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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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.

An alternative spliceosome defined by distinct snRNAs in early zebrafish embryogenesis

Johanna F. B. Pagano[1], Rob J. Dekker[1], Wim A. Ensink[1], Marina van Olst[1], Alex Bos[1], Selina van Leeuwen[1], Wim C. de Leeuw[1], Ulrike Nehrdich[2], Herman P. Spaink[2], Han Rauwerda[1], Martijs J. Jonker[1] 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 Observatorium, Leiden 2333 CE, the Netherlands

Alternative snRNAs in early zebrafish embryogenesis

snRNA, spliceosome, dual translation machinery, embryogenesis

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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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.

ABSTRACT

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

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.

INTRODUCTION

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

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

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

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

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.

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

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

RESULTS AND DISCUSSION

Cataloguing the zebrafish snRNAs

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

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.

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

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

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

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

		Chromosome														
	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	
minor snRNA																
U11	1												1			
U12	1		1													
U4atc	1									1						
U6atc	1														1	

Embryogenesis-specific snRNA variants

After cataloguing all snRNAs in the zebrafish genome, we investigated whether embryogenesis-specific snRNAs exist by examining their expression in eggs and somatic

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.

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

To check the RT-PCR-qSeq results we developed a qRT-PCR analyses for snRNA U1 and snRNA U5. This qRT-PCR analysis confirmed the exclusive presence of embryogenesis-specific snRNA variants in egg samples (Supplemental Figure SFig2).

Figure 1. Maternal-type and somatic-type snRNA sequences and expression

A: Sequence comparison of selected maternal-type (M) and somatic-type (S) snRNA variants; identical nucleotides are indicated as dots, while gaps as dashes (For sequence alignment of all snRNAs cf. Supplemental Figure SF2). The RT-PCR primers are indicated with half arrows: maternal-type specific (blue); somatic-type specific (red); and non-distinctive (green).

B: Relative expression of maternal-type (blue) and somatic-type (red) variants for each snRNA indicated by comparative percentage calculated using the RT-PCR-qSeq approach on RNA from egg and male tail tissue (MT).

Differences between maternal- and somatic-types snRNAs

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

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.

nucleotide sequences that bind to the mRNA, thus allowing each system to splice distinct introns. However, even though there are many sequence differences between the maternal-type and the somatic-type snRNAs, none of them involve the mRNA binding sites in these snRNAs (Figure 2 and Supplement File SF3). It turned out that the snRNA sequence differences are often located in specific parts of the secondary structure (Figure 2). For instance, for U1, all but one differences are located in one stem-loop and in U2 they are confined to one side of the structure (Figure 2). In general, many of the differences are found in the loops, which are thought to be specific binding locations for spliceosomal proteins. Hence, this would indicate that the embryogenesis spliceosome, besides specific snRNAs also may comprise (embryogenesis-)specific proteins.

Despite many apparently co-evolved nucleotide pairs in stems of the snRNAs (Figure 2), there seems to be only one co-evolved nucleotide pair in the interaction site between U4 and U6. However, as this nucleotide is right in the middle of the largest interaction site, it might actually prevent the binding of somatic-type U4 to maternal-type U6 and vice versa.

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

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

CONCLUDING REMARKS

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

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.

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

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

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

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.

The genomic distribution of the maternal-type and somatic-type snRNA variants is also similar to that of the small snoRNAs and 5S rRNAs in that U1, U5 and U6 have a relative high number of maternal-type snRNA genes, whereas there are only a few somatic-type genes. This correlates with the fact that the U1, U5 and U6 are the snRNA that directly interact with the mRNA. However, it is still somewhat surprising that the maternal-type U2 variant is present only in two loci. Also similar to 5S rRNA, the majority of multiple U1, U5 and U6 snRNA genes are organized in a strictly repeated manner, but with several interruptions caused by retrotransposons (26). It was also clear that some snRNA genes were only partially present in the genome (Supplemental Table ST1). Together this would be in line with the reported presence of retrotransposons in snRNA gene repeats (33,34). A quick scan of these snRNA interrupting sequences, revealed the presence of retrotransposons with and without intact open reading frames. Alike 5S loci, we expect that retrotransposons might play a role in maintaining the needed number of gene copies in the snRNA repeats.
The strict differential expression of maternal-type and somatic-type snRNAs obviously is regulated via the promotor and auxiliary sequences related to the snRNA genes and we expect that an extensive analysis of all maternal-type and somatic-type snRNA promoter regions will eventually discover the relevant differences. This holds also for mechanisms that are involved in snRNA processing and degradation. For instance, a superficial comparison suggests that the core sequences of the 3' box, which is involved in the post transcriptional processing of snRNAs (35), are different between the maternal-type and somatic-type snRNA variants (results not shown).

In many other species different snRNA variants have been found, often with differential expression during embryogenesis, such as in *Xenopus* (13), Human (36), mouse (37), sea urchin (17), *Drosophila* (12). The observed zebrafish snRNA variant system is unique in that the maternal-type snRNA is exclusively expressed in oocytes, eggs, and early embryogenesis. All other reports just mention up to 40% higher expression of some variants. This implies that by means of maternal-type snRNAs, together with the maternal-type rRNAs and maternal-type snoRNAs we previously reported, zebrafish employ a distinct translation system specifically for early embryogenesis

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.

MATERIAL AND METHODS

Biological materials

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

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

Source data

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

qRT-PCR analysis

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

RT-PCR-qSeq analysis

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

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.

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

Bioinformatics analyses

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

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

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

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.

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

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

.ACKNOWLEDGMENTS

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

SUPPLEMENTAL FILES

SF1: Zebrafish snRNA sequence alignments

SF2: Alignment of snRNA sequences to selected maternal-type and somatic-type variant sequences

SUPPLEMENTAL TABLES

ST1.xlsx: Genome distribution of zebrafish snRNAs

ST2.xlsx: Minor spliceosome snRNA read counts

ST3.xlsx: Read counts from the RT-PCR-qSeq experiment

ST4.xlsx: Primer sequences

SUPPLEMENTAL FIGURES

Sfig1.pdf: Sequence alignments of human snRNAs with maternal-type and somatic-type zebrafish snRNA

Sfig2.pdf:qRT-PCR results for snRNA U1 and snRNA U5

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.

REFERENCES

1. Chen M, Manley JL. Mechanisms of alternative splicing regulation: Insights from molecular and genomics approaches. Nat Rev Mol Cell Biol [Internet]. 2009;10(11):741–54. Available from: http://dx.doi.org/10.1038/nrm2777

2. Walter Gilbert. Why genes in pieces? Nature. 1978;271(9):501.

3. Shi Y. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. Nat Rev Mol Cell Biol [Internet]. 2017;18(11):655–70. Available from: http://dx.doi.org/10.1038/nrm.2017.86

4. Turunen JJ, Niemelä EH, Verma B, Frilander MJ. The significant other: Splicing by the minor spliceosome. Wiley Interdiscip Rev RNA. 2013;4(1):61–76.

5. Chen L, Lullo DJ, Ma E, Celniker SE, Rio DC, Doudna JA. Identification and analysis of U5 snRNA variants in Drosophila. RNA. 2005;11(10):1473–7.

6. Sontheimer EJ, Steitz JA. Three novel functional variants of human U5 small nuclear RNA. Mol Cell Biol. 2015;12(2):734–46.

7. O'Reilly D, Dienstbier M, Cowley SA, Vazquez P, Drozdz M, Taylor S, et al. Differentially expressed, variant U1 snRNAs regulate gene expression in human cells. Genome Res. 2013;23(2):281–91.

8. Boonanuntanasarn S, Panyim S, Yoshizaki G. Characterization and organization of the U6 snRNA gene in zebrafish and usage of their promoters to express short hairpin RNA. Mar Genomics [Internet]. 2008;1(3–4):115–21. Available from: http://dx.doi.org/10.1016/j.margen.2008.10.001

9. Domitrovich AM, Kunkel GR. Multiple, dispersed human U6 small nuclear RNA genes with varied transcriptional efficiencies. Nucleic Acids Res. 2003;31(9):2344–52.

10. Vazquez-Arango P, O'Reilly D. Variant snRNPs: New players within the spliceosome system. RNA Biol [Internet]. 2018;15(1):17–25. Available from: https://doi.org/10.1080/15476286.2017.1373238

11. Lo PCH, Mount SM. Drosophila melanogaster genes for U1 snRNA variants and their expression during development. Nucleic Acids Res. 1990;18(23):6971–9.

12. Lu Z, Matera AG. Developmental Analysis of Spliceosomal snRNA Isoform Expression. G3: Genes|Genomes|Genetics [Internet]. 2015;5(1):103–10. Available from: http://g3journal.org/lookup/doi/10.1534/g3.114.015735

13. Forbes DJ, Kirschner MW, Caput D, Dahlberg JE, Lund E. Differential expression of multiple U1 small nuclear RNAs in oocytes and embryos of Xenopus laevis. Cell. 1984;38(3):681–9.

14. Lund E, Bostock CJ, Dahlberg JE. The transcription of Xenopus laevis embryonic U1 snRNA genes changes when oocytes mature into eggs. Genes Dev. 1987;1(1):47–56.

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

28. Locati MD, Pagano JFB, Abdullah F, Ensink WA, van Olst M, van Leeuwen S, et al. Identifying small RNAs derived from maternal- and somatic-type rRNAs in Zebrafish Development. Genome [Internet]. 2018;gen-2017-0202. Available from: http://www.nrcresearchpress.com/doi/10.1139/gen-2017-0202

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

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

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

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

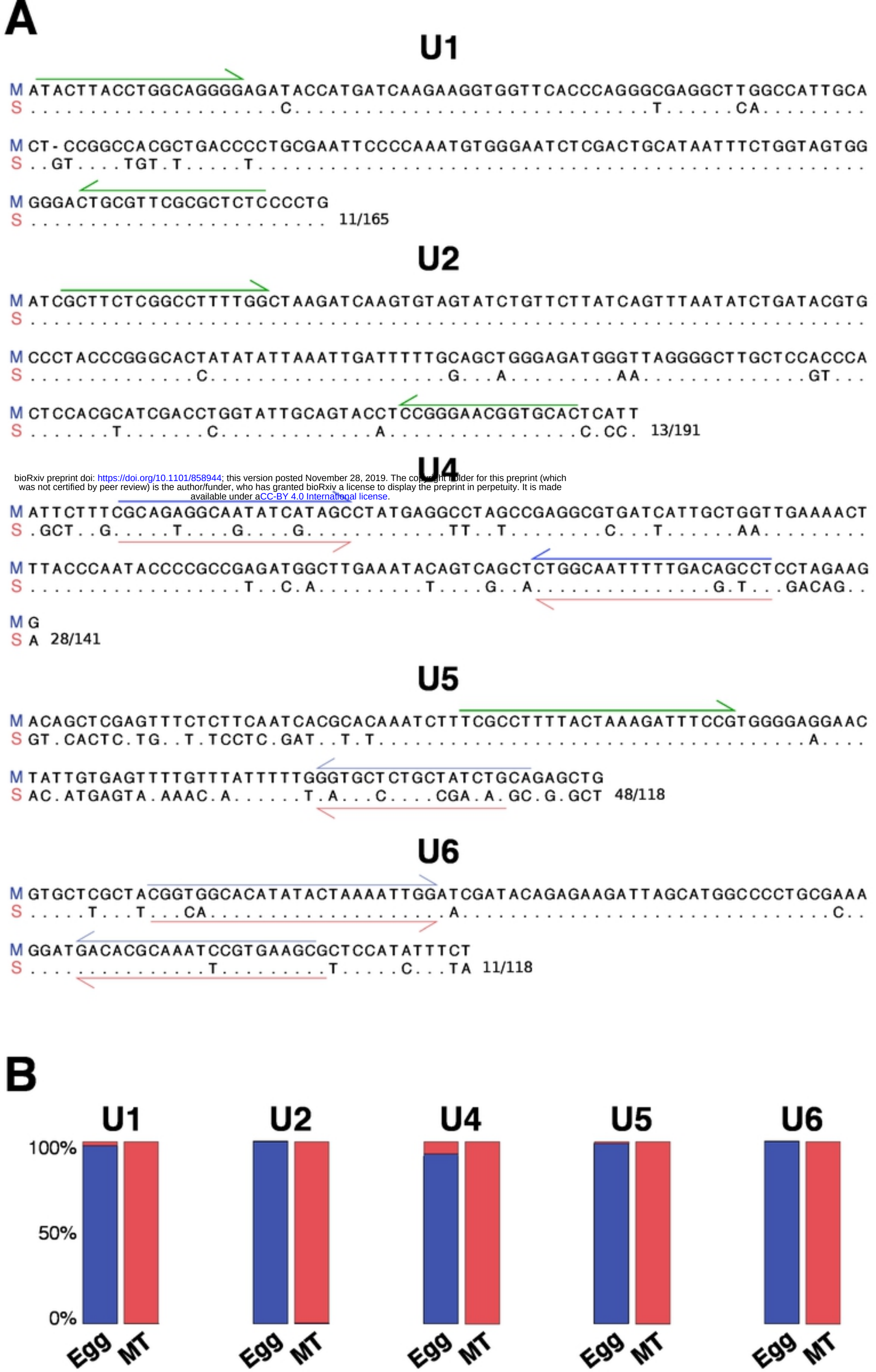
33. Kojima KK, Jurka J. A Superfamily of DNA Transposons Targeting Multicopy Small RNA Genes. PLoS One. 2013;8(7):1–7.

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

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

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

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



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.

Figure 1

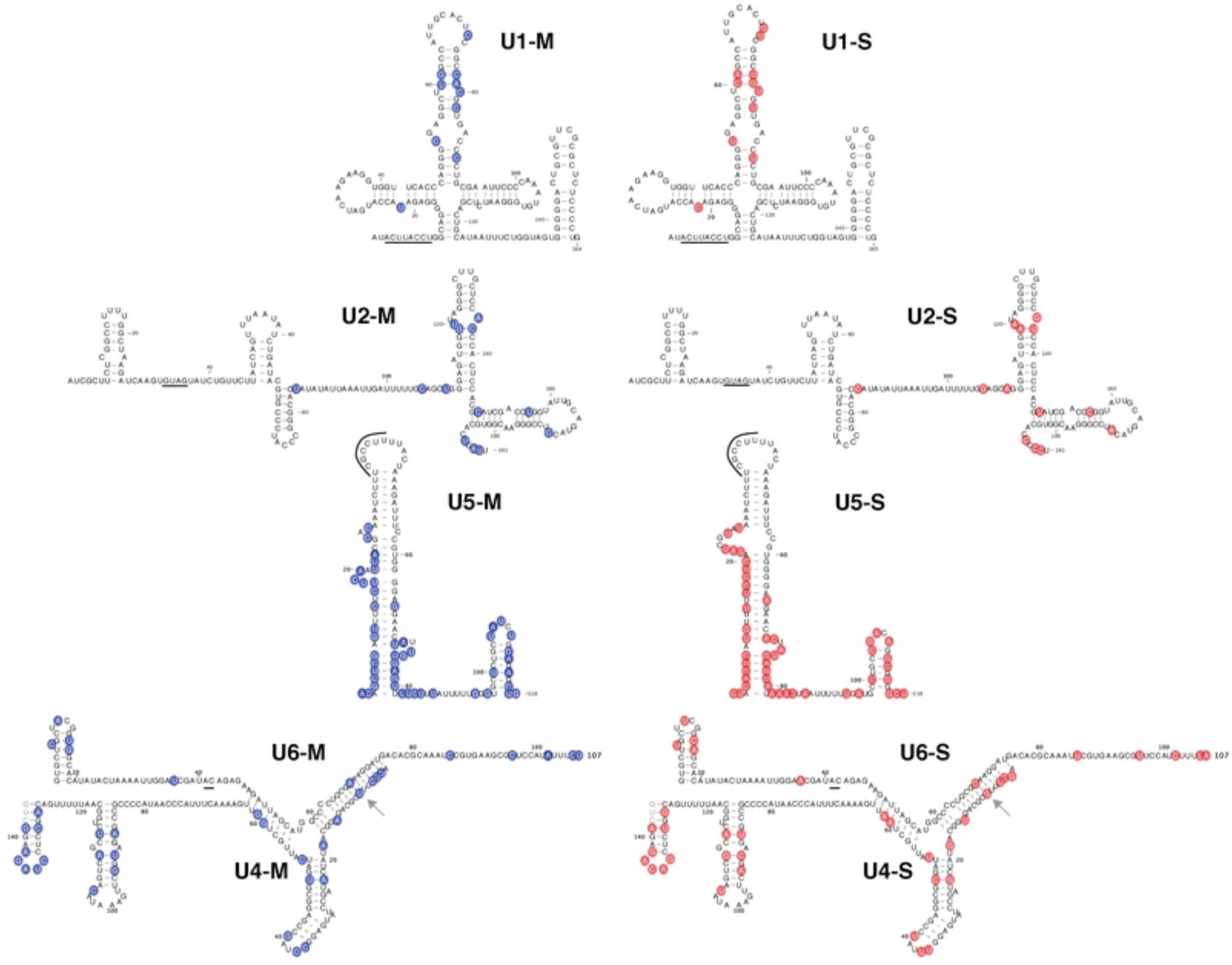


Figure 2