

An alternative spliceosome defined by distinct snRNAs in early zebrafish embryogenesis

Johanna F. B. Pagano¹, Rob J. Dekker¹, Wim A. Ensink¹, Marina van Olst¹, Alex Bos¹, Selina van Leeuwen¹, Wim C. de Leeuw¹, Ulrike Nehrdich², Herman P. Spalink², Han Rauwerda¹,
Martijn J. Jonker¹ and Timo M. Breit^{1,*}

* To whom correspondence should be addressed.

Tel: +31 20 5257058; Fax: +31 20 5257762; Email: t.m.breit@uva.nl

¹ RNA Biology & Applied Bioinformatics research group, Swammerdam Institute for Life Sciences, Faculty of Science, University of Amsterdam, Amsterdam 1090 GE, the Netherlands

² Department of Molecular Cell Biology, Institute of Biology, Leiden University, Gorlaeus Laboratories - Cell Observatory, Leiden 2333 CE, the Netherlands

Alternative snRNAs in early zebrafish embryogenesis

snRNA, spliceosome, dual translation machinery, embryogenesis

1 ABSTRACT

2 Splicing removes intronic RNA sequences are removed from pre-mRNA molecules and
3 enables, by alternative splicing, the generation of multiple unique RNA molecules from a
4 single gene. As such, splicing is an essential part of the whole translation system of a cell.
5 The spliceosome is a ribonucleoprotein complex in which five small nuclear RNAs (snRNAs)
6 are involved; U1, U2, U4, U5, and U6. For each of these snRNAs there are variant gene
7 copies present in a genome. Furthermore, in many eukaryotic species there is an
8 alternative, minor spliceosome that can splice a small number of specific introns. As we
9 previously discovered an embryogenesis-specific ribosomal system in zebrafish early
10 embryogenesis based on variant rRNA and snoRNA expression, we hypothesized that there
11 may also be an embryogenesis-specific spliceosome. An inventory of zebrafish snRNA genes
12 revealed clustered and dispersed loci for all but U2 major snRNAs. For each minor
13 spliceosome snRNA, just one gene locus was found. Since complete snRNA molecules are
14 hard to sequence, we employed a combined PCR-sequencing approach to measure the
15 individual snRNA-variant presence. Analysis of egg and male-adult samples revealed
16 embryogenesis-specific and somatic-specific variants for each major snRNA. These variants
17 have substantial sequence differences, yet none in their mRNA binding sites. Given that
18 many of the sequence differences are found in loop structures indicate possible alternative
19 protein binding. Altogether, with this study we established that the spliceosome is also an
20 element of the embryogenesis-specific translation system in zebrafish.

21

22

23

24

25

26

27

28 INTRODUCTION

29 Alternative splicing is fundamental for gene regulation and the generation of different
30 transcripts and/or proteins from an individual gene in eukaryotes (1). Splicing is executed by
31 the spliceosome and removes intronic sequences from pre-mRNA during the maturation
32 process in which the exonic sequences eventually form the mRNA (2,3).The spliceosome is a
33 molecular complex formed by hundreds of proteins and five essential small-nuclear RNAs
34 (snRNAs) that are typically located in the nucleus. The size of these small RNA molecules
35 ranges from 118 nucleotides (nt) to 191 nt. As they are uracil rich, they are called U1, U2,
36 U4, U5 and U6 snRNAs. Next to this major spliceosome, a minor (or U12 dependent)
37 spliceosome exists in many eukaryotic species, which is involved in the splicing of a relative
38 small number of specific introns (4). The snRNAs involved in the minor spliceosome are:
39 U11, U12, U4atac, and U6atac, completed by the U5 from the major spliceosome (4).

40 As splicing is at the core of the cellular translation system, the sequences of the involved
41 snRNA are highly conserved across species. At the same time, many non-canonical variants
42 and gene copies of the major snRNA genes are present within the same organisms (5–9).
43 This raises the question why these variants exist and what role they might play in
44 translation. Although expression of these variants has been extensively studied, there is still
45 not a clear understanding for the existence of these snRNA variants (10).

46 A indication to the function of variant snRNAs may lie in the fact that several studies have
47 observed differentially expression of variants during development. For instance in
48 *Drosophila* several variants are expressed during early embryogenesis, but eventually one
49 variant gradually dominates expression (11,12). Similar expression patterns for snRNA
50 variants were observed in *Xenopus* (13–15), mouse (12,16), sea urchin (17,18), and flatworm
51 (19). Comparable findings have been reported for: snRNA U2 in silk moth (20) and
52 *Dictyostelium discoideum* (21); snRNA U4 in chicken (22); and snRNA U5 in *Drosophila* (5) and
53 in human (6). snRNAs also display a tissue-preferred expression, which implicate them in
54 different disease pathologies such as several neurological diseases (7,10,23,24).

55 The fact that there are snRNAs which variants are differentially expressed during
56 embryogenesis relates to our previous findings where we discovered distinct maternal-types
57 of rRNAs and snoRNAs specifically expressed during early zebrafish embryogenesis (25).
58 These maternal-type RNAs seem to be part of a distinct early embryogenesis-specific

59 translation machinery, which is gradually replaced during embryogenesis by a somatic type.
60 Combining the snRNA and our rRNA plus snoRNA findings, lead us to hypothesize that there
61 might also be distinct maternal-type snRNAs contributing to the embryogenesis-specific
62 translation machinery.

63 As in our hands, snRNAs cannot consistently be sequenced by standard next-generation
64 sequencing protocols, probably due to strong secondary structures and ample RNA
65 modifications, we employed a wet-lab approach combining RT-PCR, for amplification, and
66 DNA sequencing, for identification, of snRNAs. Using this approach on egg and mature adult-
67 male samples, we were able to show that for each snRNA, similar to rRNA and snoRNA,
68 there are maternal-type snRNAs that are uniquely present in zebrafish eggs. This means that
69 snRNAs are a part of the zebrafish unique embryogenesis-specific translation machinery.

70 RESULTS AND DISCUSSION

71 Cataloguing the zebrafish snRNAs
72 In earlier studies we discovered cellular components related to the transcription machinery
73 that have distinct embryogenesis expression or somatic expression during zebrafish
74 development. We reported on 5S (26), 5.8S, 18S, and 28S (27), plus small-derived
75 components from those molecules (28), as well as snoRNAs (25). As it is clear that in
76 zebrafish oogenesis and early embryogenesis a different transcription machinery is
77 employed, it is obvious to investigate other components of this system. Hence, we here
78 focused on the small nuclear RNAs (snRNAs) of the spliceosome to determine whether their
79 variants also display distinct expression profiles. We started by making an inventory of all
80 snRNA elements present in the zebrafish genome based on the annotated snRNA sequences
81 from the Ensembl database (29). In total, 541 snRNA loci were retrieved for the major
82 spliceosome and seven snRNA loci for the minor spliceosome from the database. Given the
83 many snRNA loci for the major spliceosome, we compared them by sequence alignment per
84 snRNA (Supplemental Figure SF1). Several snRNA sequences appeared aberrant and it
85 turned out that these sequences partly existed of retrotransposon sequences (30), therefore
86 they were excluded (Supplemental Table ST1). To complete our reference set of snRNAs, the
87 database-derived snRNA sequences were used to explore the zebrafish genome for yet

88 unannotated major and minor spliceosome snRNA loci. This resulted in a total of 958 snRNA
89 loci (Table 1, Supplemental Table ST1 and File SFile1).

90 The alignment of the sequences within each major spliceosome snRNA showed two clusters
91 for each snRNA (Supplemental Files SF1). The distribution of these two snRNA clusters
92 coincides with a clear genomic organizational preference: either in condense genomic
93 repeats (Table 1, blue), or scattered throughout the genome (Table 1, red). Obviously, the
94 discovery of two sequence clusters for each major spliceosome snRNA is in line with the
95 possible existence of maternal-type and somatic-type snRNA spliceosomes.

96 In contrast; for the minor spliceosome, U12 snRNA is present only once in the genome,
97 which means that for U12 snRNA no maternal-type variant can exist. For all three other
98 minor spliceosome-specific snRNAs, just two loci appeared to be present in the genome
99 (Supplement Table ST1). Because the observed sequences within each of these minor
100 spliceosome snRNAs are so different, we questioned whether they are legitimate snRNA
101 loci. By analyzing the expression of all these snRNA loci, it could be determined that for each
102 minor spliceosome snRNAs, only one locus is expressed (Table 1, grey, Supplemental Table
103 2) This precludes the existence of an alternative, completely embryogenesis-specific minor
104 spliceosome and we therefore excluded these minor spliceosome snRNAs from the
105 subsequent analyses.

106

Table 1: Genomic distribution of maternal and somatic zebrafish snRNA genes

| | total # | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 11 | 13 | 17 | 19 | 20 | 21 | 25 |
|--------------------|---------|---|----|-----|---|---|---|---|---|----|----|----|----|----|----|----|
| major snRNA | | | | | | | | | | | | | | | | |
| U1 | 259 | | | 248 | | | 4 | 2 | 3 | | | | | 2 | | |
| U2 | 7 | 2 | | | | 1 | 2 | | | 2 | | | | | | |
| U4 | 26 | | 2 | | | | | | | | 23 | | | | 1 | |
| U5 | 139 | | | 132 | | | | 1 | | | | 6 | | | | |
| U6 | 520 | | 12 | 503 | | 1 | | 1 | 1 | 1 | | 1 | 1 | | | |
| minor snRNA | | | | | | | | | | | | | | | | |
| U11 | 1 | | | | | | | | | | 1 | | | | | |
| U12 | 1 | | 1 | | | | | | | | | | | | | |
| U4atc | 1 | | | | | | | | 1 | | | | | | | |
| U6atc | 1 | | | | | | | | | | | | | 1 | | |

107

108

109 Embryogenesis-specific snRNA variants

110 After cataloguing all snRNAs in the zebrafish genome, we investigated whether

111 embryogenesis-specific snRNAs exist by examining their expression in eggs and somatic

112 tissue. However, it became quickly evident that (complete) snRNA molecules, similarly to
113 tRNA, 5S, and snoRNA molecules, cannot readily be sequenced by standard smallRNA-seq
114 procedures, probably due to their robust secondary configurations, as well as possible 3'
115 modifications. Only small numbers of partial snRNA sequences were observed, which did
116 show a hint of differential snRNA expression between egg and adult tissue. To tackle this
117 technical problem, we performed a RT-PCR-qSeq analysis effectively circumventing the
118 snRNA 3' issues. For this, we selected the sequence most prominently present in the
119 genome for each cluster of every snRNAs of the major spliceosome (Figure 1A). The
120 differences between these pairs of snRNA sequences ranges from 11 nucleotides (U1 and
121 U6) to 48 nucleotides (U5) (Figure 1A, Supplemental Figure SFig1). We designed
122 (degenerated) RT-PCR-primers (Figure 1A) that amplified virtual all snRNAs, after which the
123 PCR products were sequenced and identified by mapping to the snRNA sequences. Using
124 this procedure at least 0.5 M reads were obtained for each snRNA (Supplemental Table
125 ST2). These results revealed that for each snRNA one variant accounts for almost all snRNA
126 in the egg samples, whereas the other variant makes up the snRNAs in adult zebrafish
127 (Figure 1B). Similar to rRNA and snoRNA, the snRNA variant present in egg is called
128 maternal-type versus the somatic-type variant in adult tissue (Figure 1A).
129 To check the RT-PCR-qSeq results we developed a qRT-PCR analyses for snRNA U1 and
130 snRNA U5. This qRT-PCR analysis confirmed the exclusive presence of embryogenesis-
131 specific snRNA variants in egg samples (Supplemental Figure SFig2).

132 **Figure 1. Maternal-type and somatic-type snRNA sequences and expression**
133 A: Sequence comparison of selected maternal-type (M) and somatic-type (S) snRNA variants;
134 identical nucleotides are indicated as dots, while gaps as dashes (For sequence alignment of all
135 snRNAs cf. Supplemental Figure SF2). The RT-PCR primers are indicated with half arrows: maternal-
136 type specific (blue); somatic-type specific (red); and non-distinctive (green).
137 B: Relative expression of maternal-type (blue) and somatic-type (red) variants for each snRNA
138 indicated by comparative percentage calculated using the RT-PCR-qSeq approach on RNA from egg
139 and male tail tissue (MT).

140
141 Differences between maternal- and somatic-types snRNAs
142 In order to assign any functional relevance to the embryogenesis-specific snRNAs,
143 nucleotide difference as compared to the somatic-type snRNAs were investigated (Figure
144 1A). One of the distinctions between the major and minor spliceosome snRNAs lies in their

145 nucleotide sequences that bind to the mRNA, thus allowing each system to splice distinct
146 introns. However, even though there are many sequence differences between the maternal-
147 type and the somatic-type snRNAs, none of them involve the mRNA binding sites in these
148 snRNAs (Figure 2 and Supplement File SF3). It turned out that the snRNA sequence
149 differences are often located in specific parts of the secondary structure (Figure 2). For
150 instance, for U1, all but one differences are located in one stem-loop and in U2 they are
151 confined to one side of the structure (Figure 2). In general, many of the differences are
152 found in the loops, which are thought to be specific binding locations for spliceosomal
153 proteins. Hence, this would indicate that the embryogenesis spliceosome, besides specific
154 snRNAs also may comprise (embryogenesis-)specific proteins.
155 Despite many apparently co-evolved nucleotide pairs in stems of the snRNAs (Figure 2),
156 there seems to be only one co-evolved nucleotide pair in the interaction site between U4
157 and U6. However, as this nucleotide is right in the middle of the largest interaction site, it
158 might actually prevent the binding of somatic-type U4 to maternal-type U6 and vice versa.

159

160 **Figure 2. Sequence differences between maternal-type and somatic-type snRNAs**

161 The secondary structure is presented for each maternal-type and somatic-type zebrafish snRNA.
162 These structures are an adaptation from Figure 4 in reference (4) and also indicate the specific
163 binding sites between U4 and U6. The arrows indicate the base pair that coevolved within one of the
164 U4 – U6 interaction sites. The circles indicate nucleotides that are different between the maternal-
165 type (blue) and somatic-type (red) sequences. The sites interacting with the pre-mRNA are
166 underlined.

167

168 **CONCLUDING REMARKS**

169 In this study we reveal the existence of an embryogenesis-specific major spliceosome in
170 zebrafish, consisting of at least a distinct set of U1, U2, U4, U5 and U6 snRNAs. We do not
171 know if this embryogenesis-specific spliceosome also contains embryogenesis-specific
172 proteins, yet the position of the distinguishing nucleotides in the loops of the maternal-type
173 snRNA structure would suggest this. However, a quick scan of some of the major
174 spliceosomal protein genes revealed that they are present with just one copy in the

175 zebrafish genome, which effectively rules out any embryogenesis-specific spliceosomal
176 protein gene variants. Yet, the genes for these proteins do contain introns, which have in
177 general an important function with respect to production of alternative transcripts plus
178 alternative proteins. This leads to the intriguing question, whether the maternal-type
179 snRNAs will produce embryogenesis-specific alternative-spliced transcripts and thus
180 embryogenesis-specific spliceosome proteins for the embryogenesis-specific variant of the
181 major spliceosome.

182 With respect to the minor spliceosome, which acts on different splicing sites; we did not find
183 any maternal-type snRNA variants for U11, U12, U4atc, and U6atc. This however does not
184 lead to the conclusion that no embryogenesis-specific minor spliceosome exists. On the
185 contrary, the minor spliceosome is always completed by the U5 snRNA from the major
186 spliceosome (4) and because in eggs virtually only the maternal-type U5 snRNA is present,
187 this U5 variant is used in combination with the other four ubiquitously-expressed minor
188 spliceosome snRNAs. Even though only one snRNA in this splicing system is different, we
189 would argue that this makes the whole minor spliceosome in eggs embryogenesis-specific.
190 This might also explain the observation that the maternal-type U5 snRNA sequence differs a
191 noteworthy 41% compared to the somatic-type, whereas in all other snRNAs this difference
192 is just 7% to 20%. Yet, the mRNA binding loop that holds the exons together (32) and the
193 associated stem are identical between maternal-type, somatic-type and human U5 snRNA.
194 Besides this core the zebrafish U5 snRNA variants are almost completely different (53%).
195 The functional implications remain elusive for now.

196 Although the differences between the variants of each snRNA are quite substantial, their
197 secondary structure appears to be relatively unaltered. This is also due to many co-evolved
198 nucleotides in the stem sequences. Similar to what we observed previously for the 45S
199 rRNAs, the sequence homology between the somatic-type and maternal-type snRNA
200 variants is often lower than snRNAs in distantly-related species. For instance, apart from the
201 U1 snRNA, the sequence homology between maternal-type and somatic-type zebrafish
202 snRNA variants is lower than between the somatic-type zebrafish variant and the human
203 snRNA sequences (Supplemental Figure SFig1). The considerable sequence difference of the
204 snRNA variants within the zebrafish hints on an intriguing need for two very different
205 spliceosomal systems in early embryogenesis and adult zebrafish.

206 The genomic distribution of the maternal-type and somatic-type snRNA variants is also
207 similar to that of the small snoRNAs and 5S rRNAs in that U1, U5 and U6 have a relative high
208 number of maternal-type snRNA genes, whereas there are only a few somatic-type genes.
209 This correlates with the fact that the U1, U5 and U6 are the snRNA that directly interact with
210 the mRNA. However, it is still somewhat surprising that the maternal-type U2 variant is
211 present only in two loci. Also similar to 5S rRNA, the majority of multiple U1, U5 and U6
212 snRNA genes are organized in a strictly repeated manner, but with several interruptions
213 caused by retrotransposons (26). It was also clear that some snRNA genes were only
214 partially present in the genome (Supplemental Table ST1). Together this would be in line
215 with the reported presence of retrotransposons in snRNA gene repeats (33,34). A quick scan
216 of these snRNA interrupting sequences, revealed the presence of retrotransposons with and
217 without intact open reading frames. Alike 5S loci, we expect that retrotransposons might
218 play a role in maintaining the needed number of gene copies in the snRNA repeats.
219 The strict differential expression of maternal-type and somatic-type snRNAs obviously is
220 regulated via the promotor and auxiliary sequences related to the snRNA genes and we
221 expect that an extensive analysis of all maternal-type and somatic-type snRNA promoter
222 regions will eventually discover the relevant differences. This holds also for mechanisms
223 that are involved in snRNA processing and degradation. For instance, a superficial
224 comparison suggests that the core sequences of the 3' box, which is involved in the post
225 transcriptional processing of snRNAs (35), are different between the maternal-type and
226 somatic-type snRNA variants (results not shown).
227 In many other species different snRNA variants have been found, often with differential expression
228 during embryogenesis, such as in *Xenopus* (13), Human (36), mouse (37), sea urchin (17), *Drosophila*
229 (12). The observed zebrafish snRNA variant system is unique in that the maternal-type snRNA is
230 exclusively expressed in oocytes, eggs, and early embryogenesis. All other reports just mention up to
231 40% higher expression of some variants. This implies that by means of maternal-type snRNAs,
232 together with the maternal-type rRNAs and maternal-type snoRNAs we previously reported,
233 zebrafish employ a distinct translation system specifically for early embryogenesis
234

235 MATERIAL AND METHODS

236 Biological materials

237 Adult zebrafish (strain ABTL) were handled in compliance with local animal welfare
238 regulations and maintained according to standard protocols (<http://zfin.org>). The breeding
239 of adult fish was approved by the local animal welfare committee (DEC) of the University of
240 Leiden, the Netherlands. All protocols adhered to the international guidelines specified by
241 the EU Animal Protection Directive 86/609/EEC.

242 For this study samples were used of two pools of unfertilized eggs (oocyte clutches) and two
243 whole male adult zebrafish. The harvesting of the biological materials and RNA-isolation
244 have been described previously (26,27).

245 Source data

246 In this study we use the zebrafish genome version GRCz11 and next-generation sequencing
247 data previously generated by our group (26) and available through the BioProject database
248 with accession number PRJNA347637.

249 qRT-PCR analysis

250 For snRNA U1 and U5, forward and reverse PCR primers were used from the RT-PCR-qSeq
251 analysis, and quantitative real-time PCR (qRT-PCR) probes were designed (Supplemental
252 Table ST4). Reverse transcription was done in two independent reactions for zebrafish
253 clutch (= egg pool) and adult male tail total RNA. SuperScript IV Reverse Transcriptase
254 (Thermo Fisher Scientific) was used according to the manufacturer's instructions. Separate
255 qRT-PCR analyses were performed on 10-fold dilutions of the cDNAs with the snRNA U1 and
256 U5 primer/probe combinations using a QuantStudio 3 Real-Time PCR System (Thermo Fisher
257 Scientific).

258 RT-PCR-qSeq analysis

259 Forward and reverse PCR primers were designed for maternal-type and somatic-type snRNA
260 variants, in such a way that: 1) as much as possible of the 5'-end of the full-length variants is
261 included in the final amplicon, and 2) generic primers will bind to the maternal-type, as well
262 as the somatic-type variants (Supplemental Table ST4 and Figure 1A). To avoid positive
263 results due to genomic DNA background, small RNA-enriched total RNA was treated twice
264 with 5 µl of RNase-free DNase (Qiagen) for 45 minutes at 37°C. Next, cDNA was prepared

265 from 50 ng of RNA as described in (26). For each sample, reverse transcription was primed
266 using a mixture of all reverse PCR primers. Controls without reverse transcriptase were used
267 to exclude genomic DNA contamination of the RNA in the downstream PCR. Subsequently,
268 standard PCR reactions were performed on 1 μ l of cDNA using Q5 High-Fidelity DNA
269 Polymerase (New England Biolabs) and each of the variant PCR primer pairs independently.
270 The resulting amplicons were purified using the QIAquick PCR Purification Kit (Qiagen)
271 according to the manufacturer's instructions, with the exception that a total of seven
272 volumes of solution PB was added to allow for more efficient binding of fragments <100 bp.
273 Next, the PCR products were phosphorylated using T4 PNK (New England Biolabs) and again
274 purified as described above. Afterwards, the size of the PCR product was verified on a 2200
275 TapeStation System (Agilent). From 44 ng purified phosphorylated PCR product, barcoded
276 sequencing libraries were prepared using a modified version of the Ion Xpress Plus Fragment
277 Library Kit (Thermo Fisher Scientific) as described previously (26). Massive-parallel
278 sequencing was performed on an Ion Proton System (Thermo Fisher Scientific) using an Ion
279 PI Chip Kit v3.

280 Bioinformatics analyses

281 *Known snRNA sequences.* The initial set of snRNA sequences of *D. rerio* (GRCz11) were
282 downloaded from Ensemble 95 (29) in October 2018 using Biomart by selecting snRNA as
283 *Gene type* (See the RF annotated tabs in Supplementary Table1).

284 *Discarding “contaminated” snRNA sequences.* For each of the five snRNAs a multiple
285 alignment of the downloaded sequences was made. This was done using CLC Genomic
286 Workbench with default settings (gap open cost 20, gap extension 20 end gap cost free, very
287 accurate). Via visual inspection of the alignments (Supplementary Figure SF1) the snRNA
288 sequences that contained obvious contaminating, non-snRNA sequences, such as
289 retrotransposon sequences, were discarded or truncated.

290 *Discovery of new snRNA sequences and loci.* The known snRNA sequences were then aligned
291 with the zebrafish genome using BLASTN and all filters and masking unselected. The hits
292 that were at least 95% of the query length were selected. Names were then assigned to the
293 unique sequences.

294 *Mapping of the smallRNA-seq reads.* The reads from the source data were mapped to the
295 major and minor snRNA sequences using default settings of Bowtie-2 (31).

296 *Analysis of the RT-PCR-qSeq.* To count the number of maternal-type and somatic-type reads
297 in egg and adult male zebrafish, an exact string search of the maternal-type and somatic-
298 type snRNA sequences was performed for each snRNA. Data normalization was done using
299 the total count of mapped smallRNA-seq reads per sample.

300

301 [.ACKNOWLEDGMENTS](#)

302 We acknowledge the support of The Netherlands Organization for Scientific Research
303 (NWO) grant number 834.12.003

304

305 [SUPPLEMENTAL FILES](#)

306 SF1: Zebrafish snRNA sequence alignments
307 SF2: Alignment of snRNA sequences to selected maternal-type and somatic-type variant
308 sequences

309 [SUPPLEMENTAL TABLES](#)

310 ST1.xlsx: Genome distribution of zebrafish snRNAs
311 ST2.xlsx: Minor spliceosome snRNA read counts
312 ST3.xlsx: Read counts from the RT-PCR-qSeq experiment
313 ST4.xlsx: Primer sequences

314 [SUPPLEMENTAL FIGURES](#)

315 Sfig1.pdf: Sequence alignments of human snRNAs with maternal-type and somatic-type
316 zebrafish snRNA
317 Sfig2.pdf:qRT-PCR results for snRNA U1 and snRNA U5

318 REFERENCES

- 319 1. Chen M, Manley JL. Mechanisms of alternative splicing regulation: Insights from
320 molecular and genomics approaches. *Nat Rev Mol Cell Biol* [Internet].
321 2009;10(11):741–54. Available from: <http://dx.doi.org/10.1038/nrm2777>
- 322 2. Walter Gilbert. Why genes in pieces? *Nature*. 1978;271(9):501.
- 323 3. Shi Y. Mechanistic insights into precursor messenger RNA splicing by the spliceosome.
324 *Nat Rev Mol Cell Biol* [Internet]. 2017;18(11):655–70. Available from:
325 <http://dx.doi.org/10.1038/nrm.2017.86>
- 326 4. Turunen JJ, Niemelä EH, Verma B, Frilander MJ. The significant other: Splicing by the
327 minor spliceosome. *Wiley Interdiscip Rev RNA*. 2013;4(1):61–76.
- 328 5. Chen L, Lullo DJ, Ma E, Celniker SE, Rio DC, Doudna JA. Identification and analysis of
329 U5 snRNA variants in *Drosophila*. *RNA*. 2005;11(10):1473–7.
- 330 6. Sontheimer EJ, Steitz JA. Three novel functional variants of human U5 small nuclear
331 RNA. *Mol Cell Biol*. 2015;12(2):734–46.
- 332 7. O'Reilly D, Dienstbier M, Cowley SA, Vazquez P, Drozdz M, Taylor S, et al.
333 Differentially expressed, variant U1 snRNAs regulate gene expression in human cells.
334 *Genome Res*. 2013;23(2):281–91.
- 335 8. Boonanuntanasarn S, Panyim S, Yoshizaki G. Characterization and organization of the
336 U6 snRNA gene in zebrafish and usage of their promoters to express short hairpin
337 RNA. *Mar Genomics* [Internet]. 2008;1(3–4):115–21. Available from:
338 <http://dx.doi.org/10.1016/j.margen.2008.10.001>
- 339 9. Domitrovich AM, Kunkel GR. Multiple, dispersed human U6 small nuclear RNA genes
340 with varied transcriptional efficiencies. *Nucleic Acids Res*. 2003;31(9):2344–52.
- 341 10. Vazquez-Arango P, O'Reilly D. Variant snRNPs: New players within the spliceosome
342 system. *RNA Biol* [Internet]. 2018;15(1):17–25. Available from:
343 <https://doi.org/10.1080/15476286.2017.1373238>
- 344 11. Lo PCH, Mount SM. *Drosophila melanogaster* genes for U1 snRNA variants and their
345 expression during development. *Nucleic Acids Res*. 1990;18(23):6971–9.
- 346 12. Lu Z, Matera AG. Developmental Analysis of Spliceosomal snRNA Isoform Expression.
347 *G3& Genes|Genomes|Genetics* [Internet]. 2015;5(1):103–10. Available
348 from: <http://g3journal.org/lookup/doi/10.1534/g3.114.015735>
- 349 13. Forbes DJ, Kirschner MW, Caput D, Dahlberg JE, Lund E. Differential expression of
350 multiple U1 small nuclear RNAs in oocytes and embryos of *Xenopus laevis*. *Cell*.
351 1984;38(3):681–9.
- 352 14. Lund E, Bostock CJ, Dahlberg JE. The transcription of *Xenopus laevis* embryonic U1
353 snRNA genes changes when oocytes mature into eggs. *Genes Dev*. 1987;1(1):47–56.

354 15. Lund E, Dahlberg JE. Differential accumulation of U1 and U4 small nuclear RNAs
355 during Xenopus development. *Genes Dev.* 1987;1(1):39–46.

356 16. Cheng Y, Lund E, Kahan BW, Dahlberg JE. Control of mouse U1 snRNA gene expression
357 during in vitro differentiation of mouse embryonic stem cells. *Nucleic Acids Res.*
358 1997;25(11):2197–204.

359 17. Nash MA, Sakallah S, Santiago C, Yu J-C, Marzluff F. Switch in Sea Urchin U1 RNA. *Dev
360 Biol.* 1989;296(134):289–96.

361 18. Stefanovic B, Li JM, Sakallah S, Marzluff WF. Isolation and characterization of
362 developmentally regulated sea urchin U2 snRNA genes. *Dev Biol.* 1991;148(1):284–
363 94.

364 19. Schmidt D, Reuter H, Hüttner K, Ruhe L, Rabert F, Seebeck F, et al. The Integrator
365 complex regulates differential snRNA processing and fate of adult stem cells in the
366 highly regenerative planarian Schmidtea mediterranea. *PLoS Genet.*
367 2018;14(12):e1007828.

368 20. Sierra-Montes JM, Freund A V., Ruiz LM, Szmulewicz MN, Rowold DJ, Herrera RJ.
369 Multiple forms of U2 snRNA coexist in the silk moth *Bombyx mori*. *Insect Mol Biol.*
370 2002;11(1):105–14.

371 21. Hinias A, Larsson P, Avesson L, Kirsebom LA, Virtanen A, Söderbom F. Identification of
372 the Major Spliceosomal RNAs in *Dictyostelium discoideum* Reveals Developmentally
373 Regulated U2 Variants and Polyadenylated snRNAs. *Eukaryot Cell.* 2006;5(6):924–34.

374 22. Korf GM, Botros IW, Stumph WE. Developmental and tissue-specific expression of U4
375 small nuclear RNA genes. *Mol Cell Biol.* 2015;8(12):5566–9.

376 23. Vazquez-Arango P, Vowles J, Browne C, Hartfield E, Fernandes HJR, Mandefro B, et al.
377 Variant U1 snRNAs are implicated in human pluripotent stem cell maintenance and
378 neuromuscular disease. *Nucleic Acids Res.* 2016;44(22):10960–73.

379 24. Leighton L, Bredy T. Functional Interplay between Small Non-Coding RNAs and RNA
380 Modification in the Brain. *Non-Coding RNA.* 2018;4(2):15.

381 25. Pagano JFB, Locati MD, Ensink WA, van Olst M, van Leeuwen S, de Leeuw WC, et al.
382 Maternal- and somatic-type snoRNA expression and processing in zebrafish
383 development. Submitted.

384 26. Locati MD, Pagano JFB, Ensink WA, van Olst M, van Leeuwen S, Nehrdich U, et al.
385 Linking maternal and somatic 5S rRNA types with different sequence-specific non-LTR
386 retrotransposons. *RNA [Internet].* 2017;23(4):446–56. Available from:
387 <http://rnajournal.cshlp.org/lookup/doi/10.1261/rna.059642.116>

388 27. Locati MD, Pagano JFB, Girard G, Ensink WA, van Olst M, van Leeuwen S, et al.
389 Expression of distinct maternal and somatic 5.8S, 18S, and 28S rRNA types during
390 zebrafish development. *RNA [Internet].* 2017;23(8):1188–99. Available from:
391 <http://rnajournal.cshlp.org/lookup/doi/10.1261/rna.061515.117>

392 28. Locati MD, Pagano JFB, Abdullah F, Ensink WA, van Olst M, van Leeuwen S, et al.
393 Identifying small RNAs derived from maternal- and somatic-type rRNAs in Zebrafish
394 Development. *Genome* [Internet]. 2018;gen-2017-0202. Available from:
395 <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC6200000/>

396 29. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018.
397 *Nucleic Acids Res.* 2018;46(D1):D754–61.

398 30. Kojima KK. Cross-Genome Screening of Novel Sequence-Specific Non-LTR
399 Retrotransposons: Various Multicopy RNA Genes and Microsatellites Are Selected as
400 Targets. *Mol Biol Evol* [Internet]. 2003;21(2):207–17. Available from:
401 <https://academic.oup.com/molbev/article-lookup/doi/10.1093/molbev/msg235>

402 31. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*.
403 2012 Apr;9(4):357–9.

404 32. McGrail JC, O'Keefe RT. The U1, U2 and U5 snRNAs crosslink to the 5' exon during
405 yeast pre-mRNA splicing. *Nucleic Acids Res.* 2008;36(3):814–25.

406 33. Kojima KK, Jurka J. A Superfamily of DNA Transposons Targeting Multicopy Small RNA
407 Genes. *PLoS One*. 2013;8(7):1–7.

408 34. Van Arsdell SW, Denison RA, Bernstein LB, Weiner AM, Manser T, Gesteland RF.
409 Direct repeats flank three small nuclear RNA pseudogenes in the human genome.
410 *Cell*. 1981;26(1 PART 1):11–7.

411 35. Matera AG, Wang Z. A day in the life of the spliceosome. *Nat Rev Mol Cell Biol*.
412 2014;15(2):108–21.

413 36. Lund E. Heterogeneity of human U1 snRNAs. *Nucleic Acids Res*. 1988;16(13):5813–26.

414 37. Lund E, Kahan B, Dahlberg JE. Differential control of U1 small nuclear RNA expression
415 during mouse development. *Science (80-)*. 1985;229(4719):1271–4.

416

417

A**U1**

M ATACTTACCTGGCAGGGGAGATACCATGATCAAGAAGGTGGTCACCCAGGGCGAGGCTTGGCCATTGCA
 S C T CA

M CT-CCGGCCACGCTGACCCCTGCGAATTCCCCAAATGTGGGAATCTCGACTGCATAATTCTGGTAGTGG
 S .. GT TGT. T T

M GGGACTGCGTTCGCGCTCTCCCTG
 S 11/165

U2

M ATCGCTTCTCGGCCTTTGGCTAAGATCAAGTGTAGTATCTGTTCTTATCAGTTAACATCTGATACGTG
 S

M CCCTACCCGGGCACTATATATTAAATTGATTTGCAGCTGGGAGATGGGTTAGGGCTTGCTCCACCCA
 S C G .. A AA GT

M CTCCACGCATCGACCTGGTATTGCAGTACCTCCGGGAACGGTGCACTCATT
 S T C A C. CC. 13/191

U4

bioRxiv preprint doi: <https://doi.org/10.1101/858944>; this version posted November 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

M ATTCTTCGCAGAGGCAATATCATAGCCTATGAGGCCTAGCCGAGGCCTGATCATTGCTGGTTGAAA
 S .GCT..G.....T.....G.....TT..T.....C..T.....AA.....

M TTACCCAATACCCCGCCGAGATGGCTTGAAATACAGTCAGCTCTGGCAATTGGTACAGCCTCCTAGAAG
 S T .. C. A T .. G .. A G. T .. GACAG ..

M G
 S A 28/141

U5

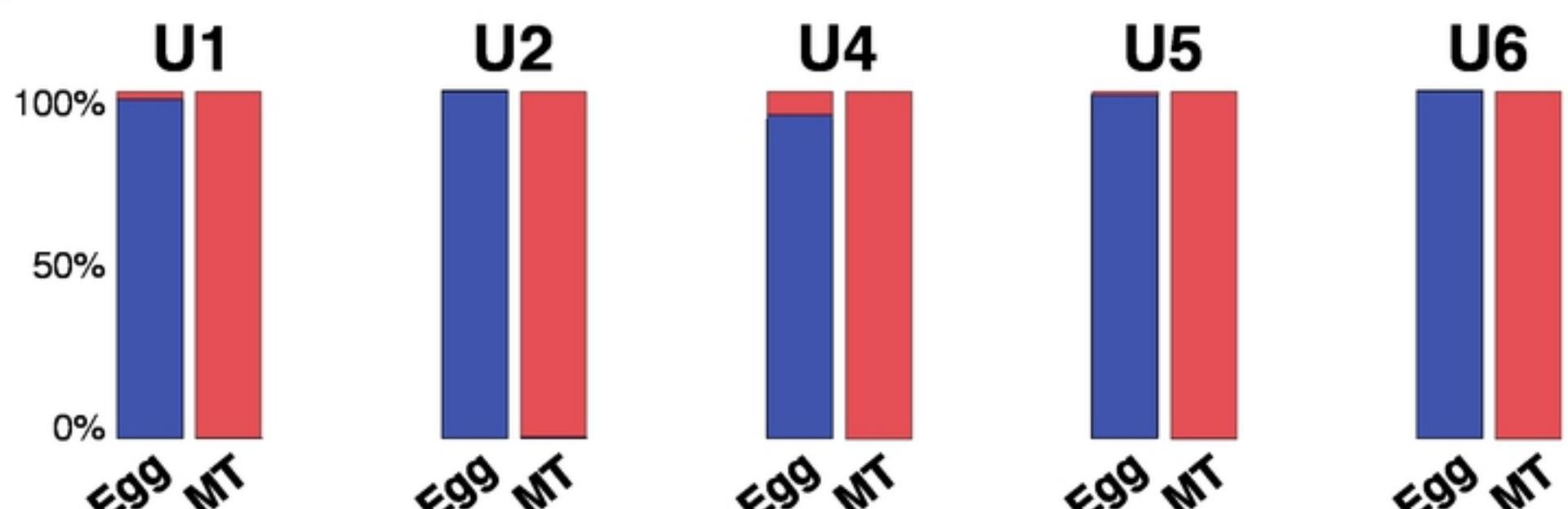
M ACAGCTCGAGTTCTCTCAATCACGCACAAATCTTCGCCTTTACTAAAGATTCCGTGGGAGGAAC
 S GT. CACTC. TG..T. TCCTC. GAT..T. T.....A.....

M TATTGTGAGTTTGTATTGGGTGCTCTGCTATCTGCAGAGCTG
 S AC. ATGAGTA. AAAC. A.....T. A .. C .. CGA. A. GC. G. GCT 48/118

U6

M GTGCTCGCTACGGTGGCACATATACTAAAATTGGATCGATACAGAGAAGATTAGCATGGCCCTGCGAAA
 S T .. T .. CA .. A .. C ..

M GGATGACACGCAAATCCGTGAAGCGCTCCATATTCT
 S T .. T .. C .. TA 11/118

B**Figure 1**

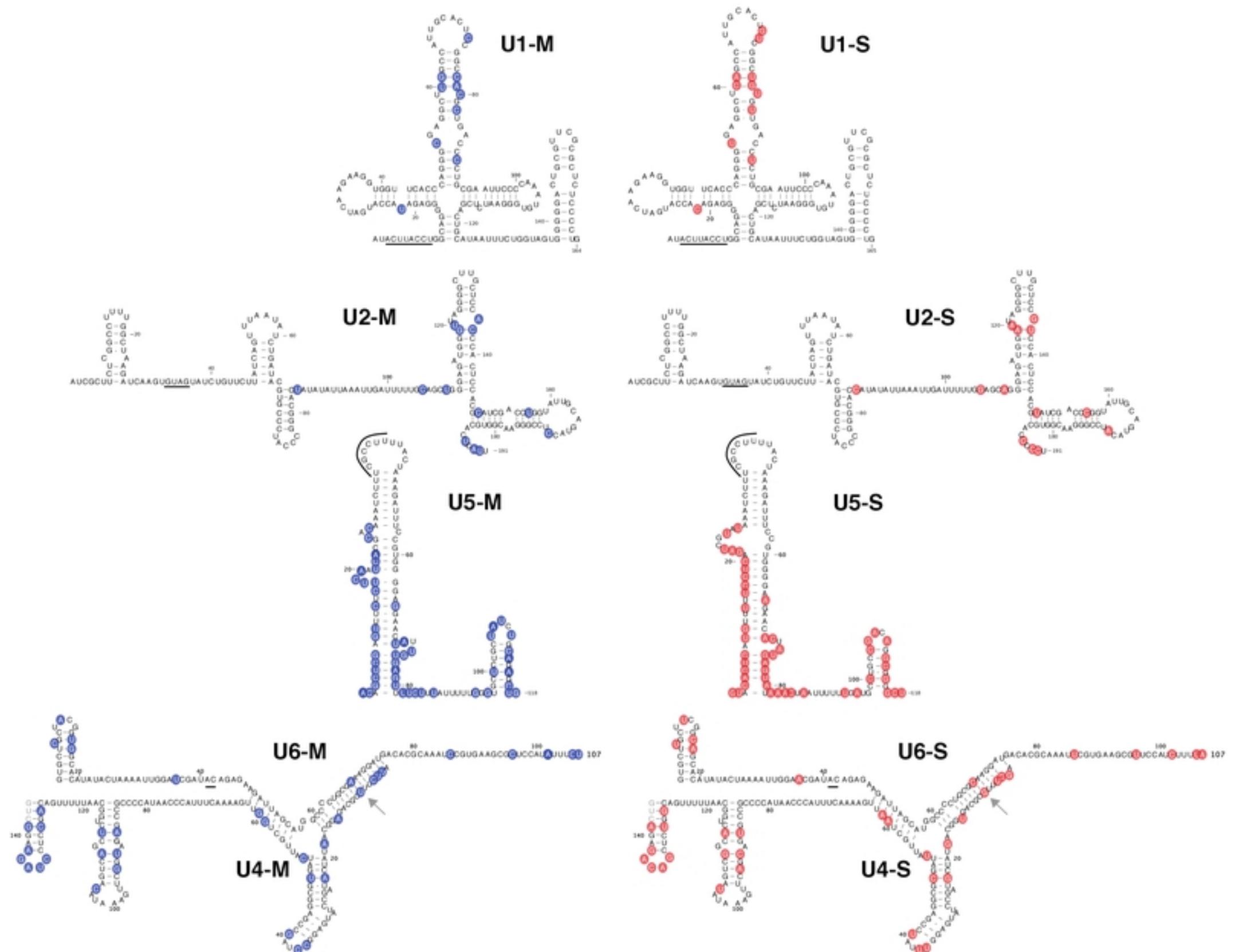


Figure 2