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[2]; Pearlman, Ronald E.[1]; Bayfield, Mark A.[1*]

[1] Department of Biology, Faculty of Science, York University, Toronto, Ontario M3J 1P3, Canada,

[2] Département de Biochimie et de Génomique Fonctionnelle, 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.

[*]Correspondence: bayfield@yorku.ca (M.A.B.)

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 classified as a genuine La protein, despite the predicted absence of the RRM1. We found that Mlp1 functions as a La protein through binding of pre-tRNAs and affecting processing in *Tetrahymena thermophila* and when expressed in fission yeast. However, unlike in other examined eukaryotes, depletion of Mlp1 results in 3'-trailer

stabilization. We also observed that 3'-trailers in *Tetrahymena thermophila* are uniquely short relative to other examined eukaryotes, and that 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*.

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 and assists with pre-tRNA folding[1]. 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[2,3]. 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[4]. 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 before RNase P endonucleolytic cleavage of the 5'-leader[5], and misfolded pre-tRNAs are also more prone to degradation through nuclear surveillance[6]. Thus, La binding to pre-tRNAs is hypothesized to establish and determine the order of 5'-leader versus 3'-trailer processing[3]. Once the 5'-leader and 3'-trailer sequences are processed, the 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[7].

Genuine La proteins are members of the La-related proteins (LARPs)[8]. 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 1A**)[9–11]. In *Tetrahymena thermophila,* the LARP7 ortholog p65 was until recently the only characterized LARP in this species[8,12], but a recent study has grouped the *Tetrahymena* Macronucleus localized protein of unknown function (Mlp1)[13] with the genuine La proteins, based on its primary sequence conservation in the LaM[12]. 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[12].

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 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[9,14], as deletion of either the LaM or RRM1 results in complete loss of binding[9,10,15,16]. 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[12]. 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.

Here, we present evidence that despite the apparent lack of the RRM1, Mlp1 functions as a genuine La protein. Using ribonucleoprotein 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*. Furthermore, we demonstrate that Mlp1 expression in *Schizosaccharomyces pombe* promotes pre-tRNA processing and tRNA mediated suppression but without typical La-associated 3'-end protection. Genetic depletion of Mlp1 in *Tetrahymena* 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 *Tetrahymena* 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. Together, 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 *Tetrahymena* that differs from those found in other studied eukaryotic systems.

Results

Mlp1 is predicted to lack an RRM adjacent to the LaM

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

Tetrahymena (**Figure 1B, Figure S1A,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 (**Figure S1B**).

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[8]. Therefore, we compared the secondary structure predictions for *Tetrahymena* La with secondary structures for different eukaryotes (**Figure 1C**). These consistently predicted an RRM fold immediately C-terminal to the LaM in all species examined, except for candidate alveolate La proteins.

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[2,3]. We immunoprecipitated Mlp1-associated RNAs followed by detection by Northern blot (**Figure 2A, Figure S2A**). 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*[17]. We conclude that based on its expected RNA target cohort, Mlp1 behaves as expected for a genuine La protein.

To investigate Mlp1-associated RNAs more extensively, we sequenced Mlp1-immunoprecipitated RNAs after removing larger 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 2B** – Mlp1). In contrast, mature tRNAs are not enriched in Mlp1-bound samples (**Figure 2B** – Mlp1; averaged enrichment values for tRNA isotypes presented in **Figure S2B**). 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[18] and found a similar enrichment for 3'-uridylate containing pre-tRNAs (**Figure 2B** – hLa; averaged enrichment values for tRNA isotypes presented in **Figure S2C**). The relative enrichment of pre-tRNAs over mature tRNAs, are strongly indicative of Mlp1 functioning as a genuine La protein in *Tetrahymena*.

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[4,19]. 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 electromobility shift assays (EMSAs), as well as versions of these that included either or both of 5'-leader or 3'-trailer extensions (**Figure 2C**). We found that Mlp1 preferentially binds pre-tRNAs

containing both 5'- and 3'-extensions, followed by pre-tRNA containing the 3'-trailer (**Figure S2D,E, Table S1**). Lowest binding affinities were consistently found for mature tRNAs, confirming the *in vivo* preferential binding of pre-tRNAs (see **Figure S2D,E, Table S1**).

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 2D-I, Figure S3A, Table 1**). To test the importance of 3'-uridylates 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)[10,16] (see **Figure 1B**, **Figure S1D**, 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 2E and F**) as well as for the LaM deletion mutant (compare **Figure 2E 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 2H**), 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 2I**). 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 we found that the predicted uridylate binding residues located in the LaM (Q11A/Y14A) were required for binding (**Figure 2J**). Additionally, binding of 3'-uridylate RNA occurs with the same affinity for Mlp1 1-250 compared to full length Mlp1, whereas affinity for uridylate RNA is lost when both domains are used individually (Mlp1 1-95 and Mlp1 95-250) (**Figure**

157 **S3B**). These data suggest that despite the apparent lack of the RRM1, Mlp1 functions in
158 a similar manner as the hLa protein in preferentially binding pre-tRNA substrates, and
159 that predicted, conserved uridylate binding residues in the LaM promote higher affinity
160 binding associated with the UUU-3'OH motif.

161 The hLa protein is known to also interact with the 5'-triphosphate containing end of the
162 nascent pre-tRNA through a short basic motif located in the C-terminal part of the
163 protein[19]. We tested the affinity of the 5'-triphosphate for Mlp1 by competition EMSA,
164 using a 5'-leader containing, 3'-trailer processed *in vitro* transcribed radioactively
165 labeled tRNA (+5'PPP) and unlabelled competitor tRNAs with the 5'-triphosphate
166 removed after phosphatase treatment (-5'PPP; +/- 5'PPP demonstrated in **Figure S3C**).
167 We found that the unlabelled –5'PPP substrate competed poorly relative to an
168 unlabelled +5'PPP substrate for the radioactive +5'PPP substrate on hLa, (**Figure 2K** –
169 compare -5'PPP and +5'PPP hLa, **Figure S3D**) but the difference in competition
170 between these RNAs on Mlp1 was smaller (**Figure 2K** – compare -5'PPP and +5'PPP
171 Mlp1, **Figure S3E**). These data are consistent with a lesser degree of 5'-leader
172 discrimination for 5'PPP in Mlp1 relative to hLa.

173

174 **Mlp1 binding to 3'-trailer RNAs is different from hLa**

175 To further study the distinct binding modes, we performed competition experiments
176 between ^{32}P-labeled uridylate RNA (U10) and unlabelled competitor pre-tRNA and
177 mature tRNA substrates. We used hLa as a control and found that, as expected from
178 previous work[4], pre-tRNAs were a stronger competitor for the uridylate binding pocket

than mature tRNAs (**Figure 3A, Figure S4A**). In contrast, uridylate RNA was competed off of Mlp1 using only low amounts of either pre-tRNA or mature tRNA (**Figure 3B, Figure S4A**), suggesting that binding of uridylates on Mlp1 is weaker compared to hLa.

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[14,15]. Previous high-resolution structural characterization of human La 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[15,16]. Notably, this uridylate is the only residue in the UUU-3'OH motif that contacts the RRM1 in hLa (**Figure 3C**). 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 variations to this sequence. 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 3D, Figure S4B**). Mutating the four last positions (4C – CUGCUGCCCC-3'OH) resulted in a total inability to compete with the wild type 4U RNA (**Figure 3D, Figure S4B**). 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 3E,**

Figure S4C). 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 for high affinity binding in the hydrophobic binding pocket between LaM and RRM1[15,16], with two hydrogen bonds formed with an aspartate (**Figure 3F**) that is conserved in Mlp1 (see **Figure 1B, Figure S1D**). 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 3G,H, Figure S4D**). 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.

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

In addition to 3'-end protection of nascent pre-tRNAs from exonucleases, La proteins also possess RNA chaperone activity to enhance correct folding of nascent pre-

tRNAs[4,20,21]. We used a tRNA-mediated suppression assay to investigate the ability of Mlp1 to rescue a misfolded suppressor tRNA[22]. We transformed Sla1p, full-length Mlp1 and multiple Mlp1 mutants into a *sla1- Schizosaccharomyces 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 4A**). 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[6]. 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 4A**, **Figure S5A**), 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 4A**), and combination of DUF3223 removal and uridylate binding inactivation (Mlp1 1-250 Q11A/Y14A) resulted in a complete loss of suppression (**Figure 4A**). These results indicate that to obtain complete tRNA mediated suppression,

246 both the conserved uridylate binding residues (Q11 and Y14) in the LaM and the C-
247 terminal DUF3223 are required.

248 To investigate 3'-end protection more directly, we extracted total RNA from Mlp1- and
249 Sla1p-expressing strains and analysed suppressor pre-tRNA Ser^{UCA} processing by
250 Northern blot (**Figure 4B**). We found that higher levels of suppressor pre-tRNA
251 stabilization in Sla1p- and Mlp1-transformed strains (lanes 2, 3) corresponded to more
252 mature suppressor tRNA which correlates with the white phenotype observed in the
253 tRNA-mediated suppression results (**Figure 4A**), as opposed to the lack of pre-tRNA
254 stabilization and subsequent mature suppressor tRNA in pRep4 control (lane 1)
255 resulting in a red phenotype (**Figure 4A,B**). Interestingly, the most abundant pre-tRNA
256 species was slightly smaller in Mlp1 transformed cells relative to Sla1p (see asterisk).
257 We also detected endogenous pre-tRNAs Lys^{CUU} processing intermediates using an
258 intron, 5'-leader and a 3'-trailer probe (**Figure 4C**). The presence of the three La-
259 dependent pathway pre-tRNA intermediates are detected after transformation of the
260 positive control Sla1p[23] (**Figure 4C** – intron probe, compare lanes 1 and 2). Notably, we
261 observed the appearance of a distinct pre-tRNA intermediate in Sla1 expressing cells
262 when using a probe specific for the 5'-leader of pre-tRNA Lys^{CUU} that did not co-migrate
263 with any of the major species observed using the intron or 3'-trailer probe (**Figure 4C** –
264 5'-leader, indicated by a double asterisk, beneath the full length pre-tRNA band). We
265 hypothesize that this species represents a subset of pre-tRNAs that are not stably
266 bound by Sla1p and whose 3'-ends have been nibbled by a 3'-exonuclease, yet have
267 retained their 5'-leaders as has been described during La-independent pre-tRNA
268 processing. As expected from the tRNA-mediated suppression assay, we found that

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 4C** – 5'-leader, 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 *Schizosaccharomyces 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 *Schizosaccharomyces pombe*, we immunoprecipitated Sla1p and Mlp1 ribonucleoprotein-complexes (**Figure S5B**) 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[24], Mlp1-associated pre-tRNAs were largely depleted for 3'-trailer containing species (**Figure 4D**), consistent with the Northern blotting results (**Figure 4B,C**). These data are consistent with our Mlp1 RNA binding data *in vitro*, indicating that Mlp1 promotes pre-tRNA processing but does not offer as strong protection of the 3'-ends of associated pre-tRNAs as other examined genuine La proteins.

Reduced Mlp1 expression in *Tetrahymena thermophila* leads to impaired 3'-trailer processing

To investigate the effect of Mlp1 depletion on pre-tRNA processing, we generated a partial *MLP1* knockout strain and confirmed genomic integration of the selection marker by Southern blot and PCR (**Figure S6A,B** and data not shown) and reduced protein expression of Mlp1 by Western blot (**Figure 5A**) and indirect immunofluorescent staining (**Figure S6C**). Notably, complete deletion of the *MLP1* locus in *Tetrahymena* macronuclei through increasing drug selection (phenotypic assortment) was not achieved, indicating that Mlp1 is likely essential in this system, similar to *Mus musculus*, *Drosophila melanogaster*, *Trypanosoma brucei* and *Arabidopsis thaliana*[25–28]. We extracted total RNA followed by pre-tRNA intermediate detection by Northern blot for pre-tRNA Ile^{UAU}, Leu^{UAA} and Val^{AAC}. We found that in absence of Mlp1, a 3'-trailer containing pre-tRNA intermediate was stabilized (**Figure 5B** – intron probe, bottom band). These data provide evidence for *Tetrahymena* 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 *Schizosaccharomyces 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.

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 2B**), 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 5C,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[29,30]. We observed that in *Tetrahymena* the number of genomic tRNA genes correlates with the expression levels of tRNAs (R^2 = 0.55, r = 0.75) (**Figure 5E**). 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 5F**) and found that when comparing tRNA expression between WT and partial KO strains the correlation was almost perfect (R^2 = 0.94, r = 0.97) (**Figure 5G**), indicating that Mlp1 depletion did not influence mature tRNA expression. Northern blot analysis for representative tRNAs confirmed equal amounts of mature tRNA between WT and partial KO strains (**Figure 5H**).

The 3'-trailer lengths are shorter in *Tetrahymena thermophila* compared to other examined eukaryotes

To explore tRNAs and pre-tRNA processing in *Tetrahymena* more extensively, we compared their predicted pre-tRNA architecture to those from other eukaryotic species. First, 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[31] that give rise to the 3'UUU-OH motif in the nascent transcript. We found that *Tetrahymena* has shorter 3'-trailer lengths than any other species analyzed, with the most common 3'-trailer length being zero nucleotides (**Figure 6A, Table 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 6B, Table 2**). We previously noted differences in enrichment of pre-tRNAs isotypes (**Figure S2B** – 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 = -0.38) (**Figure S7A**), 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 = -0.22) (**Figure S7B**).

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[32,33]. Previous studies demonstrated structural conservation of the catalytically active RNA in RNase P in *Tetrahymena*, indicating that processing occurs in an similar manner[33,34]. 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 6C**), as extensive base pairing between 5'-leader and 3'-trailer sequences inhibits 5'-leader cleavage[35]. We hypothesized that *Tetrahymena* 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

Tetrahymena thermophila, Schizosaccharomyces pombe and *Homo sapiens* and found that the most frequent nucleotide at the most 3'-residue of the 5'-leader sequence (N_{-1}) in *Tetrahymena* is an adenine (± 75%) (**Figure 6D**, **Figure S7C**). In contrast, the distribution of nucleotides at this position is more diverse in other species (**Figure 6D, Figure S7C**). 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%) (**Figure S7D**) 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 6E**, **Figure S7E**), which can pair with the first nucleotide in the 5'-leader (N_{-1}). We observed a strong lack of Watson–Crick base pairing for *Tetrahymena* 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 *Schizosaccharomyces pombe* and *Homo sapiens* (**Figure 6E**). Since shortened 3'-trailers in *Tetrahymena* 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 *Tetrahymena* (**Figure S7F**) relative to other examined species.

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[12] and our continued computational analysis have indicated that the previously characterized *Tetrahymena* 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.

In this work 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 seems 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. Using competition experiments, we

found 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 only makes 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[20,21]. Using tRNA-mediated suppression in a La null strain (*sla1-)* of *Schizosaccharomyces 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.

However, genuine La proteins are 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 *Schizosaccharomyces 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 *Tetrahymena* and performed RNA-Seq and Northern blots of endogenous tRNA species. We found that, consistent with our Mlp1 data from *Schizosaccharomyces 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}) (**Figure 7**).

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[36]. 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 to not be blocked. Thus, the apparent destabilization of 3'-trailers by Mlp1 in *Tetrahymena* 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 7**). 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 *Tetrahymena* (**Figure S8A,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[37,38]. 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 *Tetrahymena* appears to be conserved in *Tetrahymena*, indicating similar functionality (**Figure S8C,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 *Tetrahymena*, similar to budding and fission yeast.

Along with the alternate Mlp1-associated tRNA processing in *Tetrahymena*, 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 *Tetrahymena* pre-tRNAs have very short 3'-trailer sequences. Since 3'-trailers in *Tetrahymena* 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 *Tetrahymena* 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. Thus, 5'-

leaders and 3'-trailers in *Tetrahymena* may 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.

Materials and 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 **Table S6** and primary amino acid alignments were obtained using Clustal Omega (EMBL-EBI)[39], 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: (1) Asp (D), Glu (E), Asn (N), Gln (Q); (2) Lys (K), Arg (R), His (H); (3) Phe (F), Trp (W), Tyr (Y); (4) Val (V), Ile (I), Leu (L), Met (M); and (5) Ser (S), Thr (T).

Secondary structure predictions were obtained using Phyre2[40] and color coded based on predicted β-sheet or α-helices. High-resolution structures of the LaM of *Homo sapiens, Trypanosoma brucei,* and *Dictyostelium discoideum* La proteins were obtained from PDB: 2VOD, 1S29, and 2M5W, respectively. Structure of the La module (LaM+RRM1) in complex with uridylate RNA of *Homo sapiens* La was obtained from PDB: 2VOD. Structure prediction of the LaM of *Tetrahymena thermophila* was obtained using the lomets2 tool[41]. Structure of the LSm2-8 complex in complex with uridylate RNA of *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* were obtained from PDB: 6PNN and 4M7D, respectively. Modeling was done using PyMOL.

DNA constructs

The *MLP1* coding sequence encoding *Tetrahymena thermophila* La protein (NCBI Reference Sequence: XP_001019287.2) was obtained from the *Tetrahymena* Genome Database http://www.ciliate.org/[42], and codon optimized for expression in *E. coli*. gBlocks of the optimized codon sequence were ordered from integrated DNA technologies (IDT) (sequence can be found in **Table S5**).

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 6XHis-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 *Schizosaccharomyces 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 **Table S5.**

T7 DNA templates for *in vitro* transcription were generated by PCR amplification using pre- or mature tRNA specific primers (listed in **Table S5**) 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 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)[43]. Cell pellets were harvested by centrifugation for 3 minutes at 1000g. Following two washes in 10 mM Tris-HCl, pH 7.4, pellets were stored at -80°C.

PCR-amplification from *Tetrahymena 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 (http://www.ciliate.org/ - TTHERM_00384860) were cloned into pNeo4 plasmid[44] 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

539 construct was linearized using the *ScaI* restriction enzyme prior to transformation.
540 Biolistic transformation of *Tetrahymena thermophila* was performed as described
541 previously[45]. Integration of the DNA construct is based on homologous recombination
542 and transformants were grown under increasing concentration of paromomycin starting
543 at 60 µg/mL to a final concentration of 1000 µg/mL (phenotypic assortment)[46]. Correct
544 integration was determined using Southern blot and PCR.

545

546 Protein isolation from *Tetrahymena thermophila*

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

554

555 RNA-immunoprecipitation from *Tetrahymena thermophila*

556 *Tetrahymena thermophila* cell pellets from 100 mL cultures were collected at log phase
557 (0.1 - 1 x 10^6 cells/mL) and washed twice with 1X PBS. The cells were cross-linked with
558 1% formaldehyde for 10 minutes followed by quenching with 0.25 M glycine for 5
559 minutes at room temperature. Cells were washed twice in 1X PBS before lysis in 2 mL
560 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). 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. 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. 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 at 4°C. The aqueous layer obtained by centrifugation at 20,000g for 10 minutes at 4°C was ethanol precipitated overnight at -80°C. The cell 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[47,48]. Libraries were sequenced on a NextSeq500 platform (Illumina) (2x 75 bp).

tRNA-Seq processing pipeline

Details about tools and parameters of bioinformatics analyses are regrouped 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 XXX. Briefly, paired-end reads were trimmed using Trimmomatic v0.36[49], and FastQC v0.11.5 was used to evaluate read quality before and after trimming, as described previously[47]. Resulting trimmed reads were then aligned stringently (*i.e.* allowing no mismatch) to the *Tetrahymena thermophila* genome assembly version T_Thermophila_MAC_2021[50] (accessible on the Tetrahymena Genome Database at http://www.ciliate.org/system/downloads/1-upd-Genome-assembly.fasta) using STAR v2.6.1a[51] (with the parameters described previously[47] 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[47] using the genome assembly described previously and a custom *Tetrahymena 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[52] with the correct_annotation mode and default parameters. Counts were attributed to genes and normalized as transcripts per million (TPM) as previously described[47] using CoCo v0.2.5.p1 with the correct_count mode. Bedgraph files were generated using again CoCo v0.2.5.p1 (with the correct_bedgraph mode with default parameters). Differential expression analysis was performed using DESeq2[53] with default parameters and the count output of CoCo correct_count to compare the WT, Mlp1 IP and Mlp1 partial Mlp1 KO samples. Normalized counts measured in TPM and differential expression data for tRNAs are shown in **Additional Table 1**.

tRNA read fishing and binning into pre-mature 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 TTG: AATCCTCTGACCTGGGTTCGAATCCCAGTACGACCT) (**Additional Table 2**) 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 (**Table S2 and Table S3**). Fold-enrichment for Mlp1-bound tRNAs was calculated after summation of pre-tRNA counts, followed by taking the ratio of CMP data for Mlp1-immunoprecipitated tRNAs and WT input tRNAs

and displayed after log2 transformation in heatmaps (**Table S2**). Fold-enrichment for hLa-bound tRNAs from [18] 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 WT and partial KO tRNAs and displayed after log10 transformation in heatmaps (**Table S3**).

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 PAGE purified and eluted in 0.5M NaCl overnight at room temperature.

His-tagged proteins hLa, Mlp1 and Mlp1 mutants were purified from *Escherichia coli* BL21 cells using Co^{2+} beads, followed by heparin column purification. The proteins were

651 buffer exchanged in 1X PBS and quantified using Bovine Serum Albumin (BSA)
652 quantifications on SDS-PAGE gels.

653 EMSAs were performed as described[54]. Briefly, 3000 CPM of radioactive RNA
654 substrates (~ 0.1 nM) were incubated in 1X EMSA buffer [10% glycerol, 20 mM Tris-
655 HCl, pH 7.4, 100 mM KCl, 1 mM EDTA, 5 mM β-mercaptoethanol and bromophenol
656 blue] at 95°C for 5 minutes, followed by slow cooling to room temperature. For
657 competition EMSAs, unlabelled RNA was added to the radioactively labeled RNA prior
658 to incubation at 95°C. The concentration of unlabelled RNA depends on the protein
659 concentration used to bind >85% of the radioactively labeled RNA. Protein was added
660 and incubated for 30 minutes at 30°C. Then, the protein-RNA complexes were
661 immediately snap cooled on ice for 5 minutes and separated on a 8% native
662 polyacrylamide gel at 4°C at 100V. The gels were dried for 45 minutes at 80°C on a Gel
663 Dryer (Bio-Rad) and exposed to a storage phosphor screen overnight. The screens
664 were developed on a Typhoon. Quantification of bound and free RNA was done using
665 ImageJ and binding curves were fit using a nonlinear specific binding curve fitting
666 program and Kd values were calculated (GraphPad Prism).

667

668 tRNA mediated suppression assay in *Schizosaccharomyces pombe*

669 tRNA mediated suppression assays were performed as described previously[22]. In brief,
670 the *sla1- Schizosaccharomyces pombe* ySH9 strain encoding a defective UGA-
671 decoding suppressor tRNA (tRNA-Ser^{UCA}) and the *ade6-704* allele was transformed
672 using a pRep4 plasmid encoding Sla1p, Mlp1 and multiple Mlp1 mutants. Following

673 transformation of *Schizosaccharomyces pombe* suppressor strains using pRep4
674 plasmid (*ura4+*), strains were grown on selective media (Edinburgh Minimal Medium
675 (EMM) –ura –leu) and grown in liquid EMM –ura – leu to mid-log phase (OD 0.6-0.9).
676 Spotting was done by transferring 4 µL of liquid culture on low adenine-containing plates
677 (EMM –ura –leu ade10), followed by a 4-day incubation at 32°C. Yeast pellets for
678 protein purification or RNA extraction were obtained by centrifugation of mid-log phase
679 cells at 1800 g for 10 minutes following by two washes with ddH_2O.

680 Protein extraction was done by resuspension of cell pellets in NET-2 buffer [50 mM Tris-
681 HCl, pH 7.4, 150 mM NaCl, 0.05% NP40, 1 mM PMSF and 1X PIC], followed by lysis
682 through bead-beating for 2 minutes total (20 sec ON – 20 sec OFF intervals) at 4°C.
683 Cell lysates for protein analysis were obtained following centrifugation for 15 minutes at
684 20,000 g.

685 RNA extraction was completed by resuspending the cell pellets in complete RNA
686 extraction buffer A [50 mM NaOAc, pH 5.1, 10 mM EDTA, 1% SDS], followed by adding
687 37°C buffer A [50 mM NaOAc, pH 5.1]-saturated acid phenol and incubation at 65°C for
688 4 minutes with frequent vortexing. The aqueous top layer was extracted following
689 centrifugation for 3 minutes at 20,000 g and extracted again using
690 phenol:chloroform:isoamyl alcohol (25:24:1). RNA was precipitated from the aqueous
691 layer by ethanol precipitation and incubation at -80°C for at least 1 hour.

692

693 <u>RNA-immunoprecipitation and 3'RACE in *Schizosaccharomyces pombe*</u>

Sla1p- and Mlp1-transformed *Schizosaccharomyces 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 1800g, 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 KOAc, 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 20,000g 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 immunoglobin G (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. 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 (see **Table S5**), 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 on 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 *Homo sapiens, Mus musculus, Drosophila melanogaster, Arabidopsis thaliana, Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* from the Genomic tRNA Database (GtRNAdb)[55] and *Tetrahymena thermophila* sequences were obtained from the UCSC Genome Browser (https://genome.ucsc.edu/) (**Additional Table 3**). The number of tRNA genes encoded in the genome was determined for different eukaryotes for each isotype and isodecoder (**Table S4**). 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[56]). 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 *Homo sapiens, S pombe* and *Tetrahymena thermophila* using WebLogo[57] and divided based on discriminator base identity.

Immunofluorescent staining in *Tetrahymena thermophila*

Tetrahymena thermophila WT and partial Mlp1 KO strains were grown to mid-log phase and prepared for indirect immunofluorescent staining as described in[58]. Centrifugation steps including washes were performed for 3 minutes at 1000g. 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 taken 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 NaOAc, 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 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% sodium dodecyl sulfate (SDS) and 4X Denhardt's solution (Bio Basic D0062)], followed by overnight incubation with T4 Polynucleotide Kinase (PNK) ^{32}P-labeled probes (see **Table S5**). 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. To strip hybridized probe, the membrane was 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 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 [20 mM Tris-HCl, 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) 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 *Tetrahymena thermophila* cell pellets was performed as described previously[43]. 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. Next, 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 20,000g 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 WT and partial Mlp1 KO strains was digested with *EcoRI* restriction enzyme overnight at 37°C and separated by electrophoresis on 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 at 80°C on a Gel Dryer (Bio-Rad). The membrane was probed using a T4 Polynucleotide Kinase (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 biological triplicates. Mlp1 partial KO strain Northern blots and associated Western blots were performed in biological triplicates. TGIRT-sequencing of size-excluded Mlp1-RIP, WT and Mlp1 partial KO strains were performed in biological triplicates. EMSAs and competition EMSAs were performed in triplicates unless stated otherwise in figure legends. tRNA-mediates suppression assays in *Schizosaccharomyces pombe* ySH9 strains and associated Northern blots and Western blots were performed in biological triplicates unless stated otherwise in the figure legend. RNA-immunoprecipitation and associated Western blots from ySH9 and Sanger sequencing of clonal isolates derived

828 from 3'-RACE were performed in biological duplicates. Comparison of tRNA features in
829 different eukaryotes was done using the one-way ANOVA for 3'-trailer length ($F = 289.3$,
830 $DF=6$, P value < 0.0001) and mature tRNA length ($F = 0.9931$, $DF=6$, P value = 0.428).

831

832 Code availability

833 The RNAseq analysis code is available in a Snakemake workflow that can be found at
834 https://github.com/etiennefc/t_thermophila_RNA_Seq.git.

835

836 **Data availability**

837 TGIRT-sequencing has been deposited to the Gene Expression Omnibus (GEO) under
838 the accession number XXXX).

839

840 **References**

841 1. Fairley, J. A. *et al.* Human La is found at RNA polymerase III-transcribed genes in vivo.
842 *Proc. Natl. Acad. Sci. U. S. A.* **102**, 18350–18355 (2005).
843 2. Van Horn, D. J., Yoo, C. J., Xue, D., Shi, H. & Wolin, S. L. The La protein in
844 Schizosaccharomyces pombe: a conserved yet dispensable phosphoprotein that functions
845 in tRNA maturation. *RNA* **3**, 1434–1443 (1997).
846 3. Yoo, C. J. & Wolin, S. L. The Yeast La Protein Is Required for the 3 □ Endonucleolytic
847 Cleavage That Matures tRNA Precursors. **89**, 393–402 (1997).
848 4. Bayfield, M. A. & Maraia, R. J. Precursor-product discrimination by La protein during tRNA
849 metabolism. *Nat. Struct. Mol. Biol.* **16**, 430–437 (2009).

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

6. Huang, Y., Bayfield, M. A., Intine, R. V. & Maraia, R. J. Separate RNA-binding surfaces on the multifunctional La protein mediate distinguishable activities in tRNA maturation. *Nat. Struct. Mol. Biol.* **13**, 611–618 (2006).

7. Tomita, K. & Yamashita, S. Molecular mechanisms of template-independent RNA polymerization by tRNA nucleotidyltransferases. *Front. Genet.* **5**, (2014).

8. Bousquet-Antonelli, C. & Deragon, J.-M. A comprehensive analysis of the La-motif protein superfamily. *Rna* **15**, 750–764 (2009).

9. Alfano, C. *et al.* Structural analysis of cooperative RNA binding by the La motif and central RRM domain of human La protein. *Nat. Struct. Mol. Biol.* **11**, 323–329 (2004).

10. Dong, G., Chakshusmathi, G., Wolin, S. L. & Reinisch, K. M. Structure of the La motif: A winged helix domain mediates RNA binding via a conserved aromatic patch. *EMBO J.* **23**, 1000–1007 (2004).

11. Kenan, D. J. & Keene, J. D. La gets its wings. *Nat. Struct. Mol. Biol.* **11**, 303–305 (2004).

12. Deragon, J. Distribution , organization an evolutionary history of La and LARPs in eukaryotes. *RNA Biol.* **00**, 1–9 (2020).

13. Arthur, L. L. *et al.* Rapid generation of hypomorphic mutations. *Nat. Commun.* **8**, 1–15 (2017).

17. Xue, D., Rubinson, D. A., Pannone, B. K., Yoo, C. J. & Wolin, S. L. U snRNP assembly in yeast involves the La protein. *EMBO J.* **19**, 1650–1660 (2000).

18. Gogakos, T. *et al.* Characterizing Expression and Processing of Precursor and Mature Human tRNAs by Hydro-tRNAseq and PAR-CLIP. *Cell Rep.* **20**, 1463–1475 (2017).

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

20. Hussain, R. H., Zawawi, M. & Bayfield, M. A. Conservation of RNA chaperone activity of the human La-related proteins 4, 6 and 7. *Nucleic Acids Res.* **41**, 8715–8725 (2013).

21. Naeeni, A. R., Contes, M. R. & Bayfield, M. A. RNA chaperone activity of human La protein is mediated by variant RNA recognition motif. *J. Biol. Chem.* **287**, 5472–5482 (2012).

22. Porat, J. & Bayfield, M. A. Use of tRNA-Mediated Suppression to Assess RNA Chaperone Function. *Methods Mol. Biol. Clifton NJ* **2106**, 107–120 (2020).

23. Intine, R. V. A. *et al.* Control of Transfer RNA Maturation by Phosphorylation of the Human La Antigen on Serine 366. *Mol. Cell* **6**, 339–348 (2000).

24. Huang, Y., Intine, R. V., Mozlin, A., Hasson, S. & Maraia, R. J. 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. *Mol. Cell. Biol.* **25**, 621–636 (2005).

25. Arhin, G. K., Shen, S., P, I. F., Tschudi, C. & Ullu, E. Downregulation of the essential Trypanosoma brucei La protein affects accumulation of elongator methionyl-tRNA. **144**, 104–108 (2005).

26. Bai, C. & Tolias, P. P. Genetic analysis of a La homolog in Drosophila melanogaster. *Nucleic Acids Res.* **28**, 1078–1084 (2000).

27. Fleurdépine, S., Deragon, J.-M., Devic, M., Guilleminot, J. & Bousquet-Antonelli, C. A bona fide La protein is required for embryogenesis in Arabidopsis thaliana. *Nucleic Acids Res.* **35**, 3306–3321 (2007).

28. Park, J.-M. *et al.* The Multifunctional RNA-Binding Protein La Is Required for Mouse Development and for the Establishment of Embryonic Stem Cells. *Mol. Cell. Biol.* **26**, 1445–1451 (2006).

29. Behrens, A., Rodschinka, G. & Nedialkova, D. D. High-resolution quantitative profiling of tRNA abundance and modification status in eukaryotes by mim-tRNAseq. *Mol. Cell* **81**, 1802-1815.e7 (2021).
30. Tuller, T. *et al.* An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation. *Cell* **141**, 344–354 (2010).
31. Bogenhagen, D. F. & Brown, D. D. Nucleotide sequences in Xenopus 5S DNA required for transcription termination. *Cell* **24**, 261–270 (1981).
32. Leontis, N., DaLio, A., Strobel, M. & Engelke, D. Effects of tRNA-intron structure on cleavage of precursor tRNAs by RNase P from Saccharomyces cerevisiae. *Nucleic Acids Res.* **16**, 2537–2552 (1988).
33. Wu, J. *et al.* Cryo-EM Structure of the Human Ribonuclease P Holoenzyme. *Cell* **175**, 1393-1404.e11 (2018).
34. True, H. L. & Celander, D. W. Ribonuclease P of Tetrahymena thermophila *. **271**, 16559–16566 (1996).
35. Lee, Y. *et al.* Nuclear pre-tRNA terminal structure and RNase P recognition. *RNA N. Y. N* **3**, 175–185 (1997).
36. Nashimoto, M., Nashimoto, C., Tamura, M., Kaspar, R. L. & Ochi, K. The inhibitory effect of the autoantigen La on in vitro 3’ processing of mammalian precursor tRNAs. *J. Mol. Biol.* **312**, 975–984 (2001).
37. Kufel, J. & Tollervey, D. 3’-processing of yeast tRNATrp precedes 5’-processing. *RNA N. Y. N* **9**, 202–208 (2003).
38. Kufel, J., Allmang, C., Verdone, L., Beggs, J. D. & Tollervey, D. Lsm proteins are required for normal processing of pre-tRNAs and their efficient association with La-homologous protein Lhp1p. *Mol. Cell. Biol.* **22**, 5248–5256 (2002).
39. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).

40. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
41. Zheng, W. *et al.* LOMETS2: improved meta-threading server for fold-recognition and structure-based function annotation for distant-homology proteins. *Nucleic Acids Res.* **47**, W429–W436 (2019).
42. Stover, N. A., Punia, R. S., Bowen, M. S., Dolins, S. B. & Clark, T. G. Tetrahymena genome database Wiki: a community-maintained model organism database. *Database* **2012**, bas007 (2012).
43. Fillingham, J. S., Bruno, D. & Pearlman, R. E. 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 Res.* **29**, 488–498 (2001).
44. Mochizuki, K. High efficiency transformation of Tetrahymena using a codon-optimized neomycin resistance gene. *Gene* **425**, 79–83 (2008).
45. Chalker, D. L. Transformation and strain engineering of Tetrahymena. *Methods Cell Biol.* **109**, 327–345 (2012).
46. Merriam, E. V. & Bruns, P. J. Phenotypic Assortment in Tetrahymena Thermophila: Assortment Kinetics of Antibiotic-Resistance Markers, Tsa, Death, and the Highly Amplified Rdna Locus. *Genetics* **120**, 389–395 (1988).
47. Fafard-Couture, É., Bergeron, D., Couture, S., Abou-Elela, S. & Scott, M. S. Annotation of snoRNA abundance across human tissues reveals complex snoRNA-host gene relationships. *Genome Biol.* **22**, 172 (2021).
48. Nottingham, R. M. *et al.* RNA-seq of human reference RNA samples using a thermostable group II intron reverse transcriptase. *RNA N. Y. N* **22**, 597–613 (2016).
49. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma. Oxf. Engl.* **30**, 2114–2120 (2014).

50. Sheng, Y. *et al.* The completed macronuclear genome of a model ciliate Tetrahymena thermophila and its application in genome scrambling and copy number analyses. *Sci. China Life Sci.* **63**, 1534–1542 (2020).
51. Dobin, A. & Gingeras, T. R. Optimizing RNA-Seq Mapping with STAR. in *Data Mining Techniques for the Life Sciences* (eds. Carugo, O. & Eisenhaber, F.) 245–262 (Springer, 2016). doi:10.1007/978-1-4939-3572-7_13.
52. Deschamps-Francoeur, G., Boivin, V., Abou Elela, S. & Scott, M. S. CoCo: RNA-seq read assignment correction for nested genes and multimapped reads. *Bioinformatics* **35**, 5039–5047 (2019).
53. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
54. Vinayak, J. *et al.* Human La binds mRNAs through contacts to the poly(A) tail. *Nucleic Acids Res.* **46**, 4228–4240 (2018).
55. Chan, P. P. & Lowe, T. M. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.* **44**, D184–D189 (2016).
56. Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A. & Steinberg, S. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* **26**, 148–153 (1998).
57. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **14**, 1188–1190 (2004).
58. Wenkert, D. & Allis, C. D. Timing of the appearance of macronuclear-specific histone variant hv1 and gene expression in developing new macronuclei of Tetrahymena thermophila. *J. Cell Biol.* **98**, 2107–2117 (1984).

Acknowledgements

We thank R.J. Maraia and J-M. Deragon for comments on the manuscript. This work was funded by the Canadian Institutes of Health Research's Institute of Genetics (to M.A.B).

Author Contributions

K.K. performed most experiments. J.G. and R.P. provided Tetrahymena WT strains, generated the Mlp1 partial KO strain and associated confirmation (southern blot and PCR). S.A.E. performed cDNA TGIRT library preparation and sequencing and E.F-C., M.S. the associated bioinformatic analysis. K.K and M.A.B. designed the study, analysed the data, and wrote the paper.

Additional Information

Competing Interests Statement: The authors declare that they have no competing interests.

Correspondence and Materials Requests should be addressed to M.A.B.

Figure Legends

Figure 1. Alveolate La proteins are predicted to have a La motif (LaM) but not an RNA recognition motif-1 (RRM1)

(A) Schematic representation of RNA-binding domains found in La and La-related protein-7 (LARP7) from different eukaryotes. The LaM and RRM 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 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 Figure 1A. Full LaM and RRM1 domains shown in Figure S1A,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 Figure 1A.

Figure 2. Mlp1 preferentially binds pre-tRNAs partially via the 3’-terminal uridylates.

(A) Northern blot analysis of Mlp1 ribonucleoprotein-immunoprecipitated (RIP) samples from *Tetrahymena thermophila*. Mlp1 enriches pre-tRNAs more efficiently than mature tRNAs. Pre-tRNA Ile^{UAU} and Leu^{UAA} are both recognized by 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 Figure S2A.

(B) Next generation sequencing data of tRNAs using TGIRT-Seq (columns represent replicates), shown as a heatmap of log2 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.* analyzed similarly for hLa (bottom). The log2 transformed fold enrichment was calculated by taking ratios of normalized hLa-immunoprecipitated tRNA and the averaged input tRNAs CPM. The data was split between premature and mature tRNAs based on the 3'-end of the transcript (-CCA ending mature tRNAs and U-ending pre-tRNAs).

(C) Schematic representation of ^{32}P-labeled pre-tRNA intermediates containing 5'-leader and 3'-trailer, 5'-leader only, 3'-trailer only and mature tRNA used in D-I and Figure S2D,E.

(D-I) Binding curves from EMSAs comparing ^{32}P-labeled Leu^{AAG} 5'-leader, 3'-trailer containing pre-tRNA and mature Leu^{AAG} tRNA. Native gels are shown in Figure S3A 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. Native gels are shown in Figure S3A and 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). Standard deviation between replicates is shown as error bars (n=2). Native gels are shown in Figure S3D,E.

Figure 3. 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 pre-tRNA and mature tRNA for hLa (A) and Mlp1 (B). Standard deviation between replicates is shown as error bars (Mlp1 U10: n=4, tRNA: n=2, hLa U10: n=2, tRNA: n=1). Native gels are shown in Figure S4A.

(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: CUGCUGUUU<u>C</u> (U_{-1}C), CUGCUGUU<u>C</u>U (U_{-2}C), CUGCUGU<u>C</u>UU (U_{-3}C) and CUGCUG<u>CCCC</u> (4C) for hLa (D) or Mlp1 (E). Standard deviation between replicates is shown as error bars (n=2). Native gels are shown in Figure S4B,C.

(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 (degraded RNA mimic) for hLa (G) and Mlp1 (H). Standard deviation between replicates is shown as error bars (n=2). Native gels are shown in Figure S4D.

Figure 4. Mlp1 promotes tRNA mediated suppression in the absence of protection of 3'-trailers.

(A) tRNA-mediated suppression assay *Schizosaccharomyces pombe* ySH9 strain transformed with pRep4 encoded Sla1p (positive control; *S. pombe* La protein), Mlp1 or indicated Mlp1 mutants. 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). 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. 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. 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 (top band), 5'-leader processed pre-tRNA (middle band) and both 5'-leader and 3'-trailer processed pre-tRNA (bottom band) in sla1 transformants (positive control; lane 2). The same blot was probed with a 3'-trailer complementary probe, a 5'-leader complementary probe and a mature tRNA complementary probe. Black boxes correspond to a marker size of 100 nt. 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.

Figure 5. 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 knockout (KO) strain compared to the wild type (WT) strain. Loading control: histone H3 and β-actin.

(B) Northern blot detecting tRNAs Ile^{UAU}, Leu^{UAA}, Val^{CAC} and Tyr^{GUA} pre-tRNA intermediates with an intron, 5'-leader, and 3'-trailer specific probe. Pre-tRNA Ile^{UAU} and Leu^{UAA} 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 using TGIRT-Seq shown as a heatmap of log10 transformed normalized cumulative pre-tRNA counts for each uridylate tail length. For clarity, the average is plotted and the longest uridylate tail shown is U_6 (UUUUUU-

3'OH).

(D) Quantification of uridylate tail length from TGIRT-Seq data in (C). For each tRNA isodecoder, the presence of a uridylate tail length (U1, U2, etc.) was scored as present if its abundance the CPM was greater than 0.3; log10(2).

(E-F) Scatterplots of log2 transformed normalized tRNA counts (CPM) from TGIRT-Seq data from WT strains (E) and KO strains (F) plotted against the number of genes per tRNA isotype encoded in the genome (**Table S4**). 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 Tyr^{GUA} and Arg^{UCU} levels using a probe specific for the spliced mature tRNA. U5 was used as a loading control on both membranes.

Figure 6. 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 determination 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. Statistical significance ($P < 0.05$) was found between all

species, except between *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, *Schizosaccharomyces pombe* and *Arabidopsis thaliana,* and *Drosophila melanogaster, M. musculus* and *H. sapiens* after comparison using a one-way ANOVA and Tukey's multiple comparison test.

(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. No statistical significance ($P < 0.05$) was observed using a one-way ANOVA and Tukey's multiple comparison test.

(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 highlighter in dark green. The most 3'-terminal nucleotide of the 5'-leader sequence is highlighted with a light green box (N_{-1}).

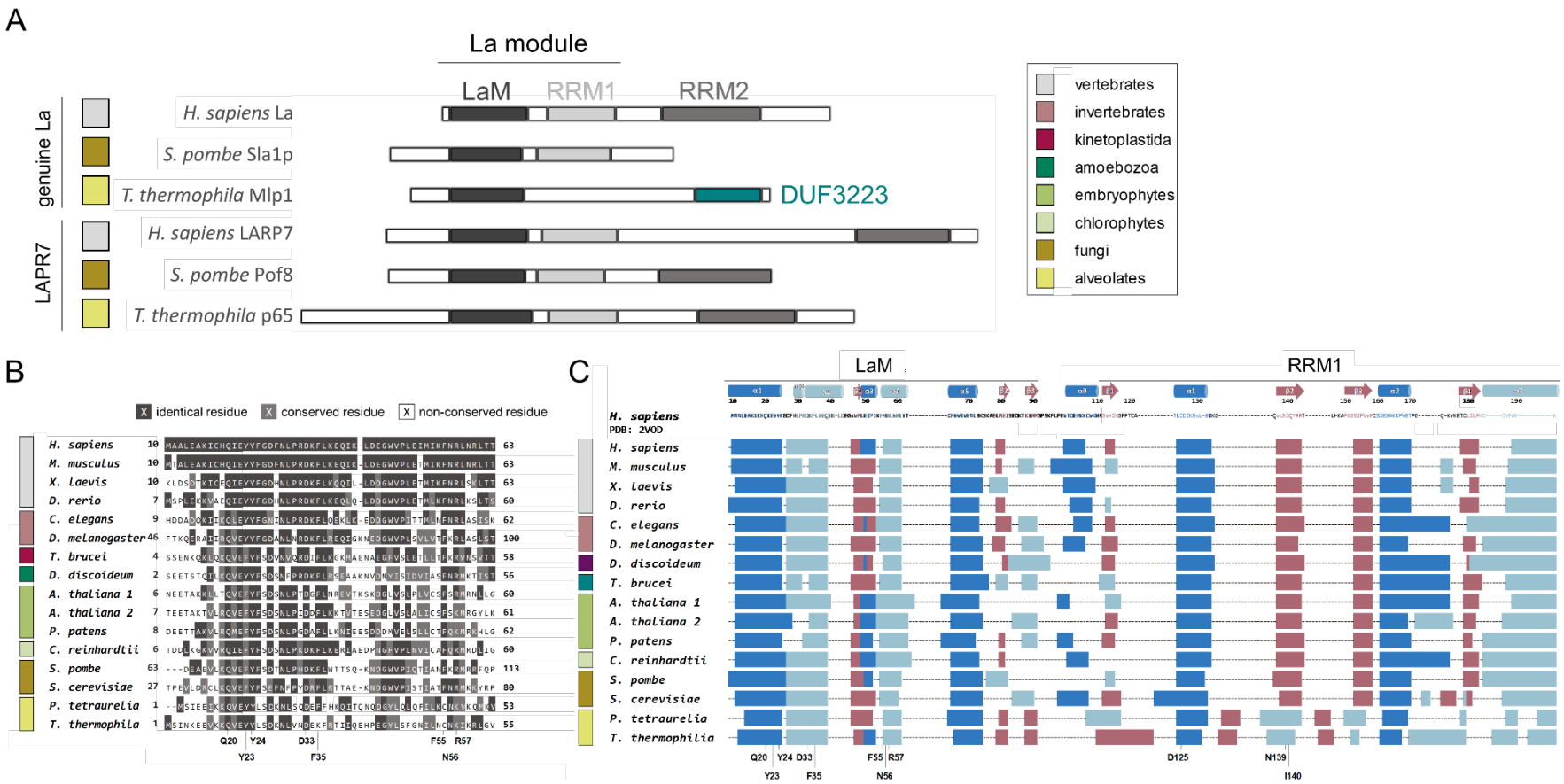
(D,E) Logo analysis of 5'-leader sequences of pre-tRNAs in *Tetrahymena thermophila, Schizosaccharomyces pombe* and *Homo sapiens* (D). The same analysis split by the discriminator base identity (E). The number of pre-tRNAs representing each condition is shown underneath each logo.

Figure 7. Model for pre-tRNA processing in *Tetrahymena thermophila*.

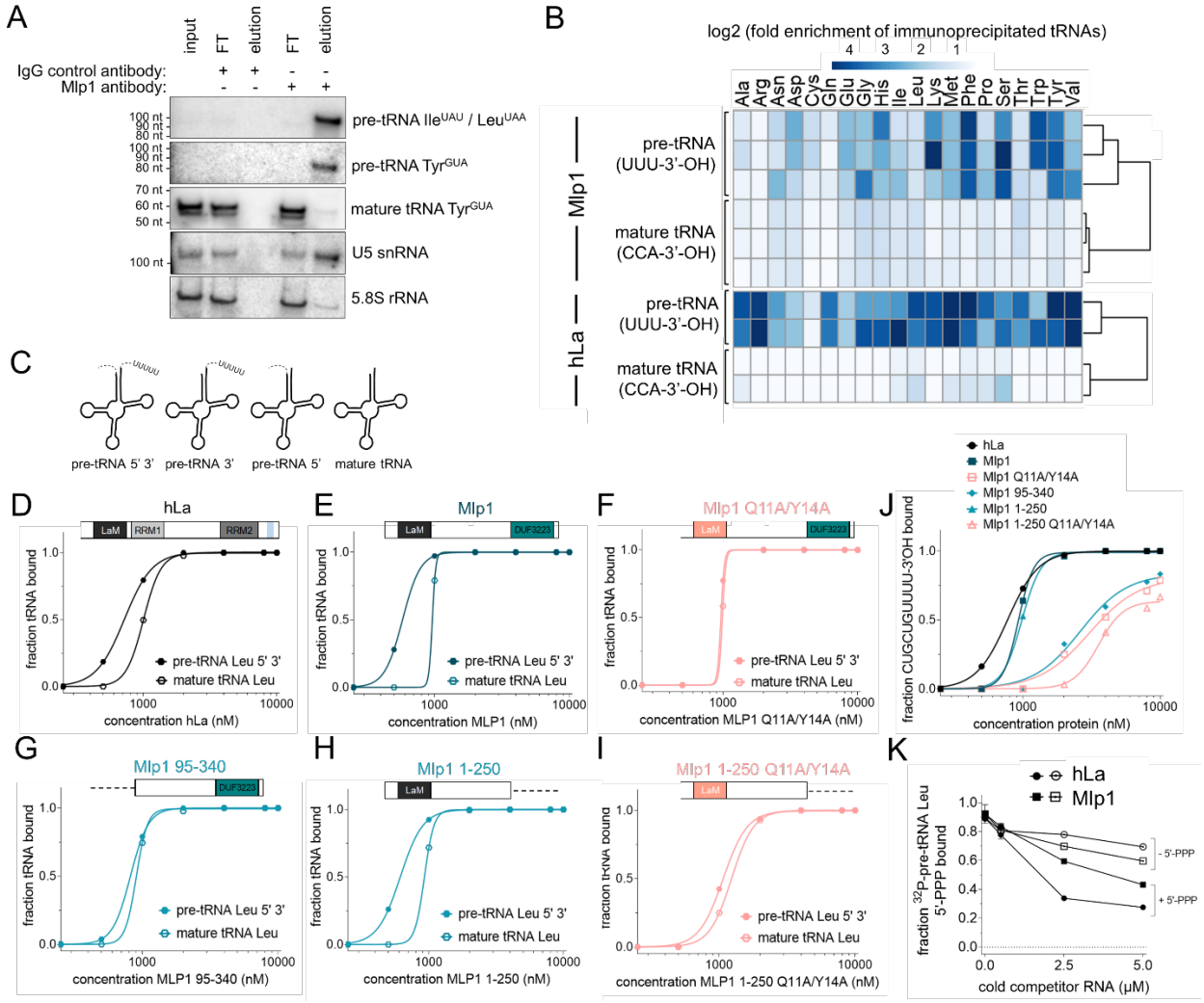
During La-dependent processing (top, left) in previously studied eukaryotes such as yeast, the La protein is the first protein to associate with pre-tRNAs on the 3'-stretch of

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 *Tetrahymena 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.

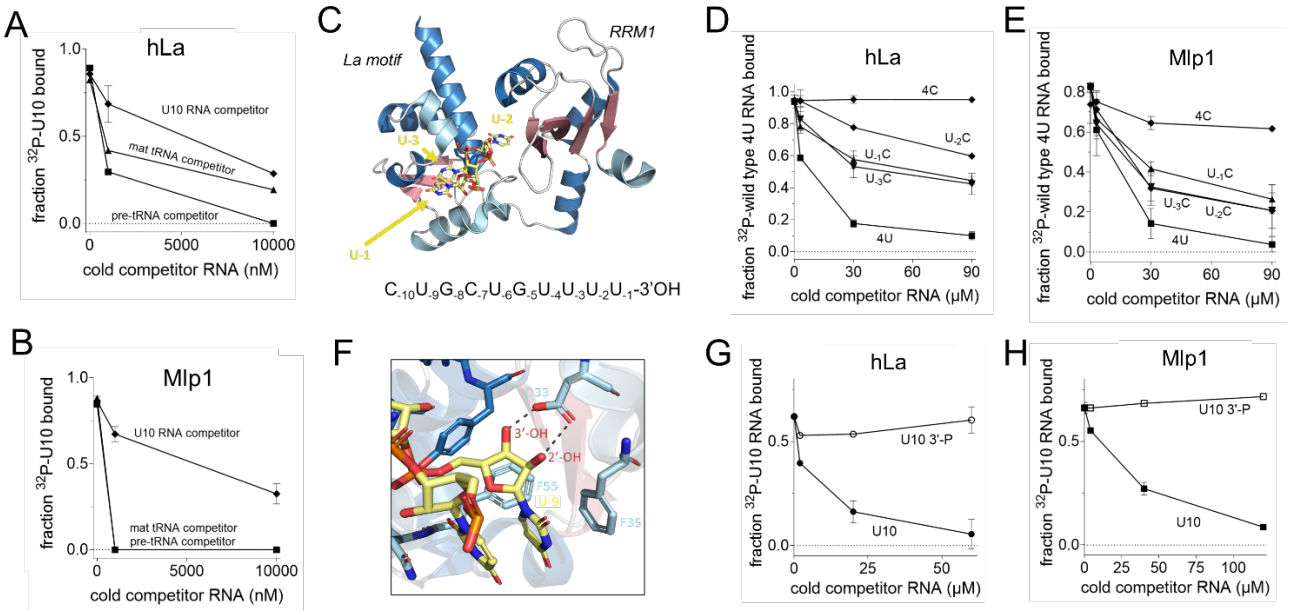
1


A
La module
LaM
RRM1
RRM2
genuine La
LAPR7
H. sapiens La
S. pombe Sla1p
T. thermophila Mlp1
DUF3223
H. sapiens LARP7
S. pombe Pof8
T. thermophila p65
vertebrates
invertebrates
kinetoplastida
amoebozoa
embryophytes
chlorophytes
fungi
alveolates
B
identical residue
conserved residue
non-conserved residue
H. sapiens
M. musculus
X. laevis
D. rerio
C. elegans
D. melanogaster
T. brucei
D. discoideum
A. thaliana 1
A. thaliana 2
P. patens
C. reinhardtii
S. pombe
S. cerevisiae
P. tetraurelia
T. thermophila
Q20
Y24
D33
F55
R57
Y23
F35
N56
C
LaM
RRM1
H. sapiens
PDB: 2VOD
M. musculus
X. laevis
D. rerio
C. elegans
D. melanogaster
D. discoideum
T. brucei
A. thaliana 1
A. thaliana 2
P. patens
C. reinhardtii
S. pombe
S. cerevisiae
P. tetraurelia
T. thermophilia
Q20
Y24
D33
F55
R57
D125
N139
Y23
F35
N56
I140

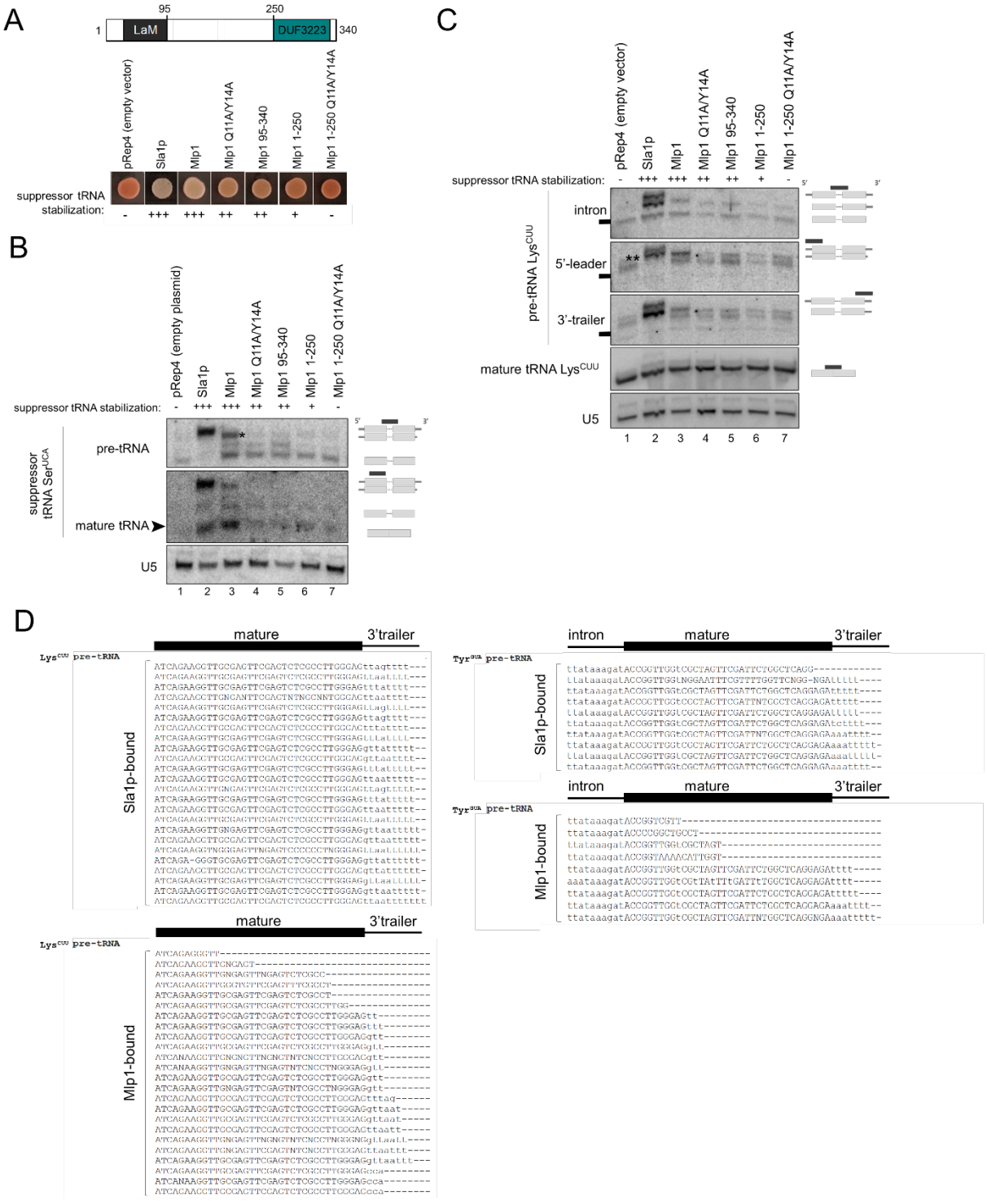
2



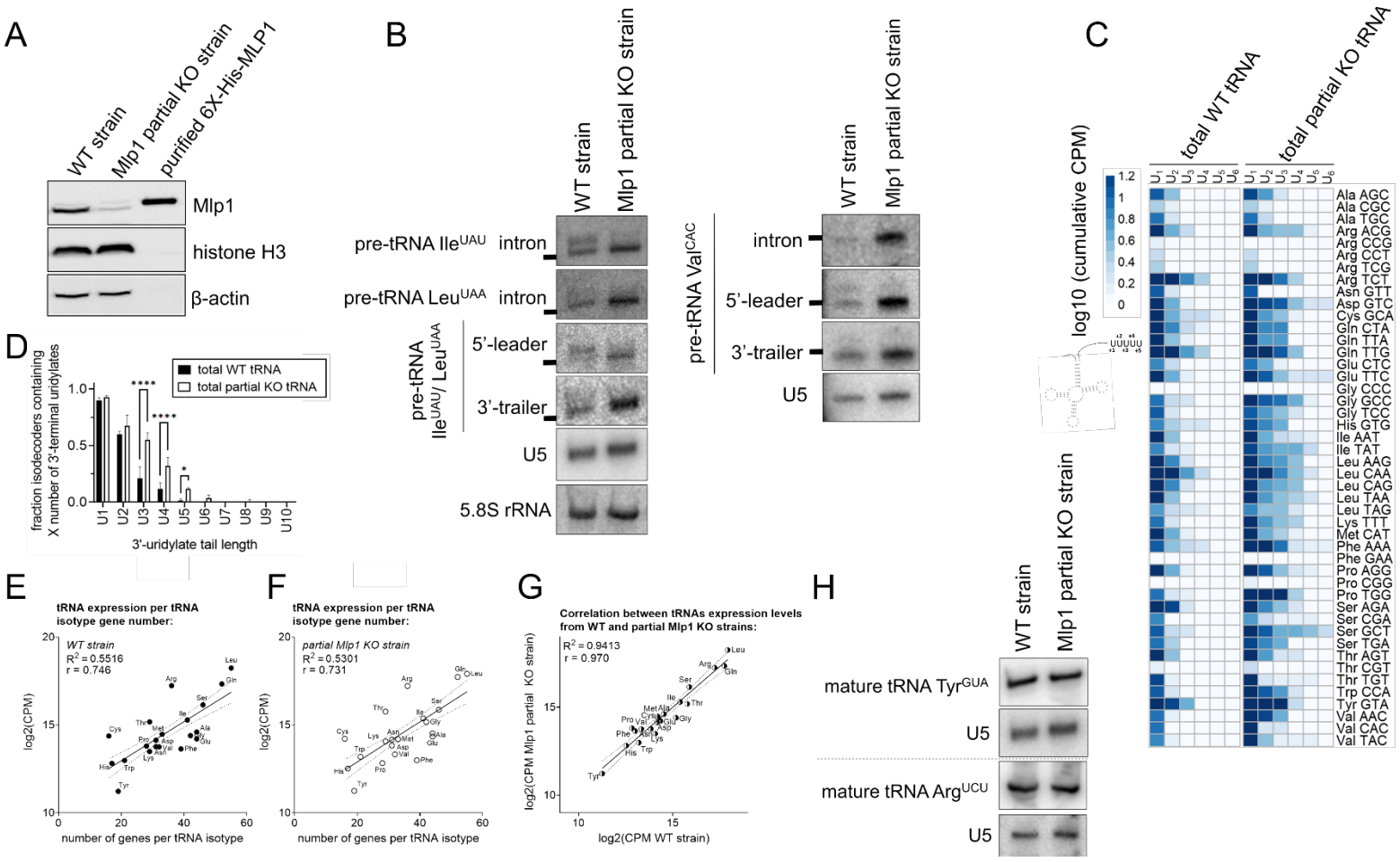
3



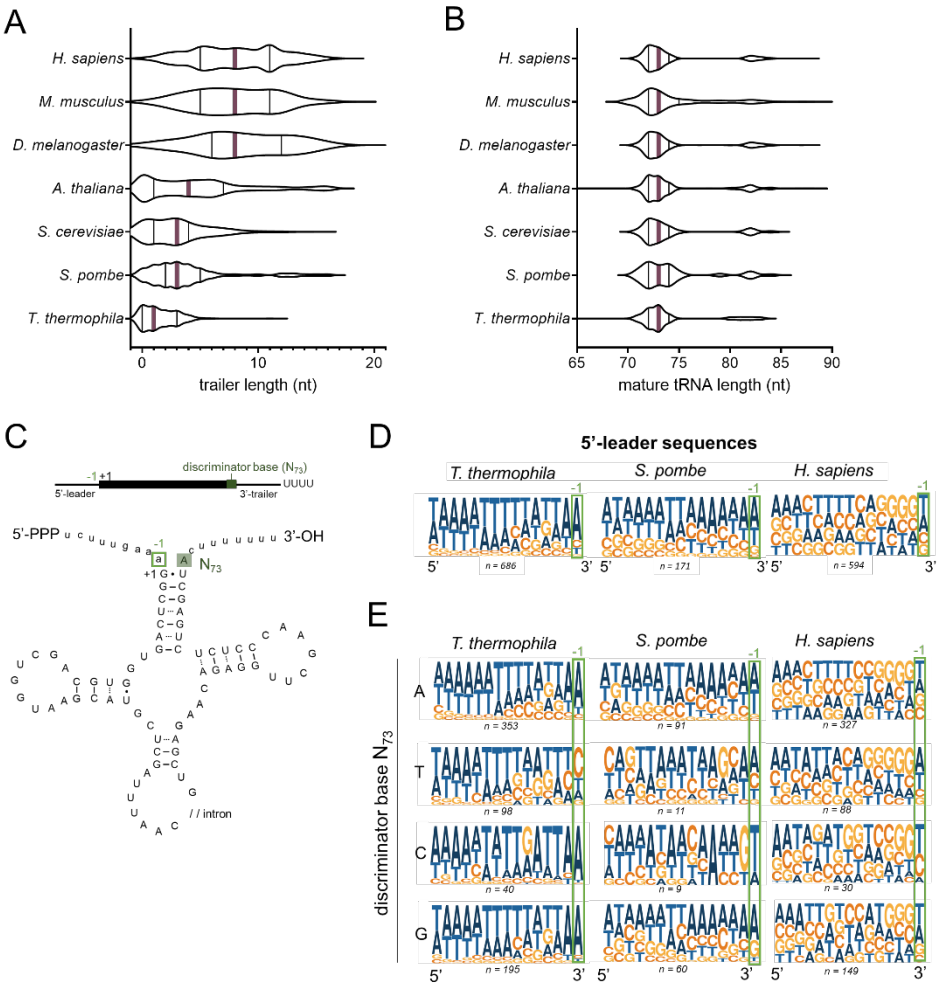
4



5


A
WT strain
Mlp1 partial KO strain
purified 6X-His-MLP1
Mlp1
histone H3
β-actin
B
WT strain
Mlp1 partial KO strain
pre-tRNA Ile^UAU intron
pre-tRNA Leu^UAA intron
pre-tRNA Ile^UAU/ Leu^UAA
5'-leader
3'-trailer
U5
5.8S rRNA
pre-tRNA Val^CAC
intron
5'-leader
3'-trailer
U5
C
total WT tRNA
total partial KO tRNA
log10 (cumulative CPM)
Ala AGC
Ala CGC
Ala TGC
Arg ACG
Arg CCG
Arg CCT
Arg TCG
Arg TCT
Asn GTT
Asp GTC
Cys GCA
Gln CTA
Gln TTA
Gln TTG
Glu CTC
Glu TTC
Gly CCC
Gly GCC
Gly TCC
His GTG
Ile AAT
Ile TAT
Leu AAG
Leu CAA
Leu CAG
Leu TAA
Leu TAG
Lys TTT
Met CAT
Phe AAA
Phe GAA
Pro AGG
Pro CGG
Pro TGG
Ser AGA
Ser CGA
Ser GCT
Ser TGA
Thr AGT
Thr CGT
Thr TGT
Trp CCA
Tyr GTA
Val AAC
Val CAC
Val TAC
D
total WT tRNA
total partial KO tRNA
fraction isodecoders containing X number of 3'-terminal uridylates
3'-uridylate tail length
E
tRNA expression per tRNA isotype gene number:
WT strain
R2 = 0.5516
r = 0.746
log2(CPM)
number of genes per tRNA isotype
F
tRNA expression per tRNA isotype gene number:
partial Mlp1 KO strain
R2 = 0.5301
r = 0.731
log2(CPM)
number of genes per tRNA isotype
G
Correlation between tRNAs expression levels from WT and partial Mlp1 KO strains:
R2 = 0.9413
r = 0.970
log2(CPM Mlp1 partial KO strain)
log2(CPM WT strain)
H
WT strain
Mlp1 partial KO strain
mature tRNA Tyr^GUA
U5
mature tRNA Arg^UCU
U5

6



7