Transcriptome-wide analysis of pseudouridylation in *Drosophila melanogaster*

Wan Song[1,3,*], Ram Podicheti[2,*], Douglas B. Rusch[2], W. Daniel Tracey[1,3,**]

[1]Gill Center for Biomolecular Research, Indiana University Bloomington, Indiana, USA

[2]Center for Genomics and Bioinformatics, Indiana University Bloomington, Indiana, USA

[3]Department of Biology, Indiana University, Bloomington, Indiana, USA

[*] Co-first authors

11 **Short title: Pseudouridine mapping in Drosophila**

12

13 **Key words:** *Drosophila melanogaster*, pseudouridine, transcriptome, Psi-seq, CMC, RluA-2 target, differential

14 expression

15

16 ** Corresponding author:

17 W. Daniel Tracey

18 The Linda and Jack Gill Center for Biomolecular Science

19 Indiana University

20 Multidisciplinary Science Building II (MSBII)

21 702 North Walnut Grove Avenue, Rm.106

22 Bloomington, IN 47405

23 Phone: (812) 856-2574

24 Email: dtracey@indiana.edu

Abstract

Pseudouridine (Psi) is one of the most frequent post-transcriptional modification of RNA. Enzymatic Psi modification occurs on rRNA, snRNA, snoRNA, tRNA, non-coding RNA and has recently been discovered on mRNA. Transcriptome-wide detection of Psi (Psi-seq) has yet to be performed for the widely studied model organism *Drosophila melanogaster*. Here, we optimized Psi-seq analysis for this species and have identified thousands of Psi modifications throughout the female fly head transcriptome. We find that Psi is widespread on both cellular and mitochondrial rRNAs. In addition, more than a thousand Psi sites were found on mRNAs. When pseudouridylated, mRNAs frequently had many Psi sites. Many mRNA Psi sites are present in genes encoding for ribosomal proteins, and many are found in mitochondrial encoded RNAs, further implicating the importance of pseudouridylation for ribosome and mitochondrial function. The 7SLRNA of the signal recognition particle is the non-coding RNA most enriched for Psi. The three mRNAs most enriched for Psi encode highly-expressed yolk proteins (Yp1, Yp2, Yp3). By comparing the pseudouridine profiles in the *RluA-2* mutant and the w^{1118} control genotype, we identified Psi sites that were missing in the mutant RNA as potential RluA-2 targets. Finally, differential gene expression analysis of the mutant transcriptome indicates a major impact of loss of RluA-2 on the ribosome and translational machinery.

41 **Introduction**

42 Pseudouridylation is one of the earliest discovered and most abundant post-transcriptional RNA modifications.
43 Pseudouridine (Psi), the C5-glycoside isomer of uridine occurs at specific nucleotide residues in rRNAs, tRNAs, snRNAs,
44 snoRNAs and other ncRNAs (Ge and Yu 2013) and more recently Psi has also been found in mRNA (Carlile et al. 2014;
45 Schwartz et al. 2014; Lovejoy et al. 2014; Li et al. 2015). Psi sites within RNAs are thought to have various functions
46 depending on the molecule. Psi sites in tRNAs are present in the anticodon stem-loop, D-stem and other conserved
47 sites and contribute to the stabilization of tertiary structure. In rRNA the known Psi sites are concentrated near the
48 decoding site, the peptidyl transferase center, and at sites of interaction with ribosomal proteins, where they are
49 thought to be important for ribosome assembly and protein synthesis (Charette and Gray 2000). In the non-coding RNA
50 (ncRNA) of telomerase a functionally important Psi residue is located in a region involved in telomerase reverse
51 transcriptase binding (Penzo et al. 2017). Incorporation of Psi into therapeutic mRNAs results in a reduced immune
52 response to the RNA and enhanced protein expression (Kauffman et al. 2016).

53 Experimental detection of pseudouridine residues has been hindered and complicated by the fact that Psi and
54 uridine share the same molecular mass and similar base-pairing properties upon reverse transcription. The most
55 frequently used method to detect this nucleoside is to treat RNA with cyclohexyl-N'-(2-morpholinoethyl)-carboiimide
56 metho-p-toluene sulfonate (CMCT). CMCT reacts with uridine-like and guanosine-like nucleotides and subsequent
57 alkaline treatment of the reacted RNA hydrolyzes CMC adducts to G and U but does not hydrolyze CMC-Psi mono-
58 adducts. The bulky CMC moiety blocks primer extension (Ho and Gilham 1967) and this block may be directly detected
59 with polyacrylamide gel electrophoresis and autoradiography in highly abundant RNA species such as rRNA. Pioneering
60 work by the Ofengand group used this reaction to detect highly conserved Psi sites in rRNA in species ranging from
61 prokaryotes to eukaryotes (Bakin and Ofengand 1993).

62 More recently next generation RNA sequencing approaches have been adopted for detection of pseudouridines
63 in lower abundance transcripts. The majority of these methods also depend on CMCT treatment since the CMC adduct
64 conjugated to Psi prevents read-through by reverse transcriptase (RT). Bioinformatic analyses of CMC treated libraries
65 identifies sites of pseudouridylation where reverse transcription is more frequently terminated relative to reads in
66 untreated libraries. This Psi-seq approach (also known as "pseudo-seq", "Ψ-seq", or "CeU-Seq") has been employed for
67 global transcriptome-wide identification of pseudouridylation sites in yeast (Carlilie et al. 2014, Schwartz et al. 2014,
68 Lovejoy et al. 2014), human cell lines (Schwartz et al. 2014, Carlilie et al. 2014, Li et al. 2015), mouse (Li et al. 2015),
69 *Toxoplasma gondii* (Nakamoto et al. 2017) and *Arabidopsis thaliana* (Sun et al. 2019).

70 The post-transcriptional isomerization of uridine to pseudouridine is catalyzed by six families of pseudouridine
71 synthases (TruA, TruB, TruD, RsuA, RluA and Pus10). The archaeal and eukaryotic TruB homologues function with
72 auxiliary guide RNAs while the rest act as stand-alone enzymes (Hamma and Ferre-D'Amare 2006; Ofengand 2002).
73 Some pseudouridine synthases have highly conserved and specific substrates whereas others show more broad and
74 complex targets (Rintala-Dempsey and Kothe 2017). In the *Drosophila melanogaster* genome, there are nine proteins
75 with an annotated pseudouridine synthase domain. *Drosophila* Nucleolar protein 60B (Nop60B), also known as *minifly*
76 (*mfl*), is a well-studied homolog of human box H/ACA RNP dyskerin (mouse NAP57 and yeast Cbf5). Nop60B/Mfl is

77 essential for *Drosophila* viability and fertility as it is required for ribosome biogenesis and the *mfl* mutant has reduced
78 pseudouridylation at several sites of 28S and 18S rRNA (Phillips et al. 1998; Giordano et al. 1999; Breznak et al. 2022).
79 Two other pseudouridine synthases (RluA-1 and RluA-2) have been studied for their roles in the negative regulation of
80 nociception (Song et al. 2020). RluA-1 is primarily expressed in multidendritic sensory neurons including the larval
81 nociceptors whereas RluA-2 is expressed ubiquitously. RluA-1 and RluA-2 belong to RluA family which have complex
82 substrate specificity in yeast and bacteria (Hoang et al. 2006).

83 Here, we have applied the Psi-seq technique to identify sites of pseudouridylation in *D. melanogaster* female
84 head transcriptome. In addition, we have investigated the pseudouridine and transcriptional profile in animals lacking
85 the ubiquitously expressed RluA family member RluA-2. Our results suggest a major role for RluA-2 for
86 pseudouridylation across a variety of RNA classes as well as large changes in gene expression that occur in animals
87 lacking RluA-2 function.

88

89 **Results:**

90 *Optimization of Psi-seq library preparation conditions for efficient detection of putative pseudouridine sites*

91 We began our experiments with optimization of a previously used strategy (Schwartz et al. 2014) and applied
92 the optimized technique to *Drosophila* head RNA. As an internal control to monitor the effectiveness of the treatment
93 and efficiency of the Psi detection, we spiked in a 150bp *in vitro* transcribed RNA oligonucleotide containing a single
94 pseudouridine at position 43 (Schwartz et al. 2014) at an early stage of library preparation. In pilot experiments we
95 identified experimental treatment conditions that allowed for robust calling of the pseudouridine of the RNA spike-in
96 (**Figure S1,** details in Materials and Methods). Our RNA samples were split into two equal halves immediately after
97 polyA enrichment with one half subjected to CMCT treatment and the other to a mock treatment. This was followed by
98 a ligation of an RNA adapter, an adapter-specific RT reaction, then a second ligation of a DNA adapter. The samples
99 were then barcoded, amplified, and pooled for Illumina sequencing (**Figure 1A**).

100 The RNA samples we used for the mapping of pseudouridylation were isolated from the heads of one day old
101 virgin female flies of an isogenized w^{1118} (isow^{1118}) strain. This genetic background was previously used for behavioral
102 characterization of *RluA-1* and *RluA-2* mutants (Song et al. 2020) and we focused on the head tissue since it would be
103 enriched for genes expressed in the nervous system which are of greatest interest to us given the nociception
104 phenotypes in the mutants. To assess repeatability, we carried out two sets of experiments. In experiment 1 (Exp 1) we
105 prepared, sequenced, and analyzed three pairs of w^{1118} libraries (with each pair containing a CMCT and mock treatment).
106 In experiment 2 (Exp 2) we performed the procedures on four pairs of w^{1118} and four pairs of *RluA-2*$^{del\text{-}FRT}$ mutant (**Table**
107 **S1 in File S1**). An average of 91% of the paired reads from the 22 libraries had both mates cleared through quality filters
108 and more than 87% of the reads could be mapped concordantly (**Table S2 in File S1**), confirming the high quality of the
109 sequencing reads for the libraries from both experiments.

110 To call potential Psi sites we determined a *termination ratio* for all nucleotide sites in the sequenced libraries.
111 The termination ratio was computed as the ratio of read counts that terminated at a specific site in CMC treated
112 samples and the total coverage in mock treated samples. Since the CMC treatment blocks reverse transcription

113 adjacent to a Psi site, a higher termination ratio in the CMC library compared to the corresponding mock library is an
114 indicator of the location of a potential Psi residue. We then used the log2 of termination ratio in the paired CMC and
115 mock libraries to define *termination fold change*. Our bioinformatics pipeline, when applied to the synthetic spike-in
116 RNA in our experiments resulted in termination fold change at the expected position adjacent to the Psi site that was
117 greater than 6 across all pairs of CMC and mock libraries (**Table S2 in File S1**). A termination fold change greater than 6
118 represents an average absolute fold change greater than 64 (since $\log_2(64) = 6$) (**Figure 1B**). This result for the RNA
119 spike-in indicates a highly efficient CMCT treatment which produced strong termination adjacent to the Psi site.

120 *Pseudouridylation of Drosophila ribosomal RNAs*

121 Previous studies have mostly employed subjectively chosen thresholds for Psi site calling, i.e. a combination of
122 "Psi ratio" (equivalent to our "termination ratio") > 0.1 and "Psi fold change" (equivalent to our "termination fold
123 change") of >3 were used to identify the candidate pseudouridine sites in yeast and human cell lines (Schwartz et al.
124 2014). We found that applying the same thresholds to our data resulted in the calling of an extremely high number of
125 sites. For instance, if we applied the thresholds of Schwartz et al (2014) we identified over 200,000 sites across seven
126 w^{1118} libraries. In addition, we found that reproducibility of sites called using this thresholding method was low when
127 comparing across replicates. The very high number of sites using the thresholding and the lack of reproducibility across
128 samples caused us to question the robustness of this calling method for our experimental conditions.

129 To overcome this limitation, we analyzed the potential Psi sites according to the statistics of the termination
130 ratios that we observed. To do this we applied a two- proportion z test to evaluate whether the termination ratio at a
131 specific site in the CMC treated sample was significantly higher than in the mocked treated sample. Sites with an
132 associated FDR<0.05 are statistically more likely to terminate than expected and were called as potential Psi sites. In
133 addition, we expect biologically meaningful sites to show reproducibility across samples, thus we required that a
134 potential Psi site was called in at least two w^{1118} replicates across the two independent experiments (details in Materials
135 and Methods). Finally, we categorized the Psi sites according to high reproducibility (identified in 6-7 paired replicates),
136 intermediate reproducibility (identified in 4-5 paired replicates), or low reproducibility (identified in 2-3 paired
137 replicates).

138 According to these criteria, we identified 735 potential Psi sites on cellular rRNA which survived the polyA
139 enrichment of our samples (**Table S3 in File S1**). Although this level of Psi on rRNA vastly exceeds the number of
140 previously identified sites, 532 of the sites (72.4%) were highly reproducible. These highly reproducible sites on cellular
141 rRNA are distributed across rRNA classes including 288 Psi sites in 28S, 16 sites in 5.8S, as well as 228 sites in 18S rRNAs
142 (**Table S3 in File S1**). In *Drosophila melanogaster*, only 57 Psi sites on the large subunit (LSU) of cytoplasmic rRNA have
143 been previously reported using the canonical primer-extension based pseudouridine detection method (Ofengand and
144 Bakin 1997) ("Ofengand site" indicated by gray vertical lines in **Figure S2)**. Using our method of Psi calling, we have
145 identified 53 of those known LSU sites (labeled as "Ofengand site" in Column H in **Table S3 in File S1**). Notably, 49/53 of
146 those known sites are highly reproducible Psi sites in our analysis (**Table S3 in File S1**). In **Figure 2**, the cytoplasmic
147 rRNA Psi sites are mapped onto the 3D based secondary structure of LSU rRNA of *Drosophila melanogaster* (RiboVision

148 version 1.15+, (Bernier et al. 2014)). As there is no obvious clustering of the sites, additional analyses will be required to
149 understand how the patterns observed relate to ribosomal function.

150 In addition to the 735 potential Psi sites identified on cytoplasmic rRNA, we also found 125 sites on
151 mitochondrial rRNA, among which 57 highly reproducible sites were identified on the mitochondrial large ribosomal
152 RNA (lrRNA) (**Table S3 in File S1**). Previously studies have identified a single pseudouridylation site in human
153 mitochondrial 16S rRNA (Psi1397) which is essential for its stability and assembly into the mitochondrial ribosomes
154 (Antonicka et al. 2017). Our results suggest a more widespread occurrence of Psi sites on mitochondrial rRNA in *D.*
155 *melanogaster*.

156 *Psi-seq identified predicted and novel Psi in snRNAs, snoRNAs, tRNAs and other ncRNAs*

157 Like the ribosome, the spliceosome is an important catalytic ribonucleoprotein machine. Numerous Psi sites
158 have been reported in the spliceosomal snRNAs U1, U2, U4, U5, U6, U12, U4atac and U6atac (Yu et al. 2011; Adachi and
159 Yu 2014). Although our protocol relied on polyA enriched RNA, we identified potential pseudouridine modifications on
160 four spliceosomal snRNAs, including 21 on U1, 11 on U2, 2 on U5 and 5 on U6 (**Table S4 in File S1**). Six of these (Psi5,
161 Psi6, Psi22 on U1, Psi44, Psi45, Psi55 on U2) were previously reported or predicted in Drosophila snRNAs (Deryusheva
162 and Gall 2009; Huang et al. 2005). Additional sites (i.e. Psi46 on U5 and Psi40 on U6) were reported at the
163 corresponding locations in human (Karijolich and Yu 2010) but have not been previously identified in Drosophila
164 spliceosomal snRNA.

165 7SK is another important and highly conserved snRNA. 7SK is pseudouridylated at position U250 by the DKC1-
166 box H/ACA RNP in human and this pseudouridylation was shown to be critical to stabilize 7SK snRNP and its function as
167 a critical regulator of the homeostasis and activity of P-TEFb, a key regulator of RNA polymerase (pol) II transcription
168 which stimulates the elongation phase (Zhao et al. 2016). This modification was identified from previous Psi-seq study
169 (Carlile et al. 2014). The 7SK snRNA homolog and a similar P-TEFb control system function in *Drosophila* (Nguyen et al.
170 2012; Gruber et al. 2008). Our pseudo-seq analysis identified a single nucleotide pseudouridylation modification in d7SK
171 snRNA, at U269 (**Table S4 in File S1**).

172 We found 10 Psi sites on the U3 RNA, a box C/D snoRNA required for rRNA processing. All the sites are distinct
173 from the four Psi sites found on human U3 using mass spectrometry (Yamaki et al. 2020). There is also a single Psi site
174 identified on snoRNA:Psi 28S-3342. Another 58 sites are localized on other ncRNAs, including 47 on 7SL, an abundant
175 cytoplasmic RNA which functions in protein secretion as a component of the signal recognition particle (SSR).
176 Previously only one Psi site (Psi211) has been reported on human 7SL. There are 11 sites on CR40469, an
177 uncharacterized lncRNAs (**Table S4 in File S1**).

178 Our data indicates 8 Psi sites in Drosophila tRNAs (**Table S4 in File S1**), 3 of which (Psi13 and Psi54 of Glu-TTC
179 and Psi27 of Val-CAC) were previously mapped on tRNAs of *D. melanogaster* (Machnicka et al. 2014).

180 All the Psi sites found on the non-coding RNAs, except for those on cytoplasmic and mitochondrial rRNAs in
181 **Table S3 in File S1**, are listed in **Table S4 in File S1** and can be sorted by the RNA biotype (column F). Note that more
182 non-coding RNA sites are likely to be identified using targeted approaches that do not rely on polyA enrichment.

183 *Psi modifications of mRNA transcripts*

184 It was previously unknown if Psi occurs on *Drosophila* mRNAs. Our analysis identified 1147 mRNA Psi sites
185 including 287 highly reproducible sites, 489 sites with intermediate reproducibility and 371 with low reproducibility
186 (**Figure 3A, Table S5 in File S1**). Interestingly, a large proportion of pseudouridylated mRNAs have multiple sites
187 identified in their transcripts (**Figure 3C**). Indeed, the 1147 Psi sites are localized on the transcripts of only 165 genes
188 with transcripts from some loci containing more than 100 sites (**Figure 3C**, **Figure 3D**, **Table S5 in File S1**).
189 In *Drosophila*, the ribosomal large subunit (RpL, 60S) consists of three rRNAs (2S, 5.8S, 28S) and 47 ribosomal
190 proteins whereas the small ribosomal subunit (RpS, 40S) consists of one ribosomal RNA (18S) and 33 ribosomal proteins
191 (Anger et al. 2013). Our analysis identified 202 Psi sites in mRNAs encoding 35 of 47 RpL proteins and 23 of 33 RpS
192 proteins. As with other mRNAs, most mRNAs encoding RPs contain at least two Psi sites. Although the vast majority
193 the ribosomal proteins are pseudouridylated, we found no Psi modifications on the mitochondrial ribosomal protein
194 mRNAs even though these were abundantly expressed in our RNA samples. Altogether, the 202 Psi sites on ribosomal
195 protein RNAs accounted for 17.6% of total pseudouridines identified on mRNAs. In addition, if adding together the 735
196 Psi sites on the cytoplasmic rRNAs and 202 sites on various ribosomal protein mRNAs, nearly 44% of the Psi sites that
197 we have identified are related to the production of cytoplasmic ribosomes.
198 Among the mRNA Psi sites, a total of 99 Psi sites were found on protein-coding genes of the mitochondrial
199 genome (**Figure 3A**, **Table S5 in File S1**). Eleven of the 13 protein encoding mRNAs in the mitochondrial genome were
200 pseudouridylated. Considering the 125 Psi sites found on the mitochondrial rRNA (**Table S3 in File S1**), we found a total
201 of 224 Psi sites on the RNAs encoded from mitochondrial genome (10.5% of the total sites), thus indicating that RNAs
202 from mitochondrial genome are significantly pseudouridylated in *Drosophila*.
203 As noted above, many of the mRNA transcripts have multiple Psi sites. Interestingly, Psi sites are found across
204 functionally related mRNA transcripts. For instance, the mRNAs encoding the yolk proteins, which are known to be
205 highly expressed in the fat bodies of adult female head (Fujii and Amrein 2002), contain the highest number of
206 pseudouridine modification of the female head transcriptome. A total of 338 Psi sites (25.5% of total mRNA Psi sites)
207 were identified on mRNAs encoding the three yolk proteins (Yp1, Yp2, and Yp3). Each of the yolk protein
208 mRNAs harbors many Psi sites (118 sites on *Yp1*, 104 on *Yp2* and 116 sites on *Yp3*) (**Figure 3E**, **Table S5 in File S1**).
209 Notably, the majority of the Psi sites are localized in the coding region and are almost absent from the 3′ UTR region of
210 the transcripts, which confirms the specificity of our Psi-seq analysis and strongly suggests for an important function of
211 pseudouridylation in the coding region for the yolk protein transcripts.
212 Transcripts of multiple genes involved in sensory function in the nervous system are also marked with Psi
213 modifications. A notable example is *ninaE*, the structural gene encoding the *Drosophila* Rhodopsin-1, with 75 Psi sites.
214 Indeed, a number of transcripts bearing multiple Psi sites play various roles in photoreceptor function, including other
215 *rhodopsins* (*Rh*) (*Rh3*: 2 Psi sites, *Rh4*: 2 Psi sites, *Rh6*: 2 Psi sites), *retinin* (18 Psi sites), and *Arrestins* (*Arr1*: 39 Psi sites,
216 *Arr2*: 53 Psi sites). Several *Odorant-binding protein* mRNAs, which function in the sensory perception of smell, also have
217 multiple Psi sites (*Obp19d*: 3 Psi sites, *Obp44a*: 12, *Obp56d*: 3, *Obp56e*: 2, *Obp99c*: 1). It is interesting that mRNAs

218 encoding two potential neuropeptide genes *Nplp2* and *Nplp3* contain 28 and 23 Psi sites, respectively (**Table S5 in File**
219 **S1**).

220 Previously, an N3-CMC-enriched pseudouridine sequencing identified pseudouridylation at U519 of the human
221 elongation factor *eEF1A1* mRNA which encodes the main protein that delivers aminoacyl-tRNAs to the ribosome. In
222 fact, *eEF1A1* is the first experimentally validated human mRNA psi site (Li et al. 2015). From our Psi-seq data, sites were
223 identified on mRNA for two of the *Drosophila* translation elongation factors (*eEF1α1* with four sites and *eEF1γ* with one
224 site) (**Table S5 in File S1**).

225 The majority of Psi sites (1006 sites, 87.7%) found on mRNAs were within the coding region. The 5′ untranslated
226 regions (UTRs) had only 98 (8.5%) sites and 61 sites (5.3%) were in 3′ UTRs (**Figure 3B**). In coding regions, Psi sites were
227 more commonly found in the second position of the codon relative to the first or third positions. For nuclear-encoded
228 mRNA, the most highly pseudouridylated codons are the CΨG and UΨG for Leucine, followed by GCΨ for Alanine and
229 ΨGG for Tryptophan (**Table 1**). Interestingly, the pattern of pseudouridylation of codons in mitochondrial genome
230 transcripts appeared to be distinct from that of nuclear genome transcripts, with the most frequent Psi-containing
231 codons encoding for Phenylalanine (F) (i.e. UUΨ and UΨU), Isoleucine (I) (i.e. AΨT), as well as Leucine (but in the
232 mitochondrial case, ΨTA was the most prominent codon for Leucine) (**Table 1**).

233 *Pseudouridylation of the UGA codon in Drosophila mitochondria*

234 A previous study proposed that Psi in mRNA stop codons may result in the suppression of translation
235 termination and enhancement of translational read-through *in vitro* protein synthesis (Karijolich and Yu 2011). We were
236 very interested in this as a set of transcripts in the *Drosophila* transcriptome are predicted to undergo stop codon
237 suppression (Jungreis et al. 2011). However, our analysis failed to identify any Psi sites on stop codons of nuclear
238 encoded genes (**Table 1**, **Table S5 in File S1**). This result suggests that pseudouridylation is unlikely to be an important
239 underlying mechanism for stop-codon readthrough in *Drosophila*.

240 However, we did find Psi sites present on the canonical "stop codon" UGA in mitochondrial encoded transcripts.
241 Out of the 68 total UGAs in mitochondrial mRNA, 5 were called as ΨGA from our analysis. This was in contrast to 3,321
242 UGAs that were present among the nuclear mRNA that did not contain a signal for ΨGA. The frequency of ΨGA codons
243 in mitochondrial RNA was significantly higher than in nuclear encoded RNA (Fishers exact test p-value = 2.81e-9).
244 Importantly, the UGA codon in mitochondria of insects and in vertebrates does not encode a stop, but instead codes for
245 tryptophan. Our analysis suggests that this codon, at least in some cases, is actually ΨGA.

246

247

248 *mRNA pseudouridylation is correlated with abundance*

249 The yolk protein mRNAs are among the most highly expressed transcripts in our female fly head libraries and
250 these were the most highly pseudouridylated genes. mRNAs encoding ribosomal proteins are also abundantly
251 expressed and were among the classes of genes with a high likelihood of Psi modification in their transcripts. Thus, we
252 explored the relationship between Psi occurrence and RNA expression level in CMC library (**Figure 4A)**. Interestingly,

253 for mRNAs there appeared to be a strong positive correlation between mRNA expression level and the number of Psi
254 sites in a particular transcript. The majority of the Psi sites are identified on the transcripts of the genes which are
255 abundantly expressed (**Figure 4B**). However, this was unlikely to be caused by a limitation coverage in calling sites
256 because we found no correlation between the read coverage and the presence of Psi sites for the non-coding RNAs,
257 such as snRNA, snoRNA, tRNA and other nc-RNA (**Figure 4A**). Our results don't allow us to infer a cause and effect for
258 the frequency of Psi on *Drosophila* mRNAs. Highly expressed mRNAs may be more stable due to the presence of Psi
259 modification, or they may be more likely to be pseudouridylated because they are more highly expressed, or both
260 factors may contribute to varying degrees.

261 *Transcriptome-wide identification of potential pseudouridylation targets of RluA-2*

262 Previously, the putative pseudouridine synthase RluA-2 was found to be broadly expressed in *Drosophila* and
263 loss of function mutation in *RluA-2* was found to cause hypersensitivity in larval pain pathways (Song et al. 2020). Thus,
264 we sought to identify potential targets for RluA-2 by comparing the pseudouridine profiles of *RluA-2* mutant mRNA
265 with w^{1118} control mRNA. A total of 423 Psi sites were not detected in any of the four *RluA-2* mutant libraries. These
266 missing sites therefore represent potential targets of RluA-2 (**Table S6 in File S1**). The candidate RluA-2 target sites
267 include 74 pseudouridylation sites on cytoplasmic and mitochondrial rRNAs (63 cellular and 11 mitochondrial), 5 on
268 snRNAs, 5 on tRNA, 9 on other ncRNAs and 328 sites on mRNAs (**Figure 5**). The mRNA Psi sites include 65 of the sites
269 encoding for components of ribosomal proteins and 42 mitochondrial genome-encoded transcripts. Among the 165
270 genes where the pseudouridylation sites were identified on w^{1118} transcripts (**Table S5 in File S1**), 115 genes were
271 missing Psi modifications in the *RluA-2* mutant. That the sites were missing in various classes of RNA suggests that the
272 RluA-2 enzyme acts on a wide range of substrates (**Table S6 in File S1**).

273 To determine if RluA-2 recognizes a particular sequence motif we ran the target sites through a motif finding
274 algorithm. Other than a dominant G downstream of the pseudouridine site there was not a significant sequence logo
275 around the modification site (**Figure S3**). This finding suggests that RluA-2 enzyme may not recognize a primary
276 sequence motif in site selectivity and may indicate that secondary structure of RNAs is critical for RluA-2 target
277 recognition (as has been found for other pseudouridine synthases (Purchal et al. 2022)).

278 *Loss of RluA-2 results in differential expression in translation-related genes and non-coding RNAs*

279 To investigate the effects of *RluA-2* on gene expression we compared the transcriptome of *RluA-2* mutant to
280 the control genetic background $isow^{1118}$ in mock treated libraries of experiment 2. The Euclidean distance of the libraries
281 demonstrated a clear clustering of samples based on genotype **(Figure S4).** A total of 2214 genes showed significant
282 change in their expression levels in the *RluA-2* vs. $isow^{1118}$ samples (FDR<0.05), including 1048 that were upregulated
283 and 1166 that were downregulated in the mutant (**Figure 6A, Table S7 in File S1**). The vast majority of differentially
284 expressed genes (94.8%) were mRNA-coding genes, which included 994 of the upregulated and 1104 of the
285 downregulated genes. Among the upregulated gene set, we found 35 RpL genes and 22 RpS genes, three elongation
286 factors and 11 initiation factors (**Figure 6B, Table S7 in File S1**). It is notable that *Drosophila* Myc (dMyc), which is a
287 major regulator of genes involved in ribosome biogenesis and translation (Gallant 2013), as well as its binding partner
288 dMax, were both upregulated in *RluA-2*. Consistent with these observations, a gene set enrichment analysis of the

differentially expressed genes showed significant enrichment in the ribosome (FDR < 0.0001) and ribosome biogenesis (FDR = 0.154) (**Figure 6C**). These results suggest a potential impact on the translation and translational machinery that is caused by loss of RluA-2.

Notably, 77 non-coding RNAs (non-protein coding anti-sense RNAs and long non-coding RNAs) were either upregulated (47) or downregulated (30) and several snoRNAs responsible for guiding pseudouridylation or methylation on 18S and 28S rRNAs were among the differentially regulated RNAs (**Table S7 in File S1**).

We investigated impact of loss of RluA-2 on the other eight putative pseudouridine synthases. Expression of *RluA-1*, the paralog of *RluA-2*, was slightly decreased which may be a consequence of loss of regulatory DNA in the *RluA-2* deletion. In contrast, *mfl* the human *Dyskerin* homolog, and *CG3709* (encoding the putative *Drosophila* Pus10), were upregulated although the fold change in expression was also slight (**Table S7 in File S1**). The expression of the other putative pseudouridine synthases is either not changed (*CG3045*, *CG4159/PUS1*, *CG34140*) or not detected in female head (*CG7849* and *CG6745/PUS7*).

Finally, we investigated whether there might be a correlation between RluA-2 dependent pseudouridylation and gene expression. If such a correlation exists, transcripts that lose sites in the mutant might show altered transcript levels. Interestingly, 67 genes with missing Psi sites in the *RluA-2* mutant also show differential expression (**Table S8 in File S1**). Among the RluA-2 targets, ribosomal protein component genes are also differentially expressed (18/21 *RpL* and 9/15 *RpS* genes). The differentially expressed mRNAs encoding ribosomal proteins all showed upregulation in the *RluA-2* mutant. A similar trend of upregulation was also observed in other RluA-2 target genes, including yolk protein mRNAs (*Yp1, Yp2, Yp3*). However, some transcripts missing Psi sites were found to be reduced in expression (i.e. *Nplp3*, *Obp19d*, *Obp44a*). The vast majority of the differentially expressed genes did not contain RluA-2 dependent Psi sites (2147, 97.0%), suggesting the majority of the differentially expressed genes do not change expression levels as a consequence of a missing pseudouridylation. This may indicate that *RluA-2* can directly affect RNA levels independently of its pseudouridine synthase activity, or the gene expression alterations may be a downstream or indirect consequence of a lack of *RluA-2* activity. Future studies investigating the RNAs that are directly bound by RluA-2 will be necessary to disentangle these alternatives.

Discussion:

For over a century studies of *Drosophila melanogaster* have led to countless insights into the underlying principles of genetics and molecular biology including several studies of pseudouridylation of RNA. Nevertheless, next generation sequencing has yet to be applied to thoroughly characterize *Drosophila* Psi. We have described here the first attempt at characterizing a *D. melanogaster* pseudouridine epitranscriptome. Our optimized library construction and bioinformatics workflow has allowed us to robustly detect previously known sites as well as to uncover a set of previously unknown Psi sites that are reproducible across biological replicates. Although our analysis has been limited to a single stage of the fly life cycle, and a single sex, our findings provide interesting insights for future studies.

Perhaps the most striking finding of our study is the widespread presence of Psi in the RNAs that encode for the many components of the cellular ribosome. Over 500 highly reproducible sites were found in rRNA which is significantly

325 more than the number of previously identified sites. The vastly increased number of candidate rRNA sites that we have
326 detected is likely due to the statistical power that is provided with next generation sequencing methods. A caveat to
327 consider, is that our methods do not allow for the explicit determination of the proportion of pseudouridylated bases
328 that occur at individual sites. Even for highly reproducible sites, it is likely that these sites occur on only a fraction of
329 rRNA molecules and additional studies will be needed for quantitative determination of site frequency. This could
330 potentially be performed with recently described HydraPsiSeq methodology (Marchand et al. 2020). Alternatively,
331 nanopore sequencing approaches would potentially allow for determination of Psi sites that occur together on
332 individual rRNA molecules (Begik et al. 2021).

333 In addition to sites on rRNA, we found that the vast majority of mRNAs encoding ribosomal proteins were
334 pseudouridylated. Although the function of the Psi sites on *Drosophila* mRNAs also awaits further analysis, an
335 interesting possibility is that codons containing Psi could favor the incorporation of non-cognate amino acids during
336 mRNA translation (Eyler et al. 2019). We found 160 Psi containing codons across mRNAs encoding for 58 ribosomal
337 proteins. If each of these codons facilitates some degree of non-cognate amino acid incorporation (Eyler et al. 2019),
338 this would allow for vast complexity of ribosomal protein isoforms. We hypothesize that such a diversity of ribosomal
339 protein subunits might allow for robustness in ribosomal function across a variety of environmental conditions which
340 could be advantageous for a poikilotherm such as *Drosophila*. Alternatively, ribosomal protein diversity could
341 contribute to cell type specific functions of the ribosome, or facilitate translation of particular mRNAs as previously
342 described for other forms of ribosome heterogeneity (Shi et al. 2017; Genuth and Barna 2018; Hopes et al. 2022).

343 The most highly pseudouridylated mRNAs (*Yp1, Yp2,* and *Yp3*) encode for the Yolk proteins. These transcripts
344 are expressed in fat bodies of the female head which is one of the places where yolk protein biogenesis is thought to
345 occur (Fujii and Amrein 2002). An important aspect of fly vitellogenesis involves synthesis and secretion of the yolk
346 proteins from fat bodies into the hemolymph circulation for transport to the ovaries (Hames and Bownes 1978). High
347 levels of pseudouridylation, especially in the coding region of the Yp transcripts, may facilitate message stability,
348 translation, or possibly plays a role in targeting the mRNA to secretory pathways. Consistent with the latter possibility,
349 the RNA with the highest percentage of Psi was the 7SL RNA which is a component of the Signal Recognition Particle,
350 the key factor in guiding mRNAs to the secretory compartments of the cell. Other mRNAs for secreted and
351 transmembrane proteins that identified in our analysis include rhodopsins (*ninaE*, *rh4*, *rh6*), *trp,* and odorant binding
352 proteins (*Obp44a*, *Obp56d*, *Obp56e*, *Obp99c*, *Obp19D*).

353 We found that transcripts expressed from the mitochondrial genome (*lrRNA, srRNA, CoI, CoII, CoIII, AtpAse6,*
354 *AtpAse8, ND2, ND5, Cyt-b)* and some nuclear encoded mitochondrial genes (*ATPsynB*, *ATPsynbeta*, *ATPsynC*, *ATPsynF,*
355 *Cyt-c-p*, *COX6B*, *COX7C*, *UQCR-C2*, *COX5A*, and *porin*) were pseudouridylated. Interestingly, the opal (UGA) cellular stop
356 codon was pseudouridylated in 5 cases on mitochondrial transcripts and we propose that this may facilitate
357 incorporation of tryptophan at ΨGA codon as part of the mitochondrial genetic code. Indeed, our finding that the ΨGA
358 codon is specific to mitochondrially encoded RNAs provides evidence in support of specificity of the Psi sites that we
359 have identified. If our method were to identify spurious sites, we would expect to find ΨGA codons in both nuclear and
360 mitochondrially encoded genes. Other potential Psi sites that were absent from our analysis are further evidence of

361 specificity (**Table 1**). Notably, we found no instances for Psi on the amber (UAG) or ochre (UAA) stop codons. Indeed,
362 many codons were detected at lower levels than expected if Psi were equally distributed across uridine nucleotides
363 (**Table 1)**.

364 An interesting pattern is apparent for nuclear and mitochondrial encoded genes and their pseudouridylation of
365 codons for Leucine. All of the seven instances for the TΨA codon (i.e. TTA Leucine) were found on mitochondrially
366 encoded mRNAs. We also found evidence for ΨTA modification of the TTA codon, and in 12/13 cases this modification
367 was identified on mitochondrial encoded transcripts. Interestingly, on nuclear encoded transcripts, the CUG codon for
368 Leucine was the codon that was most enriched for Psi. As the CUG codon for Leucine is not found at all in
369 mitochondrially encoded genes, all of the instances of CΨG codons are in nuclear encoded genes (as expected). Thus, in
370 mitochondria there is an enrichment for Psi in the UUA codon for leucine but for nuclear encoded genes the enrichment
371 is for the CUG codon for leucine. We hypothesize that this use of Psi by the cell may contribute to the appropriate
372 translation of biased codons for Leucine on mitochondrial and cellular ribosomes respectively. We note that Leucine
373 (along with Serine and Arginine) is unusual among amino acids in that it can be encoded by six codons (using the
374 canonical bases A, C, U and G). If one considers that there are 11 possible Psi containing codons for Leucine, the total
375 number of potential codons for Leucine expands to 17. The mechanistic importance of this level of diversity for Leucine
376 codons poses an interesting puzzle for the future.

377 Finally, we found that 423 Psi sites were absent from the *RluA-2* mutant RNA and these sites are thus potential
378 targets for the RluA-2 enzyme. The potential RNA targets of RluA-2 included rRNAs, mRNAs, snRNAs and tRNAs.
379 Among mRNAs, notable targets for RluA-2 included *arrestin2 (Arr2)* mRNA which was missing 24/42 sites and
380 *neuropeptide like protein 3 (NPLP3*) was missing 11/14 sites. These two targets stand out as potentially important in the
381 previously described hypersensitive nociception phenotypes of *RluA-2* mutant larvae (Song et al. 2020). Ribosomal
382 protein mRNA targets of RluA-2 are also of note as 65/160 sites on mRNAs encoding for ribosomal proteins were absent
383 in the mutant background. Interestingly, many of these same ribosomal protein mRNAs showed upregulated
384 expression in the mutant RNA libraries. We look forward to future studies that may further elaborate the impact of
385 RluA-2 dependent pseudouridylation on neuronal and ribosomal function.
386

387 **Figure legends**

388 **Figure 1 A**. Schematic of *Drosophila melanogaster* Psi-seq library construction. RNA was extracted from heads of
389 unmated female flies. An *in vitro* transcribed RNA spike-in containing a single pseudouridine at position 43 was added
390 into the RNA samples immediately after polyA selection and the optimized CMCT derivatization step (**Figure S1**) was
391 carried out prior to RNA fragmentation and subsequent library preparation. Libraries were barcoded and pooled for the
392 final Illumina sequencing. **B**. A connected box plot showing the termination fold change (log2) on each of the nucleic
393 acid position along the synthetic RNA spike-in added to the genetic background w^{1118} libraries (n=7) used for
394 pseudouridine mapping. With the optimized CMCT treatment (**Figure S1**), the termination fold change peak at the only
395 Psi (position 43) of the RNA spike-in (highlighted with the red box plot) stands out among all the other positions.

396

397 **Figure 2**. Location of pseudouridine modification sites (circled U) on cytoplasmic large subunit rRNA secondary
398 structure in *Drosophila melanogaster*. Psi residues are circled in red for high reproducible sites and in blue for previously
399 identified Offengand sites. The residues are color coded according to the "fine grained onion" scheme in which close
400 proximity to the peptidyl transferase center is indicated by cooler colors (i.e. blue) and sites far from the center are
401 indicated by warmer colors.

402

403 **Figure 3**. Features of pseudouridine sites identified from *Drosophila melanogaster* transcriptome. **A**. Psi site distribution
404 according to RNA class (mRNA, rRNA, tRNA, snRNA, snoRNA and other ncRNA). **B.** Distribution of Psi sites on mRNA
405 transcript regions (5'-UTR, coding, 3'-UTR and intron). **C.** Distribution of number of Psi sites identified per gene. **D.** The
406 10 coding genes with the highest number of Psi sites. For each gene, the numbers of Psi sites and Uridines in the
407 sequence are listed along with the normalized gene expression represented as average RPKM of the *w1118* CMC
408 libraries (n=7). **E.** Psi sites on yolk protein mRNAs. The locations of Psi sites identified on mRNAs for yolk
409 protein *Yp1*, *Yp2*, and *Yp3* are indicated with bars above the schematic of the transcripts. UTRs are marked with gray
410 shading whereas coding regions are not shaded. The height of the bars indicated by the Y axis represents the number of
411 w^{1118} replicates a Psi site is identified from (n=2~7) while red, blue and green color of the lines represents Psi sites of
412 high, intermediate, and low reproducibility, respectively.

413

414

415 **Figure 4.** Correlation between gene expression and Psi site identification. **A**. A scatter plot of gene expression *versus*
416 proportion of Psi site across pseudouridylated transcripts. The gene expression is represented as the average of log10
417 (RPKM) in w^{1118} CMC libraries (n=7) and the proportion of Psi sites is measured by the percentage of Psi sites called on
418 the total number of Uridines in various categories of RNA species. The individual data points representing the yolk
419 proteins and 7SL RNA are labeled. **B.** A plot of gene expression *versus* the number of genes with (red) or without (blue)
420 Psi sites identified on their transcripts.

421

422 **Figure 5**. Pie chart showing the distribution of number of RluA-2-dependent Psi sites in different categories of RNA
423 species (mRNA, rRNA, tRNA, snRNA, snoRNA and other ncRNA). Those potential RluA-2 target sites are missing from
424 *RluA-2* mutant compared to the w^{1118}.

425

426 **Figure 6**. Effect of loss of RluA-2 activity on gene expression in the female *Drosophila* head transcriptome. **A**. Volcano
427 plot showing the differentially expressed genes in the mock Psi-seq libraries of *RluA-2* vs w^{1118}. The $\log_2$ fold change
428 indicated the mean expression level for each gene. Each data point represents an individual gene. Red data points
429 indicate upregulated genes with $\log_2$fold change>=1.5, the blue data points indicate downregulated genes with $\log_2$fold
430 change<=1.5 in *RluA-2* vs w^{1118} (FDR<=0.05) and grey data points represent all the other genes. **B.** A hierarchically
431 clustered heatmap showing the expression pattern of the top 100 significantly differentially expressed genes between

432 *RluA-2* vs w^{1118}. Gene expression values quantified as the counts of reads mapping to the exon regions were
433 transformed by applying DESeq2's regularized-logarithm transformation. The rows representing the genes were sorted
434 in descending order of their average expression among the *RluA-2* libraries. The columns representing the individual
435 libraries were hierarchically clustered based on the gene expression. The color density on the heatmap indicates the
436 extent of deviation for each gene in each replicate library from the gene's average expression across all libraries. Yellow
437 and blue represent up- and down- regulated expression in *RluA-2*, respectively. **C.** Gene set enrichment plots show that
438 the Gene Ontology categories are enriched for the Ribosome (left) and Ribosome Biogenesis in Eukaryotes (right) in
439 the *RluA-2* mutants. The y-axis in the top portion represents enrichment score (ES) and on the x-axis are genes (vertical
440 black lines) represented in genes sets. The green line connects points of ES and genes. ES is the maximum deviation
441 from zero as calculated for each gene going down the ranked list, and represents the degree of over-representation of a
442 gene set at the top or the bottom of the ranked gene list. The middle portion shows where the members of the gene set
443 appear in the ranked list of genes. The colored band represents the degree of correlation of genes with *RluA-2* (red for
444 positive and blue for negative correlation). The ranking metric at the bottom portion shows a gene's correlation with
445 *RluA-2* (positive value) or w^{1118} (negative value). The significantly enriched gene sets are DME03010 RIBOSOME (p value
446 = 0.000; FDR = 0.000) and DME03008 RIBOSOME BIOGENESIS IN EUKARYOTES (p value = 0.000; FDR = 0.154).

Table 1. Distribution of Psi (Ψ) sites on nuclear (left) or mitochondrial (right, shaded) genome encoded codons and amino acids in *Drosophila melanogaster*.

	Encoded by nuclear genome [a]						Encoded by mitochondrial genome [b]				
Codon[1]	Amino acid[2]	Abundance[3]	Expected Count[4]	Psi count[5]	Adj. pval[6]	Codon[1]	Amino acid[2]	Abundance[3]	Expected Count[4]	Psi count[5]	Adj. pval[6]
ΨAA	STOP	5,749	1	0	1.00	ΨAA	STOP	9	0	0	1.00
ΨAG		4,827	1	0	1.00	ΨAG		1	0	0	1.00
ΨGA		3,321	1	0	1.00	ΨGA	W	68	3	5	1.00
GCΨ	**A**	107,656	26	45	1.29E-02*	GCΨ	A	88	4	3	1.00
ΨGC	C	99,092	24	21	1.00	ΨGC	C	1	0	0	1.00
ΨGU		43,290	5	9	1.00	ΨGU		20	0	1	1.00
UGΨ		43,290	5	2	1.00	UGΨ		20	0	0	1.00
GAΨ	**D**	207,714	51	23	2.93E-04***	GAΨ	D	36	2	0	1.00
ΨUC	**F**	160,233	20	31	2.77E-01	ΨUC	**F**	13	0	0	1.00
UΨC		160,233	20	14	1.00	UΨC		13	0	0	1.00
ΨUU		105,347	9	10	1.00	ΨUU		219	4	3	1.00
UΨU		105,347	9	12	1.00	UΨU		219	4	9	1.73E-01
UUΨ		105,347	9	11	1.00	**UUΨ**		219	4	12	3.51E-03**
GGΨ	**G**	96,329	24	22	1.00	GGΨ	G	28	1	1	1.00
CAΨ	**H**	79,960	20	14	1.00	CAΨ	H	51	3	0	1.00
AΨA	**I**	75,401	18	1	7.44E-06***	AΨA	I	129	6	0	6.13E-02
AΨC		164,918	40	7	2.99E-09***	AΨC		13	1	0	1.00
AΨU		127,100	16	21	1.00	AΨU		249	6	12	4.38E-01
AUΨ		127,100	16	30	1.54E-02*	AUΨ		249	6	8	1.00
CΨA	**L**	64,763	16	2	7.87E-04**	CΨA	L	18	1	0	1.00
CΨC		102,941	25	6	2.00E-04***	CΨC		0	0	0	1.00
CΨG		281,351	69	201	9.12E-42***	CΨG		0	0	0	1.00
CΨU		70,463	9	13	1.00	CΨU		22	1	0	1.00
CUΨ		70,463	9	10	1.00	CUΨ		22	1	1	1.00
ΨUA		36,665	4	1	1.00	ΨUA		335	8	12	1.00
UΨA		36,665	4	0	5.31E-01	UΨA		335	8	7	1.00
ΨUG		124,790	15	33	9.98E-04**	ΨUG		9	0	1	1.00
UΨG		124,790	15	46	1.65E-09***	UΨG		9	0	1	1.00
AΨG	**M**	174,201	43	48	1.00	AΨG	M	6	0	1	1.00
AAΨ	**N**	162,114	40	11	2.22E-06***	AAΨ	N	125	6	5	1.00
CCΨ	**P**	54,987	13	12	1.00	CCΨ	P	56	3	0	1.00
CGΨ	**R**	64,053	16	19	1.00	CGΨ	R	7	0	0	1.00
AGΨ	**S**	90,230	22	3	2.02E-05***	AGΨ	S	17	1	1	1.00
ΨCA		61,588	15	1	1.93E-04***	ΨCA		77	4	0	9.50E-01
ΨCC		144,883	36	12	1.65E-04***	ΨCC		6	0	0	1.00
ΨCG		120,743	30	9	3.59E-04***	ΨCG		3	0	0	1.00
ΨCU		54,711	7	4	1.00	ΨCU		74	2	3	1.00
UCΨ		54,711	7	9	1.00	UCΨ		74	2	1	1.00
ACΨ	**U**	76,117	19	8	2.15E-01	ACΨ	U	73	4	2	1.00
GΨA	**V**	49,291	12	0	2.52E-04***	GΨA	V	65	3	1	1.00
GΨC		100,296	25	8	4.15E-03**	GΨC		1	0	0	1.00
GΨG		204,607	50	68	2.55E-01	GΨG		6	0	0	1.00
GΨU		84,822	10	22	2.19E-02*	GΨU		52	1	2	1.00
GUΨ		84,822	10	18	4.81E-01	GUΨ		52	1	2	1.00
ΨGG	**W**	74,214	18	32	4.79E-02*	ΨGG	W	2	0	0	1.00
ΨAC	Y	136,903	34	23	1.00	ΨAC	Y	16	1	0	1.00
ΨAU		85,677	10	6	1.00	ΨAU		87	2	0	1.00
UAΨ		85,677	10	9	1.00	UAΨ		87	2	5	1.00

[a] Distribution of Psi in codons of the transcripts encoded by nuclear genome of *Drosophila melanogaster.*

[b] Distribution of Psi in codons of the transcripts encoded by mitochondrial genome of *Drosophila melanogaster.*

[1] Codon: Codon with one of the Uridines substituted with Psi site. The codons with significant Adj.pval are highlighted in bold.

[2] Amino acid: Product of translation. The amino acids/stop codons with significant Adj.pval (not shown) are highlighted in bold.

[3] Abundance: Count of the given codon in the transcriptome.

[4] Expected count: An expectation for the count of Psi sites associated with the given codon is estimated from the observed count by multiplying the total number of Psi sites on mRNA with the ratio of abundance of the given codon divided by the count of Uridines to the total count of codons in the transcriptome.

[5] Psi count: Count of the Psi sites overlapping with the given codon. When multiple transcripts of a given gene are involved, the codon from the longest transcript is considered.

[6] Adj_pval: adjusted p value. Assuming a binomial distribution for the observed Psi count, a p value indicating the cumulative probability of the count of Psi sites associated with the given codon being at least the observed count if found greater than the expected count (or at most the observed count if found less than the expected count) was calculated. The p values are then corrected for multiple testing using Bonferroni's method.

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

461 **Supplemental Materials**

462 **Supplemental Figures**

463 **Figure S1.** Optimization of CMCT treatment conditions for Psi-seq library preparations. The polyA-selected RNA was
464 subjected to either fragmentation ("Fragmented before treatment") or not ("Non-fragmented before treatment") prior
465 to the CMC/mock derivatization. The derivatization was carried out with CMC/mock treatment for 1hr or 18hr at 37°C
466 ("1hr CMC/mock" or "18hr CMC/mock"), followed by alkaline treatment for 6hr at 37°C or 75°C ("37°C hydrolysis" or
467 "75°C hydrolysis"). The Y axis represents the termination fold change while the X axis is the nucleotide position on the
468 RNA spike-in sequence. The termination fold change from the only pseudouridine at position 43 on the RNA spike-in is
469 marked with a red line whereas those from other sites are presented as black lines. The optimum condition producing
470 the highest termination fold change for Psi site and lowest background is framed with a dashed red rectangle.

471

472 **Figure S2.** The termination fold change (y axis) distribution over the large subunit (LSU) of cytoplasmic rRNA in
473 *Drosophila melanogaster* (x axis). The Psi sites identified are called with the threshold of appearing in 6 or all 7 w^{1118}
474 libraries (high reproducible site, red), 4 or 5 libraries (intermediate reproducible site, blue), and 2 or 3 libraries (low
475 reproducible site, green) from both sets of the experiments. The previously reported Psi sites on LSU identified by
476 Ofengand et al (1997) were indicated by the gray vertical lines.

477

478 **Figure S3.** Consensus sequence for the RluA-2 target sites. The X axis is the +/- 10 nucleotide sequence flanking the T
479 site where the RNA bears the pseudouridylation modification (0 position). The Y axis represent the bits score. The
480 overall probability of a nucleotide is represented by the height of the letter.

481

482 **Figure S4.** Euclidean distances of w^{1118} and *RluA-2* mutant mock libraries clearly demonstrate the clustering of samples
483 based on the genotypes. Color key indicates level of similarity between libraries.

484

485 **Supplemental File S1**

486 **Table S1**. A summary of the Psi-seq libraries (Table S1) and an overview of contents included in supplemental tables
487 (Table S2 - Table S8) for transcriptome-wide mapping of pseudouridylation and transcriptome expression analyses in
488 *Drosophila melanogaster*.

489 **Table S2**. The quality features of the sequencing data used for Psi-seq and transcriptome expression analyses in
490 *Drosophila melanogaster*.

491 **Table S3**. Psi sites identified using Psi-seq on cytoplasmic and mitochondrial rRNA in *Drosophila melanogaster*. Note
492 the cytoplasmic LSU rRNA sites reported previously are labeled as "Ofengand site" in column H.

493 **Table S4**. Psi sites detected on other types of ncRNA with Psi-seq, including tRNA, snRNA, snoRNA, and other ncRNA
494 in *Drosophila melanogaster*.

495 **Table S5**. Psi sites detected on protein-encoding mRNAs with Psi-seq in *Drosophila melanogaster*.

496 **Table S6**. Potential RluA-2 pseudouridylation targets in the adult head of *Drosophila melanogaster* with Psi-seq.
497 Pseudouridylation sites from *RluA-2* (n=4) and w^{1118} (n=4) samples from the same set of experiment (Exp 2) were
498 compared and those sites which are called in at least two libraries from both of the experiments (Exp 1 and Exp 2) in
499 w^{1118} but absent from *RluA-2* were listed as the putative RluA-2 targets.
500 **Table S7**. Differentially expressed genes in *RluA-2* compared to the genetic background isow^{1118} in the adult head of
501 *Drosophila melanogaster.* The mock samples from each Psi-seq data set were used for differential expression analysis
502 between the 4 replicates of *RluA-2* samples and 4 replicates of w^{1118} samples. Differentially expressed genes were
503 detected using DeSeq2 with the threshold of adjusted *p* value with IHW correction <0.05.
504 **Table S8**: A list of differentially expressed genes between *RluA-2* and w^{1118} which overlap with genes containing RluA-2
505 target Psi sites.
506
507 **Materials and Methods**
508 *Fly Strains and Raising Conditions*
509 The isogenic w^{1118} strain used as the control genotype was obtained from Bloomington stock center. The
510 mutant of *RluA-2*$^{del\text{-}FRT}$ used in the study was a complete deletion of the locus that was generated and previously
511 described (Song et al. 2020). All flies were reared on yeast, corn meal food in bottles or vials at an incubator with
512 controlled temperature (25°C) and humidity (70%) on a 12h light/12h dark cycle.
513 *RNA extraction from adult head tissue*
514 For Experiment 1, newly eclosed virgins from isow^{1118} were collected twice every day (once in the morning and
515 once in the afternoon) with less than 8 hours in between. For Experiment 2, newly eclosed virgins from isow^{1118} and
516 homozygous *RluA-2* $^{del\text{-}FRT}$ (Song et al. 2020) were collected in the similar way. The flies were put into vials with yeast
517 and Bloomington recipe fly medium (20-30 flies in each vial) and aged for 1d in the incubator. The flies were then snap
518 frozen in liquid nitrogen and kept at -80°C for 3 days until sufficient number of flies from each genotype had been
519 collected. The frozen flies were then decapitated and 100 heads per replicate homogenized in TRIzol reagent
520 (Invitrogen). Total RNA was prepared according to the manufacturers' recommendations.
521 *In vitro transcription of RNA "spike-in"*
522 The RNA spike-in sequence containing a single pseudouridine at position 43 as used previously (Schwartz et al.
523 2014) was *in vitro* transcribed from dsDNA template (synthesized as gBlocks gene fragment from IDT). *In vitro*
524 transcription was performed in a volume of 20ul with MEGAshortscript T7 Transcription kit (AM1354, Invitrogen), using
525 75nmol each of GTP, CTP, ATP and Psi-TP (TriLink Biotechnologies). The RNA product was purified with MEGAclear
526 Transcription Clean-Up kit (AM1908, Invitrogen) before adding into the Psi-seq libraries.
527 *CMCT treatment and RNA-seq library preparation*
528 Polyadenylated RNA was enriched from total RNA using two rounds of Oligo(dT) dynabeads (Invitrogen)
529 according to manufacturer's protocol. To each polyA-RNA sample, spike-in RNA (~3ug) was added and then the
530 samples were split into two equal samples for CMCT or mock treatment. CMCT treatment was carried out essentially as
531 described (Bakin and Ofengand 1993, 1995). However, when we used the exact conditions of prior studies (Schwartz et

532 al. 2014), we failed to obtain a distinctive signal of termination fold change at the expected Psi position of our spiked in
533 RNA (data not shown). Therefore, to optimize the CMC treatment and pseudouridylation detection efficiency for our
534 Psi-seq, we tested the impact of several main factors affecting the efficiency of CMCT derivatization, including
535 fragmentation before CMC treatment as used previously (Li et al. 2015; Carlile et al. 2014), alternative temperatures
536 and durations of CMCT and alkaline treatment. We found that a 1hr CMCT treatment at 37°C on non-fragmented RNA
537 combined with the 6 hours of alkaline hydrolysis at 37°C resulted in the highest termination fold change (>5) at the
538 expected position in the spike-in and significantly reduced noise compared to a variety of other treatment conditions
539 that we attempted (**Figure S1**). We thus implemented this optimized CMCT treatment conditions for our Psi-seq library
540 preparation.

541 Specifically, pelleted polyA+ RNA was resuspended in 30ul 200mM CMCT in BEU buffer (50mM bicine, pH8.3,
542 4mM EDTA, and 7M urea) or in 30ul BEU buffer only (for mock samples) at 37°C on a thermomixer with 300rpm rotation
543 for one hour. The CMCT reaction was stopped with 100ul of 0.3M sodium acetate and 0.1mM EDTA (pH5.6). After
544 washing, the pellet was dried, dissolved in 40ul of 50mM sodium bicarbonate (pH10.4) at 37°C for six hours. Mock
545 samples were handled identically in all steps but without CMCT treatment. RNA was then precipitated, washed and
546 dissolve in RNase-free water for library preparation.

547 Illumina sequencing libraries were prepared following the RNA ligation method as detailed (Schwartz et al.
548 2014). The single-stranded cDNA product was amplified for 9-11 cycles in a PCR reaction. Libraries were pooled and
549 sequenced on Illumina NextSeq 550 with high output of 75 bp paired end reads.

550 *Read trimming and mapping*

551 Reads were adapter trimmed and quality filtered using Trimmomatic version 0.33 (Bolger et al. 2014) setting
552 the cutoff threshold for average base quality score at 20 over a window of 3 bases, excluding the reads shorter
553 than 20 bases post-trimming (parameters: leading=20, trailing=20, sliding window=3, Minilen= 20). Alignment of the
554 reads to the reference sequence was done in two phases. Cleaned reads were first mapped to a sequence reference
555 comprising of the *Drosophila melanogaster* 18S, 5.8S, 2S and 28S ribosomal RNA gene sequences (accession: M21017.1)
556 and the sequence for RNA spike-in (Schwartz et al. 2014) using bowtie2 (Langmead and Salzberg 2012). The remaining
557 unmapped read pairs were extracted and mapped to the *D. melanogaster* genome reference release FB2017_04 (Larkin
558 et al. 2021) using a splice-aware aligner, STAR version 2.6.1a (Dobin et al. 2013).

559 *Calculation of termination ratios and termination fold change*

560 Inserts were identified by interpolating the concordantly mapped read pairs. For each insert, the 5′ end of the
561 read R2 was identified as the termination site of the product from reverse transcription. For each sample, a Termination
562 Ratio (TR) was computed as the ratio of number of reads terminating at the termination site to the total coverage at the
563 site. Also, a Termination Fold Change is computed for each site as the log2 ratio of TR in the CMC treated sample
564 (TR_{CMC}) to that in the corresponding mock treated sample (TR_{Mock}).

565 *Identification of putative Psi sites*

566 The termination sites with the TR_{CMC} being significantly larger than the corresponding TR_{Mock} were identified
567 using a two-proportion z-test.

568 Two proportion z-test at a given nucleotide position

569 H_o: $TR_{CMC} \leq TR_{Mock}$

570 H_a: $TR_{CMC} > TR_{Mock}$

571 Test statistic z is calculated as

572 $$z = \frac{TR_{CMC} - TR_{Mock}}{\sqrt{(p\,(1-p)(\frac{1}{Cvg_{CMC}} - \frac{1}{Cvg_{Mock}}))}}$$

573 where Cvg_{CMC} and Cvg_{Mock} are the read coverage values in the CMC and mock treated samples, respectively, at
574 the given nucleotide position and p is the pooled sample proportion which is calculated as

575 $$p = \frac{((Cvg_{CMC}\,TR_{CMC}) + (Cvg_{Mock}\,TR_{Mock}))}{(Cvg_{CMC} + Cvg_{Mock})}$$

576 The test is considered relevant only when there are at least 5 CMC treated reads terminating at the position and
577 another 5 continuing over the position.

578 $$(Cvg_{CMC} * TR_{CMC}) \geq 5\ AND\ \left(Cvg_{CMC} * (1 - TR_{CMC})\right) \geq 5$$

579 The p-values associated with the test static z (z-score) were adjusted for multiple-testing using the Benjamini
580 and Hochberg method. Termination sites from within the annotated gene regions with Uridine as the upstream base
581 and the adjusted p-value < 0.05 in at least two w^{1118} replicates, each from different experiments were considered
582 potential Psi sites. Sites with significant adjusted p-values in 6 or 7 w^{1118} replicates were categorized as high
583 reproducibility sites, those in 4 or 5 w^{1118} replicates as intermediate reproducibility sites and in 2 or 3 as low
584 reproducibility sites.

585 *Calculation of expected number and p value of Psi-containing codons on nuclear and mitochondrial mRNAs*

586 The probability of occurrence for a given codon with pseudouridine replacing uridine at a given position was
587 estimated as the ratio of the number of occurrences of that codon in the coding regions of the nuclear or mitochondrial
588 mRNA genes (with at least 10x coverage in the w^{1118} sample) to the total number of Uridine containing codons from
589 those regions divided by the number of Uridines within the codon. An expected number for each Psi-associated codon
590 was estimated by multiplying its probability of occurrence with the total number of Psi sites associated with the coding
591 regions (907 sites associated with nuclear mRNAs and 99 sites with mitochondrial mRNAs). P-values indicate the
592 probability of finding more than the observed number of Psi sites by chance if the observed number was greater than
593 the expected number or the probability of finding at most the observed number of Psi sites if the observed number was
594 less than or equal to the expected number as calculated assuming binomial distribution.

595 *Identification of RluA-2-dependent sites*

596 Psi sites called using our criteria described above in *RluA-2* and w^{1118} of Experiment 2 were compared. The Psi
597 sites that were called in at least two w^{1118} libraries from both of the experiments (Exp 1 and Exp 2) but not detected in
598 any of the *RluA-2* mutant libraries are RluA-2 dependent and identified as RluA-2 targets.

599 *Detection of prevalent sequence motifs*

600 For each RluA-2 dependent Psi site, a flanking sequence of 21 bp was extracted and a consensus was drawn
601 using seqLogo version 2.8 (Crooks et al. 2004) (parameters: -B 2 -T 0.25 -C 21 -k 1 -s -10 -c -n -Y).

602 *Differential gene expression analysis*

603 Mock libraries of *RluA-2* were compared against w^{1118} mock libraries to identify any significant differential gene
604 expression. Read pairs mapping concordantly and uniquely to the exon regions of the annotated genes were counted
605 using featureCounts tool ver. 2.0.0 of Subread package (Liao et al. 2014). Read alignments to antisense strand, or to
606 multiple regions on the genome or those overlapping with multiple genes were ignored (parameters: -s 2 -p -B -C).
607 Differential expression analysis was performed using DESeq2 ver. 1.12.3 (Love et al. 2014) and the p-values were
608 corrected for multiple-testing using the Benjamini–Hochberg method. Genes with adjusted p-values < 0.05 were
609 considered significantly differentially expressed.

610 *Gene set enrichment analysis*

611 Gene set enrichment analysis (GSEA) was performed on a group of predefined gene sets derived from 137
612 KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways pertaining to *Drosophila melanogaster* (Kanehisa and
613 Goto 2000) to detect any significant, concordant differences between *RluA-2* and w^{1118} gene expression. The analysis
614 was conducted using the software GSEA from Broad Institute (Subramanian et al. 2005) for which a matrix of DESeq2
615 normalized read counts for each gene were supplied as input (parameters: permute=geneset, metric= Signal2Noise,
616 scoring_scheme=weighted).

617

618 **Accession numbers**

619 Psi-Sequencing data have been deposited into the Gene Expression Omnibus (GEO) under the accession number
620 GSE213312.

621

622 **Data Availability**

623 Data generated or analyzed during this study are included in this published article (and its supplementary information
624 files). Psi-Sequence data that support the findings of this study have been deposited in the Gene Expression Omnibus
625 (GEO) with the accession number GSE213312.

626

627 **Acknowledgment**

628 The authors thank the Bloomington Drosophila Stock Center (NIH P40OD018537) and Harvard Medical School for the
629 fly stocks obtained for building the mutants used in this study. Funds to support this project were provided by the Linda
630 and Jack Gill Center for Biomolecular Science (WDT), Indiana University startup funds (WDT), NIGMS
631 grant 5R01GM086458 (WDT), and Indiana Clinical and Translational Sciences Institute (CTSI) 2015-2016 Postdoc
632 Challenge award (WS).

633

634

635

636

637

References

Adachi, H., and Y.T. Yu, 2014 Insight into the mechanisms and functions of spliceosomal snRNA pseudouridylation. *World J Biol Chem* 5 (4):398-408.

Anger, A.M., J.P. Armache, O. Berninghausen, M. Habeck, M. Subklewe *et al.*, 2013 Structures of the human and Drosophila 80S ribosome. *Nature* 497 (7447):80-85.

Antonicka, H., K. Choquet, Z.Y. Lin, A.C. Gingras, C.L. Kleinman *et al.*, 2017 A pseudouridine synthase module is essential for mitochondrial protein synthesis and cell viability. *EMBO Rep* 18 (1):28-38.

Bakin, A., and J. Ofengand, 1993 Four newly located pseudouridylate residues in Escherichia coli 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. *Biochemistry* 32 (37):9754-9762.

Bakin, A., and J. Ofengand, 1995 Mapping of the 13-Pseudouridine Residues in Saccharomyces-Cerevisiae Small-Subunit Ribosomal-Rna to Nucleotide Resolution. *Nucleic Acids Research* 23 (16):3290-3294.

Begik, O., M.C. Lucas, L.P. Pryszcz, J.M. Ramirez, R. Medina *et al.*, 2021 Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. *Nat Biotechnol* 39 (10):1278-1291.

Bernier, C.R., A.S. Petrov, C.C. Waterbury, J. Jett, F. Li *et al.*, 2014 RiboVision suite for visualization and analysis of ribosomes. *Faraday Discuss* 169:195-207.

Bolger, A.M., M. Lohse, and B. Usadel, 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15):2114-2120.

Breznak, S., Y. Peng, L. Deng, N.M. Kotb, Z. Flamholz *et al.*, 2022 Pseudouridine-dependent ribosome biogenesis regulates translation of polyglutamine proteins during *Drosophila* oogenesis. *bioRxiv*:2022.2007.2007.499147.

Carlile, T.M., M.F. Rojas-Duran, B. Zinshteyn, H. Shin, K.M. Bartoli *et al.*, 2014 Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* 515 (7525):143-+.

Charette, M., and M.W. Gray, 2000 Pseudouridine in RNA: What, where, how, and why. *Iubmb Life* 49 (5):341-351.

Crooks, G.E., G. Hon, J.M. Chandonia, and S.E. Brenner, 2004 WebLogo: a sequence logo generator. *Genome Res* 14 (6):1188-1190.

Deryusheva, S., and J.G. Gall, 2009 Small Cajal body-specific RNAs of Drosophila function in the absence of Cajal bodies. *Mol Biol Cell* 20 (24):5250-5259.

Dobin, A., C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski *et al.*, 2013 STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29 (1):15-21.

Eyler, D.E., M.K. Franco, Z. Batool, M.Z. Wu, M.L. Dubuke *et al.*, 2019 Pseudouridylation of mRNA coding sequences alters translation. *Proc Natl Acad Sci U S A* 116 (46):23068-23074.

Fujii, S., and H. Amrein, 2002 Genes expressed in the Drosophila head reveal a role for fat cells in sex-specific physiology. *Embo Journal* 21 (20):5353-5363.

Gallant, P., 2013 Myc function in Drosophila. *Cold Spring Harb Perspect Med* 3 (10):a014324.

Ge, J., and Y.T. Yu, 2013 RNA pseudouridylation: new insights into an old modification. *Trends in Biochemical Sciences* 38 (4):210-218.

Genuth, N.R., and M. Barna, 2018 The Discovery of Ribosome Heterogeneity and Its Implications for Gene Regulation and Organismal Life. *Molecular Cell* 71 (3):364-374.

Giordano, E., I. Peluso, S. Senger, and M. Furia, 1999 minifly, a Drosophila gene required for ribosome biogenesis. *J Cell Biol* 144 (6):1123-1133.

Gruber, A.R., C. Kilgus, A. Mosig, I.L. Hofacker, W. Hennig *et al.*, 2008 Arthropod 7SK RNA. *Mol Biol Evol* 25 (9):1923-1930.

Hames, B.D., and M. Bownes, 1978 Synthesis of yolk proteins in Drosophila melanogaster. *Insect Biochemistry* 8 (5):319-328.

Hamma, T., and A.R. Ferre-D'Amare, 2006 Pseudouridine synthases. *Chemistry & Biology* 13 (11):1125-1135.

Ho, N.W., and P.T. Gilham, 1967 The reversible chemical modification of uracil, thymine, and guanine nucleotides and the modification of the action of ribonuclease on ribonucleic acid. *Biochemistry* 6 (12):3632-3639.

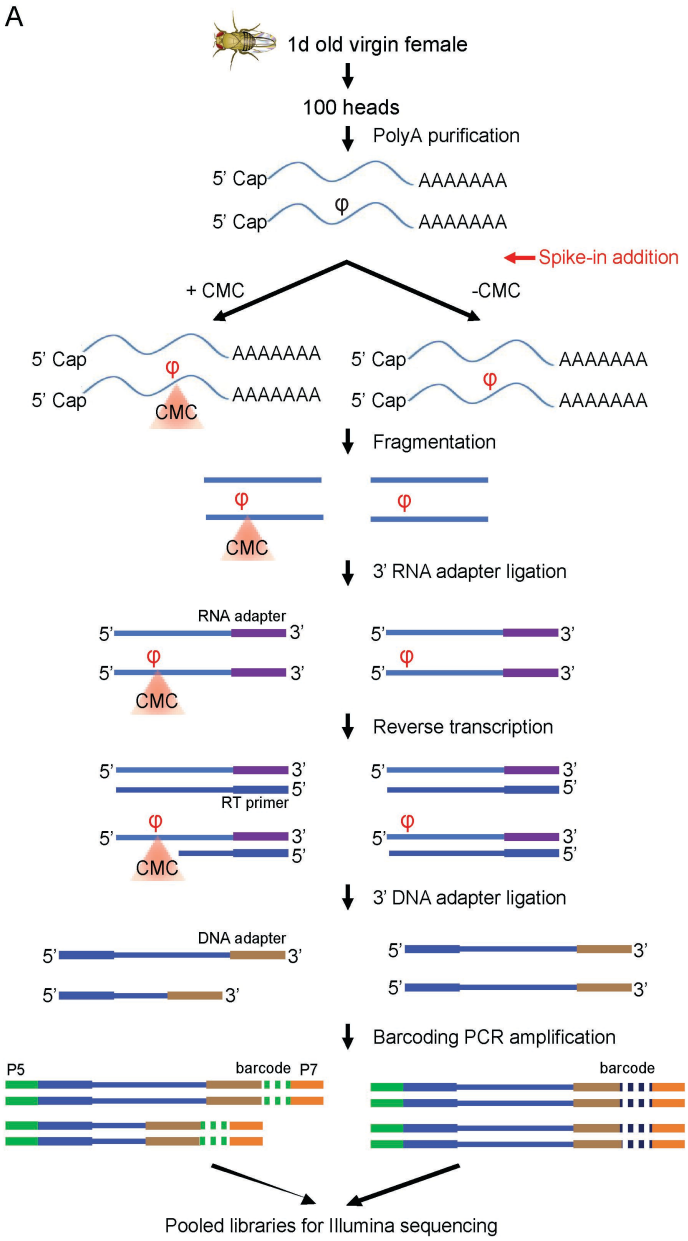
Hoang, C., J.J. Chen, C.A. Vizthum, J.M. Kandel, C.S. Hamilton *et al.*, 2006 Crystal structure of pseudouridine synthase RluA: Indirect sequence readout through protein-induced RNA structure. *Molecular Cell* 24 (4):535-545.

Hopes, T., K. Norris, M. Agapiou, C.G.P. McCarthy, P.A. Lewis *et al.*, 2022 Ribosome heterogeneity in Drosophila melanogaster gonads through paralog-switching. *Nucleic Acids Res* 50 (4):2240-2257.

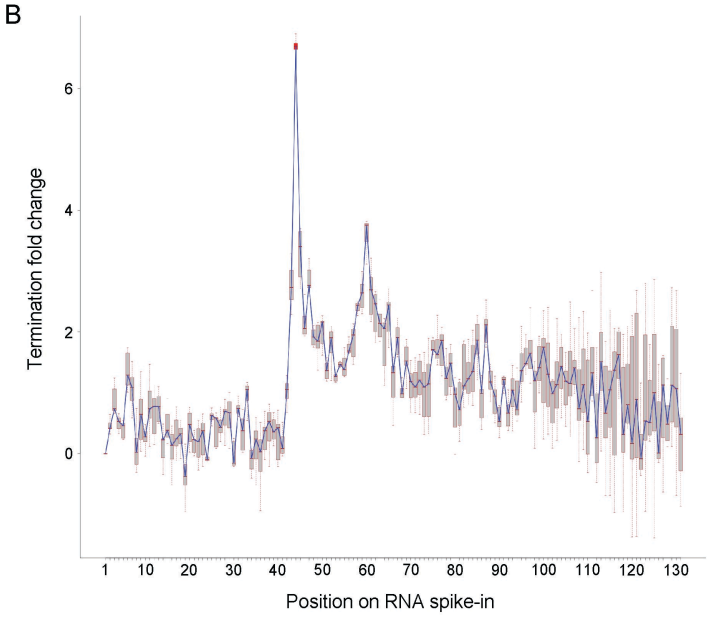
690 Huang, Z.P., H. Zhou, H.L. He, C.L. Chen, D. Liang *et al.*, 2005 Genome-wide analyses of two families of snoRNA genes
691 from Drosophila melanogaster, demonstrating the extensive utilization of introns for coding of snoRNAs. *Rna*
692 11 (8):1303-1316.
693 Jungreis, I., M.F. Lin, R. Spokony, C.S. Chan, N. Negre *et al.*, 2011 Evidence of abundant stop codon readthrough in
694 Drosophila and other metazoa. *Genome Res* 21 (12):2096-2113.
695 Kanehisa, M., and S. Goto, 2000 KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* 28 (1):27-30.
696 Karijolich, J., and Y.T. Yu, 2010 Spliceosomal snRNA modifications and their function. *Rna Biology* 7 (2):192-204.
697 Karijolich, J., and Y.T. Yu, 2011 Converting nonsense codons into sense codons by targeted pseudouridylation. *Nature*
698 474 (7351):395-+.
699 Kauffman, K.J., F.F. Mir, S. Jhunjhunwala, J.C. Kaczmarek, J.E. Hurtado *et al.*, 2016 Efficacy and immunogenicity of
700 unmodified and pseudouridine-modified mRNA delivered systemically with lipid nanoparticles in vivo.
701 *Biomaterials* 109:78-87.
702 Langmead, B., and S.L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9 (4):357-359.
703 Larkin, A., S.J. Marygold, G. Antonazzo, H. Attrill, G. Dos Santos *et al.*, 2021 FlyBase: updates to the Drosophila
704 melanogaster knowledge base. *Nucleic Acids Research* 49 (D1):D899-D907.
705 Li, X.Y., P. Zhu, S.Q. Ma, J.H. Song, J.Y. Bai *et al.*, 2015 Chemical pulldown reveals dynamic pseudouridylation of the
706 mammalian transcriptome. *Nature Chemical Biology* 11 (8):592-U593.
707 Liao, R., R. Zhang, J. Guan, and S. Zhou, 2014 A New Unsupervised Binning Approach for Metagenomic Sequences
708 Based on N-grams and Automatic Feature Weighting. *IEEE/ACM Trans Comput Biol Bioinform* 11 (1):42-54.
709 Love, M.I., W. Huber, and S. Anders, 2014 Moderated estimation of fold change and dispersion for RNA-seq data with
710 DESeq2. *Genome Biol* 15 (12):550.
711 Lovejoy, A.F., D.P. Riordan, and P.O. Brown, 2014 Transcriptome-Wide Mapping of Pseudouridines: Pseudouridine
712 Synthases Modify Specific mRNAs in S. cerevisiae. *Plos One* 9 (10).
713 Machnicka, M.A., A. Olchowik, H. Grosjean, and J.M. Bujnicki, 2014 Distribution and frequencies of post-transcriptional
714 modifications in tRNAs. *Rna Biology* 11 (12):1619-1629.
715 Marchand, V., F. Pichot, P. Neybecker, L. Ayadi, V. Bourguignon-Igel *et al.*, 2020 HydraPsiSeq: a method for systematic
716 and quantitative mapping of pseudouridines in RNA. *Nucleic Acids Research* 48 (19):e110.
717 Nguyen, D., B.J. Krueger, S.C. Sedore, J.E. Brogie, J.T. Rogers *et al.*, 2012 The Drosophila 7SK snRNP and the essential
718 role of dHEXIM in development. *Nucleic Acids Research* 40 (12):5283-5297.
719 Ofengand, J., 2002 Ribosomal RNA pseudouridines and pseudouridine synthases. *Febs Letters* 514 (1):17-25.
720 Ofengand, J., and A. Bakin, 1997 Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal
721 RNAs from representative eukaryotes, prokaryotes, archaebacteria, mitochondria and chloroplasts. *Journal of*
722 *Molecular Biology* 266 (2):246-268.
723 Penzo, M., A.N. Guerrieri, F. Zacchini, D. Trere, and L. Montanaro, 2017 RNA Pseudouridylation in Physiology and
724 Medicine: For Better and for Worse. *Genes (Basel)* 8 (11).
725 Phillips, B., A.N. Billin, C. Cadwell, R. Buchholz, C. Erickson *et al.*, 1998 The Nop60B gene of Drosophila encodes an
726 essential nucleolar protein that functions in yeast. *Mol Gen Genet* 260 (1):20-29.
727 Purchal, M.K., D.E. Eyler, M. Tardu, M.K. Franco, M.M. Korn *et al.*, 2022 Pseudouridine synthase 7 is an opportunistic
728 enzyme that binds and modifies substrates with diverse sequences and structures. *Proc Natl Acad Sci U S A* 119
729 (4).
730 Rintala-Dempsey, A.C., and U. Kothe, 2017 Eukaryotic stand-alone pseudouridine synthases - RNA modifying enzymes
731 and emerging regulators of gene expression? *Rna Biology* 14 (9):1185-1196.
732 Schwartz, S., D.A. Bernstein, M.R. Mumbach, M. Jovanovic, R.H. Herbst *et al.*, 2014 Transcriptome-wide mapping
733 reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* 159 (1):148-162.
734 Shi, Z., K. Fujii, K.M. Kovary, N.R. Genuth, H.L. Rost *et al.*, 2017 Heterogeneous Ribosomes Preferentially Translate
735 Distinct Subpools of mRNAs Genome-wide. *Molecular Cell* 67 (1):71-83 e77.
736 Song, W., S. Ressl, and W.D. Tracey, 2020 Loss of Pseudouridine Synthases in the RluA Family Causes Hypersensitive
737 Nociception in Drosophila. *G3 (Bethesda)* 10 (12):4425-4438.
738 Subramanian, A., P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert *et al.*, 2005 Gene set enrichment analysis: a
739 knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102
740 (43):15545-15550.
741 Yamaki, Y., Y. Nobe, M. Koike, Y. Yamauchi, K. Hirota *et al.*, 2020 Direct Determination of Pseudouridine in RNA by
742 Mass Spectrometry Coupled with Stable Isotope Labeling. *Analytical Chemistry* 92 (16):11349-11356.

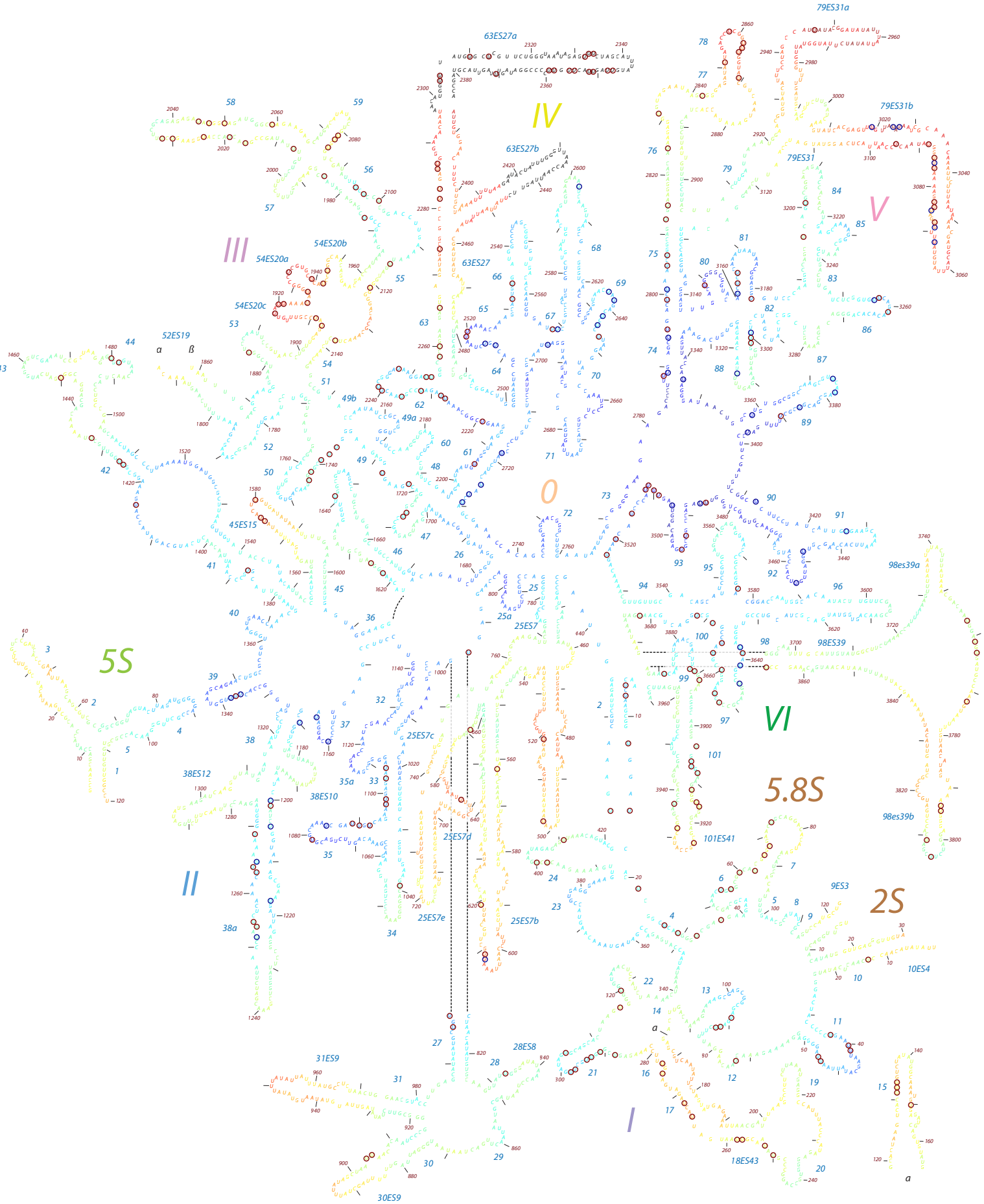
743 Yu, A.T., J. Ge, and Y.T. Yu, 2011 Pseudouridines in spliceosomal snRNAs. *Protein Cell* 2 (9):712-725.
744 Zhao, Y., J. Karijolich, B. Glaunsinger, and Q. Zhou, 2016 Pseudouridylation of 7SK snRNA promotes 7SK snRNP
745 formation to suppress HIV-1 transcription and escape from latency. *EMBO Rep* 17 (10):1441-1451.

A


1d old virgin female
100 heads
PolyA purification
5' Cap
AAAAAAA
φ
Spike-in addition
+ CMC
-CMC
CMC
Fragmentation
3' RNA adapter ligation
RNA adapter
5'
3'
Reverse transcription
RT primer
3' DNA adapter ligation
DNA adapter
Barcoding PCR amplification
P5
barcode
P7
Pooled libraries for Illumina sequencing

B


Termination fold change
0
2
4
6
1
10
20
30
40
50
60
70
80
90
100
110
120
130
Position on RNA spike-in

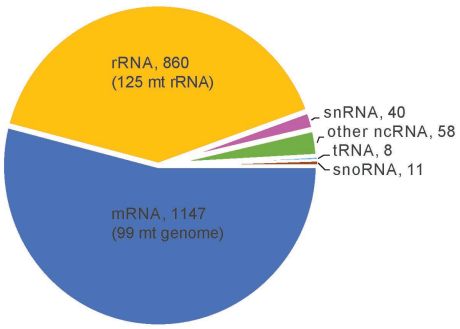


A 3D-based secondary structure, generated by RiboVision.

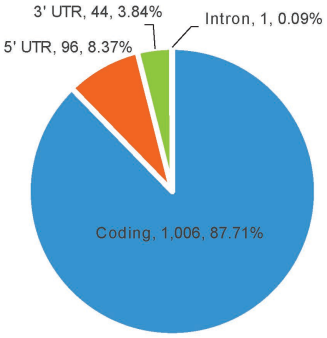
Drosophila melanogaster

large subunit ribosomal RNA

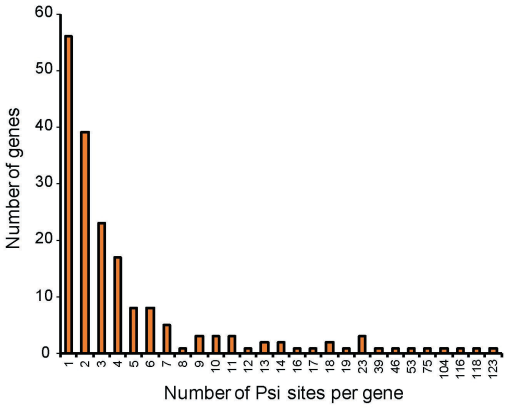
A



B



C

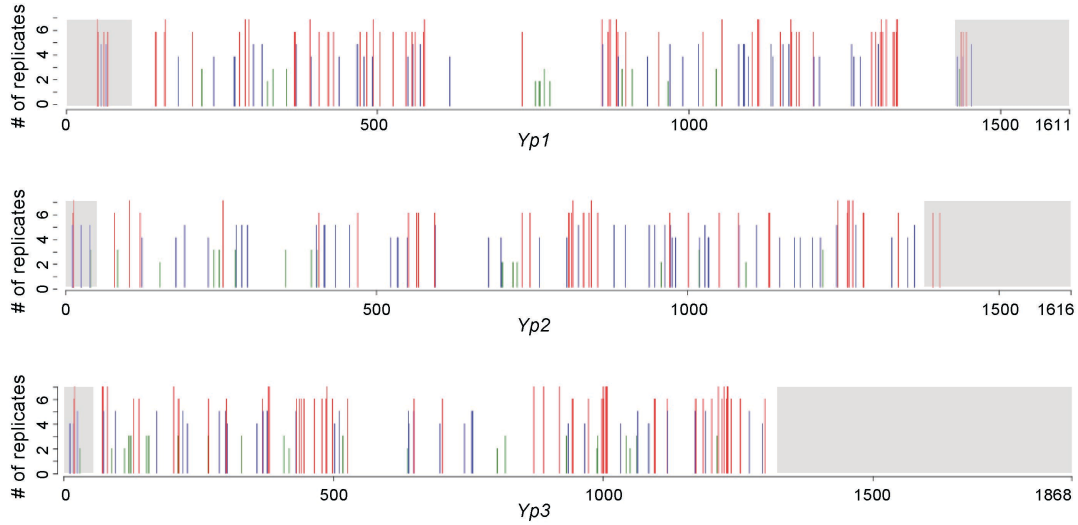


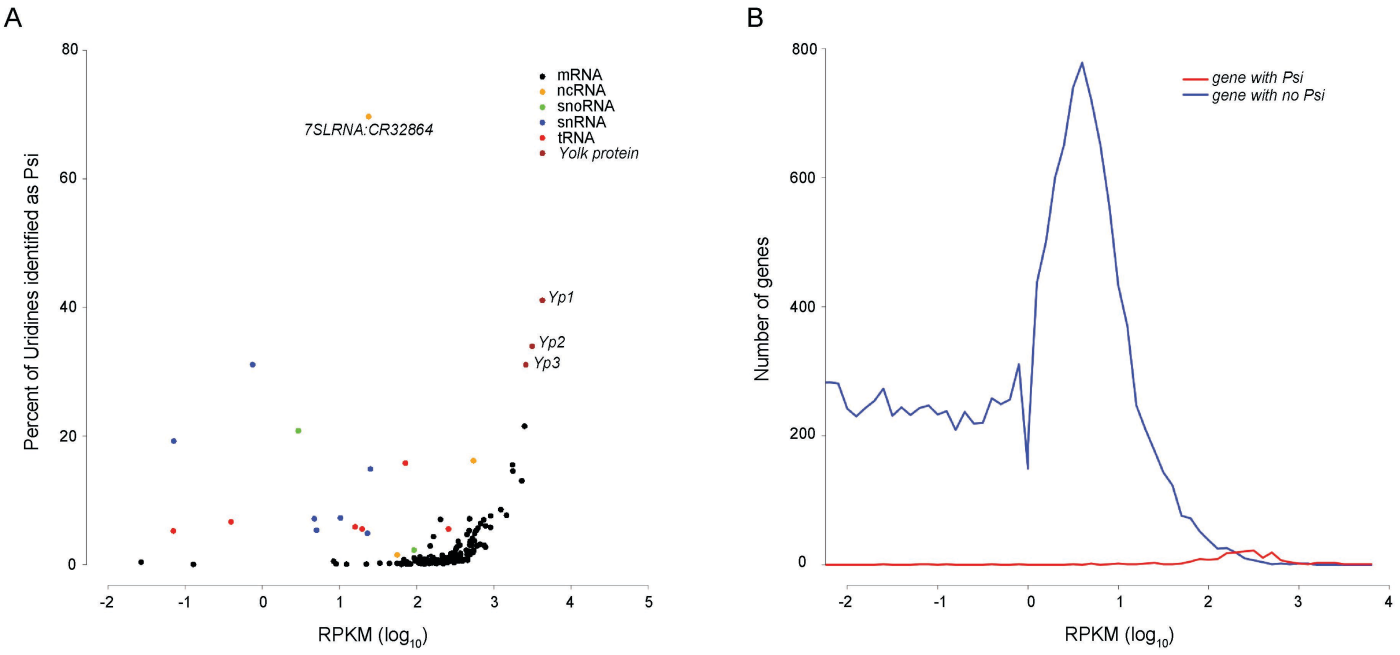
D

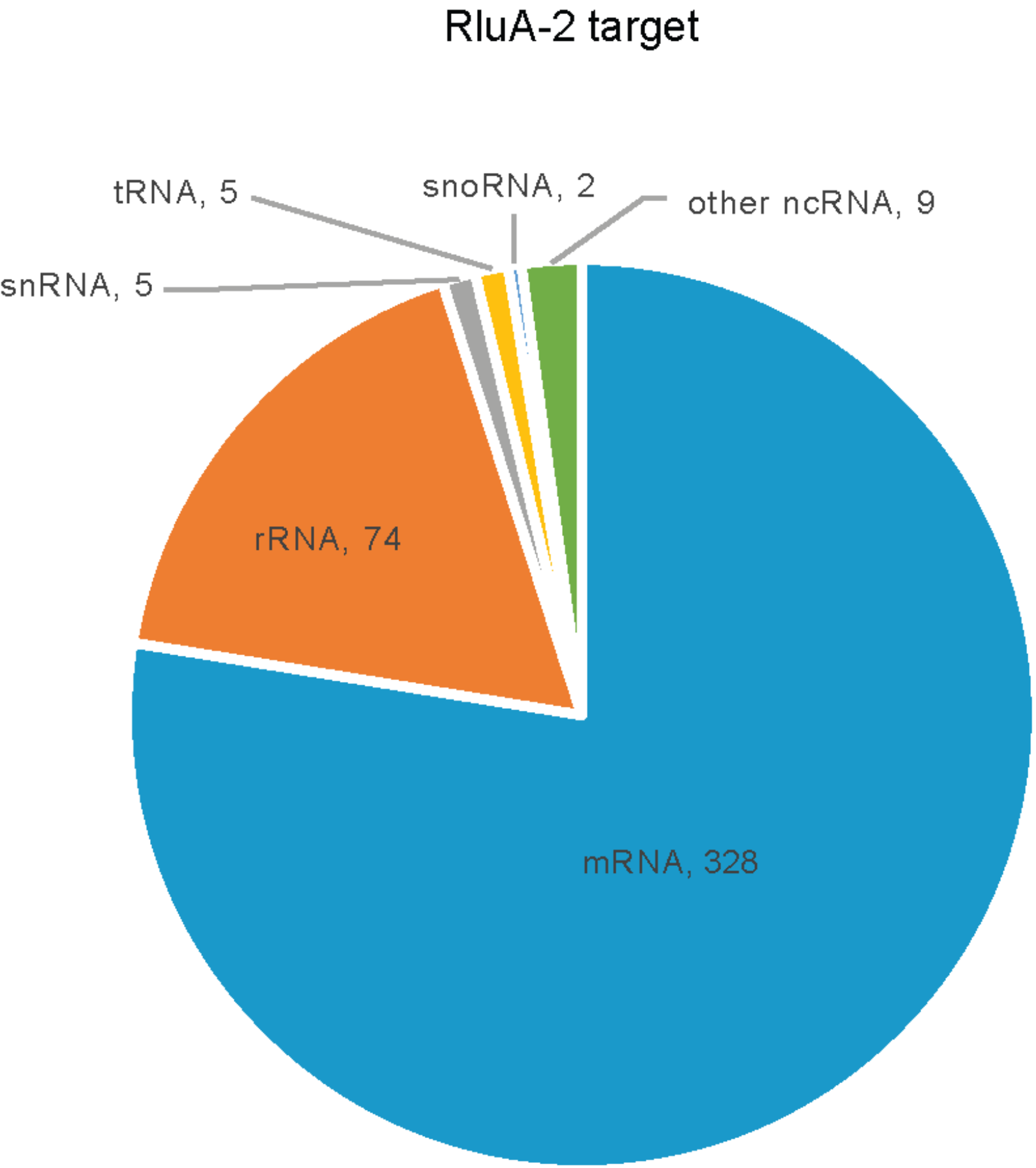
Top 10 coding genes

Gene_id	Gene symbol	Number of Psi sites	Number of Uridines	Average RPKM
FBgn000404	*Yp1*	118	287	4262
FBgn000404	*Yp3*	116	373	2598
FBgn000539	*Yp2*	104	306	3138
FBgn000294	*ninaE*	75	348	2501
FBgn000012	*Arr2*	53	341	1744
FBgn000012	*Arr1*	39	647	783
FBgn001367	mt:*CoI*	23	616	561
FBgn001367	mt:*CoIII*	23	303	912
FBgn004081	*Nplp2*	23	158	1764
FBgn026469	*Mhc*	19	1833	220

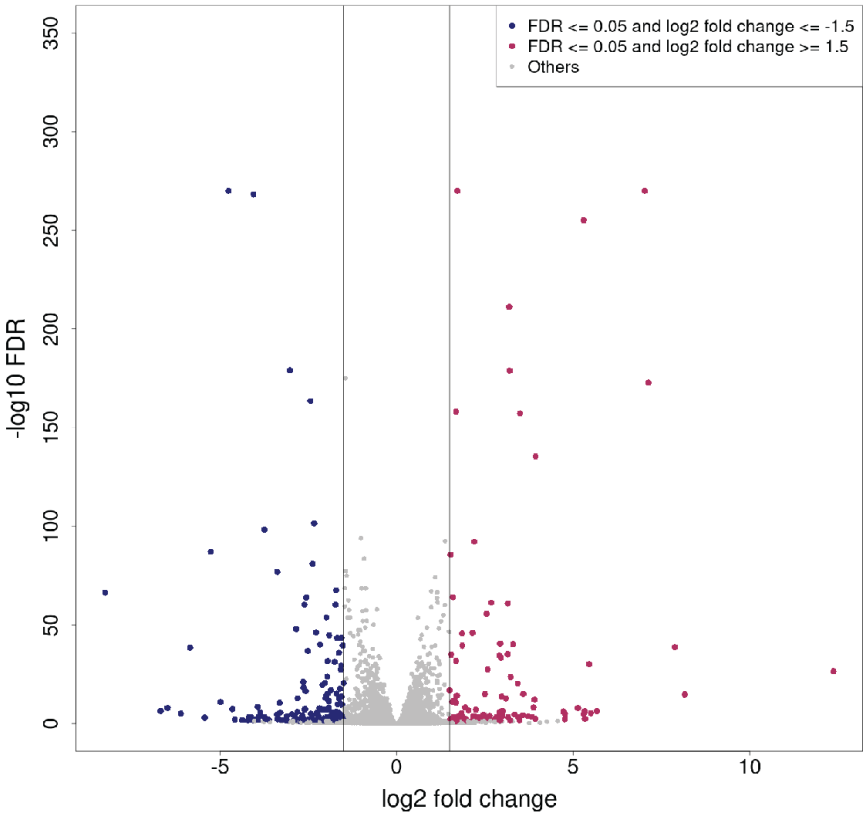
E




A
Percent of Uridines identified as Psi
RPKM (log10)
mRNA
ncRNA
snoRNA
snRNA
tRNA
Yolk protein
7SLRNA:CR32864
Yp1
Yp2
Yp3
B
Number of genes
RPKM (log10)
gene with Psi
gene with no Psi


RluA-2 target
tRNA, 5
snoRNA, 2
other ncRNA, 9
snRNA, 5
rRNA, 74
mRNA, 328

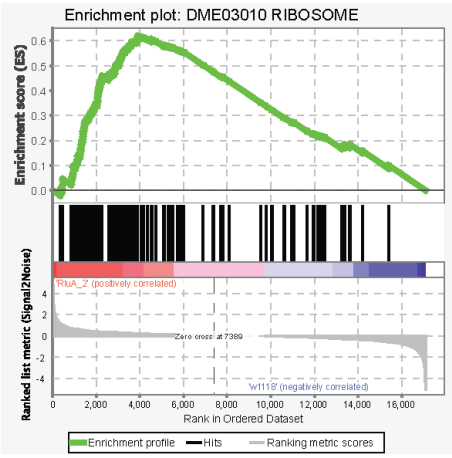
A

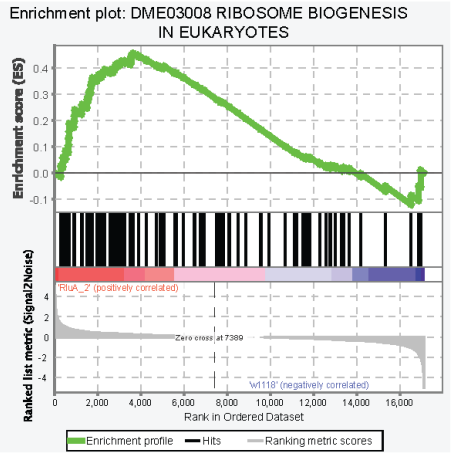

FDR <= 0.05 and log2 fold change <= -1.5
FDR <= 0.05 and log2 fold change >= 1.5
Others
-log10 FDR
log2 fold change

B

RluA-2(1)
RluA-2(2)
RluA-2(4)
RluA-2(3)
w1118(1)
w1118(4)
w1118(3)
w1118(2)

C


Enrichment plot: DME03010 RIBOSOME
Enrichment score (ES)
Ranked list metric (Signal2Noise)
'RluA_2' (positively correlated)
'w1118' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores


Enrichment plot: DME03008 RIBOSOME BIOGENESIS IN EUKARYOTES
Enrichment score (ES)
Ranked list metric (Signal2Noise)
'RluA_2' (positively correlated)
'w1118' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores