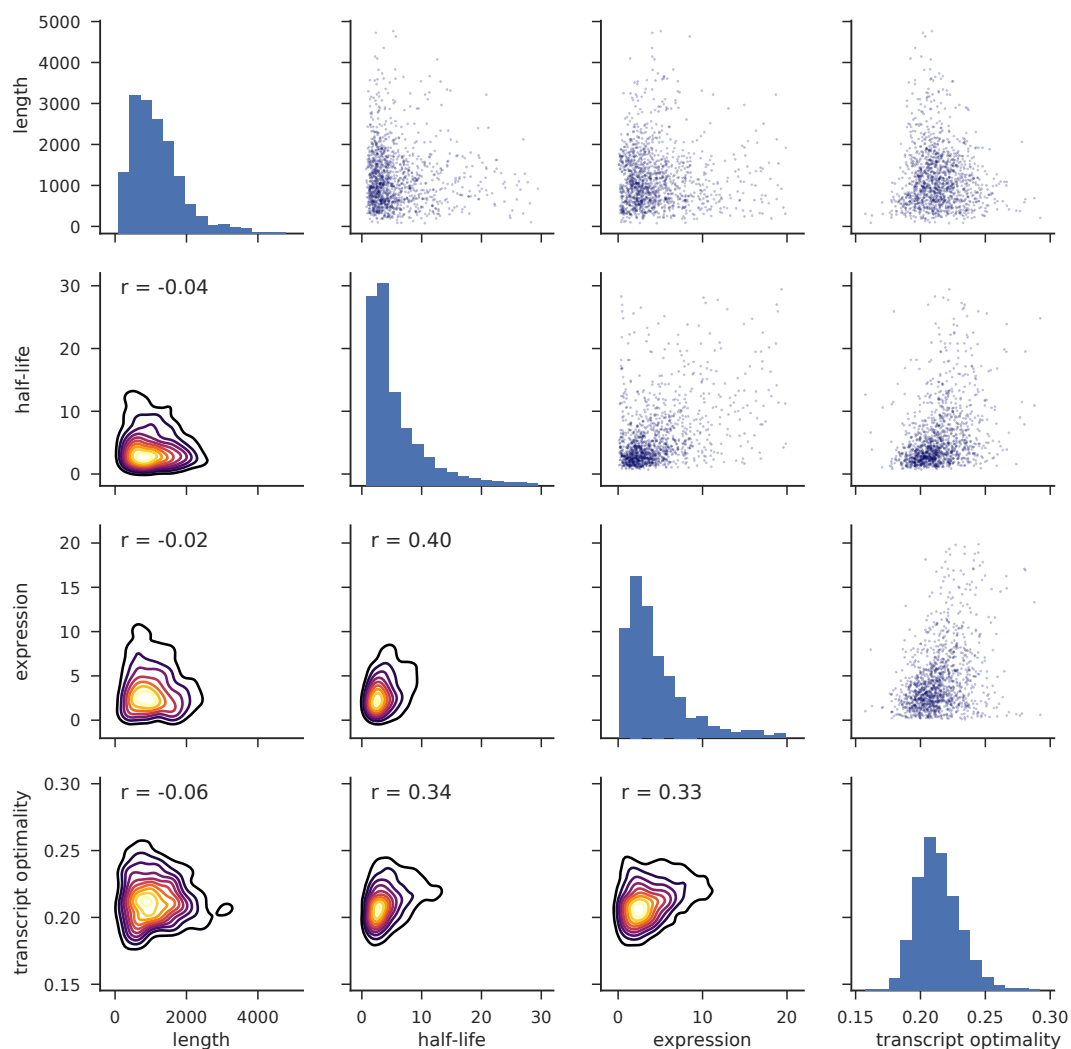


Figure 30: Distributions of transcript length, half-life, expression level and transcript optimality for yeast mRNAs.

Histograms on the diagonal show distributions of length, half-life (Methods), expression level (Baejen et al., 2017) and transcript optimality (Pechmann and

Frydman, 2013). Pairwise comparisons of features are shown as scatter plots (top right) and kernel density estimates (KDEs) of bivariate densities are shown in the bottom with Pearson correlation values (r) (Methods).

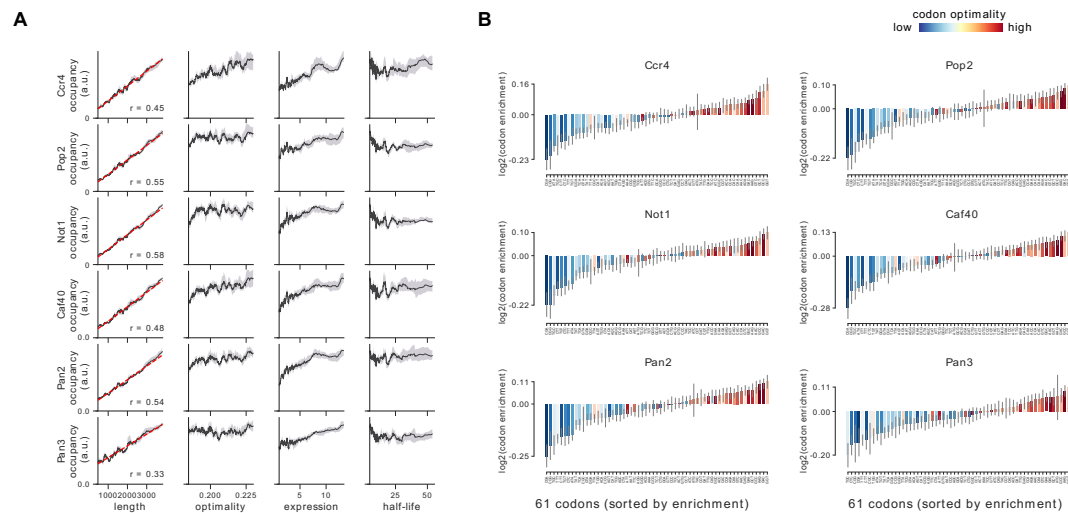


Figure 31: Occupancies of deadenylation factors (*Ccr4*, *Pop2*, *Not1*, *Caf40*, *Pan2*, and *Pan3*) compared to transcript length, optimality, expression level, and half-life.

A) To understand binding specificity of deadenylation factors, the total occupancy of each factor on a transcript is plotted against various transcript features (Grey shading: 95% confidence intervals generated by bootstrapping transcripts). B) Same analysis as in Figure 12B: Codon enrichment shows deviations in codon frequencies of transcripts bound by a degradation factor compared to each codon's frequency on all coding sequences. Each bar is colored according to its codon-optimality with highly optimal codons in dark red and highly non-optimal codons in dark blue. (Grey lines: 90% confidence intervals generated by bootstrapping coding sequences).

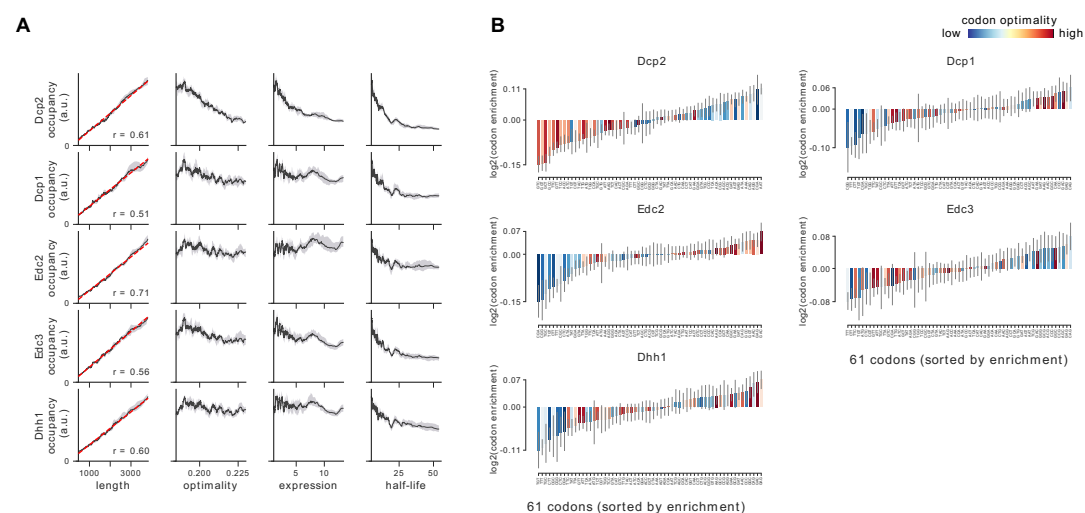


Figure 32: Occupancies of decapping factors (*Dcp2*, *Dcp1*, *Edc2*, *Edc3*, and *Dhh1*) compared to transcript length, optimality, expression level, and half-life.

A) To understand binding specificity of decapping factors, the total occupancy of each factor on a transcript is plotted against various transcript features (Grey shading: 95% confidence intervals generated by bootstrapping transcripts). B) Same analysis as in Figure 12B: Codon enrichment shows deviations in codon frequencies of transcripts bound by a degradation factor compared to each codon's frequency on all coding sequences. Each bar is colored according to its codon-optimality with highly optimal codons in dark red and highly non-optimal codons in dark blue. (Grey lines: 90% confidence intervals generated by bootstrapping coding sequences).

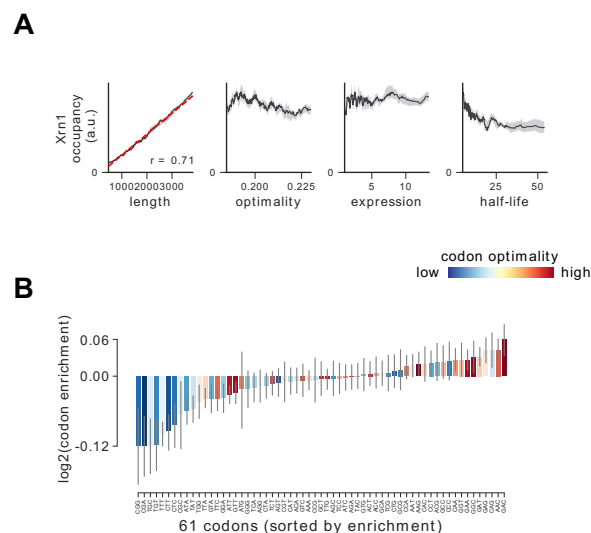


Figure 33: Occupancy of Xrn1 compared to transcript length, optimality, expression level, and half-life.

A) To understand binding specificity of Xrn1 on various mRNAs, the total occupancy of Xrn1 on a transcript is plotted against various transcript features (Grey shading: 95% confidence intervals generated by bootstrapping transcripts). B) Same analysis as in Figure 12B: Codon enrichment shows deviations in codon frequencies of transcripts bound by a degradation factor compared to each codon's frequency on all coding sequences. Each bar is colored according to its codon-optimality with highly optimal codons in dark red and highly non-optimal codons in dark blue. (Grey lines: 90% confidence intervals generated by bootstrapping coding sequences).

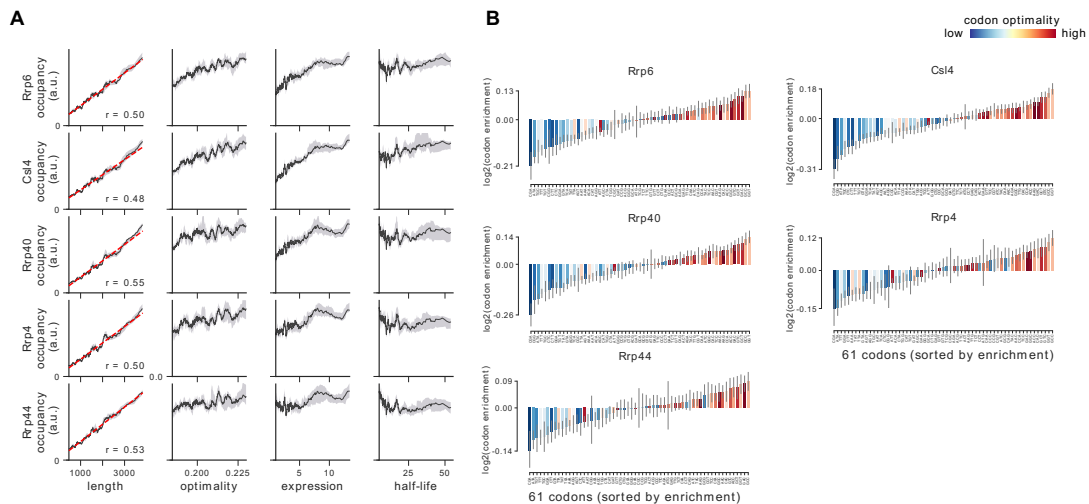


Figure 34: Occupancies of exosome components (*Rrp6*, *Csl4*, *Rrp40*, *Rrp4*, and *Rrp44*) compared to transcript length, optimality, expression level, and half-life.

A) To understand binding specificity of exosome components, the total occupancy of each factor on a transcript is plotted against various transcript features (Grey shading: 95% confidence intervals generated by bootstrapping transcripts). B) Same analysis as in Figure 12B: Codon enrichment shows deviations in codon frequencies of transcripts bound by a degradation factor compared to each codon's frequency on all coding sequences. Each bar is colored according to its codon-optimality with highly optimal codons in dark red and highly non-optimal codons in dark blue. (Grey lines: 90% confidence intervals generated by bootstrapping coding sequences).

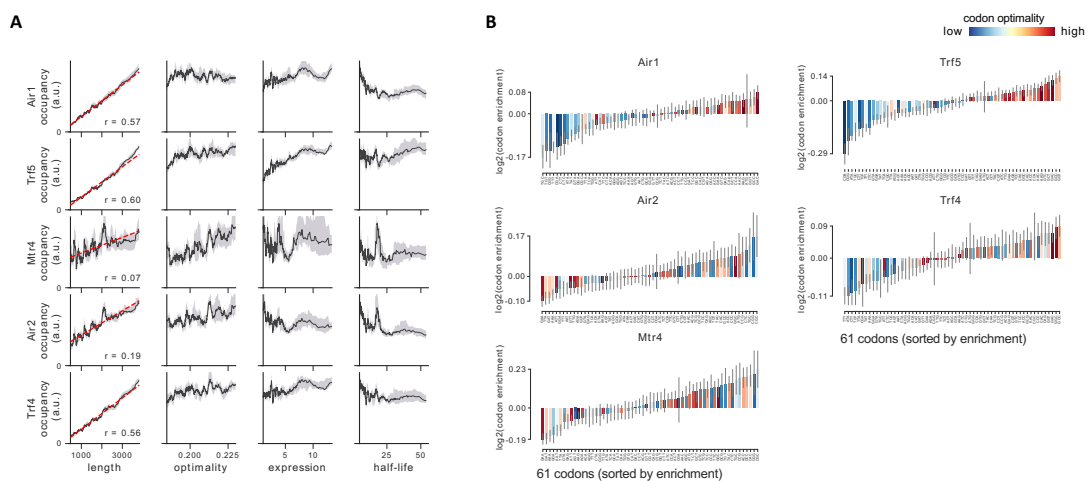


Figure 35: Occupancies for components of the TRAMP complex (*Air1*, *Trf5*, *Mtr4*, *Air2*, and *Trf4*) compared to transcript length, optimality, expression level, and half-life.

A) To understand binding specificity of TRAMP components, the total occupancy of each factor on a transcript is plotted against various transcript features (Grey shading: 95% confidence intervals generated by bootstrapping transcripts). B) Same analysis as in Figure 5B: Codon enrichment shows deviations in codon frequencies of

transcripts bound by a degradation factor compared to each codon's frequency on all coding sequences. Each bar is colored according to its codon-optimality with highly optimal codons in dark red and highly non-optimal codons in dark blue. (Grey lines: 90% confidence intervals generated by bootstrapping coding sequences).

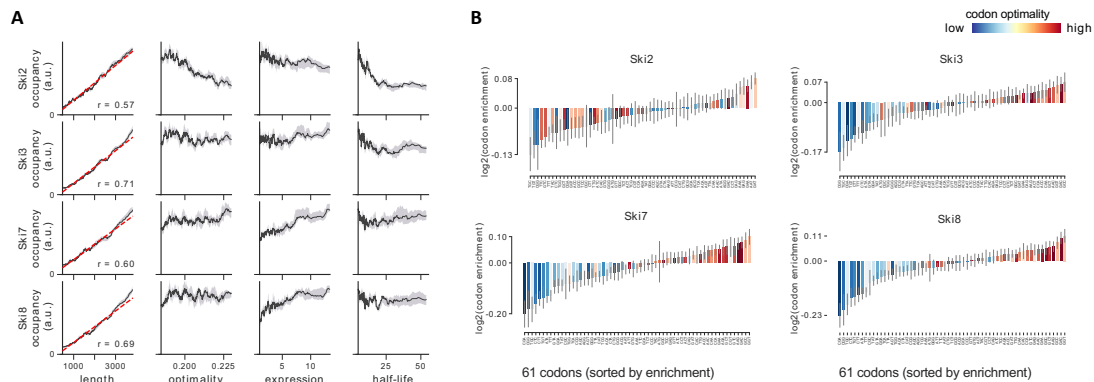


Figure 36: Occupancies for components of the Ski complex (Ski2, Ski3, Ski7, and Ski8) compared to transcript length, optimality, expression level, and half-life.

A) To understand binding specificity of factors in the Ski complex, the total occupancy of each factor on a transcript is plotted against various transcript features (Grey shading: 95% confidence intervals generated by bootstrapping transcripts). B) Same analysis as in Figure 12B: Codon enrichment shows deviations in codon frequencies of transcripts bound by a degradation factor compared to each codon's frequency on all coding sequences. Each bar is colored according to its codon-optimality with highly optimal codons in dark red and highly non-optimal codons in dark blue. (Grey lines: 90% confidence intervals generated by bootstrapping coding sequences).

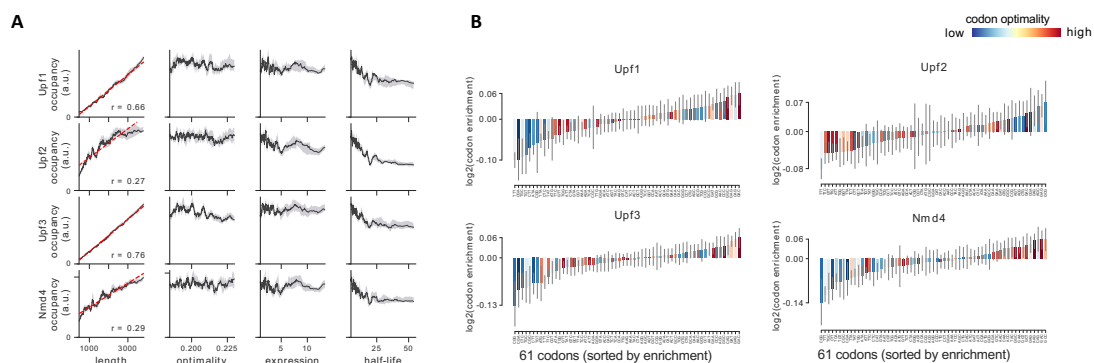


Figure 37: Occupancies for components of the NMD pathway (Upf1, Upf2, Upf3, and Nmd4) compared to transcript length, optimality, expression level, and half-life.

A) To understand binding specificity of factors in the NMD pathway, the total occupancy of each factor on a transcript is plotted against various transcript features (Grey shading: 95% confidence intervals generated by bootstrapping transcripts). B) Same analysis as in Figure 12B: Codon enrichment shows deviations in codon

frequencies of transcripts bound by a degradation factor compared to each codon's frequency on all coding sequences. Each bar is colored according to its codon-optimality with highly optimal codons in dark red and highly non-optimal codons in dark blue. (Grey lines: 90% confidence intervals generated by bootstrapping coding sequences).

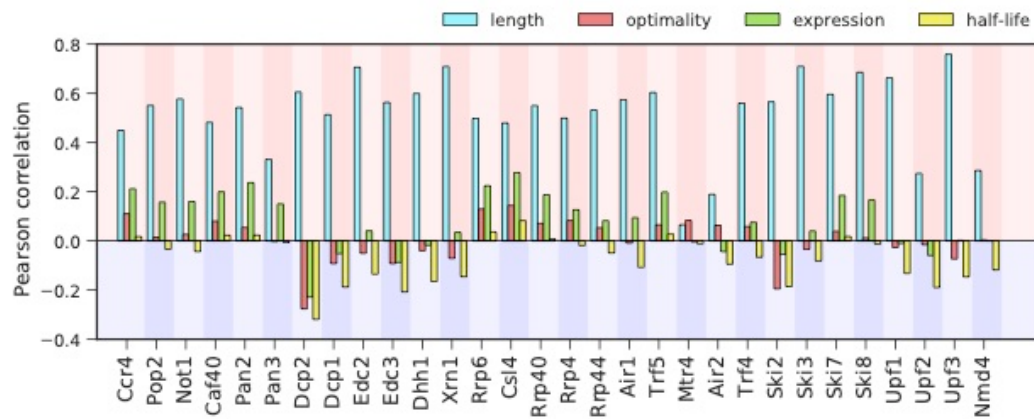


Figure 38: Correlation between binding to degradation factors and transcript length, codon-optimality, expression, and half-life.

Pearson correlation values between the binding strength of degradation factors (total occupancy over each transcript) and transcript length, transcript optimality (Pechmann and Frydman, 2013), expression level (Baejen et al., 2017), and half-life derived by multivariate linear regression analysis (Methods).

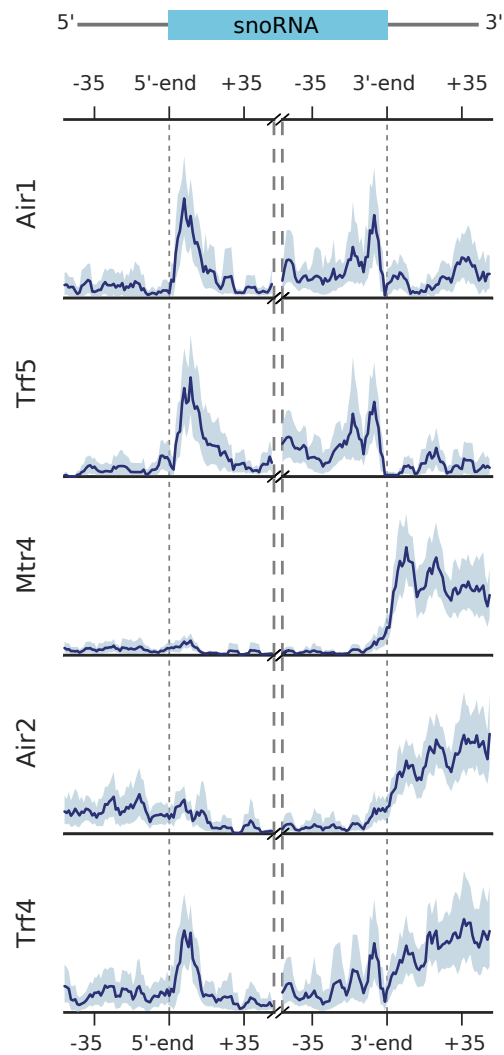


Figure 39: Metagene profiles for subunits of the TRAMP complex on snoRNA genes. Transcript averaged PAR-CLIP occupancy profiles are shown for Air1, Trf5, Mtr4, Air2, and Trf4. snoRNA genes are aligned either at their 5' end or at their 3' end (n=77). Occupancy profiles are shown over the range of ± 35 nt.

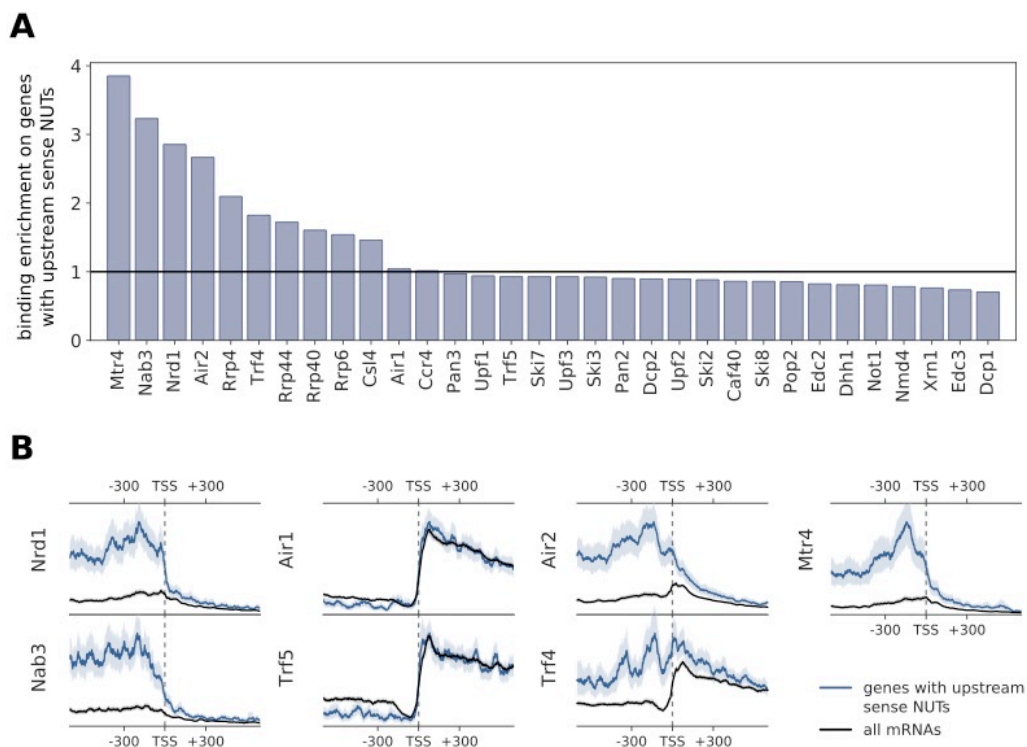


Figure 40: Comparison of binding profiles on genes containing annotated upstream sense NUTs with all mRNAs.

A) Binding enrichment of degradation factors around the TSS of genes with an upstream sense NUT. Enrichment is defined as the ratio of the average occupancy in the interval $[\pm 300 \text{ nt}]$ of the TSS on these genes that contain an upstream NUT ($n=459$) (Schulz et al., 2013) divided by the average occupancy on all genes. B) Transcript-averaged PAR-CLIP occupancy profiles for all mRNAs (black) is compared to patterns derived from genes with upstream sense NUTs (blue). Transcripts were aligned at their TSS and averaged over the interval of $[\pm 600 \text{ nt}]$. We compared Nrd1 and Nab3 profiles, known to process NUTs, with subunits of the TRAMP complex. 95% confidence intervals obtained from bootstrapping genes are shown with grey and blue shades.

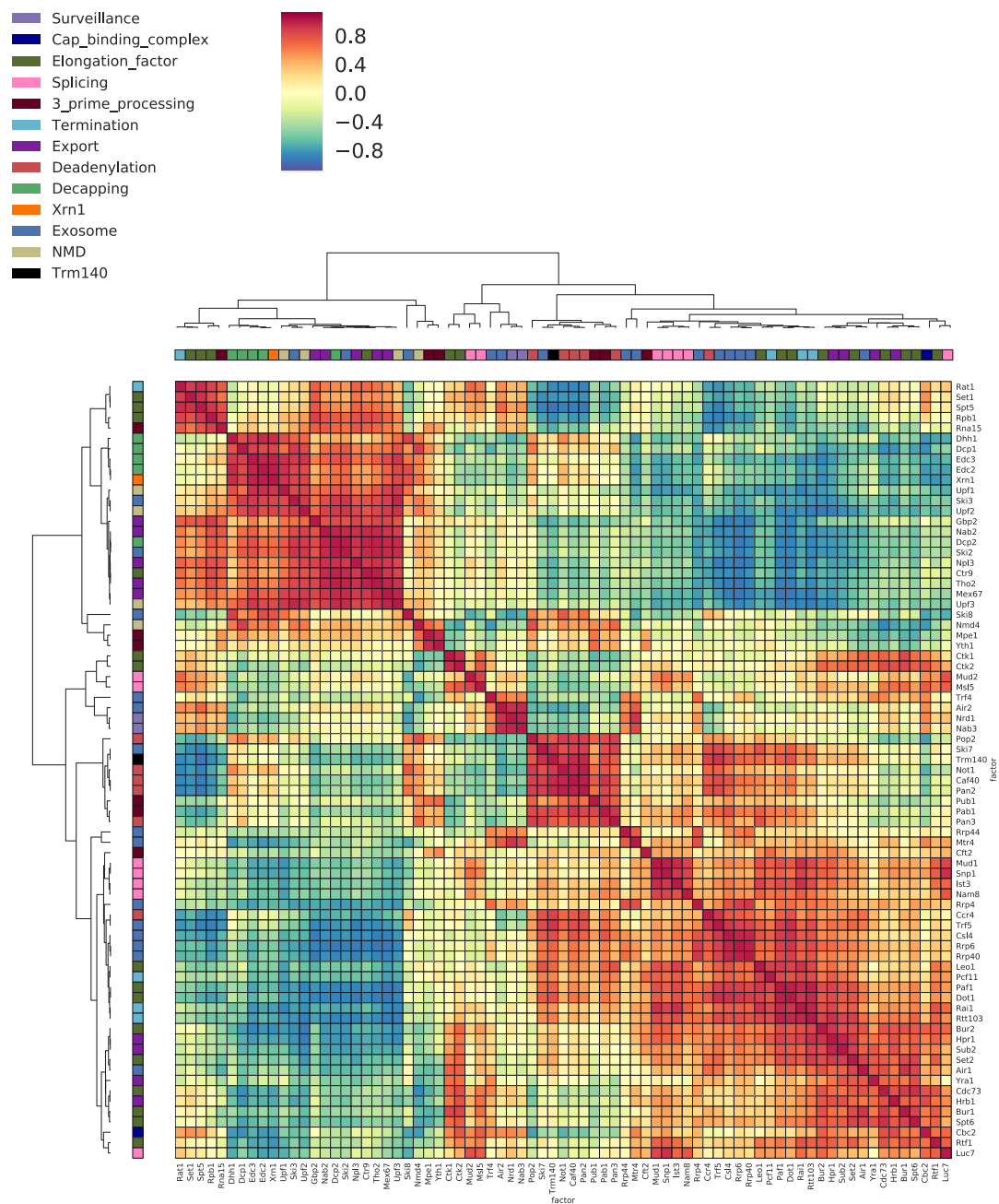


Figure 41: Co-localization coefficients for 75 RNA binding factors.

Pairwise correlation between normalized co-localization profiles of factors in a window of 40 nt centered at PAR-CLIP cross-link sites is shown here. Analysis for 75 RNA binding factors, including Trm140, 30 factors from the RNA degradation study, and 44 factors from previous studies (Baejen et al., 2017, 2014; Battaglia et al., 2017; Schulz et al., 2013). High co-localization represents binding to the same position on transcripts (marked with dark red). Factors are clustered and color coded (left and upper border) according to their general function. Trm140 is colored in black and can be found in the central cluster together with deadenylation factors.

7.1.2 Tables

Table 13: Overview of RNA processing factors and their respective published PAR-CLIP experiments.

Pathway	Factor	Details	High confidence cross-link sites	Source	Accession number
Surveillance	Nrd1	Nrd1/Nab3 complex	187164	Schulz et al., 2013	ArrayExpress: E-MTAB-1766
	Nab3		91762	Schulz et al., 2013	ArrayExpress: E-MTAB-1766
Cap binding complex	Cbc2		95530	Baejen et al., 2014	GEO: GSE59676
Elongation	Bur1	BUR kinase complex	232237		GEO: GSE81822
	Bur2		169544	Battaglia et al., 2017	GEO: GSE81822
	Ctk1	CTDK-I coomplex	257668	Battaglia et al., 2017	GEO: GSE81822
	Ctk2		209835	Battaglia et al., 2017	GEO: GSE81822
	Cdc73		234224	Battaglia et al., 2017	GEO: GSE81822
	Ctr9		5361	Battaglia et al., 2017	GEO: GSE81822
	Leo1	PAF1 complex	87202	Battaglia et al., 2017	GEO: GSE81822
	Paf1		72871	Battaglia et al., 2017	GEO: GSE81822
	Rtf1		52570	Battaglia et al., 2017	GEO: GSE81822
	Set1	Histone methyl transferase	321575	Battaglia et al., 2017	GEO: GSE81822
	Set2		90276	Battaglia et al., 2017	GEO: GSE81822
	Dot1		88966	Battaglia et al., 2017	GEO: GSE81822
	Spt5	DSIF	193823	Baejen et al., 2017	GEO: GSE79222
	Spt6	Nucleosome remodeling	204781	Battaglia et al., 2017	GEO: GSE81822
	Rpb1	RNA polymerase	153593	Baejen et al., 2017	GEO: GSE79222
	Splicing	Ist3	U2 snRNP	3899	Baejen et al., 2014
Nam8			7097	Baejen et al., 2014	GEO: GSE59676
Mud1		U1 snRNP	8512	Baejen et al., 2014	GEO: GSE59676
Snpl			6773	Baejen et al., 2014	GEO: GSE59676
Luc7			33471	Baejen et al., 2014	GEO: GSE59676
Mud2		BBP/ U2AF65	76190	Baejen et al., 2014	GEO: GSE59676
Msl5			81742	Baejen et al., 2014	GEO: GSE59676
3' processing		Pab1	Poly(A)	33422	Baejen et al., 2014
	Pub1	Poly(U)	72015	Baejen et al., 2014	GEO: GSE59676
	Rna15	CFIA	92792	Baejen et al., 2014	GEO: GSE59676
	Mpe1		3743	Baejen et al., 2014	GEO: GSE59676
	Cft2	CPF	32758	Baejen et al., 2014	GEO: GSE59676
	Yth1		6506	Baejen et al., 2014	GEO: GSE59676
Termination	Rat1	Exoribonuclease	122259	Baejen et al., 2017	GEO: GSE79222
	Rai1		23092	Baejen et al., 2017	GEO: GSE79222
	Rtt103		51954	Baejen et al., 2017	GEO: GSE79222
	Pcf11		1432	Baejen et al., 2017	GEO: GSE79222
Export	Hpr1	THO	24616	Baejen et al., 2014	GEO: GSE59676
	Tho2		3282	Baejen et al., 2014	GEO: GSE59676
	Gbp2		151941	Baejen et al., 2014	GEO: GSE59676
	Hrb1	TREX	132503	Baejen et al., 2014	GEO: GSE59676
	Mex67		10917	Baejen et al., 2014	GEO: GSE59676
	Sub2		39981	Baejen et al., 2014	GEO: GSE59676
	Yra1		46438	Baejen et al., 2014	GEO: GSE59676
	Nab2	Adaptors	99934	Baejen et al., 2014	GEO: GSE59676
Npl3	131428		Baejen et al., 2014	GEO: GSE59676	

Table 14: High confidence m3C modification sites.

The m3C modification sites including position (chr and position), number of C→T reads, unchanged reads (reference), the ratio (of transitions and reference reads), the assigned feature and gene name if annotated are shown. The m3C modification sites were called using the Trm140 PAR-CLIP sequencing data containing C→T transitions aligned to the reference genome (sacCer3, version 64.2.1). A potential m3C site was only called with a transition rate of 4% and read coverage of at least 40.

Chr	Position	C→T	Reference	Ratio	Feature	Gene
chrV	86519	104	5	0,95	mRNA	MTC7
chrIII	162694	94	16	0,85	mRNA	YCR024C-B
chrIV	1402305	71	13	0,85	mRNA	RPL27B
chrIV	1153577	82	36	0,69	mRNA	YDR341C
chrII	323668	27	18	0,60	mRNA	QDR3
chrXIII	139652	23	19	0,55	mRNA	ERV41
chrXI	46776	2538	2395	0,51	tRNA	T(CGU)
chrXVI	827581	66	76	0,46	mRNA	YPR148C
chrV	41712	83	96	0,46	mRNA	PRB1
chrXIII	159468	82	149	0,35	mRNA	SML1
chrVII	978952	19	65	0,23	mRNA	LSC2
chrVIII	303572	27	94	0,22	mRNA	TRA1
chrII	685511	9	37	0,20	mRNA	ARC40
chrXV	1017134	17	73	0,19	mRNA	PRT1
chrXV	796725	8	44	0,15	mRNA	ENV9
chrXV	829817	8	46	0,15	mRNA	VPH1
chrXII	1004572	9	54	0,14	mRNA	CNA1
chrX	538585	223	1588	0,12	tRNA	R(CCU)
chrXIV	349743	12	109	0,10	mRNA	PGA2
chrV	81963	5	52	0,09	mRNA	RAD23
chrVI	75912	6	74	0,08	mRNA	HAC1
chrXIV	654597	3	37	0,08	mRNA	ACC1
chrXIV	661501	22	273	0,07	mRNA	ACC1
chrIX	369738	4	52	0,07	mRNA	PAN1
chrVIII	292455	3	45	0,06	mRNA	HXT1
chrVIII	467020	17	266	0,06	tRNA	T(UGU)
chrV	138519	3	54	0,05	CUT/XUT	
chrXV	282163	4	76	0,05	tRNA	G(GGC)

chrVII	519348	3	58	0,05	mRNA	MSB2
chrXI	464423	6	120	0,05	mRNA	PRY2
chrXIII	887067	5	100	0,05	mRNA	GAS1
chrVII	111292	5	106	0,05	mRNA	YGL204C
chrIII	227973	45	1015	0,04	tRNA	S(CGA)
chrVII	866510	3	72	0,04	mRNA	TYS1
chrX	519511	3	72	0,04	mRNA	SSC1
chrXII	818154	13	351	0,04	XUT	

7.1.3 Spike-in sequences

7.1.3.1 Spike-in 2

GGGUGCUUUAACAAGAGGAAAUUGUGUUUUUGCCAAUUUAAGACCUAA
 UUUAUAGUUAAACCAUUAACCUUAGUUGUCCAAGGCAUAAUUAAGA
 GAGUGAGAUACAGGAUGAGCUAUUUCAGGGAGUUUUCAGUAUGCAG
 UUGCCAAGGCAGUUGCUGAUUUAGAUUUAGAUAGAAGAUUUAAAGGUU
 GUUGUCUCUGUUAUGUCCAGAGGUUCCAUAACCAUUUAAAUA
 AGAAAACUCUCCAUAUCUUCUAUGCCUCAGCAAAGUUAGCUAUAAC
 AGAGCUUUAAAUGAAUAUCCUUCAAAAGAGAAGGUAAAAGAAAGAAA
 AUAGAGCUUUGCAUCCAUAUAGUUGGAUUUAGGGAUGUUAGAUUGGAG
 UAUCCUCCAUAUCUACAAAUUGCUUUGGAUGUCCCAACUAUGGAGAAU
 UUGGAAUUUUUGUUACAAACAAUCCAAAUAGCGACCACAUCUUA
 GAGGCUGGAACACCACUAAUUA AAAAGUUUGGUUUAGAGGUUAUUGAA
 AUAUGAGAGAAUAUUUUGAUUGGCUUUUUAUGUUGCUGAUUUAAAACC
 UUAGACACUGGAAGGGUUGAGGUUAGAUUGGCAUUUGAAGCAACAGC
 UAAUUGCAGUGGCAUAAGUGGAGUAGCACCAAAUACAACAAUAAUUA
 AGCUAUCCACGAAUGUCAAAAUGUGGUUUAAUCAGCUAUUUGGAUUA
 GAUGAACGUCUCUGAACCUCAAAAUAUAUGAUUCAUUA AAAUUAAG
 CCAGAUGUUGUUUUCUUGCAUAGAGGGAUUGAUUGAGGAGACAUUUGG
 AAUUA AAAAGGAUUGGAAUUAAGGAAAACUGCUUUAUUGCAAUUGC
 UGGAGGAGUUGGUGUGGAGAAUGUUGAAGAGCUUUUAAAAGAAUAUC
 AAUAUUAUUCGUUGGUAGAGCAAUUACAAAUCAAAAGACCCAGGAA
 GAGUAAUUAGGAUUUUUAUAAACAAGAUGG

7.1.3.2 Spike-in 4

GGGUUUCGACGUUUUGAAGGAGGGUUUUUAAAGUAAUGAUCGAGAUUGA
 AAAACCAAAAUCGAAACGGUUGAAAUCAGCGACGAUGCCGAAUUUGG
 UAAGUUUGUCGUAGAGCCACUUGAGCGUGGAUAUGGUACAACUCUGG
 GUAACUCCUACGUCGUUCCUCUUAUCCUCACUCCUGGUGCCGCU
 GUAACAUCAAUCCAGAUAGAUGGUGUACUGCACGAAUUCUCGACAAU
 GAAGGCGUUGUGGAAGAUGUUACAACGAUUAUCUACACAUUAAAAG
 CUUGCAUUGAAAUCUACUCUGAUGAAGAGAAGACGCUAGAAAUUGAU
 GUACAGGGUGAAGGAACUGUAACGGCAGCUGAUUUACACACGAUAGU
 GAUGUAGAGAUUUAAAUCCUGAUCUUCAUUUCGCGACUCUUGGUGA
 GAAUGCGAGUUUCCGAGUUCGCCUACUGCUCAAAGAGGACGUGGGU
 AUACGCCUGCUGACGCAAACAAGAGAGGGCGAUCAGCCAAUCGGCGUG
 AUUCCGAUCGAUUCUUAUCUUAACGCCAGUUUCCCGUGUAUCUUAUCAG
 GUAGAGAACACUCGUGUAGGCCAAGUUGCAAACUAUGAUAAACUUAACA
 CUUGAUGUUUGGACUGAUGGAAGCACUGGACCGAAAGAAGCAAUUGC
 GCUUGGUUCAAGAUUUUAACUGAACACCUUAAUAUUAUUCGCUGGUUU
 AACUGACGAAGCUCAACAUGCUGAAAUCAUGGUUGAAGAAGAAGA
 UCAAAAAGAGAAAGUUCUUGAAAUGACAUAUGAAGAAUUGGAUCUUC
 UGUUCGUUCUUAACUGCUUAAAGCGUGCGGGUAUUAACACGGUUC
 AAGAGCUUGCGAACAAAGACGGAAGAAGAUUAUGAUGAAAGUUCGAAAUC
 UAGGACGCAAUCACUUGAAGAAGUGAAAGCGAGACUAGAAGAACUUG

GACUCGGACUUCGCAAAGACGAUUGACUAGUUUCCCUUGUGAACUAG
GAUUU UCCCGGGUAC

7.1.3.3 Spike-in 5

GGGACUGUCCUUUCAUCCAUAAGCGGAGAAAGAGGGAAUGACAUUGU
UCUUACACGGCACAAGCAGACAAAUCAACAUGGUCAUUUAGAAAUCG
GAGGUGUGGAUGCUCUCUAUUUAGCGGAGAAUAUGGUACACCUCUU
UACGUUAUUGAUGUGGCUUUAUACGUGAGCGUGCUAAAAGCUUUAAG
CAGGCGUUUAUUUCUGCAGGGCUGAAAGCACAGGUGGCAUAUGCGAG
CAAAGCAUUCUCAUCAGUCGCAAUGAUUCAGCUCGCUGAGGAAGAGG
GACUUUCUUUAGAUGUCGUAUCCGGAGGAGAGCUAUUAACGGCUGUU
GCAGCAGGCUUUCGGCAGAACGCAUCCACUUUCAUGGAAACAAUAAG
AGCAGGGAAGAACUGCGGAUGGCGCUUGAGCACCGCAUCGGCUGCAU
UGUGGUGGAUAAUUUCUAUGAAAUCGCGCUUCUUGAAGACCUAUGUAA
AGAAACGGGUCACUCCAUCGAUGUUCUUCUUCGGAUCACGCCCGGAG
UAGAAGCGCAUACGCAUGACUACAUAACGGGCCAGGAAGAUUCA
AGUUUGGUUUCGAUCUUCUAUAACGGACAAACUGAACGGGCCAUUGAAC
AAGUAUUAACAUCGGAACACAUUCAGCUGCUGGGUGUCCAUUGCCAUA
UCGGCUCGCAAUCUUUGAUACGGCCGGUUUUGUGUUAGCAGCGGAA
AAAUCUUCAAAAACUAGACGAAUGGAGAGAUUCAUAUUCAUUUGUA
UCCAAGGUGCUGAAUCUUGGAGGAGGUUUCGGCAUUCGUUAUACGGA
AGAUGAUGAACCGCUUCAUGCCACUGAAUACGUUGAAAAAUUAUCGA
AGCUGUGAAAGAAAUGCUUCCCGUUACGGUUUUGACAUUCCGGAAAU
UUGGAUCGAACCGGGCCGUUCUCUCGUGGGAGACGCAGGCACAACUC
UUUAUACGGUUGGCUCUCAAAAAGAAGUGGAUAAGCUGUACAAUCGUU
UCAUC AUUCGGCGUGCG

7.1.3.4 Spike-in 8

GGGGAUGUCCUUGGACGGGGUGGCGCAGUAUUACUGCAAGAGAGCG
GACAGAUUAGUGUGUUGGAGCCGACACAUAAGGUUCGUCCGGGGA
CCGAUCUGCAGCCUACGGGACAUUUAUCCGUAAAAGCAUGGCGCUGU
UUCGUACUUAUCGGAGGCCAGGUUUCGUCGCGGCGAGUCUCCCGAC
GACGGAGAUGGGCGUUAUCUUGGGCCGUCUCGUACUCUGUUACUU
GGCACAGAUGCGAGCCUCGUAAUGUGCAUCAGCUAAGGGCGAUUU
AUAUUGCGACGUUUGUACGGAUUCGUUACUAACGUGUUGGACGCUAG
UGGAAUAUGUGUCGUUUGGUUAGCCUACCCAUGGCUUUCGCGGCGACA
CAUGCUUAGACUCUUUCAAAACUUCGGUGAAGUUCACUCAAGCCGCGG
AGCGCCGUCGUAAUUCACUAGGGAUGGCGGUACCCGUGCCCGUCCGA
UUCGUAGCAACCUGCAUCACGAUUUUGUCUUCGGGCGACUUAUCAGA
UACGGUAAUGUAAAUAACCGGCAUUUGGGCACUUCUUGCGUUUAAGC
GGGAAAGAUCGCGAGGGCCCGCUAUUUGCGAUACUCCCAUGUCGGU
GCCGUCGCCUCUAUGUACUCGGAGACGUUAAUGCAGAGGCUAAGGAC
AAUUUACCAUGACUCGGUAAUCCGUUCGUCAAGCAGGUAGCUCGAGU
CUCCACCGGACACGUAGUGGGUUUGUAACGAUCGAUACCGAGUCUU

UUUGUCUAGUAGAACCAACCAACCAUUAAGGAGUUCACUAGCACAUCU
 UUGCGACCCGAUCGUCGGUGUGUCGCGUAAUACUUUUGUUAUGACGA
 GACAUACGCUCAAGCCCUGGGUAGCUAGUCGCGGAGGCACGUUACCG
 CGCACAACCCCUAUUCGUUUACAUGUACAUCGCAUCUGAGGUAGUACA
 CUUCCGGCGUACGUGAGUAAUUGCGCGUAAUAAGCGCGUGUUUAGCU
 GAUCCCCUCUCGUAUCGAGGUUAAGGCAGAUUAGUGCCCAGUAAUUG
 CGUUUUUUUGUCGUUGUCGCGAGAACGCGAUUUGCUCGCCAAAGC

7.1.3.5 Spike-in 9

GGGCCAGAUUACUCCAUUUCCGCCCAAGCUGCUCACAGUUAUACGGG
 CGUCGGCAUCCAGACCGUCGGCUGAUCGUGGUUUUACUAGGCUAGAC
 UAGCGUACGAGCACUAUGGUCAGUAAUUCUGGAGGAAUAGGUACCAA
 GAAAAAACGAACCUUUGGGUUCAGAGCUGUACGGUCGCACUGAACU
 CGGAUAGGUCUCAGAAAAACGAAUAUAGGCUUACGGUAGGUCCGAU
 GGCACAAAGCUUGUUCGGUAGCUGGCAUAAGAUUCCAUGCCUAGAU
 GUGAUACACGUUUCUGGAAACUGCCUCGUAUGCGACUGUCCCCGG
 GGUCAGGGCCCGCUGGUAAUUGCUGUAAAGAGGGGGCGUUGAGUCCGU
 CCGACUUCACUGCCCCCUUCAGCCUUUUGGGUCCUGUAUCCCAAUU
 CUCAGAGGUCCC GCCGUACGCUGAGGACCACCUGAAACGGGCAUCGU
 CGCUCUUCGUUGUUCGUCGACUUCUAGUGUGGAGACGAAUUGCCAGA
 AUUAUUAACUGCGCAGUUAAGGGCAGCGUCUGAGGAAGUUUGCUGCGG
 UUUCGCCUUGACCGCGGGGAAGGAGACUAACGAUAGCGACUCUGUCU
 CAGGGGAUCUGCAUAUGUUUGCAGCAUACUUUAGGUGGGCCUUGGCU
 UCCUUCGCGAGUCAAAACCGCGCAAUUAUCCCCGUCCUGAUUUACUGG
 ACUCGCAACGUGGGUCCAUCAGUUGUCCGUUAUACCAAGACGUCUAAG
 GGCGGUGUACACCCUUUUGAGCAAUGAUUUGCACAACCUGCGAUCACC
 UUAUACAGAAUUAUCAAUCAAGCUCCCCAGGAGCGGACUUGUAAGGA
 CCGCCGCUUUCGCUCGGGUCUGCGGGUUAUAGCUUUUCAGUCUCGAC
 GGGCUAGCACACAUCUGGUUGACUAGGCGCAUAGUCGCCAUUCACAG
 AUUUGCUCGGCAAUCAGUACUGGUAGGCGUUAGACCCCGUGACUCGU
 GGCUGAACGGCCGUACAA CUCGACAGCCGGUGCUUGCGUUUUACCC

7.1.3.6 Spike-in 12

GGGGCACAAGUUGCUGAAGUUGCGAGAGGGGGCGAUAAAGUGAGGCAGA
 CAGGCAUAAUUAAGAGGGGAGAGAAUUAAGCGUAGAUACUCUCCAAU
 AGUUGGUGAAGAAAUUUAUUAUGAGGCUGUUAAGCUGUAGCAACUCU
 UCCACGAGUAGGAAUUUUAGUUUUAGCUGGCUCUUUAAUUGGGAGGGA
 AGAUAACUGAAGCAGUUAAGAAUUAAGGAAAAGACUGGCAUUCCCG
 UGAUAAGCUUAAGAUUUUGGCUCUGUCCUAAGGUUGCUGAUUUUG
 GUUGUUGGAGACCCAUUGCAGGCAGGGGUUUUAGCUGUUAUGGCUAU
 UGCUGAAACAGCAAAAUUUGAUUAUAAUAAGGUUAAGGUAGGGUGCU
 AUAAAGAUAAUUUAUAAUUUUUGAUUGAAACCGAAGCGUUAGCUUUGG
 GUUAUGAAACUCCAUGAUUUUCAUUUAAUUUUUCCUAUUAAUUUUCU
 CCUAAAAGUUUCUUUAACAUAUAAUAAGGUUAAGGGAGAGCUCUAUG

AUUGUCUUCAAAAUACAAAGAUUAUUGAUGUAUUAUACUGGAGAGGUU
GUUAAAGGAAAUGUUGCAGUUGAGAGGGAUAAAAUAUCCUUUGUGGAU
UUAAAUGAUGAAAUUGAUAAAGAUAAUUGAAAAAAUAAAGGAGGAUGUUA
AAGUUUUUGACUUAAAAGGAAAAUUAUUUAUCUCCAACAUUUAUAGAUG
GGCAUAUACAUUAJAGAAUCUCCCAUCUCAUCCCAUCAGAGUUUGAGA
AAUUUGUAUUAAAAAGCGGAGUUAGCAAAGUAGUUUAUAGACCCGCAUG
AAUAGCAAUUAUUGCUGGAAAAGAAGGAAUUUUUGUUUAUGUUGAAUG
AUGCCAAAAUUUUAGAUGUCUAUGUUUAUGCUUCCUCCUGUGUUCAG
CUACAAACUUAGAAACAAGUGGAGCUGAGAUUACAGCAGA

7.2 Abbreviations

3' UTR	3' untranslated region
4tU	4-thiouracil
4sU	4-thiouridine
5' UTR	5' untranslated region
4tU	4-thiouracil
A	Adenine
A'	acceptor
Abp140	actin binding protein 140
ac4C	<i>N</i> ⁴ -acetylcytidine
ACL	anticodon loop
Air	Arginine methyltransferase-interacting RING finger protein
ALKBH	alkylation repair homolog
Arg	Arginine
Asn	Asparagine
ATP	Adenosine triphosphate
bp	base pair
BSA	bovine serum albumin
Bur	bypass UAS requirement
C	Cytosine
C32	C at position 32
Caf40	Ccr4 associated factor
Cbc2	cap binding complex 2
Ccr4	carbon catabolite repressor 4
Cdc	cell division cycle
cDNA	complementary DNA
CDS	coding sequence
Cft2	cleavage factor 2
chr	chromosome
Csl4	Cep1 synthetic lethal
CMC	<i>N</i> ³ -[<i>N</i> -cyclohexyl- <i>N'</i> - β -(4-methylmorpholinium) ethylcarbodiimide
CMV	cytomegalovirus
CRAC	Crosslinking and cDNA analysis

CTD	C-terminal domain
Ctk	Carboxy-terminal domain kinase
Ctr9	Cln three requiring
CUT	Cryptic unstable transcripts
Dcp	mRNA decapping
Dhh1	DEAD box helicase homolog
Dis3	chromosome disjunction
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
Dot1	disruptor of telomeric silencing
DTT	dithiothreitol
E	exit
<i>E. coli</i>	Escherichia coli
eCLIP	enhanced crosslinking and immunoprecipitation
ECL	enhances chemiluminescence
EDTA	ethylenediaminetetraacetic acid
Edc	enhancer of mRNA decapping
EGTA	ethyleneglycoltetraacetic acid
ERCC	external RNA control consortium
f5C	5-formylcytidine
FAM	fluorescein amidite
FBS	Fetal bovine serum
FTP	fat mass and obesity associated protein
FLP	flippase
G	Guanine
Gbp2	G-strand binding protein
GTF	General transcription factor
h	hour
HEK	human embryonic kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hm5C	5-hydroxymethylcytidine
Hrb1	Hypothetical RNA binding protein
HRP	Horseradish peroxidase
Hrp1	heterogenous nuclear ribonucleoprotein
I	Inosine
i6A37	<i>N</i> ⁶ -isopentnyladenine at position 37

IgG	immunoglobulin G
IPTG	Isopropyl- β -D-thiogalactopyranosid
Ile	Isoleucine
Ist3	increased sodium tolerance 3
hm5C	5-hydroxymethylcytidine
k	kilo
kDa	kilo Dalton
KO	knockout
L	liter
LB	<i>Lysogeny Broth</i>
LC-MS/MS	High-performance liquid chromatography-coupled triple quadrupole mass spectrometry
LDS	sample loading buffer
Leo1	left open reading frame
LIC	ligase independent cloning
lincRNA	long intergenic non-coding RNA
Luc7	Lethal unless cap-binding complex is produced
Lys	Lysine
m	milli
M	moles per liter
μ	micro
m1A	N^1 -methyladenosine
m ² ₂ G	N^2 - N^2 -dimethylguanosine
m3C	N^3 -methylcytidine
m5C	5-methylcytidine
m6A	N^6 -methyladenosine
MBP	maltose binding protein
Met	Methionine
METTL	Methyltransferase-like
Mex67	mRNA export factor of 67 kDa
min	minute
mRNA	messenger RNA
miRNA	micro RNA
MOPS	3-(<i>N</i> -morpholino)propanesulfonic acid
Mpe1	mutant Pcf11 extragenic suppressor
Mtr4	mRNA transport
Mud	Mutant U1 die

Msl5	Mud synthetic lethal
Nab	Nuclear polyadenylated RNA binding
Nam8	Nuclear accommodation of mitochondria 8
NAT10	<i>N</i> -acetyltransferase 10
ncRNA	non-coding RNA
NDR	nucleosome depleted region
NGS	Next-generation sequencing
NMD	Nonsense mediated decay
Not1	negative on TATA
Nrd1	Nuclear pre-mRNA Down-regulation
Npl3	nuclear protein localization
NUT	Nrd1-unterminated transcript
nt	nucleotide
OD	optical density
ORF	open reading frame
P	peptidyl
pA	poly adenylation
Pab1	pA binding protein 1
Paf1	RNA Pol II associated factor
PAGE	Polyacrylamide gel electrophoresis
Pan2/3	pA nuclease 2/3
PAP	pA polymerase
PAR-CLIP	Photoactivatable ribonucleoside enhanced crosslinking and immunoprecipitation
PBS	Phosphate buffered saline
Pcf11	protein 1 of cleavage and polyadenylation factor 1
PCR	polymerase chain reaction
PEG	Polyethylene glycol
PIC	Pre-initiation complex
PMSF	Phenylmethylsulfonylfluoride
PNK	polynucleotide kinase buffer
POI	Protein of interest
Pol	RNA-dependent Polymerase
Pop2	PGK promoter directed overproduction
Pub1	poly uridine binding
Pus	Pseudouridine synthase
PVDF	Polyvinylidene difluoride

Rai1	Rat1 interacting protein
Rat1	ribonucleic acid trafficking
RNA	ribonucleic acid
Rna15	pA mRNA metabolism
rRNA	ribosomal RNA
Rpb1	RNA polymerase
rpm	rounds per minute
RRACH	R is A or G; H is A,C or U
Rrp	Ribosomal RNA Processing
RT	reverse transcription
Rtf1	restores TBP function
Rtt103	Regulator of Ty1 transposition
Sen1	Splicing endonuclease
S	Sedimentation coefficient
S2P	Serine-2 phosphorylation
S5P	Serine-5 Phosphorylation
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SAM	S-adenosylmethionine
SCM	Synthetic complete medium
SDS	Sodium dodecyl sulfate
Set	SET domain containing
SGD	<i>Saccharomyces</i> genome database
Smg6	suppressor of morphological defects
Ski	super killer
SNP	single nucleotide polymorphism
Snp1	U1 small nuclear ribonucleoprotein
snRNA	small nuclear RNA
snoRNA	small nucleolar RNA
Spt	suppressor of Ty's
Sub2	suppressor of Brr1-1
SUT	stable unannotated transcript
t6A37	N ⁶ -threonyladenosine at position 37
T	Thymine
TAP	Tandem affinity purification
TBP	TATA box binding protein
TE	Translational efficiency
TEV	Tobacco etch virus

Tho2	suppressor of the transcriptional defect of Hpr1 by overexpression
Thr	Threonine
TRAMP	Trf4/5-Air1/2-Mtr4 polyadenylation complex
Trf	topoisomerase one-related function
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
Trm140	tRNA-methyltransferase 140
TRUB1	tRNA pseudouridine synthase B
tSNE	t-Distributed Stochastic Neighbor Embedding
TSS	transcription start site
TTS	transcription termination site
TT-seq	Transient transcriptome sequencing
Ψ	Pseudouridine
U	Unit
UMI	unique molecular identifier
Upf	UP frameshift
UV	Ultra violet
V	Volt
v/v	volume per volume
w/v	weight per volume
WT	wildtype
WTAP	Wilms tumor suppressor gene
YSPSTPS	Tyrosine-Serine-Proline-Serine-Threonine-Proline-Serine
YPD	Yeast extract peptone dextrose
Yra1	Yeast RNA annealing protein
YTH	YT521-B homology
Yth1	Yeast thirty kDa homolog
YTHDC	YTH domain containing
YTHDF	YTH domain family
Xrn1	exoribonuclease 1
XUT	Xrn1-untersminated transcripts

7.3 List of Figures

Figure 1: Translation by the ribosome.....	- 6 -
Figure 2: Schematic overview of RNA degradation.....	- 10 -
Figure 3: Model of the exosome complex.....	- 12 -
Figure 4: Modified nucleosides in the clover-leaf structure of the eukaryotic tRNA.	- 16 -
Figure 5: Schematic representation of known chemical modifications mapped in eukaryotic mRNA transcripts.	- 17 -
Figure 6: Structure of the N ³ -methylcytosine (m3C) modification.....	- 23 -
Figure 7: Number of PAR-CLIP cross-link sites and replicate correlation.	- 56 -
Figure 8: Distribution of degradation factor cross-link sites over the yeast transcriptome.....	- 60 -
Figure 9: Metagene analysis of degradation factor binding on protein-coding mRNAs.	- 63 -
Figure 10: Surveillance of aberrant nuclear non-coding RNA by components of the exosome and the TRAMP4 complex.....	- 65 -
Figure 11: Global co-occupancy and co-localization analysis reveals unexpected cooperation between factors from different complexes and pathways.....	- 68 -
Figure 12: Binding preferences reveal a link between decapping- mediated degradation and translation.	- 70 -
Figure 13: Location and recruitment of the decapping complex Dcp1/Dcp2 and decapping enhancers Edc3, Dhh1, and Edc2.	- 72 -
Figure 14: Schematic overview of PAR-CLIP and m3C-CLIP protocol.	- 85 -
Figure 15: PAR-CLIP of Trm140 analyzed by Western Blot and phosphor imaging.....	- 86 -
Figure 16: Identification of mutational signature of N ³ -methylcytidine (m3C) modification.	- 88 -
Figure 17: m3C modification causes reverse transcription block shown by primer extension assay on tRNA Serine.....	- 90 -
Figure 18: tRNA Arginine (CCU) contains m3C at C32 in <i>S.</i> <i>cerevisiae</i>	- 92 -
Figure 19: Presence of N ³ -methylcytidine modification in mRNA.....	- 94 -
Figure 20: m3C modification loss changes ribosome dynamics.....	- 96 -
Figure 21: Trm140 co-localizes with various RNA degradation factors.	- 98 -
Figure 22: Spearman correlation of 4tU-seq data for Trm140 KO and WT cells.	- 99 -
Figure 23: Functional characterization of m3C modification loss by using metabolic RNA labeling and sequencing in <i>S.</i> <i>cerevisiae</i>	- 100 -
Figure 24: PAR-CLIP of METTL8-3xFLAG in Flp-In™ T-REx™ 293 cells.....	- 101 -

Figure 25: METTL8 PAR-CLIP reveals localization of m3C modification in mRNA.....	- 103 -
Figure 26: Schematic drawing of Watson Crick base pairing between unmodified C and G (A) as well as H-H repulsion of m3C and G (B).....	- 107 -
Figure 27: Illustration of the m6A switch model.....	- 110 -
Figure 28: Different transcript classes have comparable U-content.....	- 141 -
Figure 29: Metagene profiles of yeast RNA degradation factors centered on translation start and stop sites in comparison to TIF-annotated TSS and pA sites.....	- 142 -
Figure 30: Distributions of transcript length, half-life, expression level and transcript optimality for yeast mRNAs.....	- 143 -
Figure 31: Occupancies of deadenylation factors (Ccr4, Pop2, Not1, Caf40, Pan2, and Pan3) compared to transcript length, optimality, expression level, and half-life.....	- 144 -
Figure 32: Occupancies of decapping factors (Dcp2, Dcp1, Edc2, Edc3, and Dhh1) compared to transcript length, optimality, expression level, and half-life.....	- 144 -
Figure 33: Occupancy of Xrn1 compared to transcript length, optimality, expression level, and half-life.....	- 145 -
Figure 34: Occupancies of exosome components (Rrp6, Csl4, Rrp40, Rrp4, and Rrp44) compared to transcript length, optimality, expression level, and half-life.....	- 146 -
Figure 35: Occupancies for components of the TRAMP complex (Air1, Trf5, Mtr4, Air2, and Trf4) compared to transcript length, optimality, expression level, and half-life.....	- 146 -
Figure 36: Occupancies for components of the Ski complex (Ski2, Ski3, Ski7, and Ski8) compared to transcript length, optimality, expression level, and half-life.....	- 147 -
Figure 37: Occupancies for components of the NMD pathway (Upf1, Upf2, Upf3, and Nmd4) compared to transcript length, optimality, expression level, and half-life.....	- 147 -
Figure 38: Correlation between binding to degradation factors and transcript length, codon-optimality, expression, and half-life.....	- 148 -
Figure 39: Metagene profiles for subunits of the TRAMP complex on snoRNA genes.....	- 149 -
Figure 40: Comparison of binding profiles on genes containing upstream sense NUTs with all mRNAs.....	- 150 -
Figure 41: Co-localization coefficients for 75 RNA binding factors.....	- 151 -

7.4 List of Tables

Table 1: Bacterial strains used in this study.	- 27 -
Table 2: Yeast strains used in this study.	- 27 -
Table 3: Human cell lines used in this study.	- 28 -
Table 4: Growth media used in this study.	- 29 -
Table 5: Media supplements used in this study.	- 29 -
Table 6: Spike-ins used in this study provided by Dr. Kristina Zumer.	- 30 -
Table 7: Primers and oligonucleotides used in this study were purchased from IDT.	- 30 -
Table 8: List of thermal cycler programs used in this study.	- 31 -
Table 9: List of plasmids used in this study.	- 32 -
Table 10: List of buffers and solutions used in this study.	- 32 -
Table 11: List of antibodies used in this study.	- 33 -
Table 12: Biological replicates for yeast PAR-CLIP experiments.	- 37 -
Table 13: Overview of RNA processing factors and their respective published PAR-CLIP experiments.	- 152 -
Table 14: High confidence m3C modification sites.	- 153 -

