Processing of non-coding RNA by Mlp1 in *Tetrahymena thermophila*

Kyra Kerkhofs

A dissertation submitted to the faculty of graduate studies in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Graduate Program in Biology
York University
Toronto, Ontario

December 2022

© Kyra Kerkhofs 2022

Abstract

RNA species are commonly transcribed as precursors and require post-transcriptional processing to become functional mature RNA transcripts. This includes the abundant cytoplasmic transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) that function in messenger RNA (mRNA) translation. Other non-coding RNAs also require extensive processing and assembly into ribonucleoproteins (RNPs), including small nuclear RNAs (snRNAs) that function in pre-mRNA splicing and small nucleolar RNAs (snoRNAs) involved in the processing and post-transcriptional modification of non-coding RNAs. The ciliate *Tetrahymena thermophila* is a highly studied eukaryotic model organism, however, many of its RNA processing pathways remain unexplored. In this work, we use molecular biology, biochemistry, cell biology and bioinformatic techniques to investigate the role of a novel La protein, Mlp1, in pre-tRNA, snRNP and snoRNP biogenesis.

Unlike previously studied genuine La proteins, Mlp1 lacks an RNA-binding domain typically required for high-affinity binding of uridylate-tailed La target RNAs. We confirm that Mlp1 performs typical La protein functions, including uridylate-dependent preferential binding of pre-tRNAs and RNA chaperone activity to promote processing and maturation of misfolded nascent pre-tRNAs. However, in contrast to pre-tRNA processing in other eukaryotes, depletion of Mlp1 results in 3′-trailer stabilization instead of rapid trimming by 3′-exonucleases, indicating that Mlp1 is linked to a fundamentally different mechanism of tRNA processing in *Tetrahymena thermophila*. We also find that Mlp1 associates with mature snRNAs lacking the typical high affinity uridylate binding site and a core protein component of the snRNP. More specifically, Mlp1 interacts with the U4/U6 di-snRNP complex, which is formed during splicing, and Mlp1 depletion affects assembly of this complex. Mlp1 depletion also results in diminished splicing efficiency of pre-mRNAs, further supporting a functional role for Mlp1 in splicing in this system. Lastly, we show that Mlp1 associates with all previously annotated snoRNAs which function as guide RNAs in post-transcriptional modification of non-coding RNAs. We use this as a metric to predict novel snoRNAs in *Tetrahymena thermophila* and validate expression. Our work demonstrates new roles for the variant genuine La protein Mlp1 in the biogenesis of non-coding RNAs critical for the translation of proteins.

In loving memory of my grandma,

Oma,

my first female role model showing my mom, sister and me

what passionate hard work can achieve

Acknowledgements

First and foremost, I would like to thank my supervisor Dr. Mark Bayfield for his constant guidance and support throughout the years. You provided me with enough space to become an independent scientist while also always being present to guide me toward the most productive path. You offered numerous opportunities to present my work at lab meetings, scientific seminars and conferences which were invaluable during my Ph.D. training. This project would not have been possible without your guidance. I would also like to acknowledge my supervisory committee, Dr. Katalin Hudak and Dr. Andrew White for their advice and constructive feedback throughout the years as part of my annual research evaluations, as well as their suggestions for this dissertation.

I would also like to thank previous Bayfield lab members for their guidance and support. In particular, Jyotsna and Jyoti. Jyotsna for providing me with extensive training and advice during my first year, making the start of my PhD considerably easier. Jyoti for patiently training me in everything I needed to work with *Tetrahymena*. Thank you to the current Bayfield lab members, Farnaz, Jennifer and Vanessa for sharing protocols and advice throughout my PhD journey, and Brittney for constant encouragement to continue to write this dissertation. The Bayfield lab undergraduate and graduate students that I worked with throughout the years, for keeping me on my toes by asking challenging questions and teaching me how to be a better mentor. I would also like to thank the other graduate students at LSB to provide advice and support. Throughout the years we went from being co-workers to cat parents and Loon Lake and Blue Mountain adventurers.

To my family in Belgium: Mama en Merel, om altijd in mij te geloven. Zelfs vanop meer dan 6000 km afstand waren jullie altijd geïnteresseerd in mijn labo avonturen. Opa, ondanks de verre afstand en zonder smartphone ben je me altijd blijven motiveren via je vele handgeschreven brieven verstuurd via de post. Papa, in een andere realiteit zou er nu zonder twijfel een afgeprinte versie van mijn 250 pagina thesis tegen de muur hebben gehangen op de een of andere manier. Je was altijd de trotste vader en dit heeft me altijd aangemoedigd om te blijven doorzetten.

And last but not least, Bruno, for always being there for me during this roller-coaster ride of a Ph.D. journey. You were there when I was at the top of the rollercoaster, and more importantly when the rollercoaster was plummeting down on some days. It did not matter if experiments went perfectly or poorly, you were always there to support me. Despite having no career in molecular biology, you were always ready to listen to my practice presentations and admire the data I generated in the lab that day. I have been incredibly fortunate to have had you by my side every day during this journey.

Table of Contents

List of Figures

Chapter I: Literature Review

Chapter II: Altered tRNA processing is linked to a distinct and unusual La protein in *Tetrahymena thermophila*

Chapter III: Genuine La protein Mlp1 functions in U4/U6 di-snRNP assembly and affects splicing in *Tetrahymena thermophila*

Chapter IV: The *Tetrahymena thermophila* La protein Mlp1 associates with box C/D and box H/ACA snoRNAs

Chapter V: Conclusion and future perspectives

List of Tables

Chapter II: Altered tRNA processing is linked to a distinct and unusual La protein in *Tetrahymena thermophila*

Chapter III: Genuine La protein Mlp1 functions in U4/U6 di-snRNP assembly and affects splicing in *Tetrahymena thermophila*

Chapter IV: The *Tetrahymena thermophila* La protein Mlp1 associates with box C/D and box H/ACA snoRNAs

List of Supplementary Figures

Chapter II: Altered tRNA processing is linked to a distinct and unusual La protein in *Tetrahymena thermophila*

Chapter III: Genuine La protein Mlp1 functions in U4/U6 di-snRNP assembly and affects splicing in *Tetrahymena thermophila*

Chapter IV: The *Tetrahymena thermophila* La protein Mlp1 associates with box C/D and box H/ACA snoRNAs

List of Supplementary Tables

Chapter II: Altered tRNA processing is linked to a distinct and unusual La protein in *Tetrahymena thermophila*

Chapter IV: The *Tetrahymena thermophila* La protein Mlp1 associates with box C/D and box H/ACA snoRNAs

List of Abbreviations

3′-RACE	3′-rapid amplification of cDNA ends
^{32}P	Phosphorus-32
AdoMet	S-adenosylmethionine
A. thaliana	*Arabidopsis thaliana*
Ala / A	Alanine
Arg / R	Arginine
Asn / N	Asparagine
Asp / D	Aspartic Acid
ATP	Adenosine Triphosphate
C. elegans	*Caenorhabditis elegans*
$CdCl_2$	Cadmium Chloride
cDNA	Complementary DNA
ChIP	Chromatin Immunoprecipitation
Ci	Curie Unit
CIP	Calf Intestinal Phosphatase
Co^{2+}	Cobalt
CPM	Counts Per Million
Ct	Cycle Threshold
CTP	Cytidine Triphosphate
D. discoideum	*Dictyostelium discoideum*
D. melanogaster	*Drosophila melanogaster*
D. rerio	*Danio rerio*
DAPI	4′,6-diamidino-2-phenylindole
di-snRNP	Dimeric snRNP
DNA	Deoxyribonucleic Acid
DTT	Dithiothreitol
DUF	Domain of Unknown Function
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic Acid
EMM	Edinburgh Minimal Medium
EMSA	Electrophoretic Mobility Shift Assay
Fe	Iron
g	Relative Centrifugal Force (RCF)

GEO	Gene Expression Omnibus
Gln / Q	Glutamine
Glu / E	Glutamic acid
H. sapiens	*Homo sapiens*
HCl	Hydrochloric Acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
$HgCl_2$	Mercuric chloride
His / H	Histidine
HRP	Horseradish Peroxidase
hLa	Human La
IgG	Immunoglobulin G
IGV	Integrated Genomics Viewer
Ile / I	Isoleucine
KCl	Potassium Chloride
Kd	Equilibrium Dissociation Constant
kDa	Kilodaltons
KO	Knockout
La	Lupus Autoantigen
LaM	La Motif
LARP	La-Related Protein
LC-MS/MS	Liquid Chromatography with Tandem Mass Spectrometry
Lhp1p	La-Homologous Protein
leu	Leucine Supplement
Leu / L	Leucine
Lys / K	Lysine
m_3G	2,2,7-trimethylguanosine
m^7G	N7-methylguanosine
M. musculus	*Mus musculus*
Met / M	Methionine
$MgCl_2$	Magnesium Chloride
Mlp1	Macronucleus Localized Protein of Unknown Function
mRNA	Messenger RNA
NaCl	Sodium Chloride
NCBI	National Center for Biotechnology Information
NMR	Nuclear Magnetic Resonance

NNS	Nrd1-Nab3-Sen1
PBS	Phosphate Buffered Saline
PCR	Polymerase chain reaction
PDB	Protein Database
pH	Potential of Hydrogen
Phe / F	Phenylalanine
PIC	Protease Inhibitor Cocktail
PMSF	Phenylmethylsulphonyl Fluoride
PNK	Polynucleotide Kinase
pre-5S rRNA	Precursor 5S Ribosomal RNA
pre-mRNA	Precursor Messenger RNA
pre-rRNA	Precursor Ribosomal RNA
pre-tRNA	Precursor Transfer RNA
qPCR	Quantitative PCR
qRT-PCR	Quantitative Reverse Transcriptase PCR
Rex1	RNA exonuclease-1
RIP-Seq	RNP-immunoprecipitation sequencing
RNA	Ribonucleic Acid
RPM	Rotations Per Minute
RT	Reverse Transcriptase
rDNA	Ribosomal DNA
rRNA	Ribosomal RNA
RNase	Ribonuclease
RNP	Ribonucleoprotein
RRM	RNA-Recognition Motif
S	Svedberg unit
S. cerevisiae	*Saccharomyces cerevisiae*
S. pombe	*Schizosaccharomyces pombe*
scaRNA	Small Cajal body-specific RNA
SDS	Sodium Dodecyl Sulfate
Ser / S	Serine
snRNA	Small Nuclear RNA
snRNP	Small Nuclear RNP
snoRNA	Small Nucleolar RNA
snoRNP	Small Nucleolar RNP

SSC	Saline Sodium Citrate
T. brucei	*Trypanosoma brucei*
T. thermophila	*Tetrahymena thermophila*
Taq	*Thermus aquaticus*
TBS	Tris-Buffered Saline
TBST	TBS + Tween20
TCA	Trichloroacetic Acid
TGIRT	Thermostable Group II Intron Reverse Transcriptase
Tgs1	Trimethylguanosine Synthase-1
Thr / T	Threonine
TPM	Transcripts Per Million
tri-snRNP	Trimeric snRNP
TRIS	tris(hydroxymethyl)aminomethane
tRNA	Transfer RNA
Trp / W	Tryptophan
Tris	Tris(hydroxymethyl)aminomethane
Tyr / Y	Tyrosine
U	Unit of Enzyme′s Catalytic Activity
ura	Uracil Supplement
USE	Upstream Sequence Element
UV	Ultraviolet
UTP	Uridine Triphosphate
UUU-3′OH	3′-terminal uridylates
Val / V	Valine
WT	Wild Type

Chapter I:

Review of Literature

1.1. Multiple non-coding RNAs are required for protein translation

RNA synthesis is achieved by three different multi-subunit RNA polymerases in eukaryotes (Carter & Drouin, 2009). RNA polymerase II transcribes all coding messenger RNAs (mRNAs) and many non-coding RNAs including small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs) (Guiro & Murphy, 2017). RNA polymerase I produces a single 47S ribosomal RNA (rRNA) (35S rRNA in yeast) which is processed into the 18S, 5.8S and 28S rRNAs (18S, 5.8S and 25S in yeast), which are components of the small and large ribosomal subunits (Klinge & Woolford, 2019). RNA polymerase III transcribes short transcripts, including the 5S rRNA which is a part of the large ribosomal subunit and transfer RNAs (tRNAs) (Dieci et al., 2007). RNA polymerase I and III are similar in that they produce a relatively small number of highly expressed transcripts (Paule & White, 2000), while RNA polymerase II transcribes all 20,000 coding genes (Guiro & Murphy, 2017).

Ribosome biogenesis involves (i) transcription, processing, and snoRNA-guided post-transcriptional modification of rRNA, (ii) mRNA transcription, splicing mediated by snRNAs, nuclear export, translation and nuclear import of ribosomal proteins, (iii) assembly of the small ribosomal subunit (40S – consisting of 18S rRNA and 33 ribosomal proteins) and the large ribosomal subunit (60S – consisting of 28S, 5.8S and 5S rRNA and 46 ribosomal proteins), and (iv) export of the small and large ribosomal subunits to the cytoplasm (**Figure 1**) (Henras et al., 2015; Lafontaine, 2015). Polypeptide production relies on coordinated function of different types of RNA and must be tightly regulated at different levels to ensure normal cell growth and proliferation (Moir & Willis, 2013). Translation is an intricate process where each puzzle piece needs to fit in exactly the right place in order to maintain homeostasis. The focus of this dissertation is on the processing and function of RNA polymerase III transcripts, including

nascent precursor tRNAs (pre-tRNAs) and the U6 snRNA and RNA polymerase II transcripts including U1, U2, U4 and U5 snRNAs and snoRNAs.

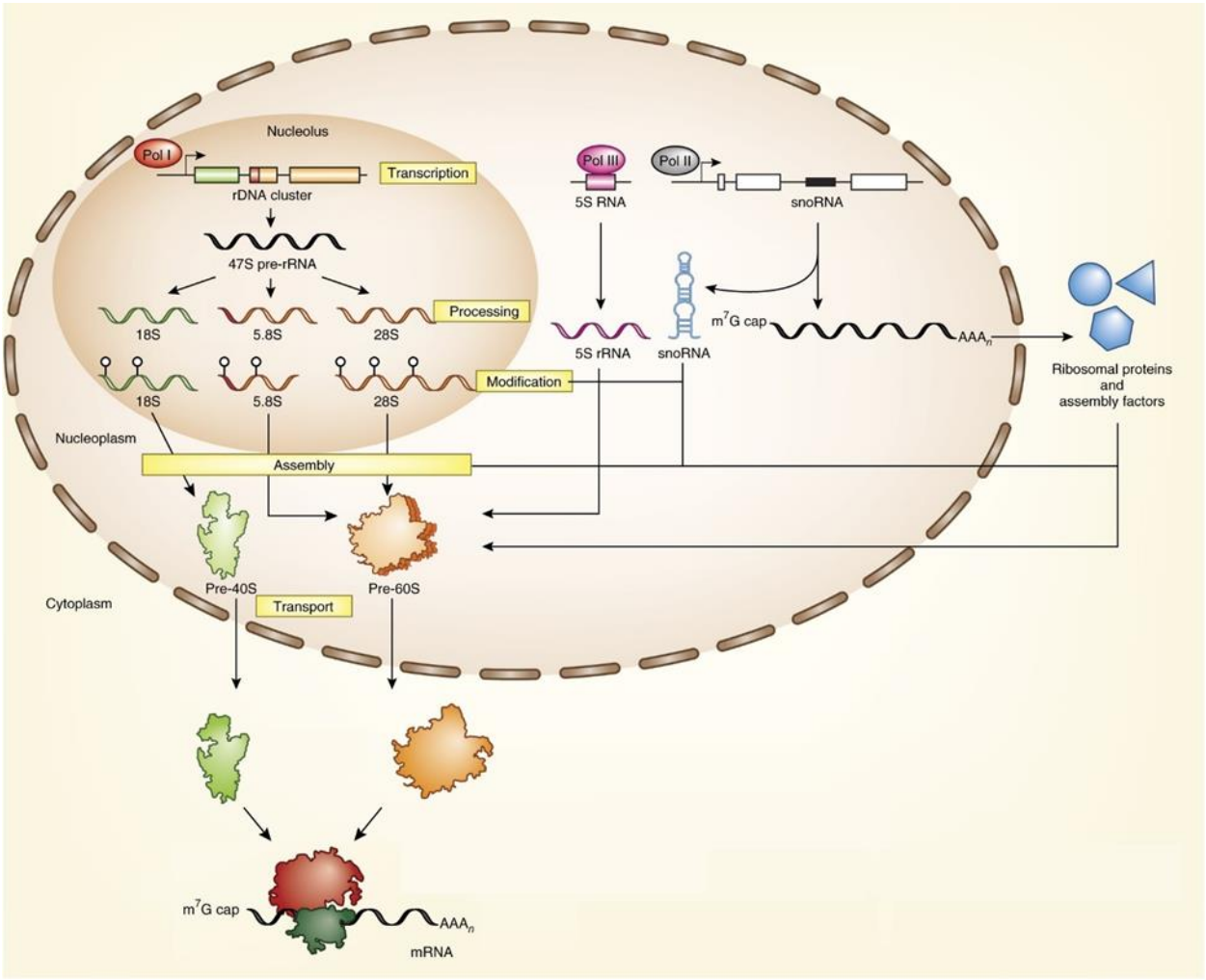


Figure 1. Protein translation requires non-coding RNAs transcribed by all three RNA polymerases. RNA polymerase II (pol II) is found in the nucleus and transcribes messenger RNAs (mRNAs) and multiple non-coding RNAs such as small nuclear RNAs (snRNAs) (not shown) and small nucleolar RNAs (snoRNAs). mRNAs encoding ribosomal proteins and assembly factors are translated in the cytoplasm and imported into the nucleus for ribosome assembly. RNA polymerase I (pol I) is located in the nucleolus and transcribes a single polycistronic transcript from ribosomal DNA (rDNA), the 47S pre-rRNA, that is processed into 18S, 5.8S and 28S rRNA. Further maturation occurs through post-transcriptional modifications by snoRNAs. RNA polymerase III (pol III) is found in the nucleus and transcribes tRNAs (not shown) and 5S rRNA which is a component of the large 60S ribosomal subunit. The small and large ribosomal subunits 40S and 60S, respectively, are assembled in the nucleus and transported into the cytoplasm for translation of mRNAs into proteins. Figure adapted from (Lafontaine, 2015).

1.2. Transcription and biogenesis of non-coding RNAs

1.2.1. RNA polymerase I transcription of rRNA and maturation

Transcription of rRNA in eukaryotes takes place in the nucleolus from ribosomal DNA (rDNA) that is encoded in multiple copies throughout the genome. For example, the yeast *Saccharomyces cerevisiae* encodes approximately 150 copies per haploid genome and humans between 300 and 400 copies per diploid genome (Henras et al., 2015). The rRNA is transcribed as a large precursor rRNA (pre-rRNA) by RNA polymerase I, containing the 18S, 5.8S and 28S rRNA (see **Figure 1**) and external (5′-ETS and 3′-ETS) and internal spacers (ITS-1 and ITS-2) (Klinge & Woolford, 2019), whereas the 5S rRNA is transcribed by RNA polymerase III (see **Figure 1**) (Dieci et al., 2007).

Processing of the nascent pre-rRNA is initiated co-transcriptionally and involves multiple complex endo- and exonucleolytic processing steps (Henras et al., 2015). In addition to the multiple endo- and exonucleolytic processing step, rRNAs molecules obtain extensive nucleotide modifications by up to approximately 200 snoRNAs (Lestrade & Weber, 2006). Most snoRNAs are transcribed by RNA polymerase II (Guiro & Murphy, 2017) and function through base pairing with the target rRNA sequence. Two classes of snoRNAs, box C/D and box H/ACA snoRNAs, direct 2′-O-methylation of the RNA backbone at the 2′-O position of the sugar ring and pseudouridylation which forms a 5′-ribosyl isomer of uridine named pseudouridine or Ψ, respectively, to the rRNA (**Figure 2**) (Decatur & Fournier, 2002). Post-transcriptional modifications are found in highly conserved rRNA regions which are often functionally important (Decatur & Fournier, 2002).

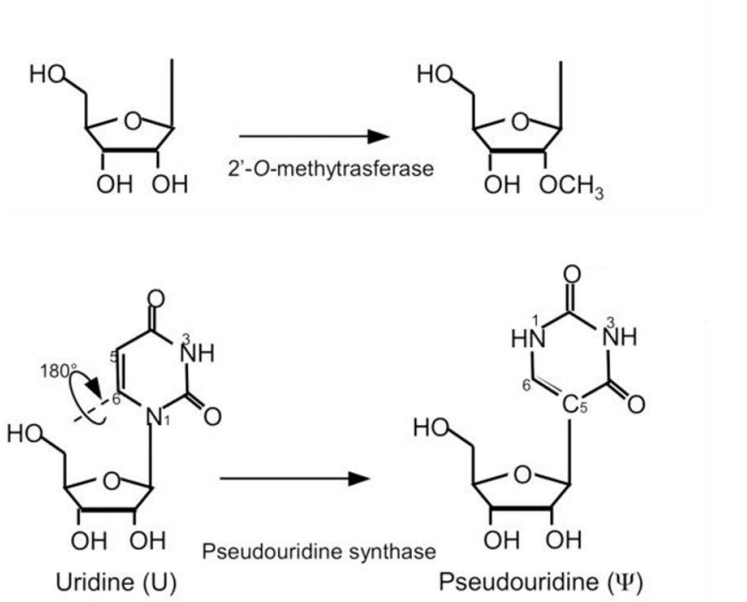


Figure 2. Schematic representation of 2′-O-methylation and pseudouridylation of RNA.
2′-O-methylation post-transcriptional modification is the addition of a methyl group at the 2′-O position of the RNA backbone by 2′-O-methyltransferases. Pseudouridine or Ψ post-transcriptional modification is a rotation isomer of uridine (U) converted by pseudouridine synthetases. Figure from (Karijolich & Yu, 2010).

1.2.2. snRNA transcription by RNA polymerase II and post-transcriptional modifications

RNA polymerase II transcribes nascent precursor mRNAs (pre-mRNAs) which obtain a 5′ N7-methylguanosine (m^7G) cap co-transcriptionally linked via an inverted 5′-5′ triphosphate bridge to the first nucleoside of the nascent transcript (**Figure 3a**) (Ghosh and Lima, 2010). The m^7G cap co-transcriptionally added to nascent mRNA has multiple functions, including protection against 5′-3′ exonucleases, efficient splicing of pre-mRNAs, polyadenylation at the 3′-end, nuclear export and efficient translation (Lewis & Izaurflde, 1997; Pabis et al., 2013). In addition to transcription of mRNAs, RNA polymerase II transcribes a large number of short non-coding RNAs, including snRNAs and snoRNAs (Guiro & Murphy, 2017). Well-characterized RNA polymerase II transcribed non-coding snRNAs include the spliceosomal snRNAs U1, U2, U4 and U5. Mature RNA polymerase II transcribed snRNA transcripts contain a 2,2,7-

trimethylguanosine (m_3G) cap which is formed by further dimethylation of the m^7G cap at the N2 guanine position of the m^7G cap structure (**Figure 3b**) (Ghosh & Lima, 2010). In addition to modification of the 5′-cap, snRNAs obtain two types of post-transcriptional modification including 2′-O-methylation and pseudouridylation which are guided by a specific subset of snoRNAs located in the Cajal bodies, named small Cajal body-specific RNAs (scaRNAs) (Jády & Kiss, 2001; Karijolich & Yu, 2010). Post-transcriptional modifications found in spliceosomal snRNAs are highly conserved throughout evolution and are located in functionally important regions for splicing, including regions of snRNA/snRNA and pre-mRNA/snRNA hybridization (**Figure 4**) (Karijolich & Yu, 2010).

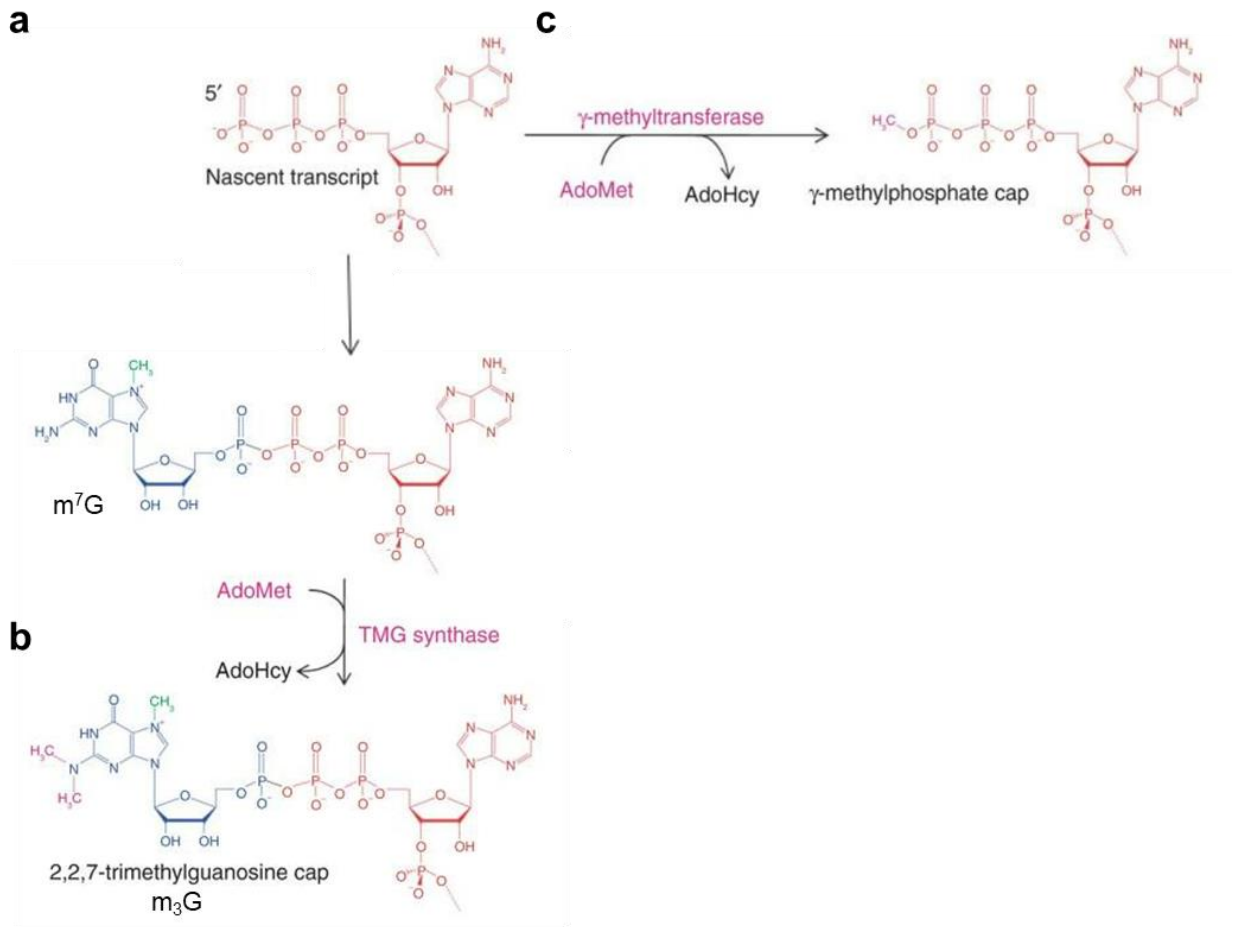


Figure legend on next page

Figure 3. Synthesis of m^7G and m_3G-cap formation in RNA polymerase II transcripts.
a Transcripts produced by RNA polymerase II co-transcriptionally obtain a 5′ N7-methylguanosine (m^7G) through linkage via an inverted 5′-5′ triphosphate bridge to the first nucleoside. **b** A subset of non-coding transcripts are further post-transcriptionally processed to obtain a 2,2,7-trimethylguanosine (m_3G) cap structure which is obtained by further dimethylation of the m^7G cap at the N2 guanine position of the m^7G cap structure. Transcripts containing a m_3G cap structure include RNA polymerase II transcribed snRNAs and some snoRNAs. **c** The γ-methylphosphate cap is formed through transfer of a methyl group on the primary transcript by γ-methyltransferases. Transcripts containing a γ-methylphosphate cap include RNA polymerase III transcript U6 snRNA. Figure adapted from (Ghosh & Lima, 2010).

The spliceosome consists of five uridine-rich snRNAs, U1, U2, U4, U5 and U6 (**Figure 4**), and hundreds of proteins (Jurica & Moore, 2003). While RNA polymerase II transcribes U1, U2, U4 and U5, the U6 snRNA is transcribed by RNA polymerase III and contains a γ-monomethyl cap (Singh & Reddy, 1989). Methylation of the γ-phosphate of the 5′-triphosphate cap is catalyzed by the γ-methyltransferase transferring a methyl group from the S-adenosylmethionine (AdoMet) to a γ-phosphate oxygen at the nascent 5′-end of the RNA polymerase III transcript (**Figure 3c**) (Gupta et al., 1990; Shimba & Reddy, 1994).

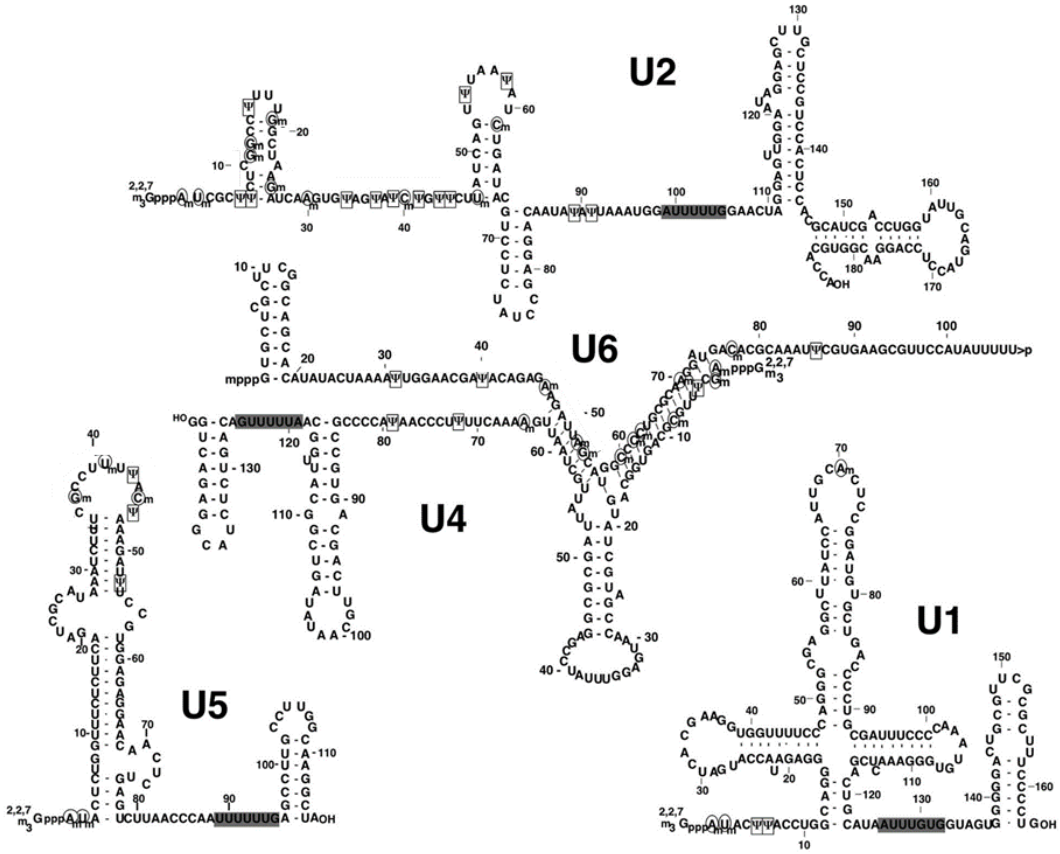


Figure 4. Schematic representation of the secondary structures of spliceosomal snRNAs.
The secondary structures of the spliceosomal small nuclear (snRNAs) are shown. Post-transcriptional modifications including pseudouridylation and 2′-O-methylations are highlighted by a rectangle or circle surrounding the individual nucleotide, respectively. The Sm-binding sites are highlighted with dark gray boxes. Figure adapted from (Karijolich & Yu, 2010).

1.2.3. snoRNA transcription by RNA polymerase II

snoRNAs are also transcribed by RNA polymerase II and are found abundantly in the nucleolus within ribonucleoprotein (RNP) complexes named snoRNPs which function as guide RNAs in post-transcriptional modification of rRNA, snRNA and tRNAs (Kiss, 2002). Two major classes of snoRNAs exist based on sequence and structural elements, including box C/D and box H/ACA snoRNAs. Box C/D snoRNAs guides 2′-O-methylation and box H/ACA guided pseudouridylation.

Different eukaryotes employ different mechanisms of snoRNA transcription, including transcription from independent genes, polycistronic transcription and processing from mRNA introns (Kufel & Grzechnik, 2019) (**Figure 5**). For example, the yeast *S. cerevisiae* expresses 76 snoRNAs (Piekna-Przybylska et al., 2007) and mostly employs transcription from independent genes. However, a small number of snoRNAs are found within mRNA introns (8 snoRNAs) and are organized in polycistrons (5 clusters encoding 17 snoRNAs) (Piekna-Przybylska et al., 2007). In contrast, humans encode the majority of snoRNA genes within long non-coding genes or introns of mRNAs in the same orientation with only a small number of human snoRNAs transcribed by RNA polymerase II as independent genes. Host genes encoding snoRNAs are often 5′-terminal oligo pyrimidine or 5′-TOP mRNAs which encode for highly expressed ribosomal proteins or proteins involved in ribosome biogenesis or translation (Ojha et al., 2020; Smith & Steitz, 1998; Tycowski et al., 1993). snoRNAs are transcribed as precursors and are required to undergo maturation at the 5′- and 3′-ends. The fully processed snoRNAs lack a poly(A) tail and are between 60 and 300 nucleotides long (Falaleeva & Stamm 2013).

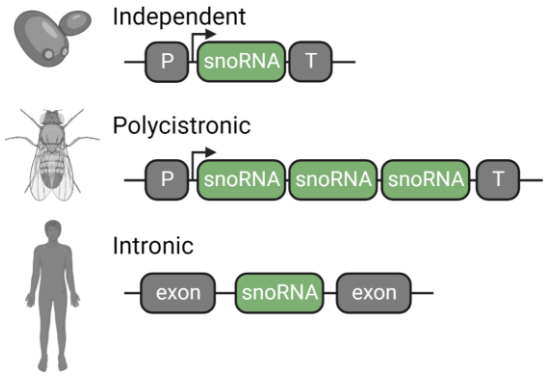


Figure 5. Genomic organization of snoRNA genes in different eukaryotic species.

Genomic organization of small nucleolar (snoRNA) transcription generally follows a similar mechanism for species within a certain branch of the evolutionary tree. The yeast *S. cerevisiae* for example encodes the majority of snoRNAs as independent genes under regulation of its own promotor and terminator, while *D. melanogaster* and *C. elegans* snoRNAs are generally transcribed as a polycistronic transcript. In vertebrates, including humans, the majority of snoRNAs are encoded within the introns of mRNAs or long non-coding RNAs. P: promotor, T: terminator. Figure adapted from (Kufel & Grzechnik, 2019) in BioRender.

1.2.4. RNA polymerase III transcription of tRNAs and other non-coding transcripts

RNA polymerase III transcribes short transcripts that make up approximately 15% (Turowski & Tollervey, 2016) of the total cellular RNAs and include many different RNAs involved in protein synthesis (*i.e.* 5S rRNA and tRNAs), pre-rRNA and pre-tRNA processing factors (*i.e.* RNase MRP and RNase P RNAs), co-translational translocation of nascent polypeptide chains during endoplasmic reticulum translation (*i.e.* 7SL RNA), transcription regulation of RNA polymerase II elongation factor P-TEFb (*i.e.* 7SK RNA) and pre-mRNA splicing (*i.e.* U6 snRNA) (Dieci et al., 2007). In some species, certain transcripts are transcribed by RNA polymerase III instead of RNA polymerase II. A few examples include the telomerase RNA functioning in telomeric DNA extension in ciliates, the U3 snoRNA in plants which is a component of the rRNA processing complex and a small number of snoRNAs *in S. cerevisiae*, *Arabidopsis thaliana*, *Caenorhabditis elegans* and *Drosophila melanogaster* (Dieci et al., 2007).

The most abundantly transcribed polymerase III gene products include 5S rRNA and tRNAs. The 5S rRNA is a part of the large ribosomal subunit (Klinge & Woolford, 2019), while tRNAs function as adaptors between the mRNA codon sequence and corresponding amino acid post-transcriptionally linked to the 3′-end of the tRNA (O′Donoghue et al., 2018). RNA polymerase III transcription of tRNAs generally uses intragenic promotors as transcription initiation elements, which are the A Box and B Box localized within the genomic tRNA sequence and the distance between these two elements is variable depending on identity of the tRNA (**Figure 6**) (Arimbasseri & Maraia, 2016). Termination occurs upon encountering the genomic stretch of deoxyadenosines functioning as the termination signal for RNA polymerase III, leading to incorporation of a 3′-terminal uridylate stretch in every nascent RNA polymerase III transcript (Arimbasseri et al., 2013). La proteins are the first to bind nascent RNA polymerase III

transcripts through association with the high-affinity 3′-uridylate RNA binding site (Fairley et al., 2005). Another typical RNA polymerase III transcript feature is an uncapped 5′-end (*i.e.* 5′-triphosphate sequence) (see **Figure 3a**) (Yan et al., 2022).

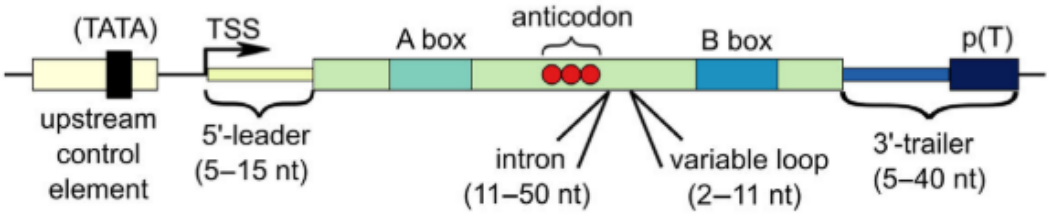


Figure 6. Schematic representation of a tRNA gene.
Nascent pre-tRNA transcription includes RNA polymerase III transcription initiation, elongation, and termination. The RNA transcription termination signal is encoded in the genome as a stretch of deoxyadenosines. This results in stuttering of the RNA polymerase and as a result, a stretch of uridylates is produced at the 3′-end of each RNA polymerase III transcript. Figure adapted from (Porter et al., 2021).

1.2.5. tRNAs are encoded at multiple copies in the genome

The human genome consists of more than 400 nuclear-encoded tRNA genes and are grouped into 20 different isotypes based on the amino acid added to the tRNA (Chan and Lowe, 2016 - http://gtrnadb.ucsc.edu GRCh37/hg19). tRNA isotypes containing different anticodons that can decode the same amino acid are named isoacceptors. For example, the alanine amino acid (codons GCU, GCC, GCG, GCA) can be decoded with the isoacceptors tRNAs Ala^{AGC}, Ala^{CGC} and Ala^{UGC}, while the phenylalanine amino acid (codons UUC and UUU) is decoded by only one isoacceptor: Phe^{GAA} (Chan and Lowe, 2016 - http://gtrnadb.ucsc.edu GRCh37/hg19). Typically, amino acids are decoded by one or more tRNA isoacceptors. The approximately 400 human tRNAs genes are divided among 48 different isoacceptor families decoding the 61 sense codons into the 20 different amino acids. For example, the above mentioned tRNAs Ala^{AGC}, Ala^{CGC} and Ala^{UGC} decode GCU, GCG, GCA synonymous codons as alanine, respectively. The GCC codon is also translated as alanine by one of these tRNA isoacceptors because of the

degeneracy of the genetic code, particularly at the wobble position of the tRNA anticodon (position 34 in the tRNA) allowing non-cognate RNA base pairing following Watson-Crick wobble rules (Chan & Lowe, 2016; Pinkard et al., 2020).

Isodecoders are tRNA isoacceptors that contain the same anticodon sequence but with different nucleotide composition in the body of the tRNA. For example, Leu^{CAG} is encoded by 9 copies in the genome, however, two mature tRNA Leu^{CAG} isodecoders exist (Chan and Lowe, 2016 - http://gtrnadb.ucsc.edu GRCh37/hg19). Regardless of the tRNA isodecoder sequence variability, the corresponding mature tRNA is required to retain the ability to fold into the typical tRNA structure to function as a transfer molecule in translation (Geslain & Pan, 2010).

1.3. Mature tRNAs adopt a classic cloverleaf secondary structure

The maturation process of a tRNA includes multiple steps, starting with transcription of the nascent tRNA by RNA polymerase III. Next, the nascent pre-tRNA is processed into a mature tRNA through removal of 5′-leader and 3′-trailer sequences, intron splicing, and CCA addition at the 3′-end by nucleotidyltransferase (**Figure 7a**) (Maraia & Lamichhane, 2011). Throughout the maturation process, post-transcription modifications are added to the tRNA that enhance correct folding of the tRNA into the classic cloverleaf secondary and L-shaped tertiary structure (Helm, 2006). The length of mature tRNAs is approximately between 74 and 93 nucleotides, and despite overall lack of sequence conservation, generally contain four or five stems that are formed through internal base pairing, including the acceptor stem, D-stem, anticodon stem, T-stem and, depending on the tRNA, a variable stem (**Figure 7b**) (Berg et al., 2019). Additionally, the tRNA structure contains a conserved D-loop, anticodon stem loop and T-loop (**Figure 7b**). Formation of the four stem structures in the tRNA require sufficient Watson–Crick A-T and C-G or G-U wobble pairs between opposing nucleotides (Berg et al., 2019).

The discriminator base is the 3′-most terminal nucleotide (N_{73}) and is found as a single nucleotide overhang that serves as the preceding nucleotide for post-transcriptional addition of the CCA sequence by ATP(CTP):tRNA nucleotidyltransferases (**Figure 7b**) (Tomita & Yamashita, 2014). The CCA sequence serves as the attachment site for amino acids transferred by aminoacyl-tRNA synthetases (Pang et al., 2014). The conserved tRNA structure is not only important for interactions with translation factors, ribosomes, and mRNAs during translation, but also for interactions with the aminoacyl-tRNA synthetases that deliver the correct amino acid to corresponding tRNA isotype (Carter & Wolfenden, 2015).

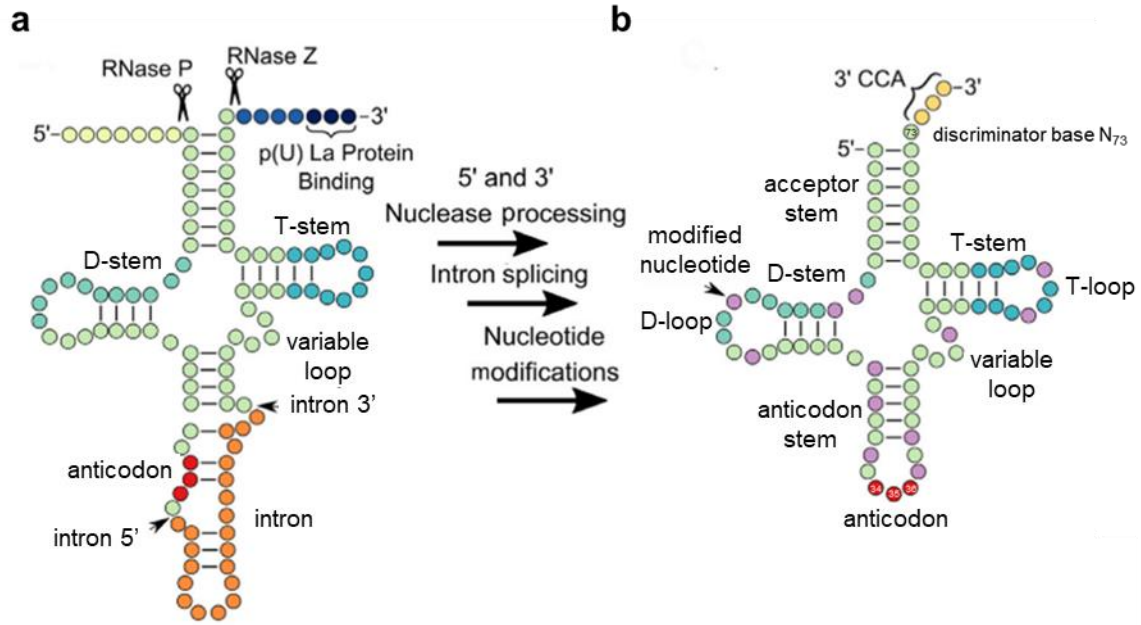


Figure legend on next page

Figure 7. Processing of nascent pre-tRNAs into mature tRNAs.

a Schematic representation of a nascent pre-tRNA. RNA polymerase III transcription results in 5′- and 3′-terminal extensions shown in *yellow* and *blue*, respectively. The La protein associates with the 3′-terminal stretch of uridylates shown in *dark blue*. Processing of nascent pre-tRNA involves 5′-leader, 3′-trailer and intron removal. Following 3′-end processing, the pre-tRNA obtains a post-transcriptionally added CCA sequence. **b** The mature tRNA secondary structure adopts a typical cloverleaf fold and the tertiary structure a L-shape (not shown). Different structural regions include loops and stems which are formed as a result of internal base pairing. These include the acceptor stem, D-stem loop, anticodon stem loop, T-stem loop and, depending on the tRNA, a variable stem loop. The anticodon triplet is in highlighted in *red.* The discriminator base at position N_{73} is located upstream of the post-transcriptionally added CCA sequence (shown in *yellow*). During the maturation process of the nascent pre-tRNA, multiple post-transcriptional modifications are obtained (highlighted in *purple*). Figure adapted from (Porter et al., 2021).

1.4. Processing of nascent pre-tRNAs in eukaryotes

1.4.1. Processing of the 5′-leader by endonuclease RNase P

In all three domains of life, processing of the 5′-leader is completed by the endonuclease RNase P through phosphodiester hydrolysis resulting in a 5′-monophosphate terminus (Gößringer et al., 2012). The RNase P enzyme consists of multiple proteins and the ribonuclease P RNA. The RNA subunit possesses catalytic activity and is highly conserved between bacteria, Archaea and eukaryotes (Chen & Pace, 1997). One major exception to standard pre-tRNA processing has been described in the parasitic archeon *Nanoarchaeum equitans* (Randau et al., 2008) in which tRNA transcription does not start upstream resulting in absence of the 5′-leader extensions. As a result, these tRNAs do not require 5′-processing by RNase P and therefore produce mature tRNAs containing a 5′-triphosphate. So far, this is the only identified species to not encode any of the components of the RNase P complex (Randau et al., 2008).

1.4.2. Processing of the 3′-trailer by either endonuclease RNase Z or exonuclease Rex1 is determined by the La protein

Once the 5′-leader has been processed by RNase P, the endonuclease RNase Z proceeds to cleave the 3′-trailer to remove the extension sequence up to the discriminator base (N_{73}, see **Figure 7**). This processing step results in a 3′-end matured pre-tRNA and concomitantly, release of the La protein from the tRNA which was bound to the 3′-terminal uridylate stretch. Generally, nascent pre-tRNA processing follows these consecutive steps in a system where La is associated with the 3′-trailer of the tRNA, termed the La-dependent pre-tRNA processing pathway (**Figure 8** – pathway 1). Here, the 5′-leader extension is removed from the nascent pre-tRNA by RNase P prior to processing of 3′-end extensions by RNase Z (Van Horn et al., 1997; Yoo & Wolin, 1997). Interestingly, in situations where La is deleted or limiting and the 3′-trailer sequences of the nascent pre-tRNA remain unprotected, the 3′ to 5′ exonuclease Rex1 rapidly processes the 3′-trailer sequence prior to 5′-trailer cleavage by RNase P (**Figure 8** – pathway 2) (Copela et al., 2008). This process is termed the La-independent pre-tRNA processing pathway.

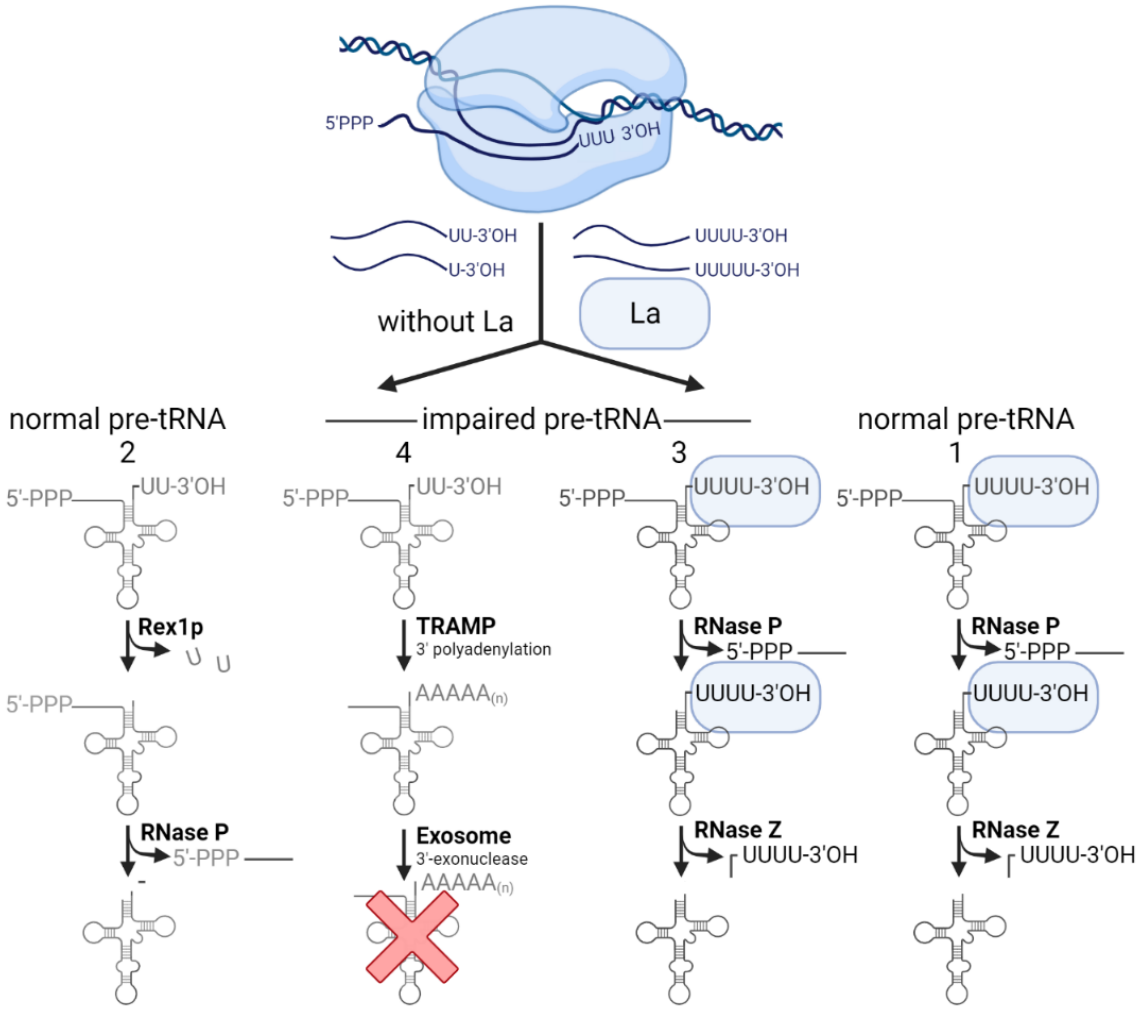


Figure 8. The La protein assists in processing of nascent pre-tRNAs.

RNA polymerase III (pol III) transcribes pre-tRNAs and produces a stretch of 3′-uridylates at the end of each nascent transcript as a result of the deoxyadenosine stretch encoded in the genome that functions as a transcription termination signal. The La protein binds the 3′-uridylate sequence provided that the stretch of uridylates is of a sufficient length. Normal pre-tRNAs are processed either by a La-dependent (# 1) or La-independent (#2) pathway resulting in end-processed tRNAs. The main difference between normal pre-tRNA processing in presence or absence of La is the order of 5′- and 3′-end processing and 3′-end processing by an endo- or exonuclease. The endonuclease RNase Z functions in the La-dependent pathway, while the exonuclease Rex1 functions in the La-independent pathway because there is no 3′-end protection of nascent pre-tRNA by La binding. In a La-independent pathway, structurally impaired pre-tRNAs (#4) are targeted for degradation by the nuclear surveillance: the misfolded pre-tRNAs are polyadenylated by the TRAMP complex followed by exosome degradation. The presence of a functional La protein enhances pre-tRNA folding and processing (#3), resulting in end-processed tRNAs. Figure adapted from (Maraia & Lamichhane, 2011) in BioRender.

As mentioned before, transcription termination by RNA polymerase III occurs after encountering a stretch of deoxyadenosines in the template DNA strand resulting in production of a 3′-terminal stretch of uridylate RNA (**Figure 8**). The minimal number of encoded deoxyadenosines for efficient RNA polymerase III transcription termination varies between eukaryotes, for example *Homo sapiens* requires only 4 deoxyadenosines, while *Schizosaccharomyces pombe* requires at least 5 deoxyadenosines and *S. cerevisiae* 6 deoxyadenosines or more (Arimbasseri et al., 2013). Typically, La proteins require between 3-6 uridylates for high affinity binding depending on the RNA polymerase III termination stretch in different species with 3 uridylates for human La (Mathews & Francoeur, 1984; Stefano, 1984) and 4-5 uridylates for *S. pombe* Sla1p (Huang et al., 2005). The number of encoded deoxyadenosines correlates with the minimal number of uridylates on the nascent pre-tRNA for high affinity La protein binding. After RNA polymerase III transcription, a heterogenous population of pre-tRNAs are present in the nucleus with varying 3′-uridylate tail lengths as a result of varying termination signal length, *i.e.* the number of deoxyadenosines encoded in the genome and stuttering of the polymerase upon encountering the deoxyadenosine termination signal. Since La proteins have a higher affinity for longer 3′-uridylate tails, nascent pre-tRNAs with a relatively shorter uridylate stretch will remain unbound when the La protein is limiting (Stefano, 1984). As a result of this preferential binding, both the La-dependent and La-independent nascent pre-tRNA processing pathway can take place simultaneously (**Figure 8** – pathways 1 and 2) and both result in the formation of a mature tRNA (Maraia & Lamichhane, 2011).

Nascent pre-tRNAs require an appropriate fold in order to be processed by the endonucleolytic tRNA processing enzymes that remove the 5′-leader and 3′-trailer. Folding of nascent pre-tRNAs takes place co-transcriptionally (Leamy et al., 2019), however, multiple different conformations could be obtained and occasionally leading to misfolded pre-tRNAs trapped in a misfolded structure (Vakiloroayaei et al., 2017). These transcripts are targeted for degradation

by the nuclear surveillance (**Figure 8** – pathway 4), but can be rescued by the RNA chaperone activity from La that functions in correct folding of pre-tRNAs (**Figure 8** – pathway 3) (Chakshusmathi et al., 2003). Another important function of the La protein is nuclear retention of tRNAs and other RNA polymerase III transcripts. The La protein contains a nuclear localization signal and retains the transcript in the nucleus through direct interactions avoiding premature export into the cytoplasm of incomplete or incorrectly processed transcripts (Bayfield et al., 2007; Intine et al., 2002).

In conclusion, the presence of the La protein on the 3′-terminal uridylates on a nascent pre-tRNA determines the usage of the 3′-end exonucleolytic or endonucleolytic maturation process, with accessibility of the 3′-end inhibited by La limiting access of the 3′-exonuclease Rex1 to the 3′-end and therefore 3′-maturation takes place through the endonucleolytic RNase Z pathway. In contrast, absence of the La protein at the 3-terminus of the nascent pre-tRNA allows 3′-exonucleases to process the 3′-trailer sequence. Both the exonucleolytic and endonucleolytic pre-tRNA processing pathway results in functional mature tRNAs, indicating that these enzymes function redundantly in processing of pre-tRNA 3′-extensions (Yoo & Wolin, 1997). The yeast La protein, Lhp1p, in *S. cerevisiae* is not essential and in *LHP1* deletion strains, the 3′-trailer sequences of nascent pre-tRNAs are processed by the 3′-exonuclease pathway (Yoo & Wolin, 1994) since the RNase Z endonuclease processing pathway is dependent on presence of Lhp1p on the 3′-uridylate extension (Yoo & Wolin, 1997).

1.4.3. Post-transcriptional modifications

Post-transcriptional modification involves the enzymatic covalent attachment of a chemical to the standard RNA bases to alter their properties in multiple ways (Smith et al., 2019). As part of the tRNA maturation process, tRNAs obtain post-transcriptional modifications at different steps during processing (Motorin & Helm, 2010). Modified nucleotides have been found on La-

associated pre-tRNAs, indicating that post-transcriptional modifications also occur in early steps of pre-tRNA processing (Hendrick et al., 1981). Nuclear-encoded tRNAs have the highest density of post-transcriptional modifications with an average of 13 modifications per molecule, while mitochondrial tRNAs only contain 5 modifications on average (Pan, 2018). tRNAs also contain the most diverse nucleotide modifications with more than 90 different types of modifications identified across archaea, bacteria, and eukaryotes (Cantara et al., 2011). Post-transcriptional modifications of tRNAs have several functions, including stability of the tRNA structure and appropriate decoding of the mRNA (El Yacoubi et al., 2012; Lorenz et al., 2017).

While the presence of modification in tRNAs is advantageous for folding and stability of the tRNA, reverse transcriptase-based sequencing methods are severely hindered by the stable and extensive secondary structure and the presence of many post-transcriptional modifications in tRNAs. Multiple methods have been developed to circumvent the block generated by modifications for complementary DNA (cDNA) extension by reverse transcriptase. First, AlkB-facilitated RNA methylation sequencing (ARM-seq) utilizes an *Escherichia coli* AlkB enzyme which removes methyl groups from tRNA to avoid generation of truncated cDNA (Cozen et al., 2015; Zheng et al., 2015). Second, a thermostable group II intron reverse transcriptase (TGIRT) possesses high processivity during cDNA generation from highly stable RNA structures and thereby enhances production of full-length cDNA transcripts (Zheng et al., 2015).

1.4.4. Splicing of intron-containing tRNAs

A subset of tRNAs contain intervening intron sequences that must be spliced before the tRNA can be a substrate for amino acid charging and therefore function as an adapter molecule in translation. The intron is typically found one nucleotide 3′-terminal to the anticodon in yeasts and vertebrates and typically contains a sequence that is complementary to the anticodon and base pairing between the intron and anticodon determines the secondary structure of the pre-

tRNA (Johnson & Abelson, 1983) (see **Figure 7**). Correct intron splicing is important to produce a full set of isoacceptors that are required to provide the translational machinery with all 20 amino acids required for polypeptide synthesis. Importantly, tRNAs retaining the intervening intron sequences are incapable of adopting an appropriate mature tRNA fold and are therefore defective in translation (Porter et al., 2021).

1.5. La proteins assist in correct folding and processing of nascent pre-tRNA through two independent binding modes

The La protein, also known as Sjogren′s Syndrome antigen-B (SS-B), was first identified as an autoantigen in patients suffering from systemic lupus erythematosus and Sjogren′s syndrome (Alspaugh & Tan, 1975; Mattioli & Reichlin, 1974). Patient antibodies were used for immunoprecipitation and revealed that La is an RNA-binding protein (Hendrick et al., 1981; Lerner et al., 1981) that in particular interacts with nascent polymerase III transcripts still containing the typical 3′-terminal uridylate sequence (encoded in the genome as an adenylate stretch at the end of the RNA polymerase III gene) (Mathews & Francoeur, 1984; Stefano, 1984). Homologues of the La protein have been identified in many eukaryotes, ranging from yeasts (Lhp1p in *S. cerevisiae* and Sla1p in *S. pombe*) (Van Horn et al., 1997; Yoo & Wolin, 1994) to metazoans (Arhin et al., 2005; Park et al., 2006; Pellizzoni et al., 1996; Scherly et al., 1993) and from insects to higher plants (Fleurdépine et al., 2007), where La is essential in metazoans and higher plants but dispensable in yeast (Fleurdépine et al., 2007; Park et al., 2006; Van Horn et al., 1997; Yoo & Wolin, 1994).

The best conserved sequence of the genuine La protein is found in the N-terminal domain, whereas the C-terminal domain is more variable between species. The conserved N-terminal domain contains the La module, an RNA-binding unit composed of two juxtaposed RNA-binding

domains, the La motif (LaM) and an RNA recognition motif (RRM) separated by a short linker (see **Figure 9** – top panel) (Teplova et al., 2006). In higher eukaryotes, the C-terminal domain of the La protein contains a second RRM (RRM2), a disordered region and a nuclear localization signal (**Figure 9**) (Jacks et al., 2003).

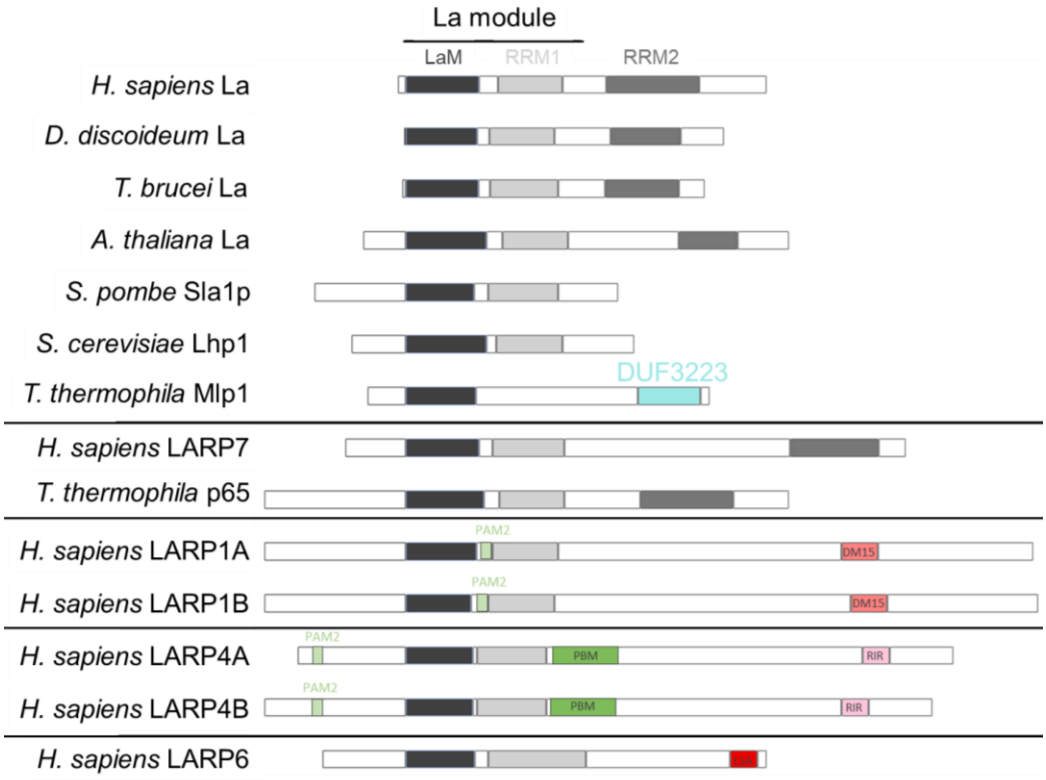


Figure 9. Schematic representation of conservation of the La module in genuine La and La-related protein family.

Domain organization of the genuine La proteins from *H. sapiens, D. discoideum, T. brucei, A. thaliana, S. pombe* (Sla1p), *S. cerevisiae* (Lhp1p) and *T. thermophila* (Mlp1) compared to LARP7 proteins from *H. sapiens* and *T. thermophila* (p65) and other La-related proteins (LARP) from *H. sapiens* (LARP1, LARP4 and LARP6). The La module is found in the N-terminal region and contains the LaM (shown in black) and the RNA recognition motif-1 (RRM1) (shown in light grey). A second RRM is found in the C-terminal region of genuine La proteins in higher eukaryotes and LARP7 proteins (shown in dark grey). LARP-1, -4, and -6 have obtained additional domains with specialized functions, including DM15: domain for binding of the 5′-terminal oligopyrimidine (TOP) containing mRNAs; PAM2: PABP-interacting motif-2; PBM: PABP binding motif; RIR: RACK1- interaction region; LSA: LaM and S1-associated motif.

1.5.1. Uridylate-specific binding of La results in 3′-end protection

For sequence-specific binding of the 3′-terminal uridylate sequence both RNA-binding domains in the La module (LaM + RRM1) are involved in formation of a binding pocket at the interdomain interface around these uridylates, without using the canonical nucleic acid interaction surfaces of the La module: the winged helix-face of the LaM and the antiparallel β-sheet surface on RRM1 (**Figure 10a**) (Kotik-Kogan et al., 2008; Teplova et al., 2006). Functional studies have demonstrated that the La module is indispensable and sufficient for binding (Goodier et al., 1997; Ohndorf et al., 2001), and structural studies confirm these findings and reveal that the La module forms a hydrophobic binding pocket at the interdomain interface of the LaM and RRM1 around the uridylate sequence (**Figure 10a**) (Kotik-Kogan et al., 2008; Teplova et al., 2006). Important residues lining the hydrophobic pocket that interact with the uridylate RNA using their side chains are Q20, Y23, Y24, D33, F35, and F55 and two additional LaM residues, N56 and R57, make contact with the uridylate RNA only by using their amide backbone structure (**Figure 10b** – U_{-1} and U_{-2} interactions). In contrast to the many interactions made by the LaM, RRM1 only makes a single contact with the uridylate RNA using the amide backbone of I140 (**Figure 10b** – U_{-2} interactions). In addition to contacts between RNA and amino acid residues, some interdomain interactions between the LaM and RRM1 have also been observed (**Figure 10b** – LaM-RRM1 interactions). Through binding of the La module (LaM + RRM1) surrounding the 3′-uridylate RNA, the 3′-end of the transcript is protected from 3′-exonucleases (Yoo & Wolin, 1997).

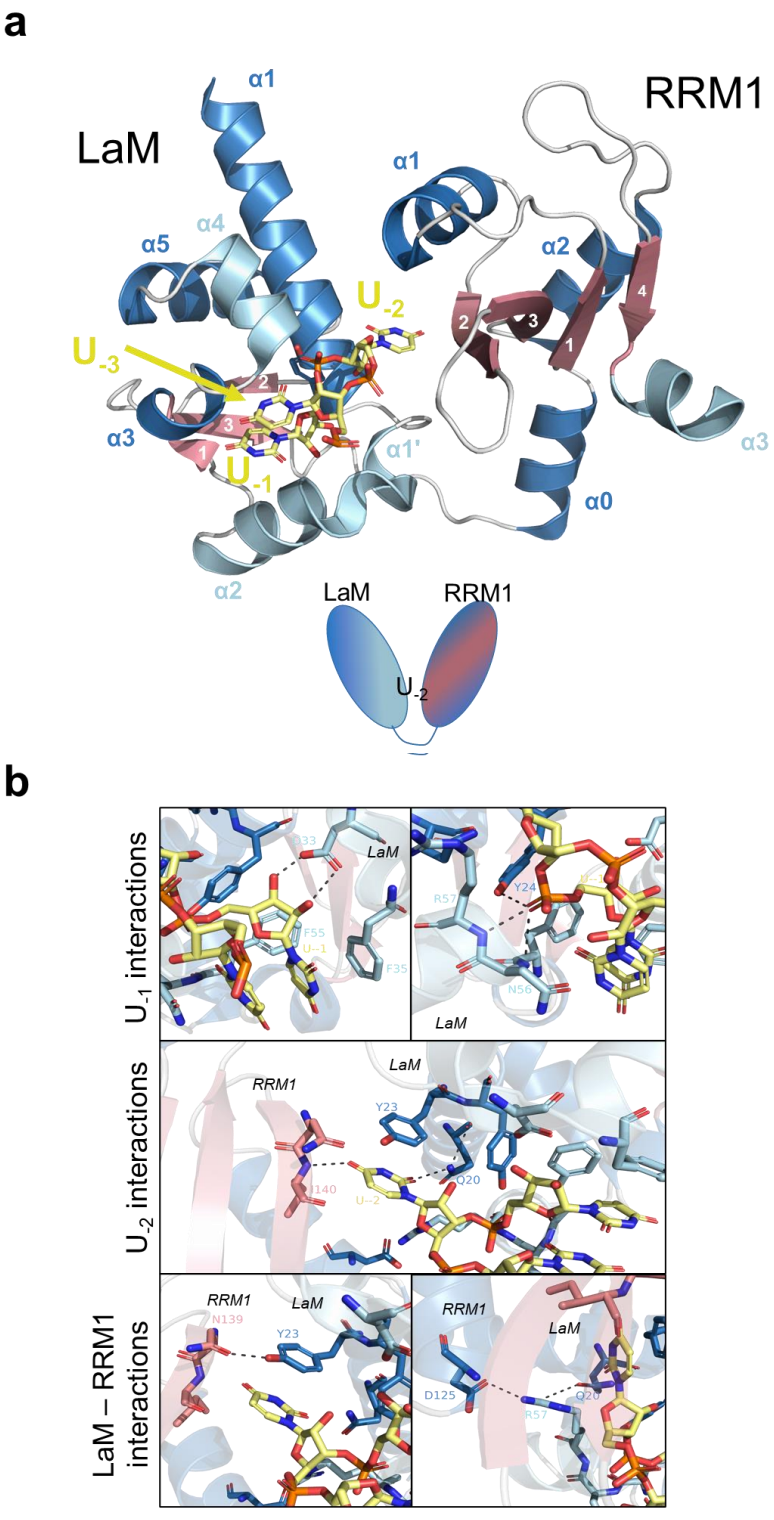


Figure legend on next page

Figure 10. The human La protein binds 3′-uridylate RNA through cooperative binding between the LaM and RRM1.

a Structure of the human La protein in complex with 5′-AUAUUUU-3′ (PDB: 2VOD). For clarity, only three uridylates at the 3′-end are shown and a negative number system was adopted with the most 3′-terminal uridylate named U_{-1}. The penultimate uridylate (U_{-2}) is found to be pointing towards the inside of the binding pocket and the 3′-terminal uridylate (U_{-1}) is pointing more towards the outside in a stacking formation with U_{-3}. Secondary structures are shown as ribbons with β-sheets colored in pink, the canonical winged-helix-fold and RRM domain α-helices shown in dark blue and La protein-specific α-helices show in light blue. RNA is shown as a stick model with carbon atoms colored in yellow. Other atoms are colored by type: oxygen in red, nitrogen in blue and phosphorous in orange. **b** Magnified views of La-U_{-1} RNA interactions: amino acids important for binding of the 3′-terminal uridylate U_{-1} are located exclusively in the LaM and interactions are strictly made to sugar-phosphate backbone features on the nucleotide (except for stacking interactions of the pyrimidine ring and the ribose with F35 and F55, respectively). D33 specifically interacts with 2′-OH and 3′-OH groups. Hydrogen bonds are formed with the phosphate group of U_{-1} by the sidechain of Y24 and both N56 and R57 use the amide backbone structures. Magnified view of La-U_{-2} RNA interactions: amino acids important for binding of U_{-2} are found in both the LaM and RRM1 and hydrogen-bonds are made with the nucleotide base of U_{-2}. Q20 makes a hydrogen bond with U_{-2} using its sidechain and I140 contacts U_{-2} with its amide backbone. The pyrimidine ring of U_{-2} is also stacked directly on Y23. Magnified view of interdomain interactions between the LaM and RRM1: interactions between amino acids are strictly formed with amino acid side chains. An interdomain contact exists between N139 in the RRM1 and Y23 in the LaM. Another interdomain interaction exists between RRM1 residue D125 and LaM residue R57, which in turn also interacts with Q20 through an intradomain interaction.

1.5.2. RNA chaperone activity of La assist in appropriate tRNA folding and maturation

The function of many non-coding RNA molecules, including tRNAs, often depends on correct folding into their precise functional conformations. RNA chaperones are proteins that can facilitate formation of functional conformations by resolving misfolds through unfolding and folding of an RNA by direct interactions (Lorsch, 2002). The La protein is an RNA chaperone that in addition to protecting the 3′-uridylate ends of pre-tRNAs from exonucleases, can also assist in RNA folding (Bayfield & Maraia, 2009; Hussain et al., 2013; Naeeni et al., 2012). The

typical RNA-binding sequences in the human La (hLa) protein RRM1 β-sheet have been found to make uridylate-independent contacts to the main body of pre-tRNAs and also possess RNA chaperone activity by assisting with tRNA folding and maturation (Bayfield & Maraia, 2009). Another contact between the hLa protein and pre-tRNA is made though interaction between the short basic motif or SBM (localized in the C-terminal part) and the 5′-triphosphate end of nascent pre-tRNA (Fan et al., 1998). The combined interaction of the hLa protein with the uridylate sequence on the 3′-end of nascent pre-tRNAs and additional binding to the body and 5′-end of the pre-tRNA helps the hLa protein identify precursor tRNAs rather than mature tRNA (Bayfield & Maraia, 2009). A proposed function of the 3′-uridylate dependent binding mode is to initiate binding via this sequence specific stretch of uridylates and thereby to recruit La-associated RNA chaperone function to these transcripts (Vakiloroayaei et al., 2017).

For example, hLa has been characterized as an RNA chaperone through an *in vitro* assay where La assists in folding of a misfolded, self-splicing intron into a functional form (Bayfield & Maraia, 2009; Belisova et al., 2005). Many *in vivo* studies have been completed in yeasts because the La protein is non-essential in these systems. However, deletion of the La protein in both the budding yeast *S. cerevisiae* (Lhp1p) (Yoo & Wolin, 1994) and fission yeast *S. pombe* (Sla1p) (Van Horn et al., 1997) has a lethal phenotype in combination with either loss of function of certain tRNA modification enzymes (Copela et al., 2006) or with a mutation in the anticodon stem of the essential tRNA Ser^{CGA}, the only tRNA isoacceptor for UCG codons in *S. cerevisiae* (Yoo & Wolin, 1997). The lethal phenotype could be rescued by a compensatory mutation that restored base pairing in the anticodon stem, indicating that Lhp1p is essential after introduction of mutations that would cause misfolding of the pre-tRNA (Yoo & Wolin, 1997).

These observations provide the basis for the tRNA mediated suppression assay in *S. pombe* where RNA chaperone function of La proteins from different species can be tested. In this

red/white assay, La proteins can be transformed into *S. pombe* ySH9 strains lacking its genuine La homolog (*sla1*-Δ) and containing a suppressor tRNA Ser^{UCA} and the *ade6-704* mRNA (**Figure 11**). When the suppressor tRNA Ser^{UCA} is functional, as a result of appropriate folding and maturation by RNA chaperone function of transformed La, translation of the full-length ADE6 enzyme can take place and as a result leads to alleviation of the accumulation of the red intermediate (**Figure 11**). In other words, yeast colonies have a white color in the presence of functional La chaperone activity and a red color in the absence thereof. The tRNA mediated suppression assay has been used previously with La proteins from different species and RNA chaperone activity has been established for all La proteins analysed (Porat & Bayfield, 2020).

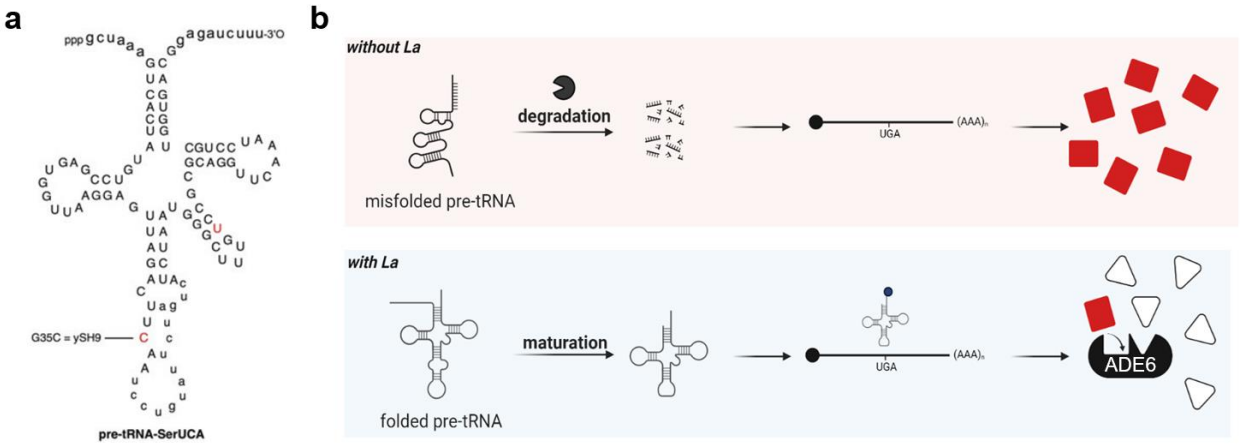


Figure legend on next page

Figure 11. tRNA-mediated suppression assay to study RNA chaperone function.
a Representation of the suppressor tRNA Ser^{UCA} with lower case letters indicating nucleotides that are going to be removed during pre-tRNA processing (5′- and 3′-end maturation and intron removal). Mutated nucleotides compared to the wild type tRNA Ser^{UGA} are indicated in red. The anticodon mutation enables the tRNA to decode a UGA stop codon by inserting a serine into the growing peptide chain, but it also results in a mismatch between the anticodon and intron sequence resulting in misfolding of the pre-tRNA and as a result degradation. Another mutation in the body of the pre-tRNA can be introduced to increase the dependency on RNA chaperone activity for correct folding of the suppressor pre-tRNA. **b** The ade6-704 mRNA contains an early stop codon UGA that can be suppressed by a functional suppressor tRNA Ser^{UCA} to produce the full-length ADE6 enzyme that has a role in adenine biosynthesis. In the presence of an RNA chaperone (in this example the La protein was transformed into *S. pombe* ySH9) the suppressor tRNA Ser^{UCA} is functional and a full-length ADE6 enzyme is produced. As a result, a red intermediate metabolite (formed in the adenine synthesis pathway) can be converted into colorless products. In contrast, when RNA chaperone function is absent, the suppressor tRNA is not functional and no ADE6 enzyme is produced leading to accumulation of a red intermediate metabolite. Figure adapted from (Porat & Bayfield, 2020) (**a**) and figure generated using BioRender (**b**).

1.6. La and La-related proteins conserved structural arrangements

The La protein is a member of the La-related proteins (LARPs), a superfamily of eukaryotic RNA-binding proteins that contain five different families: LARP1, LARP3 (genuine La), LARP4, LARP6 and LARP7 (Bousquet-Antonelli & Deragon, 2009). A common feature of the LARP family is the La module in the N-terminal part of the proteins consisting of two RNA-binding domains: the LaM and the RRM1 (see **Figure 9**) (Alfano et al., 2004; Dong et al., 2004; Kenan & Keene, 2004). While LARPs typically have a high degree of homology with the genuine La protein in the LaM-RRM arrangement, many LARPs have adopted additional functions by sequence and structure variations within the La module and by acquiring additional domains with specialized functions (Dock-Bregeon et al., 2019). Within the LARP group, LARP7 proteins are most closely related to La proteins when comparing sequence identity and function. However, in contrast to hLa that binds all nascent RNA polymerase III transcripts on their 3′-terminal uridylate tail, human LARP7 interacts almost exclusively with the non-coding RNA 7SK

(He et al., 2008; Krueger et al., 2008). The 7SK RNA is an RNA polymerase III transcript and is initially bound by the genuine La protein (Chambers et al., 1983). During 7SK RNP assembly, the La protein is replaced by LARP7 which recognizes the 3′-terminal uridylate motif of 7SK by formation of a hydrophobic binding pocket around the uridylate sequence using the LaM and RRM1 (**Figure 12b**, left) (He et al., 2008; Muniz et al., 2013; Uchikawa et al., 2015), very similar to the binding pocket formed by La upon binding of uridylate RNAs (**Figure 12a,** left) (Kotik-Kogan et al., 2008; Teplova et al., 2006). Immunoprecipitation of the La protein co-purifies the methylphosphate capping enzyme or MePCE responsible for addition of the γ-monomethyl cap on 7SK and U6 snRNA (see **Figure 3**) (Muniz et al., 2013). Transcripts containing the modified γ-monomethyl cap instead of the nascent RNA polymerase III 5′-triphosphate cap have lower binding affinity to the hLa protein indicating that the post-transcriptional modification of the cap is required for replacement of the La protein (Bhattacharya et al., 2018).

In addition to 3′-uridylate interactions, LARP7 also makes specific contacts with the 3′-hairpin of 7SK by using the α3 helix on RRM2 (**Figure 12b**, right) (Eichhorn et al., 2016). In *Tetrahymena thermophila*, the LARP7 ortholog p65 was until recently the only characterized LARP in this species (Bousquet-Antonelli & Deragon, 2009; Deragon, 2020). p65 can interact with telomerase RNA in a similar way as human LARP7 interactions with 7SK RNA: the disordered region C-terminal of RRM2 α3 forms an α-helical extension upon binding of stem IV of telomerase RNA (**Figure 12c**). Telomerase RNA is transcribed by RNA polymerase III in this species and the La module has also been proposed to interact with the 3′-uridylate motif of telomerase RNA similarly as hLa and LARP7 (Bayfield et al., 2010; Cash & Feigon, 2017; O′Connor & Collins, 2006; Prathapam et al., 2005; Singh et al., 2012). The related p43 protein found in the ciliated protozoa Euplotes is similarly critical for the assembly, nuclear retention and functionality of the telomerase complex (Aigner et al., 2003; Aigner & Cech, 2004; Möllenbeck et al., 2003). The LARP7 ortholog in fission yeast, Pof8, has also been described as a constitutive

component of the telomerase complex (Páez-Moscoso et al., 2018). In conclusion, there are structural similarities and related RNA binding modes between p65, LARP7 and La, but these factors nevertheless recognize distinct RNA substrates.

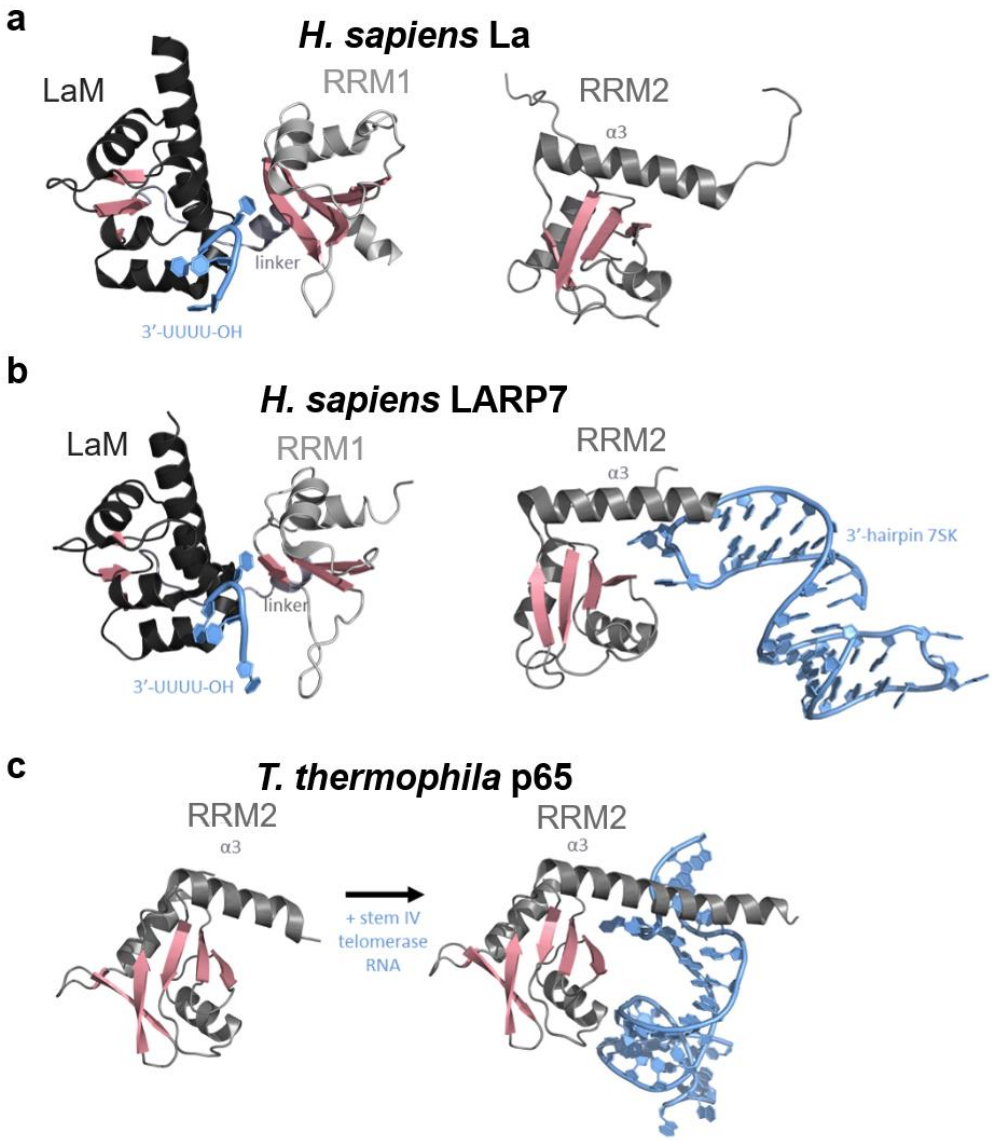


Figure legend on next page

Figure 12. Structural conservation between different member of the La-related protein family.
a Structures of the three different RNA-binding domains from the genuine La protein in *H. sapiens*. The LaM and RRM1 bind uridylate RNA inside the hydrophobic pocket formed between the two domains (PDB: 2VOD) and the structure for RRM2 (PDB: 1OWX). **b** Structures of the three different RNA-binding domains from the LARP7 protein in *H. sapiens.* Similar binding of uridylate RNA by the LaM and RRM1 forming a hydrophobic pocket around the RNA (PDB: 4WKR) and the RRM2 binding the 3′-terminal hairpin of the non-coding 7SK RNA (PDB: 6D12). **c** Structures of the RRM2 from p65 in *T. thermophila*. The C-terminal disordered region of α3 forms a helical extension upon binding of the non-coding telomerase RNA (PDB: 4EYT, 4ERD). LARP: La-related protein.

1.7. A novel type of La protein identified in alveolates lacking an RRM domain and containing a C-terminal DUF3223

A recent study has identified another protein from the LARP family in alveolates such as *T. thermophila,* and based on the primary sequence conservation of the LaM it has been grouped with the genuine La proteins (Deragon, 2020) and will be referred to as Macronucleus localized protein of unknown function (Mlp1) (Arthur et al., 2017). Interestingly, Mlp1 only contains a highly conserved LaM, unlike previously identified La proteins in other eukaryotes containing the typical LaM-RRM arrangement (**Figure 13a**). La proteins are typically expressed as single copies, including the LaM-only La protein in alveolates (**Figure 13b**). This atypical La protein has been identified in several members of the alveolates and is the first eukaryotic group in which a genuine La protein has been described that contains a LaM without an adjacent RRM1 (**Figure 13c**) (Deragon, 2020). While this is very unusual for genuine La proteins, some proteins in the LARP superfamily are known to only contain a LaM. For example, LARP1 orthologs in the fungi *S. pombe* (Slr1p) and *S. cerevisiae* (Sro9p and Slf1p) are RNA-binding proteins that contain a LaM without an adjacent RRM (Bousquet-Antonelli & Deragon, 2009; Sobel & Wolin, 1999). More recently, the human LARP1 was also described to contain a conserved LaM without presence of the adjacent RRM1 (Kozlov et al., 2022). However, the majority of

previously identified LARPs (LARP1, LARP4, LARP6 and LARP7) contain both the LaM and RRM1 RNA-binding domain (Bousquet-Antonelli & Deragon, 2009).

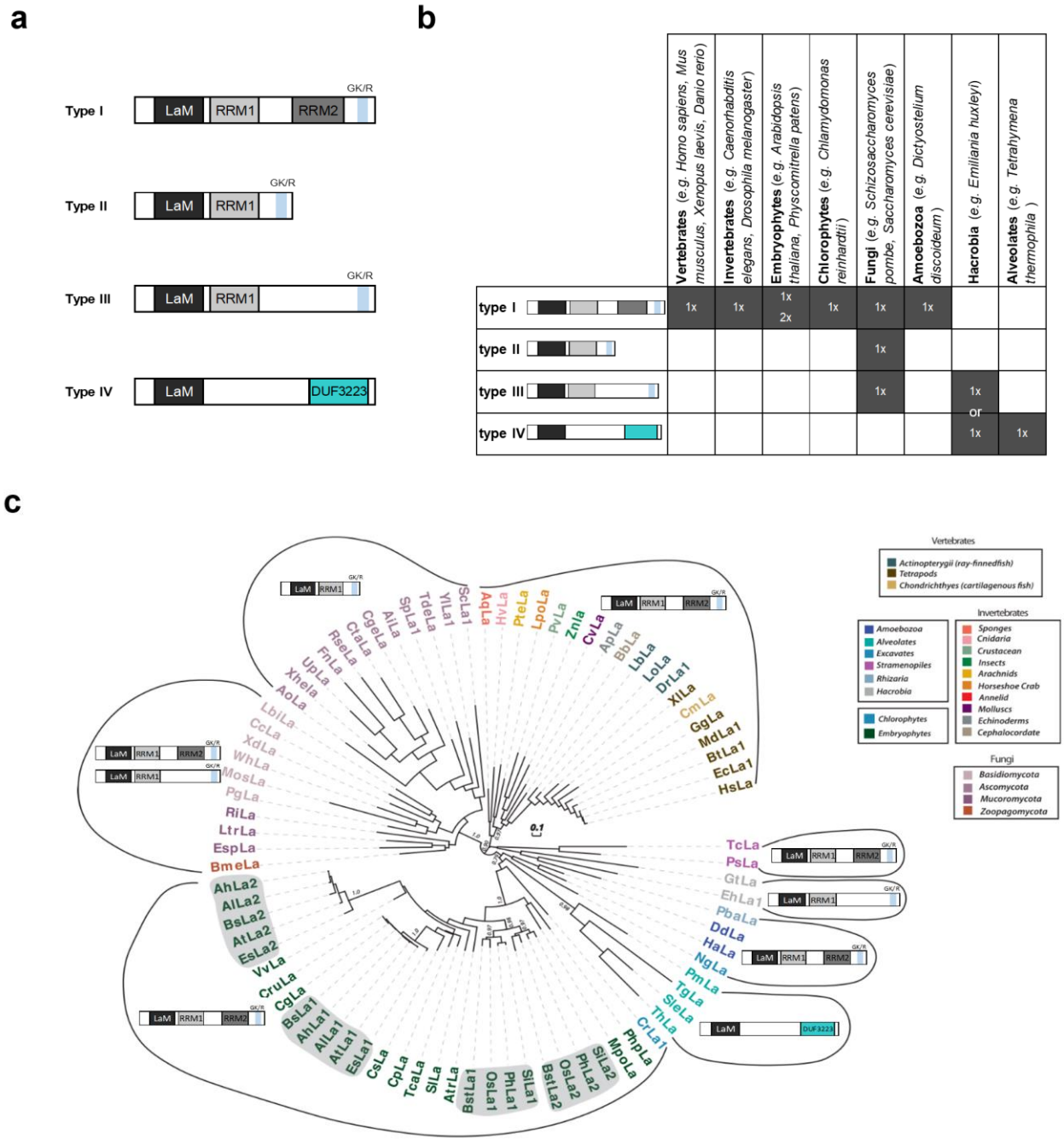


Figure legend on next page

Figure 13. Structural organization of the La protein in eukaryotes.

a The different types of genuine La proteins present in eukaryotic species. LaM: La motif, RRM1: RNA recognition motif-1, RRM2: RNA recognition motif-2, DUF3223: domain of unknown function-3223, GK/R: glycine, lysine/arginine-rich motif. **b** Overview of La proteins typically found in eukaryotic lineages. Numbering is used to indicate the number of genuine La proteins encoded in the genome. **c** Phylogenetic tree constructed using full-length La protein sequences (except for Alveolates, where the LaM alone was used). La proteins corresponding to different eukaryotic groups are color-coded corresponding to the legend on the right. Figure adapted from (Deragon, 2020)

The absence of the RRM1 in a genuine La protein is highly unexpected because in La proteins described in other eukaryotes, both RNA-binding domains in the La module (LaM + RRM1) are required for high-affinity binding of the typical uridylate RNA (see **Figure 10**). In other words, deletion of either the LaM or RRM1 results in loss of binding of uridylate RNA (Dong et al., 2004; Goodier et al., 1997). Another unusual feature of this new type of La protein is the presence of a domain of unknown function-3223 (DUF3223) in the C-terminal region. The DUF3223 has also been identified in the C-terminal region of plant-specific RNA polymerases IV/V subunits (NRPD1 and NRPE1) (Wang & Ma, 2015) and in plant proteins involved in rRNA processing (DOMINO1, DeCL and DeCL-like) (Bellaoui et al., 2003; Lahmy et al., 2004). The DUF3223 is also found many in bacteria and protists but is not found is in animals and fungi (Wang & Ma, 2015).

1.8. Statement of purpose

The discovery that formed the basis of this dissertation was the identification of a novel LaM-containing and RRM1-lacking La protein in alveolates that challenges the expectation of the structural composition of genuine La proteins. La proteins are found in almost all eukaryotes studied and are well-known to function in processing of RNA polymerase III transcripts through sequence-specific binding of 3'-terminal uridylates (Deragon, 2020; Mathews & Francoeur, 1984; Stefano, 1984). The well-described 3′-terminal uridylate high-affinity RNA binding site occurs through cooperative formation of a UUU-3′OH specific binding pocket between both RNA-binding domains found in the La module, the LaM and RRM1 (Kotik-Kogan et al., 2008; Teplova et al., 2006). We posed the question if a La protein that is structurally missing an important conserved RNA binding domain could still function as a genuine La protein in pre-tRNA binding and processing.

In Chapter II, we confirm that Mlp1 indeed functions as a genuine La protein by demonstrating uridylate-dependent preferential binding of pre-tRNAs through highly conserved residues located in the LaM and RNA chaperone activity promoting processing and maturation of misfolded pre-tRNAs using the pre-tRNA mediated suppression assay. In addition, we study tRNA processing in the yeast *S. pombe* and in *T. thermophila* and discover that Mlp1 is linked to alternate tRNA processing compared to other eukaryotes. We also describe that the architecture of pre-tRNAs in *T. thermophila* is variable from other eukaryotes, in that the 3′-trailer sequences are absent and the mature tRNA sequence immediately continues into the 3′-uridylate stretch. The discovery of an altered pre-tRNA processing pathway in the highly studied eukaryote *T. thermophila,* driven by a LaM-containing and RRM1-lacking La protein, contributed to both the pre-tRNA processing and the La protein field.

In addition to lacking the RRM1 RNA-binding domain, Mlp1 also contains a C-terminal DUF3223. We demonstrate that this domain possesses RNA chaperone activity by tRNA mediated suppression assay in Chapter II. We considered that this novel La protein containing two additional domains, the middle domain and C-terminal DUF3223, could have additional functions. In Chapter III, we further investigate the function of this variant La protein by identifying other non-coding RNA binding partners and found that Mlp1 interacts with mature snRNAs, lacking the typical high affinity uridylate binding site. In addition, we identify direct protein interactions with a protein component of the functionally active snRNPs, consistent with a function for Mlp1 in splicing. We confirm that Mlp1 indeed affects the efficiency of splicing of introns from pre-mRNAs. This is the first study describing the involvement of a genuine La protein in pre-mRNA splicing. In Chapter IV, we also find that Mlp1 interacts with all previously annotated snoRNAs and protein components of the snoRNP complex, which function as guide RNAs in post-transcriptional modification of non-coding RNAs. We use this as a metric to predict novel snoRNAs in *Tetrahymena thermophila*. Identification of all snoRNAs would contribute to a more complete understanding of post-transcriptional modifications of non-coding RNAs in *T. thermophila*.

Overall, this work demonstrates new roles for the variant genuine La protein Mlp1 in the biogenesis of non-coding RNAs in *T. thermophila*.

Chapter II:

Altered tRNA processing is linked to a distinct and unusual La protein in *Tetrahymena thermophila*

Author contributions

Altered tRNA processing is linked to a distinct and unusual La protein in *Tetrahymena thermophila*

Kerkhofs, Kyra[1]; Garg, Jyoti[1]; Fafard-Couture, Étienne[2]; Abou Elela, Sherif[3]; Scott, Michelle S.[2]; Pearlman, Ronald E.[1]; Bayfield, Mark A.[1]

[1] Department of Biology, Faculty of Science, York University, Toronto, Ontario N3J 1P3, Canada,
[2] Département de Biochimie et de Génomique Fonctionnelle (LGDP), Faculté de Médecine et des Sciences de la Santé, Université de Sherbrooke, Sherbrooke, Québec, J1E 4 K8, Canada,
[3] Département de Microbiologie et d'Infectiologie, Faculté de Médecine et des Sciences de la Santé, Université de Sherbrooke, Sherbrooke, Québec, J1E 4 K8, Canada.

Manuscript published in

Nature Communications

2022 October 10th,

Author contributions:

Jyoti Garg transformed, selected, and characterized the partial Mlp1 knockout (KO) strain in *Tetrahymena thermophila* by Southern blotting in **Supplementary Figure 7b** (original manuscript: Supplementary Figure S7b) and performed previous versions of immunofluorescent staining in **Supplementary Figure 7c** (original manuscript: Supplementary Figure S7c). Étienne Fafard-Couture performed Thermostable Group II Intron Reverse Transcriptase (TGIRT)-sequencing analysis generating transcripts per million (TPM) counts and generated a custom python script to obtain counts per million (CPM) for pre-tRNAs and mature tRNAs. These data were used to generate **Figure 15b** (original manuscript: Figure 2b) and **Figure 18c** (original manuscript: Figure 5c).

2.1. Abstract

Nascent pre-tRNAs are transcribed by RNA polymerase III and immediately bound by La proteins on the UUU-3'OH sequence, using a tandem arrangement of the La motif and an adjacent RNA recognition motif-1 (RRM1), resulting in protection from 3'-exonucleases and promotion of pre-tRNA folding. The *Tetrahymena thermophila* protein Mlp1 has been previously classified as a genuine La protein, despite the predicted absence of the RRM1. We find that Mlp1 functions as a La protein through binding of pre-tRNAs, and affects pre-tRNA processing in *Tetrahymena thermophila* and when expressed in fission yeast. However, unlike in other examined eukaryotes, depletion of Mlp1 results in 3'-trailer stabilization. The 3'-trailers in *Tetrahymena thermophila* are uniquely short relative to other examined eukaryotes, and 5'-leaders have evolved to disfavour pre-tRNA leader/trailer pairing. Our data indicate that this variant Mlp1 architecture is linked to an altered, novel mechanism of tRNA processing in *Tetrahymena thermophila*.

2.2. Introduction

RNA polymerase III transcription of nascent pre-tRNAs terminates after the synthesis of a stretch of uridylates on a 3′-trailer extension (UUU-3′OH). La is the first protein to bind these nascent pre-tRNAs on the uridylate stretch (Fairley et al., 2005) and assists with pre-tRNA folding (Bayfield & Maraia, 2009; Chakshusmathi et al., 2003; Copela et al., 2006) in a manner that can protect pre-tRNAs from 3′- to 5′exonuclease digestion (Copela et al., 2008) and rescue defective pre-tRNAs from nuclear surveillance (Huang et al., 2006). Once the nascent pre-tRNA has acquired a tRNA-like structure, the endonuclease RNase P removes the 5′-leader of La-bound pre-tRNA, followed by endonucleolytic cleavage of the 3′-trailer by RNase Z (Van Horn et al., 1997; Yoo & Wolin, 1997). As a result of 3′-end cleavage, the La protein no longer associates with the tRNA and can be recycled for processing of new nascent pre-tRNAs (Bayfield & Maraia, 2009). In addition to the La-dependent pathway, an alternative La-independent pre-tRNA processing pathway exists but the order of pre-tRNA processing is reversed: without 3′-trailer protection by La, the exonuclease Rex1 digests the 3′-trailer of the pre-tRNA rapidly before RNase P endonucleolytic cleavage of the 5′-leader (Copela et al., 2008), and misfolded pre-tRNAs are also more prone to degradation through nuclear surveillance (Huang et al., 2006). La binding to pre-tRNAs is therefore hypothesized to establish and determine the order of 5′-leader versus 3′-trailer processing (Yoo & Wolin, 1997). Once the 5′-leader and 3′-trailer sequences are processed, tRNA nucleotidyltransferase adds a CCA sequence to the discriminator base at the 3′-end of the mature tRNA which will serve as the site of amino acid charging (Tomita & Yamashita, 2014).

Genuine La proteins are members of the La-related proteins (LARPs) (Bousquet-Antonelli & Deragon, 2009). Most LARP family members, and all genuine La proteins, contain an N-terminal La module consisting of two adjacent RNA-binding domains: a La motif (LaM) and an RNA

recognition motif-1 (RRM1) (**Figure 14a**) (Alfano et al., 2004; Dong et al., 2004; Kenan & Keene, 2004). In *Tetrahymena thermophila,* the LARP7 ortholog p65 was until recently the only characterized LARP in this species (Bousquet-Antonelli & Deragon, 2009; Deragon, 2020), but a recent study has grouped the *T. thermophila* Macronucleus localized protein of unknown function (Mlp1) (Arthur et al., 2017) with the genuine La proteins, based on its primary sequence conservation in the LaM (Deragon, 2020). Interestingly, Mlp1 only contains a highly conserved LaM, unlike all previously studied La proteins which contain the tandem LaM-RRM arrangement. This atypical La protein has been identified in several members of the alveolates (Deragon, 2020). Utilizing the interesting and important model organism, the alveolate ciliate protozoan *T. thermophila* (Orias et al., 2011), the macronucleus/somatic nucleus localization and structural characteristics of TtMlp1 could provide insight into the observation that despite the apparent lack of the RRM1, Mlp1 functions as a genuine La protein in tRNA processing in the nucleus.

The absence of the RRM1 in the La module of a genuine La protein in alveolates is highly unexpected, due to the mechanism by which La proteins bind the uridylate containing 3′-trailer. Structural studies of the human La (hLa) LaM-RRM bound to UUU-3′OH revealed that both domains sandwich this RNA, and functional studies have demonstrated that the La module is indispensable and sufficient for uridylate binding (Alfano et al., 2004; Ohndorf et al., 2001), as deletion of either the LaM or RRM1 results in complete loss of binding (Alfano et al., 2004; Dong et al., 2004; Kotik-Kogan et al., 2008; Teplova et al., 2006). Another unusual feature of this new type of La protein is the presence of a domain of unknown function-3223 (DUF3223) in the C-terminal region (Deragon, 2020). Thus, questions remain as to whether Mlp1 functions as a genuine La protein, and if so, whether it uses a mode of RNA binding dissimilar to previously studied La proteins, with possible associated changes in how Mlp1 directs pre-tRNA processing.

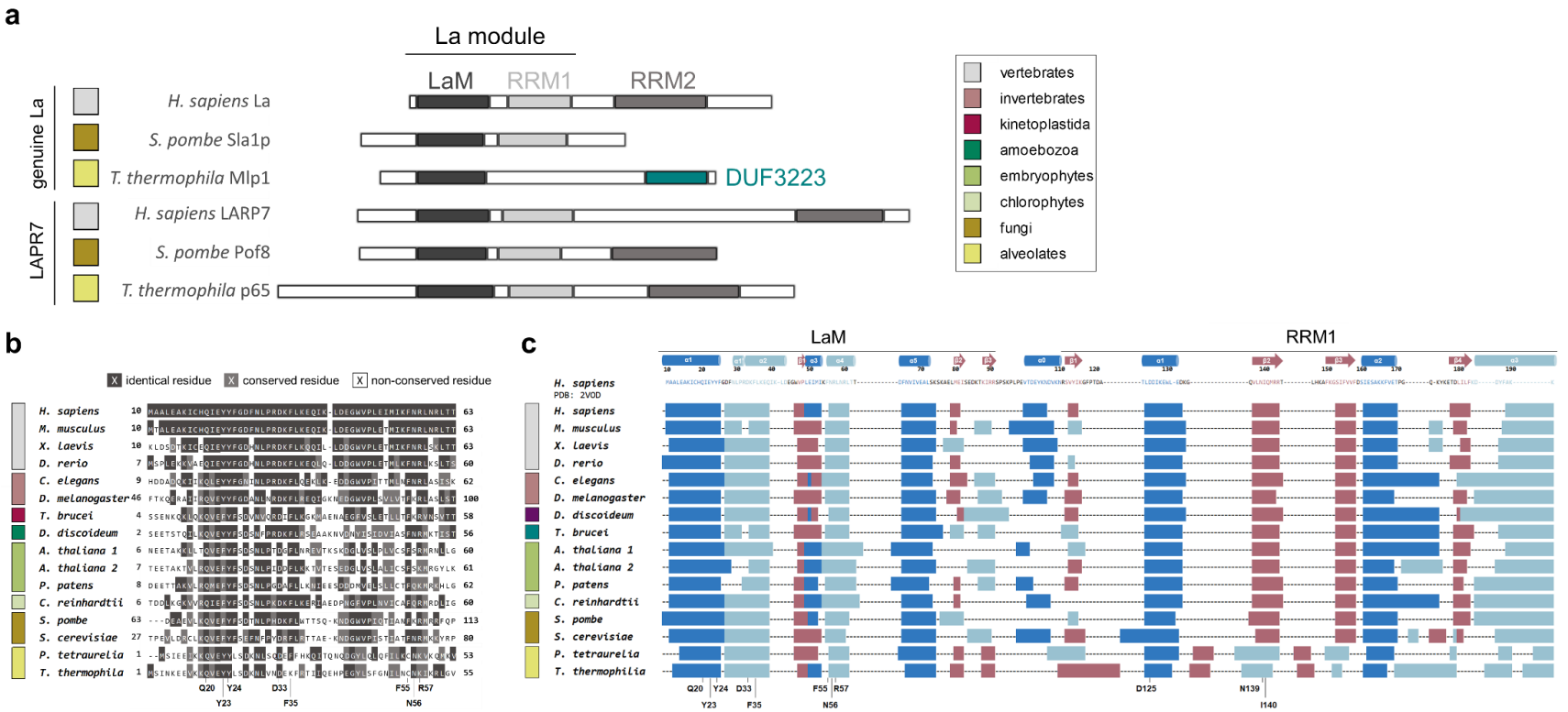


Figure 14. Alveolate La proteins are predicted to have a LaM but not an RRM1.

a Schematic representation of RNA-binding domains found in La and La-related protein-7 (LARP7) from different eukaryotes. The LaM and RRM1 together form the La module responsible for uridylate binding through formation of a hydrophobic binding pocket between the two domains. LaM: La motif, RRM: RNA recognition motif, DUF: Domain of Unknown Function. **b** Primary amino acid sequence alignments from different eukaryotic lineages showing conservation of uridylate binding residues (highlighted, bottom). A dark grey background indicates identical residues, light grey conserved residues and white indicates no conservation. Color coded legend is shown in **a**. Full LaM and RRM1 domains are shown in **Supplementary Figure 1a,b. c** Secondary structure predictions of LaM and RRM1 from different eukaryotic lineages. Predicted β-sheets shown in red and predicted α-helices in blue (dark blue: typical α-helices found in the winged-helix fold and classic RRM, light blue: inserted α-helices found in La proteins specifically) are compared against the secondary structure motif of the hLa protein crystal structure on the top (PDB: 2VOD). Location of the conserved amino acid residues important for uridylate binding are shown at the bottom. Color coded legend is shown in **a**.

Here, we present evidence that despite the apparent lack of the RRM1, Mlp1 functions as a genuine La protein. Using ribonucleoprotein (RNP) immunoprecipitation (RIP)-Seq of Mlp1, we show association with UUU-3′OH containing pre-tRNAs *in vivo*, and preferential binding of pre-tRNA substrates over mature tRNA substrates *in vitro*. In order to assess whether this altered architecture might have associated consequences in pre-tRNA processing, we tested Mlp1 function in a well-established model system and demonstrated that heterologous Mlp1 expression in *Schizosaccharomyces pombe* promotes pre-tRNA processing and tRNA mediated suppression, but without typical La-associated 3′-end protection. Consistent with this, genetic depletion of Mlp1 in *T. thermophila* reveals that in contrast to previously studied La proteins, Mlp1 destabilizes pre-tRNA 3′-ends, thus acting as a factor that accelerates 3′-end processing. In addition, when comparing pre-tRNAs found in *T. thermophila* with other eukaryotic species, we found that the 3′-trailer sequences are considerably shorter, and that 5′-leader sequences have evolved accordingly to disfavor base pairing at these shortened 3′-trailers. Our data are consistent with a model in which Mlp1 fulfills many expected functions of a genuine La protein but uses alternate RNA binding modes to promote a pre-tRNA processing pathway in *T. thermophila* that differs from those found in other studied eukaryotic systems.

2.3. Results

2.3.1. Mlp1 is predicted to lack an RRM adjacent to the LaM

A recent phylogenetic study in eukaryotes of all kingdoms identified Mlp1 as a new atypical genuine La protein in alveolates, containing a highly conserved LaM without an adjacent RRM (Deragon, 2020). To investigate primary amino acid sequence conservation, multiple sequence alignments were conducted for La proteins from different eukaryotic lineages. We confirm that residues important for uridylate binding in the LaM are mostly conserved in *T. thermophila* (**Figure 14b, Supplementary Figure 1,c-d**). In contrast, we found little to no conservation of residues in the region of the adjacent domain where the RRM1 would be located (**Supplementary Figure 1b**) (Kelley et al., 2015). When comparing primary sequences of the La module between different LARPs, the presence of an RRM1 domain could often only be inferred from secondary structure predictions (Bousquet-Antonelli & Deragon, 2009). Therefore, we compared the secondary structure predictions for *T. thermophila* La with secondary structures for different eukaryotes (**Figure 14c**). These consistently predicted an RRM fold immediately C-terminal to the LaM in all species examined, except for candidate alveolate La proteins.

2.3.2. Mlp1 preferentially binds pre-tRNAs in *Tetrahymena thermophila*

We hypothesized that should Mlp1 function as a genuine La protein, it should bind and promote the processing of RNA polymerase III transcripts, as has been demonstrated in budding and fission yeast (Van Horn et al., 1997; Yoo & Wolin, 1997). We immunoprecipitated Mlp1-associated RNAs followed by detection by Northern blot (**Figure 15a, Supplementary Figure 2a**). Using probes specific for pre-tRNA 3′-extensions, we confirmed that Mlp1 immunoprecipitation enriched pre-tRNA species relative to mature tRNAs, which were relatively

more abundant in input RNA fractions. We also observed lesser enrichment of the U5 small nuclear RNA (snRNA), which is consistent with data indicating association of La proteins with processing intermediates of U5 in *Saccharomyces cerevisiae* (Xue et al., 2000). We conclude that based on its expected RNA target cohort, Mlp1 behaves as expected for a genuine La protein.

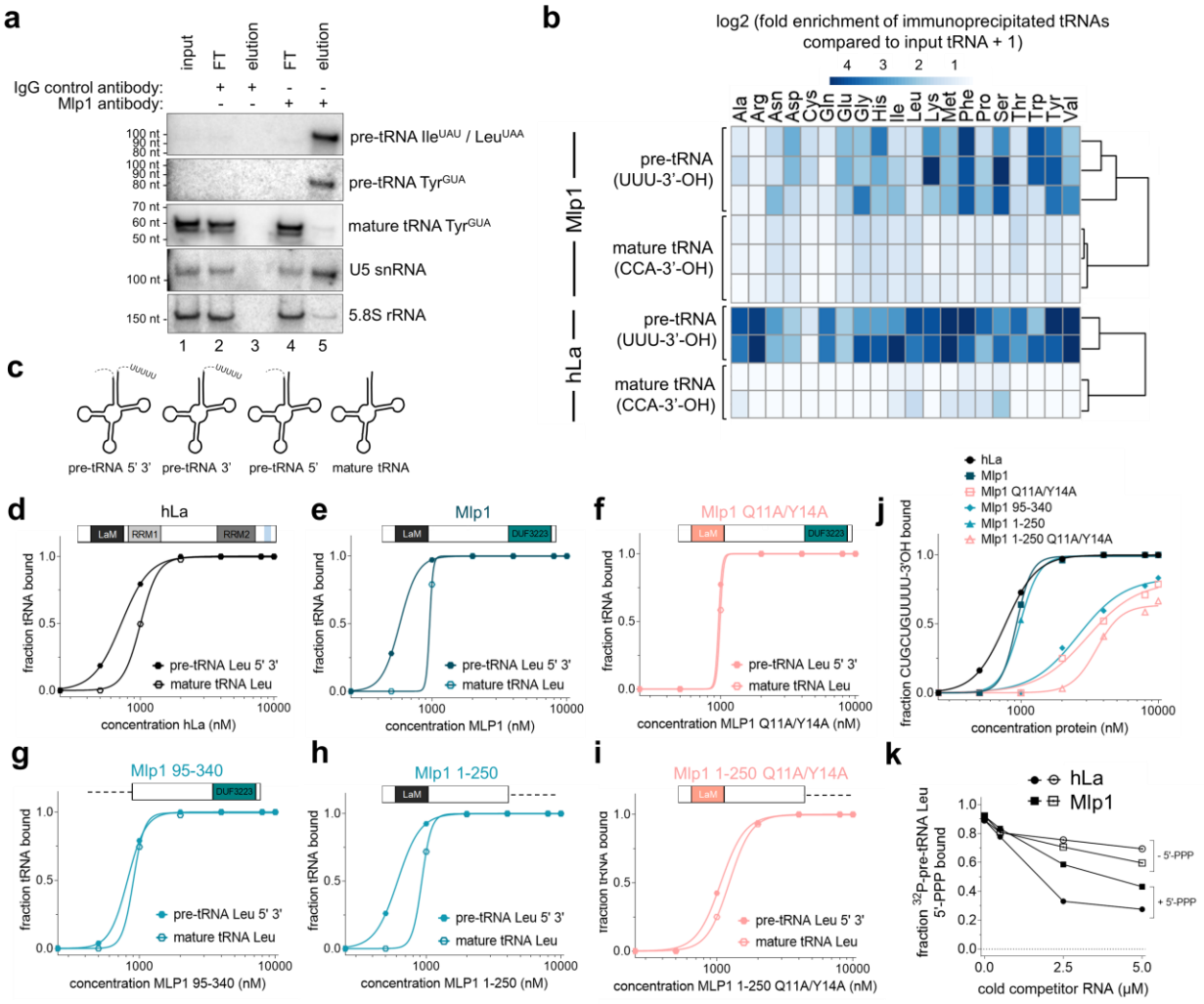


Figure legend on next page

Figure 15. Mlp1 preferentially binds pre-tRNAs partially via the 3′-terminal uridylates.

a Northern blot analysis of Mlp1 ribonucleoprotein-immunoprecipitated (RIP) samples from *T. thermophila*. Mlp1 enriches pre-tRNAs more efficiently than mature tRNAs (n=3 biologically independent samples). Pre-tRNA Ile^{UAU} and Leu^{UAA} are both recognized by the 3′-trailer probe used. 5.8S rRNA was used as a non-binding loading control. FT: flowthrough, IgG: immunoglobin G control antibody. Western blot confirming Mlp1-specific immunoprecipitation is shown in **Supplementary Figure 2a**. **b** Next generation sequencing data of immunoprecipitated pre-tRNAs (rows represent biologically independent replicates), shown as a heatmap of log2+1 transformed fold enrichment calculated by taking ratios of normalized Mlp1-immunoprecipitated tRNA and input tRNA counts per million (CPM) for different tRNA isotypes (top). Next generation tRNA sequencing data from Gogakos et al*.* 2017 analyzed similarly for hLa (bottom). The log2+1 transformed fold enrichment was calculated by taking ratios of normalized La-immunoprecipitated tRNA and the averaged input tRNAs CPM. The data were split between premature and mature tRNAs based on the 3′-end of the transcript (-CCA ending mature tRNAs and U-ending pre-tRNAs). Reproducibility between replicates for *T. thermophila* was determined by correlation plots shown in **Supplementary Figure 3**. **c** Schematic representation of ^{32}P-labeled *T. thermophila* pre-tRNA intermediates containing 5′-leader and 3′-trailer, 5′-leader only, 3′-trailer only and mature tRNA used in **d-i,k**, **Supplementary Figure 2d,e** and **Figure 16a,b**. **d-i** Binding curves from EMSAs comparing ^{32}P-labeled Leu^{AAG} 5′-leader, 3′-trailer containing pre-tRNA and mature Leu^{AAG} tRNA (n≥2 independent experiments). Native gels are shown in **Supplementary Figure 4a** and Kd quantification in **Table 1**. **j** Binding curves comparing binding of hLa, Mlp1 and Mlp1 mutants to ^{32}P-labeled CUGCUGUUUU-3′OH RNA (n≥2 independent experiments). Native gels are shown in **Supplementary Figure 4a** and Kd quantification in **Table 1**. **k** Binding curves from competition EMSA between ^{32}P-labeled 5′-leader containing pre-tRNA containing a 5′-triphosphate (+5′PPP) and unlabelled competitors +5′PPP as a positive control and 5′-triphosphate removed pre-tRNA (-5′PPP) (n=2 independent experiments). Native gels are shown in **Supplementary Figure 4d,e.** Source data are provided as a Source Data file.

To further investigate Mlp1-associated RNAs, we sequenced Mlp1-immunoprecipitated RNAs after removing RNAs > 300 nt by gel electrophoresis. As expected for a genuine La protein, we observed an enrichment of reads mapping to pre-tRNA genes carrying 3′-uridylate extensions relative to their normalized abundance in size-matched input samples (**Figure 15b** – Mlp1). In contrast, mature tRNAs are not enriched in Mlp1-bound samples (**Figure 15b** – Mlp1; individual values for tRNA isotypes presented in **Supplementary Figure 2b**). We compared our dataset with a previous study in which hLa-bound tRNA reads were obtained by photoactivatable

crosslinking and immunoprecipitation (PAR-CLIP) and where tRNA sequencing of input and immunoprecipitated RNA was done using hydro-tRNAseq (Gogakos et al., 2017) and found a similar enrichment for 3′-uridylate containing pre-tRNAs (**Figure 15b** – hLa; individual values for tRNA isotypes presented in **Supplementary Figure 2c**). The relative enrichment of pre-tRNAs over mature tRNAs, are strongly indicative of Mlp1 functioning as a genuine La protein in *T. thermophila*.

2.3.3. Mlp1 preferentially binds pre-tRNA substrates *in vitro*

The hLa protein preferentially binds pre-tRNA substrates through engagement of multiple sites, including the pre-tRNA body, 3′-trailer and 5′-leader (Bayfield & Maraia, 2009; Fan et al., 1998). To determine whether enrichment of Mlp1 associated pre-tRNAs relative to mature tRNAs correlated with changes in affinity for such ligands, we compared binding affinity of Mlp1 for *in vitro* transcribed mature tRNA versus pre-tRNA substrates using electrophoretic mobility shift assays (EMSAs), as well as versions of these that included either or both of 5′-leader or 3′-trailer extensions (**Figure 15c**). We found that Mlp1 preferentially binds pre-tRNAs containing both 5′- and 3′-extensions, followed by pre-tRNA containing the 3′-trailer. Lowest binding affinities were consistently found for mature tRNAs, confirming the *in vivo* preferential binding of pre-tRNAs (**Supplementary Figure 2d,e, Supplementary Table 1**).

To further investigate Mlp1 tRNA binding, we compared the affinity of full-length Mlp1 or Mlp1 mutants to hLa for radioactively labeled tRNA targets by EMSA (**Figure 15d-i, Supplementary Figure 4a, Table 1**). To test the importance of 3′-uridylate binding, we compared Mlp1 to an Mlp1 mutant in which conserved amino acids predicted to recognize the UUU-3′OH motif were substituted (Q11A/Y14A)(Dong et al., 2004; Teplova et al., 2006) (see **Figure 14b**, **Supplementary Figure 1d**, Q20/Y23 numbering in hLa), as well as a mutant in which the entire LaM was deleted (Mlp1 95-340). We found that the increased affinity for pre-tRNAs was lost in

the Mlp1 Q11A/Y14A mutant (compare **Figure 15e and f**) as well as for the LaM deletion mutant (compare **Figure 15e and g**). In contrast, the C-terminal DUF3223 is not important to discriminate between pre-tRNA and mature tRNAs since removal of this domain (Mlp1 1-250) maintains the binding affinity difference (**Figure 15h**), while the Q11A/Y14A mutations in the context of the deleted DUF3223 (Mlp1 1-250 Q11A/Y14A) again resulted in decreased affinity for pre-tRNAs (**Figure 15i**). To investigate UUU-3′OH binding more directly, we compared these Mlp1 mutants in protein-RNA binding experiments using a 3′-trailer sequence CUGUGUUUU-3′OH and found that the predicted uridylate binding residues located in the LaM (Q11A/Y14A) were required for binding (**Figure 15j, Supplementary Figure 4a, Table 1**). Additionally, binding of 3′-uridylate RNA occurs with the same affinity for Mlp1 1-250 compared to full length Mlp1 (**Figure 15j**), whereas affinity for uridylate RNA is lost when both domains are used individually (Mlp1 1-95 and Mlp1 95-250) (**Supplementary Figure 4b**). These data suggest that despite the apparent lack of the RRM1, Mlp1 functions in a similar manner as the hLa protein in preferentially binding pre-tRNA substrates, and that predicted, conserved uridylate binding residues in the LaM promote higher affinity binding associated with the UUU-3′OH motif.

Table 1. Kd values from EMSAs determining binding of hLa, Mlp1 and Mlp1 mutants to typical La protein RNA targets.

Kd: equilibrium dissociation constant, S.D.: standard deviation, N.D.: binding affinity not determined, N.A.: relative binding calculation not applicable. Color code corresponds to **Figure 15d-j**.

	Protein	**CUGCUGUUUU-3'OH K_d ± S.D. (nM)**	**pre-tRNA Leu K_d ± S.D. (nM)**	**mature tRNA Leu K_d ± S.D. (nM)**	**Relative binding K_{rel} (pre-tRNA vs mature tRNA)**
	hLa	772 ± 31	720 ± 44	1002 ± 16	1.39
	Mlp1	926 ± 27	586 ± 15	831 ± 33	1.42
	Mlp1 Q11A/Y14A	3038 ± 416	965 ± 16	990 ± 15	1.03
	Mlp1 95-340	2616 ± 412	700 ± 62	902 ± 59	1.29
	Mlp1 1-250	982 ± 40	613 ± 53	927 ± 29	1.51
	Mlp1 1-250 Q11A/Y14A	3554 ± 380	1068 ± 51	1232 ± 118	1.15
	Mlp1 1-95	> 40,000	N.D.	N.D.	N.A.
	Mlp1 95-250	> 40,000	N.D.	N.D.	N.A.

The hLa protein is known to also interact with the 5′-triphosphate containing end of the nascent pre-tRNA through a short basic motif located in the C-terminal part of the protein (Fan et al., 1998). We tested the affinity of the 5′-triphosphate for Mlp1 by competition EMSA, using a 5′-leader containing, 3′-trailer processed *in vitro* transcribed radioactively labeled tRNA (+5′PPP) and unlabelled competitor tRNAs with the 5′-triphosphate removed after phosphatase treatment (-5′PPP; +/- 5′PPP demonstrated in **Supplementary Figure 4c**). The unlabelled –5′PPP substrate competed poorly relative to an unlabelled +5′PPP substrate for the radioactive +5′PPP substrate on hLa (**Figure 15k** – compare -5′PPP and +5′PPP hLa, **Supplementary Figure 4d**), but the difference in competition between these RNAs on Mlp1 was smaller (**Figure 15k** – compare -5′PPP and +5′PPP Mlp1, **Supplementary Figure 4e**). These data are consistent with a lesser degree of 5′-leader discrimination for 5′PPP in Mlp1 relative to hLa.

2.3.4. Mlp1 binding to 3′-trailer RNAs is different from hLa

To further study the distinct binding modes, we performed competition experiments between ^{32}P-labeled uridylate RNA (U10) and unlabelled competitor pre-tRNA and mature tRNA substrates. We used hLa as a control and found that, as expected from previous work (Bayfield & Maraia, 2009), pre-tRNAs were a stronger competitor for the uridylate binding pocket than mature tRNAs (**Figure 16a, Supplementary Figure 5a**). In contrast, uridylate RNA was competed off Mlp1 using only low amounts of either pre-tRNA or mature tRNA (**Figure 16b, Supplementary Figure 5a**), suggesting that binding of uridylates on Mlp1 is weaker compared to hLa.

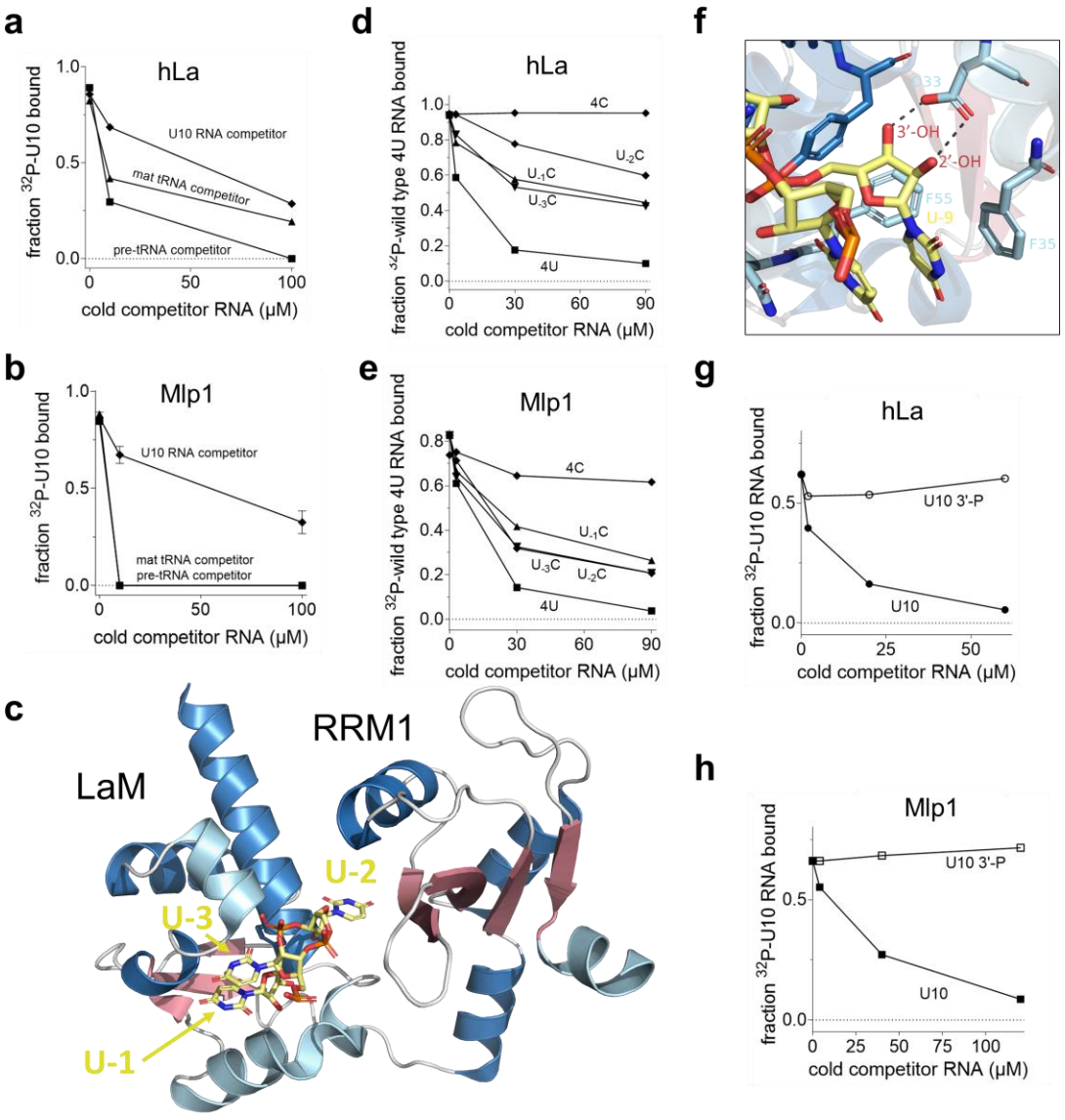


Figure legend on next page

Figure 16. Mlp1 does not discriminate 3′-uridylates as stringently as hLa.

a,b Binding curves from competition EMSA between ^{32}P-labeled uridylate RNA (U10) and unlabelled competitor U10 RNA (control), 5′-leader and 3′-trailer containing *T. thermophila* pre-tRNA Leu^{AAG} and mature tRNA Leu^{AAG} for hLa (**a**) and Mlp1 (**b**). Error bars represent the mean +/- standard deviation (Mlp1 U10: n=4, tRNA: n=2, hLa U10: n=2, tRNA: n=1 independent experiments). Native gels are shown in **Supplementary Figure 5a**. **c** Three-dimensional representation of the hLa protein in complex with uridylate RNA with the last three terminal nucleotides shown (UUU-3′OH) (PDB: 2VOD). The penultimate uridylate U_{-2} is positioned in between the LaM and RRM1, while the 3′-terminal uridylate U_{-1} is positioned more towards the outside in stacking formation with U_{-3}. RNA is shown in yellow; β-sheets are shown in red and α-helices shown in blue (dark blue: typical α-helices found in the winged-helix fold and classic RRM, light blue: inserted α-helices specifically found in La proteins). Image generated using PyMOL. **d,e** Binding curves from competition EMSAs using P^{32}-labeled wild type CUGCUGUUUU (4U) and unlabelled 4U and mutant RNAs: CUGCUGUUUC (U_{-1}C), CUGCUGUUCU (U_{-2}C), CUGCUGUCUU (U_{-3}C) and CUGCUGCCCC (4C) for hLa (**d**) or Mlp1 (**e**) (n=2 independent experiments). Native gels are shown in **Supplementary Figure 5c,d**. **f** Magnified views of hLa and 3′-terminal uridylate U_{-1} interactions (PDB: 2VOD) demonstrating the importance of the 3′-end. Protein carbon backbones have the same color coding as in c, oxygen is shown in red, nitrogen is shown in blue and phosphorous is shown in orange. RNA is shown in yellow. Image generated using PyMOL. **g,h** Binding curves from competition EMSA between P^{32}-labeled U10-3′OH and unlabelled U10-3′OH (control) and U10-3′-P for hLa (**g**) and Mlp1 (**h**) (n=2 independent experiments). Native gels are shown in **Supplementary Figure 5e**. Source data are provided as a Source Data file.

Previous high-resolution structural characterization of hLa bound to UUU-3′OH established that the penultimate uridylate (U**U**U-3′OH) has the greatest importance for sequence specific, high affinity binding (Kotik-Kogan et al., 2008; Teplova et al., 2006). Notably, this uridylate is the only residue in the UUU-3′OH motif that contacts the RRM1 in hLa (**Figure 16c**). Previous work established that hLa has no binding affinity for 10-mers C10, G10 or A10, however, 20-mer versions of G20 and A20 have a considerable binding affinity for hLa (Vinayak et al., 2018). To compare the importance of the position of the penultimate uridylate (two nucleotides from the 3′-end: U_{-2}), we performed competition EMSAs for hLa and Mlp1 using the radioactively labeled UUU-3′OH containing RNA CUGCUGUUUU-3′OH RNA (hence referred to as wild type 4U) and unlabelled RNAs carrying specific U to C variations to the terminal uridylates in this sequence,

as previous work indicated the U to C substitution had the greatest effect on RNA substrate discrimination by hLa (Teplova et al., 2006), and we confirmed complete lack of binding affinity of Mlp1 for C10 (**Supplementary Figure 5b**). As expected, mutating the penultimate uridylate (U_{-2}) into a cytidylate ($U_{-2}C$ – CUGCUGUUCU-3′OH) resulted in this RNA functioning as a poor competitor with hLa, whereas changes at the most 3′-terminal position ($U_{-1}C$) and the third last position ($U_{-3}C$) has a lesser effect on competition capability (**Figure 16d, Supplementary Figure 5c**). Mutating the four last positions (4C – CUGCUGCCCC-3′OH) resulted in the inability to compete with the wild type 4U RNA (**Figure 16d, Supplementary Figure 5c**). In contrast, competition for Mlp1 between radioactively labeled 4U and any unlabelled variant RNA competitor ($U_{-1}C$, $U_{-2}C$ and $U_{-3}C$) showed similar competition levels, indicating that the position of the penultimate uridylate U_{-2} is not as important for interactions between Mlp1 and 3′-trailer sequences (**Figure 16e, Supplementary Figure 5d**). Interestingly, unlabelled competitor 4C was more capable of competition with wild type 4U RNA for binding on Mlp1. These data are also consistent with Mlp1 discrimination for uridylate RNA not occurring as strongly as for hLa, and that the specificity of binding to the penultimate uridylate is weaker or absent for Mlp1.

The previous high-resolution work on hLa in complex with UUU-3′OH containing RNA also revealed the importance of the 2′- and 3′-hydroxyls (-OH) for high affinity binding in the hydrophobic binding pocket between LaM and RRM1 (Kotik-Kogan et al., 2008; Teplova et al., 2006), with two hydrogen bonds formed with an aspartate (**Figure 16f**) that is conserved in Mlp1 (see **Figure 14b, Supplementary Figure 1d,** D33 numbering in hLa). To investigate the importance of the free 3′-terminal hydroxyl groups for Mlp1 binding, we used competition EMSAs comparing a regular U10 and a 3′-phosphorylated unlabeled competitor RNA. We found that for both hLa and Mlp1 the phosphorylated substrate makes a poor competitor, confirming that the presence of a 3′-hydroxyl end is important for interactions between Mlp1 and uridylates (**Figure 16g,h, Supplementary Figure 5e**). Together, these results demonstrate that Mlp1

shows the same preference and LaM amino acid dependence for pre-tRNAs and UUU-3′OH binding as hLa, but that the altered architecture correlates with diminished discrimination for these substrates relative to hLa.

2.3.5. RNA chaperone function of Mlp1 in an *Schizosaccharomyces pombe* based heterologous system

Our computational and biochemical work suggested that Mlp1 differs from other examined La proteins in its binding and discrimination of RNA targets. We have previously used a well-established, heterologous tRNA-mediated suppression system in *S. pombe* to investigate the function of both yeast and human La in the promotion of La-dependent pre-tRNA processing. This system can report on La-dependent 3′-end protection of nascent pre-tRNAs from exonucleases, as well as RNA chaperone activity to enhance correct folding of nascent pre-tRNAs through the ability to rescue a misfolded suppressor tRNA *in vivo* (Bayfield & Maraia, 2009; Hussain et al., 2013; Naeeni et al., 2012; Porat & Bayfield, 2020). To test whether the altered Mlp1 architecture correlated with differences in pre-tRNA processing, we transformed *S. pombe* La protein Sla1p, full-length Mlp1 and multiple Mlp1 mutants into a *sla1- S. pombe* strain (ySH9) which encodes a defective stop codon UGA-decoding suppressor tRNA (tRNA-Ser^{UCA}) as well as the *ade6-704* allele, which is decoded by a tRNA-Ser^{UCA} to suppress red pigment accumulation during growth on low adenine. Successful suppression in this system relies on the presence of a La protein or other suitable RNA chaperone, resulting in white colonies versus red colonies in unsuppressed cells.

When comparing Sla1p-transformants to full-length Mlp1-transformants, we found that Mlp1 can stabilize the defective suppressor pre-tRNA similar to Sla1p (**Figure 17a**). Maturation of the defective suppressor tRNA can occur via 3′-terminal protection of uridylates from exonucleases, or general RNA chaperone activity, or a combination of these to assist with folding of pre-tRNAs

(Huang et al., 2006). To test the importance of the uridylate binding residues for suppression, we compared the Mlp1 Q11A/Y14A mutant and the LaM deletion mutant (Mlp1 95-340) to wild type Mlp1 and observed a pink phenotype indicating intermediate suppression levels despite equal levels of protein expression (**Figure 17a**, **Supplementary Figure 6a**), suggesting that these uridylate binding residues function in maturation of the pre-tRNA *in vivo*. Next, we studied the function of the C-terminal DUF3223 and found that removal of this domain (Mlp1 1-250) led to near complete loss of suppression (**Figure 17a**), and combination of DUF3223 removal and uridylate binding inactivation (Mlp1 1-250 Q11A/Y14A) resulted in a complete loss of suppression (**Figure 17a**). These results indicate that to obtain complete tRNA mediated suppression, both the conserved uridylate binding residues (Q11 and Y14) in the LaM and the C-terminal DUF3223 are required.

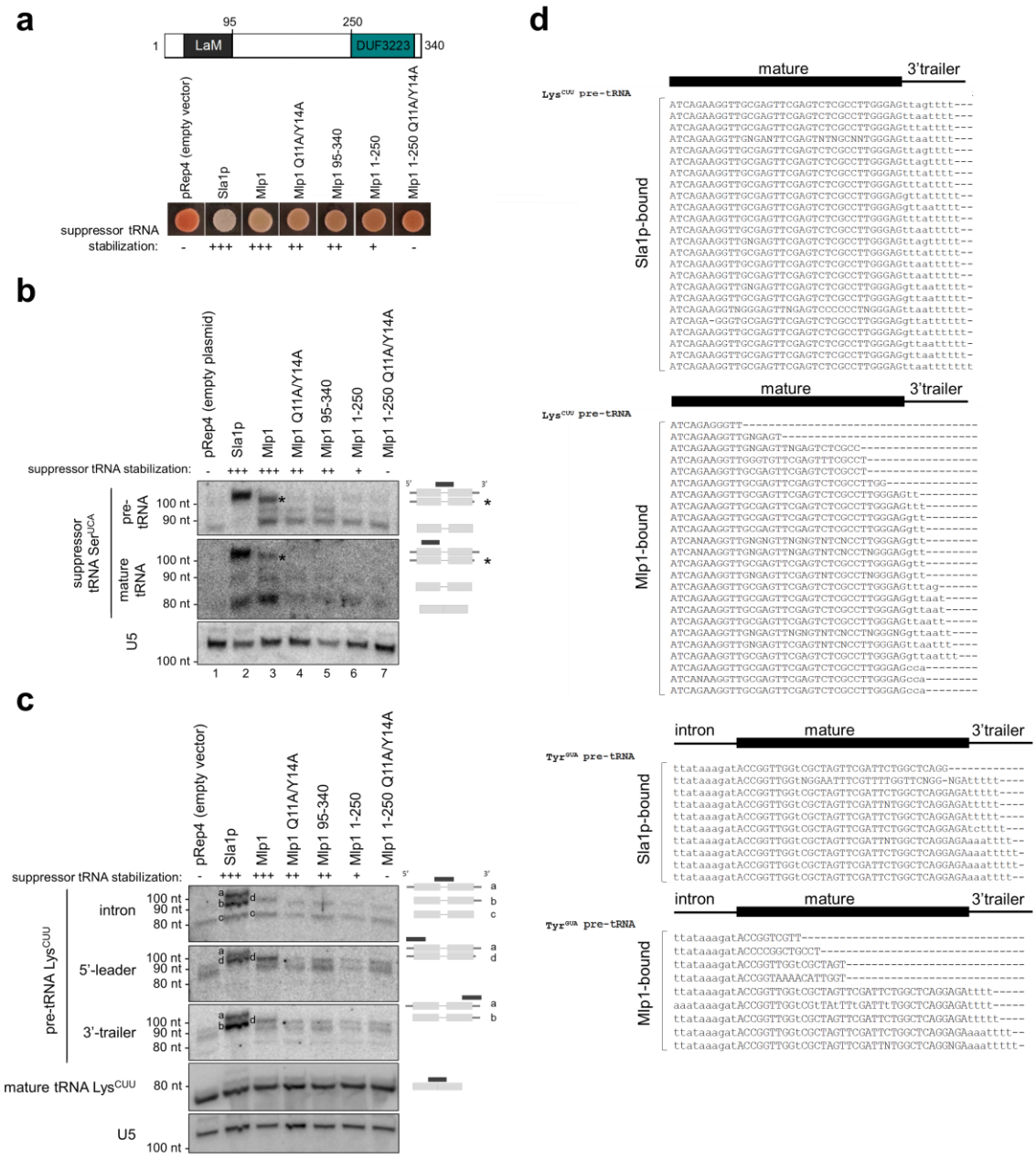


Figure legend on next page

Figure 17. Mlp1 promotes tRNA mediated suppression in the absence of protection of 3′-trailers.
a tRNA-mediated suppression assay of *S. pombe* ySH9 strain transformed with pRep4 encoded Sla1p (positive control; *S. pombe* La protein), Mlp1 or indicated Mlp1 mutants (n=3 biologically independent samples). Protein expression levels are shown by Western blot in **Supplementary Figure 6a**. Levels of suppressor tRNA stabilization is based on colour of the colonies and annotated as follows: full suppression (+++), near full suppression (++), intermediate suppression (+), no suppression (-). Top: diagram showing domain architecture of Mlp1. **b** Northern blot to determine 3′-end protection of suppressor pre-tRNA-Ser^{UCA} in ySH9 transformants shown in **a** (n=2 biologically independent samples). Accumulation of suppressor pre-tRNA Ser^{UCA} was determined using an intron complementary probe and mature tRNA using a tRNA body probe leading to detection of both pre-tRNA and mature tRNA. The asterisk indicates a 3′-processed pre-tRNA intermediate. Excess unlabeled probe complementary to endogenous Ser^{UGA} was added to avoid cross-reaction between suppressor tRNA and endogenous tRNA. U5: loading control. **c** Northern blot to determine 3′-end protection of endogenous pre-tRNA Lys^{CUU} in ySH9 transformants shown in a, with suppressor tRNA stabilization levels shown above the blot (n=3 biologically independent samples). Accumulation of pre-tRNA Lys^{CUU} intermediates was determined using an intron complementary probe which detects unprocessed 5′-leader and 3′-trailer containing pre-tRNA (band a), 5′-leader processed pre-tRNA (band b) and both 5′-leader and 3′-trailer processed pre-tRNA (band c) in Sla1p transformants (positive control; lane 2). The 3′-processed, 5′-leader containing pre-tRNA intermediate is annotated as band d. The same blot was probed with a 3′-trailer complementary probe, a 5′-leader complementary probe and a mature tRNA complementary probe. U5: loading control. **d** Sanger sequencing of clonal isolates corresponding to cDNAs derived from 3′-terminal sequences from Sla1p- and Mlp1-immunoprecipitated pre-tRNAs in ySH9 as transformed in **a-c** (n=2 biologically independent samples). Western blot of Sla1p- and Mlp1-immunoprecipitations shown in **Supplementary Figure 6b**. Source data are provided as a Source Data file.

To investigate 3′-end protection more directly, we extracted total RNA from Mlp1- and Sla1p-expressing strains and analysed suppressor pre-tRNA Ser^{UCA} processing by Northern blot (**Figure 17b**). We found that higher levels of suppressor pre-tRNA stabilization in Sla1p- and Mlp1-transformed strains (lanes 2, 3) corresponded to more mature suppressor tRNA which correlates with the white phenotype observed in the tRNA-mediated suppression results (see **Figure 17a**), as opposed to the lack of pre-tRNA stabilization and subsequent mature suppressor tRNA in pRep4 control (lane 1) resulting in a red phenotype (**Figure 17a,b**). Interestingly, the most abundant pre-tRNA species was slightly smaller in Mlp1 transformed cells relative to Sla1p (see asterisk). We also detected endogenous pre-tRNAs Lys^{CUU} processing intermediates using an intron, 5′-leader and a 3′-trailer probe (**Figure 17c**). The presence of the three La-dependent pathway pre-tRNA intermediates are detected after transformation of the positive control Sla1p, as have been described previously (Intine et al., 2000) (**Figure 17c** – intron probe, compare lanes 1 and 2, indicated by a, b and c). Notably, we observed the appearance of a distinct pre-tRNA intermediate in Sla1p expressing cells when using a probe specific for the 5′-leader of pre-tRNA Lys^{CUU} that did not co-migrate with any of the major species observed using the intron or 3′-trailer probe (**Figure 17c** – 5′-leader, indicated by d, beneath the full length pre-tRNA band). We hypothesize that this species represents a subset of pre-tRNAs that are not stably bound by Sla1p and whose 3′-ends have been nibbled by a 3′-exonuclease, yet have retained their 5′-leaders as has been described during La-independent pre-tRNA processing. As expected from the tRNA-mediated suppression assay, Mlp1 also stabilized endogenous precursor pre-tRNA Lys^{CUU}, however, this precursor tRNA co-migrated with the 5′-leader containing, 3′-trailer exonuclease processed intermediate observed in Sla1p -transformants (**Figure 17c** – 5′-leader, band d, compare lane 2 and 3), consistent with Mlp1 stabilizing a 5′-leader containing, 3′-processed or nibbled pre-tRNA intermediate. As expected, all Mlp1 mutants defective in tRNA-mediation suppression were defective in

stabilizing pre-tRNA intermediates. Together, these data are consistent with Mlp1 engaging pre-tRNAs and promoting tRNA-mediated suppression in *S. pombe*, but with impaired protections of pre-tRNA 3′-ends relative to Sla1p.

To further compare Mlp1 and Sla1p function in the processing of pre-tRNA intermediates by Mlp1 in *S. pombe*, we immunoprecipitated Sla1p and Mlp1 RNP-complexes (**Supplementary Figure 6b**) and sequenced the 3′-ends of their associated pre-tRNA Lys^{CUU} and pre-tRNA Tyr^{GUA} by 3′-rapid amplification of cDNA ends (3′-RACE). While Sla1p-associated pre-tRNAs were enriched for species containing primarily four and five nucleotide long uridylate 3′-trailers, as has been described previously (Huang et al., 2005), Mlp1-associated pre-tRNAs were largely depleted for 3′-trailer containing species (**Figure 17d**), consistent with the Northern blotting results (**Figure 17b,c**), with a minority of Mlp1 immunoprecipitated pre-tRNAs ending in 3′-CCA, indicating Mlp1 can also bind pre-tRNAs after CCA addition but prior to nuclear export and intron removal. These data are consistent with our Mlp1 RNA binding data *in vitro*, indicating that Mlp1 promotes pre-tRNA processing but with impaired 3′ discrimination and lower protection of the 3′-ends of associated pre-tRNAs relative to other examined genuine La proteins.

2.3.6. Mlp1 depletion in *Tetrahymena thermophila* leads to impaired 3′-trailer processing

To investigate the effect of Mlp1 depletion on pre-tRNA processing in *T. thermophila*, we generated a partial *MLP1* knockout strain and confirmed genomic integration of the selection marker by Southern blot and PCR (**Supplementary Figure 7a,b** and data not shown) and reduced protein expression of Mlp1 by Western blot (**Figure 18a**) and indirect immunofluorescent staining (**Supplementary Figure 7c**). Complete deletion of the *MLP1* locus in *T. thermophila* macronuclei through increasing drug selection (phenotypic assortment

(Merriam & Bruns, 1988)) was not achieved, indicating that Mlp1 is likely essential in this system, similar to *Mus musculus*, *Drosophila melanogaster*, *Trypanosoma brucei* and *Arabidopsis thaliana* (Arhin et al., 2005; Bai & Tolias, 2000; Fleurdépine et al., 2007; Park et al., 2006). We extracted total RNA followed by pre-tRNA intermediate detection by Northern blot for pre-tRNA Ile^{UAU}, Leu^{UAA} and Val^{AAC}. In the absence of Mlp1, a 3′-trailer containing pre-tRNA intermediate was stabilized (**Figure 18b** – intron probe, bottom band). These data provide evidence for *T. thermophila* being the first eukaryote in which La protein levels correlate inversely with 3′-trailer stabilization. These results are also consistent with our heterologous expression data in *S. pombe* indicating impaired protection of pre-tRNA 3′-trailers by Mlp1. These data suggest that the variant domain architecture of Mlp1 relative to other genuine La proteins is associated with an altered pre-tRNA processing pathway in this system.

.

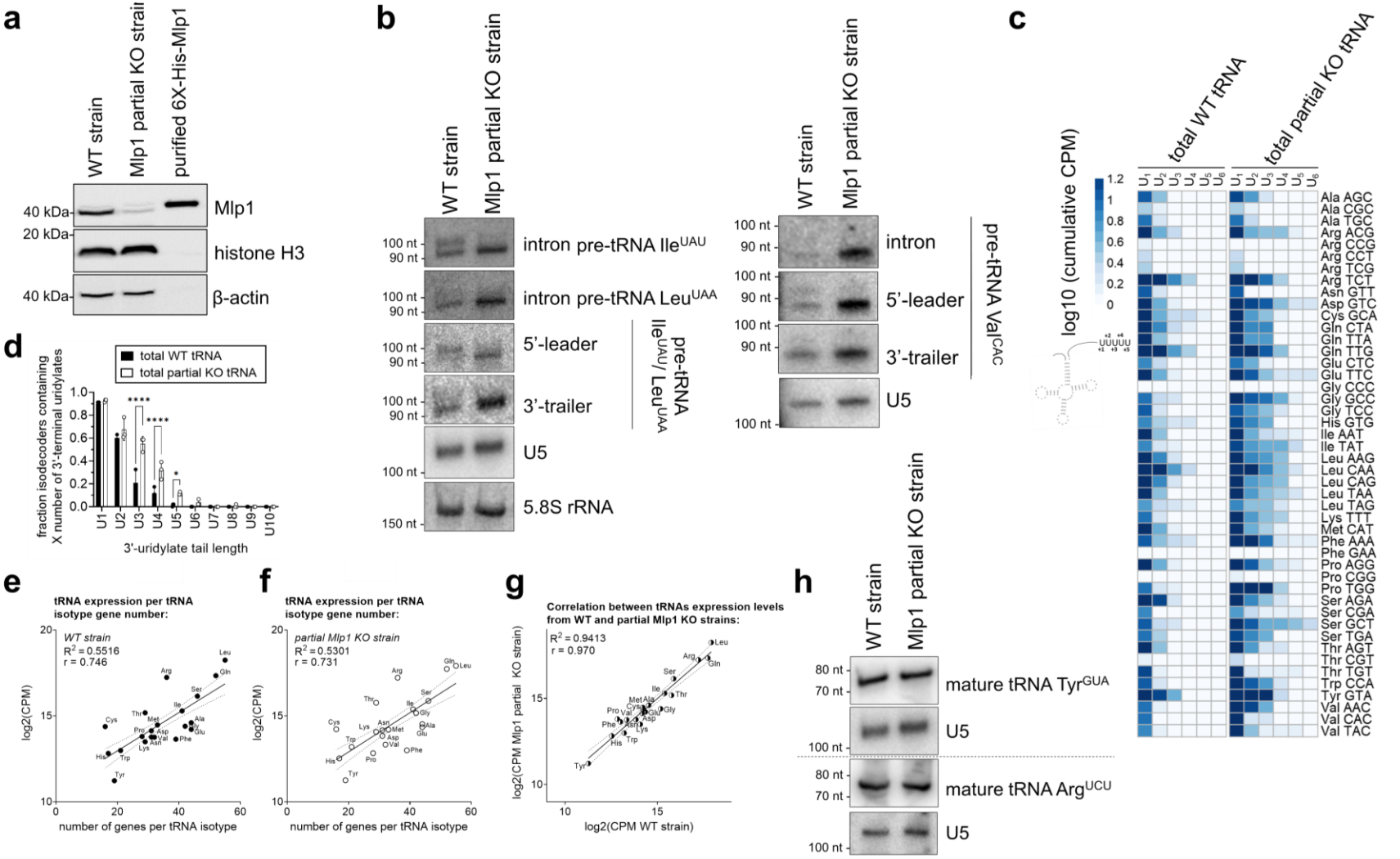


Figure legend on next page

Figure 18. Depletion of Mlp1 results in altered pre-tRNA processing in *Tetrahymena thermophila* with normal levels of mature tRNAs.
a Western blot confirming decreased protein expression of Mlp1 in the partial Mlp1 knockout (KO) strain compared to the wild type (WT) strain (n=3 biologically independent samples). Loading control: histone H3 and β-actin. Purified 6X-His-Mlp1 was analysed to control for antibody specificity. **b** Northern blot detecting tRNAs IleUAU, LeuUAA and ValCAC pre-tRNA intermediates with an intron, 5′-leader, and 3′-trailer specific probe (n=3 biologically independent samples). Pre-tRNA IleUAU and LeuUAA are both recognized by 3′-trailer and 5′-leader probe used. Black boxes correspond to a marker of the same size. U5 snRNA and/or 5.8S rRNA were used as loading controls. **c** Next generation sequencing data shown as a heatmap of log10 transformed normalized cumulative pre-tRNA counts for each uridylate tail length (n=3 biologically independent replicates). For clarity, the average is plotted and the longest uridylate tail shown is U6 (UUUUUU-3′OH). **d** Quantification of uridylate tail length from next generation sequencing data (n=3 biologically independent samples) in **c**. Error bars represent the mean +/- standard deviation. A two-way ANOVA with Bonferroni′s correction for multiple comparisons was performed to determine statistical significance: **** $P < 0.0001$, * $P < 0.05$= 0.0491. For each tRNA isoacceptor, the presence of a uridylate tail length (U1, U2, etc.) was scored as present if its abundance in CPM was greater than 0.3; log10(2). **e,f** Scatterplots of log2 transformed normalized tRNA counts (CPM) from next generation sequencing data from WT strains (**e**) and partial Mlp1 KO strains (**f**) plotted against the number of genes per tRNA isotype encoded in the genome (see Supplementary Additional Data 1). The correlation was calculated, and a positive correlation was observed for both WT and KO strains ($r = 0.746$ and $r = 0.731$, respectively) indicating that more tRNA gene copies result in higher tRNA expression. **g** Plotting of log2 transformed normalized WT and partial Mlp1 KO counts (CPM) gives a strong positive correlation ($r = 0.970$). **h** Northern blot analysis detecting mature tRNA TyrGUA and ArgUCU levels using a probe specific for the spliced mature tRNA (n=3 biologically independent samples). U5 was used as a loading control on both membranes. Source data are provided as a Source Data file.

To study the effect of Mlp1 depletion on a transcriptome-wide scale, we performed total RNA-sequencing of small RNA-enriched samples, similar to our Mlp1 RIP-Seq (see **Figure 15b**), in which counts for each pre-tRNA isoacceptor were determined for each unique uridylate tail length at the 3′-end of the tRNA. We found that the average length of the 3′-uridylate tail increased for the majority of tRNA species as a result of Mlp1 depletion (**Figure 18c,d**). These findings indicate that Mlp1 promotes removal of the 3′-trailer and that when Mlp1 is limiting, the 3′-trailer sequences are not processed as efficiently. Together, these data are consistent with Mlp1 accelerating 3′-end processing, relative to other examined species in which genuine La proteins stabilize 3′-trailers.

Previous work has demonstrated that in eukaryotes, the number of genomic tRNA gene copies correlates strongly with the expression of mature tRNA transcripts and thus can function as a predictor for tRNA expression levels (Behrens et al., 2021; Tuller et al., 2010). We observed that in *T. thermophila* the number of genomic tRNA genes correlates with the expression levels of tRNAs (R^2 = 0.55, r = 0.75) (**Figure 18e**). To investigate if depletion of Mlp1 results in changes of total tRNA levels, we performed the same analysis for our Mlp1 partial knockout strain (R^2 = 0.53, r = 0.73) (**Figure 18f**) and found that when comparing tRNA expression between wild type and partial Mlp1 knockout strains the correlation was very strong (R^2 = 0.94, r = 0.97) (**Figure 18g**), indicating that Mlp1 depletion did not influence mature tRNA expression. Northern blot analysis for representative tRNAs confirmed similar amounts of mature tRNA between wild type and partial Mlp1 knockout strains (**Figure 18h**), reminiscent of minimal differences in mature tRNA abundance (despite changes to pre-tRNA processing) in conditional mouse La knockout cells (Gaidamakov et al., 2014).

2.3.7. The 3′-trailer lengths are shorter in *Tetrahymena thermophila* relative to other eukaryotes

To explore tRNAs and pre-tRNA processing in *T. thermophila* more extensively, we compared their predicted pre-tRNA architecture to those from other eukaryotic species. We compared 3′-trailer lengths of each pre-tRNA, as determined by the number of nucleotides between the discriminator base and the genomic stretch of at least four consecutive thymines (Bogenhagen & Brown, 1981) that give rise to the 3′UUU-OH motif in the nascent transcript. We found that *T. thermophila* has shorter 3′-trailer lengths than any other species analyzed, with the most common 3′-trailer length being zero nucleotides (**Figure 19a, Table 2, Supplementary Data 2**). Identical trends were observed for genomic stretches/candidate terminators of five, six or seven consecutive thymines (**Supplementary Table 2, Supplementary Data 2**). A similar analysis for the total length of the mature tRNA revealed that, as expected, the average length of mature tRNAs is the same as other examined species (**Figure 19b, Table 2**). We previously noted differences in enrichment of pre-tRNAs isotypes (**Supplementary Figure 2b** – pre-tRNAs) in our Mlp1 RIP-seq results and hypothesized that the 3′-trailer length could be a determinant for binding affinity. We plotted fold enrichment for pre-tRNAs between Mlp1-immunoprecipitated and input tRNAs against 3′-trailer length and found that these are moderately negatively correlated (R^2 = 0.14, r = -0.47) (**Supplementary Figure 8a**), suggesting that nascent pre-tRNAs containing a longer 3′-trailer sequence may have a slightly lower binding affinity for Mlp1. When comparing this to the hLa data, we found a lesser correlation (R^2 = 0.05, r = -0.23) (**Supplementary Figure 8b**).

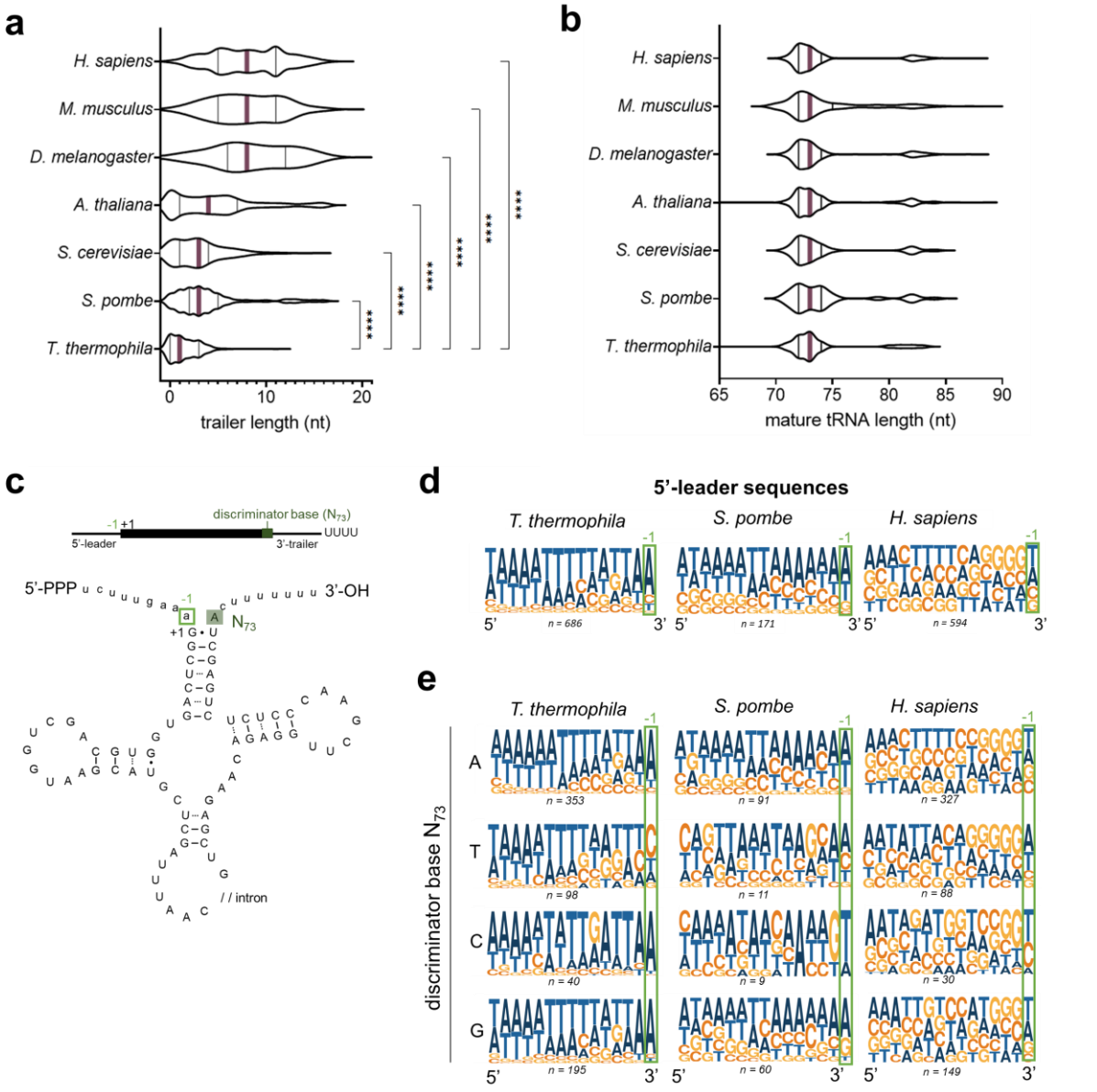


Figure legend on next page

Figure 19. The 3′-trailer length of *Tetrahymena thermophila* tRNAs is considerably shorter compared to other eukaryotes and affects the N_{-1} composition of the 5′-leader.

a Genome-wide analysis of 3′-trailer lengths as determined by the number of nucleotides between the discriminator base (the most 3′-terminal nucleotide in the mature tRNA upstream of the posttranscriptional added CCA) and a genomic stretch of four Ts of different eukaryotes. Violin plots shows the median as a full purple line and the quartiles as full black lines. A one-way ANOVA with Tukey's correction for multiple comparisons was performed to determine statistical significance: **** $P < 0.0001$. Data summary shown in **Table 2**. **b** Genome-wide analysis of mature tRNA length for different eukaryotes. Violin plots show the median as a full purple line and the quartiles as full black lines. A one-way ANOVA with Tukey's correction for multiple comparisons was performed to determine statistical significance: no statistical significance ($P < 0.05$) was observed. Data summary shown in **Table 2**. **c** Schematic representation of a pre-tRNA. The mature tRNA sequence is shown as a black rectangle and written in uppercase (UCGA) in the tRNA cartoon. Pre-tRNA specific sequences, including the 5′-leader and 3′-trailer are shown as a black line and written in lowercase (ucga) in the tRNA cartoon. The discriminator base (N_{73}) is the most 3′-terminal nucleotide of the mature tRNA sequence highlighted in dark green. The most 3′-terminal nucleotide of the 5′-leader sequence is highlighted with a light green box (N_{-1}). **d,e** Logo analysis of 5′-leader sequences of pre-tRNAs in *T. thermophila, S. pombe* and *H. sapiens* (**d**). The same analysis split by the discriminator base identity (**e**). The number of pre-tRNAs representing each condition is shown underneath each logo. Source data are provided as a Source Data file.

Table 2. Summary of 3′-trailer and mature tRNA sequence lengths in different eukaryotic species.

S.D.: standard deviation, nt: nucleotides.

Species	Average of mature tRNA length ± S.D. (nt)	Number of tRNAs analysed	Average of 3'-trailer length ± S.D. (nt)	Number of tRNAs analysed (3'-trailer lengths with >20 nt excluded)
H. sapiens	74 ± 3.4	594	8.1 ± 3.7	412
M. musculus	74 ± 3.7	811	7.9 ± 3.6	343
D. melanogaster	74 ± 3.6	295	8.5 ± 4.0	247
A. thaliana	74 ± 3.8	642	4.6 ± 4.3	532
S. cerevisiae	74 ± 3.5	275	3.0 ± 2.6	269
S. pombe	74 ± 3.4	171	3.3 ± 2.8	162
T. thermophila	74 ± 3.3	686	1.5 ± 1.5	686

Processing of the 5′-leader by RNase P is based on tertiary structure recognition of the tRNA followed by endonucleolytic cleavage at the –1/+1 position (Leontis et al., 1988; Wu et al., 2018). Previous studies demonstrated structural conservation of the catalytically active RNA in

RNase P in *T. thermophila*, indicating that processing occurs in a similar manner (True & Celander, 1996; Wu et al., 2018). Optimal efficiency of RNase P cleavage is dependent on the presence of a bulge that includes the last nucleotide of the 5′-leader (the N_{-1} position; **Figure 19c**), as extensive base pairing between 5′-leader and 3′-trailer sequences inhibits 5′-leader cleavage (Lee et al., 1997). We hypothesized that *T. thermophila* tRNAs should have adapted 5′-leader sequences to ensure the generation of a mismatched bulge against the shorter 3′-trailer sequences. We determined the most common 5′-leader sequences by logo generation for *T. thermophila, S. pombe* and *Homo sapiens* and found that the most frequent nucleotide at the most 3′-residue of the 5′-leader sequence (N_{-1}) in *T. thermophila* is an adenine (± 75%) (**Figure 19d**, **Supplementary Figure 8c**). In contrast, the distribution of nucleotides at this position is more diverse in other species (**Figure 19d, Supplementary Figure 8c**). At the 3′-end of the mature transcript, the discriminator base (N_{73}) is mostly adenine (± 50%), followed by guanine as second most frequent nucleotide (± 25%) (**Supplementary Figure 8d**) which is common for all eukaryotes in this study.

We then split our data based on the identity of the discriminator base (N_{73}) (**Figure 19e**, **Supplementary Figure 8e**), which can pair with the first nucleotide in the 5′-leader (N_{-1}). We observed a strong lack of Watson–Crick base pairing for *T. thermophila* between the opposing nucleotides in the discriminator base pairing with the typical adenine (A_{73} – A_{-1}; C_{73} – A_{-1} and G_{73} – A_{-1}). Conversely, tRNAs containing a thymine as discriminator base (T_{73}) strongly avoided an adenine as the N_{-1} 5′-leader base to avoid base pairing. This partitioned discrimination is not evident for *S. pombe* and *H. sapiens* (**Figure 19e**). Since shortened 3′-trailers in *T. thermophila* should have a greater dependence on leader-trailer bulges occurring through the discriminator base, these data are thus consistent with the 3′ most nucleotide of the 5′-leader having a greater evolutionary pressure to avoid base pairing with the discriminator base which is seen most strongly for *T. thermophila* (**Supplementary Figure 8f**) relative to other examined species.

2.4. Discussion

Previously studied genuine La proteins contain a conserved tandem arrangement of a LaM and RRM1 collectively referred to as a La module. Previous phylogenetic predictions (Deragon, 2020) and our continued computational analysis have indicated that the previously characterized *T. thermophila* protein Mlp1 may group with the genuine La proteins, despite the predicted lack of the La module′s RRM1 domain. In addition to the absence of the RRM1, Mlp1 is predicted to have a previously uncharacterized domain of unknown function (DUF3223), which is absent in all other examined members of the LARP superfamily. Since this arrangement of a genuine La protein is unprecedented, it was not clear how this protein might perform La associated functions and what effects this might have, if any, on the processing of La RNA targets. We present data consistent with Mlp1 functioning as a genuine La protein. Using *in vitro* binding assays, we demonstrate that the identity of the Mlp1 residues Q11 and Y14, analogous to UUU-3′OH binding residues in hLa, are also required for high affinity binding of 3′-uridylate containing trailer sequences. We demonstrate that binding of uridylate RNA is retained when both the predicted LaM (Mlp1 1-95) and middle domain (Mlp1 95-250) are included**,** however, uridylate binding does not occur when either the LaM (Mlp1 1-95) or the middle domain (Mlp1 95-250) are tested in isolation. Given the relative paucity of contacts from the RRM relative to the LaM during hLa binding to UUU-3′OH, it is likely that the 95-250 region might serve an analogous function, making contacts important for UUU-3′OH binding despite the absence of the RRM fold.

While the Mlp1 LaM and middle domain combine to support UUU-3′OH binding similar to the classic La module, other key differences exist. Competition experiments indicated that short UUU-3′OH containing trailers were more easily displaced from Mlp1 by pre-tRNA and mature tRNA substrates, relative to hLa, suggesting that Mlp1 binding to UUU-3′OH may be less stable

compared to La proteins with the classic LaM-RRM1 arrangement. The previous hLa-UUU-3′OH co-crystals revealed that RRM1 makes only a single contact with the penultimate uridylate RNA (U_{-2}) during UUU-3′OH binding, and it is this uridylate that is recognized with the greatest specificity. Unlike hLa, the $U_{-2}C$ RNA was no less effective a competitor against the 4U RNA than the $U_{-3}C$ and $U_{-1}C$ RNAs when Mlp1 was tested. The impaired ability of Mlp1 to discriminate the U_{-2} residue, as well as uridylates more generally, suggest that the lack of the RRM1 results in a relatively lower ability of Mlp1 to differentiate UUU-3′OH containing RNAs.

Previous work has demonstrated that the La module (LaM+RRM1) of different LARPs, and more specifically their RRM regions, possess RNA chaperone activity in the absence of 3′-uridylate protection from exonucleases (Hussain et al., 2013; Naeeni et al., 2012). Using tRNA-mediated suppression in a La null strain (*sla1-)* of *S. pombe*, we demonstrated that while the uridylate binding residues in the LaM of Mlp1 are important for suppression, the DUF3223 region also promotes the correct folding of defective pre-tRNAs. These data raise the possibility that the DUF3223 region of Mlp1 may serve an analogous function in RNA chaperone activity previously associated with the RRM1 and other C-terminal regions of genuine La proteins.

Genuine La proteins are however well known to also stabilize 3′-trailer containing intermediates through high-affinity binding of the 3′-uridylate tail and thereby providing protection against 3′-exonucleases. Using Northern blotting and RIP-3′-RACE, we demonstrated that while *S. pombe* La stabilizes 5′-leader, 3′-trailer, and intron-containing pre-tRNA intermediates, Mlp1 stabilized a 5′-leader and intron containing, but 3′-trailer lacking, pre-tRNA species, suggesting that Mlp1 may accelerate 3′-processing of nascent pre-tRNAs, relative to other La proteins. This could imply that binding of Mlp1 to the 3′-ends of the nascent pre-tRNAs is not as strong as compared to Sla1p, leaving the RNA exposed to 3′-exonucleases. Alternatively, Mlp1 could increase access of the 3′-trailer to the endonuclease RNase Z.

To better understand how the altered architecture of Mlp1 might influence pre-tRNA processing, we generated a partial *MLP1* knockout strain in *T. thermophila* and performed small RNA-sequencing and Northern blots of endogenous tRNA species. We found that, consistent with our Mlp1 data from *S. pombe*, Mlp1 appears to promote the removal of 3′-trailers, as depletion of Mlp1 *in vivo* leads to a greater relative abundance of UUU-3′OH trailer extensions. In yeast, depletion of La leads to exonucleolytic nibbling and a lower abundance of UUU-3′OH ends, while in the presence of La the 3′-end is stabilized until endonucleolytic cleavage by RNase Z directly 3′ to the discriminator base (N_{73}).

Previous work has demonstrated that hLa impedes access by RNase Z, delaying the 3′-endonucleolytic and resulting in 5′-leader processing proceeding 3′-trailer processing (Nashimoto et al., 2001). We hypothesize that the presence of an RRM1 in hLa could be contributing to this function, seeing that in Mlp1, a La protein lacking the RRM1, the 3′-processing appears not to be blocked. Thus, the apparent destabilization of 3′-trailers by Mlp1 in *T. thermophila* suggests that the function of La in this system, differing from other examined La proteins, may be to increase access of the pre-tRNA 3′-trailer to the tRNA processing machinery (**Figure 20**). The difference in processing of pre-tRNAs is likely caused by Mlp1 since the 3′-exonuclease Rex1p and 3′-endonuclease RNase Z are conserved in *T. thermophila* (**Supplementary Figure 9a,b**). An alternate but not mutually exclusive hypothesis may be that reduction of Mlp1 levels may alter access of pre-tRNA 3′-ends to the LSm2-8 complex, and that this may be linked to 3′-trailer accumulation relative to wild-type cells, as previous work in yeast has also linked the LSm2-8 complex to the processing of pre-tRNA 3′-ends (Kufel et al., 2002; Kufel & Tollervey, 2003). To confirm that the LSm2-8 complex functions similar as in other eukaryotic systems, we compared primary amino acid sequence alignments and found that the LSm2-8 complex in *T. thermophila* appears to be conserved, indicating similar functionality (**Supplementary Figure 9c,d**). We also observed that Mlp1 depletion did not lead to changes in

mature tRNA expression levels, suggesting that an Mlp1-independent tRNA maturation pathway also likely exists in *T. thermophila*, similar to budding and fission yeast.

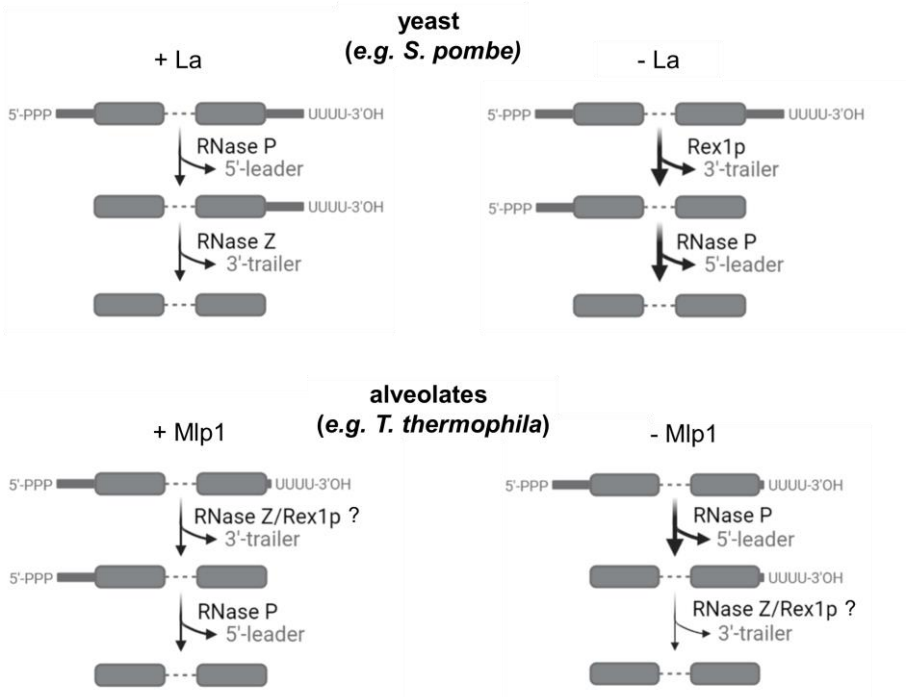


Figure 20. Model for pre-tRNA processing in *Tetrahymena thermophila*.

During La-dependent processing (top, left) in previously studied eukaryotes such as yeast, the La protein is the first protein to associate with pre-tRNAs on the 3′-stretch of uridylates generated by RNA polymerase III transcription termination. Binding of La provides protection from degradation by 3′-exonucleases, assists with tRNA folding through RNA chaperone activity and stabilizes the nascent pre-tRNA. The next step in tRNA processing is endonucleolytic removal of the 5′-leader by RNase P, followed by an endonucleolytic cut by RNase Z resulting in removal of the 3-trailer sequence and the La protein bound to the uridylate stretch. In contrast, during La-independent processing of pre-tRNAs (top, right), the 3′-trailer is rapidly removed first by 3′-exonucleases such as Rex1p, followed by RNase P processing resulting in an end-matured tRNA. Our data from *T. thermophila* point towards a pre-tRNA processing model in which Mlp1-dependent processing (bottom, left) 3′-trailers are processed efficiently. Mlp1-independent processing results in accumulation of pre-tRNAs containing unprocessed 3′-trailer sequences, indicating that Mlp1 is required for efficient 3′-end processing unlike other eukaryotes. Image created with BioRender.com. The Publication License is provided as a Source Data file.

Along with the alternate Mlp1-associated tRNA processing in *T. thermophila*, we observed several differences in features of pre-tRNA 5′-leaders and 3′-trailers compared to other eukaryotic species. Our genome-wide 3′-trailer length analysis in eukaryotes revealed that *T. thermophila* pre-tRNAs have very short 3′-trailer sequences. Since 3′-trailers in *T. thermophila* are dominated by the uridylate stretch, we studied the prevalence of the 3′-most terminal residue of the 5′-leader (N_{-1}) and found that *T. thermophila* pre-tRNAs appear to have evolved to avoid perfect matches with the discriminator base (N_{73}) (preceding the uridylate tail), ensuring a lack of complete base pairing to enhance RNase P cleavage in the context of a minimized 3′-trailer sequence. Further upstream in the 5′-leader (N_{-2}) is often an adenosine, but then typically uridylates, which would not be expected to pair with the 3′UUU-OH motif. The 5′-leaders and 3′-trailers in *T. thermophila* may therefore have evolved to have minimal base pairing between pre-tRNA 5′-leaders and 3′-trailers, compared to other eukaryotes in which a bulge proximal to the mature tRNA ends is often followed by a paired region closer to the 5′-leader and 3′-trailer extremities. For example, human pre-tRNAs have an approximately equal distribution of N_{-1} nucleotide for discriminator bases A, T and G. The most prevalent N_{-1} nucleotide for discriminator bases A and T, are T and A respectively, indicating that human pre-tRNAs, which generally contain a longer 3′-trailer sequence, have more flexibility in N_{-1} sequence to introduce a mismatch bulge at the N_{-1} site for optimal 5′-leader processing by RNase P. It will be interesting to further investigate whether this alternate arrangement is linked to the altered functional roles described here for Mlp1 in this system.

2.5. Methods

Conservation analysis.

Accession numbers used to obtain primary amino acid sequences from the National Center for Biotechnology Information (NCBI) for La, Rex1p, RNase Z/ELAC 2 and LSm2-8 complex are shown in **Supplementary Table 3** and primary amino acid alignments were obtained using Clustal Omega (EMBL-EBI) (Sievers et al., 2011), followed by analysis using a custom python script to annotate identical and conserved amino acids using human protein sequences as a reference. Amino acids were grouped as conserved based on side chain diversity: (i) Asp (D), Glu (E), Asn (N), Gln (Q); (ii) Lys (K), Arg (R), His (H); (iii) Phe (F), Trp (W), Tyr (Y); (iv) Val (V), Ile (I), Leu (L), Met (M); and (v) Ser (S), Thr (T).

Secondary structure predictions were obtained using Phyre2 (Kelley et al., 2015) and color coded based on predicted β-sheet or α-helices. High-resolution structures of the LaM of *H. sapiens, T. brucei,* and *D. discoideum* La proteins were obtained from PDB: 2VOD [https://www.rcsb.org/structure/2vod], 1S29 [https://www.rcsb.org/structure/1s29], and 2M5W [https://www.rcsb.org/structure/2M5W], respectively. Structure of the La module (LaM+RRM1) in complex with uridylate RNA of *H. sapiens* La was obtained from PDB: 2VOD [https://www.rcsb.org/structure/2vod]. Structure prediction of the LaM of *T. thermophila* was obtained using the lomets2 tool (Zheng et al., 2019). Structure of the LSm2-8 complex in complex with uridylate RNA of *S. pombe* and *S. cerevisiae* were obtained from PDB: 6PNN [https://www.rcsb.org/structure/6PNN] and 4M7D [https://www.rcsb.org/structure/4M7D], respectively. Modeling was done using PyMOL.

DNA constructs.

The *MLP1* sequence encoding *T. thermophila* La protein (NCBI Reference Sequence: XP_001019287.2) was obtained from the Genome Database http://www.ciliate.org/

(TTHERM_00384860) (Stover et al., 2012), and codon optimized for expression in *Escherichia coli*. gBlocks of the optimized codon sequence were ordered from integrated DNA technologies (IDT) (sequence can be found in **Supplementary Data 3**).

Mlp1 and Mlp1 mutants used for *in vitro* electromobility shift assays (EMSAs) were cloned in the *NheI* and *BamHI* restriction enzyme sites of the pET28a vector using the plasmid encoded N-terminal 6X-His-tag for protein purification. pET28a hLa was previously cloned in the *NcoI* and *BamHI* sites including a reverse primer-encoded 6X-His tag. Mlp1 and Mlp1 mutants used for *S. pombe* transformations were cloned in the *SalI* and *BamHI* sites in pRep4 plasmid (ura+) incorporating a 6X-His tag N-terminally in the forward primer during PCR amplification. All primers are listed in **Supplementary Data 3.**

T7 DNA templates for *in vitro* transcription were generated by PCR amplification using pre- or mature tRNA specific primers (listed in **Supplementary Data 3**) to obtain DNA templates containing an upstream T7 promotor. The DNA template was gel purified using 7 M urea denaturing 10% polyacrylamide gel and DNA was eluted from the gel by overnight rotation in 150 mM sodium acetate in 50% phenol:chloroform:isoamylalcohol (25:24:1) at 4°C. The aqueous layer obtained by centrifugation at 20,000 g for 10 minutes at 4°C was ethanol precipitated overnight at -80°C.

***Tetrahymena thermophila* cultivation and knockout strain generation.**

Liquid cultures were grown to mid-log phase (0.1 - 1 x 10^6 cells/mL) at 30°C shaking at 90 RPM in SPP media (1% proteose peptone, 0.1% yeast extract, 0.2% glucose, 0.003% Fe-EDTA) (Fillingham et al., 2001). Cell pellets were harvested by centrifugation for 3 minutes at 1000 g. Following two washes in 10 mM Tris-HCl, pH 7.4, pellets were stored at -80°C.

PCR-amplification from *T. thermophila* genomic DNA to obtain flanking 5′- and 3′-regions was completed using primers containing *KpnI/XhoI* and NotI/SacI, respectively. The flanking regions of the *MLP1* gene (TTHERM_00384860 [http://ciliate.org/index.php/feature/details/TTHERM_00384860]) were cloned into pNeo4 plasmid (Mochizuki, 2008) flanking the paromomycin (Neo4) drug resistance cassette using restriction enzyme sites *KpnI XhoI* and *SacI NotI*, respectively. The Neo4 cassette is located downstream of the $CdCl_2$ inducible metallothionein (MTT1) promoter. The resulting pNeo4 *MLP1* knockout plasmid DNA construct was linearized using the *ScaI* restriction enzyme prior to transformation. Biolistic transformation of *T. thermophila* was performed as described previously (Chalker, 2012). Integration of the DNA construct is based on homologous recombination and transformants were grown under increasing concentration of paromomycin starting at 60 µg/mL to a final concentration of 1000 µg/mL (phenotypic assortment) (Merriam & Bruns, 1988). Correct integration was determined using Southern blot and PCR. Partial MLP1 knockout strain available on request.

Protein isolation from *Tetrahymena thermophila.*

T. thermophila cell pellets were resuspended in 10% Trichloroacetic acid (TCA) in 1X PBS, followed by incubation at –20°C overnight to enhance protein precipitation. The white fluffy protein pellet was collected by centrifugation for 15 minutes at 10,000 g at 4°C, then washed twice with 100% ice-cold acetone. The pellet was air-dried and resuspended in 50µL/mL culture 2.5X loading dye [5X loading dye: 5% β-mercaptoethanol (v/v), 0.02% bromophenol blue (w/v), 30% glycerol (v/v), 10% sodium dodecyl sulfate (SDS) (w/v), 250 mM Tris-HCl, pH 6.8].

RNP-immunoprecipitation from *Tetrahymena thermophila.*

T. thermophila cell pellets from 100 mL cultures were collected at log phase (0.1 - 1 x 10^6 cells/mL) and washed twice with 1X PBS. The cells were cross-linked with 1% formaldehyde for

10 minutes followed by quenching with 0.25 M glycine for 5 minutes at room temperature. Cells were washed twice in 1X PBS before lysis in 2 mL buffer A [30 mM Tris-HCl pH 7.4, 150 mM NaCl, 20 mM KCl, 2 mM $MgCl_2$, 0.1% Triton-X100, 1 mM phenylmethylsulfonyl fluoride (PMSF), 1X Halt Protease Inhibitor Cocktail (PIC) (ThemoFisher Scientific), 100U SUPERase In RNase Inhibitor (ThermoFisher Scientific)] followed by sonication (25% amplification with 15s intervals for 4 minutes). The lysate was clarified by centrifugation at 20,000 g for 45 minutes at 4°C and treated with 10 U TurboDNase. The lysates were pre-cleared using rabbit isotype immunoglobin G (IgG) control-bound Protein G Dynabeads rotating for 1 hour at 4°C (12 µg / 50 µL beads slurry). Immunoprecipitations were performed using an affinity purified rabbit anti-Mlp1 antibody (ThermoFisher Scientific – Custom Antibodies) and a rabbit IgG control coupled to Protein G Dynabeads rotating 1 hour at 4°C (12 µg / 50 µL beads slurry). The Protein G Dynabeads were washed 5 times using buffer A (without PIC, PMSF and RNase inhibitor). Input and eluted RNA was isolated following reverse cross-linking for 45 minutes at 70°C in buffer B [50 mM Tris-HCl pH 7.4, 5 mM EDTA pH 8.0, 10 mM DTT, 1% SDS], followed by Trizol extraction.

tRNA library preparation and TGIRT sequencing and analysis.

Total and Mlp1-associated RNAs were size-selected (20-300 nucleotides) on a 7 M urea denaturing 10% polyacrylamide gel. RNA was eluted from the gel by overnight rotation in 150 mM sodium acetate, pH 5.1 in 50% phenol:chloroform:isoamylalcohol (25:24:1) at 4°C. The aqueous layer obtained by centrifugation at 20,000 g for 10 minutes at 4°C was ethanol precipitated overnight at -80°C. The pellet was washed once with 70% ethanol, air dried and resuspended in RNase-free H_2O. To promote sequencing of highly structured tRNA species carrying modifications that are inhibitory to cDNA synthesis, we used TGIRT-III reverse transcriptase (InGex) template-switching to prepare cDNA libraries as described previously (Fafard-Couture et al., 2021; Nottingham et al., 2016). Libraries were sequenced on a NextSeq500 platform (Illumina) (2x 75 bp).

tRNA-Seq processing pipeline.

Details about tools and parameters of bioinformatic analyses are provided in a reproducible Snakemake workflow that can be found at https://github.com/etiennefc/t_thermophila_RNA_Seq.git and are also described below. The raw data (.fastq files) are available for download on the Gene Expression Omnibus (GEO) under the accession number GSE199642 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE199642]. Briefly, paired-end reads were trimmed using Trimmomatic v0.36 (Bolger et al., 2014), and FastQC v0.11.5 was used to evaluate read quality before and after trimming, as described previously (Fafard-Couture et al., 2021). Resulting trimmed reads were then aligned stringently (*i.e.* allowing no mismatch) to the *T. thermophila* genome assembly version T_Thermophila_MAC_2021 (Sheng et al., 2020) (accessible on the Genome Database at http://www.ciliate.org/system/downloads/1-upd-Genome-assembly.fasta) using STAR v2.6.1a (Dobin & Gingeras, 2016) (with the parameters described previously (Fafard-Couture et al., 2021) and also the following parameter and value: --outFilterMismatchNmax 0. The index needed to align reads to the genome was produced using STAR v2.6.1a as described previously (Fafard-Couture et al., 2021) using the genome assembly described previously and a custom *T. thermophila* annotation (.gtf) file available at https://zenodo.org/record/6391187#.YkH9v-fMK3A. The annotation file was built by converting .gff files (one for protein-coding genes, one for tRNA genes and one for 5S ribosomal RNA genes; these .gff files are also available at https://zenodo.org/record/6391187#.YkH9v-fMK3A) into .gtf files and by concatenating these files into one final (.gtf) annotation file using custom bash scripts. This annotation was corrected for embedded genes using CoCo v0.2.5.p1 (Deschamps-Francoeur et al., 2019) with the correct_annotation mode and default parameters. Counts were attributed to genes and normalized as transcripts per million (TPM) as previously described (Fafard-Couture et al., 2021) using CoCo v0.2.5.p1 with the correct_count mode.

Bedgraph files were generated using CoCo v0.2.5.p1 (with the correct_bedgraph mode with default parameters). Differential expression analysis was performed using DESeq2 (Love et al., 2014) with default parameters and the count output of CoCo correct_count to compare the wild type, Mlp1 immunoprecipitation and Mlp1 partial Mlp1 knockout samples. Normalized counts measured in TPM and differential expression data for tRNAs are shown in **Supplementary Data 4**. Reproducibility of tRNA counts in biological replicates was determined generating correlation plots of the TPM after filtering reads with low counts in R.

tRNA read fishing and binning into pre-tRNAs and mature tRNAs.

Raw counts for pre-tRNA (3′-UUU) and mature tRNAs (3′-CCA) were generated using custom python scripts. A list of unique sequences was generated for each tRNA isoacceptor (*e.g.* Gln^{UUG}: AATCCTCTGACCTGGGTTCGAATCCCAGTACGACCT) (**Supplementary Data 5**) and used to obtain ("fish") all reads from the unmapped raw sequence files (.fastq file format). Each sequence was grouped in corresponding bins based on the 3′-end of the reads: -CCA (mature tRNA), 1U, 2U, 3U, 4U, 5U, 6U, 7U, 8U, 9U or 10U (premature tRNA). Raw counts for each tRNA isotype were normalized as counts per million (CPM) by dividing raw counts by the total number of fished read per bin for each replicate divided by 10^6 (**Supplementary Data 6 and Supplementary Data 7**). Fold-enrichment for Mlp1-bound tRNAs was calculated after summation of pre-tRNA counts, followed by taking the ratio of CPM data for Mlp1-immunoprecipitated tRNAs and wild type input tRNAs and displayed after log2 transformation in heatmaps (**Supplementary Data 6**). Fold-enrichment for hLa-bound tRNAs from Gogakos et al. (Gogakos et al., 2017) was obtained from the GEO database under accession number GSE95683 where counts for pre-tRNAs and mature tRNAs were normalized as CPM and transformed identically as Mlp1-bound tRNAs. The cumulative abundance in CPM for each pre-tRNA isoacceptor was calculated for wild type and partial knockout tRNAs and displayed after log10 transformation in heatmaps (**Supplementary Data 7**).

Electromobility shift assays (EMSA).

U10 and CUGCUGUUUU (20 pmol) were chemically synthesized (Integrated DNA Technologies (IDT)) and 20 pmol RNA was radioactively labeled using [γ-^{32}P]-ATP (PerkinElmer, 10mCi/ml) and 10 units T4 Polynucleotide Kinase (PNK) enzyme (New England Biolabs, cat#M0201S) for 2 hours at 37°C in 1X T4 PNK buffer (New England Biolabs, cat#B0201S). Radioactively labeled tRNAs were produced by T7 *in vitro* transcription in the presence of [α-^{32}P]-UTP (PerkinElmer, 10mCi/ml) using PCR products containing an upstream T7 promotor (see DNA constructs). Dephosphorylation of the 5′-triphosphate pre-tRNA was done using 5 units QuickCIP (New England Biolabs, cat#M0525S) for 30 minutes at 37°C in 1X rCutSmart Buffer (New England Biolabs, cat# B6004S), followed by Trizol extraction. All radioactively labeled RNAs were purified on a denaturing polyacrylamide gel and eluted in 0.5M NaCl overnight at room temperature.

His-tagged proteins hLa, Mlp1 and Mlp1 mutants were purified from *E. coli* BL21 cells using Co^{2+} beads, followed by heparin column purification. The proteins were buffer exchanged in 1X PBS and quantified using Bovine Serum Albumin (BSA) quantifications on a SDS polyacrylamide gel.

EMSAs were performed as described (Vinayak et al., 2018). Briefly, 3000 CPM of radioactive RNA substrates (~ 0.1 nM) were incubated in 1X EMSA buffer [10% glycerol, 20 mM Tris-HCl, pH 7.4, 100 mM KCl, 1 mM EDTA, 5 mM β-mercaptoethanol and bromophenol blue] at 95°C for 5 minutes, followed by slow cooling to room temperature. For competition EMSAs, unlabelled RNA was added to the radioactively labeled RNA prior to incubation at 95°C. The concentration of unlabelled RNA depends on the protein concentration used to bind >85% of the radioactively labeled RNA. Protein was added and incubated for 30 minutes at 30°C. The protein-RNA complexes were immediately snap cooled on ice for 5 minutes and separated on a 8% native

polyacrylamide gel at 4°C at 100V. The gels were dried for 45 minutes at 80°C on a Gel Dryer (Bio-Rad) and exposed to a storage phosphor screen overnight. The screens were developed on a Typhoon imager. Quantification of bound and free RNA was done using ImageJ and binding curves were fit using a nonlinear specific binding curve fitting program and Kd values were calculated (GraphPad Prism).

tRNA mediated suppression assay in *Schizosaccharomyces pombe.*

tRNA mediated suppression assays were performed as described previously (Porat & Bayfield, 2020). In brief, the *sla1- S. pombe* ySH9 strain encoding a defective UGA-decoding suppressor tRNA (tRNA-Ser^{UCA}) and the *ade6-704* allele was transformed using a pRep4 plasmid encoding Sla1p, Mlp1 and multiple Mlp1 mutants. Following transformation of *S. pombe* suppressor strains using pRep4 plasmid (*ura4+*), strains were grown on selective media (Edinburgh Minimal Medium (EMM) –ura –leu) and grown in liquid EMM –ura – leu to mid-log phase (OD 0.6-0.9). Spotting was done by transferring 4 µL of liquid culture on low adenine-containing plates (EMM –ura –leu ade10), followed by a 4-day incubation at 32°C. Yeast pellets for protein purification or RNA extraction were obtained by centrifugation of mid-log phase cells at 1800 g for 10 minutes following by two washes with ddH_2O.

Protein extraction was done by resuspension of cell pellets in NET-2 buffer [50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.05% NP40, 1 mM PMSF and 1X PIC], followed by lysis through bead-beating for 2 minutes total (20 sec ON – 20 sec OFF intervals) at 4°C. Cell lysates for protein analysis were obtained following centrifugation for 15 minutes at 20,000 g.

RNA extraction was completed by resuspending the cell pellets in complete RNA extraction buffer A [50 mM sodium acetate, pH 5.1, 10 mM EDTA, 1% SDS], followed by adding 37°C buffer A [50 mM sodium acetate, pH 5.1]-saturated acid phenol and incubation at 65°C for 4 minutes with frequent vortexing. The aqueous top layer was extracted following centrifugation

for 3 minutes at 20,000 g and extracted again using phenol:chloroform:isoamyl alcohol (25:24:1). RNA was precipitated from the aqueous layer by ethanol precipitation and incubation at -80°C for at least 1 hour.

RNP-immunoprecipitation and 3′-RACE in *Schizosaccharomyces pombe.*

Sla1p- and Mlp1-transformed *S. pombe sla1-* strains (ySH9) were grown to mid-log phase (OD 0.6-0.9) in EMM –ura –leu. The culture was cross-linked at 200 RPM in 0.5% formaldehyde at room temperature for 20 minutes, followed by adding 200 mM glycine for 10 minutes. The yeast pellet was collected by centrifugation for 10 minutes at 4000 RPM (Beckman Coulter JLA9.1000 rotor), washed with ddH_2O and collected by centrifugation for 10 minutes at 1800 g, followed by one wash in resuspension buffer [1.2% polyvinylpyrrolidone (PVP), 20 mM HEPES, pH 7.4, 1 mM PMSF, 1X PIC, 1 mM DTT]. The yeast pellet was flash frozen in liquid nitrogen as a continuous bead, followed by cryogrinding in liquid nitrogen using a mortar and pestle. Yeast powder was lysed in RNP buffer [20 mM HEPES, pH 7.4, 110 mM potassium acetate, 100 mM NaCl, 0.5% Triton-X100, 0.1% Tween-20, 1 mM PMSF, 1X PIC and 0.05 U/µL RNase inhibitor]. The lysate was clarified by centrifugation at 20,000 g for 10 minutes at 4°C. Next, 0.005 U/µL TurboDNase (ThermoFisher Scientific AM2238) was added, and the cell lysate was pre-cleared using Protein G Dynabeads coated with rabbit isotype IgG control rotating 1 hour at 4°C. Immunoprecipitations were performed using an affinity purified rabbit anti-Mlp1 and anti-Sla1p antibody (ThermoFisher Scientific – Custom Antibodies) coated to Protein G Dynabeads rotating 1 hour at 4°C (12 µg / 50 µL beads slurry). As a control, the same antibodies were used for immunoprecipitations from ySH9 transformed with empty pRep4 plasmid. The Protein G Dynabeads were washed 5 times using RNP buffer (without PIC, PMSF and RNase inhibitor). RNA was isolated following reverse cross-linking for 45 minutes at 70°C in buffer B [50 mM Tris-HCl pH 7.4, 5 mM EDTA pH 8.0, 10 mM DTT, 1% SDS], followed by Trizol extraction.

RNA samples were polyadenylated and reverse transcribed into cDNA using qScript® microRNA cDNA Synthesis Kit (QuantaBio). Using a pre-tRNA intron-specific forward primer (**Supplementary Data 3**), the substrate of interest was PCR-amplified using *Taq* DNA polymerase in combination with the reverse PerfeCTa Universal PCR Primer (QuantaBio) which anneals with the oligo-dT adapter sequence incorporated during cDNA synthesis. The PCR products were purified through a 1% agarose gel and ligated into a pGEM-T Easy Vector Systems (Promega) plasmid followed by transformation in *E. coli* cells. Plasmid DNA was extracted, and sequences determined by Sanger sequencing at the Hospital for Sick Children – The Centre for Applied Genomics (TCAG).

tRNA 5′-leader and 3′-trailer computational analysis.

A custom python script was used to scrape tRNA information for *H. sapiens, M. musculus, D. melanogaster, A. thaliana, S. cerevisiae* and *S. pombe* from the Genomic tRNA Database (GtRNAdb) (Chan & Lowe, 2016) and *T. thermophila* sequences were obtained from the UCSC Genome Browser (https://genome.ucsc.edu/) (**Supplementary Data 2**). The number of tRNA genes encoded in the genome was determined for different eukaryotes for each isotype and isoacceptor (**Supplementary Data 1**). Trailer lengths were calculated as the number of nucleotides found between the discriminator base, the last annotated mature tRNA nucleotide, and a stretch of minimum four Ts in the genomic DNA. Trailer lengths longer than 20 nucleotides were excluded from the analysis. Mature tRNA sequence lengths were determined starting at nucleotide +1 and ending at discriminator base N73, excluding introns (official tRNA numbering as described previously (Sprinzl et al., 1998)). Statistical significance ($P < 0.05$) was calculated using a one-way ANOVA and Tukey′s multiple comparison test. The 5′-leader logo were generated for *H. sapiens, S. pombe* and *T. thermophila* using WebLogo (Crooks et al., 2004) and divided based on discriminator base identity.

Immunofluorescent staining in *Tetrahymena thermophila.*

T. thermophila wild type and partial Mlp1 knockout strains were grown to mid-log phase and prepared for indirect immunofluorescent staining as described (Wenkert & Allis, 1984). Centrifugation steps including washes were performed for 3 minutes at 1000 g. Briefly, fixative (two parts saturated $HgCl_2$ and one part 95% ethanol) was added to the cell suspension and incubated for 5 minutes at room temperature, followed by collecting and resuspending the cell pellet in 100% ice-cold methanol twice. The cell pellet was washed with PBS and incubated with primary rabbit anti-Mlp1 antibody (1:500 dilution) rotating overnight at 4°C. The cell pellet was washed three times with PBS, followed by incubation with secondary goat anti-rabbit IgG (1:10,000, ThermoFisher Scientific A11008) rotating for 1 hour at room temperature. The cell pellet was washed three times with PBS. The cell suspension was dropped on a coverslip, airdried and mounted with Vectashield Antifade Mounting Medium with DAPI (Vector Laboratories) onto a microscope slide. Microscopy images were obtained at 63x magnification on a LSM700 confocal laser scanning microscope (Zeiss) and processed in ZEN3.3 (blue edition).

Northern blotting.

RNA was obtained from storage solution at -80°C [100% ethanol slurry containing 150 mM sodium acetate, pH 5.1 and 30 µg GlycoBlue Coprecipitant (ThermoFisher Scientific AM9515)] by centrifugation at 20,000 g for 10 minutes. The RNA pellet was washed with 70% ethanol and airdried, followed by resuspension in RNase-free ddH_2O. Samples were prepared by adding an equal volume of 2X formamide loading dye [80% deionized formamide, 0.06% (w/v) bromophenol blue, 0.06% (w/v) xylene cyanol, 10 mM EDTA, pH 8.0], followed by incubation at 95°C for 5 minutes and snap cooling on ice. RNA samples were separated on a 7 M urea denaturing polyacrylamide gel at 4°C at 100V and transferred onto a charged nylon transfer

membrane (PerkinElmer), followed by cross-linking using a UV Stratalinker and drying at 80°C for 15 minutes on a Gel Dryer (Bio-Rad).

The membranes were hybridized for 2 h at the corresponding probe melting temperature (Tm) -10°C in hybridization buffer [6X SSC (1X SSC: 150 mM NaCl and 15 mM sodium citrate), 1% SDS and 4X Denhardt′s solution (Bio Basic D0062)], followed by overnight incubation with T4 PNK ^{32}P-labeled probes (**Supplementary Data 3**). Following three 20-minute washes in wash buffer [2X SSC and 0.1% SDS] the membrane was exposed to a storage phosphor screen overnight and developed on a Typhoon imager. To strip hybridized probe, membranes were incubated three times with stripping buffer [0.1X SSC and 0.1% SDS] for 20 minutes each at 70°C.

Western blotting.

Protein concentration was obtained using Bradford assays (ThermoFisher Scientific, cat#23300). Protein samples were resuspended in 1X loading dye [5X loading dye: 5% β-mercaptoethanol (v/v), 0.02% bromophenol blue (w/v), 30% glycerol (v/v), 10% SDS (w/v), 250 mM Tris-HCl, pH 6.8] and were incubated at 95°C for 10 minutes, separated by electrophoresis on a 12% SDS polyacrylamide gel and transferred onto a nitrocellulose membrane. The membrane was blocked in 0.5% (w/v) skim milk powder in Tris-Buffered Saline (TBS) [20 mM Tris-HCl pH 7.4, 150 mM NaCl] + 0.1% Tween20 (TBST) for 1 hour at room temperature (or overnight at 4°C), followed by incubation with primary antibodies in TBST for 1 hour at room temperature (or overnight at 4°C). The membrane was washed 5 times with TBST, followed by incubation with HRP-conjugated secondary antibodies at 1:10,000 dilutions and incubated for 1 hour at room temperature. Primary antibodies used in this study: mouse anti-beta actin (abcam ab8224), rabbit anti-histone H3 (abcam ab1791), mouse anti-His (abcam ab18184) and affinity purified antibodies (ThermoFisher Scientific – Custom Antibodies) rabbit anti-Mlp1 and rabbit

anti-Sla1p. Secondary HRP-coupled antibodies used in this study: goat anti-rabbit IgG (Cell Signaling Technology 7074) and horse anti-mouse IgG (Cell Signaling Technology 7076).

Genomic DNA extraction and Southern blotting in *Tetrahymena thermophila.*

Genomic DNA extraction from *T. thermophila* cell pellets was performed as described previously (Fillingham et al., 2001). Briefly, cell pellets from 25 mL cultures were harvested and resuspended in 0.5 mL 10 mM Tris-HCl, pH 7.4, followed by addition of 3.5 mL urea buffer [42% (w/v) urea, 350 mM NaCl, 10 mM Tris, pH 7.4, 10 mM EDTA, 1% SDS, 0.1 mg/mL Proteinase K] and incubated for 5 minutes at 50°C. DNA was extracted twice with an equal volume of phenol:chloroform:isoamylalcohol (25:24:1), followed by chloroform:isoamylalcohol (24:1) extraction. One-third volume of 5 M NaCl was added to the aqueous phase and the DNA was precipitated with an equal volume of isopropanol. The DNA was pelleted by centrifugation at 20,000 g for 10 minutes at 4°C and the DNA pellet resuspended in 50 µL Tris-EDTA (TE), pH 8.0 [10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0]. The DNA suspension was treated with RNase A (10 mg/mL) overnight at 55°C and stored at -20°C.

Genomic DNA from wild type and partial Mlp1 knockout strains was digested with *EcoRI* restriction enzyme overnight at 37°C and separated by electrophoresis through a 1% agarose gel. The DNA was transferred onto a nitrocellulose membrane by capillary forces, followed by cross-linking using a UV Stratalinker and drying for 15 minutes at 80°C on a Gel Dryer (Bio-Rad). The membrane was probed using a T4 PNK ^{32}P-labeled PCR product as described in "Northern blotting".

Data reproducibility and statistics.

Mlp1 RIP Northern blot and associated Western blot analysis were performed in biologically independent triplicates. Mlp1 partial KO strain Northern blots and associated Western blots were performed in biologically independent triplicates. TGIRT-sequencing of size-excluded

Mlp1-RIP, WT and Mlp1 partial KO strains were performed in biologically independent triplicates. EMSAs and competition EMSAs were performed in at least independent experimental duplicates unless stated otherwise in figure legends. tRNA-mediated suppression assays in *S. pombe* ySH9 strains and associated Northern blots and Western blots were performed in biologically independent triplicates unless stated otherwise in the figure legend. RNP-immunoprecipitation and associated Western blots from ySH9 and Sanger sequencing of clonal isolates derived from 3′-RACE were performed in biologically independent duplicates. Statistical analysis was performed on data derived from three or more independent replicates. Error bars represent the mean +/- standard deviation (S.D.) of at least three independent replicates. Comparison of tRNA features in different eukaryotes was done using the one-way ANOVA for 3′-trailer length ($F = 289.3$, $DF=6$, P value < 0.0001) and mature tRNA length ($F = 0.9931$, $DF=6$, P value $= 0.428$).

Data availability

TGIRT-sequencing has been deposited to the Gene Expression Omnibus (GEO) under the accession number GSE199642. NCBI Mlp1 protein sequence: https://www.ncbi.nlm.nih.gov/protein/XP_001019287.2/; MLP1 reference gene: http://ciliate.org/index.php/feature/details/TTHERM_00384860. Source data are provided with this paper.

Code availability

Code for the analysis of TGIRT-sequencing and tRNA read fishing and binning into pre-tRNAs and mature tRNAs available in a Snakemake workflow that can be found at https://github.com/etiennefc/t_thermophila_RNA_Seq.git. Code for scraping tRNA sequences from the tRNA database http://gtrnadb.ucsc.edu/ is available at https://github.com/brunohenderyckx/trna-database-scraper.

2.6. Supplementary Figures and Tables

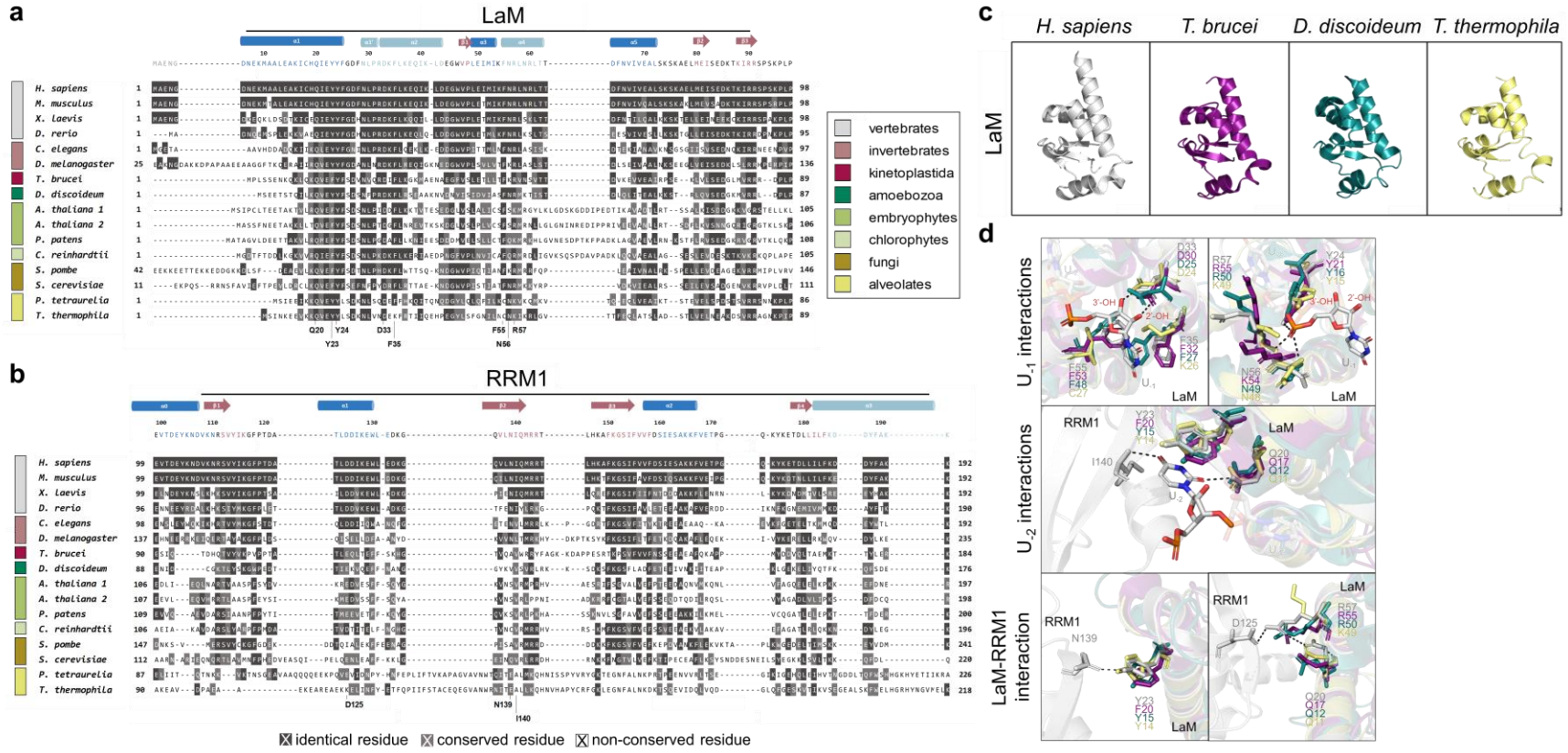


Figure legend on next page

Supplementary Figure 1. Multiple sequence alignments of La proteins from several eukaryotes demonstrate high conservation of the LaM and absence of the RRM1 in alveolates.

a,b Primary amino acid alignments of full RNA-binding domains LaM (**a**) and RRM1 (**b**) from different eukaryotic lineages. A dark grey background indicates identical residues, light grey conserved residues and white indicates no conservation. Most uridylate binding residues are conserved in the LaM of alveolates, whereas residues in the RRM1 are more variable. Important uridylate binding residues are shown at the bottom. The secondary structure motif of the high-resolution hLa protein structure is shown at the top with β-sheets shown in red and α-helices in blue (dark blue: typical α-helices found in the winged-helix fold and classic RRM, light blue: inserted α-helices found in La proteins specifically) (PDB: 2VOD). **c** High-resolution structures of the LaM in different species: *H. sapiens* in white (PDB: 2VOD) (Kotik-Kogan et al., 2008), *T. brucei* in purple (PDB: 1S29) (Dong et al., 2004), *D. discoideum* in teal (PDB: 2M5W) (Apostolidi et al., 2014) and the predicted tertiary structure *for T. thermophila* in yellow (Zheng et al., 2019). **d** Magnified views of La-U_{-1}, La-U_{-2} and LaM-RRM interactions. U_{-1} is the most 3′-terminal uridylate possessing 2′-OH and 3′-OH ends. Most uridylate binding residues located in the LaM are conserved in *T. thermophila,* except for F35 and F55 (human La numbering) which bind the most 3′-terminal U_{-1}. Carbon atoms within the RNA (U_{-1} and U_{-2}) from the *H. sapiens* La structure are colored in white with other atoms colored by type: oxygen in red; nitrogen in blue and phosphorous in orange. Hydrogen bonds are shown as black dashed lines.

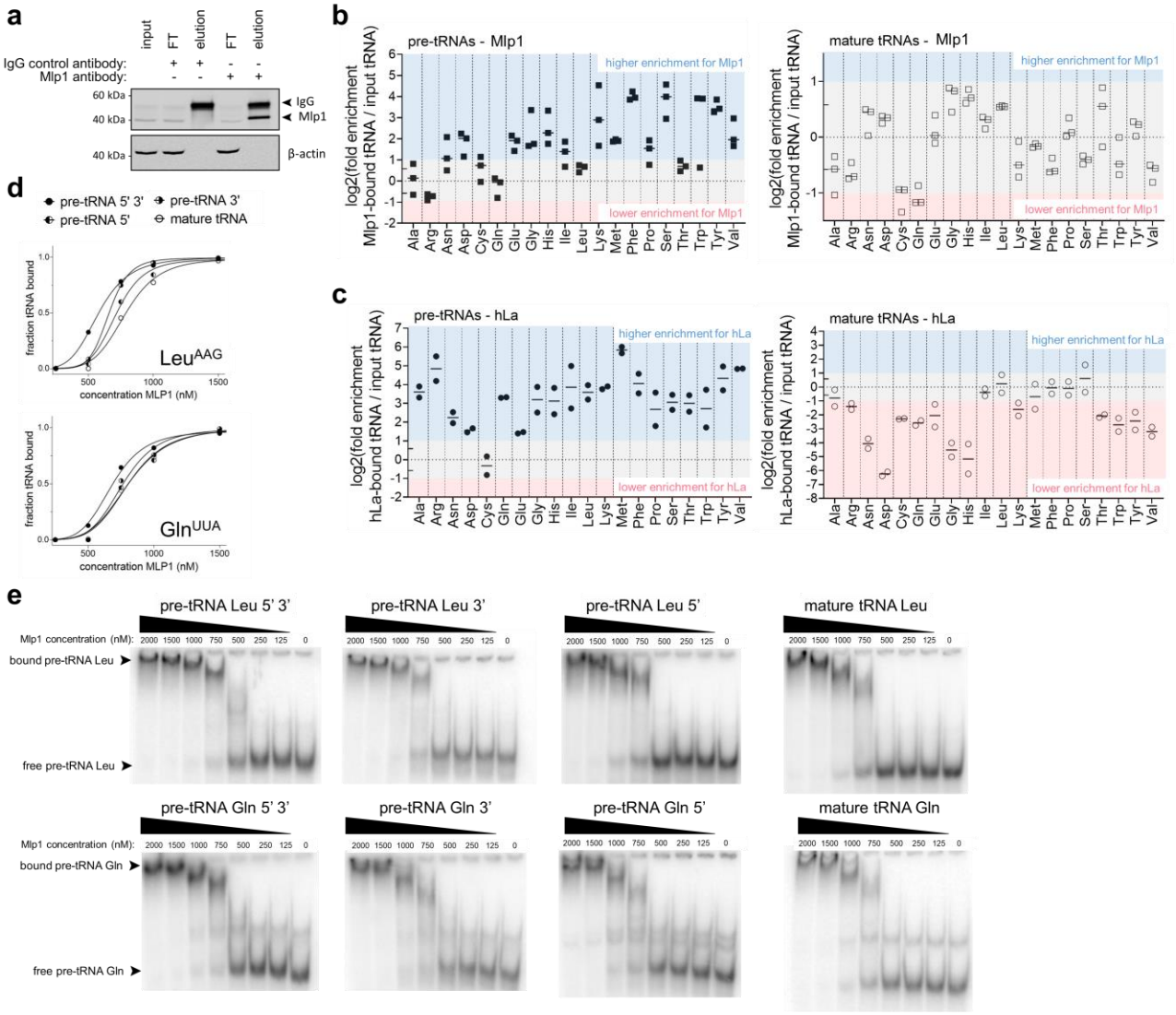
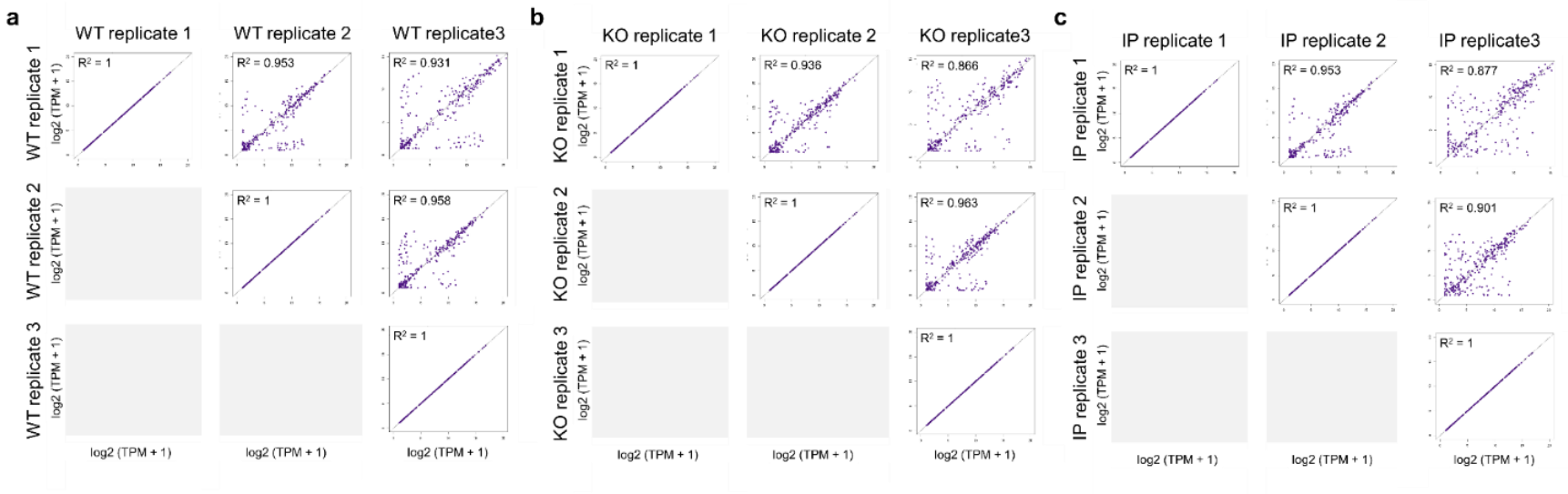


Figure legend on next page

Supplementary Figure 2. Mlp1 demonstrates preferential binding of certain tRNA isotypes and unprocessed pre-tRNAs.

a Western blot confirming Mlp1-specific immunoprecipitation from *T. thermophila* using an affinity purified rabbit anti-Mlp1 antibody and a rabbit isotype immunoglobin G (IgG) control tRNAs (n=3 biologically independent samples). The Mlp1-specific antibody was used for both ribonucleoprotein-immunoprecipitation (RNP-IP) and western blotting. The heavy chain (50 kDa) of the antibodies was detected with the secondary anti-rabbit antibody and shown as IgG. Loading control: β-actin. **b,c** Next generation sequencing data of tRNAs split by tRNA isotypes for Mlp1 (n=3 biologically independent samples) (**b**) and hLa from Gogakos et al. 2017 (n=2 biologically independent samples) (**c**). Fold enrichment is shown as the log2 transformed ratio between Mlp1- or hLa-immunoprecipitated tRNAs and input tRNAs. The small horizontal line represents the average value between replicates. Both Mlp1 and hLa have a higher binding affinity for pre-tRNAs compared to mature tRNAs. **d,e** Binding curves from EMSAs (**d**) and corresponding native gels (**e**) comparing binding affinity of Mlp1 between ^{32}P-labeled pre-tRNA containing 5′- and 3′-extensions or 5′- or 3′-extensions and mature tRNA for Leu^{AAG} and Gln^{UUA} (n=3 independent experiments). The highest binding affinity is found for 5′- and 3′-end containing pre-tRNAs. See **Supplementary Table 1** for Kd quantifications. Source data are provided as a Source Data file.



Supplementary Figure 3. TGIRT-sequencing replicates reproducibility determined by correlation plots.

a-c Biological replicates represented as correlation plots following removal of low count reads. Comparison between replicates for tRNA TPM from wild type (WT) (**a**), partial Mlp1 knockout (KO) (**b**), and Mlp1-associated (**c**) samples. R^2 was calculated using the cor function in R. TPM: transcripts per million.

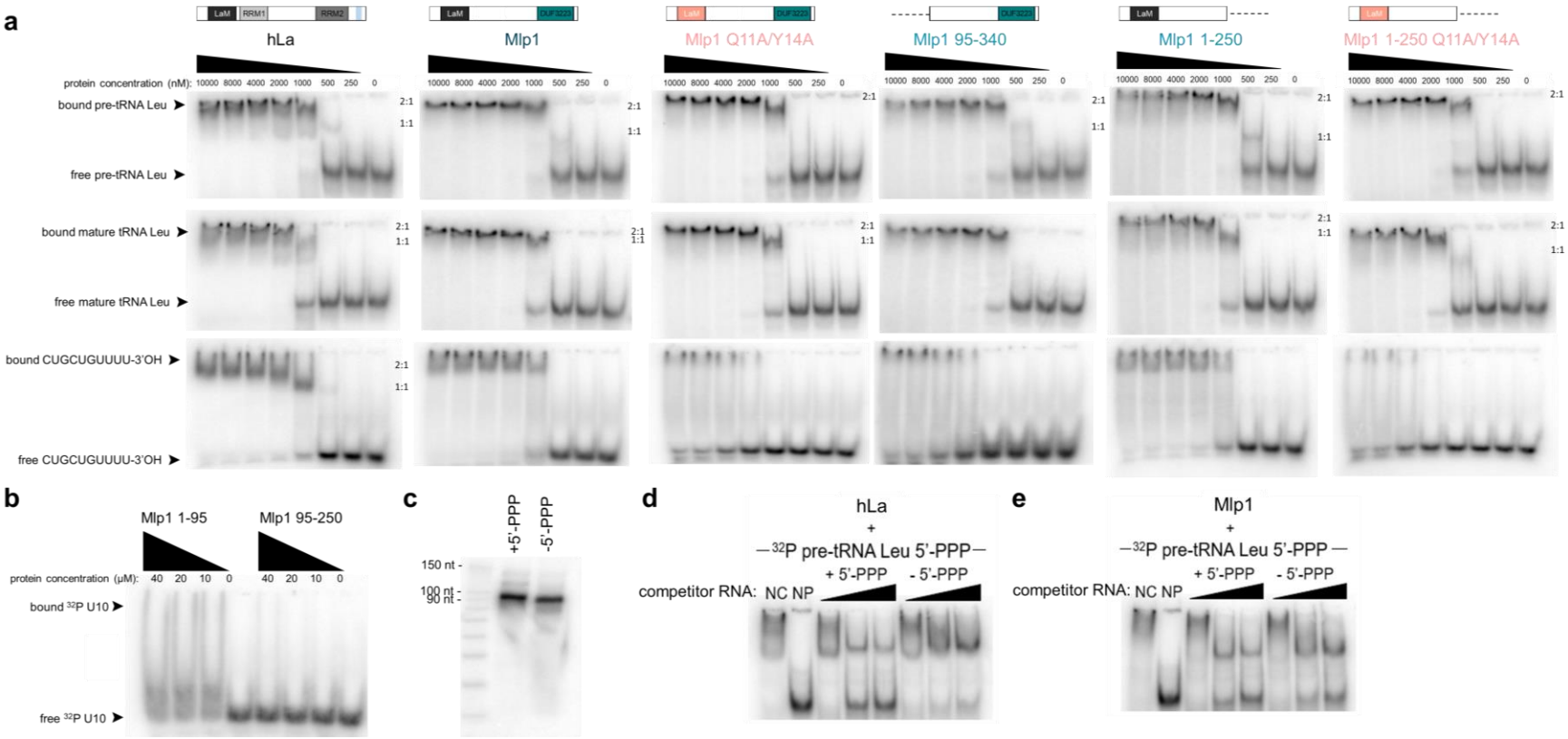
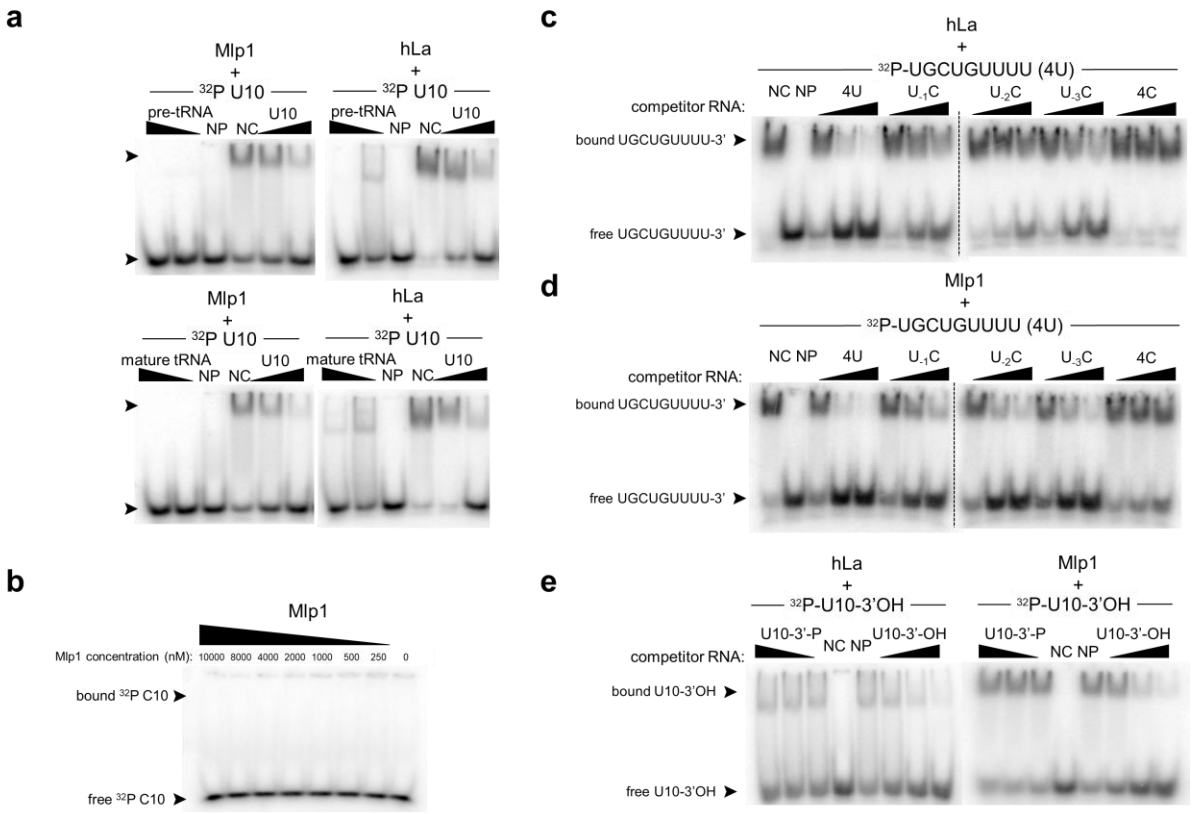


Figure legend on next page

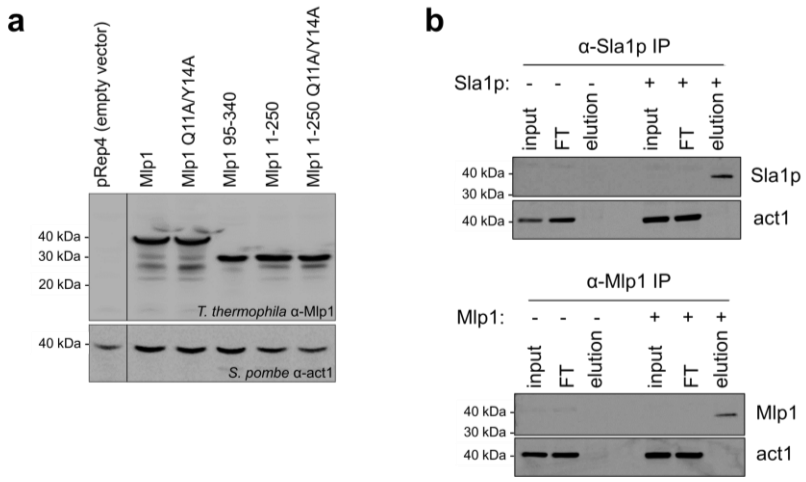
Supplementary Figure 4. Mlp1 binding to typical La protein target RNAs.

a Native gels comparing binding to ^{32}P-labeled pre-tRNAs, mature tRNAs and a 3′-trailer sequence CUGCUGUUUU-3′OH (n≥2 independent experiments). Unbound RNA is found at the bottom of the gel, while protein-RNA complexes shift upwards. The numbering on the side of the gels indicate binding events. A single protein bound to the RNA is indicated at 1:1 and two proteins bound to the same RNA is denoted at 2:1. **b** Native gels comparing binding to ^{32}P-labeled U10 RNA for mutants Mlp1 1-95 (LaM only) and Mlp1 95-250 (middle domain only) (n=2 independent experiments). **c** Denaturing gel of *in vitro* transcribed ^{32}P-labeled 5′-leader containing, 3′-trailer processed pre-tRNAs produced side-by-side as unlabelled-competitor pre-tRNAs (lacking ^{32}P-labeled labeling) used in (**d,e**) to confirm successful removal of the 5′-triphosphate group (n=1 independent experiment). **d,e** Native gels comparing unlabelled-competitors 5′-triphosphate containing pre-tRNA (+5′PPP) and dephosphorylated pre-tRNA (-5′PPP) for ^{32}P-labeled +5′PPP binding on hLa (**d**) or Mlp1 (**e**) (n=2 independent experiments). Competition is seen as a decrease in protein-RNA complex and increase in unbound RNA. NP: no protein lane, NC: no unlabeled competitor lane.



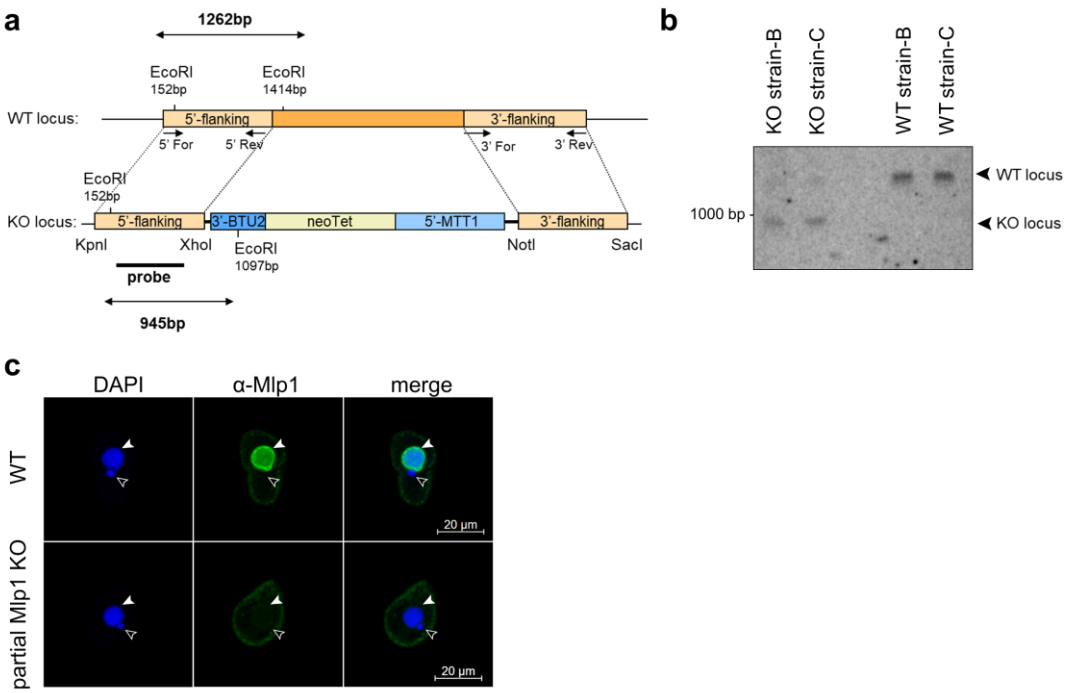
Supplementary Figure 5. Mlp1 binding to uridylate RNA does not discriminate between the position of the uridylate.

a Native gels comparing unlabelled-competitors U10 (positive control), pre-tRNA and mature tRNA for ^{32}P-labeled U10 binding (Mlp1 U10: n=4, tRNA: n=2, hLa U10: n=2, tRNA: n=1 independent experiments). Competition is seen as a decrease in protein-RNA complex and increase in unbound RNA. NP: no protein lane, NC: no competitor lane. **b** Native gel from EMSA for Mlp1 and ^{32}P-labeled C10 (n=2 independent experiments). **c,d** Native gels comparing unlabelled-competitors CUGCUGUUUU (4U), CUGCUGUUUC (U$_{-1}$C), CUGCUGUUCU (U$_{-2}$C), CUGCUGUCUU (U$_{-3}$C) and CUGCUGCCCC (4C) for ^{32}P-labeled 4U wild type RNA binding on hLa (**c**) or Mlp1 (**d**) (n=2 independent experiments). Competition is seen as a decrease in protein-RNA complex and increase in unbound RNA. NP: no protein lane, NC: no competitor lane. **e** Native gels comparing a 3′-phosphorylated substrate (U10-3′-P) and normal 3′-OH containing substrate (U10-3′OH) for binding of ^{32}P-labeled U10-3′-OH. Competition is seen as a decrease in protein-RNA complex and increase in unbound RNA (n=2 independent experiments). No competition was observed with the phosphorylated target for either Mlp1 or hLa. NP: no protein lane, NC: no competitor lane.



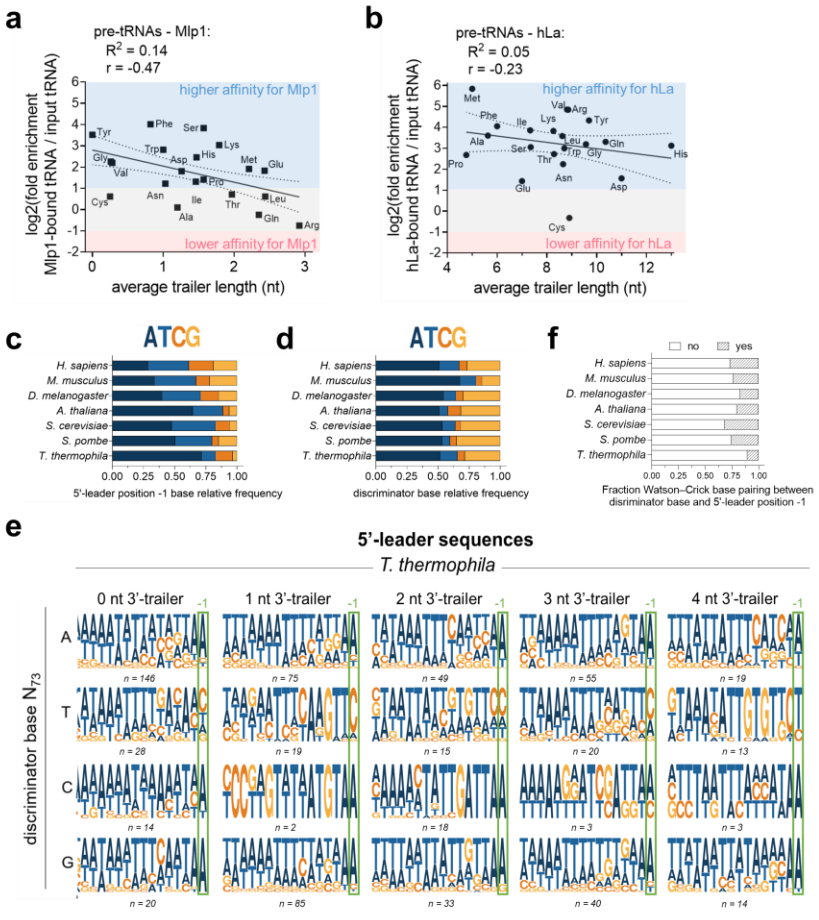
Supplementary Figure 6. tRNA mediated suppression protein expression in *Schizosaccharomyces pombe*.

a Western blot confirming Mlp1 protein expression levels in *S. pombe* following pRep4 plasmid transformation in the tRNA-mediated suppression assay (n=3 biologically independent samples). Loading control: act1. **b** Western blot confirming immunoprecipitation (IP) of Sla1p and Mlp1 from Sla1p- and Mlp1-transformed *S. pombe* ySH9 tRNA suppressor strains. pRep4 empty vector transformed strains were used as a control for background immunoprecipitation using Sla1p and Mlp1 antibodies (n=2 biologically independent samples). Loading control: act1. FT: flow through. Source data are provided as a Source Data file.



Supplementary Figure 7. Confirmation of the partial Mlp1 *Tetrahymena thermophila* knockout strain.

a Schematic overview of the *T. thermophila* wild type (WT) locus encoding endogenous Mlp1 and knockout (KO) locus following homologous recombination between 5′- and 3′-flanking regions of the genome with transformation plasmid encoding the neomycin (neoTet) selection marker. Restriction enzymes used for cloning 5′- and 3′-flanking regions for homologous recombination are shown below the KO locus. Genomic DNA digest before Southern blotting was performed with *EcoRI* restriction enzyme generating a 1262 bp fragment for the WT locus and a 945 bp fragment for the KO locus. **b** Southern blot of genomic DNA digested with restriction enzyme *EcoRI* probed with ^{32}P-labeled PCR-generated probe reveals almost complete KO of Mlp1 (see schematic in (**a**)) (n=2 biologically independent samples). **c** Indirect immunofluorescent staining of *T. thermophila* WT and partial Mlp1 KO strains using the polyclonal rabbit anti-Mlp1 antibody. Nuclei were stained using DAPI (n=2 biologically independent samples). Full white arrows are denoting the transcriptionally active macronucleus and empty white arrows show the transcriptionally inactive micronucleus. Source data are provided as a Source Data file.



Supplementary Figure 8. The effect of short 3′-trailer sequences on 5′-leader composition and Mlp1 binding affinity.

a,b Next generation sequencing data of Mlp1-bound pre-tRNAs (n=3 biologically independent samples) (**a**) and hLa-bound pre-tRNAs (n=2 biologically independent samples) (**b**) split by tRNA isotypes compared against 3′-trailer lengths. **c** Distribution of the nucleotide at the most 3′-terminal position in the 5′-leader (N_{-1}) of pre-tRNAs from different eukaryotes. *T. thermophila* has a high frequency of adenosines in this position. **d** Distribution of the nucleotide found as the discriminator base (N_{73}) in different eukaryotes. Distribution is equal between different species. **e** Logo analysis of *T. thermophila* 5′-leader sequences split by the discriminator base identity and 3′-trailer length. the most 3′-terminal position in the 5′-leader (N_{-1}) is highlighted with a green box. The number of analyzed pre-tRNAs is shown underneath each logo. **f** Distribution of pre-tRNAs containing a perfect Watson-Crick base pairing between the discriminator base (N_{73}) preceding the 3′-trailer and most 3′-terminal 5′-leader nucleotide (N_{-1}). Source data are provided as a Source Data file.

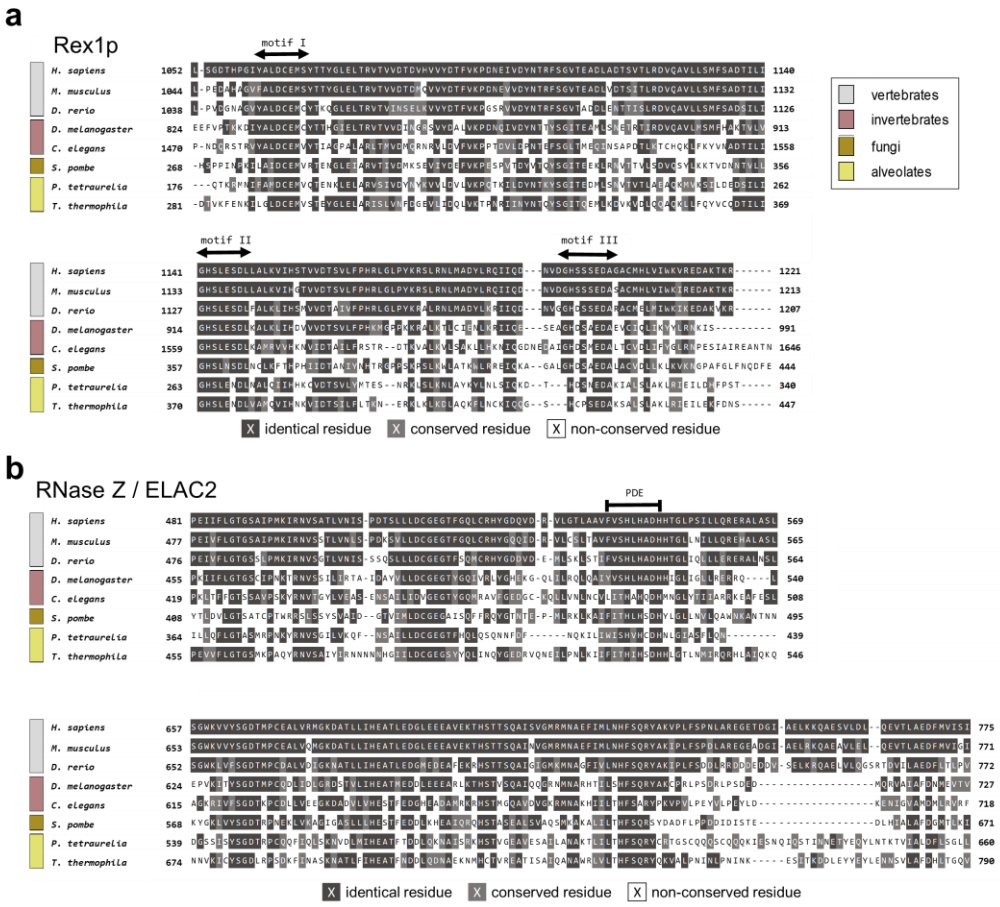


Figure continued on next page

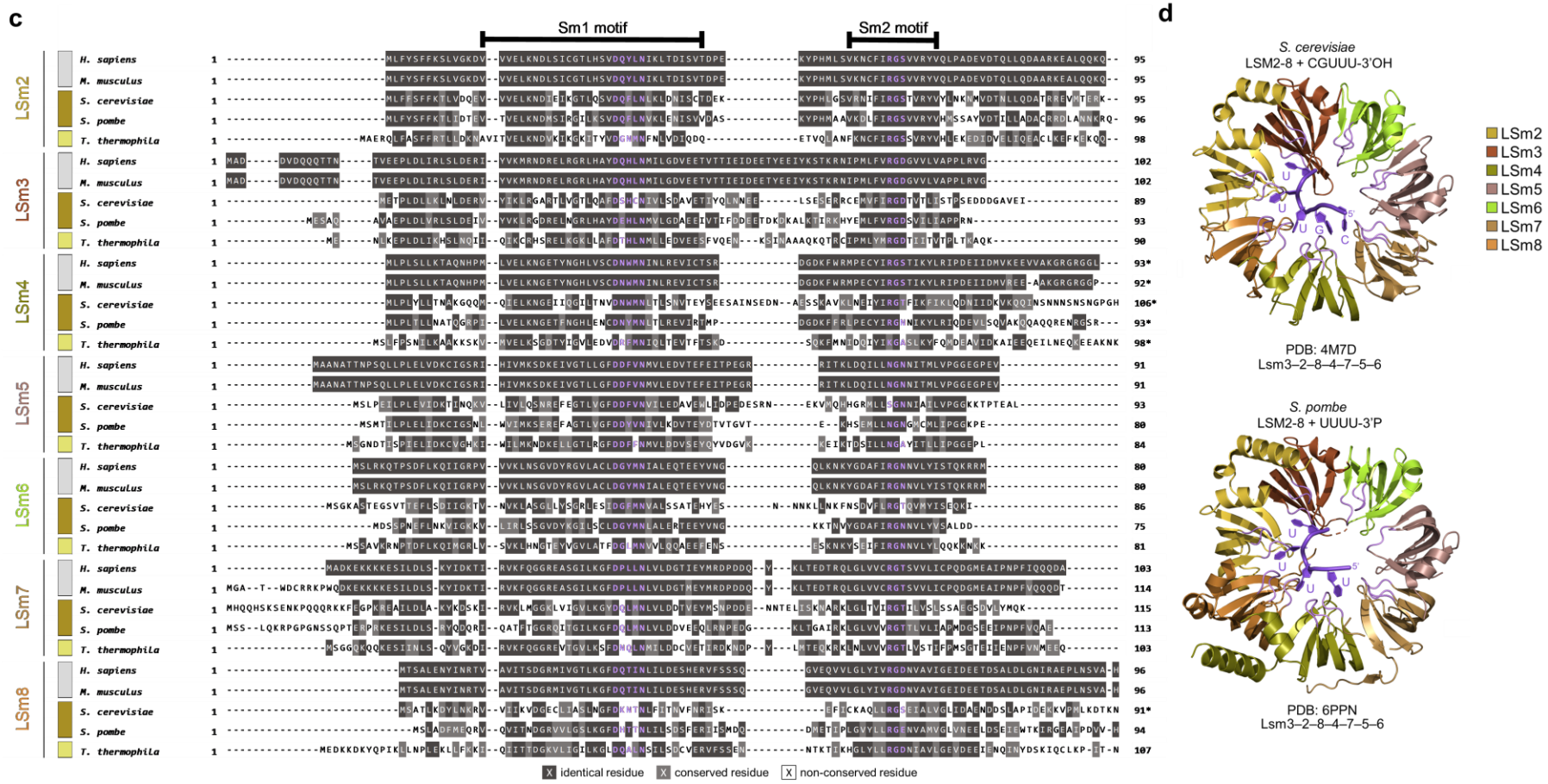


Figure legend on next page

Supplementary Figure 9. Primary sequence alignments of the 3′-exonuclease Rex1p, 3′-endonuclease RNase Z / ELAC2 and LSm2-8 complex from different eukaryotic species.

a-c Primary amino acid alignments from different eukaryotic lineages revealing conservation of the 3′-exonuclease Rex1p (a), 3′-endonuclease RNase Z/ELAC2 (b) and LSm2-8 complex (c). Regions important for Rex1p function are highlighted at motif I, motif II and motif III. The RNase Z catalytic domain is shown as PDE. LSm2-8 complex conserved RNA interacting residues are highlighted in purple and conserved Sm1 and Sm2 motifs are shown. A dark grey background indicates identical residues, light grey conserved residues and white indicates no conservation. d High-resolution structures of the LSm2-8 complex in *S. cerevisiae* (PDB: 4M7D) and *S. pombe* (PDB: 6PPN) in complex with 3′-uridylate RNA shown in dark purple. The conserved RNA interacting residues, highlighted in purple in c, are also shown in light purple and are found in the loops.

Supplementary Table 1. Kd values from EMSAs determining binding of Mlp1 to different processed pre-tRNA intermediates and mature tRNA.

K_d: equilibrium dissociation constant, S.D.: standard deviation.

	Leu^{AAG} K_d ± S.D. (nM)	**Gln^{UUA} K_d ± S.D. (nM)**
pre-tRNA 5' 3'	586 ± 15	687 ± 44
pre-tRNA 3'	658 ± 36	740 ± 40
pre-tRNA 5'	714 ± 21	804 ± 16
mature tRNA	831 ± 33	801 ± 62

Supplementary Table 2. Summary of 3′-trailer length in different eukaryotic species calculated using different RNA polymerase III termination signals.

S.D.: standard deviation, nt: nucleotides.

	Termination signal TTTT or $(T)_4$		Termination signal TTTTT or $(T)_5$		Termination signal TTTTTT or $(T)_6$		Termination signal TTTTTT or $(T)_7$	
Species	**Average of 3'-trailer length ± S.D. (nt)**	**Number of tRNAs analysed (3'-trailer lengths with >20 nt excluded)**	**Average of 3'-trailer length ± S.D. (nt)**	**Number of tRNAs analysed (3'-trailer lengths with >20 nt excluded)**	**Average of 3'-trailer length ± S.D. (nt)**	**Number of tRNAs analysed (3'-trailer lengths with >20 nt excluded)**	**Average of 3'-trailer length ± S.D. (nt)**	**Number of tRNAs analysed (3'-trailer lengths with >20 nt excluded)**
H. sapiens	8.1 ± 3.7	412	8.4 ± 3.6	140	8.2 ± 3.9	65	7.4 ± 3.7	45
M. musculus	7.9 ± 3.6	343	7.9 ± 3.5	141	7.9 ± 3.4	55	7.6 ± 3.0	39
D. melanogaster	8.5 ± 4.0	247	8.2 ± 3.9	214	8.1 ± 3.9	162	7.1 ± 3.4	79
A. thaliana	4.6 ± 4.3	532	4.7 ± 4.3	441	5.2 ± 4.4	251	4.9 ± 4.1	133
S. cerevisiae	3.0 ± 2.6	269	3.5 ± 3.0	266	3.6 ± 3.2	214	3.9 ± 3.6	132
S. pombe	3.3 ± 2.8	162	4.3 ± 3.8	154	3.9 ± 3.7	86	4.2 ± 4.5	36
T. thermophila	1.5 ± 1.5	686	1.7 ± 1.9	674	2.2 ± 2.9	545	2.5 ± 3.3	312

Supplementary Table 3. NCBI accession numbers of primary amino acid sequences used for conservation analysis.

NCBI: National Center for Biotechnology Information.

		La	Rex1p	RNase Z / ELAC2	LSm2	LSm3
Vertebrates	*Homo sapiens*	NP_003133.1	NP_065746.3	NP_060597.4	NP_067000.1	NP_055278.1
	Mus musculus	NP_033304.1	NP_080128.2	NP_001349912.1	NP_001103571.1	NP_080585.1
	Xenopus laevis	NP_001081021.1				
	Danio rerio	NP_955841	NP_001119888.1	NP_001243133.1		
Invertebrates	*Caenorhabditis elegans*	NP_491411.1	NP_498135.2	NP_001023110.1		
	Drosophila melanogaster	NP_477014.1	NP_001034073.1	NP_724916.1		
Amoebozoa	*Dictyostelium discoideum*	XP_640625.1				
Kinetoplastid	*Trypanosoma brucei*	XP_822491.1				
Embryophytes	*Arabidopsis thaliana*	NP_567904.1 NP_178106.2				
Chlorophytes	*Physcomitrella patens*	XP_024359573.1				
	Chlamydomonas reinhardtii	XP_042914771.1				
Fungi	*Schizosaccharomyces pombe*	NP_593315.1	NP_594627.2	NP_595514.1	NP_588459.1	NP_595747.1
	Saccharomyces cerevisiae	NP_010232.1			NP_009527.1	NP_013543.3
Alveolates	*Paramecium tetraurelia*	XP_001439258.1	XP_001435945.1	XP_001346883.1		
	Tetrahymena thermophila	XP_001019287.2	XP_001033374.2	XP_001031902.2	XP_012654399	XP_012656392.1

		LSm4	LSm5	LSm6	LSm7	LSm8
Vertebrates	*Homo sapiens*	NP_036453.1	NP_036454.1	NP_009011.1	NP_057283.1	NP_057284.1
	Mus musculus	NP_056631.2	NP_079796.1	NP_001177933.1	NP_001346078.1	NP_598700.1
	Xenopus laevis					
	Danio rerio					
Invertebrates	*Caenorhabditis elegans*					
	Drosophila melanogaster					
Amoebozoa	*Dictyostelium discoideum*					
Kinetoplastid	*Trypanosoma brucei*					
Embryophytes	*Arabidopsis thaliana*					
Chlorophytes	*Physcomitrella patens*					
	Chlamydomonas reinhardtii					
Fungi	*Schizosaccharomyces pombe*	NP_001342832.1	NP_596373.1	NP_594380.1	NP_588340.1	NP_588509.1
	Saccharomyces cerevisiae	NP_011037.3	NP_011073.1	NP_010666.2	NP_014252.2	NP_012556.2
Alveolates	*Paramecium tetraurelia*					
	Tetrahymena thermophila	XP_001008519.2	XP_012655775.1	XP_012653392.1	XP_001031297.1	XP_001470817.1

Chapter III:

Genuine La protein Mlp1 functions in U4/U6 di-snRNP assembly and affects splicing in *Tetrahymena thermophila*

Author contributions

Genuine La protein Mlp1 functions in U4/U6 di-snRNP assembly and affects splicing in *Tetrahymena thermophila*

Kerkhofs, Kyra; Laganas, Jarred, Pearlman, Ronald E.; Bayfield, Mark A.

Department of Biology, Faculty of Science, York University, Toronto, Ontario N3J 1P3, Canada,

Manuscript in preparation

Author contributions:

Jarred Laganas and Ronald E. Pearlman provided annotated genomic locations of snRNAs in *Tetrahymena thermophila*.

3.1. Abstract

Splicing of pre-mRNAs removing non-coding intron sequences is an essential step in pre-mRNA maturation and is catalyzed by the spliceosome. The spliceosome consists of numerous protein components and the U1, U2, U4, U5 and U6 small nuclear RNAs (snRNAs) which are assembled into ribonucleoprotein (RNP) complexes. During splicing, the snRNPs function individually (U1 and U2), as a heterodimer (U4/U6) or as a heterotrimer (U4/U6·U5). Formation of the dimeric U4/U6 snRNP (di-snRNP) involves extensive RNA-RNA interactions between U4 and U6 snRNA. Previous work in the yeast *Saccharomyces cerevisiae* demonstrated that biogenesis of RNA polymerase II transcribed snRNAs involves exo- and endonucleolytic processing into a mature snRNA and addition of a 2,2,7-trimethyl guanosine (m_3G) cap by the trimethylguanosine synthase-1 (Tgs1) enzyme. snRNA processing intermediates contain a 3′-terminal stretch of uridylates which is a high-affinity binding site for La proteins. We found that *Tetrahymena thermophila* nascent snRNAs lack an internal uridylate sequence and that instead, the *Tetrahymena thermophila* La protein, Mlp1, interacts with mature m_3G-cap containing snRNAs. More specifically, we found that Mlp1 interacts with the U4/U6 di-snRNP and that Mlp1 depletion affects the assembly of this complex and diminishes splicing efficiency of pre-mRNAs. Our work is consistent with a functional role for the genuine La protein Mlp1 in pre-mRNA splicing in *Tetrahymena thermophila*.

3.2. Introduction

Protein coding genes in eukaryotes are transcribed by RNA polymerase II in the nucleus as precursor messenger RNAs (pre-mRNAs), and the majority contain non-coding intron sequences that interrupt the coding sequence. Removal of these introns is required for translation of functional proteins in a process known as splicing which is catalyzed by the spliceosome in the nucleus (Gehring & Roignant 2021). The spliceosome contains more than 100 proteins and five uridine-rich small nuclear RNAs (U snRNAs), namely the U1, U2, U4, U5 and U6 snRNAs (Chen & Moore, 2015) which are assembled through a multi-step processes in multiple subcellular compartments into four large small nuclear ribonucleoprotein (snRNP) complexes: the U1, U2, U4/U6 and U5 snRNPs (Gehring & Roignant, 2021). The snRNPs function in catalyzing the splicing reaction by excising the intron through recognition of conserved sequences at the intron boundaries (Villa et al., 2002). The average number of introns per gene differs strongly between different eukaryotic species (Xiong et al., 2012). For example, humans contain 12.4 introns per gene, while the yeast *Saccharomyces cerevisiae* contains only 0.06 introns per gene. The alveolates *Paramecium tetraurelia* and *Tetrahymena thermophila* contain 2.3 and 3.6 introns per gene, respectively (Xiong et al., 2012), indicating that most mRNA transcripts require splicing.

The snRNPs function either individually (U1 and U2), as a heterodimer (U4/U6) or as a heterotrimer (U4/U6·U5). The first step involves complementary base-pairing of the U1 and U2 snRNP with the conserved 5′-splice site and the branch point sequences of the intron, respectively (**Figure 21** – A complex). Early in the spliceosome assembly steps, the U4 and U6 snRNPs form a single particle through extensive RNA-RNA base-pairing to form the dimeric U4/U6 snRNP (di-snRNP) (**Figure 22a**). Then, the U4/U6 complex associates with the U5 snRNP forming the trimeric U4/U6·U5 snRNP (tri-snRNP) (**Figure 22b**), which in turn

associates with the assembling spliceosome on the pre-mRNA (**Figure 21** – pre-catalytic B complex) leading to rearrangements resulting in dissociation of the U1 and U4 snRNPs, the formation of new extensive RNA-RNA interactions between the U6 and U2 snRNPs, and interactions between the pre-mRNA and U6 and U5 (**Figure 21** – active spliceosome B^{act}/B* complex). Following formation of the active spliceosome, the first transesterification splicing reaction occurs resulting in formation of the C complex which then catalyzes the second splicing step resulting in excision of the lariat intron and associated U2, U5 and U6 snRNPs (**Figure 21** – P complex) and ligation of the mRNA exons. Once the lariat intron is released, together with remaining U2, U5 and U6 snRNPs, the snRNPs are recycled for assembly of the next catalytically active spliceosome (Gehring & Roignant, 2021).

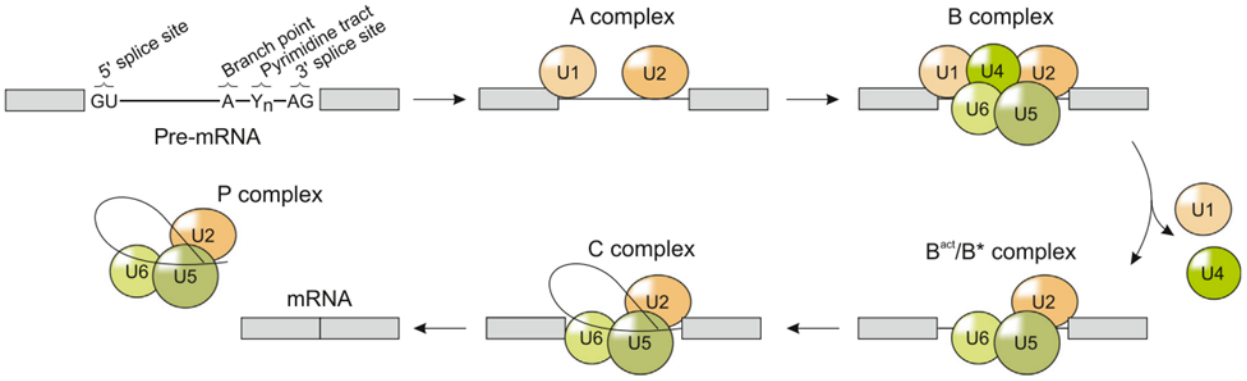


Figure 21. The function of snRNPs in splicing of pre-mRNAs.

Precursor mRNA (pre-mRNA) maturation involves intron excision by the spliceosome. The grey boxes represent coding exons and the black line the non-coding intron sequences. Formation of the A complex is achieved following binding of the U1 snRNP to the 5′-splice site and the U2 snRNP to the branch point, followed by association of the U4/U6·U5 tri-snRNP leading to formation of the B complex. The U4/U6·U5 tri-snRNP is formed through extensive RNA-RNA interactions between U4 and U6 snRNAs, while U5 association is independent of RNA-RNA interactions. Following formation of the pre-catalytically active B complex, multiple rearrangements take place leading to removal of the U1 and U4 snRNPs and formation of the catalytically active B* complex which performs the first transesterification splicing step leading to formation of the C complex. The C complex completes the splicing reaction by a second splicing step leading to removal of the lariat intron and ligation of the mRNA exons. Figure from (Gehring & Roignant, 2021).

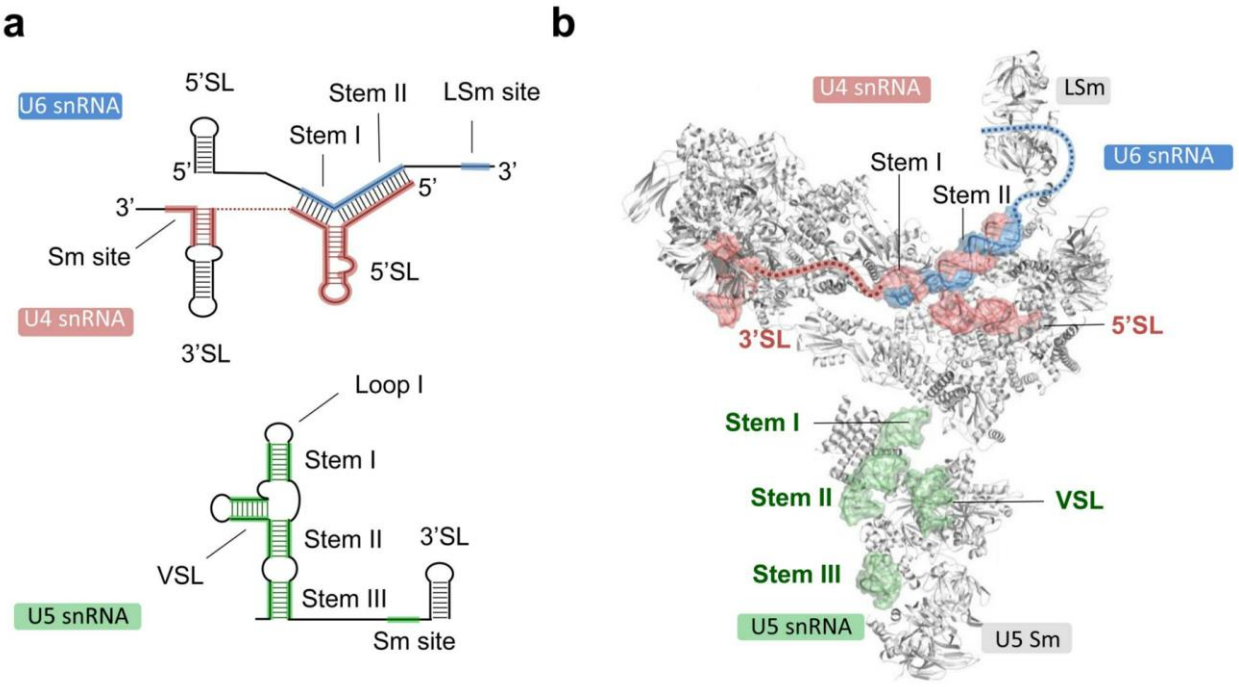


Figure 22. Organization of the U4/U6·U5 tri-snRNP.

a Secondary structure of U4/U6 di-snRNAs and U5 snRNA. Each snRNA is assembled into snRNPs through association with several proteins (not shown): U4 and U5 associate with heteroheptameric rings of Sm proteins and U6 associates with the LSm2-8 ring. In addition, multiple snRNP specific proteins are present. Sm site: conserved binding site for Sm proteins, LSm site: conserved binding site for LSm proteins, SL: stem loop, VSL: variable stem loop. **b** High-resolution structure of a U4/U6·U5 tri-snRNP which is formed through RNA-RNA interactions between the U4 and U6 snRNA, while association of U5 snRNP is independent on RNA-RNA interactions. Figure adapted from (Nguyen et al., 2015).

The spliceosomal snRNAs U1, U2, U4 and U5 are non-polyadenylated transcripts that are transcribed by RNA polymerase II from a specialized promotor (Richard & Manley 2009). Transcription termination in the yeast *S. cerevisiae* is mediated by the non-coding RNA specific Nrd1-Nab3-Sen1 (NNS) complex through co-transcriptional binding of RNA-binding proteins Nrd1 and Nab3 to consensus sequences GUA[A/G] and UCUU, respectively (Carroll et al., 2004). As a result of Nrd1 and Nab3 binding to the nascent snRNA, the RNA polymerase II transcription rate decreases and the nascent snRNA is released from the DNA by the helicase Sen1 (Carroll et al., 2004; Steinmetz et al., 2001; Steinmetz & Brow, 1996). Next, the snRNA

obtains a short poly(A) tail intended to recruit 3′-exonucleases for processing of the 3′-end (Allmang et al., 1999; van Hoof et al., 2000). Alternatively, nascent snRNAs are endonucleolytically processed downstream of the mature snRNA sequence resulting in a 3′-terminal stretch of uridylates (Chanfreau et al., 1997; Seipelt et al., 1999). La proteins have been shown to associate with these truncated nascent snRNAs ending in 3′-uridylate sequences (Stefano, 1984; Xue et al., 2000). Nascent RNA polymerase II transcribed snRNAs obtain a N7-methylguanosine (m^7G) cap which is hypermethylated into a 2,2,7-trimethyl guanosine (m_3G) cap by the trimethylguanosine synthase-1 (Tgs1) enzyme located in the nucleolus of *S. cerevisiae*, as well as multiple post-transcriptional modifications including 2′-O-methylations and pseudouridylations (Karijolich & Yu, 2010; Mouaikel et al., 2002). Addition of the m_3G-cap on snRNA is required for efficient transport back into the nucleus where pre-mRNA splicing occurs (Mouaikel et al., 2002).

In contrast, the Nab3 homologue is not expressed and Senataxin, the DNA/RNA helicase Sen1 homologue, is not involved in snRNA transcription termination in human cells (Suraweera et al., 2009). Instead, transcription termination of snRNAs in vertebrates is mediated by the Integrator Complex which endonucleolytically cleaves the nascent snRNAs upstream of the 3′-box *cis*-acting sequence, enabling termination of RNA polymerase II transcription (Baillat et al., 2005). Following transcription termination, nascent snRNAs are transported to the cytoplasm where they assemble with seven Sm core proteins (**Figure 23**) (Mattaj, 1986). Upon Sm snRNP assembly, the m_3G-cap is synthesized, 3′-terminal additional nucleotides are processed and the snRNP complex is imported into the nucleus (**Figure 23**) (Jády et al., 2003; Mattaj, 1986).

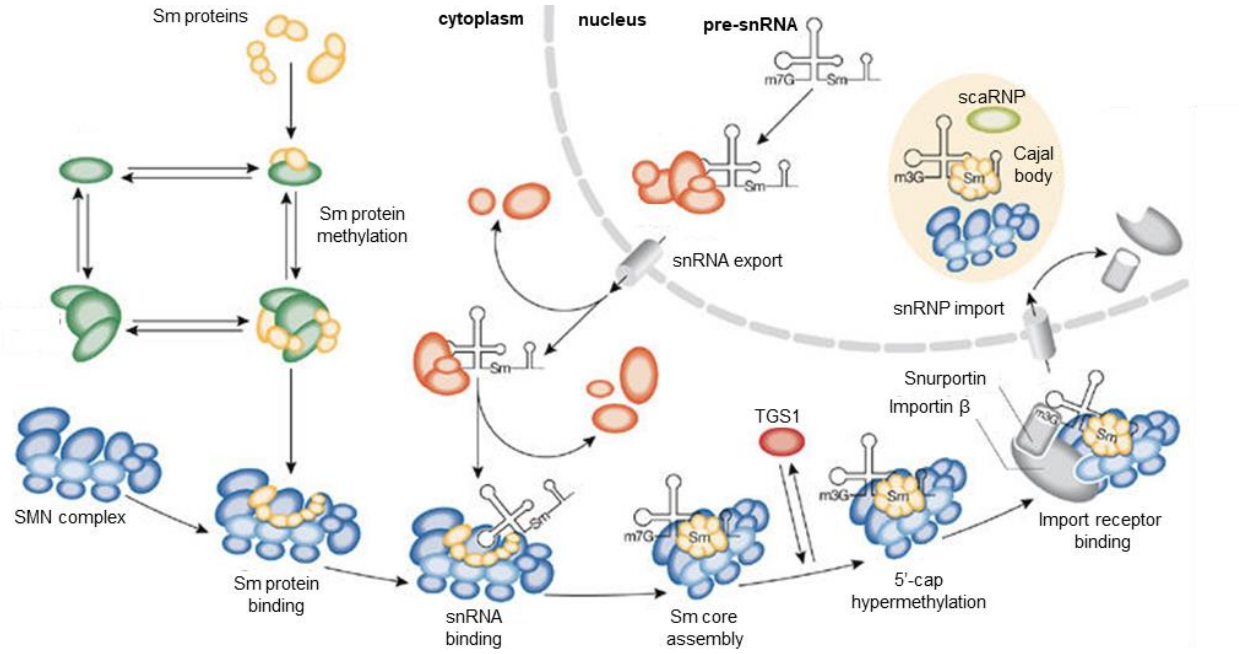


Figure 23. Spliceosomal snRNP complex biogenesis occurs in the multiple subcellular compartments.

In humans, RNA polymerase II transcribed small nuclear RNAs (snRNAs) undergo biogenesis in both the cytoplasm and nucleus, while in yeast snRNA biogenesis takes place in the nucleus and nucleolus (not shown). Following transcription, the nascent N7-methylguanosine (m^7G)-cap containing snRNAs are exported into the cytoplasm, where the SMN complex assembles with seven Sm proteins which then replaces the snRNA export complex on the snRNA. The Sm proteins interact with the conserved Sm-binding sites within the snRNAs. Next, the Sm core assembly interacts with the trimethylguanosine synthase-1 (Tgs1) converting the m^7G cap into a 2,2,7-trimethylguanosine (m_3G) cap structure through hypermethylation (see **Figure 3**), followed by exonucleolytic trimming at the 3′-end. The m_3G-cap structure and the Sm core function as nuclear localization signals leading to recruitment of the Snurportin-1 import adapter and interactions with Importin β both through Snurportin-1 and the Sm core. This assembly results in nuclear import of the small nuclear ribonucleoprotein (snRNP) complex. Further snRNP remodelling take place in Cajal bodies. Figure adapted from (Pellizzoni, 2007).

The spliceosomal snRNA U6 differs from U1, U2, U4 and U5 snRNAs in several aspects. First, U6 snRNA is transcribed by RNA polymerase III and the transcription termination signal is encoded in the genome as a stretch of deoxyadenosines in the coding strand which result in a 3′-terminal stretch of uridylates in the nascent RNA (Dieci et al., 2007; Kunkel et al., 1986; Reddy et al., 1987). Second, while RNA polymerase II transcribed snRNAs obtain a m^7G cap

which is hypermethylated into a m_3G-cap by the Tgs1 enzyme (Mouaikel et al., 2002), U6 snRNA obtains a 5′-triphosphate which is converted into a γ-monomethyl phosphate cap by γ-methyltransferases (Shimba & Reddy, 1994; Singh & Reddy, 1989).

Lastly, each snRNA assembles into snRNP complexes through assembly with a set of snRNP-specific proteins. The Sm proteins are found in each U1, U2, U4 and U5 snRNP complex, with the seven core Sm proteins (B/B′, D_1, D_2, D_3, E, F and G), containing a highly conserved Sm motif that assemble into a ring-shaped heteroheptameric complex binding the consensus Sm-binding site located near the 3′-end of these snRNAs (Raker et al., 1996, 1999). On the other hand, U6 snRNA lacks this consensus Sm-binding site and instead assembles with seven structurally related Sm-Like (LSm) proteins (LSm2-8) that also form a heteroheptameric ring-shaped complex through binding the 3′-uridylate tail (Achsel et al., 1999). In addition to the core proteins which are found in each U1, U2, U4, U5 and U6 snRNP, multiple snRNP specific proteins are present in a particular snRNP (Rappsilber et al., 2002).

Prior to LSm2-8 complex binding to the 3′-terminal end of U6 snRNA, the La protein associates with nascent U6 snRNAs. La proteins are highly conserved eukaryotic RNA-binding proteins that are well known to function as RNA chaperones in small RNA biogenesis, including pre-tRNAs, pre-5S rRNA and U6 snRNA (Rinke & Steitz, 1982, 1985). A common feature of these transcripts is transcription by RNA polymerase III which results in generation of 3′-terminal uridylates during termination to which La proteins bind (Stefano, 1984). Generally, La proteins are the first proteins to associate with nascent RNA polymerase III transcripts, followed by replacement by other proteins as the RNA matures (Guddat et al., 1990; Terns et al., 1992). La proteins are the first proteins to associate with U6 snRNA following RNA polymerase III transcription through binding of the 3′-terminal uridylate tail (Pannone et al., 2001). Following La binding and stabilization of nascent U6 snRNA, the 3′-uridylate tail is elongated and

subsequently trimmed which results in formation of five uridylates and a terminal 2′3′-cyclophosphate, which reduces La protein affinity more than 10-fold leading to replacement of La by the LSm2-8 complex (Achsel et al., 1999; Lund & Dahlberg, 1992; Stefano, 1984; Terns et al., 1992). The La protein is non-essential in yeast (Van Horn et al., 1997; Yoo & Wolin, 1994), however, La knockout yeast strains are synthetically lethal with either loss of function of certain tRNA modification enzymes (Copela et al., 2006) or with a mutation in the anticodon stem of the essential tRNA Ser^{CGA}, indicating an important function of La proteins in correct pre-tRNA folding and stability. Similarly, a lethal phenotype is obtained in La knockout yeast strains combined with a mutation in members of the Sm and LSm protein family (Pannone et al., 1998; Xue et al., 2000), suggesting a shared function between Lhp1p and Sm and LSm proteins. La proteins bind uridylate sequences through coupled binding of the highly conserved La motif (LaM) and adjacent RNA recognition motif-1 (RRM1) (Teplova et al., 2006). A distinct La protein lacking the RRM1 was discovered in alveolates (Deragon, 2020) and functional assays confirmed that despite missing the highly conserved RRM1, the alveolate *T. thermophila* La protein Mlp1 functions as a genuine La protein (Kerkhofs et al., 2022).

In this study, we predict that snRNAs transcribed by RNA polymerase II in *T. thermophila* use a yeast-like transcription termination pathway based on computational observations made in the downstream genomic sequence. We describe the presence of a conserved Nab3 RNA-binding site, however, without presence of a predicted Nrd1 site. Interestingly, we identified a novel UAAUCAAACAAAA motif in close proximity to the Nab3 RNA-binding site. Through related observations, we found that *T. thermophila* nascent RNA polymerase II transcribed snRNAs lack an internal uridylate sequence that could serve as a high-affinity binding site for La proteins. Therefore, we further investigated if Mlp1 would function in snRNA biogenesis. We found that Mlp1 interacts with mature m_3G-cap containing snRNAs by RNP-immunoprecipitations in combination with RNA-sequencing and Northern blotting. We also

demonstrate that Mlp1 interacts with the U4/U6 di-snRNP and that Mlp1 depletion affects the assembly of the U4/U6 di-snRNP and levels of intron retention from pre-mRNAs. Our work is consistent with a role for Mlp1 in pre-mRNA splicing in *T. thermophila*.

3.3. Results

3.3.1. RNA polymerase II transcribed snRNA genes contain a Nab3 binding site and a newly discovered motif at the 3′-end in *Tetrahymena thermophila*

Previous work demonstrated that the *S. cerevisiae* La protein associates with RNA polymerase II snRNA processing intermediates terminating in 3′-uridylate sequences (Xue et al., 2000). To determine if the *T. thermophila* genome encodes a potential 3′-uridylate processing intermediate binding site, we determined the genomic location of *cis*-acting transcription termination signals in order to estimate the length of the 3′-extensions of nascent snRNA transcripts. For this, we used a detailed view in the Integrated Genomics Viewer (IGV) browser of our previously published small RNA-sequencing dataset for total RNA from the *T. thermophila* wild type strain (Kerkhofs et al. 2022) using newly annotated chromosomal locations of spliceosomal RNAs of the downstream genomic region of U1, U2, U4 and U5 snRNAs (**Figure 24**, **Table 3**). In the yeast *S. cerevisiae*, transcription termination is initiated through binding of Nrd1 and Nab3 to consensus sequences GUA[A/G] and UCUU, respectively, on the nascent RNA (Carroll et al., 2004). Multiple UCUU Nab3 binding sites were identified in the immediate downstream genomic region, however, no GUA[A/G] Nrd1 binding site was found (**Figure 24** – consensus RNA polymerase II termination signal sequences highlighted in orange).

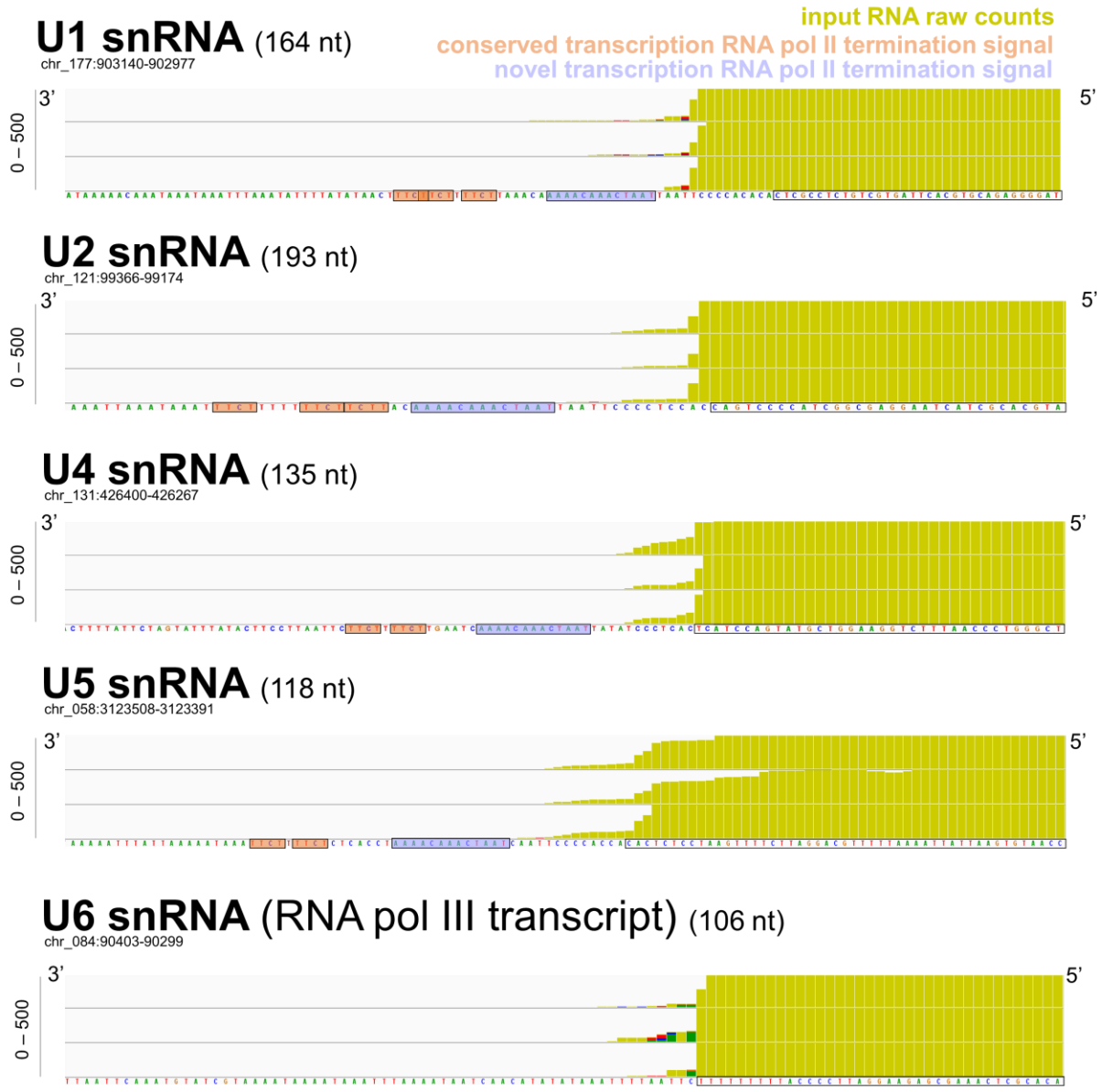


Figure 24. The snRNA transcription termination region contains consensus binding sites for Nab3/UGUU but not Nrd1/GUA[A/G], as well as a newly identified motif.

Zoomed display of downstream chromosomal locations encoding spliceosomal snRNAs. Expression levels are shown as raw counts on the y-axis (n=3 biological replicates). Mature snRNA sequences are boxed based on a previous study (Orum et al., 1991). Transcription termination sequences for RNA-binding proteins Nrd1 (GTA[A/G]) and Nab3 (UCUU) are highlighted in orange. Note that only Nab3 sequences exist. The newly discovered UAAUCAAACAAAA motif is highlighted in purple. Image was generated in Integrative Genomics Viewer (IGV).

Table 3. Chromosomal location of snRNAs in *Tetrahymena thermophila.*

Mature snRNA sequences were obtained from a previous study (Orum et al., 1991).

snRNA	Length (nt)	Sequence	Chromosome location	Strand	NCBI accession number
U1	164	ACTTACCTGGCTGGAGTTTGCTATCGATCATGAAGGGTAGCGGCTTAGGGTGGAGCAGGTCATTGCACAAAAGATGTCTGTAATACCTTATTGTTCCCCGTGCGGGGAACCGAAACAGCACAATTTCTGCTAGGGGAGACGTGCACTTAGTGCTGTCTCCGCTC-3'	chr_177:903140-902977	-	X58845.1
U2	193	ATACCTTCTCGGCCTTTTGGCTAAGATCAAGTGTAGTATCTGTTCTTATCAGTGTGAAAACTGATACTGTCCCTACTAGGGACATGTGGTTTCACATTAATTTTTCACAGGGGTCGGATTTACTAGTGGCTTGCCCTAGTCCCGACGCGGTTGCCCTTGGCATGCACGCTACTAAGGAGCGGCTACCCCTGAC-3'	chr_121:99366-99174	-	X58842.1
U4	135	NACCTCGCGACAGGGGCAATATAGCAGCAAGTGACGGTTAACTGATGCGCTATTATTGCTAGTTGAAAACTACTTCAATAAGTGGAAACGACGCTTGCGTCGGGTCCCAATTTCTGGAAGGTCGTATGACCTACT-3'	chr_131:426400-426267	-	X58844.1
U5	118	ATCACAGAACTCAGCTCATTACGCTTTAATTTTTCGCCTTTTACTAAAGATTACCGTGGGCTGGGTTTACCAATGTGAATTATTAAAATTTTTGCAGGATTCTTTTGAATCCTCTCAC-3'	chr_058:3123508-3123391	-	X58841.1
U6	106	NGAACCCGAAAGGGTCATCCGTTAAAATTGGAACGATACAGAGAAGATTAGCATGGCCCCTGCGCGAAGGATGACACGCTCAAAGCGAGAAGGATTCCCCATTTTT-3'	chr_084:90403-90299	-	X58843.1

To determine if additional, previously unidentified consensus sequences might be located downstream of the genes encoding the RNA polymerase II transcribed snRNAs, we performed a genomic search up to 200 nucleotides downstream of the U1, U2, U4 and U5 snRNAs genes and performed a Gapped Local Alignment of Motifs (GLAM2) search within this region. We identified a UAAUCAAACAAAA motif located immediately downstream of the 3′-end of the RNA polymerase II transcribed snRNA reads and immediately upstream of a Nab3/UCUU sequence (**Figure 24**). This motif is absent from the RNA polymerase III transcribed U6 snRNA (**Figure 24**), but for this we noted an extended uridylate stretch which functions as the RNA polymerase III termination signal and would be expected to form the canonical binding site for Mlp1. We also studied the downstream sequences of several RNA polymerase II transcribed small nucleolar RNAs (snoRNAs) and did not observe a UAAUCAAACAAAA motif (data not shown).

Overall, these results suggest that RNA polymerase II transcribed snRNAs contain a Nab3 binding site (UCUU) and a newly described motif immediately downstream of the mature snRNA 3′-end. Our further analysis did not reveal any apparent uridylate stretches downstream of the mature 3′-end that might function as a high-affinity Mlp1 binding site in a nascent snRNA processing intermediate, as previously described in yeast (Xue et al., 2000).

3.3.2. Mlp1 interacts with mature snRNPs

To investigate if Mlp1 interacts with nascent snRNAs, we reanalysed our previously published small RNA-sequencing dataset from Mlp1-associated and input RNA (Kerkhofs et al. 2022). We found that Mlp1 highly enriches the spliceosomal snRNAs U1, U2, U4, U5 and U6 (**Figure 25a**). While U6 snRNA is transcribed by RNA polymerase III (Dieci et al., 2007; Kunkel et al., 1986; Reddy et al., 1987) and thus contains the typical La/Mlp1 high-affinity 3′-terminal stretch of uridylates (Kerkhofs et al. 2022), the U1, U2, U4 and U5 snRNAs are transcribed by RNA polymerase II. Investigation of the mapped 3′- and 5′- ends of these snRNA transcripts revealed

that Mlp1 associated reads align with the mature snRNA reads (**Figure 25a**), as opposed to 3′-uridylate tailed snRNA intermediates as described in *S. cerevisiae* (Xue et al., 2000). To validate the small RNA-sequencing data, we immunoprecipitated Mlp1-associated RNAs followed by detection by Northern blot using probes specific for mature spliceosomal snRNAs (**Figure 25b**). We confirmed that Mlp1 specifically enriches spliceosomal U snRNAs compared to the Immunoglobin G (IgG) control **Figure 25b**, compare lanes 3 and 5). We also concluded that Mlp1 interacts with mature snRNA since the Mlp1-bound RNA are the same size as the most abundant transcript in the input RNA lane (**Figure 25b** – compare lanes 1 and 5). In addition, we confirmed specific Mlp1-enrichment of snRNAs by qRT-PCR (**Figure 25c**). We compared fold enrichment of Mlp1- and IgG-immunoprecipitated RNA between RNA polymerase II transcribed snRNAs (**Figure 25c** – light blue) and RNA polymerase III transcribed transcripts, including U6 snRNA, 5S rRNA and pre-tRNAs (**Figure 25c**, grey) and found that these are strongly enriched compared to negative control 17S rRNA transcribed by RNA polymerase I (**Figure 25c**, dark blue).

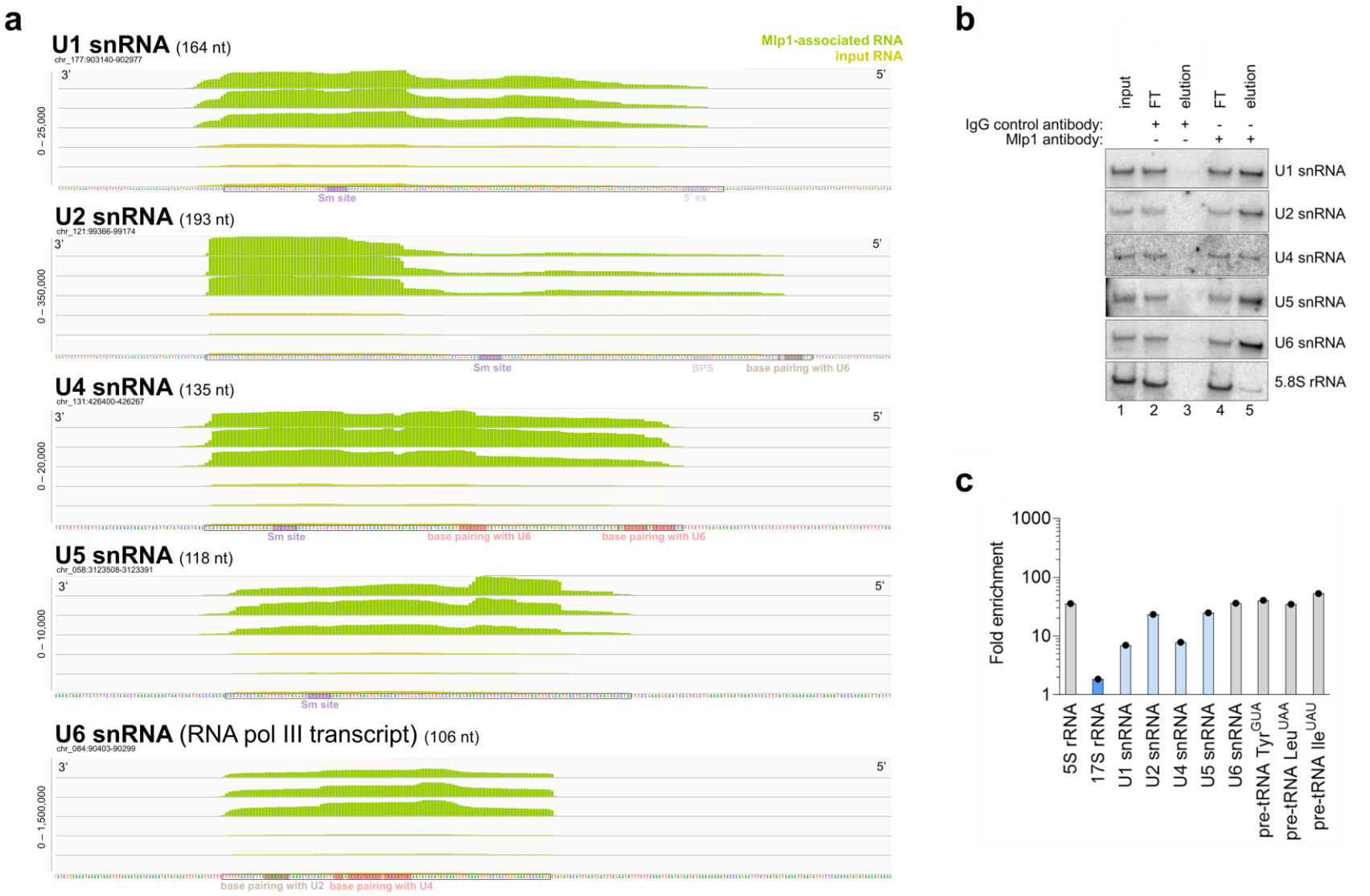


Figure legend on next page

Figure 25. Mlp1 interacts with mature spliceosomal snRNAs.

a Next generation RNA sequencing view of chromosomal locations of spliceosomal snRNAs. Expression levels are shown as raw counts on the y-axis. Mlp1-associated RNA is shown in green (top three rows) and total RNA is shown in yellow (bottom three rows) (n=3 biological replicates). Mature snRNA sequences are boxed based on a previous study (Orum et al., 1991). snRNA-snRNA interactions are highlighted in red/beige, Sm protein binding sites in purple and snRNA-mRNA interactions in light purple. 5′SS: 5′ splice site, BPS: branch point sequence. Image was generated in Integrative Genomics Viewer (IGV). **b** Northern blot analysis of Mlp1 ribonucleoprotein-immunoprecipitated (RIP) samples from *T. thermophila*. Mlp1 enriches all components of the major spliceosome. 5.8S rRNA was used as a negative non-binding control (n=3 biological replicates). IgG: immunoglobin G negative control. **c** qRT-PCR analysis of Mlp1 RIP samples from *T. thermophila* with RNA polymerase I transcript (17S rRNA) highlighted in dark blue, RNA polymerase II transcripts (U1, U2, U4 and U5 snRNAs) in light blue and RNA polymerase III transcripts in grey. Fold enrichment was calculated using the ΔCt method ($Ct_{RIP} - Ct_{input}$), followed by fold enrichment calculation by taking Mlp1-RIP / IgG-RIP ratios (n=1 biological replicate). IgG: immunoglobin G negative control.

To investigate if Mlp1 also interacts with a protein component of snRNPs, we performed endogenous Mlp1 immunoprecipitation in combination with Liquid Chromatography with Tandem Mass Spectrometry (LC-MS/MS). We confirmed specific Mlp1-immunoprecipitation compared to IgG-immunoprecipitation control by Western blot (**Supplementary Figure 10a**). To test for Mlp1-specific protein-protein interactions, immunoprecipitation was performed in the presence of benzonase, which degrades both DNA and RNA. We confirmed successful immunoprecipitation of protein interactors by silver staining (**Supplementary Figure 10b**). Next, we performed LC-MS/MS and statistical analysis of identified peptides and found significant enrichment of a few Mlp1-interacting proteins compared to the immunoglobin G (IgG) control sample (**Figure 26a**). Interestingly, one of the enriched proteins is Sm-B, which is member of the heteroheptameric rings of Sm proteins found surrounding each RNA polymerase II transcribed snRNA (**Figure 26b,c** – Sm-B highlighted in orange) (Leung et al., 2011). Sm-B protein makes contacts with two uridylate RNAs found within the highly conserved U4 Sm motif 5′-AUUUUUG-3′ using side chains of methionine-38 (M38) and asparagine-39 (N39) (**Figure**

26c,d). We compared primary amino acid sequences for Sm-B between several eukaryotic species and found that these residues are conserved in *T. thermophila* (**Figure 26e**), consistent with this protein functioning similarly.

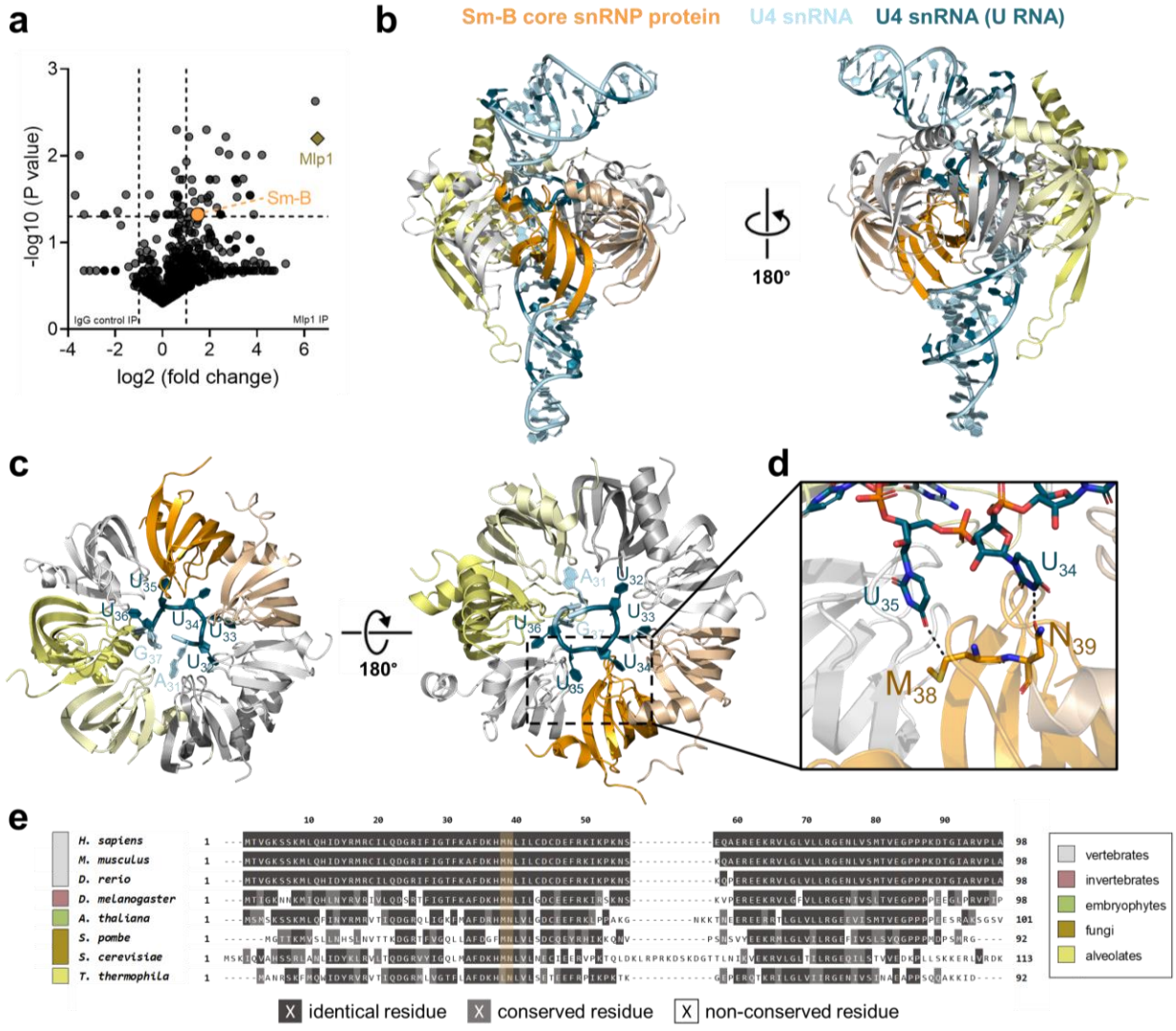


Figure legend on next page

Figure 26. Mlp1 interacts with core spliceosomal protein Sm-B.

a Mass spectrometry data represented as a volcano plot to identify significantly enriched Mlp1-associated proteins compared to an IgG negative control immunoprecipitation. Dotted lines indicate significantly enriched proteins (-log10 < 0.05 or) on the y-axis and significantly enriched proteins (log 2-fold change > 2) on the x-axis. Mlp1 (bait) is shown as a diamond shape in dark yellow (n=2 biological replicates). **b,c** Front, back (**b**), top and bottom views (**c**) of high-resolution structure of the human U4 snRNP core containing SmE–SmG–SmD3–SmB–SmD1–SmD2–SmF. Mlp1-protein interactor SmB is highlighted in orange and RNA is shown in light blue with uridylate RNA highlighted in dark blue enabling visualization of the highly conserved Sm site AUUUUUG located in the middle of the heteroheptameric ring (PDB: 4WZJ). For clarity, only the Sm site AUUUUUG is shown in **c**. **d** Zoomed panel showing hydrogen bonds between Sm-B and uridylate RNA within the Sm motif of U4 snRNA. Carbon atoms are colored based on the molecule (uridylate RNA dark blue and Sm-B protein orange) with other atoms colored by type: oxygen in red; nitrogen in blue, phosphorous in orange and sulfur in yellow. Hydrogen bonds are shown as black dashed lines. **e** Primary amino acid sequence alignments from different eukaryotic lineages showing conservation of the Sm-B protein. Residues important for uridylate binding, as shown in **d**, are highlighted in orange. A dark grey background indicates identical residues, light grey conserved residues and white indicates no conservation.

3.3.3. Depletion of Mlp1 has no effect on snRNA stability and subcellular localization

Since Mlp1 interacts with mature snRNAs, we investigated if Mlp1 depletion affects transcript abundance. To test this, we compared transcript abundances between the *T. thermophila* wild type strain and previously described partial Mlp1 knockout (KO) strain (Kerkhofs et al. 2022) by Northern blot (**Figure 27a**) and qRT-PCR (**Figure 27b**). We found that Mlp1 depletion has no substantial effect on the abundance of snRNAs, indicating that Mlp1 is not involved in the steady-state stabilization or maturation of these transcripts. As a control, we determined the mRNA levels for Mlp1 using qRT-PCR and confirm that Mlp1 mRNA levels in the Mlp1 partial knockout strain were very low (**Figure 27b**).

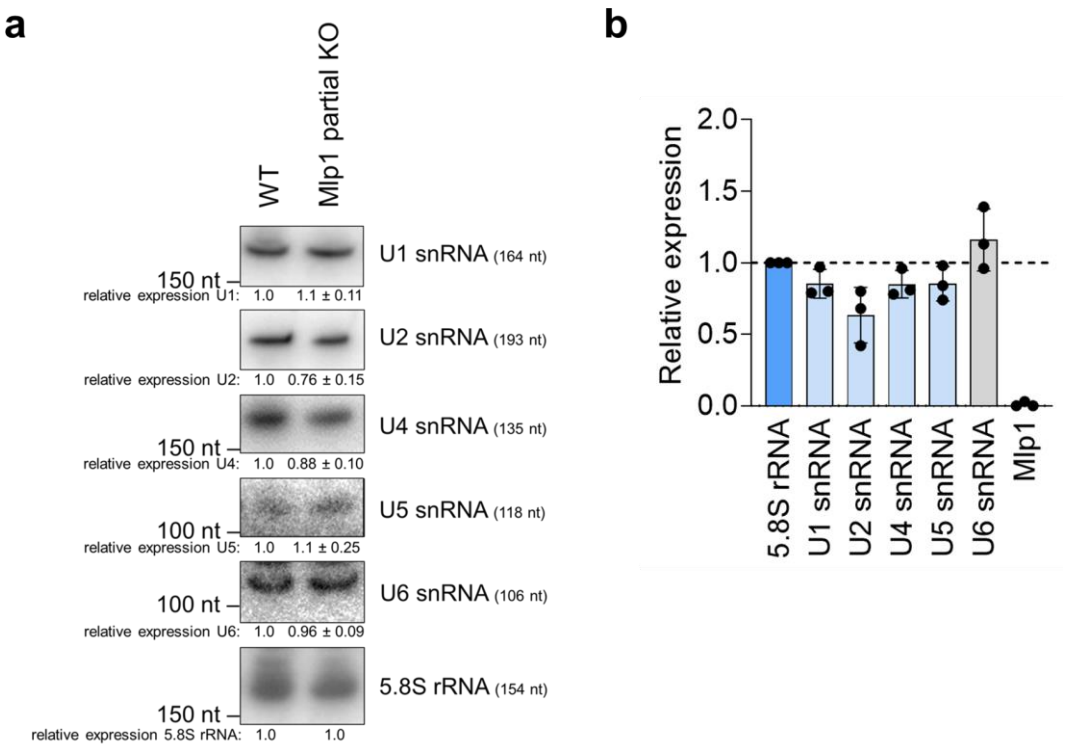


Figure 27. Mlp1 depletion has no effect on snRNA stability.

a Northern blot detecting mature U1, U2, U4, U5 and U6 snRNA in wild type (WT) and partial Mlp1 knockout (KO) strains. Loading control: 5.8S rRNA. Quantification of Northern blot samples (n=3 biological replicates) is shown below. Band intensity was determined using ImageJ relative to the 5.8S rRNA loading control and relative expression was calculated by taking the ratio between the Mlp1 partial knockout and wild type strains. **b** qRT-PCR analysis of WT and partial Mlp1 KO strains. RNA polymerase I transcript (5.8S rRNA) is highlighted in dark blue, RNA polymerase II transcripts (U1, U2, U4 and U5 snRNAs) in light blue and RNA polymerase III transcripts in grey. Fold enrichment was calculated using the ΔCt method ($Ct_{5.8S\ rRNA} - Ct_{RNA}$), followed by fold enrichment calculation by taking partial Mlp1 KO / WT ratios (n=3 biological replicates).

In vertebrates, RNA polymerase II transcribed snRNAs are transported into the cytoplasm where they assemble with the Sm proteins (Mattaj, 1986). This step is required for dimethylation of the m^7G cap to form the m_3G-cap structure and subsequent reimport into the nucleus (Mattaj, 1986). In the yeast *S. cerevisiae* snRNAs obtain a m_3G-cap structure in the nucleolus (Mouaikel et al., 2002). On the other hand, the RNA polymerase III transcribed U6 snRNA consistently is

only found in the nucleus for biogenesis and transiently travels through the nucleolus to obtain post-transcriptional modifications (Didychuk et al. 2018). La proteins have been described to function in retention of pre-tRNAs in the nucleus to complete correct pre-tRNA processing. Previous work demonstrated that La mutants that mislocalize to the cytoplasm have been shown to have dysfunctional processing, where unprocessed 5′-leader and 3′-trailer containing pre-tRNAs are found in the cytoplasm. Since tRNA intron splicing in yeast takes place in the cytoplasm, the pre-tRNAs were spliced prior to 5′- and 3′-end processing leading to aberrant tRNA maturation (Bayfield et al., 2007; Intine et al., 2002). We tested if Mlp1 depletion affects subcellular localization of snRNAs by cellular fractionation separating the cytoplasmic compartment from the nucleus. We confirmed cellular fractionation by Western blot using cytoplasmic marker β-actin and nuclear marker histone H3 (**Figure 28a**) and determined the distribution of snRNAs by Northern blotting using snRNA-specific probes. We found that Mlp1 depletion has no effect on cellular distribution of snRNAs (**Figure 28b**), indicating that Mlp1 is not required for retaining these non-coding RNAs in the nucleus.

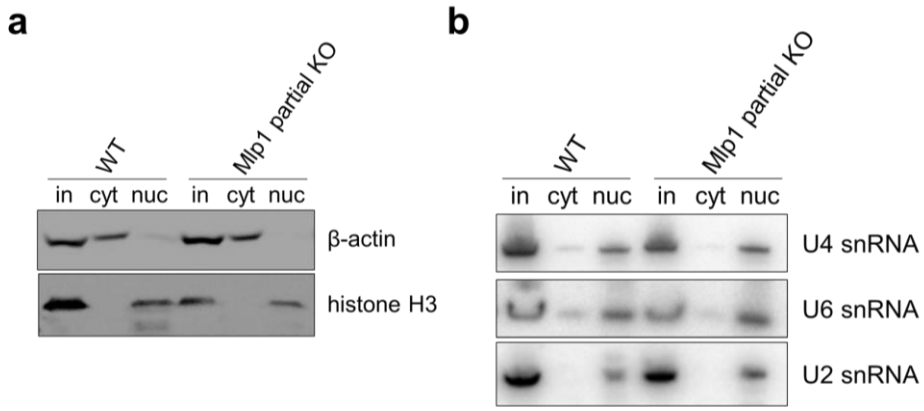


Figure 28. Mlp1 depletion has no effect on subcellular localization of snRNAs.

a Western blot confirming cellular fractionation of wild type (WT) and partial Mlp1 knockout (KO) strains between cytoplasm (cyt) and nucleus (nuc) using cellular compartment specific antibodies. β-actin was used as a control for the cytoplasmic fraction and histone H3 for the nuclear fraction (n=3 biological replicates). **b** Northern blot analysis of RNAs following cellular fractionation (n=3 biological replicates).

3.3.4. The *Tetrahymena thermophila* Tgs1 enzyme is only partially conserved

RNA polymerase II transcribed snRNAs obtain a m_3G-cap structure synthesized by the highly conserved Tgs1 enzyme (Mouaikel et al., 2002, 2003). Early work in *T. thermophila* described that its snRNAs contain a 5′-cap that is distinct from the well-described mammalian and yeast m_3G-caps. This distinct 5′-cap is resistant to periodate oxidation and β-elimination, which requires free 2′- and 3′-hydoxyl groups in the ribose, indicating that these residues are likely modified (Pedersen et al., 1985). First, we confirmed conservation of the Tgs1 enzyme responsible for conversion of the m^7G cap into a m_3G-cap structure (Mouaikel et al., 2002). The Tgs1 enzyme functions as a dimethyl transferase catalyzing two successive methyl group transfers each from one S-adenosyl-L-methionine (AdoMet) onto the m^7G-cap structure of the nascent non-coding RNA (Monecke et al., 2009). A high-resolution structure revealed the spatially adjacent positions of these two substrates within the catalytic domain of the human Tgs1 enzyme (Monecke et al. 2009) (**Figure 29a**). Residues important for interactions with AdoMet and m^7G substrates are highlighted in blue and orange, respectively (**Figure 29b,c**). To determine if the *T. thermophila* Tgs1 enzyme is conserved, we compared primary amino acid sequences between in *T. thermophila* and other eukaryotic species (**Figure 29d**). Strikingly, the *T. thermophila* residues required for AdoMet interactions are highly conserved (**Figure 29d** – highlighted in blue), while residues important for m^7G substrate recognition are variable (**Figure 29d** – highlighted in orange). We compared the predicted AlphaFold (Jumper et al., 2021) structure of *T. thermophila* Tgs1 to the high-resolution human TGS1 structure (Monecke et al., 2009) and found that *T. thermophila* Tgs1 predicted structure generally adopts a similar fold, however, one main difference is the absence of an α-helix in *T. thermophila* important for interactions with the m^7G substrate (**Supplementary Figure 11**). These data are consistent with previous work suggesting that the 5′-ends of snRNAs in *T. thermophila* may have a different structure from the m_3G-cap found in other eukaryotes, even if the presence of a Tgs1 homolog in this system suggests that that this alternate structure might still be related to an m_3G-cap.

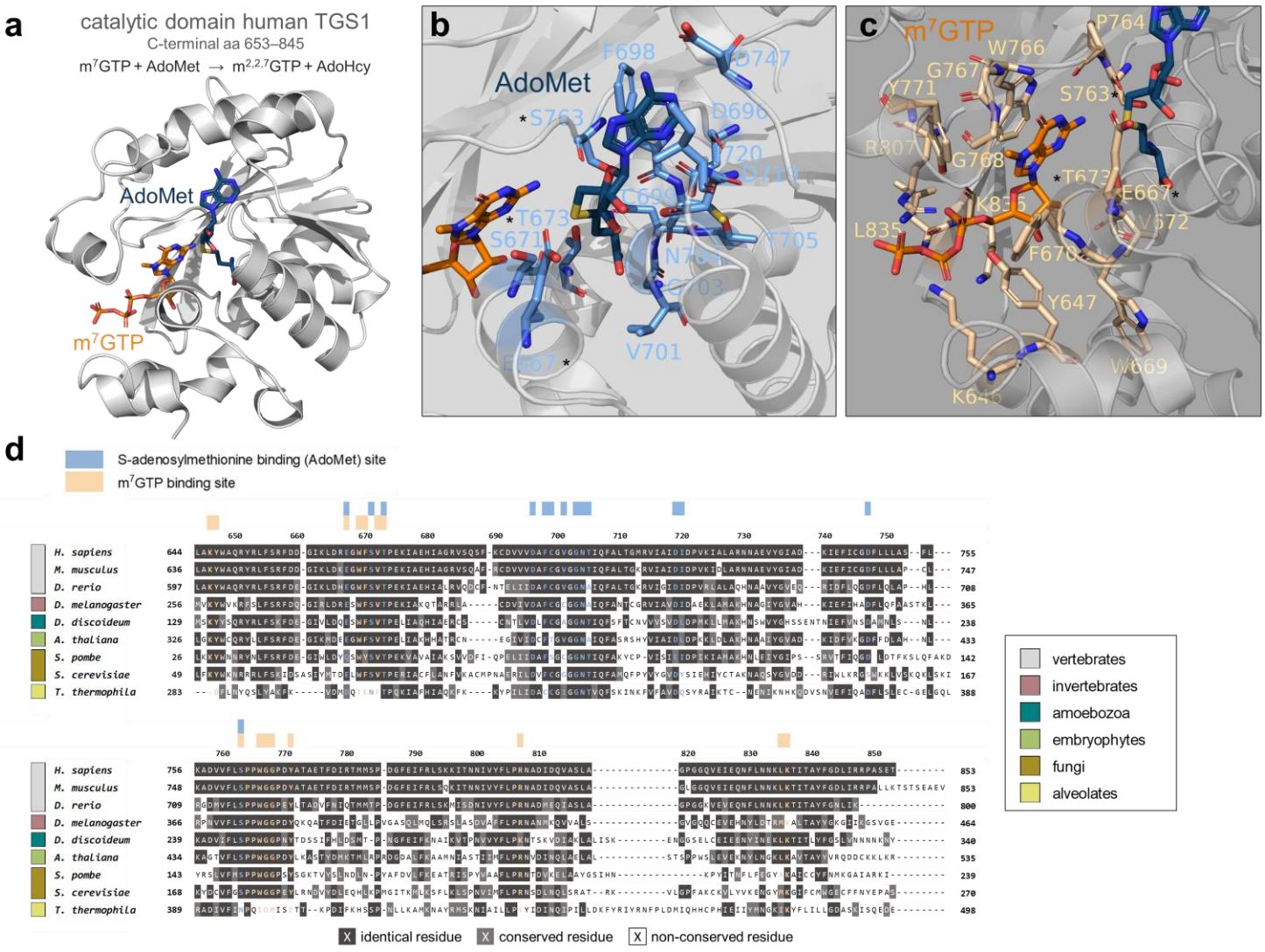


Figure legend on next page

Figure 29. Conservational analysis of the m_3G-capping Tgs1 enzyme.

a High-resolution structure from (Monecke et al., 2009) (PDB: 3GDH) reveals location of substrates S-adenosylmethionine (AdoMet) and m^7GTP within the trimethyl synthetase-1 (TGS1) enzyme. **b,c** Zoomed panels highlight residues involved in AdoMet (blue) (**b**) and m^7GTP interactions (orange) (**c**). Residues with a dual function in AdoMet and m^7GTP interactions are indicated with an asterisk. Carbon atoms are colored based on the molecule (AdoMet in dark blue and m^7GTP in orange) with other atoms colored by type: oxygen in red; nitrogen in blue, phosphorous in orange and sulfur in yellow. **d** Primary amino acid alignments of the catalytic domain of trimethylguanosine synthase-1 (Tgs1) from different eukaryotic lineages. A dark grey background indicates identical residues, light grey conserved residues and white indicates no conservation. Amino acid residues involved in AdoMet and m^7GTP binding are highlighted in blue and orange, respectively.

Next, we tested if m_3G or m_3G-like capped structures exist in *T. thermophila* by RNA-immunoprecipitation using an anti-m_3G antibody from total RNA. We noted an enrichment of bands whose size is consistent with snRNAs in the anti-m_3G elution lanes, but loss of a smear typically associated with tRNAs when RNA was visualized using ethidium bromide (EtBr) staining, consistent with some transcripts containing cap structure recognized by the anti-m_3G antibody (**Figure 30a**). Next, we performed Northern blotting for specific non-coding RNAs and found that RNA polymerase II transcribed snRNAs were immunoprecipitated by the anti-m_3G antibody, whereas an RNA polymerase III transcribed mature tRNA containing a 5′-triphosphate cap was not (**Figure 30b**). Unexpectedly, we saw enrichment for the RNA polymerase III transcribed U6 snRNA, which in other species obtains a γ-monomethyl phosphate instead of a m_3G-cap (**Figure 30b**) (Singh & Reddy, 1989). This observation could be due to the fact that the majority of cellular U6 is found in the context of the U4/U6 di-snRNP, possibly leading to non-specific association of U6 through U4 in the m_3G-cap antibody elution.

Having established that *T. thermophila* snRNAs are enriched by the anti-m_3G antibody, we then tested if Mlp1 associates with mature snRNAs containing this cap structure. To investigate this, we performed a similar immunoprecipitation using an anti-m_3G antibody but instead from RNA previously immunoprecipitated from total RNA using an Mlp1-specific antibody. Detection of

Mlp1-associated/m_3G-antibody immunoprecipitated RNA by Northern blot revealed that Mlp1 associates with snRNAs that are immunoprecipitated with the anti-m_3G-cap antibody (**Figure 30d**), suggesting that Mlp1 binds the mature form of snRNAs that contain an m_3G or m_3G-like capped 5′-end.

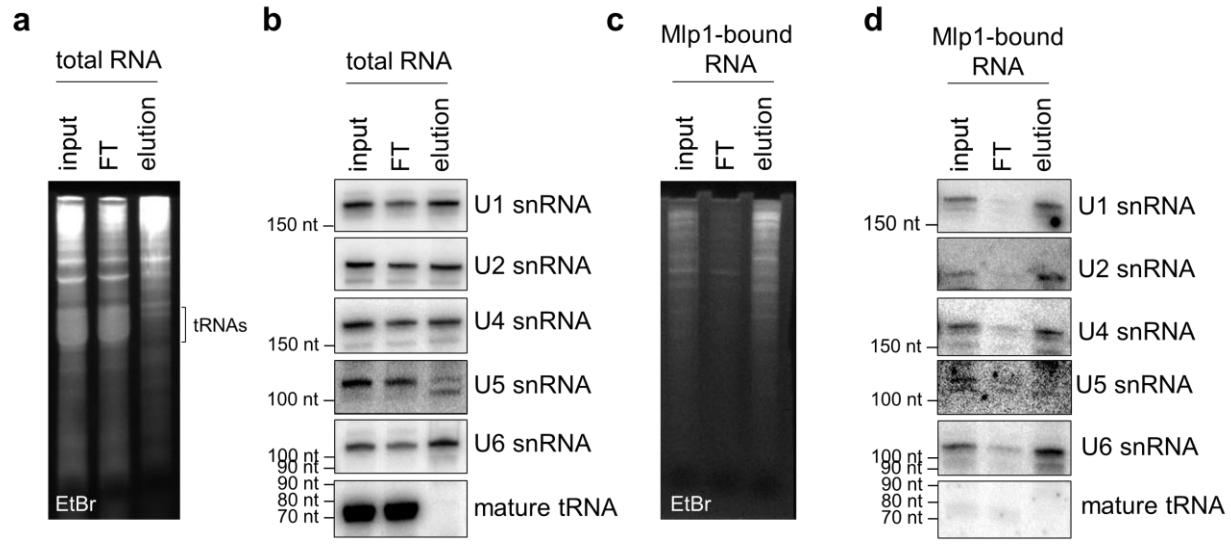


Figure 30. Mlp1 associates with m_3G-cap containing snRNAs.

a,c EtBr staining of denaturing urea gels from an RNA-immunoprecipitation using a m_3G antibody from total RNA (**a**) and Mlp1-associated RNA (**c**). EtBr: ethidium bromide. **b,d** Northern blot analysis of denaturing urea gels from **a,c** detecting m_3G-cap containing snRNAs (U1, U2, U4 and U5), negative control γ-monomethyl phosphate capped U6 snRNA and negative control 5′-triphosphate containing mature tRNA (n=1 biological replicate).

3.3.5. snRNP assembly is affected upon Mlp1 depletion

The observations that Mlp1 interacts with mature snRNAs and a core snRNP protein led us to investigate a potential role for Mlp1 in snRNP assembly. To test this, we performed fractionation of total wild type and partial Mlp1 knockout lysates on 10-30% glycerol gradients to separate the snRNPs based on density, where a defect or decrease in snRNP assembly could be observed as a shift of the snRNA towards the lighter (top) glycerol fractions as detected by Northern blot, as demonstrated previously (Makarova et al., 2002; Tanackovic et al., 2011). First, we

determined the distribution of total RNA by ethidium bromide (EtBr) staining (**Figure 31a**). Northern blot analysis revealed that upon Mlp1 depletion, the tested snRNAs were shifted slightly towards the lighter fractions, indicating that either fewer snRNAs are assembled into functional snRNP complexes or that the formed complexes are incomplete relative to wild type (**Figure 31b**). We used the distribution of mature tRNA as a loading control since Mlp1 is a nuclear protein and depletion of Mlp1 has no effect on mature tRNA levels (Kerkhofs *et. al*. 2022). Next, we used Western blot to determine the distribution of the Mlp1 protein on the 10-30% gradients and found that that Mlp1 is present in fractions containing snRNAs, consistent with Mlp1 associating with snRNP complexes (**Figure 31c**).

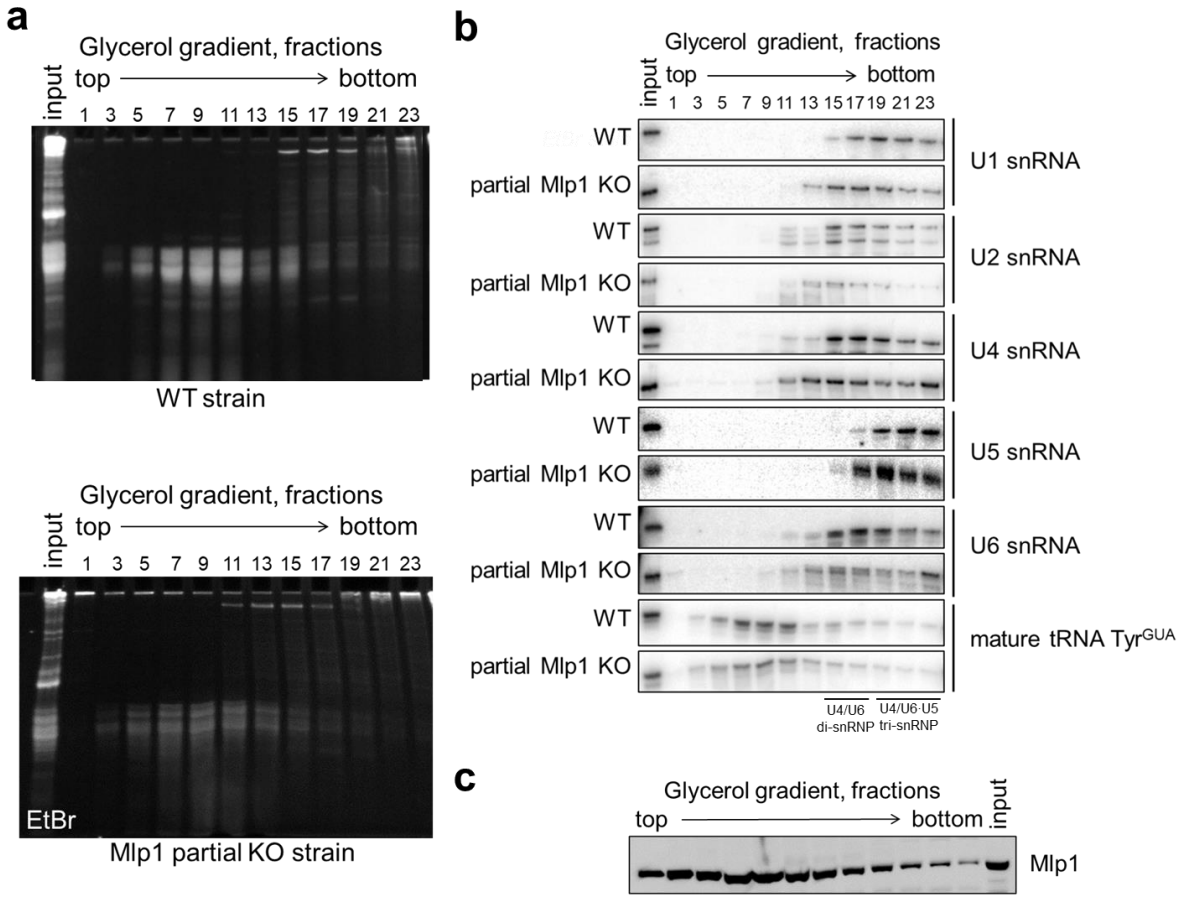


Figure legend on next page

Figure 31. Mlp1 depletion affects snRNP assembly.

a Ethidium Bromide (EtBr) stain to detect total RNA distributions on the Northern blot in **b**. **b** Northern blot analysis of 10-30% glycerol gradient. Fractions were collected and RNA was extracted using phenol:chloroform:isoamylalcohol extraction and ethanol precipitation (n=3 biological replicates). Localization of U4/U6 di-snRNP and U4/U6·U5 tri-snRNP are shown below based on co-localization patterns of U4, U6 and U5 snRNAs in the wild type strain. **c** Western blot to detect distribution of Mlp1 confirming co-localization of Mlp1 with higher molecular weight snRNP complexes (n=3 biological replicates).

3.3.6. Mlp1 associates with the U4/U6 di-snRNPs during splicing

Next, we determined if Mlp1 functions in the formation of spliceosomal complexes. During splicing, the U4 snRNA base-pairs with U6 RNA forming the U4/U6 snRNA, followed by formation of the U4/U6·U5 tri-snRNP (see **Figure 22**). To test association of Mlp1 with the U4/U6 complex we performed native RNP-immunoprecipitation of Mlp1. We confirmed Mlp1-specific immunoprecipitation by Western blot (**Figure 32a**). The immunoprecipitated RNA was separated on a native polyacrylamide gel under conditions that preserve RNA-RNA interactions. We performed Northern blotting using a U4-specific probe and found that Mlp1 strongly enriches the U4/U6 complex compared to the free U4 snRNAs (**Figure 32b**). Next, we determined if Mlp1 depletion affects the formation of the U4/U6 complex by comparing wild type and partial Mlp1 knockout strains. We confirmed Mlp1-specific depletion by Western blot (**Figure 32c**). Total RNA was separated on a native polyacrylamide gel and probed using U4 and U6-specific probes. We found that U4/U6 complex formation was strongly reduced in partial Mlp1 knockout strains compared to the partial Mlp1 knockout strain (**Figure 32d,e**). The next step in spliceosome assembly includes assembly of U5 snRNP with U4/U6 through protein-protein and RNA-protein interactions rather than RNA-RNA interactions, leading to formation of the U4/U6·U5 tri-snRNP (Gehring & Roignant, 2021). However, total RNA was extracted using Trizol which denatures proteins and would therefore not retain the U4/U6·U5 tri-snRNP complex (Rio et al., 2010). As expected, we found that U5 was not found as a higher molecular weight

complex in these native gels (**Figure 32d**). To form the catalytically active spliceosome, the U4/U6 RNA-RNA interactions are unwound and U1 snRNA at the 5′-splice site is displaced by U6 (see **Figure 21**), leading to RNA-RNA interactions between U2 and U6 snRNAs (Gehring & Roignant, 2021). To investigate if Mlp1 depletion affects the levels of RNA-RNA interaction between U2 and U6, we detected U2 snRNA by Northern blot of a native polyacrylamide gel and found no observed effect (data not shown).

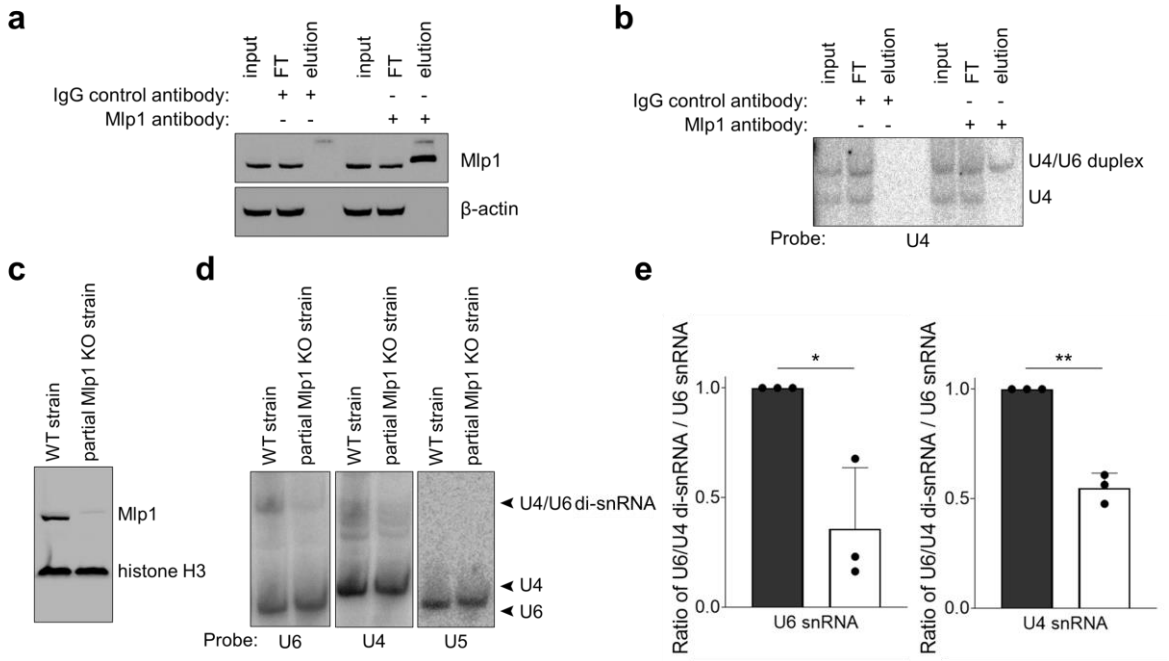


Figure 32. Mlp1 associates with the U4/U6 di-snRNP and affects assembly.

a Western blot confirming Mlp1-immunoprecipitation from *T. thermophila*. Loading control: β-actin. **b** Native polyacrylamide gel of Mlp1 ribonucleoprotein-immunoprecipitated samples in **a** from *T. thermophila* (n=2 biological replicates). Mlp1 enriches the U4/U6 snRNA di-snRNP rather than the individual U4 snRNA. FT: flowthrough, IgG: immunoglobin G control antibody. **c** Western blot analysis confirming Mlp1 depletion. Loading control: histone H3. **d** Native polyacrylamide gel from wild type (WT) and partial Mlp1 knockout (KO) strains in **c** (n=3 biological replicates). **e** Quantification of native gels in **d**. Statistical significance (P-value) was calculated using a paired t-test. P value < 0.05: *, P value < 0.01: **.

3.3.7. Mlp1 depletion leads to pre-mRNA splicing defects

To investigate if decreased snRNP and U4/U6 di-snRNP assembly caused by Mlp1 depletion affects splicing, we tested for aberrant mRNA splicing by investigating two pre-mRNAs that

have been previously described as being prone to intron retention in *T. thermophila* (Tian et al., 2017). Specifically, we investigated splicing of the pre-mRNAs for NLRC3 (TTHERM_00327090) and a hypothetical protein (TTHERM_00526420) by RT-PCR amplification, using primers that would amplify both the intron-containing pre-mRNA and spliced mature mRNA (**Figure 33a,c**). Comparison of the unspliced and spliced amplification products between wild type and partial Mlp1 knockout strain revealed that splicing efficiency is decreased upon Mlp1 depletion (**Figure 33b,d**). These data are consistent with the observed defect in U4/U6 di-snRNP formation leading to associated splicing defects in Mlp1 depleted cells.

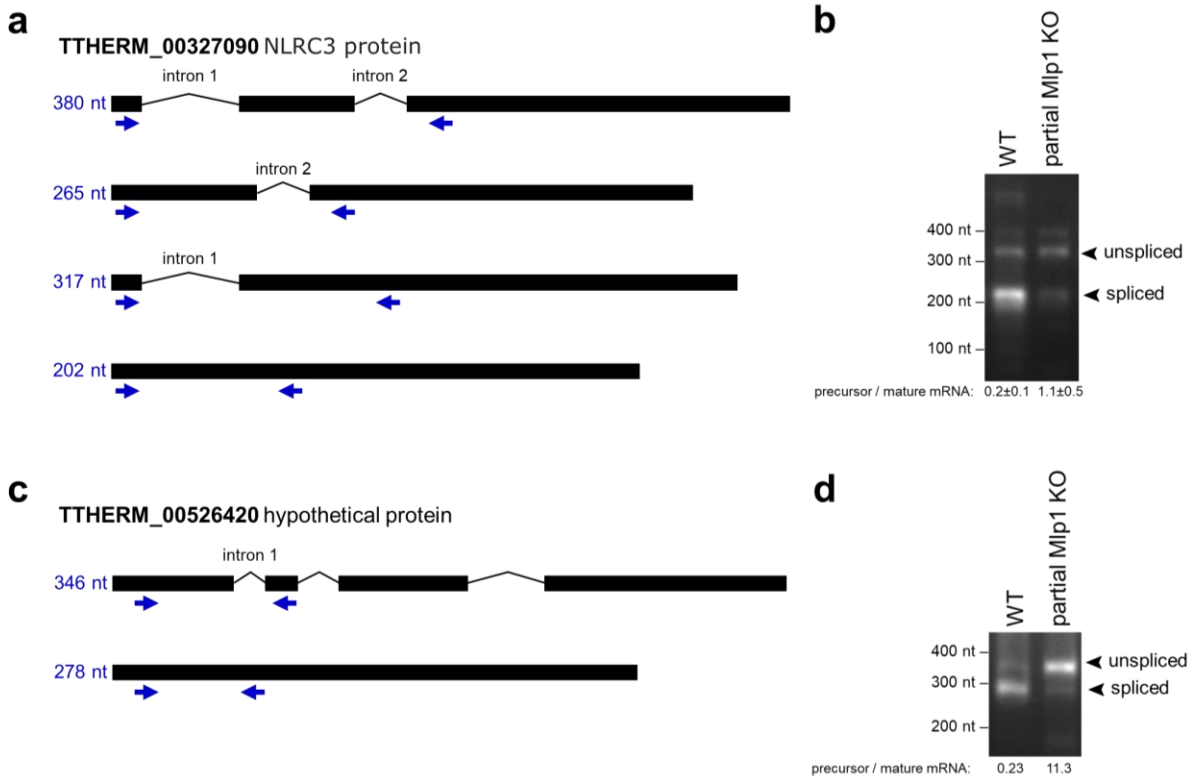


Figure 33. Pre-mRNA splicing is affected by Mlp1 depletion.

a, c Schematics of different mRNAs intermediates ranging from unspliced (containing introns) to fully spliced (containing no introns). Black rectangles represent exons and thin black line represent introns. Blue arrows indicate primers used for PCR from cDNA and the corresponding amplicon size is shown on the left of each transcript. **b, d** PCR products separated on an 1% agarose gel. The upper band represents the unspliced pre-mRNA and lower band the spliced mature mRNA. Ratios between upper (unspliced) and lower (spliced) bands were calculated and are represented below the image (n=2 biological replicates in **b** and n=1 biological replicate in **d**).

3.4. Discussion

Splicing of pre-mRNAs is an essential step in pre-mRNA maturation and is catalyzed by the spliceosome which consists of multiple proteins and U1, U2, U4, U5 and U6 snRNAs which are assembled through multi-step processes into four large RNP complexes: the U1, U2, U4/U6 and U5 snRNPs (Gehring & Roignant, 2021). In this study, we compared the conservation of expected RNA polymerase II snRNA transcription termination elements in *T. thermophila* to that of budding yeast. Transcription termination in *S. cerevisiae* is mediated by the NNS complex, consisting of RNA-binding proteins Nrd1 and Nab3 and the helicase Sen1. Nrd1 and Nab3 function through binding of highly conserved *cis*-acting binding sites GUA[A/G] and UCUU, respectively, present in the nascent RNA transcript, which in turn leads to a decrease in RNA polymerase II transcription rate and nascent snRNA release from the DNA by helicase Sen1 (**Figure 34a**) (Allmang et al., 1999; Carroll et al., 2004). We performed a genomic search up to 200 nucleotides downstream of the snRNA genes and identified multiple repeats of the Nab3 binding site UCUU in the downstream genomic region, proximal to the 3′-end of the corresponding mature snRNA transcripts. However, the Nrd1 binding site GUA[A/G] was not detected. Consistent with this, a distantly related Nab3 ortholog is described in the Ciliates Genome Database (https://ciliate.org) as RRM44 polypyrimidine tract-binding protein (TTHERM_00685970), while the Nrd1 ortholog could not be identified in *T. thermophila.* The presence of the consensus Nab3 UCUU binding site may indicate that RNA polymerase II transcription of snRNAs in *T. thermophila* utilizes a transcription termination pathway more similar to yeast as opposed to vertebrates.

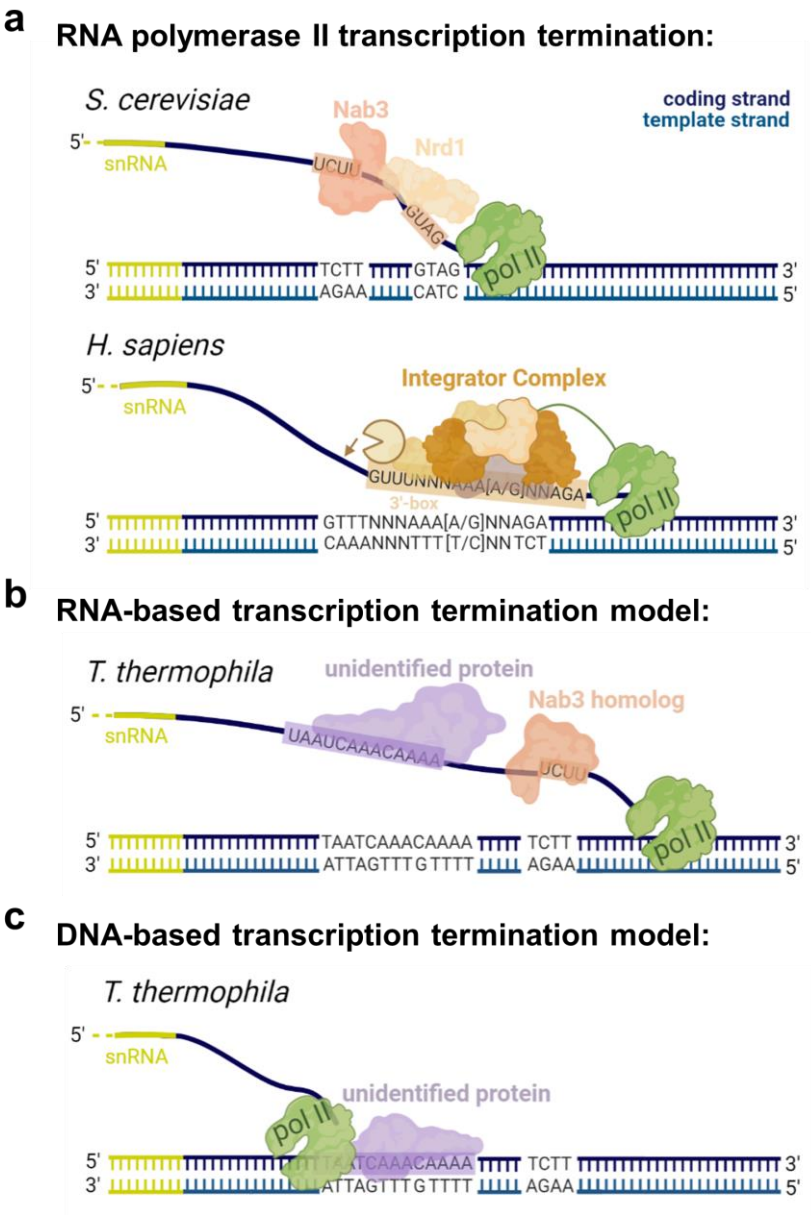


Figure legend on next page

Figure 34. RNA polymerase II transcribed snRNA transcription termination model.
Schematic representation of the 3′-downstream region of RNA polymerase II transcribed snRNAs. The mature snRNA in both DNA and RNA is highlighted in yellow. Conserved RNA motifs and corresponding RNA-binding proteins Nab3, Nrd1 and Integrator complex are shown in orange. Novel RNA or DNA motif and corresponding RNA- or DNA-binding protein are shown in purple. RNA polymerase II (pol II) is shown in green. **a** Previous work described the RNA polymerase II transcription termination mechanisms in the yeast *S. cerevisiae* and in humans. *S. cerevisiae* initiates transcription termination through co-transcriptional binding of RNA-binding proteins Nrd1 and Nab3 to consensus sequences GUA[A/G] and UCUU, respectively. In humans, transcription termination is completed by endonucleolytic cleavage by the Integrator Complex of the nascent snRNAs upstream of the 3′-box *cis*-acting element. **b** The hypothesized RNA-based transcription mechanism for *T. thermophila* is similar to the *S. cerevisiae* where the Nab3 homolog, RRM44 polypyrimidine tract-binding protein, binds the conserved UCUU sequence and the yet to be identified *trans*-acting protein binds the novel UAAUCAAACAAAA, leading to transcription termination. **c** The hypothesized DNA-based transcription termination involved binding of the unidentified *trans*-acting factor to the novel TAATCAAACAAAA or TTTGTTTGATTA on the template strand, leading to a block in RNA polymerase II progression and as a result transcription termination. Image generated using BioRender.

Interestingly, we identified a novel UAAUCAAACAAAA motif immediately downstream of each RNA polymerase II transcribed snRNA gene. This novel motif is also consistently located less than 10 nucleotides upstream to the UCUU sequence, which could indicate a similar transcription termination mechanism utilized in yeast where Nab3 and Nrd1 form heterodimers to increase RNA binding affinity through cooperative binding (**Figure 34a,b**) (Hobor et al., 2011). Alternatively, this novel motif could instead function as a *cis*-acting DNA element recruiting a termination factor leading to transcription termination through blocking RNA polymerase II progression on the DNA (**Figure 34c**). For example, Reb1 is involved in rRNA transcription termination by RNA polymerase I in the yeast *S. cerevisiae* through binding to 11 nucleotides within a 53 base pair DNA sequence found within the 3′-downstream region resulting in termination and release of the nascent transcript (Németh et al., 2013). To further investigate the function of this novel motif in *T. thermophila*, we are going to determine if the novel UAAUCAAACAAAA motif is transcribed and thus present in nascent snRNA transcripts by

Northern blotting with specific probes against that region within the snRNA 3′-end. Without transcription of this novel motif, a DNA-based transcription termination model is more likely. To identify a more specific function, we are going to produce a biotin-tagged DNA and RNA sequence of this novel motif and flanking sequences and identify interacting proteins by mass spectrometry. These experiments will allow us to design functional assays in *T. thermophila* to further study snRNA transcription termination by generating knockout strains for the potential *trans*-acting factor(s) and strains carrying point mutations within the novel motif.

Nucleolar small RNAs (snoRNAs) are also transcribed by RNA polymerase II and employ a similar Nab3/Nrd1/Sen1 mechanism for transcription termination of individually encoded snoRNAs in yeast (Steinmetz et al., 2001). We searched the downstream genomic region of several snoRNAs for transcription termination signals and were only able to find the UCUU sequence, while the GUA[A/G] and novel UAAUCAAACAAAA motif were absent. This is consistent with this novel motif and the potential corresponding *trans*-acting factor being used specifically for snRNA expression.

In humans, the Nab3/Nrd1/Sen1 transcription termination system is not conserved. Instead, the Integrator Complex endonucleolytically cleaves the nascent snRNA upstream of the 13-16 nucleotide 3′-box *cis*-acting sequence (**Figure 34a**) (Baillat et al., 2005). The 3′-box $GTTTN_{0-3}AAA[A/G]NNAGA$ motif is highly conserved in all RNA polymerase II transcribed snRNAs (Hernandez, 1985; Yuo et al., 1985) and is not conserved in yeast (Uguen & Murphy, 2003). The Integrator complex contains multiple subunits and is associated with the phosphorylated C-terminal domain of RNA polymerase II (Mendoza-Figueroa et al., 2020). A previous study performed a proteome search comparing sequence conservation against human integrator subunits of more than 1000 eukaryotic species and found that the components of the Integrator Complex have the highest conservation in animals (70%, *i.e.* the percentage of animals

containing all integrator subunits IntS1-14). In contrast, 18% of plants, 10% other eukaryotes and only 2% of fungi contained all integrator subunits (Kirstein et al., 2021). This indicates that alveolates such as *T. thermophila* likely do not employ the Integrator Complex for processing of RNA polymerase II transcripts, however, further studies are required to confirm this.

We investigated if the *T. thermophila* La protein Mlp1 could function similarly as the *S. cerevisiae* Lhp1p protein, where following transcription termination, the 3′-ends are endonucleolytically processed downstream of the mature snRNA sequence resulting in processing intermediates of RNA polymerase II transcribed snRNAs ending in a 3′-terminal stretch of uridylates that are bound by Lhp1p (Chanfreau et al., 1997; Seipelt et al., 1999; Xue et al., 2000). Based on the computational transcription termination analysis, we concluded that *T. thermophila* nascent RNA polymerase II transcribed snRNAs lack an internal uridylate sequence that could potentially be exposed during processing to function as a high affinity binding site for La proteins. Taken together, the absence of Nrd1/GUA[A/G] sequence and an internal uridylate sequence upstream of transcription termination signal(s) UCUU and the novel motif, indicate that *T. thermophila* snRNA transcription and processing likely proceeds through an alternative pathway compared to the previously studied yeast *S. cerevisiae*.

We established that Mlp1 interacts with mature snRNAs rather than nascent transcripts or processing intermediates by RNP-immunoprecipitation followed by Northern blotting by confirming that Mlp1-associated snRNAs have the same size as the most abundant mature snRNA transcript in total RNA. In addition, we obtained chromosomal locations of snRNAs and reanalyzed our previously published small RNA-sequence dataset to confirm that Mlp1 enriches mature snRNA as opposed to longer precursor versions of these transcripts. Using co-immunoprecipitation of Mlp1, we identified direct protein interactors by mass spectrometry and found that Mlp1 associates with the Sm-B protein, which is a core Sm protein that is part of the

heteroheptameric ring formed around the Sm-binding site of each RNA polymerase II transcribed snRNA (Leung et al., 2011). The Sm heteroheptameric ring assembles on nascent snRNAs and is required for m_3G-capping of the snRNA by Tgs1 and remains associated during splicing of pre-mRNAs (Mattaj, 1986; Will & Lührmann, 2011). We failed to identify other members of the Sm heteroheptameric ring, potentially because Mlp1 could associate with the complex through direct interactions with Sm-B only. We hypothesize that Mlp1 is associated with the entire Sm heteroheptameric ring, but indirect interactions were lost during our sample preparation. In addition, multiple uncharacterized proteins were identified as significantly enriched protein-proteins interactors for Mlp1, and further investigation is required to identify these proteins through conservational analysis (NCBI Resource Coordinators, 2016).

To identify at which stage Mlp1 associates with the snRNP we first investigated if Mlp1-associated snRNAs have already obtained the m_3G-cap structure (or m_3G-like cap structure). First, we performed a conservational analysis of the catalytical domain of the Tgs1 enzyme which converts the m^7G-cap into the m_3G-cap structure through hypermethylation using two consecutive methyl group transfers from AdoMet substrates (Monecke et al., 2009; Mouaikel et al., 2002). Strikingly, we found that the AdoMet substrate binding site was highly conserved, while the m^7G substrate binding site was highly variable in *T. thermophila*. The highly conserved primary amino acid sequence for both substrate binding domains in other eukaryotes likely indicates that these are required for optimal activity of the catalytic domain of Tgs1.

Early work in *T. thermophila* described the presence of a distinct 5′-cap on snRNAs that is divergent from the classic m_3G-cap found on snRNAs in other eukaryotes (Pedersen et al., 1985). More specifically, this cap structure was hypothesized to have modifications on the 2′- and 3′-hydoxyl groups in the ribose (Pedersen et al., 1985). A potential link between a semi-conserved Tgs1 enzyme and an alternative 5′-m_3G-like cap structure might exist in *T.*

thermophila but would require further investigation. Nonetheless, RNA-immunoprecipitations performed with a m_3G-specific antibody was capable of enriching expected RNA targets, including the RNA polymerase II transcribed snRNAs. Specificity of the antibody against transcripts lacking a m_3G cap structure was obtained from lack of detection of mature tRNAs which contains a monophosphate at the 5′-terminus (Gößringer et al., 2012). Since m_3G-antibodies are able to distinguish between transcripts containing a m_3G-cap and m^7G cap structure (Reuter et al., 1984), this indicates that the alternative 5′-cap structure likely resembles a m_3G-cap with some potential minor changes outside the m_3G region.

Mlp1 association with both RNA polymerase II transcripts snRNAs U1, U2, U4 and U5 and a direct protein-protein interaction with the core snRNP protein Sm-B led us to investigate if Mlp1 depletion affects snRNP assembly. We performed glycerol gradient centrifugation of wild type and partial Mlp1 knockout lysates and found that the snRNPs sediment in lighter fractions, indicating that Mlp1 functions in assembly of these snRNPs or that the snRNP complexes are incomplete. Next, we investigated if Mlp1 depletion would have an effect on any of the spliceosome assembly steps. An important step during spliceosome assembly is formation of U4/U6·U5 tri-snRNP which associates with the U1 and U2 snRNPs assembled on the intron sequence (Gehring & Roignant, 2021). Assembly of U4/U6 snRNP involves extensive RNA-RNA interactions between U4 and U6 snRNA leading to formation of the U4/U6 snRNP, which is followed by recruitment of the U5 snRNP. Multiple factors are found within the U4/U6·U5 tri-snRNP complex, including the Sm-proteins associated with U4 and U5 snRNAs and LSm-proteins which are found at the 3′-end of the U6 snRNA. In addition, multiple other proteins are found within the U4/U6·U5 tri-snRNP complex (Hardin et al., 2015). We found that Mlp1 immunoprecipitated the U4/U6 di-snRNA complex as opposed to the unpaired snRNA and that upon Mlp1 depletion, the formation of the U4/U6 di-snRNP complex decreased. We hypothesize that Mlp1 functions as an RNA chaperone to promote the folding and assembly of the U4/U6 di-

snRNP, which then results in U5 snRNP association. Inappropriate assembly during this step caused by absence of Mlp1 would then manifest in reduced splicing efficiency, consistent with our intron-retention splicing assay, where we found reduced splicing efficiency in the partial Mlp1 knockout strain compared to the wild type strain. Currently, we are unable to conclude if Mlp1 also influences assembly of the U4/U6·U5 tri-snRNP complex. To determine reduced assembly of the U4/U6·U5 tri-snRNP we are going to compare whole cell extracts separated on a native polyacrylamide gel between wild type and partial Mlp1 knockout strains, as performed previously (Pannone et al., 1998). In addition, immunoprecipitation of a tri-snRNP specific protein and quantification of associated snRNAs between wild type and partial Mlp1 knockout strains would also elucidate Mlp1 function in U4/U6·U5 tri-snRNP assembly.

Mlp1 interacts with all mature spliceosomal snRNAs and affects assembly of all snRNPs moderately. To elucidate if Mlp1, in addition to U4/U6 snRNP assembly, also affects other steps during spliceosome assembly, we are going to perform an *in vitro* spliceosome assembly with wild type and partial Mlp1 knockout extracts using a radioactively labeled model pre-mRNA which contains all splicing signals required for spliceosome assembly and intron removal, as performed in (Tanackovic et al., 2011). This assay allows us to observe formation of the different spliceosome assembly complexes, including A complex, B complex and C complex over time (see **Figure 21**).

In addition of association of Mlp1 with RNA polymerase II spliceosomal snRNAs U1, U2, U4 and U5, Mlp1 also interacts with RNA polymerase III U6 snRNA. This observation is not unexpected since U6 is an RNA polymerase III transcript which obtains the high-affinity 3′-uridylate sequence for La binding. An important outstanding question that remains to be answered is if Mlp1 associates directly with the U4 snRNA for example, or if association is through binding of U6 through the 3′-uridylate sequence resulting in detection of U4. In order to provide an answer

to this question, we could perform cross-linking and immunoprecipitation (CLIP), followed by an RNase digest step to determine direct interaction of Mlp1 with mature snRNAs. Another outstanding question is how Mlp1 interacts with the U1, U2, U4 and U5 mature snRNAs. This could be either through binding of secondary structures or through specific binding with the internal uridylate sequences that comprise of the Sm-binding sites. This site is also where the Sm heteroheptameric ring complex binds the snRNA. We have not yet tested the ability of Mlp1 for binding with an internal uridylate sequence. Future work therefore includes *in vitro* electrophoretic mobility shift assays (EMSAs) using mature snRNA sequences to identify binding affinity for Mlp1. Mutational analysis of the snRNA through disruption of secondary structures and mutation of the internal uridylate-rich Sm-binding site is going to determine how Mlp1 associates with these RNAs.

In conclusion, this study presents evidence that the structurally distinct genuine La protein Mlp1 is involved in splicing of pre-mRNAs in *T. thermophila*. The association of Mlp1 with both snRNAs and a core Sm protein could indicate that Mlp1 assists in assembly of the snRNP, which would explain the downstream effects observed upon Mlp1 depletion, including decreased assembly of snRNPs, decreased formation of the U4/U6 di-snRNP through RNA-RNA interactions and a decrease in splicing efficiency of pre-mRNAs. Previous work in yeast demonstrated that Lhp1p stabilized nascent U6 snRNA transcripts (Pannone et al., 1998) and that together with a mutant core Sm protein Sm-D, Lhp1p is required for U4/U6 di-snRNP and U4/U6·U5 tri-snRNP accumulation at lower temperatures (Xue et al., 2000). Another study identified spliceosome-associated factors in human cells through isolation of pre-mRNA bound catalytically active spliceosomes followed by mass spectrometry. More than 300 proteins were identified, including all previously described essential human splicing factors and 96 novel proteins, including the human La protein (Rappsilber et al., 2002). These data are consistent with a potentially conserved function for La proteins in pre-mRNA splicing.

3.5. Methods

Motif prediction.

We obtained DNA sequences for U1, U2, U4 and U5 up to 200 nucleotides downstream of the gene annotations (U1: chr_177:903140-902977, U2: chr_121:99366-99174, U4: chr_131:426400-426267, U5: chr_058:3123508-3123391) using the UCSC genome browser. Next, GLAM2 (Gapped Local Alignment of Motifs) (MEME Suite 5.5.0) was used to determine novel motifs in these genomic sequences.

Tetrahymena cultivation and protein extraction.

T. thermophila was grown in liquid SPP media [1% proteose peptone, 0.1% yeast extract, 0.2% glucose, 0.003% Fe-EDTA] at 30°C shaking at 90 RPM to mid-log phase (0.1 - 1 x 10^6 cells/mL) (Fillingham et al., 2001). Cell pellets were harvested by centrifugation for 3 minutes at 1000 g, followed by two washes in 1X PBS, pH 7.4. Pellets were stored at -80°C unless stated otherwise.

Cell pellets were resuspended in 10% Trichloroacetic acid (TCA) in 1X PBS. Following complete homogenization, the samples were incubated overnight at -20°C to enhance protein precipitation. The protein pellet was collected by centrifugation for 15 minutes at 10,000 g at 4°C, followed by two washes in 100% ice-cold acetone. The pellet was air-dried and resuspended in 50µL/mL culture 2.5X loading dye [5X loading dye: 5% β-mercaptoethanol (v/v), 0.02% bromophenol blue (w/v), 30% glycerol (v/v), 10% sodium dodecyl sulfate (SDS) (w/v), 250 mM Tris-HCl, pH 6.8].

RNP- and m_3G-cap immunoprecipitations.

Cell pellets from 100 mL cultures were collected at log phase (0.1 - 1 x 10^6 cells/mL) and washed twice with 1X PBS prior to cross-linking with 1% formaldehyde for 10 minutes, followed

by quenching with 0.25 M glycine for 5 minutes at room temperature. Next, cells were washed twice in 1X PBS and store at -80°C until further use. Cell pellets were lysed in 2 mL ice-cold buffer A [30 mM Tris-HCl pH 7.4, 150 mM NaCl, 20 mM KCl, 2 mM MgCl2, 0.1% Triton-X100, 1 mM phenylmethylsulfonyl fluoride (PMSF), 1X Halt Protease Inhibitor Cocktail (PIC) (ThemoFisher Scientific), 100U SUPERase In RNase Inhibitor (ThermoFisher Scientific)] followed by sonication (25% amplification with 15s intervals for 4 minutes). Lysates were obtained by centrifugation at 20,000 g for 45 min at 4°C.

For RNP-immunoprecipitation, the lysates were pre-cleared using rabbit IgG isotype control-bound Protein G Dynabeads (12 µg / 50 µL beads slurry) rotating for 1 hour at 4°C. Immunoprecipitations were performed using an affinity purified rabbit anti-Mlp1 antibody (ThermoFisher Scientific – Custom Antibodies) and a rabbit IgG control coupled to Protein G Dynabeads rotating 1 hour at 4°C (12 µg / 50 µL beads slurry). The Protein G Dynabeads were washed 5 times using buffer A (without PIC, PMSF and RNase inhibitor). Input and eluted RNA was isolated following reverse cross-linking for 45 minutes at 70°C in buffer B [50 mM Tris-HCl pH 7.4, 5 mM EDTA pH 8.0, 10 mM DTT, 1% SDS], followed by Trizol extraction. RNA samples were prepared for analysis on a denaturing urea polyacrylamide gel followed by Northern blotting or as input RNA for m_3G-cap RNA immunoprecipitations.

Native immunoprecipitation of Mlp1 was completed following the RNP-immunoprecipitation protocol, with one change to avoid denaturation of RNA-RNA complexes. Instead of reverse cross-linking, the RNA samples were proteinase K treated (40 µg in 0.5% SDS, 30 minutes rotating at room temperature) (ThermoFisher Scientific, O0491) to remove protein from the samples. Next, proteinase K treated samples were Trizol extracted to isolate RNA and prepared for native gel electrophoresis.

Immunoprecipitations of m_3G-capped RNA was performed as described previously (Hayes et al., 2018) with a few minor adaptations. In brief, RNA samples obtained from input and Mlp1 RNP-immunoprecipitation were resuspended in NET-2 buffer [50 mM Tris-HCl, pH 7.4; 150 mM NaCl; 0.05% NP40] and immunoprecipitated using an anti-m_3G antibody (Sigma, MABE302) coupled to Protein A Dynabeads rotating overnight at 4°C (10 µg / 50 µL beads slurry). The Protein A Dynabeads were washed 5 times using NET-2 buffer, followed by resuspension in G-50 buffer [20 mM Tris, pH 7.4; 300 mM sodium acetate, 2 mM EDTA, pH 8.0; 0.25% SDS]. RNA was isolated using phenol:chloroform:isoamylalcohol extraction (25:24:1). The aqueous layer obtained by centrifugation at 20,000 g for 10 minutes at 4°C was ethanol precipitated in 150 mM sodium acetate, pH 5.1 and 30 µg GlycoBlue (ThermoFisher Scientific) for at least 1 hour at -80°C.

Northern blotting.

RNA samples for 7 M urea denaturing urea 15% polyacrylamide gels were prepared by adding an equal volume of 2X formamide loading dye [80% deionized formamide, 0.06% (w/v) bromophenol blue, 0.06% (w/v) xylene cyanol, 10 mM EDTA, pH 8.0]. To fully denature the RNA, samples were placed at 95°C for 5 minutes, immediately followed by snap cooling on ice. RNA samples for 8% non-denaturing native polyacrylamide gels were resuspended in 1X native loading dye [10% glycerol and bromophenol blue] without heating.

The RNA was separated on a denaturing or non-denaturing polyacrylamide gel at 4°C at 100V. Next, the gel was stained with 10 µg Ethidium Bromide (ThermoFisher Scientific, 15585011) for 5 minutes at room temperature and developed under UV irradiation, transferred onto a charged nylon transfer membrane (PerkinElmer, NEF1017001PK), followed by cross-linking using a UV Stratalinker and drying at 80°C for 15 minutes on a Gel Dryer (Bio-Rad). Next, the membranes were hybridized for 2 h at the corresponding probe melting temperature (Tm) -10°C in

hybridization buffer [6X SSC (1X SSC: 150 mM NaCl and 15 mM sodium citrate), 1% SDS and 4X Denhardt's solution (Bio Basic D0062)], followed by overnight incubation with T4 Polynucleotide Kinase (PNK) ^{32}P-labeled probes (**Table 4**). Prior to exposure to a storage phosphor screen overnight, the membranes were washed 3 times for 20 minutes in wash buffer [2X SSC and 0.1% SDS]. Storage phosphor screens were developed on a Typhoon imager. To strip hybridized probe, membranes were incubated three times with stripping buffer [0.1X SSC and 0.1% SDS] for 20 minutes each at 70°C.

Table 4. Northern blot probes used for snRNA detection.

Primer name	Tm (°C)	Sequence
U1 snRNA	61.0	5' TCCCCTAGCAGAAATTGTGCTGTT 3'
U2 snRNA	58.5	5' AGCCACTAGTAAATCCGACCCC 3'
U4 snRNA	58.0	5' CGACCTTCCAGAAATTGGGACC 3'
U5 snRNA	59.0	5' CACATTGAAAAACCCAGCTCACGG 3'
U6 snRNA	56.9	5' AAAAATGGGGAATCCTTCTCGCT 3'
5.8S rRNA	54.6	5' CTGCAATTCGCATTGCGT 3'
mature tRNA Tyr^{GTA}	57.7	5' GACCACCGGATTTACAGTCCG 3'

qRT-PCR.

Total RNA was extracted from *T. thermophila* wild type and partial Mlp1 knockout cell pellets using Trizol extraction. The purified RNA samples were resuspended in 1X Reaction Buffer with 2 units of Turbo DNase (ThermoFisher Scientific, cat# AM2238) and incubated for 30 minutes at 37°C. The DNase-treated samples were subsequently Trizol extracted to inactivate DNase activity. Next, 50 ng/µL DNase-treated RNA was reverse transcribed using the iScript cDNA Synthesis Kit (Bio-Rad, cat#1708891) according to manufacturer's instructions. The cDNA was diluted to 25 ng (in 12.5 µL per technical triplicate) and quantified using the SensiFAST SYBR No-Rox kit (Bioline, BIO-98005) with 12.5 µM of each primer (**Table 5**) (1x forward, 1x reverse) using following qPCR settings: 95°C for 5 minutes and 35 cycles of 5 sec at 95°C and 15 sec at

60°C, followed by a melting curve analysis up to 99°C to confirm amplification of a single amplicon. For RNA-immunoprecipitation, fold enrichment was calculated using the ΔCt method ($Ct_{RIP} - Ct_{input}$), followed by fold enrichment calculation by taking Mlp1-RIP / IgG-RIP ratios. For comparison between wild type and partial Mlp1 knockout strains, fold enrichment was calculated using the ΔCt method ($Ct_{5.8S\ rRNA} - Ct_{RNA}$), followed by fold enrichment calculation by taking partial Mlp1 knockout / wild type ratios.

Table 5. qRT-PCR primers used for snRNA detection.

	Primer name	Sequence	
	5S rRNA For	5' GTCGGCCATACTAAGGTGAAAAC	3'
	5S rRNA Rev	5' GGCTGTCGACACTGGGACTT	3'
	17S rRNA For	5' GGCCGTTCTTAGTTGGTGGA	3'
	17S rRNA Rev	5' TCGCAGTCCAGAAGGTTCAC	3'
	U1 snRNA For	5' CTTACCTGGCTGGAGTTTGCT	3'
	U1 snRNA Rev	5' TCCCCTAGCAGAAATTGTGCTGTT	3'
	U2 snRNA For	5' CCTTCTCGGCCTTTTGGCTAAG	3'
	U2 snRNA Rev	5' AGCCACTAGTAAATCCGACCCC	3'
	U4 snRNA For	5' CTCGCGACAGGGGCAATATAG	3'
	U4 snRNA Rev	5' CGACCTTCCAGAAATTGGGACC	3'
	U5 snRNA For	5' TCACAGAACTCAGCTCATTACGA	3'
	U5 snRNA Rev	5' CACATTGAAAAACCCAGCTCACGG	3'
	U6 snRNA For	5' CCCGAAAGGGTCATCCGTT	3'
	U6 snRNA Rev	5' AAAAATGGGGAATCCTTCTCGCT	3'
	pre-tRNA Tyr^{GUA} For	5' CTGTAAAGTTTGTAGTCATCCGGTG	3'
	pre-tRNA Tyr^{GUA} Rev	5' AAAAAAAATCCGAACTCCCGGG	3'
	pre-tRNA Leu^{UAA} For	5' GTTCATCCTCCTACATTTAGAAAAGAGTGTCG	3'
	pre-tRNA Ile^{UAU} For	5' GCTGACCAACCTTATAAAAGTGTGTCG	3'
	pre-tRNA Leu^{UAA}/Ile^{UAU}Rev	5' AAAAAAAAAGTAGCTCAGAGAGGGTTC	3'
	Mlp1 For	5' AACGTCCATGCCCCTTACTG	3'
	Mlp1 Rev	5' ACACCGTTGTAGTGACGACC	3'

Protein-protein immunoprecipitation and mass spectrometry.

Protein-protein immunoprecipitation to identify direct interactors of Mlp1 was performed as described in (Nabeel-Shah et al., 2021) using a rabbit anti-Mlp1 affinity purified antibody for endogenous immunoprecipitation of Mlp1 (ThermoFisher Scientific – Custom Antibodies) (previously characterized in Kerkhofs et al. 2022). In brief, cell pellets from 500 mL liquid *T. thermophila* culture were thawed on ice and lysed using 15 mL of 1X lysis buffer [40 mM Tris-HCl, pH 8.0; 600 mM NaCl; 0.2% NP-40; 1 mM MgCl2; 1 mM PMSF and 1X PIC (ThermoFisher Scientific)]. The lysate was treated with 50 units of benzonase nuclease (Millipore, 70746) rotating for 30 minutes at 4°C, followed by centrifugation at 16,100 g for 30 minutes at 4°C to clarify the lysate. Immunoprecipitations were performed using rabbit anti-Mlp1 and rabbit IgG control coupled to Protein G Dynabeads rotating 2 hours at 4°C (12 µg / 50 µL beads slurry). The Protein G Dynabeads were washed once using high-salt IPP300 wash buffer [10 mM Tris-HCl, pH 8.0; 300 mM NaCl, 0.1% NP-40], followed by two washes with IPP100-NP-40 wash buffer [10 mM Tris-HCl, pH 8.0; 100 mM NaCl, 0.1% NP-40] and two washes with IPP100 wash buffer [10 mM Tris-HCl, pH 8.0; 100 mM NaCl]. Next, the Protein G Dynabeads were resuspended in freshly prepared 0.5 M NH_4OH and rotated for 20 minutes at room temperature. The samples were lyophilized using a vacuum concentrator (Eppendorf Vacufuge). Next, samples were fragmented using in-solution tryptic digestion and LC-MS/MS analysis was performed by the IRCM Proteomics Discovery Platform on a Q Exactive HF. Data was analyzed using Scaffold v5.2.0 using following settings: 95% protein threshold, 2 minimum number of peptides and 5% False Discovery Rate (FDR) peptide threshold, identifying 1086 proteins. Total Spectrum Counts were used for statistical analysis using a one-tailed t-test. Fold change was calculated by taking the ratio between Mlp1- and IgG-immunoprecipitated samples. Volcano plots were generated using GraphPad Prism v9.0. Silver staining was performed following

protein separation a 12% SDS polyacrylamide gel using the ProteoSilver Silver Stain Kit (Millipore Sigma, PROTSIL1) following manufacturer′s instructions.

Cellular fractionation and Western blotting.

T. thermophila cell pellets from 25 mL cultures were collected at log phase (0.1 - 1 x 10^6 cells/mL) (without freezing) and washed once with 1X PBS and once with lysis buffer without Triton X-100 [10 mM Tris-HCl, pH 7.4; 5 mM $MgCl_2$; 10 mM KCl]. Next, the pellet was lysed in 500 µL ice-cold complete lysis buffer [10 mM Tris-HCl, pH 7.4; 5 mM $MgCl_2$; 10 mM KCl, 0.05% Triton X-100, 1X PIC (ThemoFisher Scientific), 40 U SUPERase In RNase Inhibitor (ThermoFisher Scientific)] by centrifugation at 2500 g for 5 minutes at 4°C. The supernatant, containing the cytoplasmic fraction, was carefully collected and the pellet, containing nuclei, was washed with lysis buffer without Triton X-100 twice to remove any non-specific cytoplasmic components. Fractions were divided into two equal part for RNA extraction using Trizol and Northern blot analysis and protein analysis using Western blot.

Prior to separation on a 12% SDS polyacrylamide gel, samples were resuspended in 1X SDS loading dye [5X SDS loading dye: 5% β-mercaptoethanol (v/v), 0.02% bromophenol blue (w/v), 30% glycerol (v/v), 10% SDS (w/v), 250 mM Tris-HCl, pH 6.8] and incubated at 95°C for 10 minutes. Next, the gel was transferred onto a nitrocellulose membrane and blocked in 0.5% (w/v) skim milk powder in Tris-Buffered Saline [20 mM Tris-HCl pH 7.4, 150 mM NaCl] + 0.1% Tween20 (TBST) for 1 hour at room temperature (or overnight at 4°C). Primary antibodies were added in TBST for 1 hour at room temperature (or overnight at 4°C) and washed 5 times in TBST. Secondary HRP-conjugated secondary antibodies were added at 1:10,000 dilutions and incubated for 1 hour at room temperature. Primary antibodies used in this study: mouse anti-beta actin (abcam ab8224), rabbit anti-histone H3 (abcam ab1791), and affinity purified antibody (ThermoFisher Scientific – Custom Antibodies) rabbit anti-Mlp1 (previously characterized in

Kerkhofs et al. 2022). Secondary HRP-coupled antibodies used in this study: goat anti-rabbit IgG (Cell Signaling Technology 7074) and horse anti-mouse IgG (Cell Signaling Technology 7076).

Conservation analysis.

Accession numbers were obtained from the National Center for Biotechnology Information (NCBI) to obtain primary amino acid sequences for Sm-B in *H. sapiens* (NP_003082.1), *M. musculus* (NP_033251.1), *D. rerio* (NP_991230.1), *D. melanogaster* (NP_001188780.1), *A. thaliana* (NP_001032011.1), *S. pombe* (NP_594151.1), *S. cerevisiae* (NP_010946.3), and *T. thermophila* (XP_001031807.1). NCBI accession numbers for the TGS1 enzyme included: *H. sapiens* (NP_079107.6), *M. musculus* (NP_473430.3), *D. rerio* (XP_003197913.2), *D. melanogaster* (NP_732316.2), *D. discoideum* (XP_636670.1), *A. thaliana* (NP_973978.2), *S. pombe* (NP_593095.3), *S. cerevisiae* (NP_015168.1) and *T. thermophila* (XP_001021693.2). Primary amino acid alignments were obtained using Clustal Omega (EMBL-EBI) (Sievers et al., 2011) and annotation of identical and conserved amino acids was completed using a custom python script. Amino acids were grouped as conserved based on side chain diversity: (i) Asp (D), Glu (E), Asn (N), Gln (Q); (ii) Lys (K), Arg (R), His (H); (iii) Phe (F), Trp (W), Tyr (Y); (iv) Val (V), Ile (I), Leu (L), Met (M); and (v) Ser (S), Thr (T). High-resolution structure of the *H. sapiens* U4 snRNP core was obtained from PDB: 4WZJ and *H. sapiens* TGS1 enzyme in complex with m^7GTP and AdoMet substrates was obtained from PDB: 3GDH. Modeling was done using PyMOL. Structure prediction of the catalytic domain (amino acids 283–498) of *T. thermophila* Tgs1 (http://www.ciliate.org/ TTHERM_00151600) was obtained using the AlphaFold Colab notebook.

Glycerol gradients.

T. thermophila cell pellets from 50 mL cultures were lysed in 500 µL lysis buffer [50 mM Tris-HCl, pH 7.4; 250 mM KCl; 10 mM $MgCl_2$; 0.5 mM DTT; 0.2 mM EDTA, pH 8.0; 10% glycerol; 1 mM PMSF, 1X PIC (ThemoFisher Scientific), 25U SUPERase In RNase Inhibitor (ThermoFisher Scientific)]. The lysate was obtained by centrifugation at 20,000 g for 30 minutes at 4°C. We prepared 5-step 10-30% glycerol gradients in buffer C [20 mM HEPES, pH 7.4; 150 mM KCl; 1 mM MgOAc; 0.1% NP40 and 0.5 mM PMSF] in advance and kept at 4°C overnight to diffuse into a linear gradient. The lysate was layered gently on top of the glycerol gradient and centrifuged for 20 hours at 35,000 RPM (Beckman SW41Ti rotor) at 4°C.

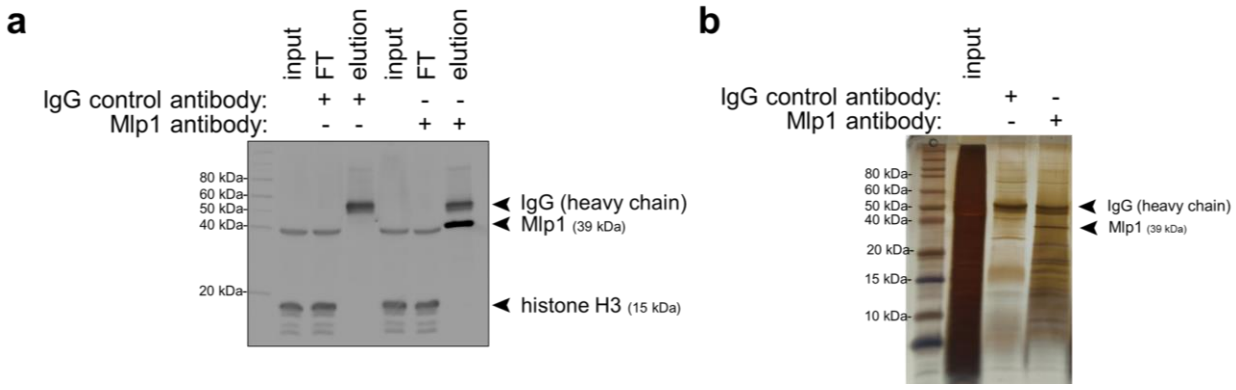
Fractions were collected (400 uL/fraction) and RNA was isolated for Northern blot analysis using phenol:chloroform:isoamyl alcohol (25:24:1) extraction (1:1 volume) followed by ethanol precipitation of the aqueous layer in 150 mM sodium acetate, pH 5.1 and 30 µg GlycoBlue (ThermoFisher Scientific) for at least 1 hour at -80°C. The RNA pellet was washed with 70% ethanol and airdried, followed by resuspension in RNase-free ddH_2O. Protein samples for Western blot analysis were obtained by Trichloroacetic acid TCA precipitation where the final concentration was brought to 10%, followed by incubation at –20°C overnight. The protein pellet was collected by centrifugation for 15 minutes at 10,000 g at 4°C, then washed twice with 100% ice-cold acetone. The pellet was air-dried and resuspended in 50µL/mL culture 2.5X loading dye [5X loading dye: 5% β-mercaptoethanol (v/v), 0.02% bromophenol blue (w/v), 30% glycerol (v/v), 10% SDS (w/v), 250 mM Tris-HCl, pH 6.8].

Splicing assay.

Total RNA was isolated from *T. thermophila* wild type and partial Mlp1 knockout cell pellets using Trizol extraction, followed by TurboDNase (2U) treatment in 1X Reaction Buffer (ThermoFisher Scientific, cat# AM2238) for 30 minutes at 37°C and Trizol extracted to remove

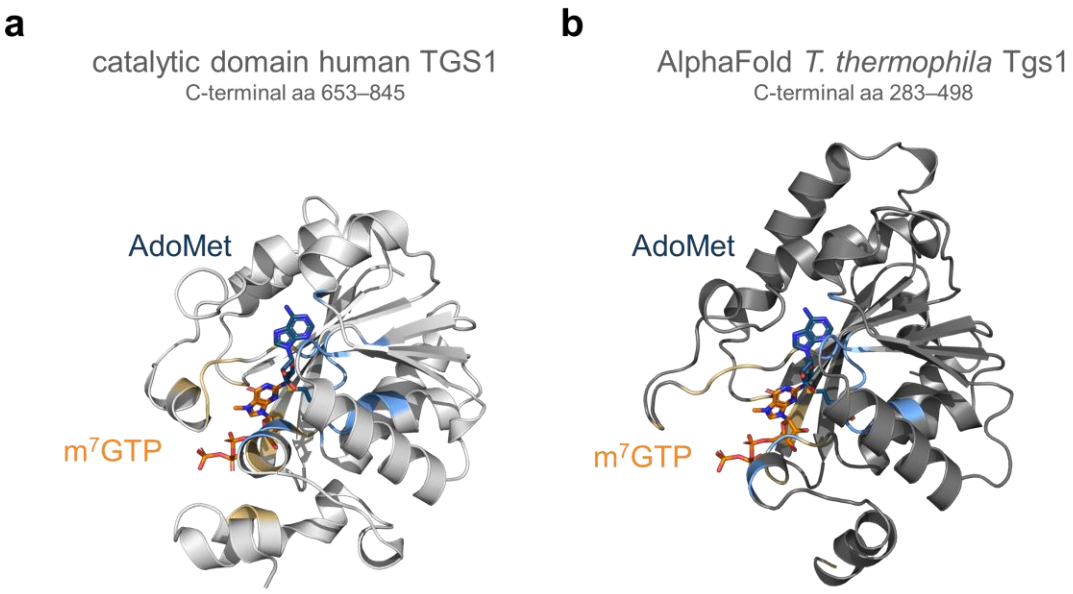
the DNase enzyme. Next, 50 ng/µL DNase-treated RNA was reverse transcribed using the iScript cDNA Synthesis Kit (Bio-Rad, cat#1708891) according to manufacturer′s instructions. *Taq* DNA Polymerase (New England Biolabs, M0273L) was used to amplify DNA fragments from 5 ng cDNA according to manufacturer′s instructions using following PCR settings: initial denaturation for 5 min at 95°C, 30 cycles of 30 sec at 95°C, 30 sec at 45°C and 45 sec at 68°C, followed by a final extension of 5 min at 68°C. DNA amplicons were resuspended in 1X TrackIt Cyan/Orange Loading Buffer and separated by electrophoresis through a 1% agarose gel containing 50 µg / 100 mL Ethidium Bromide (ThermoFisher Scientific, 15585011).

2.7. Supplementary Figures



Supplementary Figure 10. Mlp1 protein-protein immunoprecipitation validation.

a Western blot to confirm Mlp1-specific immunoprecipitation from *T. thermophila* using an affinity purified rabbit anti-Mlp1 antibody and a rabbit isotype immunoglobin G (IgG) control. The Mlp1-specific antibody was used for both ribonucleoprotein-immunoprecipitation (RNP-IP) and western blotting (n=2 biological replicates). The heavy chain (50 kDa) of the antibodies was detected with the secondary anti-rabbit antibody and shown as IgG. Loading control: histone H3. **b** Silver stain of benzonase treated Mlp1-immunoprecipitation for identification of protein-protein interactors (n=1 biological replicate). IgG: immunoglobin G, negative control.



Supplementary Figure 11. Structure comparison of the Tgs1 catalytic domain between human and *Tetrahymena thermophila.*

a High-resolution structure of human TGS1 (PDB: 3GDH) with substrates AdoMet and m^7GTP binding sites highlighted in blue and orange, respectively. **b** AlphaFold predicted structure of *T. thermophila* Tgs1 with substrates AdoMet and m^7GTP binding sites highlighted in blue and orange, respectively. Highlighted residues are based on primary sequence alignments shown in **Figure 29b** and AdoMet and m^7GTP spatial orientation obtained from alignment with human TGS1 (PDB: 3GDH). AdoMet: S-adenosylmethionine, aa: amino acids.

Chapter IV:

The *Tetrahymena thermophila* La protein Mlp1 associates with box C/D and box H/ACA snoRNAs

Author contributions

The *Tetrahymena thermophila* La protein Mlp1 associates with box C/D and box H/ACA snoRNAs

Kerkhofs, Kyra[1]; Fafard-Couture, Étienne[2]; Laganas, Jarred[1]; Scott, Michelle S.[2]; Pearlman, Ronald E.[1]; Bayfield, Mark A.[1]

[1] Department of Biology, Faculty of Science, York University, Toronto, Ontario N3J 1P3, Canada,
[2] Département de Biochimie et de Génomique Fonctionnelle (LGDP), Faculté de Médecine et des Sciences de la Santé, Université de Sherbrooke, Sherbrooke, Québec, J1E 4 K8, Canada,

Manuscript in preparation

Author contributions:

Étienne Fafard-Couture performed Thermostable Group II Intron Reverse Transcriptase (TGIRT)-sequencing analysis generating transcripts per million (TPM) and counts per million (CPM) data for small nucleolar RNAs (snoRNAs). Étienne Fafard-Couture completed snoRNA predictions using rFam, snoscan and SnoReport2.0. Jarred Laganas and Ronald E. Pearlman provided annotated genomic locations of previously published snoRNAs in *Tetrahymena thermophila*.

4.1. Abstract

Small nucleolar RNAs (snoRNAs) are located in the nucleolus as small nucleolar ribonucleoproteins (snoRNPs) and function as guide RNAs in post-transcriptional modification of rRNA, snRNA and tRNAs. The box C/D snoRNAs guide 2′-O-methylation and H/ACA snoRNAs guide pseudouridylation. We discovered that Mlp1, the genuine La protein in *Tetrahymena thermophila*, interacts with both box C/D and box H/ACA snoRNAs that were previously described. We found that Mlp1 depletion has no effect on snoRNA transcript abundancy. We used Mlp1 association with all previously identified snoRNAs as a metric to identify novel snoRNAs in *Tetrahymena thermophila* in combination with bioinformatics tools leading to prediction of 18 high-confidence and highly expressed box C/D snoRNAs. In addition to mature snoRNA binding, we found that Mlp1 co-immunoprecipitates core snRNP proteins from both box C/D and box H/ACA snoRNP complexes, indicating that Mlp1 might function in snoRNP assembly or activity. We tested snoRNP assembly using glycerol gradients, however, we found that Mlp1 depletion has no effect on assembly. We tested snoRNP functionality using a RNase H-dependent 2′-O-methylation detection assay to determine post-transcriptional modification levels in rRNA. Again, we found no difference between wild type and partial Mlp1 knockout strains. Our results indicate that Mlp1 associates with snoRNAs in *Tetrahymena thermophila*, and that this can be used to predict new snoRNAs in this system.

4.2. Introduction

In eukaryotes, rRNA is transcribed in the nucleolus by RNA polymerase I as a pre-rRNA containing the 18S, 5.8S and 28S rRNAs and including the external (5′-ETS and 3′-ETS) and internal (ITS-1 and ITS-2) spacers. Maturation of rRNA involves removal of the external and internal spaces by nucleolytic processing and multiple post-transcriptional modifications by small nucleolar RNAs (snoRNAs) in the nucleolus (Klinge & Woolford 2019). Post-transcriptional modifications include 2′-O-methylation and pseudouridylation in functionally important regions of the ribosome, including the peptidyl transferase center, the decoding center (A, P and E sites) and regions important for intersubunit interactions (Decatur & Fournier, 2002).

RNA polymerase II transcribes most snoRNAs as precursors that are required to undergo maturation at the 5′- and 3′-ends. The fully processed snoRNAs lack a poly(A) tail and are between 60 and 300 nucleotides long (Falaleeva & Stamm 2013). snoRNAs complex with proteins to generate small nucleolar ribonucleoproteins (snoRNPs), which localize to the nucleolus and direct post-transcriptional modifications on target RNAs (Kiss, 2001). Three classes of snoRNAs exist based on structural elements. One class contains only one member, the MRP RNA which functions as the RNA subunit of the mitochondrial RNase mitochondrial RNA processing enzyme complex (Tollervey & Kiss, 1997). The two other snoRNA classes have more members and comprise the box C/D and box H/ACA snoRNAs, which are classified based on the presence of conserved sequences (*i.e. "*boxes") (**Figure 35**) (Kiss, 2002). Most of the box C/D snoRNAs guide 2′-O-methylation, while most H/ACA snoRNAs guide pseudouridylation (Kiss, 2002). Both 2′-O-methylation and pseudouridylation are associated with stabilization of double-stranded RNA structure (Decatur & Fournier, 2002).

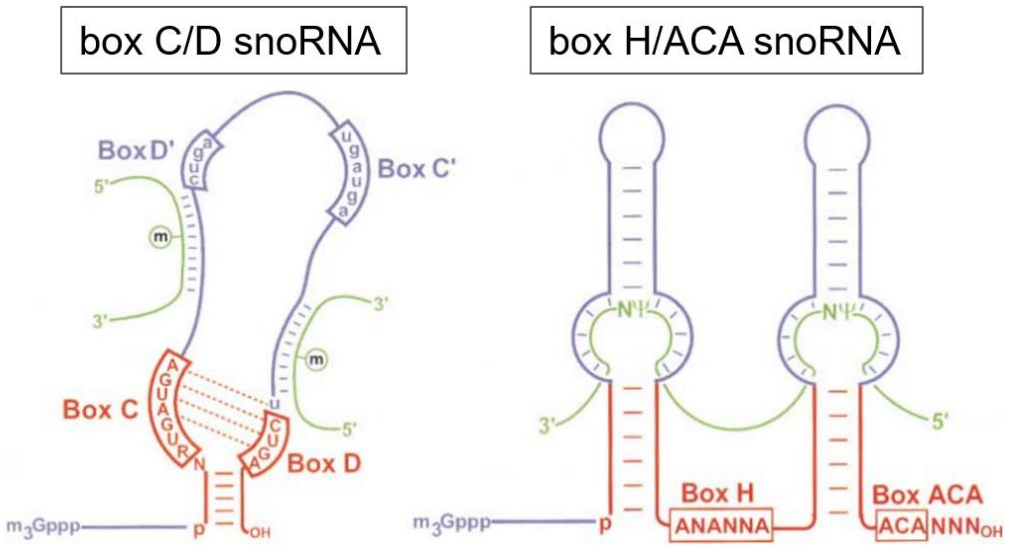


Figure 35. Schematic representation of structure and function of box C/D and box H/ACA snoRNAs.

The box C/D snoRNA adopts a closed stem loop containing two highly conserved box C (RUGAUGA) and box D (CUGA) motifs shown in red and two less conserved box C′ and box D′ located in the loop region. The box H/ACA snoRNAs contain a highly conserved ACA sequence near the 3′-end and two stem loops that are connected by the H box (ANANN motif) with the highly conserved box H and box ACA highlighted in red. Target rRNAs are shown in green and are 2′-O-methylated and pseudouridylated by box C/D and box H/ACA snoRNAs, respectively. Figure adapted from (Kiss, 2002).

The box C/D snoRNA forms a closed stem loop containing two highly conserved box C (RUGAUGA) and box D (CUGA) motifs and two less conserved box C′ and box D′ motifs located in the loop region. The box H/ACA snoRNAs typically contain an ACA sequence near the 3′-end and two stem loops that are connected by the H box (ANANN motif). Both C/D and H/ACA snoRNAs assemble with distinct sets of proteins into snoRNP complexes. The box C/D snoRNPs contain four core proteins, including the methyltransferase Nop1 (fibrillarin in vertebrates), Nop56, Nop58 (or Nop5) and Snu13 (Cahill et al., 2002). Nop1/fibrillarin protein is the catalytic subunit of the box C/D snoRNP (Jansen, 1991). Box H/ACA snoRNAs likewise contain four core proteins: the pseudouridine synthetase Cbf5p (dyskerin in vertebrates), Gar1,

Nhp2 and Nop10 (Cahill et al., 2002). The catalytic subunit of H/ACA snoRNPs is Cbf5p/dyskerin.

Most snoRNAs function as guide snoRNP and contain regions of complementary to other cellular RNA substrates. The box C/D snoRNAs guide the methylation of the 2′ hydroxyl of the ribose (2-O′-methylation) and contain 10 – 21 complementarity nucleotides to the mature rRNA. The corresponding snoRNA-rRNA hybrid places the box D and D′ of the snoRNA five nucleotides from the target nucleotide for methylation (**Figure 35**) (Cavaillé et al., 1996; Kiss-László et al., 1996). For isomerization of uridine to pseudouridine by the box H/ACA snoRNAs, 3 – 10 nucleotide snoRNA/rRNA duplexes are formed within the stem loop separated by two nucleotides (**Figure 35** – indicated in green as NΨ) (Ganot et al., 1997; Ni et al., 1997).

A selection of snoRNAs, including the box C/D snoRNAs U3, U8, U14 and U22 (Kass et al., 1990; Li et al., 1990; Peculis, 1997; Tycowski et al., 1994), and the box H/ACA snoRNAs U17 and snR10 (Enright et al., 1996; Venema et al., 1997) are involved in processing of the pre-rRNA through nucleolytic cleavage. These snoRNAs function as RNA chaperones through base pairing interactions with the rRNA resulting in the correct rRNA conformation to induce nucleolytic cleavage (Henras et al., 2015; Steitz & Tycowski, 1995). The box C/D snoRNA U14 has a dual function serving as a methylation guide snoRNA and a pre-rRNA chaperone for 18S rRNA processing (Dunbar & Baserga, 1998; Li et al., 1990).

Tetrahymena thermophila contains large inverted repeats of approximately 9000 rDNA copies in the transcriptionally active macronucleus, which is produced from the parental chromosome in the micronucleus from a single copy through amplification during macronucleus formation (Kapler, 1993). Mature rRNA products in *T. thermophila* include the 17S, 5.8S and $26S_{\alpha}$ and $26S_{\beta}$ rRNAs (Kister et al., 1983). The $26S_{\alpha}$ and $26S_{\beta}$ are produced because of a cleavage site in the 26S rRNA, thereby producing two parts which are approximately the same size as the 17S

rRNA (Kister et al., 1983). Fragmented rRNA incorporation in the ribosomes has been described in multiple unicellular eukaryotes. For example, *Trypanosoma brucei* contains a 26S rRNA equivalent that is processed into six separate transcripts (Barth et al., 2008; Campbell et al., 1987)

A previous study identified 38 box C/D and 28 box H/ACA snoRNAs in the ciliate *T. thermophila* (Andersen & Nielsen, 2012). In this work, we have discovered that Mlp1, the genuine La protein in *T. thermophila,* associates with all previously identified snoRNAs. Therefore, the aim of this study was to expand identification of novel snoRNAs using our previously published small RNA-sequencing dataset (Kerkhofs et al. 2022). Using bioinformatics tools, we predicted novel box C/D snoRNAs and confirmed their expression by obtaining expression levels from our small RNA sequencing dataset. We confirmed that Mlp1 associates with mature snoRNAs rather than precursor snoRNAs and further investigated Mlp1 function in snoRNP biogenesis. We found that Mlp1 associates with protein components of the snRNP complex, however, depletion of Mlp1 has no effect on transcript stability or snRNP assembly. We also tested 2′-O-methyltation levels on rRNA for novel predicted snoRNAs but observed no defects upon Mlp1 depletion. Thus, the function of Mlp1 in snoRNP biogenesis or function is still unclear.

4.3. Results

4.3.1. Mlp1 interacts with mature box C/D and H/ACA snoRNAs

La proteins are known to associate with transcripts produced by RNA polymerase III containing 3′-terminal uridylates (Mathews & Francoeur, 1984; Stefano, 1984). The *Saccharomyces cerevisiae* La protein Lhp1p has also been shown to associate with RNA polymerase II transcript processing intermediates containing internal 3′-uridylate sequences that are exposed during processing. These transcripts include spliceosomal RNAs U1, U2, U4 and U5 (Inada & Guthrie, 2004; Xue et al., 2000). In addition, an RNP-immunoprecipitation of Lhp1p followed by microarray analysis described association with snoRNAs (Inada & Guthrie, 2004). Using the non-coding small RNA (60-100 nucleotides) dataset from Gogakos et al. 2017 we established that the human La (hLa) does not associate with snoRNAs (**Figure 36a** – compare input RNAs and hLa-bound RNAs, **Supplementary Table 4**). Next, we reanalysed our previously published small RNA-sequencing dataset (< 300 nucleotides) from Mlp1-associated and input RNA (Kerkhofs et al. 2022) using annotated chromosomal locations of previously published snoRNAs (Andersen & Nielsen, 2012). We obtained normalized counts for rRNA, tRNA, snoRNA and mRNAs in transcripts per million (TPM) for both total and Mlp1-immunoprecipitated RNA and found that box C/D and box H/ACA snoRNA abundance in total RNAs is less than 1% and 0.1%, respectively, while Mlp1-immunoprecipitation enriches box C/D and box H/ACA snoRNA abundancy to approximately 11% and more than 1%, respectively (**Figure 36b** – compare input RNAs and Mlp1-bound RNAs, **Supplementary Table 4**). These relative abundancies indicate that Mlp1 specifically enriches both box C/D and box H/ACA snoRNAs.

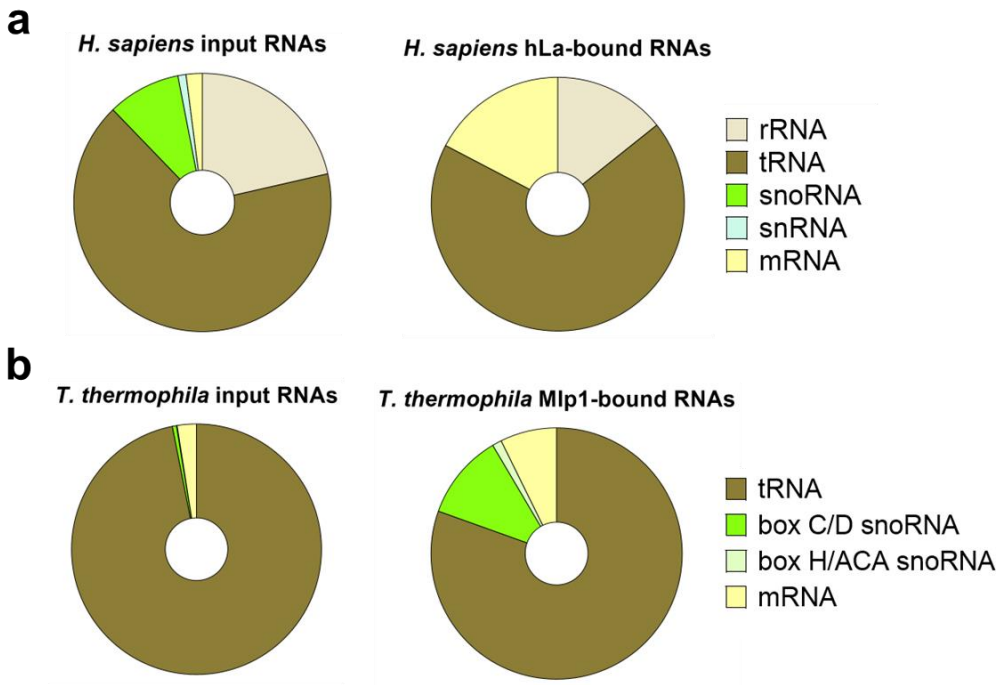


Figure 36. Mlp1 enriches snoRNAs in *Tetrahymena thermophila*.

a,**b** Pie charts representing averaged percentages of transcripts per million (TPM) reads (*i.e.* normalized RNA sequencing counts) for input RNA and hLa- (**a**) and Mlp1-associated (**b**) small non-coding RNAs and mRNA (n=3 biological replicates for Mlp1-immunoprecipitation and input and n=2 biological replicates for hLa-immunoprecipitation and n=4 for input). Raw data shown in **Supplementary Table 4**. hLa data was obtained from (Gogakos et al., 2017), Mlp1 data was obtained from (Kerkhofs et al. 2022).

Since Lhp1p has previously been shown to associate with snRNA processing intermediates without interacting with the mature snRNAs, we determined if Mlp1 associates with precursor snoRNAs or the fully processed mature snoRNAs. Using read alignments to the genome from Mlp1-bound snoRNAs and input snoRNAs, we were able to determine that the Mlp1 reads aligned strongly with the input reads, consistent with Mlp1 interacting with mature snoRNAs and not a processing intermediate (**Figure 37a**). We also confirmed interaction with the mature snoRNA by Northern blot, since the Mlp1-bound snoRNAs are the same size as the most abundant transcript in the input snoRNA lane (**Figure 37b** – compare lanes 1 and 5). Specific enrichment of snoRNAs by Mlp1 was shown by comparison against a Immunoglobin G (IgG)

immunoprecipitation control **Figure 37b**, compare lanes 3 and 5). In addition, we confirmed specific Mlp1-enrichment of snoRNAs by qRT-PCR (**Figure 37c**) where we compared fold enrichment of Mlp1- and IgG-immunoprecipitated for snoRNAs and previously described RNA polymerase III transcribed La protein targets: 5S and pre-tRNAs (**Figure 37c**). We observed a strong enrichment compared to negative control 17S rRNA transcribed by RNA polymerase I.

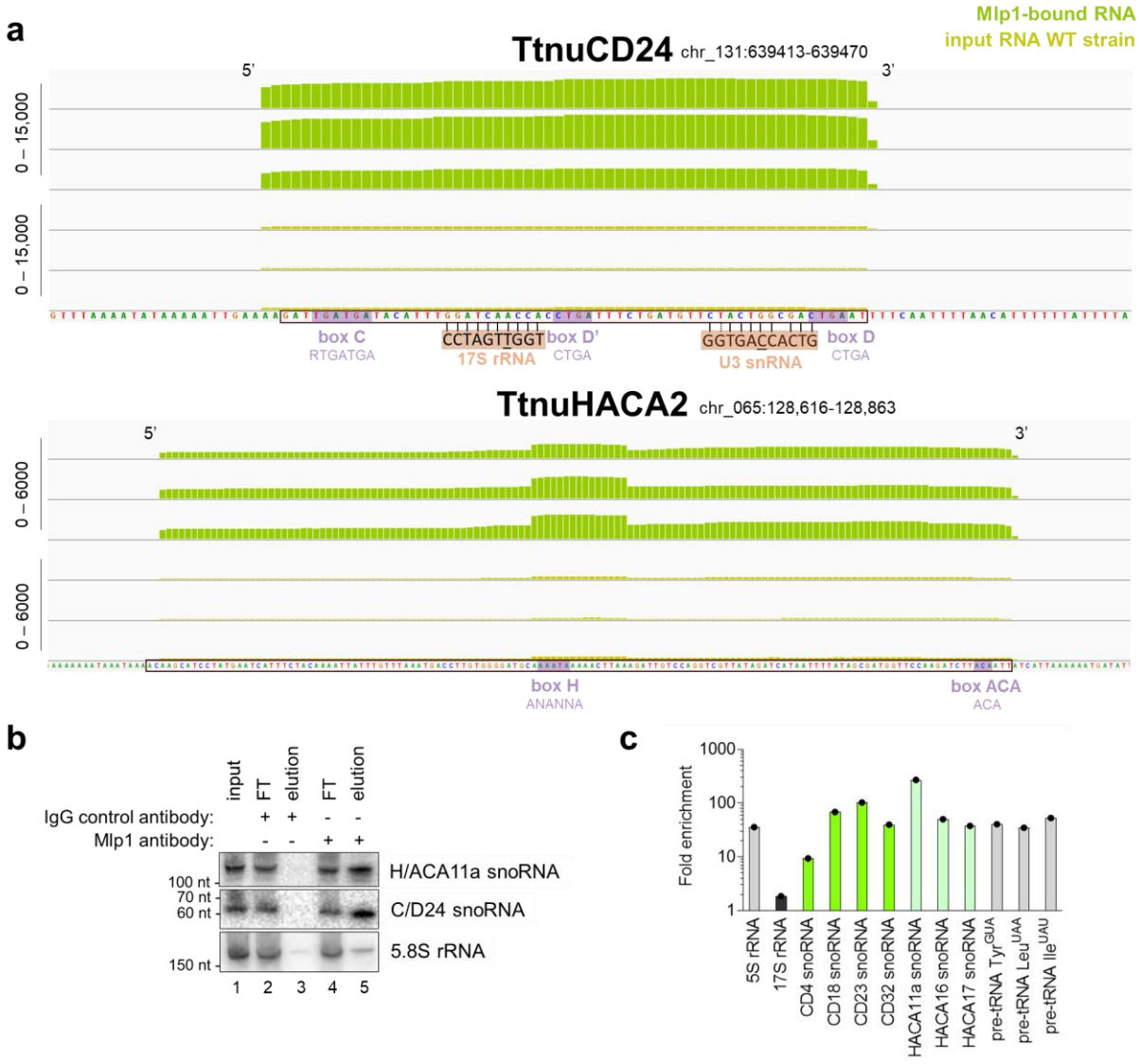


Figure legend on next page

Figure 37. Mlp1 interacts with mature snoRNAs *in vivo*.

a Chromosomal locations encoding representative examples of a box C/D and box H/ACA snoRNA. Expression levels are shown as raw counts on the y-axis. Mlp1-associated RNA is shown in green (top three rows) and total RNA is shown in yellow (bottom three rows). Mature snRNA sequences are boxed based on a previous study (Andersen & Nielsen, 2012). Typical box C, box D, box H and box ACA sequences are highlighted in purple. Complementary sequences of RNA modification targets are shown underneath the snoRNA sequence and are highlighted in orange. The nucleotide target for modification is underlined. R: purine (guanine (G) or adenine (A)), N: any nucleotide. Image was generated in Integrative Genomics Viewer (IGV). **b** Northern blot analysis of Mlp1 ribonucleoprotein-immunoprecipitated (RIP) samples from *T. thermophila*. Mlp1 enriches box C/D4 and box H/ACA258 snoRNAs specifically compared to negative IgG control (n=2 biological replicates). 5.8S rRNA was used as a negative non-binding control. IgG: immunoglobin G. **c** qRT-PCR analysis of Mlp1 RIP samples from *T. thermophila* with RNA polymerase I transcript (17S rRNA) highlighted in dark grey, box C/D and box H/ACA snoRNAs in green and RNA polymerase III transcripts in light grey. Fold enrichment was calculated using the ΔCt method ($Ct_{RIP} - Ct_{input}$), followed by fold enrichment calculation by taking Mlp1-RIP / IgG-RIP ratios (n=1 biological replicate). IgG: immunoglobin G negative control.

Mature snoRNAs are located in the nucleolus, together with ribosomal DNA (rDNA), pre-rRNA at various processing stages and multiple rRNA maturation and processing factors (Kiss, 2002). In *T. thermophila*, between 500-1000 nucleoli are present which are found at the inner surface of the nuclear envelope (Nielsen et al., 1992). From indirect immunofluorescent staining in *T. thermophila*, we have previously established localization of Mlp1 in the macronucleus (Kerkhofs et al. 2022). In addition, Mlp1 is also found at the periphery of the macronucleus where the nucleolus is located (**Figure 38**).

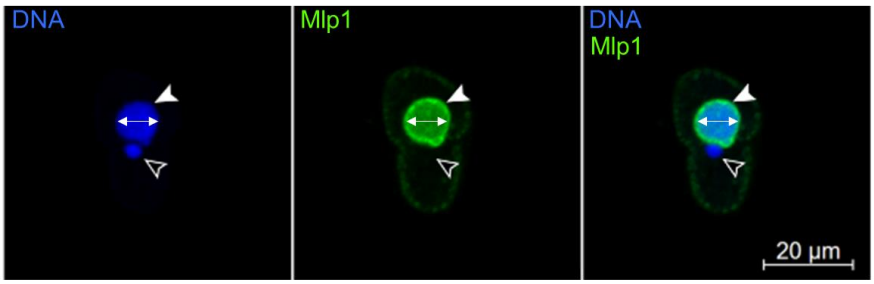


Figure 38. Mlp1 localizes to nucleoli.

Reannotation of **Supplementary Figure 7c** in Chapter II. Indirect immunofluorescent staining of *T. thermophila* wild type strains using the polyclonal rabbit anti-Mlp1 antibody. Nuclei were stained using DAPI. Full white arrows are denoting the transcriptionally active macronucleus and empty white arrows show the transcriptionally inactive micronucleus. The two-sided white arrow was added on top of the image to illustrate Mlp1 localization to nucleoli.

4.3.2. Annotation of novel expressed snoRNAs in the *Tetrahymena thermophila* genome based on Mlp1-enrichment

Since Mlp1 is located in the nucleolus and associates with both box C/D and box H/ACA snoRNAs, we next determined if Mlp1 enriches all previously published 38 box C/D and 28 box H/ACA snoRNAs (Andersen & Nielsen, 2012). We used the small RNA-sequencing dataset and after filtering low abundance reads (average TPM < 1), we calculated fold enrichment of Mlp1-associated snoRNAs relative to input snoRNAs and found that Mlp1 strongly enriches all previously described box C/D and box H/ACA snoRNAs (**Figure 39**). From our previous work, we know that Mlp1 preferentially binds pre-tRNAs compared to mature tRNAs (Kerkhofs et al. 2022). Therefore, we included averaged fold enrichments for all pre-tRNAs and mature tRNA isotypes and plotted these side-by-side with the snoRNA data for comparison (**Figure 39**).

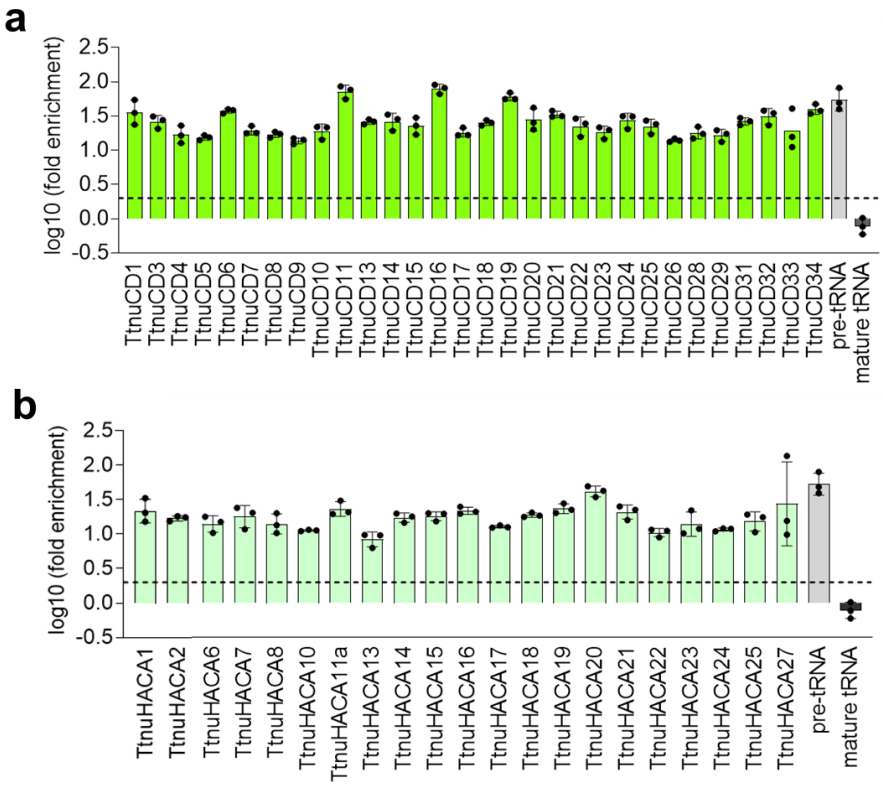


Figure 39. Mlp1 strongly enriches all previously annotated box C/D and box H/ACA snoRNAs.

a,b RNA-sequencing dataset representation of fold enrichment from Mlp1-immunoprecipitated samples for previously annotated *T. thermophila* box C/D (**a**) and box H/ACA (**b**) snoRNAs (n=3 biological replicates). Fold enrichment was calculated by taking the ratio of Mlp1-immunoprecipitated normalized reads and input RNA normalized reads, followed by log10 transformation. Pre-tRNA and mature tRNA fold-enrichment data was obtained from Kerkhofs et al. 2022, averaged and plotted as a positive control and negative control for Mlp1 binding, respectively. The dotted line represents a fold-enrichment of > 2 or log10(2) = 0.3.

Having established that Mlp1 associates with every previously annotated box C/D and box H/ACA snoRNA, we hypothesized that we could also use this dataset as a metric for prediction of novel snoRNAs in the *T. thermophila* genome. First, we used snoRNA prediction tools snoscan and snoreport2.0 (de Araujo Oliveira et al., 2016; Lowe & Eddy, 1999) which can reliably predict novel box C/D snoRNAs. We identified 74 high confidence predicted box C/D snoRNAs in the *T. thermophila* genome. We obtained chromosomal locations and reanalysed the small RNA-sequencing dataset from input and Mlp1-associated RNA to obtain read counts for these novel snoRNAs. We filtered the dataset to retain highly expressed snoRNA in input samples (average TPM input RNA > 10) and obtained 18 novel high-confidence and highly expressed box C/D snoRNAs (**Table 6**). We calculated fold enrichment of Mlp1-associated snoRNAs relative to input snoRNAs and found that Mlp1 strongly enriches all 18 novel predicted box C/D (**Supplementary Figure 12**). As a comparison, averaged fold enrichments for all pre-tRNAs and mature tRNA isotypes are plotted within the same graph (**Supplementary Figure 12**).

Table 6. Novel predicted high-confidence and highly expressed box C/D snoRNAs.

Box C/D snoRNAs predicted using snoscan and snoreport2.0 filtered for high confidence (score > 14) and high expression (TPM input > 10). Box C, box D and box D′ sequences are highlighted in purple and rRNA complementary regions are highlighted in orange.

T. thermophila snoRNA				*T. thermophila* rRNA	
Name	**Chromosomal location**	**Strand**	**Sequence**	**rRNA**	**Target position**
snoscan_499	chr_022:659743-659826	-	TTGGG**ATGATGA**TCATAAATTCTCGACGACGGTCTTCATAATATACATGTTGATATAGAGGATTC**CTGA**CAC	26S	Gm8107
snoscan_3686	chr_169:486432-486529	+	TTTGG**ATGATG**GTTTAAACTTTCGTTCTACTGAAGCATAACATTTCGGTTGTGCTGATTGAGGATATCTTTGTTATAAA**CTGA**CAA	17S	Am3486
snoscan_2790	chr_123:346633-346782	+	TCAAC**TGATGA**TAAATATTCCAGTTCTGCTGCTGAAAACAGCTTGAAGATAACTTTGCTTACCTTT**CTGA**CCT	26S	Am5909
snoscan_3331	chr_146:4460-4648	-	TCAAC**TGATG**CATAACTATGAATATGATTTCTGATTAAATTGTTCATAAATACCATCTTTCGG**GA**CTGACAA	26S	Gm5576
snoscan_1369	chr_065:380204-380379	+	TCTTA**ATGATGA**TATTTAGAGTCTCACGACGTCTGATAAAATGTGACGCAACTACACTCTC**CTGA**TCA	26S	Cm8109
snoscan_981	chr_049:119088-119155	-	TTAT**ATGATGA**TAACTTACAGACCTGTTCTTAAATATGTGAGGACAAAATCCATAGCCAACG**CTGA**C	17S	Gm3933
snoscan_107	chr_003:64065-64131	+	CAAC**TGATGA**CAAATTTATCCCGTATGCTAAAAGCTATGATGGAATTAGATTACAATTCT**CTGA**AA	26S	Um4894
snoscan_1736	chr_073:1071908-1071982	-	TCAAC**TGATGA**CAAAACGATGGGCGGACCAAATCCCAAGTATGTGGACCAATAAATCGGTGAACAATCC**CTGA**C	26S	Um8067
snoscan_2791	chr_123:347846-347987	+	TATAGT**TGATGA**CAATCATATTTGCTACTCTTGACTCGAAAGGGGTTGATGATTATACCACCAAGATCT**CTGA**AA	26S	Cm6210
snoscan_1372	chr_065:380965-381067	+	TTTT**ATGATGA**ATATTTGTCTCGTTCGTCTGAATTCTATTGAAGAAATCACATAACGGAAGC**CTGA**TTA	17S	Am3832
snoscan_2027	chr_096:272835-272902	+	TCAA**ATGATG**CATTTTATTTTTGTATGATTACTTATGATTCTAAATAATCACCATCTTTCGA**CTGA**TAA	26S	Am5578
snoscan_3027	chr_131:639405-639478	+	AAGAT**TGATGA**TACATTTGGATCAACCACCTGATTTCTGATGTTCTACTGGCGA**CTGA**AT	17S	Um2542
snoscan_3545	chr_160:277995-278058	-	TTA**ATGATGA**TAAAAGCCAAAAACGTGACCATATTCTGGACGTATACAATGATAGGC**CTGA**TT	26S	Am8032
snoscan_1371	chr_065:380654-380734	+	AGTA**ATGATGA**AAATTTTCTGACACCGTTGATATTTTATGAAAATAAACTCTACTAGCATA**CTGA**TT	26S	Gm7941
snoscan_2789	chr_123:346102-346310	+	TATTC**ATGATGA**TATAAATATGGAATTACCGTCTGAATTTAGATTGATGGAAAAATTTTGCTAAA**CTGA**TA	17S	Um3106
snoscan_2122	chr_099:1393874-1393969	-	GTTT**AGGATGA**CAAATTTCCCGTGTTTCTGAAACACATTATTTAAATTCATGATTAAAACCTTTGTTGAGTTTTCG**CTGA**CT	pre-rRNA	Am147
snoscan_3386	chr_151:36597-36674	+	TCATT**TGATGA**ACAAAATACTTATCTACATTGTATGAGTAAATGATTAAAATTTCATTA**CTGA**AAA	26S	Um6658
snoscan_500	chr_022:660255-660320	-	TTAA**ATGATGA**TTTATTTTTTCATGACAAGTTTTACCATGGATGTTTTGTCATAGTTAC**ACTGA**TTA	26S	Am6997
			box C RTGATGA **box D'** CTGA **box D** CTGA		

4.3.3. Mlp1 depletion has no significant effect on snoRNA abundance

Since Mlp1 interacts with mature snoRNAs, we investigated if Mlp1 depletion affects their transcript abundance. To determine snoRNA expression levels, we obtained RNA-sequencing counts for wild type and a partial Mlp1 knockout strain from the previously published small RNA-sequencing dataset (Kerkhofs et al. 2022). Relative expression was calculated by taking the ratio between normalized TPM for partial Mlp1 knockout and wild type RNA (**Figure 40a,b**). The relative expression for both box C/D and box H/ACA snoRNAs is around 1, indicating that Mlp1 depletion has no strong effect on snoRNA abundance. We validated these results by qRT-PCR (**Figure 40c**) and Northern blot (**Figure 40d**). As a control, we determined the mRNA levels for Mlp1 by qRT-PCR and confirm that Mlp1 mRNA levels in the Mlp1 partial knockout strain are very low (**Figure 40c**).

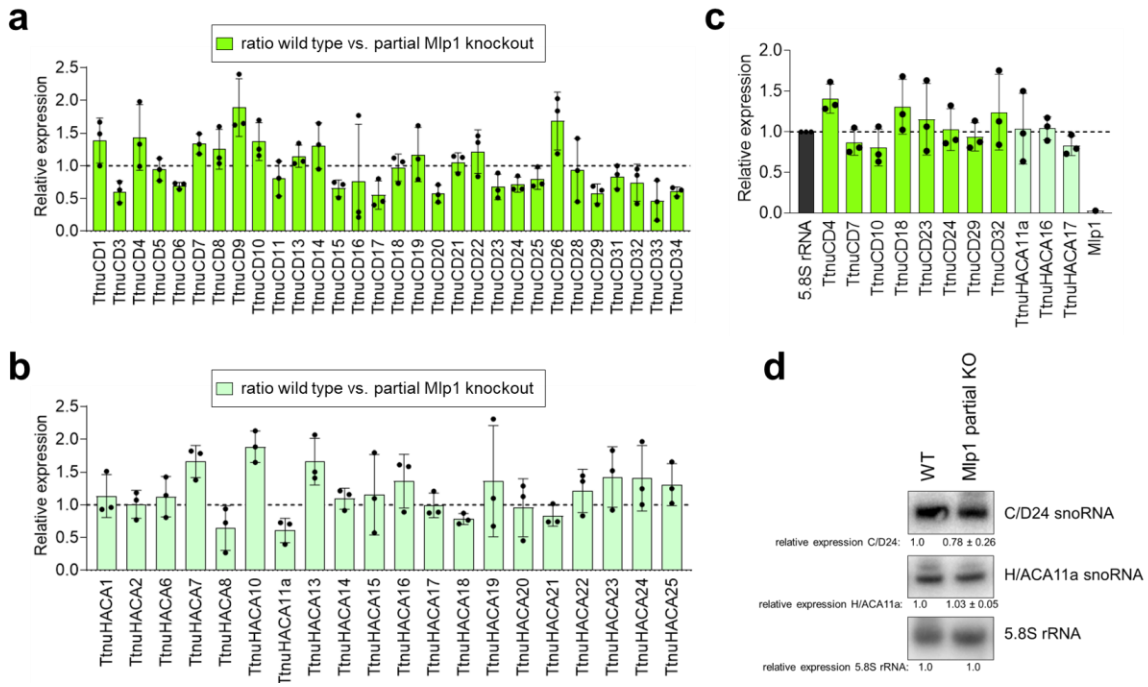


Figure legend on next page

Figure 40. Mlp1 depletion has no substantial effect on snoRNA expression.

a,b RNA-sequencing dataset representation of fold enrichment from wild type (WT) and partial Mlp1 knockout (KO) strains from *T. thermophila* of box C/D (**a**) and box H/ACA (**b**) snoRNAs (n=3 biological replicates). Fold enrichment was calculated by taking the ratio of KO and WT RNA normalized reads in TPM. **c** qRT-PCR analysis of wild type (WT) and partial Mlp1 knockout (KO) strains. RNA polymerase I transcript (5.8S rRNA) is highlighted in dark grey and snoRNAs in green. Fold enrichment was calculated using the ΔCt method ($Ct_{5.8S\ rRNA} - Ct_{RNA}$), followed by fold enrichment calculation by taking partial Mlp1 KO / WT ratios (n=3 biological replicates). **d** Northern blot detecting box C/D24 and H/ACA snoRNA in WT and partial Mlp1 KO strains. Loading control: 5.8S rRNA. Quantification of Northern blot samples (n=3 biological replicates) is shown below. Band intensity was determined using ImageJ relative to the 5.8S rRNA loading control and relative expression was calculated by taking the ratio between the Mlp1 partial knockout and wild type strains.

4.3.4. Mlp1 interacts with the H/ACA RNP complex proteins

RNA polymerase II transcribes snoRNAs as precursors which are required to undergo maturation at the 5′- and 3′-ends (Falaleeva & Stamm, 2013). snoRNAs assembly with four protein components into the functionally active snoRNP complexes. The box H/ACA snoRNPs consists of Nop10, Cbf5 (or Dyskerin in vertebrates), Gar1 and Nhp2 (**Figure 41a**) and the box C/D snoRNAs of fibrillarin, Nop56, Nop58 and 15.5 kDa protein (**Figure 41c**) (Cahill et al., 2002). Since our data indicate that Mlp1 associates with mature snoRNAs rather than precursor snoRNAs, we hypothesized that Mlp1 might function in the assembly of the mature snoRNA into the snoRNP complex. To test Mlp1 association with the snoRNP complex, we performed an endogenous Mlp1-immunoprecipitation and identified potential protein interactors with Liquid Chromatography with Tandem Mass Spectrometry (LC-MS/MS) (**Figure 41b,d**). The samples were treated with benzonase to enable identification of direct protein-protein interactors. We performed statistical analysis of identified peptides and found significant enrichment of a few Mlp1-interacting proteins compared to the immunoglobin G (IgG) control sample, including Nop10 **Figure 41b** – highlighted in yellow). The other H/ACA snRNP proteins components are enriched by Mlp1-immunoprecipitation but did not reach the statistical significance threshold.

The box C/D snoRNPs also contain four core proteins, however, not every component was identified in *T. thermophila* (Snu13 was not annotated) and Nop1/Fibrillarin was not significantly enriched, while Nop56 and Nop58 were enriched, but not significantly **Figure 41d**).

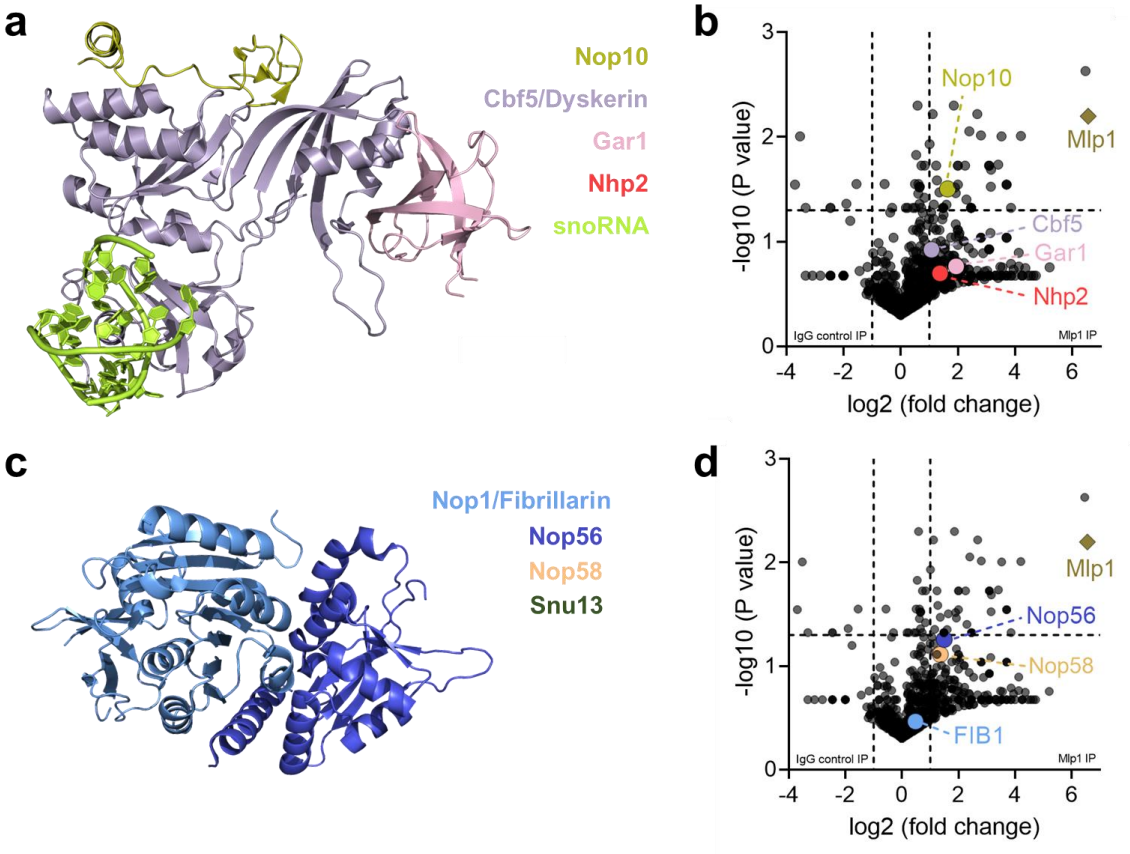


Figure legend on next page

Figure 41. Mlp1 directly interacts with H/ACA snoRNP Nop10 protein.

a,c High-resolution structure of the archaeal Cbf5-Nop10-Gar1 complex (Nhp2 is lacking from the structure) bound to a 17-mer box H/ACA snoRNA sequence (PDB: 3MQK) (**a**) and the *S. cerevisiae* Nop1/fibrillarin-Nop56 complex (Nop58 and Snu13 are missing from the structure) (PDB: 6ZDT) (**c**). **b,d** Mass spectrometry data represented as a volcano plot to identify significantly enriched Mlp1-associated proteins compared to an IgG negative control immunoprecipitation (n=2 biological replicates). Dotted lines indicate significantly enriched proteins (-log10 < 0.05 or) on the y-axis and significantly enriched proteins (log 2-fold change > 2) on the x-axis. Box H/ACA (**b**) and C/D (**d**) snRNP protein components are highlighted in the same color scheme as **a** and **c**, respectively. Mlp1 (bait) is shown as a diamond shape in dark yellow.

4.3.5. snoRNP assembly is minimally affected upon Mlp1 depletion

The observations that Mlp1 directly interacts with both snoRNAs and the Nop10 protein component of the snoRNP complex led us to investigate a potential role for Mlp1 in snoRNP assembly. To test snoRNP complex assembly we fractionated total lysates from wild type and partial Mlp1 knockout strains on 10-30% glycerol gradients. We determined the distribution of total RNA by ethidium bromide (EtBr) staining (**Figure 42a**). Northern blot analysis revealed that upon Mlp1 depletion, snoRNAs are only slightly shifted towards the lighter fractions (**Figure 42b**). As a loading control, we used the distribution of mature tRNAs since Mlp1 is a nuclear protein and depletion of Mlp1 has no effect on mature tRNA levels (Kerkhofs *et. al*. 2022).

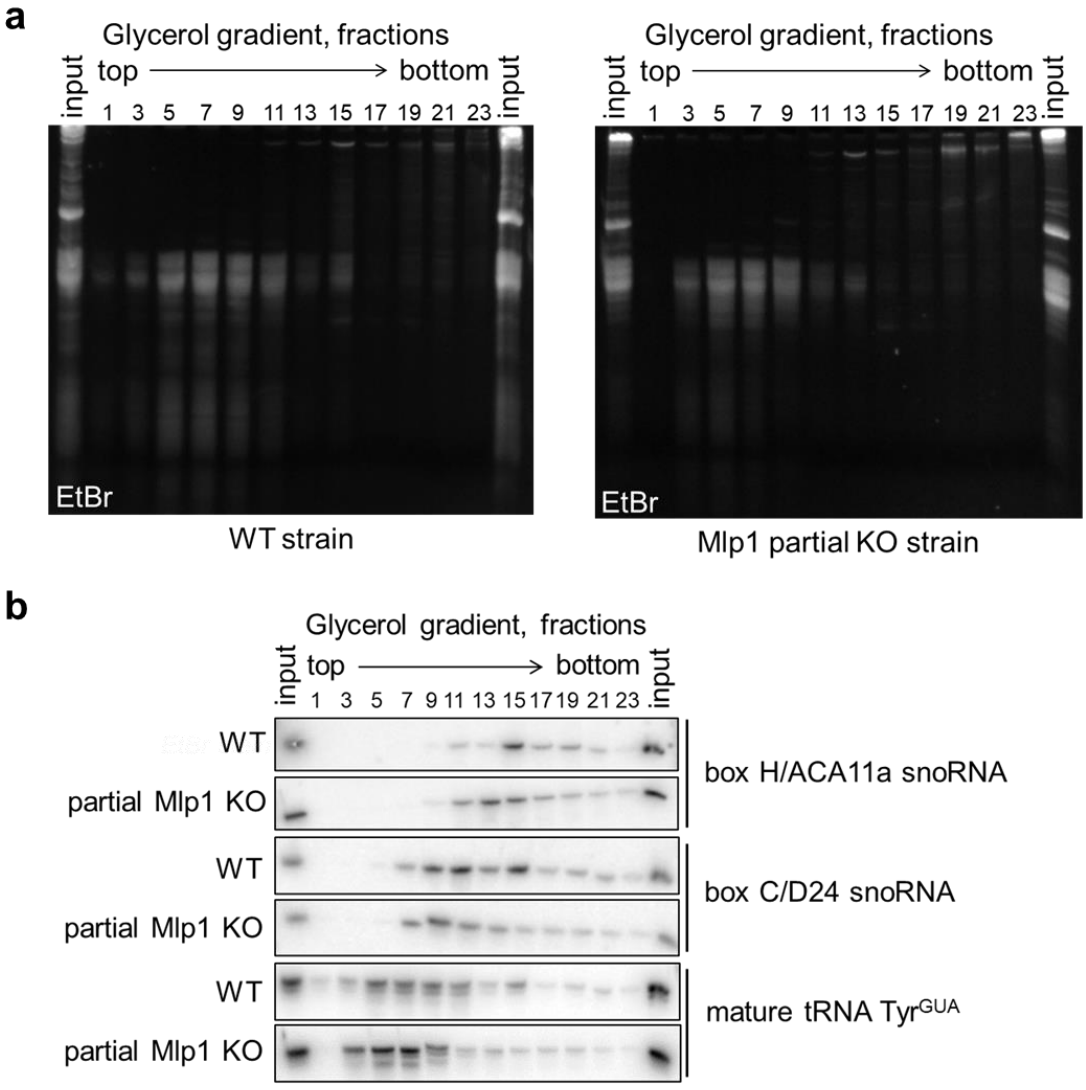


Figure 42. Mlp1 depletion affects snoRNP assembly slightly.

a Ethidium Bromide (EtBr) stain to detect total RNA distributions on the Northern blot in **b**. **b** Northern blot analysis of 10-30% glycerol gradient. Fractions were collected and RNA was extracted using phenol:chloroform:isoamyl alcohol extraction and ethanol precipitation (n=3 biological replicates).

4.3.6. Testing 2′-O-methylation sites of novel snoRNAs in *Tetrahymena thermophila* rRNA

The majority of snoRNAs functions in adding post-transcriptional modification on rRNA. We therefore obtained rRNA modification site predictions using snoscan and snoreport2.0 (de Araujo Oliveira et al., 2016; Lowe & Eddy, 1999) for our novel snoRNAs (see **Table 6**). We investigated the functionality of a few novel predicted box C/D snoRNAs by investigating the predicted post-transcriptional modifications on rRNA. For this, we used an RNase H-dependent 2′-O-methylation detection assay (Yu et al., 1997). This method utilized chimeric oligonucleotides that consist of complementary sequences against the rRNA 2′-O-methylation site with four central deoxyribonucleotides flanked by ribonucleotides containing 2′-O-methylation modifications (**Figure 43a**). Following binding of the chimeric oligonucleotide to the rRNA, RNase H is added which specifically digests the RNA sequence in RNA/DNA hybrids, except when the RNA contains a 2′-O-methylation (Yu et al., 1997). Therefore, if the rRNA contains a 2′-O-methylation, RNase H is not going to produce two separate cleavage products (**Figure 43a**).

To visualize undigested *T. thermophila* rRNA, we included a control where no chimeric oligonucleotide was added (**Figure 43b** – wild type strain, lane 1). The 26S rRNA in *T. thermophila* is incorporated into the ribosomes as two individual parts, $26S_{\alpha}$ and $26S_{\beta}$, which are approximately the same size as the 17S rRNA (Kister et al., 1983). We included two positive controls to visualize RNase H cleavage products by adding a deoxyoligonucleotide against the 17S and 26S rRNA sequence (**Figure 43b** – wild type strain, lanes 2 and 4, respectively). Next, we tested if the predicted box C/D snoRNAs, snoscan_3686 (17S rRNA target), snoscan_2791 (26S rRNA target) and snoscan_3331 (26S rRNA target), were capable of adding a 2′-O-methylation on rRNA (see **Table 6**). We designed chimeric oligonucleotides as shown in **Figure 43a** and found that the rRNA was RNase H digested completely for snoscan_3686 and

snoscan_2791, but only slightly digested for snoscan_3331 (**Figure 43b** – wild type strain, lanes 3, 5 and 6, respectively). This indicates that the first two snoRNAs likely do not add a 2′-O-methylation on the predicted target, or alternatively that these chimeras need to be further optimized, while the snoscan_3331 modifies most rRNA molecules at the predicted site.

Since Mlp1 associates with mature snoRNAs and protein components of the snRNP complexes, we next determined if Mlp1 depletion affects post-transcriptional modification levels on rRNA. We performed the same assay in an Mlp1 partial knockout strain but did not observe any changes for the only functional snoscan_3331 snoRNA compared to the wild type strain (**Figure 43b** – compare lanes 6 between wild type and partial Mlp1 knockout strain), indicating that Mlp1 depletion has no effect on the functionality of this particular box C/D snoRNA.

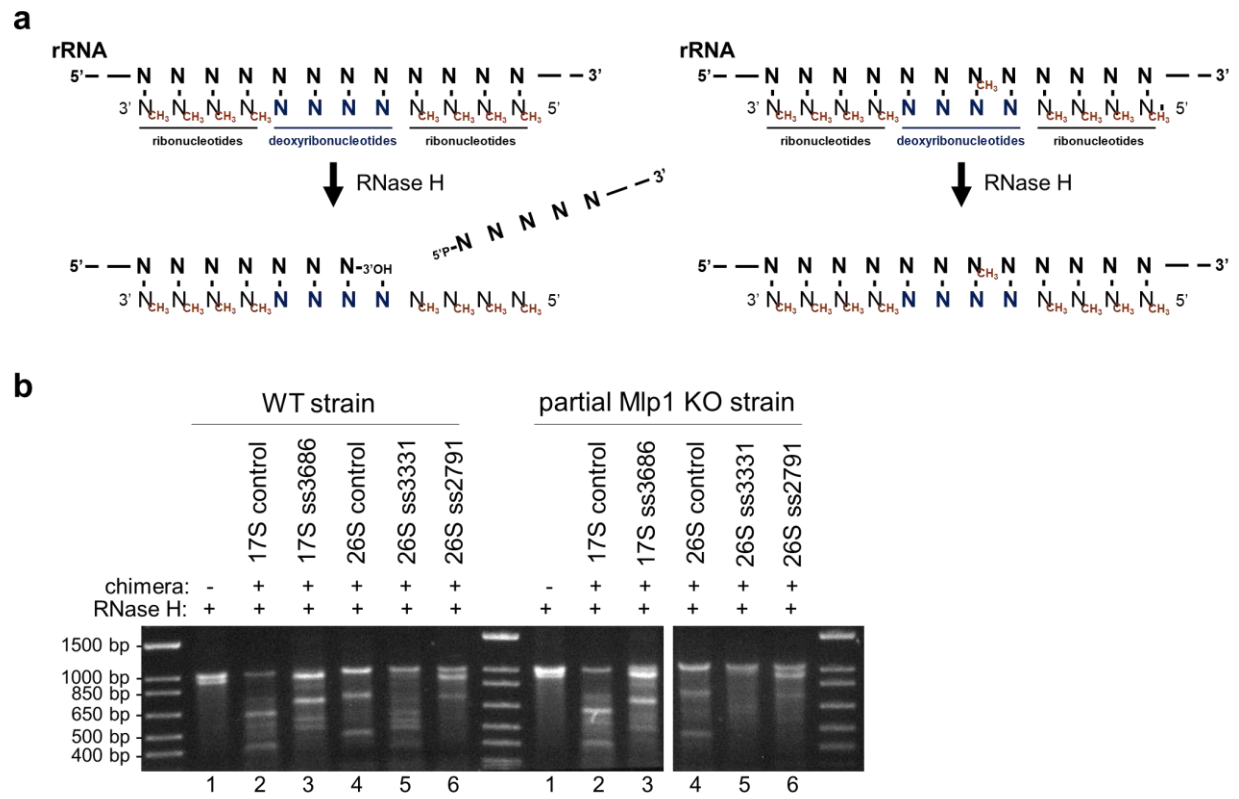


Figure 43. RNase H assay to determine 2′-O-methylation sites on rRNA in *T. thermophila*.

a Schematic representation of chimeric oligonucleotides binding to the corresponding complementary rRNA sequence. The chimeric oligonucleotide consists of four deoxynucleotides complementary to the 2′-O-methylated nucleotide, flanked by ribonucleotides containing a 2′-O-methylation modification. RNase H digest of the RNA strand in the RNA/DNA duplex only occurs when no 2′-O-methylation is found on the rRNA. Figure adapted from (Yu et al., 1997) **b** Total RNA (500 ng) from wild type (WT) and partial Mlp1 knockout (KO) strain incubated with 100-fold excess chimeric oligonucleotides, followed by RNase H digest (n=1 biological replicate). The RNA was separated on a 2% agarose gel using a double stranded DNA marker.

4.4. Discussion

La proteins are well known to associate with nascent RNA polymerase III transcripts such as pre-tRNAs and pre-5S rRNA through binding of the typical 3′-terminal uridylates produced by transcription termination (Mathews & Francoeur, 1984; Stefano, 1984). The second class of non-coding RNA La protein binding targets include RNA polymerase II spliceosomal snRNA transcripts that lack the presence of a 3′-terminal uridylate tail, including U1, U2, U4 and U5. In the yeast *S. cerevisiae* however, these transcripts have processing intermediates where the 3′-ends have been processed to end in 3′-UUU sequences which is the high affinity binding sequence for La proteins (Xue et al., 2000). Similarly, the precursor form of U3 snoRNA in *S. cerevisiae* contains a short tract of 3′-terminal uridylates during processing that form the binding site for La proteins (Kufel et al., 2000). Previous work described a role for the *S. cerevisiae* yeast La protein Lhp1p in stabilization of a 3′-extended precursor intermediate of box C/D U3 snoRNA. Lhp1p-immunoprecipitation revealed that this U3 precursor snoRNA lacks the 2,2,7-trimethylguanosine (m_3G) cap and is not yet assembled into a snoRNP, whereas the mature U3 snoRNA contains a trimethyl guanosine cap and associates with snoRNP proteins Nop1,Nop56 and Nop58 (Kufel et al., 2000). The U3 snoRNA is classified as a box C/D snoRNA and is involved in early endonucleolytic processing of 18S rRNA in both yeast and vertebrates (Hughes & Ares, 1991; Kass et al., 1990)

The data presented in this study suggest that Mlp1 is involved in the function of RNA mature snoRNAs, rather than stabilization of snoRNA precursors. To test Mlp1 association with mature snoRNAs, we reanalyzed our previously published small RNA-sequencing dataset (Kerkhofs et al. 2022) where we performed Mlp1-immunoprecipitation followed by next generation sequencing. Mature snoRNAs are found as a snoRNP complex containing multiple proteins. The box H/ACA snoRNPs consists of Nop10, Cbf5 (or Dyskerin in vertebrates), Gar1 and Nhp2

and the box C/D snoRNAs of fibrillarin, Nop56, Nop58 and 15.5 kDa protein (Cahill et al., 2002). In addition, Mlp1 also interacts with the Nop10 protein components of H/ACA snoRNP complex. We also identified enrichment of the remaining box H/ACA snoRNP protein components Cbf5, Gar1 and Nhp2, and the box C/D snoRNP protein components Nop56 and Nop58, however, these interactions were not statistically significant. Mass spectrometry analysis has some limitations to keep in mind. For example, small proteins would generate few detectable tryptic peptides, and together with a relatively low abundance could lead to these proteins not being identified within the complex peptide mixture generated for mass spectrometry detection (Rappsilber et al., 2002). In addition, not all snoRNP proteins have been annotated in the *T. thermophila* proteome.

Overall, Mlp1 association with mature snoRNAs and direct protein-protein interaction with components of the snRNP complex indicates that Mlp1 might function in snoRNP assembly or activity. We therefore tested assembly of the mature snoRNA into a functional snoRNP complex containing the four snoRNP complex specific proteins between wild type and partial Mlp1 knockout strains. We performed glycerol gradients to separate snoRNP complexes based on density, however, upon Mlp1 depletion, we did not observe a strong shift of the snoRNA towards lighter fractions indicating that Mlp1 has a small or no effect on snoRNP assembly. We also tested if Mlp1 depletion affects 2′-O-methylation on rRNA by comparing post-transcriptional modification levels using an RNA-DNA chimeric oligonucleotide-directed RNase H assay. Our preliminary results indicate that two novel predicted snoRNAs do not add a 2′-O-methylation on the predicted target, or alternatively that these chimeras need to be further optimized. The third novel snoRNA test, provided a post-transcriptional modification on most of the rRNA molecules at the predicted site, however, we did not observe any difference in post-transcriptional modification levels upon comparison of partially Mlp1 depleted with wild type RNA.

While we did not specifically investigate the U3 snoRNA, we confirm that similar to other box C/D and box H/ACA snoRNAs, U3 snoRNA is strongly enriched in Mlp1-associated RNA samples compared to input RNA (data not shown). To further investigate the function of Mlp1 in U3 snoRNP functionally, we could compare 17S pre-rRNA processing intermediates between wild type and partial Mlp1 knockout strains. The U3 snoRNP consists of the core box C/D snoRNP proteins, fibrillarin, Nop56, Nop58 and 15.5 kDa protein, and an additional U3-55k protein which specifically interact with U3 and not other box C/D snoRNAs (Lukowiak et al., 2000). From our current data investigating box C/D24 and box H/ACA11 snoRNAs, we cannot exclude that Mlp1 depletion has no effect on U3 snoRNP assembly and as a result rRNA processing.

In this study, we predicted a total of 18 high-confidence novel box C/D snoRNAs and the rRNA positions for 2′-O-methylation. We confirmed expression of these snoRNAs using the small RNA-sequencing dataset. The high confidence and highly expressed snoRNAs are annotated in **Table 6**. However, certain snoRNAs that are not expressed under normal wild type conditions could be expressed during starvation, mating or during cellular stress conditions to specifically regulate condition-specific post-transcriptional modifications on the rRNA target tailoring an appropriate response in protein translation. The complete novel predicted box C/D snoRNA list can be found in **Supplementary Table 5**.

Future work includes further expansion of both box C/D and box H/ACA snoRNA identification by using bioinformatics tools to construct a *de novo* transcriptome of the Mlp1-associated RNA dataset. This method is frequently used to quantify transcript abundancies from RNA-sequencing data in species lacking a sequenced genome (Li & Dewey, 2011). Since Mlp1 associates with every previously annotated box C/D and box H/ACA snoRNA, we can use this as a metric to identify all expressed snoRNAs in *T. thermophila*. The advantage of using the

Mlp1-associated RNA dataset avoids the problem of relatively low abundancies of some snoRNAs which would avoid detection in the input RNA dataset. Another advantage of this method would be identification of any previously unidentified Mlp1-targets.

The algorithm used in snoRNA prediction tools is based on a combination of secondary structure prediction and machine learning to identify primary sequence conservation of box C, box D, box H and box ACA sequences (de Araujo Oliveira et al., 2016; Hertel et al., 2008). The primary difference between snoreport, which was used for predictions of the original snoRNA (Andersen & Nielsen, 2012), and snoreport2.0, was inclusion of the sequence of potential target sites within ribosomal or spliceosomal RNAs (de Araujo Oliveira et al., 2016; Hertel et al., 2008). This enhances the specificity of novel snoRNA prediction because snoRNAs function as guide RNAs and thus contain a complementary region against the potential target sites (Kiss, 2002). The disadvantage of including of target sequences within snoRNA prediction tools is reduced identification of orphan snoRNAs which lack a complementary sequence and are thus not involved in post-transcriptional modifications of rRNA and other non-coding RNAs. While the function of orphan snoRNAs largely remains unknown, some functional studies have been performed and have identified important functions. For example, orphan snoRNA HBII-52/SNORD115 lacks any complementary against non-coding RNA targets, however, was found to have complementarity to an alternatively spliced exon of the serotonin receptor. A function in regulation of alternative splicing was established for this snoRNA (Kishore & Stamm, 2006). In conclusion, identification of novel targets through a *de novo* transcriptome assembly of the Mlp1-associated RNA dataset ensures the inclusion of these orphan snoRNAs.

Transcription of snoRNAs is different between eukaryotic species. Different modes of transcription include transcription from independent genes, polycistronic transcription and processing from mRNA introns (Kufel & Grzechnik, 2019). Previous work identified 38 box C/D

and 28 box H/ACA snoRNAs in *T. thermophila* and described that the majority of snoRNAs (62%) are found within intergenic regions indicating that these snoRNAs are transcribed from independent genes. The remaining 38% are divided between intron encoded (9%), exon encoded (13%), intron/exon border encoded (3%) or exon/5′-UTR/3′UTR encoded (13%) (Andersen & Nielsen, 2012). Approximately 25% of all *T. thermophila* snoRNAs are encoded as polycistronic clusters. A large snoRNA cluster was identified on chromosome 123 expressing box C/D7, C/D8, C/D9, C/D10, C/D11, C/D12, C/D13, C/D14, C/D15 and C/D16 snoRNAs with spacing in between adjacent snoRNA genes ranging between 187 and 421 base pairs (**Figure 44**) (Andersen & Nielsen, 2012). Based on the chromosomal location of the novel predicted box C/D snoRNAs, we identified a few new snoRNA clusters on chromosomes 22, 65 and 181. A consensus upstream sequence element (USE), 5′-AAACCCATAA-3′, located between 54 and 179 base pairs upstream of intergenic snoRNAs was identified previously. This consensus sequence is also found in the upstream genomic region of snRNAs in *T. thermophila* (Andersen & Nielsen, 2012). In contrast, more than 50% of the clustered snoRNAs had a divergent consensus sequence, indicating that some clustered snoRNAs are co-transcribed as polycistronic transcripts (Andersen & Nielsen, 2012). From the 75 newly predicted snoRNAs listed in **Supplementary Table 5**, 7 contained the exact USE sequence and 12 slightly diverging sequences such as 5′-NNACCCATAA and 5′-AAACCCATNN, indicating that the snoRNAs have conserved expression elements.

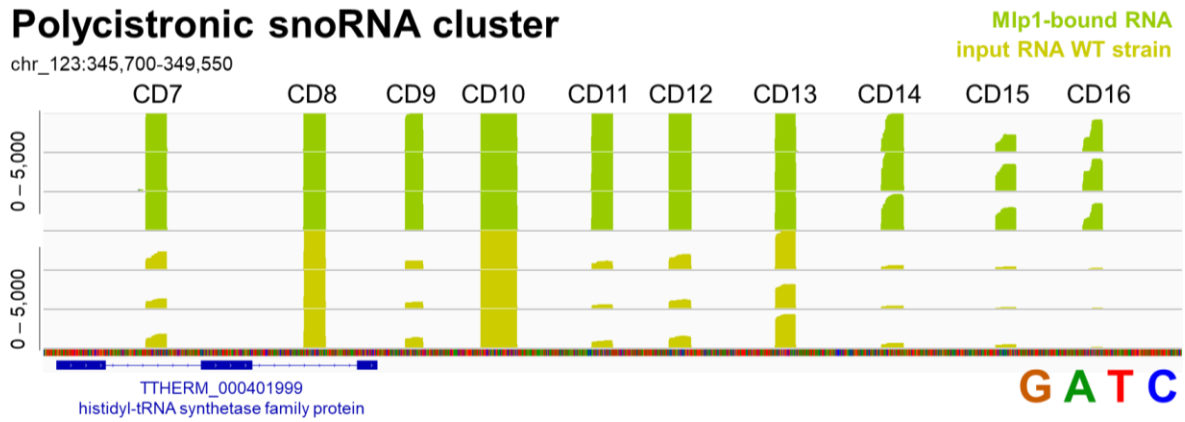


Figure 44. Polycistronic snoRNA clusters.

Genomic view of snoRNA cluster previously identified (Andersen & Nielsen, 2012). Expression levels are shown as raw counts on the y-axis. Image was generated in Integrative Genomics Viewer (IGV).

We noted that, in addition to snoRNAs, Mlp1 appears to enrich mRNA as well (see **Figure 36b**). However, it is likely that at least a part of these mRNAs are predicted protein coding genes based on the higher GC-content relative to the AT-rich surrounding genomic DNA in *T. thermophila* (Eisen et al., 2006). The overall genomic AT% of *T. thermophila* is approximately 75% (Nielsen et al., 1992). Therefore, these regions could encode unannotated ncRNAs instead of protein coding genes. For example, the T3-4 snRNA is encoded on the reverse strand within a predicted protein coding gene (TTHERM_00048812) (**Figure 45**) and without appropriate annotation of the chromosomal location of this non-coding RNA within the protein coding gene, the reads mapping to this genomic location will be counted towards mRNA genes instead of non-coding RNAs. The 127 nucleotide T3-4 snRNA has previously been described as a novel class of nucleolar RNA in *T. thermophila* (Nielsen et al., 1992) and more recently identified as box H/ACA24 snoRNA containing complementary guide sequences against 26S rRNA and pre-rRNA (Andersen & Nielsen, 2012).

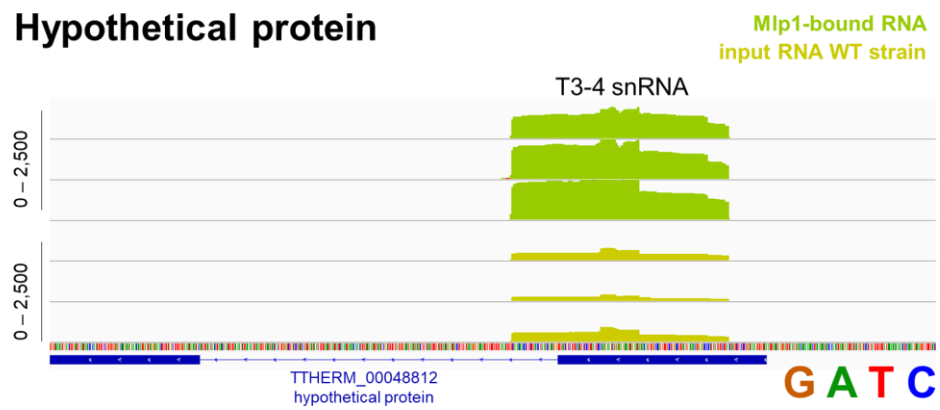


Figure 45. Unannotated non-coding RNAs are located within protein coding genes.

Genomic view of overlapping non-coding RNA and predicted protein coding genes. Image was generated in Integrative Genomics Viewer (IGV).

4.5. Methods

RNP-immunoprecipitation

Cell pellets from 100 mL *T. thermophila* cultures were collected at log phase (0.1 - 1 x 10^6 cells/mL) and washed twice with 1X PBS. Cross-linking with 1% formaldehyde for 10 minutes was followed by quenching with 0.25 M glycine for 5 minutes at room temperature. Next, the cells were washed twice in 1X PBS before lysis in 2 mL buffer A [30 mM Tris-HCl pH 7.4, 150 mM NaCl, 20 mM KCl, 2 mM MgCl2, 0.1% Triton-X100, 1 mM phenylmethylsulfonyl fluoride (PMSF), 1X Halt Protease Inhibitor Cocktail (PIC) (ThemoFisher Scientific), 100U SUPERase In RNase Inhibitor (ThermoFisher Scientific)] followed by sonication (25% amplification with 15s intervals for 4 minutes). Lysate clarification was completed by centrifugation at 20,000 g for 45 minutes at 4°C. Next, the lysated were pre-cleared using rabbit IgG isotype control-bound Protein G Dynabeads rotating for 1 hour at 4°C (12 µg / 50 µL beads slurry). Immunoprecipitations were performed using an affinity purified rabbit anti-Mlp1 antibody (ThermoFisher Scientific – Custom Antibodies) and a rabbit IgG control coupled to Protein G Dynabeads rotating 1 hour at 4°C (12 µg / 50 µL beads slurry). The Protein G Dynabeads were washed 5 times using buffer A (without PIC, PMSF and RNase inhibitor). Input and eluted RNA was isolated following reverse cross-linking for 45 minutes at 70°C in buffer B [50 mM Tris-HCl pH 7.4, 5 mM EDTA pH 8.0, 10 mM DTT, 1% sodium dodecyl sulfate (SDS)]. RNA was isolated by Trizol extraction.

Northern blotting.

RNA samples were prepared by adding an equal volume of 2X formamide loading dye [80% deionized formamide, 0.06% (w/v) bromophenol blue, 0.06% (w/v) xylene cyanol, 10 mM EDTA, pH 8.0] and incubation at 95°C for 5 minutes, followed by immediate snap cooling on ice. Samples were separated on a 7 M urea denaturing polyacrylamide gel at 4°C at 100V and

transferred onto a charged nylon transfer membrane (PerkinElmer). Next, the membranes were cross-linked using a UV Stratalinker and dried for 15 minutes at 80°C on a Gel Dryer (Bio-Rad).

The membranes were hybridized for 2 h at the corresponding probe melting temperature (Tm) -10°C in hybridization buffer [6X SSC (1X SSC: 150 mM NaCl and 15 mM sodium citrate), 1% SDS and 4X Denhardt′s solution (Bio Basic D0062)], followed by overnight incubation with Polynucleotide Kinase (PNK) ^{32}P-labeled probes. Oligonucleotides used for Northern blotting: TthnuCD24 5′- ATTCAGTCGCCAGTAGAACATCAGA-3′, TtnuHACA11a 5′-AAATGTCCTTATCTTGGTATCTTTT-3′, 5.8S rRNA 5′- CTGCAATTCGCATTGCGT-3′ and mature tRNA Tyr^{GUA} 5′-GACCACCGGATTTACAGTCCG-3′. Following overnight incubation, the membranes were washed three times for 20-minutes in wash buffer [2X SSC and 0.1% SDS], followed by membrane exposure to a storage phosphor screen overnight and developing on a Typhoon imager. To strip hybridized probes, membranes were incubated three times with stripping buffer [0.1X SSC and 0.1% SDS] for 20 minutes each at 70°C.

qRT-PCR.

Total RNA was extracted from *T. thermophila* wild type and partial Mlp1 knockout using Trizol extraction. The purified RNA samples were resuspended in 1X Reaction Buffer with 2 units of Turbo DNase (ThermoFisher Scientific, cat# AM2238) and incubated for 30 minutes at 37°C. The DNase-treated samples were subsequently Trizol extracted to inactivate DNase activity. Next, 50 ng/µL DNase-treated RNA was reverse transcribed using the iScript cDNA Synthesis Kit (Bio-Rad, cat#1708891) according to manufacturer′s instructions. The cDNA was diluted to 25 ng (in 12.5 µL per technical triplicate) and quantified using the SensiFAST SYBR No-Rox kit (Bioline, BIO-98005) with 12.5 µM of each primer (**Table 7**) (1x forward, 1x reverse) using following qPCR settings: 95°C for 5 minutes and 35 cycles of 5 sec at 95°C and 15 sec at 60°C, followed by a melting curve analysis up to 99°C to confirm amplification of a single amplicon.

Table 7. qRT-PCR primers used for snoRNA detection.

Primer name	Sequence	
5S rRNA For	5' GTCGGCCATACTAAGGTGAAAAC	3'
5S rRNA Rev	5' GGCTGTCGACACTGGGACTT	3'
pre-tRNA Tyr^{GUA} For	5' CTGTAAAGTTTGTAGTCATCCGGTG	3'
pre-tRNA Tyr^{GUA} Rev	5' AAAAAAAATCCGAACTCCCGGG	3'
pre-tRNA Leu^{UAA} For	5' GTTCATCCTCCTACATTTAGAAAAGAGTGTCG	3'
pre-tRNA Ile^{UAU} For	5' GCTGACCAACCTTATAAAAGTGTGTCG	3'
pre-tRNA Leu^{UAA}/Ile^{UAU}Rev	5' AAAAAAAAAGTAGCTCAGAGAGGGTTC	3'
17S rRNA For	5' GGCCGTTCTTAGTTGGTGGA	3'
17S rRNA Rev	5' TCGCAGTCCAGAAGGTTCAC	3'
CD4 snoRNA For	5' TGATGATAACATCCAGCTCACTTTG	3'
CD4 snoRNA Rev	5' TTGTCAGAGTATAGGAAAGGTTTTT	3'
CD7 snoRNA For	5' CATGATGATATAAATATGGAATTAC	3'
CD7 snoRNA Rev	5' TATCAGTTTAGCAAAATTTTTCCAT	3'
CD10 snoRNA For	5' TTGATGACAACTCGCGTCCATACTC	3'
CD10 snoRNA Rev	5' TTTTCAGGATTGCAGAACAAAGTCA	3'
CD18 snoRNA For	5' TCGTTGATGAGCATTTGTGTCCGTG	3'
CD18 snoRNA Rev	5' TTTCAGAAATGTGTATTAATCACGTT	3'
CD23 snoRNA For	5' ATGATGATTCAAGATAGGGAC	3'
CD23 snoRNA Rev	5' TTATCAGGATCTACTATCGATTATG	3'
CD24 snoRNA For	5' TTGATGATACATTTGGATCAACCAC	3'
CD24 snoRNA Rev	5' ATTCAGTCGCCAGTAGAACATCAGA	3'
CD29 snoRNA For	5' AATGATGACAAAATTCCCAAAATGC	3'
CD29 snoRNA Rev	5' TATCAGTGGCAAATTGGATTAACCT	3'
CD32 snoRNA For	5' CATGATGTATAAAACGCATGG	3'
CD32 snoRNA Rev	5' AATCAGACGATTGGTATGGATTCAT	3'
HACA11a snoRNA For	5' GGGTAACAGAGTCATTAACTATA	3'
HACA11a snoRNA Rev	5' AAATGTCCTTATCTTGGTATCTTTT	3'
HACA16 snoRNA For	5' GCGAGAACTACTAATAGTTTGTT	3'
HACA16 snoRNA Rev	5' TAATGTCTGAATTCGGAAATTACCA	3'
HACA17 snoRNA For	5' GAACACTATCCTTGCTCATTC	3'
HACA17 snoRNA Rev	5' AAATGTCTGAACATGCAATTACCAC	3'
Mlp1 For	5' AACGTCCATGCCCCTTACTG	3'
Mlp1 Rev	5' ACACCGTTGTAGTGACGACC	3'

Protein-protein immunoprecipitation and mass spectrometry.

The same LC-MS/MS dataset described in Chapter III was used for identification of Mlp1-interacting snoRNP proteins (see p. 139: Method section – Protein-protein immunoprecipitation and mass spectrometry).

Glycerol gradients.

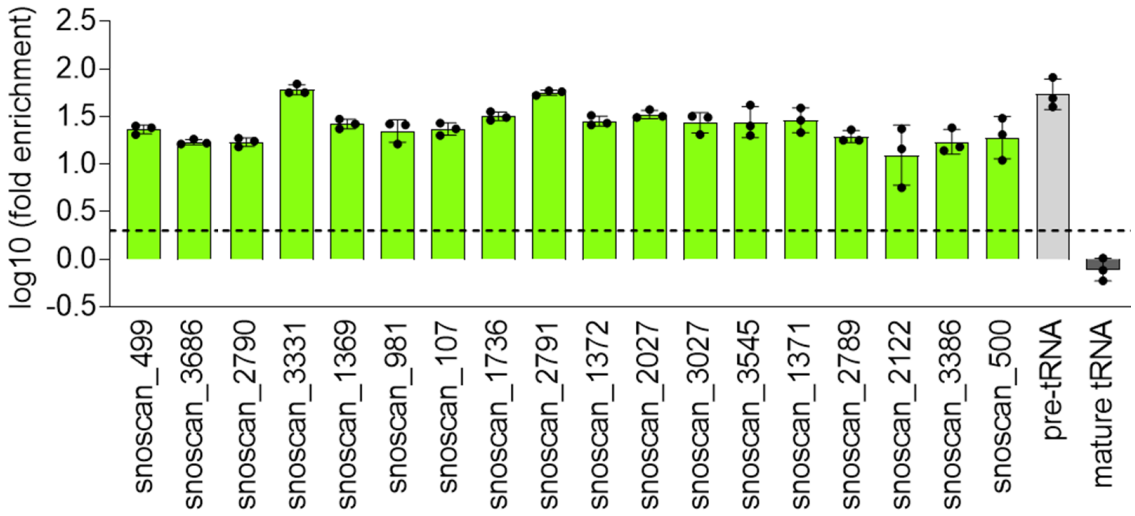
Cell pellets from 50 mL *T. thermophila* cultures were lysed in 500 µL lysis buffer [50 mM Tris-HCl, pH 7.4; 250 mM KCl; 10 mM $MgCl_2$; 0.5 mM DTT; 0.2 mM EDTA, pH 8.0; 10% glycerol; 1 mM PMSF, 1X PIC (ThemoFisher Scientific), 25U SUPERase In RNase Inhibitor (ThermoFisher Scientific)]. Lysates were clarified by centrifugation at 20,000 g for 30 minutes at 4°C and carefully layered on a 5-step 10-30% glycerol gradients in buffer C [20 mM HEPES, pH 7.4; 150 mM KCl; 1 mM MgOAc; 0.1% NP40 and 0.5 mM PMSF]. The glycerol gradient was kept at 4°C overnight to diffuse into a linear gradient prior to loading sample. Lysates were fractionated by centrifugation for 20 hours at 35,000 RPM (Beckman SW41Ti rotor) at 4°C and fractions collected for further RNA and protein analysis (400 uL/fraction). RNA extraction was completed using phenol:chloroform:isoamyl alcohol (25:24:1) extraction (1:1 volume). The aqeous layer was ethanol precipitated in 150 mM NaOAc, pH 5.1 and 30 µg GlycoBlue (ThermoFisher Scientific) for at least 1 hour at -80°C. Protein samples for Western blot analysis were obtained by Trichloroacetic acid (TCA) precipitation at a 10% final concentration, followed by incubation at –20°C overnight to enhance protein precipitation. Protein pellets were collected by centrifugation for 15 minutes at 10,000 g at 4°C, then washed twice with 100% ice-cold acetone. The pellet was air-dried and resuspended in 50µL/mL culture 2.5X loading dye [5X loading dye: 5% β-mercaptoethanol (v/v), 0.02% bromophenol blue (w/v), 30% glycerol (v/v), 10% SDS (w/v), 250 mM Tris-HCl, pH 6.8].

RNA-DNA chimeric oligonucleotide-directed RNase H assay.

Total RNA was extracted from *T. thermophila* wild type and partial Mlp1 knockout using Trizol extraction and a total of 500 ng was added to 100 μM chimeric oligonucleotide, incubated for 5 minutes at 95°C and slow cooled to 37°C to allow annealing of the chimeric oligonucleotide to the rRNA. Next, 1X RNase H Reaction Buffer and 1.25 units of RNase H (NEB, cat# M0297S) were added in a total volume of 10 μL. The RNase H reaction mix was incubated at 37°C for 30 minutes. The reaction was stopped by adding an equal volume of 2X formamide loading dye [80% deionized formamide, 0.06% (w/v) bromophenol blue, 0.06% (w/v) xylene cyanol, 10 mM EDTA, pH 8.0], followed by incubation at 95°C for 5 minutes and snap cooling on ice. RNA samples were separated on a 2% agarose [20 g/L agarose, 1X TBE (100 mM Tris, 100 mM boric acid, 2 mM EDTA), 0.05 μL/mL ethidium bromide] run in 1X TBE electrophoresis buffer.

Oligonucleotides used for the RNA-DNA chimeric oligonucleotide-directed RNase H assay include 17S rRNA RNase H control 5′-TTTCGCATTAGTTAGTCC-3′; 26S rRNA RNase H control 5′- CATTTAAGTAAACTCCGC-3′; snoscan_3686 RNase H mUmCGTTCmUmUmGmAmUmUmAmAmUmGmAmA; snoscan_2791 RNase H mCmAAGATmCmUmCmAmCmUmAmAmUmCmGmC; snoscan_3331 RNase H mUmUTCGGmGmUmCmCmUmAmAmCmAmAmUmU.

4.6. Supplementary Figures and Tables



Supplementary Figure 12. Mlp1 strongly enriches all novel predicted box C/D snoRNAs.

RNA-sequencing dataset representation of fold enrichment from Mlp1-immunoprecipitated samples for novel predicted *T. thermophila* box C/D snoRNAs (n=3 biological replicates). Fold enrichment was calculated by taking the ratio of Mlp1-immunoprecipitated normalized reads and input RNA normalized reads, followed by log10 transformation. Pre-tRNA and mature tRNA fold-enrichment data were obtained from Kerkhofs et al. 2022, averaged and plotted as a positive control and negative control for Mlp1 binding, respectively. The dotted line represents a fold-enrichment of > 2 or log10(2) = 0.3.

Supplementary Table 4. Comparison of relative expression of small non-coding RNAs and mRNA in Mlp1- and hLa-RNP immunoprecipitations.

Data obtained from Gogakos et al. 2017 (hLa dataset) and Kerkhofs et al. 2022 (Mlp1 dataset.

	hLa dataset					
	input rep1	**input rep2**	**input rep3**	**input rep4**	**IP rep2**	**IP rep3**
rRNA	21.4%	68.8%	57.7%	73.2%	14.3%	5.1%
tRNA	66.3%	20.8%	30.9%	17.5%	68.4%	50.0%
snoRNA	9.2%	3.1%	4.1%	3.1%	0.0%	0.0%
snRNA	1.0%	2.1%	2.1%	1.0%	0.0%	0.0%
mRNA	2.0%	4.2%	4.1%	4.1%	17.3%	44.9%

	Mlp1 dataset					
	input rep1	**input rep2**	**input rep3**	**IP rep1**	**IP rep2**	**IP rep3**
rRNA	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
tRNA	96.9%	97.0%	96.6%	80.5%	80.4%	78.7%
box C/D snoRNA	0.5%	0.6%	0.6%	11.0%	11.1%	12.0%
box H/ACA snoRNA	0.1%	0.1%	0.1%	1.2%	1.1%	1.5%
mRNA	2.5%	2.4%	2.6%	7.3%	7.4%	7.8%

Supplementary Table 5. Complete table of novel predicted box C/D snoRNAs.

	T. thermophila snoRNA			*T. thermophila* rRNA	
	Name	**Chromosomal location**	**Strand**	**rRNA**	**Target position**
#1	snoscan_499	chr_022:659743-659826	-	26S	Gm8107
#2	snoscan_3686	chr_169:486432-486529	+	17S	Am3486
#3	snoscan_2790	chr_123:346633-346782	+	26S	Am5909
#4	snoscan_3331	chr_146:4460-4648	-	26S	Gm5576
#5	snoscan_1369	chr_065:380204-380379	+	26S	Cm8109
#6	snoscan_981	chr_049:119088-119155	-	17S	Gm3933
#7	snoscan_107	chr_003:64065-64131	+	26S	Um4894
#8	snoscan_1736	chr_073:1071908-1071982	-	26S	Um8067
#9	snoscan_2791	chr_123:347846-347987	+	26S	Cm6210
#10	snoscan_1372	chr_065:380965-381067	+	17S	Am3832
#11	snoscan_2027	chr_096:272835-272902	+	26S	Am5578
#12	snoscan_3027	chr_131:639405-639478	+	17S	Um2542
#13	snoscan_3545	chr_160:277995-278058	-	26S	Am8032
#14	snoscan_1371	chr_065:380654-380734	+	26S	Gm7941
#15	snoscan_2789	chr_123:346102-346310	+	17S	Um3106
#16	snoscan_3492	chr_156:538144-538259	N.D.	17S	Am3224
#17	snoscan_2122	chr_099:1393874-1393969	-	pre-rRNA	Am147
#18	snoscan_3386	chr_151:36597-36674	+	26S	Um6658
#19	snoscan_2793	chr_123:349220-349323	N.D.	26S	Um7940
#20	snoscan_500	chr_022:660255-660320	-	26S	Am6997
#21	snoscan_1836	chr_082:209334-209417	N.D.	pre-rRNA	Am9556
#22	snoscan_4172	chr_181:6917-7125	N.D.	17S	Um3603
#23	snoscan_4177	chr_181:14136-14344	N.D.	17S	Um3603
#24	snoscan_3264	chr_139:107080-107173	N.D.	26S	Gm5552
#25	snoscan_1320	chr_061:154353-154444	N.D.	26S	Gm5281

Table continued on next page

	T. thermophila snoRNA			*T. thermophila* rRNA	
	Name	**Chromosomal location**	**Strand**	**rRNA**	**Target position**
#26	snoscan_4185	chr_181:18851-19162	N.D.	26S	Gm8221
#27	snoscan_633	chr_026:325555-325678	N.D.	pre-rRNA	Um279
#28	snoscan_3394	chr_151:195415-195535	N.D.	pre-rRNA	Am7200
#29	snoscan_3750	chr_173:161869-161962	N.D.	pre-rRNA	Gm2054
#30	snoscan_4170	chr_181:6106-6295	N.D.	5.8S	Gm4435
#31	snoscan_4169	chr_181:5734-5934	N.D.	26S	Cm4800
#32	snoscan_275	chr_010:72360-72469	N.D.	17S	Gm3834
#33	snoscan_3873	chr_177:335044-335137	N.D.	26S	Um6299
#34	snoscan_4181	chr_181:15652-15861	N.D.	26S	Um5025
#35	snoscan_1220	chr_058:1260902-1260968	N.D.	26S	Am5202
#36	snoscan_3231	chr_138:446790-446871	N.D.	26S	Um8345
#37	snoscan_4179	chr_181:14966-15155	N.D.	17S	Gm4435
#38	snoscan_1972	chr_090:439154-439271	N.D.	26S	Um4783
#39	snoscan_1608	chr_069:19900-19992	N.D.	pre-rRNA	Am535
#40	snoscan_4126	chr_180:867248-867330	N.D.	pre-rRNA	Am245
#41	snoscan_2971	chr_130:244997-245090	N.D.	17S	Gm3497
#42	snoscan_4000	chr_179:1160004-1160112	N.D.	pre-rRNA	Cm4408
#43	snoscan_3145	chr_135:1619052-1619153	N.D.	pre-rRNA	Gm758
#44	snoscan_3353	chr_148:309037-309127	N.D.	26S	Gm8304
#45	snoscan_3716	chr_170:629077-629177	N.D.	pre-rRNA	Um8617
#46	snoscan_3660	chr_168:150934-151097	N.D.	26S	Um5547
#47	snoscan_1993	chr_092:15306-15429	N.D.	pre-rRNA	Um190
#48	snoscan_2669	chr_119:641871-641970	N.D.	17S	Gm2957
#49	snoscan_1840	chr_082:258824-258929	N.D.	pre-rRNA	Am7241
#50	snoscan_1896	chr_084:213295-213402	N.D.	pre-rRNA	Um1049

Table continued on next page

	T. thermophila snoRNA			*T. thermophila* rRNA	
	Name	**Chromosomal location**	**Strand**	**rRNA**	**Target position**
#51	snoscan_2098	chr_099:714764-714842	N.D.	pre-rRNA	Am9999
#52	snoscan_2851	chr_125:1301-1427	N.D.	pre-rRNA	Um1077
#53	snoscan_1211	chr_058:950911-951037	N.D.	pre-rRNA	Am8777
#54	snoscan_308	chr_012:308692-308771	N.D.	pre-rRNA	Um1102
#55	snoscan_1837	chr_082:210680-210770	N.D.	26S	Gm5794
#56	snoscan_885	chr_042:68385-68480	N.D.	17S	Gm4047
#57	snoscan_3310	chr_142:293517-293612	N.D.	pre-rRNA	Gm1995
#58	snoscan_1990	chr_091:170013-170142	N.D.	17S	Um3601
#59	snoscan_1906	chr_087:4424-4546	N.D.	pre-rRNA	Um19
#60	snoscan_1270	chr_058:2625369-2625468	N.D.	pre-rRNA	Am9999
#61	snoscan_923	chr_043:293713-293824	N.D.	26S	Gm5382
#62	snoscan_2055	chr_097:392875-392992	N.D.	26S	Am6195
#63	snoscan_3395	chr_151:197005-197089	N.D.	pre-rRNA	Um10218
#64	snoscan_10	chr_001:215682-215777	N.D.	pre-rRNA	Um973
#65	snoscan_109	chr_003:99303-99387	N.D.	pre-rRNA	Am426
#66	snoscan_2314	chr_105:1466170-1466284	N.D.	pre-rRNA	Am9563
#67	snoscan_1738	chr_074:74263-74374	N.D.	pre-rRNA	Um9910
#68	snoscan_3211	chr_138:117042-117141	N.D.	26S	Am5428
#69	snoscan_2511	chr_112:1457978-1458065	N.D.	pre-rRNA	Am427
#70	snoscan_1984	chr_090:662562-662690	N.D.	pre-rRNA	Um10056
#71	snoscan_2323	chr_105:1741303-1741403	N.D.	pre-rRNA	Am774
#72	snoscan_3612	chr_164:117447-117508	N.D.	26S	Am5098
#73	snoscan_3526	chr_158:428554-428626	N.D.	26S	Am6139
#74	snoscan_3455	chr_154:15378-15465	N.D.	pre-rRNA	Am10249

Chapter V:

Conclusions and future perspectives

5.1. The discovery of a novel type of La protein lacking a highly conserved RRM domain

The La protein was originally discovered as an autoantigen in patients suffering from the autoimmune diseases systemic lupus erythematosus and Sjogren′s syndrome in 1974 (Alspaugh & Tan, 1975; Mattioli & Reichlin, 1974). A few years later, in 1981, immunoprecipitations using patient antibodies led to hLa being classified as an RNA binding protein (Hendrick et al., 1981; Lerner et al., 1981). In 1984, RNA targets were identified as nascent polymerase III transcripts still containing the typical 3′-terminal uridylate sequence which is encoded in the genome as a deoxyadenosines stretch at the end of the RNA polymerase III gene (Mathews & Francoeur, 1984; Stefano, 1984). More than ten years later, functional studies using human La (hLa) mutants revealed that the N-terminal La module is necessary and sufficient for high-affinity binding of transcripts containing a 3′-terminal stretch of uridylates (Goodier et al., 1997; Ohndorf et al., 2001).

High-resolution structures obtained in 2004 of the hLa protein revealed that the La module contains a La motif (LaM) and RNA recognition motif (RRM) fold (Alfano et al., 2004). The LaM adopts an elaborated version of a winged helix fold which is commonly found in DNA and RNA binding proteins (Alfano et al., 2004; Gajiwala & Burley, 2000), and the RRM adopt a classical RRM-fold consisting of a four-stranded antiparallel β-sheet and two α-helices, with an additional C-terminal α3-helix (Alfano et al., 2004). The RRM is the most common RNA binding domain in eukaryotes and is also found in prokaryotes and viruses in lower abundance. Conserved sequences within RRMs include the RNP1 and RNP2 motifs that are located in the central strands (β1 and β3) of the antiparallel β-sheet. Accordingly, the putative single-stranded RNA binding site is located on the β-sheet surface (Maris et al., 2005). A major discovery in the La protein field came from the co-crystal structure of the hLa module in complex with a 3′-uridylate

containing RNA in 2006, where the RNA was found to be bound without using the conserved β-sheet within RRM1. Instead, it is a lateral side of the RRM1 β-sheet together with the winged helix in the LaM that interacts with the UUU-3′OH containing RNA (Kotik-Kogan et al., 2008; Teplova et al., 2006). Binding of uridylate-tailed RNA to these unanticipated surfaces keeps the expected RNA-binding surface available for additional interactions. The typical RNA-binding sequences in the hLa protein RRM1 β-sheet have since been found to make uridylate-independent contacts to the main body of pre-tRNAs and also possess RNA chaperone activity by assisting with tRNA folding and maturation (Bayfield & Maraia, 2009; Huang et al., 2006; Naeeni et al., 2012).

It is now expected in the genuine La protein field that high-affinity binding of 3′-terminal uridylates requires the presence of both RNA-binding domains found in the La module to cooperatively bind through formation of a UUU-3′OH specific binding pocket (Maraia & Bayfield, 2006). The recently proposed discovery of a novel type of La protein in alveolates, including *Tetrahymena thermophila,* therefore challenged this presumption (Deragon, 2020). This evolutionary study included eukaryotic species from all major eukaryotic lineages and grouped the LaM-containing and RRM1-lacking alveolate La-like proteins with the genuine La proteins (Deragon, 2020). We confirmed the absence of the conservation of the LaM and absence of RRM1 domain in the *T. thermophila* La protein Mlp1 through primary and predicted secondary structure analysis (Kerkhofs et al. 2022). More recently, we have gathered preliminary Nuclear Magnetic Resonance (NMR) data in collaboration with the Conte group at King′s College London that we hope will corroborate the three-dimensional conservation of the LaM and absence of a classic RRM1 structure in the region adjacent to the LaM (Kerkhofs, K., Jarvis, J., Conte, M. and Bayfield, M., unpublished data). In addition to solving the structures of the LaM and adjacent domain, we have biochemical and NMR-based RNA binding experiments planned to identify which residues are involved in RNA interactions. This study will further elucidate how

an RRM1-lacking genuine La protein interacts with the typical La protein RNA target, 3′-terminal ending uridylate RNA.

5.1.1. Mlp1 functions as a genuine La protein

We demonstrated that Mlp1 functions as a typical La protein, despite the apparent lack of the highly conserved RRM1, in the following aspects: (i) preferential binding of pre-tRNAs over mature tRNAs both *in vitro* and *in vivo*, (ii) uridylate-specific binding using highly conserved residues in the La motif and (iii) RNA chaperone activity to promote processing and maturation of misfolded nascent pre-tRNAs using a tRNA mediated suppression assay.

Genuine La proteins interact with nascent RNA polymerase III transcripts through high affinity binding of the 3′-terminal uridylate sequences using highly conserved residues located in the LaM (Kotik-Kogan et al., 2008; Teplova et al., 2006). We demonstrated that, similar to the previously established hLa uridylate binding deficient mutant Q20/Y24/D33R (Huang et al., 2006; Vinayak et al., 2018), mutation of conserved LaM-residues Q11A/Y14A in Mlp1 results in loss of binding to uridylate containing RNA targets *in vitro*.

Previous work described that RNA chaperone activity in the hLa protein is located within the conserved RRM1 β-sheet, the additional α3-helix in RRM1, the RRM1 loop-3 and in C-terminal regions (Bayfield et al., 2021; Bayfield & Maraia, 2009; Huang et al., 2006; Naeeni et al., 2012). While these domains are not present in the *T. thermophila* La protein Mlp1, we found that that heterologous expression of Mlp1 in *Schizosaccharomyces pombe* promotes pre-tRNA processing and tRNA mediated suppression in a tRNA-mediated suppression assay. In order to map the structural region of Mlp1 important for RNA chaperone activity, we performed this assay using a range of deletion mutants, including a LaM deletion mutant (Mlp1 95-340) and C-terminal domain of unknown function-3223 (DUF3223) deletion mutant (Mlp1 1-340). We found that upon deletion of either domain, pre-tRNA mediated suppression was impaired, but a

stronger effect was observed in the DUF3223 deletion mutant. This suggests that the C-terminal domain of Mlp1 in *T. thermophila* may substitute for the RNA chaperone activity normally linked to the RRM that is required for appropriate folding of misfolded pre-tRNAs.

5.1.2. Pre-tRNA processing in *Tetrahymena thermophila* is divergent from other eukaryotic systems

Previous work has demonstrated that La proteins bind to nascent pre-tRNAs on their 3′-uridylate tail and thereby stabilize these nascent polymerase III transcripts through protection against degradation by 3′-exonucleases (Copela et al., 2006; Fairley et al., 2005; Yoo & Wolin, 1997). In addition, interactions to the body of the tRNA induces folding into the correct conformation through RNA chaperone activity and provides a platform for post-transcriptional modifications by tRNA modifying enzymes. During maturation of polymerase III transcripts, the 3′-uridylate trailer is removed through endonucleolytic processing by RNase Z and the La protein is released. On the other hand, when the La protein has been deleted or is limiting, the 3′-trailer sequences of the nascent pre-tRNA remain unprotected, resulting in rapid processing by 3′-exonucleases prior to 5′-trailer cleavage by RNase P (Copela et al., 2008). In absence of the La protein, misfolded pre-tRNAs are more prone to degradation through nuclear surveillance (Huang et al., 2006).

We studied nascent pre-tRNA processing in the heterologous *sla1- S. pombe* tRNA mediated suppression system and found that Mlp1 promotes folding and maturation of the misfolded nascent suppressor pre-tRNA, however, with diminished protection of 3′-trailer sequences. This could indicate that protection of the 3′-ends of the nascent pre-tRNAs is not as strong for Mlp1 compared to Sla1p, leaving the RNA exposed to 3′-exonucleases. We also studied pre-tRNA processing in *T. thermophila* where we found that upon depletion of Mlp1, the 3′-trailer sequences of nascent pre-tRNAs are not as efficiently processed compared to a wild type strain containing normal levels of Mlp1. It was expected that depletion of Mlp1 from *T. thermophila*

should increase accessibility of the 3′-trailer to 3′-exonucleases and result in accelerated 3′-end processing. However, we observed the opposite: depletion of Mlp1 results in more unprocessed pre-tRNA 3′-trailer sequences. Our current hypothesis is that another unidentified protein could be functioning together with Mlp1 to promote processing *in vivo* (**Figure 46**). When Mlp1 is depleted, the unidentified protein remains associated with the 3′-uridylate containing end of the pre-tRNA, thereby providing protection against 3′-exonucleases (**Figure 46**). However, without RNA chaperone activity associated with Mlp1, the cleavage site for RNase Z remains inaccessible thereby delaying the 3′-processing speed resulting in accumulation of pre-tRNAs retaining their 3′-ends. For future work, we have performed Liquid Chromatography with Tandem Mass Spectrometry (LC-MS/MS) of endogenous Mlp1 immunoprecipitations to identify potential protein-protein interactions that might confirm this model, and some candidate interactors have been validated through *in vitro* pulldown interaction assays (data not shown). We are also developing a system in which we test for pre-tRNA 3′-end protection by co-expression of Mlp1 and candidate Mlp1 protein interactors in *sla1- S. pombe* cells. These data should validate if a cooperative interaction between Mlp1 and another protein increases 3′-end protection of nascent pre-tRNAs.

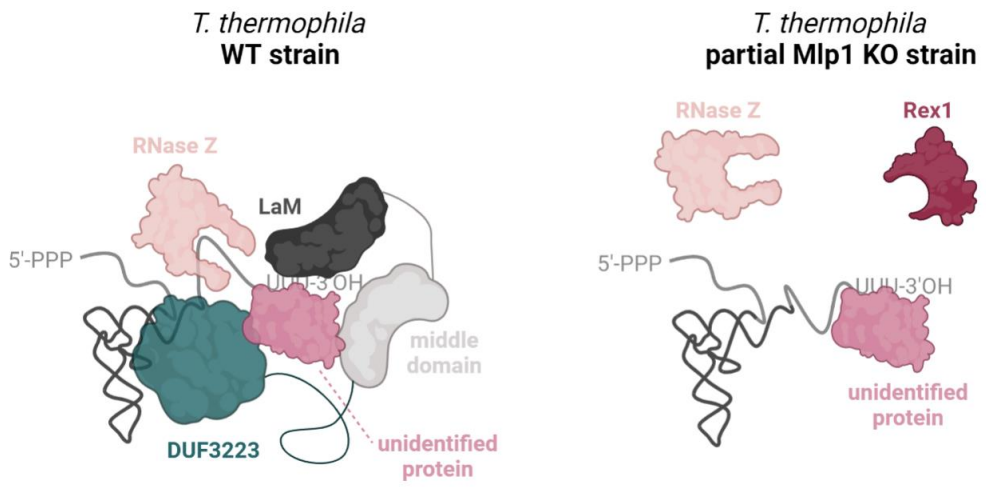


Figure 46. Pre-tRNA processing model of Mlp1 and an unidentified protein functioning in 3′-trailer accessibility for RNase Z cleavage and blocking Rex1 trimming in *Tetrahymena thermophila*.
Schematic representation of pre-tRNA processing in wild type (WT) and partial Mlp1 knockout (KO) strains in *T. thermophila*. The mature tRNA sequence is shown in black and 5′-leader and 3′-trailer sequences are shown in dark grey. In presence of Mlp1, RNA chaperone activity folds the pre-tRNA to ensure accessibility of the 3′-cleavage site for endonuclease RNase Z. The unidentified protein is hypothesized to associate with the 3′-trailer sequence. In absence of Mlp1, the unidentified protein remains associated with the 3′-trailer sequence, thereby blocking accessibility for trimming by the 3′-exonuclease Rex1. In addition, without the RNA chaperone activity of Mlp1, accessibility of the 3′-trailer sequence for RNase Z cleavage is delayed resulting in extended 3′-trailer sequences in Mlp1 depleted *T. thermophila* strains. Mlp1 is depicted containing three domains, including the LaM (shown in black), the middle domain (shown in light grey) and the C-terminal DUF3223 (shown in teal). Endonuclease RNase Z shown in light pink, exonuclease Rex1 shown in red. Image generated using BioRender.

5.2. RNA polymerase II transcription termination regulation of non-coding RNAs in *Tetrahymena thermophila*

Non-coding RNAs assume a large proportion of the transcriptome, including RNA polymerase I transcribed rRNA, RNA polymerase III transcribed tRNAs and RNA polymerase II transcribed snRNAs and snoRNAs (Boivin et al., 2018). The genomic organization of these non-coding RNAs is very different between transcript types and also between species (Zhang et al., 2019).

RNA polymerase II transcribed snRNA are present at multiple copies in vertebrate genomes. For example, the U1 snRNA gene is found at four copies in humans (O′Reilly et al., 2013). Our preliminary identification of snRNA genes in the *T. thermophila* genome resulted in identification of multiple genomic copies for U1, U2, U5 and U6 snRNAs (data not shown). A manual search using the small RNA sequencing dataset in the Integrated Genomics Viewer (IGV) revealed multiple identical copies in close proximity of these annotated snRNA genes, indicating that the current chromosomal locations of spliceosomal snRNAs we have obtained are incomplete and a more in-depth search is required to identify all chromosomal locations of *T. thermophila* snRNA genes.

Our analysis of the downstream genomic sequence of snRNA genes revealed that *T. thermophila* lacks any Nrd1/GUA[A/G] transcription termination binding site, while multiple Nab3/UCUU binding sites were identified. In addition, we identified a potential novel consensus sequence, UAAUCAAACAAAA, immediately downstream of the 3′-end of the RNA polymerase II transcribed snRNAs and immediately upstream of a Nab3/UCUU sequence. Our manual search for U1 snRNA confirmed that most U1 genes contain this novel motif, however, we observed a single variant U1 snRNA gene which has two nucleotide substitutions (sequence variation underlined: UAAUUAAAAAAAA). If these variant snRNA genes are expressed normally, we could study the effect of mutations in the novel motif by investigating RNA polymerase II run-through transcription products in *T. thermophila* by Northern blot. To further study this novel motif, *T. thermophila* strains carrying specific point mutants within the potential transcription termination motif could be generated and U snRNA transcription termination in *T. thermophila* can be elucidated.

Computational methods to predict snRNA genomic locations are a valuable tool to study snRNA gene sequences and surrounding *cis*-acting gene expression elements such as promoter

elements. It would be interesting to obtain the location of all U snRNA and study 3′-end processing of snRNAs with the presence/absence of this novel motif. However, throughout evolution multiple pseudogenes arose which are imperfect copies originally derived from functional genes that have obtained mutations, deletions or insertions that would disrupt their expression and/or function (Balakirev & Ayala, 2003). Therefore, to validate expression of snRNAs at a specific genomic location, an RNA polymerase II ChIP assay would be required for U1, U2, U4 and U5 snRNA and an RNA polymerase III ChIP assay for U6 snRNA validation. A study on U1 snRNA pseudogenes in humans revealed that multiple U1 pseudogenes are actively transcribed by RNA polymerase II through ChIP analysis and undergo assembly into functional snRNP (O′Reilly et al., 2013).

snoRNAs are also transcribed by RNA polymerase II and are encoded as a single copy either as individual genes, polycistronic genes or processed from pre-mRNA introns. The majority (approximately 60%) of previously identified snoRNA in *T. thermophila* are encoded within intergenic regions indicating that these snoRNAs are transcribed from independent genes (Andersen & Nielsen, 2012). The remaining 40% of snoRNAs are encoded within coding protein genes. A large polycistronic snoRNA cluster was also identified on chromosome 123 expressing box C/D7, C/D8, C/D9, C/D10, C/D11, C/D12, C/D13, C/D14, C/D15 and C/D16 snoRNAs. However, almost no work has been completed on elucidating transcription mechanisms of these snoRNA genes besides identification of a consensus upstream sequence element (USE), 5′-AAACCCATAA-3′, typically found in independently transcribed snoRNA genes (Andersen & Nielsen, 2012). More than half of the clustered snoRNAs have a divergent consensus sequence, which indicates that some clustered snoRNAs are co-transcribed as polycistronic transcripts. However, no experimental evidence has been provided to support these bioinformatics-based predictions.

5.3. Mlp1 functions in RNA polymerase II transcribed small non-coding RNA biogenesis

The best-characterized RNA targets of the La protein are nascent pre-tRNAs. However, La proteins are known to bind multiple other non-coding RNA targets, including other RNA polymerase III transcripts that contain 3′-uridylate tails, such as the 5S rRNA (Rinke & Steitz, 1982), spliceosomal U6 RNA (Rinke & Steitz, 1985), the signal recognition particle (SRP) RNA (Leung et al., 2014), cytoplasmic Y RNA (Simons et al., 1996) and vault RNAs (vRNAs) (Kickhoefer et al., 2002). Besides RNA polymerase III transcript RNA targets, the *Saccharomyces cerevisiae* La protein Lhp1p has been shown to associate with RNA polymerase II snRNA processing intermediates terminating in 3′-uridylate sequences (Xue et al., 2000). In addition to establishing a role for Mlp1 as a genuine La protein in altered pre-tRNA processing in *T. thermophila*, we also described interactions of Mlp1 with mature snRNA and snoRNAs, rather than precursor molecules as described previously (Maraia & Intine 2002). In addition to Mlp1 binding of mature snRNAs and snoRNA, we also identified direct protein-protein interactors by Mlp1-immunoprecipitation coupled to LC-MS/MS. These proteins included core proteins found in functional snRNPs (*i.e.* Sm-B) and snoRNPs (*i.e.* Nop10). The association of Mlp1 with both RNA and protein components of assembled RNPs led us to investigate a functional role of Mlp1 in splicing and the addition of post-transcriptional modifications on rRNA.

3.5.1. Mlp1 affects pre-mRNA splicing through decreased assembly of the U4/U6 di-snRNP

During splicing, the U4 and U6 snRNP undergo extensive RNA-RNA base pairing to form a single particle. Then, the U5 snRNP is recruited, joining the U4/U6 complex to form the trimeric U4/U6·U5 snRNP (tri-snRNP) which associates with the assembling spliceosome on the pre-

mRNA (Gehring & Roignant, 2021). Assembly of U6 snRNA into a functional U6 snRNP following transcription by RNA polymerase III involves multiple steps. First, the La protein associates with the 3′-terminal uridylate stretch on the nascent U6 snRNA transcript produced during RNA polymerase III transcription (Pannone et al., 1998). The LSm2-8 complex replaces La at the 3′-end of U6 snRNA which is required for nuclear retention of the U6 snRNP (Spiller et al., 2007). U6 snRNA obtains a γ-monomethyl cap (Singh & Reddy, 1989) and multiple additional post-transcriptional modifications, including 2′-O-methylations and pseudouridylations in the nucleolus (Ganot et al., 1999; Lange & Gerbi, 2000). Interestingly, the yeast *S. cerevisiae* U6 snRNA remains unmodified and thus also retains a strictly nuclear localization (Bertrand et al., 1998; Didychuk et al., 2018; Massenet et al., 1999). Previous work in *T. thermophila* identified snoRNAs that possess significant sequence complementarity to U6 snRNA sequences (Andersen & Nielsen, 2012), indicating that U6 snRNA in *T. thermophila* likely also obtains post-transcriptional modifications. Two 2′-O-methylation sites in the 5′-stem loop of U6 snRNA were validated by primer extension analysis (Andersen & Nielsen, 2012). Both box C/D21 and C/D30 snoRNAs, responsible for 2′-O-methylation of U6 snRNA, also contain complementary guide sequences against 17S rRNA and U6 snRNA (Andersen & Nielsen, 2012). Post-transcriptional modification of the RNA polymerase II transcribed snRNAs U1, U2, U4 and U5 takes place in a separate subnuclear compartment named Cajal bodies by Small Cajal body-specific RNAs (scaRNAs) (Jády et al., 2003). Localization of scaRNAs to Cajal-body involves the CAB-box localization signals which contains a UGAG sequence. This sequence is required and sufficient for localization of RNAs to Cajal bodies (Richard et al., 2003). In *T. thermophila*, scaRNAs were not identified, however, complementary guide sequences against U2, U4 and U5 were identified within box C/D and box H/ACA snoRNAs that guide post-transcriptional modifications in rRNA (Andersen & Nielsen, 2012). This could imply that RNA

polymerase II transcribed snRNAs obtain post-transcriptional modifications in the nucleolus of *T. thermophila*, however, further studies are required to confirm this.

Previous work in human cells has shown that assembly of the dimeric U4/U6 snRNP (di-snRNP) and U4/U6·U5 tri-snRNP takes place in Cajal bodies (Novotný et al., 2011; Staněk & Neugebauer, 2004), which also contain scaRNAs guiding post-transcriptional modifications on U1, U2, U4 and U5 (Jády et al., 2003). The yeast *S. cerevisiae* Prp24 (SART3 or p110 in humans) has been demonstrated to be important for assembly of the U4/U6 di-snRNP (Rader & Guthrie, 2002; Raghunathan & Guthrie, 1998). Free U6 snRNA contains a 3′-internal stem-loop which needs to be unfolded during di-snRNP formation since the nucleotides within this stem-loop are required for base pairing with U4 snRNA (**Figure 47**) (Fortner et al., 1994). Prp24 associates with the U6 snRNP through binding of the 3′-internal stem-loop of U6 snRNA and functions as an RNA chaperone to disrupt this sequence to enhance U6 base pairing with U4 snRNA through RNA chaperone activity. Once base pairing between U6 and U4 snRNAs has been established, Prp24 is removed from the di-snRNP (**Figure 47**) (Montemayor et al., 2018). In addition, Prp24 also interacts with the LSm2-8 complex bound to the 3′-end of U6 snRNA through contacts using the highly conserved SNFFL box found in the C-terminal region of Prp24 (**Figure 47**) (Rader & Guthrie, 2002). In the N-terminal region, Prp24 contains four RRMs, including RRM1-3 which adopt a classic RRM fold (two α-helices and a four-stranded antiparallel β-sheet containing two conserved RNP motifs in the central strands β1 and β3) and RRM4 adopting a non-canonical occluded RRM structure (oRRM) where two additional α-helices obscure the β-sheet face (Martin-Tumasz et al., 2011). The function of the C-terminal oRRM4 is to unwind the U6 3′-internal stem loop during U4/U6 di-snRNP assembly (Martin-Tumasz et al., 2011).

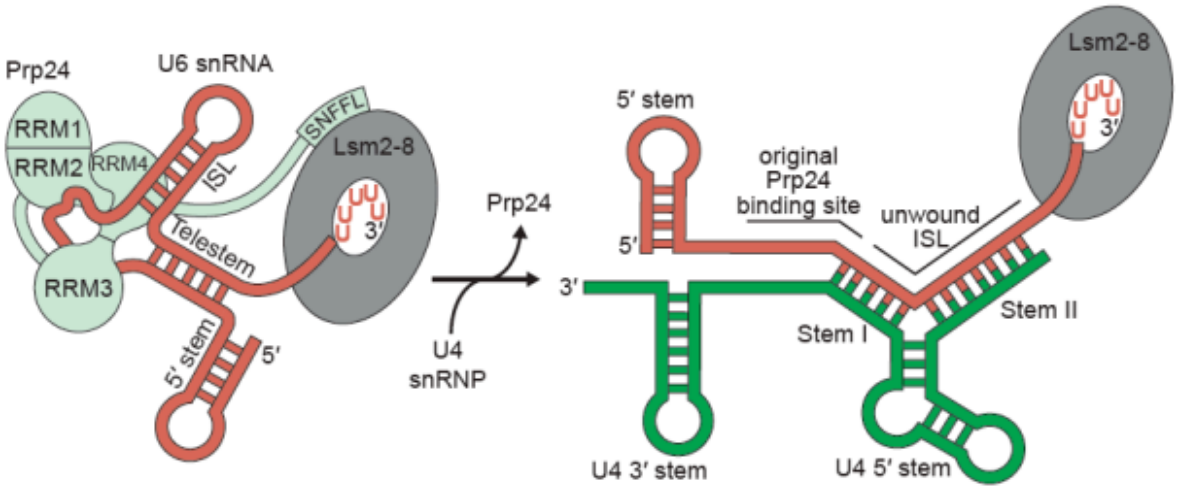


Figure 47. Schematic representation of U4/U6 di-snRNP assembly.

Assembly of the U4/U6 di-snRNP takes place in a stepwise manner in Cajal bodies. Prp24/SART3 catalyzes annealing of the U4/U6 di-snRNP through rearrangement of the 3′-internal stem loop (ISL) structure. Prp24 makes contacts with the LSm 2-8 complex using its C-terminal SNFFL box and RRM4 (not represented in this image). The LSm2-8 complex associates with the 3′-end of U6 snRNA. U4 snRNP proteins are not shown for clarity. Figure from (Montemayor et al., 2018).

Previous work revealed that deletion of the SNFFL box in Prp24 leads to a decrease in assembly of the U4/U6 di-snRNP (Rader & Guthrie, 2002). Similar to the Prp24 SNFFL box deletion mutant, we demonstrated that Mlp1 depletion results in decreased assembly of the U4/U6 di-snRNP. Both Prp24 and La proteins function as ATP-independent RNA chaperones, potentially indicating that Mlp1 could function similarly to Prp24 in reorganizing RNA structures for snRNP assembly (Ohndorf et al., 2001; Raghunathan & Guthrie, 1998). However, a notable difference between Mlp1 and Prp24 is that Mlp1 immunoprecipitates the RNA-RNA base paired U4/U6 di-snRNP, while Prp24 is removed upon base pairing of U4 and U6 snRNA since the Prp24 binding site on U6 has been rearranged to allow for RNA-RNA interactions with U4 snRNA (see **Figure 47**). An outstanding question is whether Mlp1 also associates, and as a result affects assembly, with the U4-U6•U5 tri-snRNPs which are assembled through association of U4/U6 di-snRNP with U5 snRNP through protein-protein interactions between

Prp31 and Prp6, respectively (Novotný et al., 2015). Future work will be needed to confirm which specific snRNP combinations are associated with Mlp1.

We studied the high-resolution structure of the U6 snRNP containing U6 snRNA, LSm2-8 and Prp24 components (**Figure 48a**) (Montemayor et al., 2018) to understand Prp24 conservation in *T. thermophila*. More specifically, we determined amino acid residues important for Prp24-U6 snRNA and Prp24-LSm2-8 complex interactions to investigate whether *T. thermophila* might express a functional Prp24 protein. Conservation analysis of a candidate *T. thermophila* Prp24 protein (http://www.ciliate.org/ TTHERM_000267919) reveals conservation of the C-terminal SNFFL box, which interacts with the LSm2-8 complex through direct protein-protein interactions with LSm5 and LSm7 (**Figure 48b,d**). However, the Prp24 oRRM4 region, which contacts Lsm2 in yeast (Montemayor et al., 2018), is less conserved between species (**Figure 48c,d**). Mutation of this region has no effect on Prp24 binding to U6, but had a two-fold decreased affinity of U6•Prp24 to LSm2-8 binding *in vitro* (Montemayor et al., 2018). The majority of contacts to the U6 snRNA 3′-internal stem loop are made by Prp24 RRM2 and RRM3 which are more highly conserved between different eukaryotic species, including the candidate *T. thermophila* Prp24 (**Figure 48d**). (Montemayor et al., 2018). Overall, *T. thermophila* possesses a conserved Prp24, however, future work is needed to confirm expression and function.

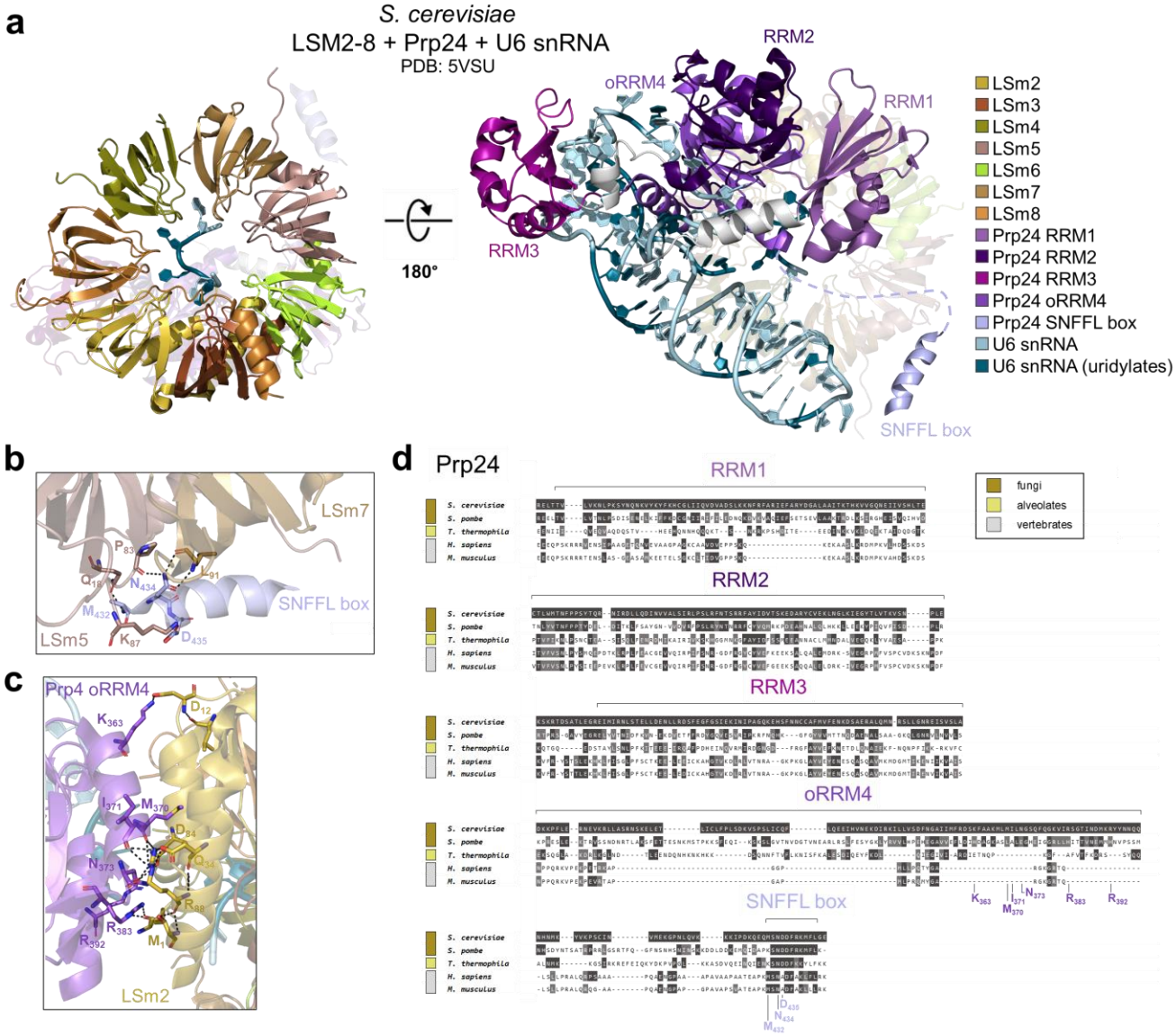


Figure caption on next page

Figure 48. *Tetrahymena thermophila* Prp24 contains conserved RNA- and protein-binding domains important for function.

a High-resolution structure from (Montemayor et al., 2018) (PDB: 5VSU) of the yeast U6 snRNP complex including U6 snRNA (shown in light blue, with uridylate RNA highlighted in dark blue), the LSm2-8 complex forming a ring-structure surrounding the 3′-end of the U6 snRNA (shown in different shades of brown, yellow, orange and green), and Prp24 (shown in different shaded of purple) interacting with U6 snRNA mostly using RRM2 and RRM3 and the LSm2-8 complex through Prp24 SNFFL box interactions with LSm5 and LSm7 and Prp24 oRRM4 interaction with LSm2. **b,c** Zoomed panels highlight residues involved in interactions between Prp24 SNFFL box, LSm5 and LSm7 (**b**) and between Prp24 oRRM4 and LSm2 (**c**). Carbon atoms are colored based on the molecule (different shades of purple for Prp24, yellow for LSm2 and different shades of brown for LSm5 and LSm7) with other atoms colored by type: oxygen in red; nitrogen in blue, phosphorous in orange and sulfur in yellow. Hydrogen bonds are shown as black dashed lines. **d** Primary amino acid alignments of Prp24 proteins from different eukaryotic lineages. Amino acids involved in protein-protein interactions with LSm2-8 complex proteins shown in **b,c** are annotated below. A dark grey background indicates identical residues, light grey conserved residues and white indicates no conservation.

The 5′-stem loop structure is not required for U6 snRNP assembly or U4/U6 di-snRNP formation in yeast as shown by *in vitro* assembly of the U6 snRNP (Montemayor et al., 2018). However, the *T. thermophila* U6 snRNA 5′-stem loop has a different secondary structure compared to other eukaryotes (**Figure 49**). It will be interesting to test whether these variations in Prp24 and the U6 snRNA are related to the involvement of a genuine La protein in U4/U6 di-snRNP maturation in *T. thermophila* compared to other eukaryotic species.

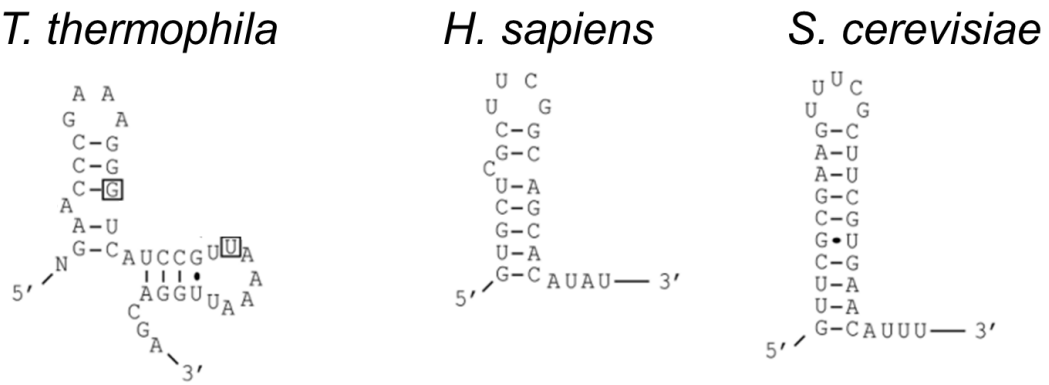


Figure 49. The secondary structure of 5′-stem loop in *Tetrahymena thermophila* is different from other eukaryotes.

Comparison of secondary structures of the U6 snRNA 5′-terminal stem loop between H*. sapiens, S. cerevisiae* and *T. thermophila*. Experimentally confirmed 2′-O-methylates nucleotides are boxed in *T. thermophila*. Figure adapted from (Andersen & Nielsen, 2012).

3.5.2. Mlp1 associates with box C/D and box H/ACA snoRNPs

Box H/ACA snoRNAs provide pseudouridylation and box C/D snoRNAs provide 2′-O-methylation post-transcriptional modifications on rRNA, snRNA and tRNAs within functional regions of the non-coding RNAs (Decatur & Fournier, 2002; Karijolich & Yu, 2010). All these transcripts function directly or indirectly in translation of proteins. Misregulation of pseudouridylation has been described to have major functional effects. This includes ribosome function (King et al., 2003), spliceosome function (Dönmez et al., 2004; Valadkhan & Manley, 2003; Yang et al., 2005) and small nuclear RNP assembly (Yu et al., 1998).

One of the major targets for pseudouridylation and 2′-O-methylation post-transcriptional modifications by snoRNAs is rRNA with 99 modifications in *S. cerevisiae* and 196 modifications in human previously identified (Decatur & Fournier, 2002). In our study, we established that Mlp1 associates with both mature snoRNAs and core protein components of the catalytically active snRNP. We predicted novel snoRNAs and their predicted post-transcriptional modifications in rRNA. To investigate if Mlp1 has a functional role in rRNA post-transcriptional modifications, we performed an RNA-DNA chimeric oligonucleotide-directed RNase H assay using novel predicted snoRNAs which will need further optimization. However, we hypothesize that Mlp1 depletion could affect the addition of certain post-transcriptional modification on rRNA or pre-rRNA processing through nucleolytic cleavage.

An outstanding question about the previously identified snoRNAs (Andersen & Nielsen, 2012) and newly predicted snoRNAs is the spatial position of the rRNA targets within the ribosome. To address this question, we are going to obtain the high-resolution structures for the 40S and 60S ribosomal subunits from the Protein Database (PDB: 4V5O, 4BTS and PDB: 4V8P, respectively) (Klinge et al., 2011; Rabl et al., 2011; Weisser et al., 2013) and map post-transcriptional modifications within the three-dimensional structure. This information will allow identifications of post-transcriptional modification in functionally important regions of the ribosome and allow us to study the most interesting snoRNAs in the future.

3.5.3. Mlp1 is essential in *Tetrahymena thermophila*

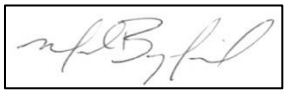
We established that a complete knockout of the Mlp1 protein results in a lethal phenotype based on phenotypic assortment (Merriam & Bruns, 1988) of alleles under increasing concentrations of selection drug during amitotic division of the macronucleus in *T. thermophila.* However, we have not established whether the essential function of Mlp1 is required for pre-tRNA processing, snRNP assembly, snoRNA binding or through a currently unidentified function of Mlp1. A first

experiment to inquire if a function in pre-tRNAs metabolism is important for viability is to generate a uridylate deficient binding (Mlp1 Q11A/Y14A) knock-in mutant strain and assess viability. This is hinging on the assumption that these residues are not important for snRNA or snoRNA interactions or are used for any direct protein-protein interactions of previously identified snRNP and snRNP protein binding partners of Mlp1. Alternatively, replacement of Mlp1 with mutations in other regions important for various classes of RNA binding or partner protein interaction might shed light on why Mlp1 is essential in this system.

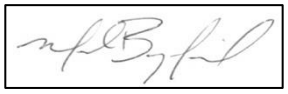
Overall, Mlp1 appears to have a central role in non-coding RNA biogenesis which ultimately is important for the translation of proteins. In Chapter II we demonstrated that Mlp1 directly binds and assists with maturation of nascent pre-tRNA in *T. thermophila* which eventually function as adapter molecules in translation, matching the correct amino acid with corresponding codon within the mRNA. Before the mRNA can be translated, the non-coding introns need to be spliced from the nascent pre-mRNA, which is completed by the spliceosome. In Chapter III, we demonstrate that Mlp1 associates with mature snRNAs, a core protein component of the snRNP and upon depletion, affects assembly of the U4/U6 snRNP and impedes splicing. Translation of mRNAs is completed by ribosomes which are assembled from mature rRNAs that have undergone processing and have obtained post-transcriptional modifications by snoRNAs. In Chapter IV, we found that Mlp1 associates with mature snoRNAs and core protein components of the snoRNPs. In addition to post-transcriptional modification of rRNA, snoRNAs are also involved in addition of post-transcriptional modification of tRNAs and snRNAs. Thus, Mlp1 represents a central node in the post-transcriptional processing and expression of factors that are essential for the protein translation apparatus.

Statement of Contributions

Kyra Kerkhofs (author) has prepared all aspects of this dissertation, designed, and performed all experiments, with exceptions of the author contributions for each chapter listed below.

Mark A. Bayfield, supervisor 

Chapter II has been published in a peer reviewed journal and Chapter III and IV are being prepared for publication. Permission to use published materials is given by the author and supervisor:

Mark A. Bayfield, supervisor 

Kyra Kerkhofs, author 

Chapter II: Altered tRNA processing is linked to a distinct and unusual La protein in *Tetrahymena thermophila*

Kerkhofs, Kyra[1]; Garg, Jyoti[1]; Fafard-Couture, Étienne[2]; Abou Elela, Sherif[3]; Scott, Michelle S.[2]; Pearlman, Ronald E.[1]; Bayfield, Mark A.[1]

Author contributions:

Jyoti Garg transformed, selected, and characterized the partial Mlp1 knockout (KO) strain in *Tetrahymena thermophila* by Southern blotting in **Supplementary Figure 7b** (original manuscript: Supplementary Figure S7b) and performed previous versions of immunofluorescent staining in **Supplementary Figure 7c** (original manuscript: Supplementary Figure S7c). Étienne Fafard-Couture performed Thermostable Group II Intron Reverse Transcriptase (TGIRT)-sequencing analysis generating transcripts per million (TPM) counts and generated a custom python script to obtain counts per million (CPM) for pre-tRNAs and mature tRNAs. These data were used to generate **Figure 15b** (original manuscript: Figure 2b) and **Figure 18c** (original manuscript: Figure 5c).

Chapter III: Genuine La protein Mlp1 functions in U4/U6 di-snRNP assembly and affects splicing in *Tetrahymena thermophila*

Kerkhofs, Kyra; Laganas, Jarred, Pearlman, Ronald E.; Bayfield, Mark A.

Author contributions:

Jarred Laganas and Ronald E. Pearlman provided annotated genomic locations of snRNAs in *Tetrahymena thermophila*.

Chapter IV: The *Tetrahymena thermophila* La protein Mlp1 associates with box C/D and box H/ACA snoRNAs

Kerkhofs, Kyra[1]; Fafard-Couture, Étienne[2]; Laganas, Jarred[1]; Scott, Michelle S.[2]; Pearlman, Ronald E.[1]; Bayfield, Mark A.[1]

Author contributions:

Étienne Fafard-Couture performed Thermostable Group II Intron Reverse Transcriptase (TGIRT)-sequencing analysis generating transcripts per million (TPM) and counts per million (CPM) data for small nucleolar RNAs (snoRNAs). Étienne Fafard-Couture completed snoRNA predictions using rFam, snoscan and SnoReport2.0. Jarred Laganas and Ronald E. Pearlman provided annotated genomic locations of previously published snoRNAs in *Tetrahymena thermophila*.

Additional contributions that were made throughout my Ph.D. that are not included in this dissertation:

Respiratory Syncytial Virus infection changes the cellular translational landscape toward AU-rich transcripts to enhance viral protein production

Kerkhofs, K.; Bayfield M.A. Publication in preparation.

Human La protein is required for efficient translation of a selection of mRNAs during stress granule inducing stress

Kerkhofs, K.; Bayfield M.A. Publication in preparation.

Heterogeneity in leukemia cells that escape drug-induced senescence

Miller, D.; Kerkhofs, K.; Madahar, S.S.; Bayfield, M.A.; Benchimol, S.

Leukemia. Publication under review.

La proteins couple use of sequence-specific and non-specific binding modes to engage RNA substrates.

Bayfield, M.A.; Vinayak, J.; Kerkhofs, K.; Mansouri-Noori, F.

RNA Biol. 2021 Feb;18(2):168-177. doi: 10.1080/15476286.2019.1582955. Epub 2019 Mar 18. PMID: 30777481; PMCID: PMC7928037.

C-lection by the DM15 Motif Gets LARP1 to the TOP.

Bayfield, N.A.; Kerkhofs, K.; Mansouri-Noori, F.

Structure. 2019 Dec 3;27(12):1737-1739. doi: 10.1016/j.str.2019.11.010. PMID: 31801095.

References

Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M., & Lührmann, R. (1999). A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *The EMBO Journal*, *18*(20), 5789–5802. https://doi.org/10.1093/emboj/18.20.5789

Aigner, S., & Cech, T. R. (2004). The Euplotes telomerase subunit p43 stimulates enzymatic activity and processivity in vitro. *RNA (New York, N.Y.)*, *10*(7), 1108–1118. https://doi.org/10.1261/rna.7400704

Aigner, S., Postberg, J., Lipps, H. J., & Cech, T. R. (2003). The Euplotes La motif protein p43 has properties of a telomerase-specific subunit. *Biochemistry*, *42*(19), 5736–5747. https://doi.org/10.1021/bi034121y

Alfano, C., Sanfelice, D., Babon, J., Kelly, G., Jacks, A., Curry, S., & Conte, M. R. (2004). Structural analysis of cooperative RNA binding by the La motif and central RRM domain of human La protein. *Nature Structural & Molecular Biology*, *11*(4), 323–329. https://doi.org/10.1038/nsmb747

Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E., & Tollervey, D. (1999). Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *The EMBO Journal*, *18*(19), 5399–5410. https://doi.org/10.1093/emboj/18.19.5399

Alspaugh, M. A., & Tan, E. M. (1975). Antibodies to cellular antigens in Sjögren's syndrome. *The Journal of Clinical Investigation*, *55*(5), 1067–1073. https://doi.org/10.1172/JCI108007

Andersen, K. L., & Nielsen, H. (2012). Experimental identification and analysis of macronuclear non-coding RNAs from the ciliate Tetrahymena thermophila. *Nucleic Acids Research*, *40*(3), 1267–1281. https://doi.org/10.1093/nar/gkr792

Apostolidi, M., Vourtsis, D. J., Chasapis, C. T., Stathopoulos, C., Bentrop, D., & Spyroulias, G. A. (2014). 1H, 15N, 13C assignment and secondary structure determination of two domains of La protein from D. discoideum. *Biomolecular NMR Assignments*, *8*(1), 47–51. https://doi.org/10.1007/s12104-012-9450-6

Arhin, G. K., Shen, S., Pérez, I. F., Tschudi, C., & Ullu, E. (2005). Downregulation of the essential Trypanosoma brucei La protein affects accumulation of elongator methionyl-tRNA. *Molecular and Biochemical Parasitology*, *144*(1), 104–108. https://doi.org/10.1016/j.molbiopara.2005.06.006

Arimbasseri, A. G., & Maraia, R. J. (2016). RNA Polymerase III Advances: Structural and tRNA Functional Views. *Trends in Biochemical Sciences*, *41*(6), 546–559. https://doi.org/10.1016/j.tibs.2016.03.003

Arimbasseri, A. G., Rijal, K., & Maraia, R. J. (2013). Transcription termination by the eukaryotic RNA polymerase III. *Biochimica Et Biophysica Acta*, *1829*(3–4), 318–330. https://doi.org/10.1016/j.bbagrm.2012.10.006

Arthur, L. L., Chung, J. J., Jankirama, P., Keefer, K. M., Kolotilin, I., Pavlovic-djuranovic, S., Chalker, D. L., Grbic, V., Green, R., Menassa, R., True, H. L., Skeath, J. B., & Djuranovic, S. (2017). Rapid generation of hypomorphic mutations. *Nature Communications*, *8*, 1–15. https://doi.org/10.1038/ncomms14112

Bai, C., & Tolias, P. P. (2000). Genetic analysis of a La homolog in Drosophila melanogaster. *Nucleic Acids Research*, *28*(5), 1078–1084. https://doi.org/10.1093/nar/28.5.1078

Baillat, D., Hakimi, M.-A., Näär, A. M., Shilatifard, A., Cooch, N., & Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell*, *123*(2), 265–276. https://doi.org/10.1016/j.cell.2005.08.019

Balakirev, E. S., & Ayala, F. J. (2003). Pseudogenes: Are They "Junk" or Functional DNA? *Annual Review of Genetics*, *37*(1), 123–151. https://doi.org/10.1146/annurev.genet.37.040103.103949

Barth, S., Shalem, B., Hury, A., Tkacz, I. D., Liang, X.-H., Uliel, S., Myslyuk, I., Doniger, T., Salmon-Divon, M., Unger, R., & Michaeli, S. (2008). Elucidating the role of C/D snoRNA in rRNA processing and modification in Trypanosoma brucei. *Eukaryotic Cell*, *7*(1), 86–101. https://doi.org/10.1128/EC.00215-07

Bayfield, M. A., Kaiser, T. E., Intine, R. V., & Maraia, R. J. (2007). Conservation of a Masked Nuclear Export Activity of La Proteins and Its Effects on tRNA Maturation. *Molecular and Cellular Biology*, *27*(9), 3303–3312. https://doi.org/10.1128/MCB.00026-07

Bayfield, M. A., & Maraia, R. J. (2009). Precursor-product discrimination by La protein during tRNA metabolism. *Nature Structural & Molecular Biology*, *16*(4), 430–437. https://doi.org/10.1038/nsmb.1573

Bayfield, M. A., Vinayak, J., Kerkhofs, K., & Mansouri-Noori, F. (2021). La proteins couple use of sequence-specific and non-specific binding modes to engage RNA substrates. *RNA Biology*, *18*(2), 168–177. https://doi.org/10.1080/15476286.2019.1582955

Bayfield, M. A., Yang, R., & Maraia, R. J. (2010). Conserved and divergent features of the structure and function of La and La-related proteins (LARPs). *Biochimica Et Biophysica Acta*, *1799*(5–6), 365–378. https://doi.org/10.1016/j.bbagrm.2010.01.011

Behrens, A., Rodschinka, G., & Nedialkova, D. D. (2021). High-resolution quantitative profiling of tRNA abundance and modification status in eukaryotes by mim-tRNAseq. *Molecular Cell*, *81*(8), 1802-1815.e7. https://doi.org/10.1016/j.molcel.2021.01.028

Belisova, A., Semrad, K., Mayer, O., Kocian, G., Waigmann, E., Schroeder, R., & Steiner, G. (2005). RNA chaperone activity of protein components of human Ro RNPs. *RNA (New York, N.Y.)*, *11*(7), 1084–1094. https://doi.org/10.1261/rna.7263905

Bellaoui, M., Keddie, J. S., & Gruissem, W. (2003). DCL is a plant-specific protein required for plastid ribosomal RNA processing and embryo development. *Plant Molecular Biology*, *53*(4), 531–543. https://doi.org/10.1023/B:PLAN.0000019061.79773.06

Berg, M. D., Giguere, D. J., Dron, J. S., Lant, J. T., Genereaux, J., Liao, C., Wang, J., Robinson, J. F., Gloor, G. B., Hegele, R. A., O'Donoghue, P., & Brandl, C. J. (2019). Targeted sequencing reveals expanded genetic diversity of human transfer RNAs. *RNA Biology*, *16*(11), 1574–1585. https://doi.org/10.1080/15476286.2019.1646079

Bertrand, E., Houser-Scott, F., Kendall, A., Singer, R. H., & Engelke, D. R. (1998). Nucleolar localization of early tRNA processing. *Genes & Development*, *12*(16), 2463–2468.

Bhattacharya, R., Perumal, K., Sinha, K., Maraia, R., & Reddy, R. (2018). Methylphosphate Cap Structure in Small RNAs Reduces the Affinity of RNAs to La Protein. *Gene Expression*, *10*(5–6), 243–253.

Bogenhagen, D. F., & Brown, D. D. (1981). Nucleotide sequences in Xenopus 5S DNA required for transcription termination. *Cell*, *24*(1), 261–270. https://doi.org/10.1016/0092-8674(81)90522-5

Boivin, V., Deschamps-Francoeur, G., Couture, S., Nottingham, R. M., Bouchard-Bourelle, P., Lambowitz, A. M., Scott, M. S., & Abou-Elela, S. (2018). Simultaneous sequencing of coding and noncoding RNA reveals a human transcriptome dominated by a small number of highly expressed noncoding genes. *RNA (New York, N.Y.)*, *24*(7), 950–965. https://doi.org/10.1261/rna.064493.117

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bousquet-Antonelli, C., & Deragon, J.-M. (2009). A comprehensive analysis of the La-motif protein superfamily. *Rna*, *15*(5), 750–764. https://doi.org/10.1261/rna.1478709

Cahill, N. M., Friend, K., Speckmann, W., Li, Z.-H., Terns, R. M., Terns, M. P., & Steitz, J. A. (2002). Site-specific cross-linking analyses reveal an asymmetric protein distribution for a box C/D snoRNP. *The EMBO Journal*, *21*(14), 3816–3828. https://doi.org/10.1093/emboj/cdf376

Campbell, D. A., Kubo, K., Clark, C. G., & Boothroyd, J. C. (1987). Precise identification of cleavage sites involved in the unusual processing of trypanosome ribosomal RNA. *Journal of Molecular Biology*, *196*(1), 113–124. https://doi.org/10.1016/0022-2836(87)90514-6

Cantara, W. A., Crain, P. F., Rozenski, J., McCloskey, J. A., Harris, K. A., Zhang, X., Vendeix, F. A. P., Fabris, D., & Agris, P. F. (2011). The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Research*, *39*(Database issue), D195–D201. https://doi.org/10.1093/nar/gkq1028

Carroll, K. L., Pradhan, D. A., Granek, J. A., Clarke, N. D., & Corden, J. L. (2004). Identification of cis elements directing termination of yeast nonpolyadenylated snoRNA transcripts. *Molecular and Cellular Biology*, *24*(14), 6241–6252. https://doi.org/10.1128/MCB.24.14.6241-6252.2004

Carter, C. W., & Wolfenden, R. (2015). TRNA acceptor stem and anticodon bases form independent codes related to protein folding. *Proceedings of the National Academy of Sciences*, *112*(24), 7489–7494. https://doi.org/10.1073/pnas.1507569112

Carter, R., & Drouin, G. (2009). Structural differentiation of the three eukaryotic RNA polymerases. *Genomics*, *94*(6), 388–396. https://doi.org/10.1016/j.ygeno.2009.08.011

Cash, D. D., & Feigon, J. (2017). Structure and folding of the Tetrahymena telomerase RNA pseudoknot. *Nucleic Acids Research*, *45*(1), 482–495. https://doi.org/10.1093/nar/gkw1153

Cavaillé, J., Nicoloso, M., & Bachellerie, J. P. (1996). Targeted ribose methylation of RNA in vivo directed by tailored antisense RNA guides. *Nature*, *383*(6602), 732–735. https://doi.org/10.1038/383732a0

Chakshusmathi, G., Kim, S. D., Rubinson, D. A., & Wolin, S. L. (2003). A La protein requirement for efficient pre-tRNA folding. *The EMBO Journal*, *22*(24), 6562–6572. https://doi.org/10.1093/emboj/cdg625

Chalker, D. L. (2012). Transformation and strain engineering of Tetrahymena. *Methods in Cell Biology*, *109*, 327–345. https://doi.org/10.1016/B978-0-12-385967-9.00011-6

Chambers, J. C., Kurilla, M. G., & Keene, J. D. (1983). Association between the 7 S RNA and the lupus La protein varies among cell types. *The Journal of Biological Chemistry*, *258*(19), 11438–11441.

Chan, P. P., & Lowe, T. M. (2016). GtRNAdb 2.0: An expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Research*, *44*(D1), D184–D189. https://doi.org/10.1093/nar/gkv1309

Chanfreau, G., Elela, S. A., Ares, M., & Guthrie, C. (1997). Alternative 3′-end processing of U5 snRNA by RNase III. *Genes & Development*, *11*(20), 2741–2751.

Chen, J. L., & Pace, N. R. (1997). Identification of the universally conserved core of ribonuclease P RNA. *RNA*, *3*(6), 557–560.

Chen, W., & Moore, M. J. (2015). Spliceosomes. *Current Biology: CB*, *25*(5), R181-183. https://doi.org/10.1016/j.cub.2014.11.059

Copela, L. A., Chakshusmathi, G., Sherrer, R. L., & Wolin, S. L. (2006). The La protein functions redundantly with tRNA modification enzymes to ensure tRNA structural stability. *RNA*, *12*(4), 644–654. https://doi.org/10.1261/rna.2307206

Copela, L. A., Fernandez, C. F., Sherrer, R. L., & Wolin, S. L. (2008). Competition between the Rex1 exonuclease and the La protein affects both Trf4p-mediated RNA quality control and pre-tRNA maturation. *RNA*, *14*(6), 1214–1227. https://doi.org/10.1261/rna.1050408

Cozen, A. E., Quartley, E., Holmes, A. D., Hrabeta-Robinson, E., Phizicky, E. M., & Lowe, T. M. (2015). ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nature Methods*, *12*(9), 879–884. https://doi.org/10.1038/nmeth.3508

Crooks, G. E., Hon, G., Chandonia, J.-M., & Brenner, S. E. (2004). WebLogo: A Sequence Logo Generator. *Genome Research*, *14*(6), 1188–1190. https://doi.org/10.1101/gr.849004

de Araujo Oliveira, J. V., Costa, F., Backofen, R., Stadler, P. F., Machado Telles Walter, M. E., & Hertel, J. (2016). SnoReport 2.0: New features and a refined Support Vector Machine to improve snoRNA identification. *BMC Bioinformatics*, *17*(18), 464. https://doi.org/10.1186/s12859-016-1345-6

Decatur, W. A., & Fournier, M. J. (2002). RRNA modifications and ribosome function. *Trends in Biochemical Sciences*, *27*(7), 344–351. https://doi.org/10.1016/s0968-0004(02)02109-6

Deragon, J. (2020). Distribution , organization an evolutionary history of La and LARPs in eukaryotes. *RNA Biology*, *00*(00), 1–9. https://doi.org/10.1080/15476286.2020.1739930

Deschamps-Francoeur, G., Boivin, V., Abou Elela, S., & Scott, M. S. (2019). CoCo: RNA-seq read assignment correction for nested genes and multimapped reads. *Bioinformatics*, *35*(23), 5039–5047. https://doi.org/10.1093/bioinformatics/btz433

Didychuk, A. L., Butcher, S. E., & Brow, D. A. (2018). The life of U6 small nuclear RNA, from cradle to grave. *RNA*, *24*(4), 437–460. https://doi.org/10.1261/rna.065136.117

Dieci, G., Fiorino, G., Castelnuovo, M., Teichmann, M., & Pagano, A. (2007). The expanding RNA polymerase III transcriptome. *Trends in Genetics*, *23*(12), 614–622. https://doi.org/10.1016/j.tig.2007.09.001

Dobin, A., & Gingeras, T. R. (2016). Optimizing RNA-Seq Mapping with STAR. In O. Carugo & F. Eisenhaber (Eds.), *Data Mining Techniques for the Life Sciences* (pp. 245–262). Springer. https://doi.org/10.1007/978-1-4939-3572-7_13

Dock-Bregeon, A.-C., Lewis, K. A., & Conte, M. R. (2019). The La-related proteins: Structures and interactions of a versatile superfamily of RNA-binding proteins. *RNA Biology*, *0*(0), 1–16. https://doi.org/10.1080/15476286.2019.1695712

Dong, G., Chakshusmathi, G., Wolin, S. L., & Reinisch, K. M. (2004). Structure of the La motif: A winged helix domain mediates RNA binding via a conserved aromatic patch. *EMBO Journal*, *23*(5), 1000–1007. https://doi.org/10.1038/sj.emboj.7600115

Dönmez, G., Hartmuth, K., & Lührmann, R. (2004). Modified nucleotides at the 5' end of human U2 snRNA are required for spliceosomal E-complex formation. *RNA (New York, N.Y.)*, *10*(12), 1925–1933. https://doi.org/10.1261/rna.7186504

Dunbar, D. A., & Baserga, S. J. (1998). The U14 snoRNA is required for 2'-O-methylation of the pre-18S rRNA in Xenopus oocytes. *RNA (New York, N.Y.)*, *4*(2), 195–204.

Eichhorn, C. D., Chug, R., & Feigon, J. (2016). HLARP7 C-terminal domain contains an xRRM that binds the 3′ hairpin of 7SK RNA. *Nucleic Acids Research*, *44*(20), 9977. https://doi.org/10.1093/nar/gkw833

Eisen, J. A., Coyne, R. S., Wu, M., Wu, D., Thiagarajan, M., Wortman, J. R., Badger, J. H., Ren, Q., Amedeo, P., Jones, K. M., Tallon, L. J., Delcher, A. L., Salzberg, S. L., Silva, J. C., Haas, B. J., Majoros, W. H., Farzad, M., Carlton, J. M., Smith, R. K., … Orias, E. (2006). Macronuclear genome sequence of the ciliate Tetrahymena thermophila, a model eukaryote. *PLoS Biology*, *4*(9), 1620–1642. https://doi.org/10.1371/journal.pbio.0040286

El Yacoubi, B., Bailly, M., & de Crécy-Lagard, V. (2012). Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annual Review of Genetics*, *46*, 69–95. https://doi.org/10.1146/annurev-genet-110711-155641

Enright, C. A., Maxwell, E. S., Eliceiri, G. L., & Sollner-Webb, B. (1996). 5'ETS rRNA processing facilitated by four small RNAs: U14, E3, U17, and U3. *RNA (New York, N.Y.)*, *2*(11), 1094–1099.

Fafard-Couture, É., Bergeron, D., Couture, S., Abou-Elela, S., & Scott, M. S. (2021). Annotation of snoRNA abundance across human tissues reveals complex snoRNA-host gene relationships. *Genome Biology*, *22*(1), 172. https://doi.org/10.1186/s13059-021-02391-2

Fairley, J. A., Kantidakis, T., Kenneth, N. S., Intine, R. V., Maraia, R. J., & White, R. J. (2005). Human La is found at RNA polymerase III-transcribed genes in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(51), 18350–18355. https://doi.org/10.1073/pnas.0506415102

Falaleeva, M., & Stamm, S. (2013). Processing of snoRNAs as a new source of regulatory non-coding RNAs snoRNA fragments form a new class of functional RNAs. *BioEssays :*

News and Reviews in Molecular, Cellular and Developmental Biology, *35*(1), 46–54. https://doi.org/10.1002/bies.201200117

Fan, H., Goodier, J. L., Chamberlain, J. R., Engelke, D. R., & Maraia, R. J. (1998). 5′ Processing of tRNA Precursors Can Be Modulated by the Human La Antigen Phosphoprotein. *Molecular and Cellular Biology*, *18*(6), 3201–3211.

Fillingham, J. S., Bruno, D., & Pearlman, R. E. (2001). Cis-acting requirements in flanking DNA for the programmed elimination of mse2.9: A common mechanism for deletion of internal eliminated sequences from the developing macronucleus of Tetrahymena thermophila. *Nucleic Acids Research*, *29*(2), 488–498. https://doi.org/10.1093/nar/29.2.488

Fleurdépine, S., Deragon, J.-M., Devic, M., Guilleminot, J., & Bousquet-Antonelli, C. (2007). A bona fide La protein is required for embryogenesis in Arabidopsis thaliana. *Nucleic Acids Research*, *35*(10), 3306–3321. https://doi.org/10.1093/nar/gkm200

Fortner, D. M., Troy, R. G., & Brow, D. A. (1994). A stem/loop in U6 RNA defines a conformational switch required for pre-mRNA splicing. *Genes & Development*, *8*(2), 221–233. https://doi.org/10.1101/gad.8.2.221

Gaidamakov, S., Maximova, O. A., Chon, H., Blewett, N. H., Wang, H., Crawford, A. K., Day, A., Tulchin, N., Crouch, R. J., Morse, H. C., Blitzer, R. D., & Maraia, R. J. (2014). Targeted deletion of the gene encoding the La autoantigen (Sjögren's syndrome antigen B) in B cells or the frontal brain causes extensive tissue loss. *Molecular and Cellular Biology*, *34*(1), 123–131. https://doi.org/10.1128/MCB.01010-13

Gajiwala, K. S., & Burley, S. K. (2000). Winged helix proteins. *Current Opinion in Structural Biology*, *10*(1), 110–116. https://doi.org/10.1016/S0959-440X(99)00057-3

Ganot, P., Bortolin, M.-L., & Kiss, T. (1997). Site-Specific Pseudouridine Formation in Preribosomal RNA Is Guided by Small Nucleolar RNAs. *Cell*, *89*(5), 799–809. https://doi.org/10.1016/S0092-8674(00)80263-9

Ganot, P., Jády, B. E., Bortolin, M. L., Darzacq, X., & Kiss, T. (1999). Nucleolar factors direct the 2'-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA. *Molecular and Cellular Biology*, *19*(10), 6906–6917. https://doi.org/10.1128/MCB.19.10.6906

Gehring, N. H., & Roignant, J.-Y. (2021). Anything but Ordinary – Emerging Splicing Mechanisms in Eukaryotic Gene Regulation. *Trends in Genetics*, *37*(4), 355–372. https://doi.org/10.1016/j.tig.2020.10.008

Geslain, R., & Pan, T. (2010). Functional analysis of human tRNA isodecoders. *Journal of Molecular Biology*, *396*(3), 821–831. https://doi.org/10.1016/j.jmb.2009.12.018

Ghosh, A., & Lima, C. D. (2010). Enzymology of RNA cap synthesis. *Wiley Interdisciplinary Reviews. RNA*, *1*(1), 152–172. https://doi.org/10.1002/wrna.19

Gogakos, T., Brown, M., Garzia, A., Meyer, C., Hafner, M., & Tuschl, T. (2017). Characterizing Expression and Processing of Precursor and Mature Human tRNAs by Hydro-tRNAseq and PAR-CLIP. *Cell Reports*, *20*(6), 1463–1475. https://doi.org/10.1016/j.celrep.2017.07.029

Goodier, J. L., Fan, H., & Maraia, R. J. (1997). A carboxy-terminal basic region controls RNA polymerase III transcription factor activity of human La protein. *Molecular and Cellular Biology*, *17*(10), 5823–5832.

Größringer, M., Helmecke, D., & Hartmann, R. K. (2012). Characterization of RNase P RNA Activity. In J. S. Hartig (Ed.), *Ribozymes: Methods and Protocols* (pp. 61–72). Humana Press. https://doi.org/10.1007/978-1-61779-545-9_5

Guddat, U., Bakken, A. H., & Pieler, T. (1990). Protein-mediated nuclear export of RNA: 5S rRNA containing small RNPs in xenopus oocytes. *Cell*, *60*(4), 619–628. https://doi.org/10.1016/0092-8674(90)90665-2

Guiro, J., & Murphy, S. (2017). Regulation of expression of human RNA polymerase II-transcribed snRNA genes. *Open Biology*, *7*(6), 170073. https://doi.org/10.1098/rsob.170073

Gupta, S., Busch, R. K., Singh, R., & Reddy, R. (1990). Characterization of U6 small nuclear RNA cap-specific antibodies. Identification of gamma-monomethyl-GTP cap structure in 7SK and several other human small RNAs. *The Journal of Biological Chemistry*, *265*(31), 19137–19142.

Hardin, J. W., Warnasooriya, C., Kondo, Y., Nagai, K., & Rueda, D. (2015). Assembly and dynamics of the U4/U6 di-snRNP by single-molecule FRET. *Nucleic Acids Research*, *43*(22), 10963–10974. https://doi.org/10.1093/nar/gkv1011

Hayes, K. E., Barr, J. A., Xie, M., Steitz, J. A., & Martinez, I. (2018). Immunoprecipitation of Tri-methylated Capped RNA. *Bio-Protocol*, *8*(3), e2717–e2717.

He, N., Jahchan, N. S., Hong, E., Li, Q., Bayfield, M. A., Maraia, R. J., Luo, K., & Zhou, Q. (2008). A La-related protein modulates 7SK snRNP integrity to suppress P-TEFb-dependent transcriptional elongation and tumorigenesis. *Molecular Cell*, *29*(5), 588–599. https://doi.org/10.1016/j.molcel.2008.01.003

Helm, M. (2006). Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Research*, *34*(2), 721–733. https://doi.org/10.1093/nar/gkj471

Hendrick, J. P., Wolin, S. L., Rinke, J., Lerner, M. R., & Steitz, J. A. (1981). Ro small cytoplasmic ribonucleoproteins are a subclass of La ribonucleoproteins: Further characterization of the Ro and La small ribonucleoproteins from uninfected mammalian cells. *Molecular and Cellular Biology*, *1*(12), 1138–1149.

Henras, A. K., Plisson-Chastang, C., O'Donohue, M.-F., Chakraborty, A., & Gleizes, P.-E. (2015). An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdisciplinary Reviews. RNA*, *6*(2), 225–242. https://doi.org/10.1002/wrna.1269

Hernandez, N. (1985). Formation of the 3' end of U1 snRNA is directed by a conserved sequence located downstream of the coding region. *The EMBO Journal*, *4*(7), 1827–1837.

Hertel, J., Hofacker, I. L., & Stadler, P. F. (2008). SnoReport: Computational identification of snoRNAs with unknown targets. *Bioinformatics*, *24*(2), 158–164. https://doi.org/10.1093/bioinformatics/btm464

Hobor, F., Pergoli, R., Kubicek, K., Hrossova, D., Bacikova, V., Zimmermann, M., Pasulka, J., Hofr, C., Vanacova, S., & Stefl, R. (2011). Recognition of Transcription Termination Signal by the Nuclear Polyadenylated RNA-binding (NAB) 3 Protein. *Journal of Biological Chemistry*, *286*(5), 3645–3657. https://doi.org/10.1074/jbc.M110.158774

Huang, Y., Bayfield, M. A., Intine, R. V., & Maraia, R. J. (2006). Separate RNA-binding surfaces on the multifunctional La protein mediate distinguishable activities in tRNA maturation. *Nature Structural & Molecular Biology*, *13*(7), Article 7. https://doi.org/10.1038/nsmb1110

Huang, Y., Intine, R. V., Mozlin, A., Hasson, S., & Maraia, R. J. (2005). Mutations in the RNA Polymerase III Subunit Rpc11p That Decrease RNA 3′ Cleavage Activity Increase 3′-Terminal Oligo(U) Length and La-Dependent tRNA Processing. *Molecular and Cellular Biology*, *25*(2), 621–636. https://doi.org/10.1128/mcb.25.2.621-636.2005

Hughes, J. M., & Ares, M. (1991). Depletion of U3 small nucleolar RNA inhibits cleavage in the 5' external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA. *The EMBO Journal*, *10*(13), 4231–4239. https://doi.org/10.1002/j.1460-2075.1991.tb05001.x

Hussain, R. H., Zawawi, M., & Bayfield, M. A. (2013). Conservation of RNA chaperone activity of the human La-related proteins 4, 6 and 7. *Nucleic Acids Research*, *41*(18), 8715–8725. https://doi.org/10.1093/nar/gkt649

Inada, M., & Guthrie, C. (2004). Identification of Lhp1p-associated RNAs by microarray analysis in Saccharomyces cerevisiae reveals association with coding and noncoding RNAs. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(2), 434–439. https://doi.org/10.1073/pnas.0307425100

Intine, R. V. A., Sakulich, A. L., Koduru, S. B., Huang, Y., Pierstorff, E., Goodier, J. L., Phan, L., & Maraia, R. J. (2000). Control of Transfer RNA Maturation by Phosphorylation of the Human La Antigen on Serine 366. *Molecular Cell*, *6*(2), 339–348. https://doi.org/10.1016/S1097-2765(00)00034-4

Intine, R. V., Dundr, M., Misteli, T., & Maraia, R. J. (2002). Aberrant nuclear trafficking of La protein leads to disordered processing of associated precursor tRNAs. *Molecular Cell*, *9*(5), 1113–1123. https://doi.org/10.1016/S1097-2765(02)00533-6

Jacks, A., Babon, J., Kelly, G., Manolaridis, I., Cary, P. D., Curry, S., & Conte, M. R. (2003). Structure of the C-terminal domain of human La protein reveals a novel RNA recognition motif coupled to a helical nuclear retention element. *Structure (London, England: 1993)*, *11*(7), 833–843.

Jády, B. E., Darzacq, X., Tucker, K. E., Matera, A. G., Bertrand, E., & Kiss, T. (2003). Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal body following import from the cytoplasm. *The EMBO Journal*, *22*(8), 1878–1888. https://doi.org/10.1093/emboj/cdg187

Jády, B. E., & Kiss, T. (2001). A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. *The EMBO Journal*, *20*(3), 541–551. https://doi.org/10.1093/emboj/20.3.541

Jansen, R. P. (1991). Evolutionary conservation of the human nucleolar protein fibrillarin and its functional expression in yeast. *The Journal of Cell Biology*, *113*(4), 715–729.

Johnson, P. F., & Abelson, J. (1983). The yeast tRNATyr gene intron is essential for correct modification of its tRNA product. *Nature*, *302*(5910), 681–687. https://doi.org/10.1038/302681a0

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard,

A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), Article 7873. https://doi.org/10.1038/s41586-021-03819-2

Jurica, M. S., & Moore, M. J. (2003). Pre-mRNA Splicing: Awash in a Sea of Proteins. *Molecular Cell*, *12*(1), 5–14. https://doi.org/10.1016/S1097-2765(03)00270-3

Kapler, G. M. (1993). Developmentally regulated processing and replication of the Tetrahymena rDNA minichromosome. *Current Opinion in Genetics & Development*, *3*(5), 730–735. https://doi.org/10.1016/s0959-437x(05)80091-7

Karijolich, J., & Yu, Y.-T. (2010). Spliceosomal snRNA modifications and their function. *RNA Biology*, *7*(2), 192–204.

Kass, S., Tyc, K., Steitz, J. A., & Sollner-Webb, B. (1990). The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing. *Cell*, *60*(6), 897–908. https://doi.org/10.1016/0092-8674(90)90338-f

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, *10*(6), 845–858. https://doi.org/10.1038/nprot.2015.053

Kenan, D. J., & Keene, J. D. (2004). La gets its wings. *Nature Structural and Molecular Biology*, *11*(4), 303–305. https://doi.org/10.1038/nsmb0404-303

Kerkhofs, K., Garg, J., Fafard-Couture, É., Abou Elela, S., Scott, M. S., Pearlman, R. E., & Bayfield, M. A. (2022). Altered tRNA processing is linked to a distinct and unusual La protein in Tetrahymena thermophila. *Nature Communications*, *13*(1), Article 1. https://doi.org/10.1038/s41467-022-34796-3

Kickhoefer, V. A., Poderycki, M. J., Chan, E. K. L., & Rome, L. H. (2002). The La RNA-binding Protein Interacts with the Vault RNA and Is a Vault-associated Protein*. *Journal of Biological Chemistry*, *277*(43), 41282–41286. https://doi.org/10.1074/jbc.M206980200

King, T. H., Liu, B., McCully, R. R., & Fournier, M. J. (2003). Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Molecular Cell*, *11*(2), 425–435. https://doi.org/10.1016/s1097-2765(03)00040-6

Kirstein, N., Gomes Dos Santos, H., Blumenthal, E., & Shiekhattar, R. (2021). The Integrator complex at the crossroad of coding and noncoding RNA. *Current Opinion in Cell Biology*, *70*, 37–43. https://doi.org/10.1016/j.ceb.2020.11.003

Kishore, S., & Stamm, S. (2006). The snoRNA HBII-52 Regulates Alternative Splicing of the Serotonin Receptor 2C. *Science*, *311*(5758), 230–232. https://doi.org/10.1126/science.1118265

Kiss, T. (2001). Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *The EMBO Journal*, *20*(14), 3617–3622. https://doi.org/10.1093/emboj/20.14.3617

Kiss, T. (2002). Small Nucleolar RNAs: An Abundant Group of Noncoding RNAs with Diverse Cellular Functions. *Cell*, *109*(2), 145–148. https://doi.org/10.1016/S0092-8674(02)00718-3

Kiss-László, Z., Henry, Y., Bachellerie, J. P., Caizergues-Ferrer, M., & Kiss, T. (1996). Site-specific ribose methylation of preribosomal RNA: A novel function for small nucleolar RNAs. *Cell*, *85*(7), 1077–1088. https://doi.org/10.1016/s0092-8674(00)81308-2

Kister, K. P., Müller, B., & Eckert, W. A. (1983). Complex endonucleolytic cleavage pattern during early events in the processing of pre-rRNA in the lower eukaryote, Tetrahymena thermophila. *Nucleic Acids Research*, *11*(11), 3487–3502. https://doi.org/10.1093/nar/11.11.3487

Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., Arpagaus, S., & Ban, N. (2011). Crystal Structure of the Eukaryotic 60S Ribosomal Subunit in Complex with Initiation Factor 6. *Science*, *334*(6058), 941–948. https://doi.org/10.1126/science.1211204

Klinge, S., & Woolford, J. L. (2019). Ribosome assembly coming into focus. *Nature Reviews Molecular Cell Biology*, *20*(2), Article 2. https://doi.org/10.1038/s41580-018-0078-y

Kotik-Kogan, O., Valentine, E. R., Sanfelice, D., Conte, M. R., & Curry, S. (2008). Structural Analysis Reveals Conformational Plasticity in the Recognition of RNA 3′ Ends by the Human La Protein. *Structure*, *16*(6), 852–862. https://doi.org/10.1016/j.str.2008.02.021

Kozlov, G., Mattijssen, S., Jiang, J., Nyandwi, S., Sprules, T., Iben, J. R., Coon, S. L., Gaidamakov, S., Noronha, A. M., Wilds, C. J., Maraia, R. J., & Gehring, K. (2022). Structural basis of 3′-end poly(A) RNA recognition by LARP1. *Nucleic Acids Research*, *50*(16), 9534–9547. https://doi.org/10.1093/nar/gkac696

Krueger, B. J., Jeronimo, C., Roy, B. B., Bouchard, A., Barrandon, C., Byers, S. A., Searcey, C. E., Cooper, J. J., Bensaude, O., Cohen, E. A., Coulombe, B., & Price, D. H. (2008). LARP7 is a stable component of the 7SK snRNP while P-TEFb, HEXIM1 and hnRNP A1 are reversibly associated. *Nucleic Acids Research*, *36*(7), 2219–2229. https://doi.org/10.1093/nar/gkn061

Kufel, J., Allmang, C., Chanfreau, G., Petfalski, E., Lafontaine, D. L., & Tollervey, D. (2000). Precursors to the U3 small nucleolar RNA lack small nucleolar RNP proteins but are stabilized by La binding. *Molecular and Cellular Biology*, *20*(15), 5415–5424. https://doi.org/10.1128/MCB.20.15.5415-5424.2000

Kufel, J., Allmang, C., Verdone, L., Beggs, J. D., & Tollervey, D. (2002). Lsm proteins are required for normal processing of pre-tRNAs and their efficient association with La-homologous protein Lhp1p. *Molecular and Cellular Biology*, *22*(14), 5248–5256. https://doi.org/10.1128/MCB.22.14.5248-5256.2002

Kufel, J., & Grzechnik, P. (2019). Small Nucleolar RNAs Tell a Different Tale. *Trends in Genetics*, *35*(2), 104–117. https://doi.org/10.1016/j.tig.2018.11.005

Kufel, J., & Tollervey, D. (2003). 3'-processing of yeast tRNATrp precedes 5'-processing. *RNA (New York, N.Y.)*, *9*(2), 202–208. https://doi.org/10.1261/rna.2145103

Kunkel, G. R., Maser, R. L., Calvet, J. P., & Pederson, T. (1986). U6 small nuclear RNA is transcribed by RNA polymerase III. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(22), 8575–8579.

Lafontaine, D. L. J. (2015). Noncoding RNAs in eukaryotic ribosome biogenesis and function. *Nature Publishing Group*, *22*(1), 11–19. https://doi.org/10.1038/nsmb.2939

Lahmy, S., Guilleminot, J., Cheng, C.-M., Bechtold, N., Albert, S., Pelletier, G., Delseny, M., & Devic, M. (2004). DOMINO1, a member of a small plant-specific gene family, encodes a protein essential for nuclear and nucleolar functions. *The Plant Journal: For Cell and Molecular Biology*, *39*(6), 809–820. https://doi.org/10.1111/j.1365-313X.2004.02166.x

Lange, T. S., & Gerbi, S. A. (2000). Transient nucleolar localization Of U6 small nuclear RNA in Xenopus Laevis oocytes. *Molecular Biology of the Cell*, *11*(7), 2419–2428. https://doi.org/10.1091/mbc.11.7.2419

Leamy, K. A., Yamagami, R., Yennawar, N. H., & Bevilacqua, P. C. (2019). Single-nucleotide control of tRNA folding cooperativity under near-cellular conditions. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(46), 23075–23082. https://doi.org/10.1073/pnas.1913418116

Lee, Y., Kindelberger, D. W., Lee, J. Y., McClennen, S., Chamberlain, J., & Engelke, D. R. (1997). Nuclear pre-tRNA terminal structure and RNase P recognition. *RNA (New York, N.Y.)*, *3*(2), 175–185.

Leontis, N., DaLio, A., Strobel, M., & Engelke, D. (1988). Effects of tRNA-intron structure on cleavage of precursor tRNAs by RNase P from Saccharomyces cerevisiae. *Nucleic Acids Research*, *16*(6), 2537–2552. https://doi.org/10.1093/nar/16.6.2537

Lerner, M. R., Boyle, J. A., Hardin, J. A., & Steitz, J. A. (1981). Two novel classes of small ribonucleoproteins detected by antibodies associated with lupus erythematosus. *Science (New York, N.Y.)*, *211*(4480), 400–402.

Lestrade, L., & Weber, M. J. (2006). SnoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Research*, *34*(Database issue), D158-162. https://doi.org/10.1093/nar/gkj002

Leung, A. K. W., Nagai, K., & Li, J. (2011). Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. *Nature*, *473*(7348), Article 7348. https://doi.org/10.1038/nature09956

Leung, E., Schneider, C., Yan, F., Mohi-El-Din, H., Kudla, G., Tuck, A., Wlotzka, W., Doronina, V. A., Bartley, R., Watkins, N. J., Tollervey, D., & Brown, J. D. (2014). Integrity of SRP RNA is ensured by La and the nuclear RNA quality control machinery. *Nucleic Acids Research*, *42*(16), 10698–10710. https://doi.org/10.1093/nar/gku761

Lewis, J. D., & Izaurflde, E. (1997). The Role of the Cap Structure in RNA Processing and Nuclear Export. *European Journal of Biochemistry*, *247*(2), 461–469. https://doi.org/10.1111/j.1432-1033.1997.00461.x

Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*(1), 323. https://doi.org/10.1186/1471-2105-12-323

Li, H. D., Zagorski, J., & Fournier, M. J. (1990). Depletion of U14 small nuclear RNA (snR128) disrupts production of 18S rRNA in Saccharomyces cerevisiae. *Molecular and Cellular Biology*, *10*(3), 1145–1152. https://doi.org/10.1128/mcb.10.3.1145-1152.1990

Lorenz, C., Lünse, C. E., & Mörl, M. (2017). tRNA Modifications: Impact on Structure and Thermal Adaptation. *Biomolecules*, *7*(2), 35. https://doi.org/10.3390/biom7020035

Lorsch, J. R. (2002). RNA Chaperones Exist and DEAD Box Proteins Get a Life. *Cell*, *109*(7), 797–800. https://doi.org/10.1016/S0092-8674(02)00804-8

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lowe, T. M., & Eddy, S. R. (1999). A computational screen for methylation guide snoRNAs in yeast. *Science (New York, N.Y.)*, *283*(5405), 1168–1171. https://doi.org/10.1126/science.283.5405.1168

Lukowiak, A. A., Granneman, S., Mattox, S. A., Speckmann, W. A., Jones, K., Pluk, H., van Venrooij, W. J., Terns, R. M., & Terns, M. P. (2000). Interaction of the U3-55k protein with U3 snoRNA is mediated by the Box B/C motif of U3 and the WD repeats of U3-55k. *Nucleic Acids Research*, *28*(18), 3462–3471.

Lund, E., & Dahlberg, J. E. (1992). Cyclic 2',3'-phosphates and nontemplated nucleotides at the 3' end of spliceosomal U6 small nuclear RNA's. *Science (New York, N.Y.)*, *255*(5042), 327–330. https://doi.org/10.1126/science.1549778

Makarova, O. V., Makarov, E. M., Liu, S., Vornlocher, H.-P., & Lührmann, R. (2002). Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6·U5 tri-snRNP formation and pre-mRNA splicing. *The EMBO Journal*, *21*(5), 1148–1157. https://doi.org/10.1093/emboj/21.5.1148

Maraia, R. J., & Bayfield, M. A. (2006). The La Protein-RNA complex surfaces. *Molecular Cell*, *21*(2), 149–152. https://doi.org/10.1016/j.molcel.2006.01.004

Maraia, R. J., & Intine, R. V. (2002). La protein and its associated small nuclear and nucleolar precursor RNAs. *Gene Expression*, *10*(1–2), 41–57.

Maraia, R. J., & Lamichhane, T. N. (2011). 3′ processing of eukaryotic precursor tRNAs. *Wiley Interdisciplinary Reviews. RNA*, *2*(3), 362–375. https://doi.org/10.1002/wrna.64

Maris, C., Dominguez, C., & Allain, F. H.-T. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *The FEBS Journal*, *272*(9), 2118–2131. https://doi.org/10.1111/j.1742-4658.2005.04653.x

Martin-Tumasz, S., Richie, A. C., Clos, L. J., Brow, D. A., & Butcher, S. E. (2011). A novel occluded RNA recognition motif in Prp24 unwinds the U6 RNA internal stem loop. *Nucleic Acids Research*, *39*(17), 7837–7847. https://doi.org/10.1093/nar/gkr455

Massenet, S., Motorin, Y., Lafontaine, D. L. J., Hurt, E. C., Grosjean, H., & Branlant, C. (1999). Pseudouridine Mapping in the *Saccharomyces cerevisiae* Spliceosomal U Small Nuclear RNAs (snRNAs) Reveals that Pseudouridine Synthase Pus1p Exhibits a Dual Substrate Specificity for U2 snRNA and tRNA. *Molecular and Cellular Biology*, *19*(3), 2142–2154. https://doi.org/10.1128/MCB.19.3.2142

Mathews, M. B., & Francoeur, A. M. (1984). La antigen recognizes and binds to the 3'-oligouridylate tail of a small RNA. *Molecular and Cellular Biology*, *4*(6), 1134–1140.

Mattaj, I. W. (1986). Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell*, *46*(6), 905–911. https://doi.org/10.1016/0092-8674(86)90072-3

Mattioli, M., & Reichlin, M. (1974). Heterogeneity of RNA protein antigens reactive with sera of patients with systemic lupus erythematosus. Description of a cytoplasmic nonribosomal antigen. *Arthritis and Rheumatism*, *17*(4), 421–429.

Mendoza-Figueroa, M. S., Tatomer, D. C., & Wilusz, J. E. (2020). The Integrator Complex in Transcription and Development. *Trends in Biochemical Sciences*, *45*(11), 923–934. https://doi.org/10.1016/j.tibs.2020.07.004

Merriam, E. V., & Bruns, P. J. (1988). Phenotypic Assortment in Tetrahymena Thermophila: Assortment Kinetics of Antibiotic-Resistance Markers, Tsa, Death, and the Highly Amplified Rdna Locus. *Genetics*, *120*(2), 389–395.

Mochizuki, K. (2008). High efficiency transformation of Tetrahymena using a codon-optimized neomycin resistance gene. *Gene*, *425*(1–2), 79–83. https://doi.org/10.1016/j.gene.2008.08.007

Moir, R. D., & Willis, I. M. (2013). Regulation of Pol III Transcription by Nutrient and Stress Signaling Pathways. *Biochimica et Biophysica Acta*, *1829*(3–4), 361–375. https://doi.org/10.1016/j.bbagrm.2012.11.001

Möllenbeck, M., Postberg, J., Paeschke, K., Rossbach, M., Jönsson, F., & Lipps, H. J. (2003). The telomerase-associated protein p43 is involved in anchoring telomerase in the nucleus. *Journal of Cell Science*, *116*(Pt 9), 1757–1761. https://doi.org/10.1242/jcs.00351

Monecke, T., Dickmanns, A., & Ficner, R. (2009). Structural basis for m7G-cap hypermethylation of small nuclear, small nucleolar and telomerase RNA by the dimethyltransferase TGS1. *Nucleic Acids Research*, *37*(12), 3865–3877. https://doi.org/10.1093/nar/gkp249

Montemayor, E. J., Didychuk, A. L., Yake, A. D., Sidhu, G. K., Brow, D. A., & Butcher, S. E. (2018). Architecture of the U6 snRNP reveals specific recognition of 3′-end processed U6 snRNA. *Nature Communications*, *9*(1), Article 1. https://doi.org/10.1038/s41467-018-04145-4

Motorin, Y., & Helm, M. (2010). TRNA Stabilization by Modified Nucleotides. *Biochemistry*, *49*(24), 4934–4944. https://doi.org/10.1021/bi100408z

Mouaikel, J., Bujnicki, J. M., Tazi, J., & Bordonné, R. (2003). Sequence–structure–function relationships of Tgs1, the yeast snRNA/snoRNA cap hypermethylase. *Nucleic Acids Research*, *31*(16), 4899–4909.

Mouaikel, J., Verheggen, C., Bertrand, E., Tazi, J., & Bordonné, R. (2002). Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus. *Molecular Cell*, *9*(4), 891–901. https://doi.org/10.1016/s1097-2765(02)00484-7

Muniz, L., Egloff, S., & Kiss, T. (2013). RNA elements directing in vivo assembly of the 7SK/MePCE/Larp7 transcriptional regulatory snRNP. *Nucleic Acids Research*, *41*(8), 4686–4698. https://doi.org/10.1093/nar/gkt159

Nabeel-Shah, S., Garg, J., Kougnassoukou Tchara, P.-E., Pearlman, R. E., Lambert, J.-P., & Fillingham, J. (2021). Functional proteomics protocol for the identification of interaction partners in Tetrahymena thermophila. *STAR Protocols*, *2*(1). https://doi.org/10.1016/j.xpro.2021.100362

Naeeni, A. R., Conte, M. R., & Bayfield, M. A. (2012). RNA Chaperone Activity of Human La Protein Is Mediated by Variant RNA Recognition Motif*. *Journal of Biological Chemistry*, *287*(8), 5472–5482. https://doi.org/10.1074/jbc.M111.276071

Nashimoto, M., Nashimoto, C., Tamura, M., Kaspar, R. L., & Ochi, K. (2001). The inhibitory effect of the autoantigen La on in vitro 3’ processing of mammalian precursor tRNAs. *Journal of Molecular Biology*, *312*(5), 975–984. https://doi.org/10.1006/jmbi.2001.5026

NCBI Resource Coordinators. (2016). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *44*(D1), D7-19. https://doi.org/10.1093/nar/gkv1290

Németh, A., Perez-Fernandez, J., Merkl, P., Hamperl, S., Gerber, J., Griesenbeck, J., & Tschochner, H. (2013). RNA polymerase I termination: Where is the end? *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1829*(3), 306–317. https://doi.org/10.1016/j.bbagrm.2012.10.007

Nguyen, T. H. D., Galej, W. P., Bai, X., Savva, C. G., Newman, A. J., Scheres, S. H. W., & Nagai, K. (2015). The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature*, *523*(7558), Article 7558. https://doi.org/10.1038/nature14548

Ni, J., Tien, A. L., & Fournier, M. J. (1997). Small Nucleolar RNAs Direct Site-Specific Synthesis of Pseudouridine in Ribosomal RNA. *Cell*, *89*(4), 565–573. https://doi.org/10.1016/S0092-8674(00)80238-X

Nielsen, H., Orum, H., & Engberg, J. (1992). A novel class of nucleolar RNAs from Tetrahymena. *FEBS Letters*, *307*(3), 337–342. https://doi.org/10.1016/0014-5793(92)80708-o

Nottingham, R. M., Wu, D. C., Qin, Y., Yao, J., Hunicke-Smith, S., & Lambowitz, A. M. (2016). RNA-seq of human reference RNA samples using a thermostable group II intron reverse transcriptase. *RNA (New York, N.Y.)*, *22*(4), 597–613. https://doi.org/10.1261/rna.055558.115

Novotný, I., Blažíková, M., Staněk, D., Herman, P., & Malinsky, J. (2011). In vivo kinetics of U4/U6·U5 tri-snRNP formation in Cajal bodies. *Molecular Biology of the Cell*, *22*(4), 513–523. https://doi.org/10.1091/mbc.E10-07-0560

Novotný, I., Malinová, A., Stejskalová, E., Matějů, D., Klimešová, K., Roithová, A., Švéda, M., Knejzlík, Z., & Staněk, D. (2015). SART3-Dependent Accumulation of Incomplete Spliceosomal snRNPs in Cajal Bodies. *Cell Reports*, *10*(3), 429–440. https://doi.org/10.1016/j.celrep.2014.12.030

O'Connor, C. M., & Collins, K. (2006). A Novel RNA Binding Domain in Tetrahymena Telomerase p65 Initiates Hierarchical Assembly of Telomerase Holoenzyme. *Molecular and Cellular Biology*, *26*(6), 2029–2036. https://doi.org/10.1128/MCB.26.6.2029-2036.2006

O'Donoghue, P., Ling, J., & Söll, D. (2018). Transfer RNA function and evolution. *RNA Biology*, *15*(4–5), 423–426. https://doi.org/10.1080/15476286.2018.1478942

Ohndorf, U.-M., Steegborn, C., Knijff, R., & Sondermann, P. (2001). Contributions of the Individual Domains in Human La Protein to Its RNA 3′-End Binding Activity *. *Journal of Biological Chemistry*, *276*(29), 27188–27196. https://doi.org/10.1074/jbc.M102891200

Ojha, S., Malla, S., & Lyons, S. M. (2020). snoRNPs: Functions in Ribosome Biogenesis. *Biomolecules*, *10*(5), 783. https://doi.org/10.3390/biom10050783

O'Reilly, D., Dienstbier, M., Cowley, S. A., Vazquez, P., Drozdz, M., Taylor, S., James, W. S., & Murphy, S. (2013). Differentially expressed, variant U1 snRNAs regulate gene expression in human cells. *Genome Research*, *23*(2), 281–291. https://doi.org/10.1101/gr.142968.112

Orias, E., Cervantes, M. D., & Hamilton, E. P. (2011). Tetrahymena thermophila , a unicellular eukaryote with separate germline and somatic genomes. *Research in Microbiologoy*, *162*(6), 578–586. https://doi.org/10.1016/j.resmic.2011.05.001

Orum, H., Nielsen, H., & Engberg, J. (1991). Spliceosomal small nuclear RNAs of Tetrahymena thermophila and some possible snRNA-snRNA base-pairing interactions. *Journal of Molecular Biology*, *222*(2), 219–232. https://doi.org/10.1016/0022-2836(91)90208-n

Pabis, M., Neufeld, N., Steiner, M. C., Bojic, T., Shav-Tal, Y., & Neugebauer, K. M. (2013). The nuclear cap-binding complex interacts with the U4/U6·U5 tri-snRNP and promotes spliceosome assembly in mammalian cells. *RNA (New York, N.Y.)*, *19*(8), 1054–1063. https://doi.org/10.1261/rna.037069.112

Páez-Moscoso, D. J., Pan, L., Sigauke, R. F., Schroeder, M. R., Tang, W., & Baumann, P. (2018). Pof8 is a La-related protein and a constitutive component of telomerase in fission yeast. *Nature Communications*, *9*(1), 587. https://doi.org/10.1038/s41467-017-02284-8

Pan, T. (2018). Modifications and functional genomics of human transfer RNA. *Cell Research*, *28*(4), Article 4. https://doi.org/10.1038/s41422-018-0013-y

Pang, Y. L. J., Poruri, K., & Martinis, S. A. (2014). tRNA synthetase: TRNA Aminoacylation and beyond. *Wiley Interdisciplinary Reviews. RNA*, *5*(4), 461–480. https://doi.org/10.1002/wrna.1224

Pannone, B. K., Kim, S. D., Noe, D. A., & Wolin, S. L. (2001). Multiple functional interactions between components of the Lsm2-Lsm8 complex, U6 snRNA, and the yeast La protein. *Genetics*, *158*(1), 187–196. https://doi.org/10.1093/genetics/158.1.187

Pannone, B. K., Xue, D., & Wolin, S. L. (1998). A role for the yeast La protein in U6 snRNP assembly: Evidence that the La protein is a molecular chaperone for RNA polymerase III transcripts. *The EMBO Journal*, *17*(24), 7442–7453. https://doi.org/10.1093/emboj/17.24.7442

Park, J.-M., Kohn, M. J., Bruinsma, M. W., Vech, C., Intine, R. V., Fuhrmann, S., Grinberg, A., Mukherjee, I., Love, P. E., Ko, M. S., DePamphilis, M. L., & Maraia, R. J. (2006). The Multifunctional RNA-Binding Protein La Is Required for Mouse Development and for the Establishment of Embryonic Stem Cells. *Molecular and Cellular Biology*, *26*(4), 1445–1451. https://doi.org/10.1128/MCB.26.4.1445-1451.2006

Paule, M. R., & White, R. J. (2000). Transcription by RNA polymerases I and III. *Nucleic Acids Research*, *28*(6), 1283–1298.

Peculis, B. A. (1997). The sequence of the 5’ end of the U8 small nucleolar RNA is critical for 5.8S and 28S rRNA maturation. *Molecular and Cellular Biology*, *17*(7), 3702–3713.

Pedersen, N., Hellung-Larsen, P., & Engberg, J. (1985). Small nuclear RNAs in the ciliate Tetrahymena. *Nucleic Acids Research*, *13*(11), 4203–4224. https://doi.org/10.1093/nar/13.11.4203

Pellizzoni, L. (2007). Chaperoning ribonucleoprotein biogenesis in health and disease. *EMBO Reports*, *8*(4), 340–345. https://doi.org/10.1038/sj.embor.7400941

Pellizzoni, L., Cardinali, B., Lin-Marq, N., Mercanti, D., & Pierandrei-Amaldi, P. (1996). A Xenopus laevis homologue of the La autoantigen binds the pyrimidine tract of the 5’ UTR

of ribosomal protein mRNAs in vitro: Implication of a protein factor in complex formation. *Journal of Molecular Biology*, *259*(5), 904–915. https://doi.org/10.1006/jmbi.1996.0368

Piekna-Przybylska, D., Decatur, W. A., & Fournier, M. J. (2007). New bioinformatic tools for analysis of nucleotide modifications in eukaryotic rRNA. *RNA*, *13*(3), 305–312. https://doi.org/10.1261/rna.373107

Pinkard, O., McFarland, S., Sweet, T., & Coller, J. (2020). Quantitative tRNA-sequencing uncovers metazoan tissue-specific tRNA regulation. *Nature Communications*, *11*, 4104. https://doi.org/10.1038/s41467-020-17879-x

Porat, J., & Bayfield, M. A. (2020). Use of tRNA-Mediated Suppression to Assess RNA Chaperone Function. *Methods in Molecular Biology (Clifton, N.J.)*, *2106*, 107–120. https://doi.org/10.1007/978-1-0716-0231-7_6

Porter, J. J., Heil, C. S., & Lueck, J. D. (2021). Therapeutic promise of engineered nonsense suppressor tRNAs. *Wiley Interdisciplinary Reviews. RNA*, *12*(4), e1641. https://doi.org/10.1002/wrna.1641

Prathapam, R., Witkin, K. L., O'Connor, C. M., & Collins, K. (2005). A telomerase holoenzyme protein enhances telomerase RNA assembly with telomerase reverse transcriptase. *Nature Structural & Molecular Biology*, *12*(3), 252–257. https://doi.org/10.1038/nsmb900

Rabl, J., Leibundgut, M., Ataide, S. F., Haag, A., & Ban, N. (2011). Crystal Structure of the Eukaryotic 40S Ribosomal Subunit in Complex with Initiation Factor 1. *Science*, *331*(6018), 730–736. https://doi.org/10.1126/science.1198308

Rader, S. D., & Guthrie, C. (2002). A conserved Lsm-interaction motif in Prp24 required for efficient U4/U6 di-snRNP formation. *RNA*, *8*(11), 1378–1392.

Raghunathan, P. L., & Guthrie, C. (1998). A Spliceosomal Recycling Factor That Reanneals U4 and U6 Small Nuclear Ribonucleoprotein Particles. *Science*, *279*(5352), 857–860. https://doi.org/10.1126/science.279.5352.857

Raker, V. A., Hartmuth, K., Kastner, B., & Lührmann, R. (1999). Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner. *Molecular and Cellular Biology*, *19*(10), 6554–6565. https://doi.org/10.1128/MCB.19.10.6554

Raker, V. A., Plessel, G., & Lührmann, R. (1996). The snRNP core assembly pathway: Identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. *The EMBO Journal*, *15*(9), 2256–2269.

Randau, L., Schröder, I., & Söll, D. (2008). Life without RNase P. *Nature*, *453*(7191), 120–123. https://doi.org/10.1038/nature06833

Rappsilber, J., Ryder, U., Lamond, A. I., & Mann, M. (2002). Large-Scale Proteomic Analysis of the Human Spliceosome. *Genome Research*, *12*(8), 1231–1245. https://doi.org/10.1101/gr.473902

Reddy, R., Henning, D., Das, G., Harless, M., & Wright, D. (1987). The capped U6 small nuclear RNA is transcribed by RNA polymerase III. *The Journal of Biological Chemistry*, *262*(1), 75–81.

Reuter, R., Appel, B., Bringmann, P., Rinke, J., & Lührmann, R. (1984). 5'-Terminal caps of snRNAs are reactive with antibodies specific for 2,2,7-trimethylguanosine in whole cells

and nuclear matrices. Double-label immunofluorescent studies with anti-m3G antibodies and with anti-RNP and anti-Sm autoantibodies. *Experimental Cell Research*, *154*(2), 548–560. https://doi.org/10.1016/0014-4827(84)90179-4

Richard, P., Darzacq, X., Bertrand, E., Jády, B. E., Verheggen, C., & Kiss, T. (2003). A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. *The EMBO Journal*, *22*(16), 4283–4293. https://doi.org/10.1093/emboj/cdg394

Richard, P., & Manley, J. L. (2009). Transcription termination by nuclear RNA polymerases. *Genes & Development*, *23*(11), 1247–1269. https://doi.org/10.1101/gad.1792809

Rinke, J., & Steitz, J. A. (1982). *Precursor Molecules of Both Human 5s Ribosomal RNA and Transfer RNAs Are Bound by a Cellular Protein Reactive with Anti-La Lupus Antibodies*. *29*(May), 149–159.

Rinke, J., & Steitz, J. A. (1985). Association of the lupus antigen La with a subset of U6 snRNA molecules. *Nucleic Acids Research*, *13*(7), 2617–2629. https://doi.org/10.1093/nar/13.7.2617

Rio, D. C., Ares, M., Hannon, G. J., & Nilsen, T. W. (2010). Purification of RNA Using TRIzol (TRI Reagent). *Cold Spring Harbor Protocols*, *2010*(6), pdb.prot5439. https://doi.org/10.1101/pdb.prot5439

Scherly, D., Stutz, F., Lin-Marq, N., & Clarkson, S. G. (1993). La proteins from Xenopus laevis. CDNA cloning and developmental expression. *Journal of Molecular Biology*, *231*(2), 196–204.

Seipelt, R. L., Zheng, B., Asuru, A., & Rymond, B. C. (1999). U1 snRNA is cleaved by RNase III and processed through an Sm site-dependent pathway. *Nucleic Acids Research*, *27*(2), 587–595. https://doi.org/10.1093/nar/27.2.587

Sheng, Y., Duan, L., Cheng, T., Qiao, Y., Stover, N. A., & Gao, S. (2020). The completed macronuclear genome of a model ciliate Tetrahymena thermophila and its application in genome scrambling and copy number analyses. *Science China. Life Sciences*, *63*(10), 1534–1542. https://doi.org/10.1007/s11427-020-1689-4

Shimba, S., & Reddy, R. (1994). Purification of human U6 small nuclear RNA capping enzyme. Evidence for a common capping enzyme for gamma-monomethyl-capped small RNAs. *The Journal of Biological Chemistry*, *269*(17), 12419–12423.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*, 539. https://doi.org/10.1038/msb.2011.75

Simons, F. H., Rutjes, S. A., van Venrooij, W. J., & Pruijn, G. J. (1996). The interactions with Ro60 and La differentially affect nuclear export of hY1 RNA. *RNA*, *2*(3), 264–273.

Singh, M., Wang, Z., Koo, B.-K., Patel, A., Cascio, D., Collins, K., & Feigon, J. (2012). Structural basis for telomerase RNA recognition and RNP assembly by the holoenzyme La family protein p65. *Molecular Cell*, *47*(1), 16–26. https://doi.org/10.1016/j.molcel.2012.05.018

Singh, R., & Reddy, R. (1989). Gamma-monomethyl phosphate: A cap structure in spliceosomal U6 small nuclear RNA. *Proceedings of the National Academy of Sciences*, *86*(21), 8280–8283. https://doi.org/10.1073/pnas.86.21.8280

Smith, C. M., & Steitz, J. A. (1998). Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5'-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. *Molecular and Cellular Biology*, *18*(12), 6897–6909. https://doi.org/10.1128/MCB.18.12.6897

Smith, T. S., Zoltek, M. A., & Simon, M. D. (2019). Reengineering a tRNA Methyltransferase To Covalently Capture New RNA Substrates. *Journal of the American Chemical Society*, *141*(44), 17460–17465. https://doi.org/10.1021/jacs.9b08529

Sobel, S. G., & Wolin, S. L. (1999). Two yeast La motif-containing proteins are RNA-binding proteins that associate with polyribosomes. *Molecular Biology of the Cell*, *10*(11), 3849–3862. https://doi.org/10.1091/mbc.10.11.3849

Spiller, M. P., Boon, K.-L., Reijns, M. A. M., & Beggs, J. D. (2007). The Lsm2-8 complex determines nuclear localization of the spliceosomal U6 snRNA. *Nucleic Acids Research*, *35*(3), 923–929. https://doi.org/10.1093/nar/gkl1130

Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A., & Steinberg, S. (1998). Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Research*, *26*(1), 148–153. https://doi.org/10.1093/nar/26.1.148

Staněk, D., & Neugebauer, K. M. (2004). Detection of snRNP assembly intermediates in Cajal bodies by fluorescence resonance energy transfer. *The Journal of Cell Biology*, *166*(7), 1015–1025. https://doi.org/10.1083/jcb.200405160

Stefano, J. E. (1984). Purified lupus antigen La recognizes an oligouridylate stretch common to the 3' termini of RNA polymerase III transcripts. *Cell*, *36*(1), 145–154.

Steinmetz, E. J., & Brow, D. A. (1996). Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. *Molecular and Cellular Biology*, *16*(12), 6993–7003. https://doi.org/10.1128/MCB.16.12.6993

Steinmetz, E. J., Conrad, N. K., Brow, D. A., & Corden, J. L. (2001). RNA-binding protein Nrd1 directs poly(A)-independent 3′-end formation of RNA polymerase II transcripts. *Nature*, *413*(6853), Article 6853. https://doi.org/10.1038/35095090

Steitz, J. A., & Tycowski, K. T. (1995). Small RNA chaperones for ribosome biogenesis. *Science (New York, N.Y.)*, *270*(5242), 1626–1627. https://doi.org/10.1126/science.270.5242.1626

Stover, N. A., Punia, R. S., Bowen, M. S., Dolins, S. B., & Clark, T. G. (2012). Tetrahymena genome database Wiki: A community-maintained model organism database. *Database*, *2012*, bas007. https://doi.org/10.1093/database/bas007

Suraweera, A., Lim, Y., Woods, R., Birrell, G. W., Nasim, T., Becherel, O. J., & Lavin, M. F. (2009). Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation. *Human Molecular Genetics*, *18*(18), 3384–3396. https://doi.org/10.1093/hmg/ddp278

Tanackovic, G., Ransijn, A., Thibault, P., Abou Elela, S., Klinck, R., Berson, E. L., Chabot, B., & Rivolta, C. (2011). PRPF mutations are associated with generalized defects in spliceosome formation and pre-mRNA splicing in patients with retinitis pigmentosa. *Human Molecular Genetics*, *20*(11), 2116–2130. https://doi.org/10.1093/hmg/ddr094

Teplova, M., Yuan, Y.-R., Phan, A. T., Malinina, L., Ilin, S., Teplov, A., & Patel, D. J. (2006). Structural Basis for Recognition and Sequestration of UUUOH 3′ Temini of Nascent RNA Polymerase III Transcripts by La, a Rheumatic Disease Autoantigen. *Molecular Cell*, *21*(1), 75–85. https://doi.org/10.1016/j.molcel.2005.10.027

Terns, M. P., Lund, E., & Dahlberg, J. E. (1992). 3'-end-dependent formation of U6 small nuclear ribonucleoprotein particles in Xenopus laevis oocyte nuclei. *Molecular and Cellular Biology*, *12*(7), 3032–3040. https://doi.org/10.1128/mcb.12.7.3032-3040.1992

Tian, M., Yang, W., Zhang, J., Dang, H., Lu, X., Fu, C., & Miao, W. (2017). Nonsense-mediated mRNA decay in Tetrahymena is EJC independent and requires a protozoa-specific nuclease. *Nucleic Acids Research*, *45*(11), 6848–6863. https://doi.org/10.1093/nar/gkx256

Tollervey, D., & Kiss, T. (1997). Function and synthesis of small nucleolar RNAs. *Current Opinion in Cell Biology*, *9*(3), 337–342. https://doi.org/10.1016/S0955-0674(97)80005-1

Tomita, K., & Yamashita, S. (2014). Molecular mechanisms of template-independent RNA polymerization by tRNA nucleotidyltransferases. *Frontiers in Genetics*, *5*. https://www.frontiersin.org/article/10.3389/fgene.2014.00036

True, H. L., & Celander, D. W. (1996). *Ribonuclease P of Tetrahymena thermophila *. 271*(28), 16559–16566.

Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., & Pilpel, Y. (2010). An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation. *Cell*, *141*(2), 344–354. https://doi.org/10.1016/j.cell.2010.03.031

Turowski, T. W., & Tollervey, D. (2016). Transcription by RNA polymerase III: Insights into mechanism and regulation. *Biochemical Society Transactions*, *44*(5), 1367–1375. https://doi.org/10.1042/BST20160062

Tycowski, K. T., Shu, M. D., & Steitz, J. A. (1993). A small nucleolar RNA is processed from an intron of the human gene encoding ribosomal protein S3. *Genes & Development*, *7*(7a), 1176–1190. https://doi.org/10.1101/gad.7.7a.1176

Tycowski, K. T., Shu, M. D., & Steitz, J. A. (1994). Requirement for intron-encoded U22 small nucleolar RNA in 18S ribosomal RNA maturation. *Science (New York, N.Y.)*, *266*(5190), 1558–1561. https://doi.org/10.1126/science.7985025

Uchikawa, E., Natchiar, K. S., Han, X., Proux, F., Roblin, P., Zhang, E., Durand, A., Klaholz, B. P., & Dock-Bregeon, A.-C. (2015). Structural insight into the mechanism of stabilization of the 7SK small nuclear RNA by LARP7. *Nucleic Acids Research*, *43*(6), 3373–3388. https://doi.org/10.1093/nar/gkv173

Uguen, P., & Murphy, S. (2003). The 3′ ends of human pre-snRNAs are produced by RNA polymerase II CTD-dependent RNA processing. *The EMBO Journal*, *22*(17), 4544–4554. https://doi.org/10.1093/emboj/cdg430

Vakiloroayaei, A., Shah, N. S., Oeffinger, M., & Bayfield, M. A. (2017). The RNA chaperone La promotes pre-TRNA maturation via indiscriminate binding of both native and misfolded targets. *Nucleic Acids Research*, *45*(19), 11341–11355. https://doi.org/10.1093/nar/gkx764

Valadkhan, S., & Manley, J. L. (2003). Characterization of the catalytic activity of U2 and U6 snRNAs. *RNA (New York, N.Y.)*, *9*(7), 892–904. https://doi.org/10.1261/rna.5440303

van Hoof, A., Lennertz, P., & Parker, R. (2000). Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. *Molecular and Cellular Biology*, *20*(2), 441–452. https://doi.org/10.1128/MCB.20.2.441-452.2000

Van Horn, D. J., Yoo, C. J., Xue, D., Shi, H., & Wolin, S. L. (1997). The La protein in Schizosaccharomyces pombe: A conserved yet dispensable phosphoprotein that functions in tRNA maturation. *RNA*, *3*(12), 1434–1443.

Venema, J., Bousquet-Antonelli, C., Gelugne, J. P., Caizergues-Ferrer, M., & Tollervey, D. (1997). Rok1p is a putative RNA helicase required for rRNA processing. *Molecular and Cellular Biology*, *17*(6), 3398–3407.

Villa, T., Pleiss, J. A., & Guthrie, C. (2002). Spliceosomal snRNAs: Mg2+-Dependent Chemistry at the Catalytic Core? *Cell*, *109*(2), 149–152. https://doi.org/10.1016/S0092-8674(02)00726-2

Vinayak, J., Marrella, S. A., Hussain, R. H., Rozenfeld, L., Solomon, K., & Bayfield, M. A. (2018). Human La binds mRNAs through contacts to the poly(A) tail. *Nucleic Acids Research*, *46*(8), 4228–4240. https://doi.org/10.1093/nar/gky090

Wang, Y., & Ma, H. (2015). Step-wise and lineage-specific diversification of plant RNA polymerase genes and origin of the largest plant-specific subunits. *The New Phytologist*, *207*(4), 1198–1212. https://doi.org/10.1111/nph.13432

Weisser, M., Voigts-Hoffmann, F., Rabl, J., Leibundgut, M., & Ban, N. (2013). The crystal structure of the eukaryotic 40S ribosomal subunit in complex with eIF1 and eIF1A. *Nature Structural & Molecular Biology*, *20*(8), Article 8. https://doi.org/10.1038/nsmb.2622

Wenkert, D., & Allis, C. D. (1984). Timing of the appearance of macronuclear-specific histone variant hv1 and gene expression in developing new macronuclei of Tetrahymena thermophila. *The Journal of Cell Biology*, *98*(6), 2107–2117. https://doi.org/10.1083/jcb.98.6.2107

Will, C. L., & Lührmann, R. (2011). Spliceosome Structure and Function. *Cold Spring Harbor Perspectives in Biology*, *3*(7), a003707. https://doi.org/10.1101/cshperspect.a003707

Wu, J., Niu, S., Tan, M., Huang, C., Li, M., Song, Y., Wang, Q., Chen, J., Shi, S., Lan, P., & Lei, M. (2018). Cryo-EM Structure of the Human Ribonuclease P Holoenzyme. *Cell*, *175*(5), 1393-1404.e11. https://doi.org/10.1016/j.cell.2018.10.003

Xiong, J., Lu, X., Zhou, Z., Chang, Y., Yuan, D., Tian, M., Zhou, Z., Fu, C., Orias, E., & Miao, W. (2012). *Transcriptome Analysis of the Model Protozoan, Tetrahymena thermophila , Using Deep RNA Sequencing*. *7*(2), 1–13. https://doi.org/10.1371/journal.pone.0030630

Xue, D., Rubinson, D. A., Pannone, B. K., Yoo, C. J., & Wolin, S. L. (2000). U snRNP assembly in yeast involves the La protein. *The EMBO Journal*, *19*(7), 1650–1660. https://doi.org/10.1093/emboj/19.7.1650

Yan, B., Tzertzinis, G., Schildkraut, I., & Ettwiller, L. (2022). Comprehensive determination of transcription start sites derived from all RNA polymerases using ReCappable-seq. *Genome Research*, *32*(1), 162–174. https://doi.org/10.1101/gr.275784.121

Yang, C., McPheeters, D. S., & Yu, Y.-T. (2005). Psi35 in the branch site recognition region of U2 small nuclear RNA is important for pre-mRNA splicing in Saccharomyces cerevisiae. *The Journal of Biological Chemistry*, *280*(8), 6655–6662. https://doi.org/10.1074/jbc.M413288200

Yoo, C. J., & Wolin, S. L. (1994). La proteins from Drosophila melanogaster and Saccharomyces cerevisiae: A yeast homolog of the La autoantigen is dispensable for growth. *Molecular and Cellular Biology*, *14*(8), 5412–5424.

Yoo, C. J., & Wolin, S. L. (1997). *The yeast La protein is required for the 3' endonucleolytic cleavage that matures tRNA precursors*. *89*, 393–402.

Yu, Y. T., Shu, M. D., & Steitz, J. A. (1997). A new method for detecting sites of 2'-O-methylation in RNA molecules. *RNA*, *3*(3), 324–331.

Yu, Y. T., Shu, M. D., & Steitz, J. A. (1998). Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *The EMBO Journal*, *17*(19), 5783–5795. https://doi.org/10.1093/emboj/17.19.5783

Yuo, C. Y., Ares, M., & Weiner, A. M. (1985). Sequences required for 3′ end formation of human U2 small nuclear RNA. *Cell*, *42*(1), 193–202. https://doi.org/10.1016/S0092-8674(85)80115-X

Zhang, P., Wu, W., Chen, Q., & Chen, M. (2019). Non-Coding RNAs and their Integrated Networks. *Journal of Integrative Bioinformatics*, *16*(3), 20190027. https://doi.org/10.1515/jib-2019-0027

Zheng, G., Qin, Y., Clark, W. C., Dai, Q., Yi, C., He, C., Lambowitz, A. M., & Pan, T. (2015). Efficient and quantitative high-throughput tRNA sequencing. *Nature Methods*, *12*(9), Article 9. https://doi.org/10.1038/nmeth.3478

Zheng, W., Zhang, C., Wuyun, Q., Pearce, R., Li, Y., & Zhang, Y. (2019). LOMETS2: Improved meta-threading server for fold-recognition and structure-based function annotation for distant-homology proteins. *Nucleic Acids Research*, *47*(W1), W429–W436. https://doi.org/10.1093/nar/gkz384