Maternal- and Somatic-type snoRNA Expression and Processing in Zebrafish Development

Johanna F. B. Pagano[1+], Mauro D. Locati[1+], Wim A. Ensink[1], Marina van Olst[1], Selina van Leeuwen[1], Wim C. de Leeuw[1], Ulrike Nehrdich[2], Herman P. Spaink[2], Han Rauwerda[1], Martijs J. Jonker[1], Rob J. Dekker[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

[+] First two authors contributed equally to this publication

[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

Maternal- and somatic-type snoRNAs in zebrafish

U3 U8 snoRNA, maternal expression, RNA processing, embryogenesis

ABSTRACT

Small nucleolar RNAs (snoRNAs) are non-coding RNAs that play an important role in the complex maturation process of ribosomal RNAs (rRNAs). SnoRNAs are categorized in classes, with each class member having several variants present in a genome. Similar to our finding of specific rRNA expression types in zebrafish embryogenesis, we discovered preferential maternal- and somatic-expression for snoRNAs. Most snoRNAs and their variants have higher expression levels in somatic tissues than in eggs, yet we identified three snoRNAs; U3, U8 and snoZ30 of which specific variants show maternal- or somatic-type expression. For U3 and U8 we also found small-derived snoRNAs that lack their 5' rRNA recognition part and are essentially Domain II hairpin structures (U-DII). These U-DII snoRNAs from variants showed similar preferential expression, in which maternal-type variants are prominently expressed in eggs and subsequently replaced by a somatic-type variants during embryogenesis. This differential expression is related to the organization in tandem repeats (maternal type) or solitary (somatic-type) genes of the involved U snoRNA loci. The collective data showed convincingly that the preferential expression of snoRNAs is achieved by transcription regulation, as well as through RNA processing. Finally, we observed small-RNAs derived from internal transcribed spacers (ITSs) of a U3 snoRNA loci that via complementarity binding, may be involved in the biosynthesis of U3-DII snoRNAs. Altogether, the here described maternal- and somatic-type snoRNAs are the latest addition to the developing story about the dual ribosome system in zebrafish development.

INTRODUCTION

Small nucleolar RNAs (snoRNAs) are a class of non-coding RNA molecules of variable length (the majority being 60-200 nucleotides long), found in archaea and eukaryotes (1). SnoRNAs

are thought to mainly be involved in post-transcriptional modifications and maturation of ribosomal RNAs (rRNAs) (2,3). However, recently additional functions have been ascribed to specific snoRNAs, from regulation of mRNA editing and splicing (4) to post-transcriptional gene silencing (5,6). SnoRNAs do not possess any intrinsic catalytic or modification activity, but act both as a scaffold for partner proteins, forming small nucleolar ribonucleoproteins (snoRNPs) and as guide for target specificity (7). Based on base-pairing interactions with their target RNA, snoRNAs can thus direct the associated catalytic protein subunits to accurately modify a specific RNA site (8).

In general, eukaryotic genomes can contain up to 200+ unique snoRNA genes (9). Based on the presence of conserved sequence motifs, the majority of snoRNAs are classified into two distinct classes: box C/D snoRNAs, which guide 2’-O-methylation of ribose, and box H/ACA snoRNAs, which are involved in the isomerization of specific uridine residues to pseudouridine (1). C/D snoRNAs are defined by the presence of two conserved motifs, the C box (UGAUGA) and the D box (CUGA), found near the 5’- and 3’-end, respectively (10). In the folded C/D snoRNA molecule, these two motifs are in close proximity of each other by means of a hairpin structure and serve as a binding site for interacting proteins (11–13). In addition, many C/D box snoRNAs may have less-well-conserved copies of the C and D motifs (called C' and D'), which are also involved in the interaction with specific proteins (14–16). A conserved region of 7-20 nucleotides upstream of the D (and the D’, if present) box interacts via base-complementarity with the rRNA, pinpointing the to-be-methylated RNA base, which is usually the 5th nucleotide from the D or D’ box (17–20). Additional base interactions with the target rRNA can stimulate methylation by up to five-fold (16). H/ACA snoRNAs typically feature a secondary structure consisting of two hairpins linked by a hinge region that contains the H-box (ANANNA), followed by a short tail with the ACA-box (ACA) (21). The

49 hairpin regions contain internal bulges known as pseudouridylation pockets in which a
50 conserved region of 6-20 nucleotides is complementary to the target, but leaves a uridine
51 residue unpaired, marking it for enzymatic modification (22,23). However, more and more
52 snoRNAs are discovered that don't follow this classification. For example, there are snoRNAs
53 that have both C/D box and H/ACA box (24,25), or C/D snoRNAs that are not involved in 2'-
54 O-methylation, like snoRNA U3 and U8, which instead are essential in the processing of 18S
55 and 28S, 5.8S rRNAs, respectively (26–28).
56 Most vertebrate snoRNA genes are located in introns of genes that usually encode for
57 proteins related to ribosome biogenesis and protein synthesis (29) (Figure 1A). After
58 transcription of these genes by RNA polymerase II, the formation of functional snoRNAs
59 requires the processing of intronic RNA sequences, which are released during pre-mRNA
60 splicing (2). In vertebrates, only a few snoRNAs genes have an independent promoter and
61 are directly transcribed by RNA polymerase II or III (30,31). snoRNA genes are present either
62 as solitary snoRNA genes, or as clusters of multiple coding units; such clusters can consist of
63 the same or different snoRNA genes (9) (Figure 1A).
64 An increasing number of recently identified snoRNAs show differential expression among
65 different cell types and tissues, suggesting a role in distinct physiological processes (6). For
66 example, several studies discovered that certain snoRNAs are expressed mainly in the brain
67 where they might affect organ function and/or development (32,33).
68 Our group has recently described the presence of two distinct types of rRNAs expressed
69 during zebrafish development: a maternal-type, exclusively accumulated during oogenesis,
70 and replaced throughout the embryogenesis by a somatic-type, which is the only one rRNA
71 type present in adult somatic tissue (34,35). Since maternal- and somatic-type rRNAs show
72 ample sequence differences combined with the fact that snoRNAs interact with rRNA in a

73 sequence specific way, we hypothesized that there also might exist specific maternal- and
74 somatic-type snoRNAs. For this, we investigated the expression of the snoRNAome
75 throughout zebrafish development from egg to adult, by small-RNA-seq and discovered that
76 indeed complete and partial transcripts of snoRNA variants exist that show either maternal-
77 or somatic-preferential expression. Moreover, we determined that this developmentally
78 regulated expression is likely regulated both at the level of transcription, as well as RNA
79 processing. Together, with our recent finding that also a maternal-type spliceosome with
80 specific snRNA variants exists for early embryogenesis (36) everything points to the
81 existence of a comprehensive dual translation system in zebrafish embryogenesis.
82

83 RESULTS AND DISCUSSION

84 Cataloguing the zebrafish snoRNAs

85 In our previous studies (34,35), we identified two distinct types of rRNAs in zebrafish
86 development: maternal and somatic. The maternal-type rRNAs make up virtually all the
87 rRNA present in oocytes and are gradually, yet completely, replaced by the somatic-type
88 rRNAs during embryogenesis. There are, for each rRNA species, 5S, 5.8S, 18S and 28S,
89 significant sequence differences between the two types of rRNA, indicative of a substantial
90 functional difference. Given that snoRNAs are intricately involved in the complex
91 maturation process of rRNA via sequence specific interactions, we investigated whether
92 there are specific snoRNAs being co-expressed with maternal- and somatic-type rRNAs.
93 As the zebrafish genome is quite well annotated, we started by making an inventory of the
94 known genomic snoRNA sequences. In general, snoRNAs are categorized in two main
95 families based on the presence of conserved sequence motifs: the C/D box and H/ACA box
96 (37) and they can be found in the databases snOPY (38) and Ensemble 89 (39)

97 (Supplemental Table ST1 and Supplemental Figure SF1). Collectively, 257 snoRNA loci were
98 identified on the 25 zebrafish chromosomes. However, since some of these loci contained
99 identical sequences, 250 zebrafish snoRNA sequences were found to be unique
100 (Supplemental File SF1). In contrast, several snoRNAs appear to be present as different
101 variants, which we labeled with a unique identifier (Supplemental Table ST1). Most snoRNA
102 loci are present in introns (*intronic, 81%*) and rely for their expression on the transcription of
103 the associated genes (Figure 1A and 1C). Yet, a few snoRNA loci, mainly belonging to the
104 C/D box snoRNAs, are expressed as independent transcriptional units, consisting of one (*4%*)
105 or several (*15%*) snoRNA genes (Figure 1A and 1C, Table 1 and Supplemental Figure SF1).
106 Most transcriptionally self-regulated snoRNAs have been named U3 and U8 (40). The
107 snoRNA loci appear to be non-randomly distributed over the chromosomes, but no
108 apparent co-location with known rRNA sequences was observed (Supplemental Table ST1).

109 **Figure 1. Genomic organization and RNA molecules from zebrafish snoRNA genes.**
110 A: Different configurations of genomics organization of snoRNA genes illustrated by
111 exemplifying variants (upper row) and the different forms of RNA transcripts (middle rows).
112 For the U snoRNA genes also derived transcripts are indicated (lower row): snoRNA-DII
113 (green) are transcripts that exist just of the Domain-II part of the U snoRNA gene; sitsRNA
114 (yellow) are snoRNA-ITS-derived-small-RNAs that originate from ITS regions in clusters of
115 the tandemly repeated U snoRNA genes. The dotted lines indicate possible RNA processing
116 routes for snoRNA-DII transcripts.
117 B: The various snoRNA boxes are indicated in a schematic representation of the U3 snoRNA
118 (left panel). The possible interaction of a chromosome 14 U3 sitsRNA with a complete U3
119 snoRNA is indicated (middle panel), which may play a role in processing the latter at the
120 position of indicated by the red arrow to produce a U3-DII transcript. A schematic
121 representation of the U3-DII transcript found in this study (right panel).
122 C: Several characteristics of zebrafish snoRNA genes organized by their H/ACA or CD boxes.
123 [1]Distribution plot of lengths of snoRNA genes (Supplemental Figure SF1);
124 [2]snoU2_19 and SNORD94_a are present twice in the genome;
125 [3]U3_l and U3_k were found two and five times in the genome, respectively.
126

127 ## Developmental-stage-specific expression of two snoRNA types

128 After the identification and annotation of all snoRNAs in the zebrafish genome (GRCz10), we
129 investigated whether, similar to rRNAs, maternal- and somatic-type snoRNAs exist. For this,
130 we analyzed the differential expression between egg and adult zebrafish of the 204
131 expressed snoRNAs found by small-RNA-seq. The distribution of snoRNAs based on their
132 differential expression was clearly bimodal, with no snoRNA-variant being equally expressed
133 in egg and adult zebrafish (Figure 2A). Hence, all these snoRNAs displayed preferential
134 expression, as either maternal- (n = 18) or somatic-type (n = 186) snoRNA (Figure 2), albeit
135 less absolute as previously observed with maternal- and somatic-type rRNAs (Supplemental
136 Table ST2). The maternal-type snoRNAs turned out to be variants of just three snoRNAs: U3,
137 U8 and snoZ30, each of which also has one somatic-type variant (Figure 2B). Remarkably,
138 for the U3 and U8 snoRNAs, the variants in tandem repeats showed a maternal-preferential
139 expression profile, whereas the solitary variants were somatic-preferential (Figure 2B). For
140 both U3 and U8 in tandem repeats, there were some variants that showed no expression
141 (Supplemental Figure SF2). For the snoZ30 snoRNAs, the maternal- and somatic-type
142 variants are also organized in a different way on the genome with the maternal-type being a
143 solitary snoRNA, whereas the somatic-type is located intronically (Figure 2B). Thus, there is
144 a clear association between the genome organization of the snoRNA variants and their
145 expression during development.

146 There are ample sequence differences between the maternal and somatic snoRNA types
147 (Figure 3A). We used these sequence differences to confirm the developmental-stage
148 preferential expression of maternal- and somatic-type snoRNAs, by qRT-PCR-analysis for
149 several selected maternal- and somatic-type U3, U8 and snoZ30 snoRNAs in egg and adult
150 tissue. The qRT-PCR results are in in line with the small-RNA-seq results (Figure 3C).

151 To further characterize the shift from maternal-type snoRNA expression to somatic-type
152 expression, we determined the expression of the selected maternal- and somatic snoRNAs
153 during twelve stages of zebrafish embryogenesis. Similar to rRNAs, the predominant
154 maternal-type snoRNAs in eggs are, during embryogenesis, gradually replaced by somatic-
155 type snoRNAs (Figure 3B), extending the zebrafish dual ribosome system with snoRNAs.
156
157 **Figure 2. Identifying maternal and somatic types snoRNAs.**
158 A: Distribution plot of differential snoRNA expression between zebrafish egg and adult-male tissue.
159 Maternal-type snoRNA (blue), 18 snoRNA variants with a positive fold change, i.e. with a higher
160 expression in egg than in adult-male tissue; somatic-type snoRNA (red), 186 snoRNA variants with a
161 negative fold change.
162 B: Overview of maternal- and somatic-type snoRNA variants of U3, U8 and snoZ30 genes.
163 [1] Names of the associated variants within each of the indicated snoRNA locus.
164 [2] Genomic organization: intronic, solitary or in clusters of tandemly repeated (tand rep) genes.
165 Chromosome number (chrom) plus the number (n) of variant genes in a snoRNA locus.
166 [3] Normalized NGS-read counts of the snoRNA variants (cf. Supplemental Table ST2) in the same
167 cluster of tandemly repeated genes were added up.
168
169 **Figure 3. Expression of maternal- and somatic-type snoRNAs in zebrafish development.**
170 **A**: Sequence alignment of relevant maternal-type (M) and somatic-type (S) snoRNA variants;
171 identical nucleotides are indicated as dots, while gaps as dashes (For sequence alignment of all U
172 snoRNA cf. Supplemental Figure SF3). The qRT-PCR primers (panel C, Supplemental Table ST4) are
173 indicated with half arrows specific for the maternal-type (blue) or specific somatic-type (red). The
174 grey boxed sequences in the U3 and U8 snoRNAs indicate the absent regions for the observed U
175 snoRNA-DII transcripts (Supplemental Figures SF2 and SF3).
176 **B**: Relative expression of maternal-type (blue) and somatic-type (red) snoRNA variants indicated by
177 comparative percentage of U3-DII, U8-DII, and snoZ30 NGS-reads, respectively. Prot-mouth:
178 protruding-mouth; Adult FT: adult female-tail; Adult MWB: adult-male whole-body.
179 **C**: Relative expression of maternal-type (blue) and somatic-type (red) snoRNA variants indicated by
180 comparative percentage qRT-PCR of U3-DII, U8-DII, and snoZ30 NGS, respectively for qRT-PCR
181 analyses on zebrafish eggs and adult male-tail (Adult MT) tissue using RT-PCR primers as indicated in
182 panel A (Supplemental Table ST3).
183 **D**: Relative expression of maternal-type (blue) and somatic-type (red) U3 and U8 snoRNA variants
184 indicated by comparative percentages of RT-PCR-qSeq on zebrafish eggs and adult male-tail (A MT)
185 tissue using RT-PCR primers and subsequent quantitative sequencing (Supplemental Table ST4).
186

187 Structural characteristics of maternal- and somatic-types snoRNAs

188 A significant characteristic of the observed U3 and U8 snoRNA sequences was their length.
189 Whereas all snoZ30 NGS-reads mapped to the whole genes (98 nt) (Supplemental Table
190 ST1), almost all U3 NGS-reads (~130 nt.) represented just the 3' part (Domain II (DII), Figures
191 1A and B), including the C/D box, of the ~215 nt U3 genes (Supplemental Figure SF3). Hence,
192 these U3 NGS-reads are missing the U3 5' part, including the A and A' boxes, which guides
193 the snoRNA in interactions with target RNAs. Consequently, the length of these U3-DII
194 snoRNAs is quite similar to all other H/ACA box snoRNAs (Supplemental Figure SF1).
195 Similarly, almost all U8 NGS-reads (~86 nt) represented just the 3' part (Domain II, Figures
196 1A and 4), with C/D box, of the ~131 nt U8 genes, again missing the guiding 5' part. Whether
197 complete U transcripts are truly absent or merely go undetected is addressed in the next
198 section. The length of the U8-DII snoRNAs is quite similar to most C/D box snoRNAs (Table 1
199 and Supplemental Figures SF1 and SF3). U3-DII and U8-DII snoRNAs are reported before
200 (13,41) and are likely a result of the removal of the mono-methylated (m^7G) 5' cap by the
201 decapping complex, which is regulated by PNRC1 (42) although we cannot completely rule
202 out that they also could originate from an alternative transcriptional start site (Figure 1A).
203 The U-DII snoRNAs are not merely degradation products of the complete U3 and U8
204 snoRNAs, as can be seen by their gradual disappearance during the embryogenesis (Figure
205 2B) and by the strictly defined read coverage (Supplemental Figure SF3). Thus, beside an
206 earlier reported dominant-negative effect on ribosome biogenesis (10) , we speculate that
207 these partial U-DII snoRNAs may exert other specific functions. Since their guiding parts are
208 missing (1,43), such function would be different from that of the traditional U snoRNA.
209

210 As a starting point for finding biological relevance of the observed maternal- and somatic-
211 expression profiles for different snoRNAs, we compared the sequences of the maternal-type
212 to the somatic-type variants of the U3, U8, and snoZ30 snoRNAs. There are quite some
213 sequence differences (8%, 16%, and 36%, respectively), which may be indicative for
214 functional diversity (Figure 3A and Supplemental Figure SF4). One obvious difference could
215 relate to the usual target RNAs for each snoRNA.
216 Several potential snoRNA-rRNA interactions could be pinpointed by complementary
217 sequences, in line with those reported in literature (1). However, many are in the 5' guiding
218 part of the involved snoRNAs, which is absent in our reads (Supplemental Figure SF5).
219 Conversely, many (known) interactions are found in the ETS regions of the rRNA genes
220 (Supplemental Figure SF5). Yet we observed that the involved maternal-type U-DII snoRNAs
221 are present throughout embryogenesis, much like the maternal-type rRNAs. This, combined
222 with the assumption that maternal-type rRNAs are only transcribed and processed during
223 oogenesis and somatic-type rRNAs during embryogenesis and adulthood, it seems unlikely
224 that maternal-type snoRNAs have a role in the processing of maternal-type rRNAs. Hence,
225 the observed U-DII snoRNA likely have other functions, much like snoZ30 that binds to
226 snRNA U6 (44). Given these constraints we were unable to identify any promising
227 interactions between the many possible interactions of U-DII snoRNAs and rRNAs
228 (Supplemental Figure SF5). It should be noted that such interactions are hard to find given
229 the sometimes seemingly feeble reverse complement base pairing in these situations.
230

231 ## Preferential expression of complete snoRNAs

232 The fact that virtually only U-DII snoRNA sequences were found by small-RNAseq, did not
233 match with the current knowledge of complete U3 and U8 snoRNAs, as well as a qRT-PCR

234 analysis which showed their abundant presence in the tested zebrafish samples (result not
235 shown). Given that we sequenced nearly all zebrafish tissues lead to the conclusion that,
236 with the NGS platform (Ion Proton) employed in this study, we are not able to sequence
237 complete U3 and U8 snoRNAs. This is also the case for many 5S rRNAs plus almost all tRNAs
238 and is likely caused by the cumulative effect of 5'/3' modifications, strong secondary
239 structures, and modified nucleotides.
240 To still determine whether the complete U snoRNAs also display preferential maternal or
241 somatic expression, we developed a PCR-based strategy that focused on the missing 5' parts
242 of the U snoRNA genes. To overcome the high similarity between the U3 variants, as well as
243 between the U8 variants, this approach starts with a RT-PCR, using as much as possible
244 generic PCR primers (Supplemental Table ST4), after which sequencing of the PCR products
245 reveals the distribution of variants (Supplemental Table ST4). This RT-PCR-qSeq approach
246 revealed that the complete U3 and U8 snoRNAs display the same expression patterns as the
247 shorter U3DII and U8DII (Figure 3D and Supplemental Table ST4).
248

249 ## Possible involvement of ITS sequences in U3-DII processing

250 While investigating the maternal-type U3 snoRNA tandem repeats, we observed additional
251 small RNAs that originate from the ITS regions between the snoRNA genes in the U3
252 genomic regions. These small RNAs either have their own promoter, or the U snoRNA genes
253 in tandem repeats generate one transcript which is later on processed into individual
254 mature U snoRNAs. As these small RNAs are derived from the U-ITS region, we named them
255 snoRNA-ITS-small-RNAs (sitsRNA). For instance, for the U3 loci on chromosome 14, we
256 detected essentially five different sitsRNAs (25 nt to 30 nt) that came from a highly
257 conserved region of 265 nt, just 379 nt upstream of each U3 variant sequence. We noticed

258 that there are several sitsRNAs from the various U snoRNA in tandem repeats that show
259 complementarity with U snoRNA sequences, in particular, one 26 nt sitsRNAs from the U3
260 loci on chromosome 14 showed two regions (each 7 nt long) that are reverse complement
261 to U3 snoRNA a sequence in the 5' part and one in the D-II region (Figure 1B). This raises the
262 intriguing possibility that this particular sitsRNAs may somehow be involved in the
263 processing from full length U3 snoRNA to U3-DII snoRNA as the locations of interaction
264 between the sitsRNAs and the U3 snoRNA span the cutting site (Figure 1B). Although similar
265 sites were found in the other U3 and U8 clusters, we were unable to find another sitsRNAs
266 that would interact like this. Still we feel there are enough indications that warrant further
267 investigation of the possible role of U-ITS sequences in snoRNA processing.

268 CONCLUSION

269 In this study, similar to our previously report on developmental-specific expression of
270 rRNAs, we observed a specific subset of snoRNAs; U3, U8 and snoZ30 of which variants
271 show distinct expression profiles during early zebrafish embryogenesis. All other snoRNAs
272 are about eight times higher preferentially expressed in non-embryonic developmental
273 stages, which may be a logical consequence of the fact that the rRNAs are already processed
274 in an egg, thus only requiring snoRNAs for other tasks than rRNA maturation during early
275 embryogenesis.
276 We discovered next to the complete U snoRNAs, also U3-DII and U8-DII partial snoRNAs,
277 which miss their 5' rRNA recognizing part and essentially consist of just the Domain-II
278 hairpin structures. These U-DII partial snoRNAs variants also showed maternal- or somatic
279 preferential expression, which correlated nicely with their genomic organization in tandem
280 repeats and solitary, respectively. The complete versus partial U snoRNAs show the same
281 preferential expression.

282 Though, at least part of the detected differential gene expression is likely caused by
283 associated promotors of the involved snoRNA genes, we were unable to find any sequences
284 in the 200 bp upstream promoter region that could discriminate the maternal- from the
285 somatic-type snoRNA genes (results not shown).
286 While the function of the U-DII snoRNAs is still unclear, given that the DII part of complete U
287 snoRNAs is known to bind to several proteins (45,46), the intact hairpin in a U-DII snoRNA
288 hints at a role as ribonucleoprotein. In any case, since they seem to be strictly regulated, U-
289 DII snoRNAs probably have a significant role in the zebrafish embryogenesis.
290 How they come about is another fascinating puzzle. We observed small RNAs, which
291 originate from the ITS regions of U3 snoRNA loci, that show convincing complementarity
292 with U3 sequences. This raises the possibility that they somehow may be involved in the
293 biosynthesis of the U3-DII snoRNAs. In any case, there is a notable analogy between the
294 relatively-small RNA-processing snoRNAs that are located in introns of genes usually
295 involved in ribosome biogenesis and small RNAs, which are located in the ITS of snoRNAs
296 they possibly support processing.
297

298 MATERIAL AND METHODS

299 Biological materials, RNA-isolation and small-RNA-seq

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

305 For this study we used samples of two pools of unfertilized eggs (oocyte clutches) and two
306 male-adult zebrafish tails. The harvesting of the biological materials and RNA-isolation have
307 been described previously in (34) and (35).

308 ### Source data

309 Next-generation data previously generated in our group (34) and available through the
310 BioProject database with accession number PRJNA347637 has been used in this study with
311 respect to i) Three pools of unfertilized eggs (oocytes); ii) one embryo at each of the 12
312 developmental stages: 64 cells (2 hours post-fertilization (hpf)); high stage (3.3 hpf); 30%
313 epiboly stage (4.7 hpf); 70% epiboly stage (7 hpf); 90% epiboly stage (9 hpf); 4-somite stage
314 (11.3 hpf); 12-somite stage (15 hpf); 22-somite stage (20 hpf); prim-5 stage (24 hpf); prim-16
315 (31 hpf); long-pec stage (48 hpf); protruding-mouth stage (72 hpf); and iii) one whole-body
316 male-adult zebrafish sample.

317 ### qRT-PCR analysis

318 Forward and reverse PCR primers were designed for the maternal-type snoRNA genes U3_k,
319 U8_d and snoZ30_b, and for the somatic-type snoRNA genes U3_a, U8_e and snoZ30_a
320 (Supplemental Table ST3). Reverse transcription was done in two independent reactions
321 primed with the combined reverse PCR primers of either the three maternal-type variants
322 or the three somatic-type variants. Both reactions were performed on zebrafish egg pool
323 and whole-body adult male total RNA, in a total of four reactions. SuperScript IV Reverse
324 Transcriptase (Thermo Fisher Scientific) was used according to the manufacturer's
325 instructions. Quantitative real-time PCR (qRT-PCR) was performed on 10-fold dilutions of
326 the cDNA using a QuantStudio 3 Real-Time PCR System (Thermo Fisher Scientific).

327 ## RT-PCR-qSeq analysis

328 Forward and reverse PCR primers were designed for all U3 and U8 variants, in such a way
329 that: 1) as much as possible of the 5'-end of the full-length variants is included in the final
330 amplicon, and 2) generic primers were selected that bind to the maternal-type, as well as
331 the somatic-type variants (Supplemental Table ST4). cDNA was prepared as described above
332 and used in regular PCR reactions for each of the variants independently. Amplification was
333 performed using the Q5 High-Fidelity DNA Polymerase (New England Biolabs). The resulting
334 amplicons were purified using the QIAquick PCR Purification Kit (Qiagen) and their size was
335 verified on a 2200 TapeStation System (Agilent). Barcoded sequencing libraries were
336 prepared using a modified version of the Ion Xpress Plus Fragment Library Kit (Thermo
337 Fisher Scientific). Massive-parallel sequencing was performed on an Ion Proton System
338 (Thermo Fisher Scientific) using an Ion PI Chip Kit v3.

339 ## Bioinformatics analyses

340 *Known snoRNA sequences.* Known snoRNA sequences of *D. rerio* were downloaded from
341 Ensemble 91 (39) and from snOPY (38) in May 2017. A union was made of these two set of
342 sequences and sequence annotations. Characters were added to the snoRNA names to
343 uniquely distinguish the multiple variants of the same snoRNA and a FASTA file containing
344 only the unique snoRNA sequences was created.

345 *Zebrafish snoRNA similarity tree.* A hierarchical clustering was used to compare the snoRNA
346 sequences. The dendrogram was made by using the *hclust* and *stringDist* functions from the
347 R version 3.2.1 package 'stats' and 'Biostrings' respectively (47)

348 *Mapping NGS-reads.* At both 5' and 3' end of each snoRNA sequence, 5 Ns were added to
349 facilitate the alignments in the NGS-read mapping. NGS-reads longer than 20 nt from all
350 experiments, were mapped against the unique snoRNA sequences (Supplemental File

351 SFile1) using Bowtie2 (48) with the following settings: *-np* to 0, - - *score-min* to L, -1, -0.3 for
352 zebrafish in order to limit the maximal amount of mismatches to 5%. SAMtools v1.2 (49)
353 was used to convert the alignment to the BAM file format and to retrieve the mapped NGS-
354 read counts per snoRNA sequence. NGS-reads that were smaller than 50% of the length of a
355 snoRNA sequence were discarded.
356 *Analysis of NGS-read mapping results.* The NGS-read counts were scaled using total raw 5.8S
357 RNA NGS-read counts for each sample (35). The average of the three egg and the three
358 adult-male technical replicates was taken. A cutoff of at least 100 NGS-read count for egg
359 and adult-male combined was used to determine whether a snoRNA sequence was present
360 and to available for the analysis of differential gene expression, which was calculated in a
361 log2 scale of the egg over adult-male NGS-read counts.
362 *Analysis of the RT-PCR-qSeq reads.* A subsequence defined by the primers was selected for
363 U3, U8 snoRNA (*Supplemental Table 4*). These subsequences were then used to search
364 exact matching reads in the FASTQ files generated by the RT-PCR-qSeq experiments. The
365 number of exact matches is reported as read count for each variant of U3 and U8 for all
366 samples. The read counts belonging to the same tissues (clutch and adult male) for each
367 variant were added up. The percentage of each variant was calculated based on total tissue
368 reads.
369 *snoRNA-rRNA interactions.* BLASTn (50) was used to detect possible U3 and U8 snoRNA-
370 rRNA base-pairing interactions starting from known interactions in human. The word-size
371 parameter of BLASTn was set to 7 and of the resulting alignments the forward-reverse were
372 selected with at least 10 matching base-pairs. For snoRNA-ITS interaction analysis, this was
373 set to 7 matching base pairs.
374

ACKNOWLEDGMENTS

We are grateful for the valuable input by (anonymous) reviewers, which allowed us to improve our manuscript. We acknowledge the support of The Netherlands Organization for Scientific Research (NWO) grant number 834.12.003

REFERENCES

1. Watkins NJ, Bohnsack MT. The box C/D and H/ACA snoRNPs: Key players in the modification, processing and the dynamic folding of ribosomal RNA. Wiley Interdiscip Rev RNA. 2012;3(3):397–414.
2. Kiss T. Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. EMBO J. 2001;20(14):3617–22.
3. Jorjani H, Kehr S, Jedlinski DJ, Gumienny R, Hertel J, Stadler PF, et al. An updated human snoRNAome. Nucleic Acids Res. 2016;44(11):5068–82.
4. Falaleeva M, Pages A, Matuszek Z, Hidmi S, Agranat-Tamir L, Korotkov K, et al. Dual function of C/D box small nucleolar RNAs in rRNA modification and alternative pre-mRNA splicing. Proc Natl Acad Sci [Internet]. 2016;113(12):E1625–34. Available from: http://www.pnas.org/lookup/doi/10.1073/pnas.1519292113
5. Scott MS, Ono M. From snoRNA to miRNA: Dual function regulatory non-coding RNAs. Biochimie [Internet]. 2011;93(11):1987–92. Available from: http://dx.doi.org/10.1016/j.biochi.2011.05.026
6. Brameier M, Herwig A, Reinhardt R, Walter L, Gruber J. Human box C/D snoRNAs with miRNA like functions: Expanding the range of regulatory RNAs. Nucleic Acids Res. 2011;39(2):675–86.
7. Ellis JC, Brown DD, Brown JW. The small nucleolar ribonucleoprotein (snoRNP) database. Rna. 2010;16(4):664–6.
8. Sloan KE, Warda AS, Sharma S, Entian KD, Lafontaine DLJ, Bohnsack MT. Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. RNA Biol [Internet]. 2017;14(9):1138–52. Available from: https://doi.org/10.1080/15476286.2016.1259781
9. Matera AG, Terns RM, Terns MP. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nat Rev Mol Cell Biol [Internet]. 2007 Mar [cited 2014 Jul 11];8(3):209–20. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17318225
10. Samarsky DA, Fournier MJ, Singer RH, Bertrand E. The snoRNA box C/D motif directs nucleolar targeting and also couples snoRNA synthesis and localization. EMBO J. 1998;17(13):3747–57.
11. Watkins NJ. Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. Rna. 1998;4(12):1549–68.
12. Watkins NJ, Ségault V, Charpentier B, Nottrott S, Fabrizio P, Bachi A, et al. A common

core RNP structure shared between the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP. Cell. 2000;103(3):457–66.
13. Newman DR, Kuhn JF, Shanab GM, Maxwell ES. Box C/D snoRNA-associated proteins: Two pairs of evolutionarily ancient proteins and possible links to replication and transcription. Rna. 2000;6(6):861–79.
14. Decatur WA, Fournier MJ. RNA-guided Nucleotide Modification of Ribosomal and Other RNAs. J Biol Chem. 2003;278(2):695–8.
15. Filipowicz W, Pogačić V. Biogenesis of small nucleolar ribonucleoproteins. Curr Opin Cell Biol. 2002;14(3):319–27.
16. Van Nues RW, Granneman S, Kudla G, Sloan KE, Chicken M, Tollervey D, et al. Box C/D snoRNP catalysed methylation is aided by additional pre-rRNA base-pairing. EMBO J [Internet]. 2011;30(12):2420–30. Available from: http://dx.doi.org/10.1038/emboj.2011.148
17. Nicoloso M, Qu LH, Michot B, Bachellerie JP. Intron-encoded, antisense small nucleolar RNAs: The characterization of nine novel species points to their direct role as guides for the 2'-O-ribose methylation of rRNAs. J Mol Biol. 1996;260(2):178–95.
18. Kiss-László Z, Bachellerie J-P, Caizergues-Ferrer M, Kiss T, Henry Y. Site-Specific Ribose Methylation of Preribosomal RNA: A Novel Function for Small Nucleolar RNAs. Cell. 1996;85(7):1077–88.
19. Cavaille J, Nicoloso M, Bachellerie J. Targeted ribose methylation of RNA in vivo directed by tailored antisense RNA guides. 1996;383(October):732–5.
20. Chen CL, Perasso R, Qu LH, Amar L. Exploration of Pairing Constraints Identifies a 9 Base-pair Core within Box C/D snoRNA-rRNA Duplexes. J Mol Biol. 2007;369(3):771–83.
21. Ganot P, Caizergues-Ferrer M, Kiss T. The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. Genes Dev. 1997;
22. Ganot P, Bortolin ML, Kiss T. Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. Cell. 1997;89(5):799–809.
23. Kehr S, Bartschat S, Tafer H, Stadler PF, Hertel J. Matching of soulmates: Coevolution of snoRNAs and their targets. Mol Biol Evol. 2014;31(2):455–67.
24. Jády BE, Kiss T. A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. EMBO J. 2001;20(3):541–51.
25. Marz M, Gruber AR, Höner Zu Siederdissen C, Amman F, Badelt S, Bartschat S, et al. Animal snoRNAs and scaRNAs with exceptional structures. RNA Biol. 2011;8(6).
26. Peculis BA, Steitz JA. Disruption of U8 nucleolar snRNA inhibits 5.8S and 28S rRNA processing in the Xenopus oocyte. Cell. 1993;73(6):1233–45.
27. Hughes JMX. Functional Base-pairing Interaction Between Highly Conserved Elements of U3 Small Nucleolar RNA and the Small Ribosomal Subunit RNA. J Mol Biol [Internet]. 1996;259(4):645–54. Available from: https://doi.org/10.1006/jmbi.1996.0346
28. Peculis BA, DeGregorio S, McDowell K. The U8 snoRNA gene family: Identification and characterization of distinct, functional U8 genes in Xenopus. Gene. 2001;274(1–2):83–92.
29. Lui L, Lowe T. Small nucleolar RNAs and RNA-guided post-transcriptional modification. Essey Biochem. 2013;(54):53–77.
30. Dieci G, Preti M, Montanini B. Eukaryotic snoRNAs: A paradigm for gene expression

flexibility. Genomics [Internet]. 2009;94(2):83–8. Available from: http://dx.doi.org/10.1016/j.ygeno.2009.05.002

31. Bratkovič T, Rogelj B. Biology and applications of small nucleolar RNAs. Vol. 68, Cellular and Molecular Life Sciences. 2011. p. 3843–51.

32. Bachellerie JP, Cavaillé J, Hüttenhofer A. The expanding snoRNA world. Vol. 84, Biochimie. 2002. p. 775–90.

33. Bratkovič T, Rogelj B. The many faces of small nucleolar RNAs. Biochim Biophys Acta - Gene Regul Mech. 2014;1839(6):438–43.

34. 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

35. 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

36. Pagano JFB, Dekker RJ, Ensink WA, van Olst M, Bos A, van Leeuwen S, et al. An alternative spliceosome defined by distinct snRNAs in early zebrafish embryogenesis. Not Publ.

37. Dupuis-Sandoval F, Poirier M, Scott MS. The emerging landscape of small nucleolar RNAs in cell biology. Wiley Interdiscip Rev RNA. 2015;6(4):381–97.

38. Yoshihama M, Nakao A, Kenmochi N. snOPY: a small nucleolar RNA orthological gene database. BMC Res Notes. 2013;6(1):426.

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

40. Olson MOJ. The Nucleolus. 1st ed. Olson MOJ, editor. Springer US; 2004. 364 p.

41. Watkins NJ, Leverette RD, Xia L, Andrews MT, Stuart Maxwell E. Elements essential for processing intronic U14 snoRNA are located at the termini of the mature snoRNA sequence and include conserved nucleotide boxes C and D. RNA. 1996;

42. Gaviraghi M, Vivori C, Pareja Sanchez Y, Invernizzi F, Cattaneo A, Santoliquido BM, et al. Tumor suppressor PNRC 1 blocks r RNA maturation by recruiting the decapping complex to the nucleolus . EMBO J. 2018;

43. Dutca LM, Gallagher JEG, Baserga SJ. The initial U3 snoRNA:pre-rRNA base pairing interaction required for pre-18S rRNA folding revealed by in vivo chemical probing. Nucleic Acids Res. 2011;39(12):5164–80.

44. Qu LH, Zhou H, Chen YQ, Du YP. The Schizosaccharomyces pombe mgU6-47 gene is required for 2 ’-O-methylation of U6 snRNA at A41. Nucleic Acids Res. 2002;30(4):894–902.

45. Watkins NJ, Lemm I, Ingelfinger D, Schneider C, Hoßbach M, Urlaub H, et al. Assembly and maturation of the U3 snoRNP in the nucleoplasm in a large dynamic multiprotein complex. Mol Cell. 2004;16(5):789–98.

46. Granneman S, Kudla G, Petfalski E, Tollervey D. Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. Proc Natl Acad Sci [Internet]. 2009;106(24):9613–8. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.0901997106

47. R Development Core Team. R: A Language and Environment for Statistical Computing. R Found Stat Comput Vienna Austria [Internet]. 2016;0:{ISBN} 3-900051-07-0.

509 Available from: http://www.r-project.org/
510 48. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods.
511 2012 Apr;9(4):357–9.
512 49. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
513 Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug;25(16):2078–9.
514 50. Madden T. Chapter 16 : The BLAST Sequence Analysis Tool. NCBI Handbook[internet].
515 2002;1–15.
516
517

518 SUPPLEMENTAL FIGURES

519 SF1.pdf: Zebrafish snoRNA characteristics.

520 SF2.pdf: Tandem repeat DII snoRNA expression distribution.

521 SF3.pdf: Maternal- and somatic-type snoRNA read coverage.

522 SF4.pdf: Sequence alignments of snoRNA variants.

523 SF5.pdf: Possible snoRNA-rRNA interactions

524

525 SUPPLEMENTAL TABLES AND FILES

526 SFile1.fa : Zebrafish snoRNA sequences

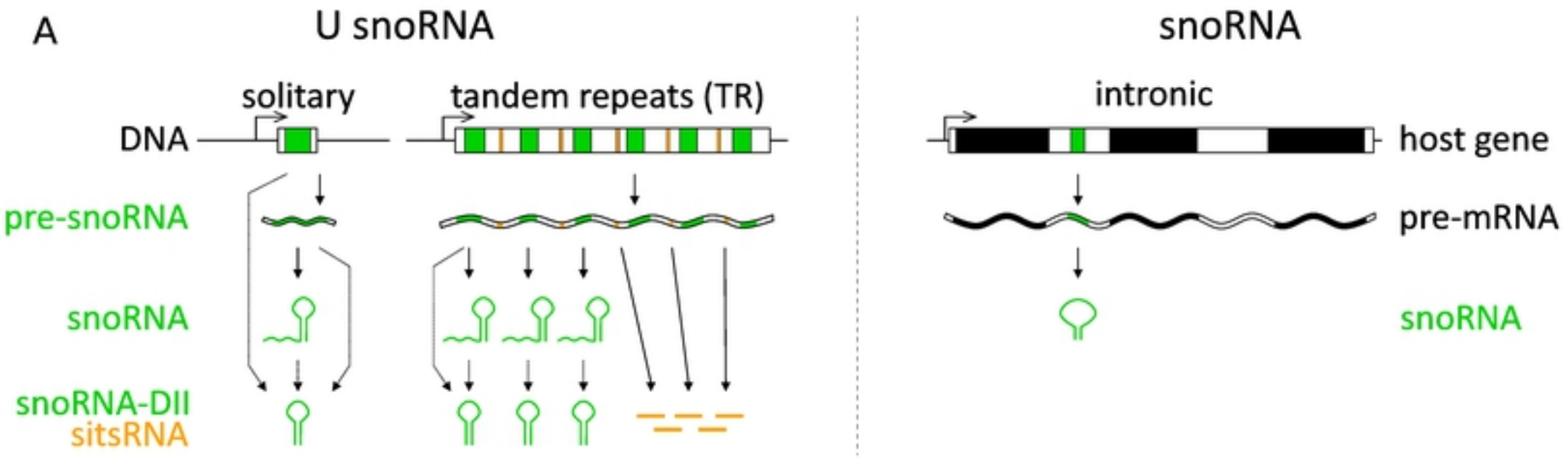
527 ST1.xlsx: Genome annotation of zebrafish snoRNAs

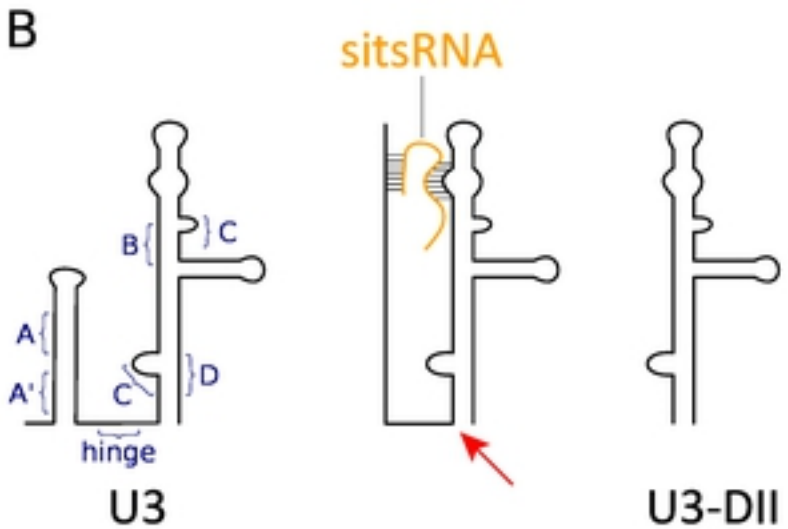
528 ST2.xlsx: Read counts from rna-seq experiment mapping to zebrafish snoRNAs

529 ST3.xlsx: qRT-PCR primers

530 ST4.xlsx: RT-PCR-qSeq experiments set-up and results.

531





C

Characteristics of zebrafish snoRNA genes

snoRNA		Avg. length[1]	Genomic organization			Total
Box	Subset		Intronic	Solitary	TR	
H/ACA	all	141	89	3	0	**92**
C/D	rest	91	119	3	0	**122[2]**
C/D	U3	215	0	3	35	**38[3]**
C/D	U8	131	0	1	4	**5**
			81%	***4%***	***15%***	***257***

Figure1

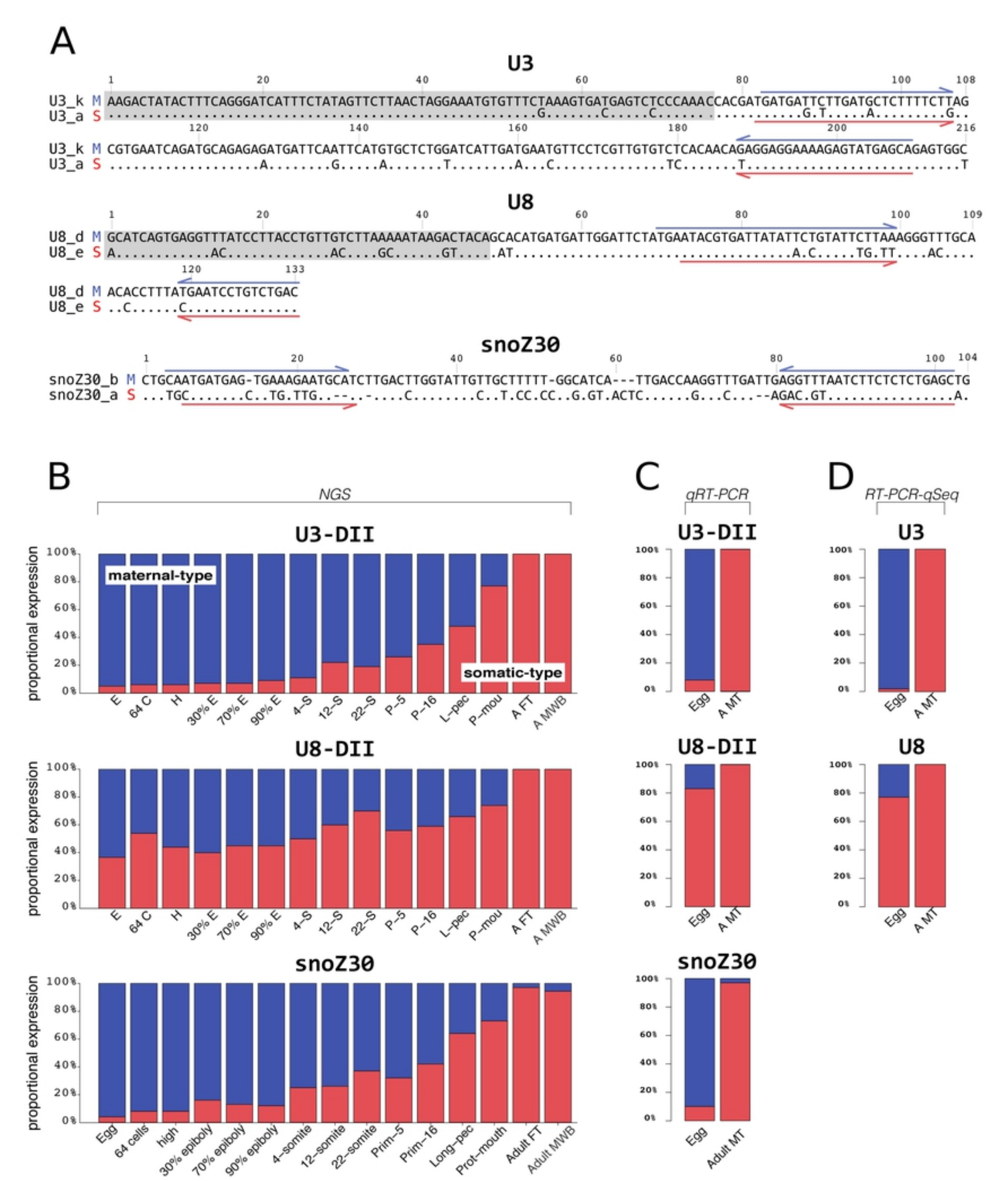
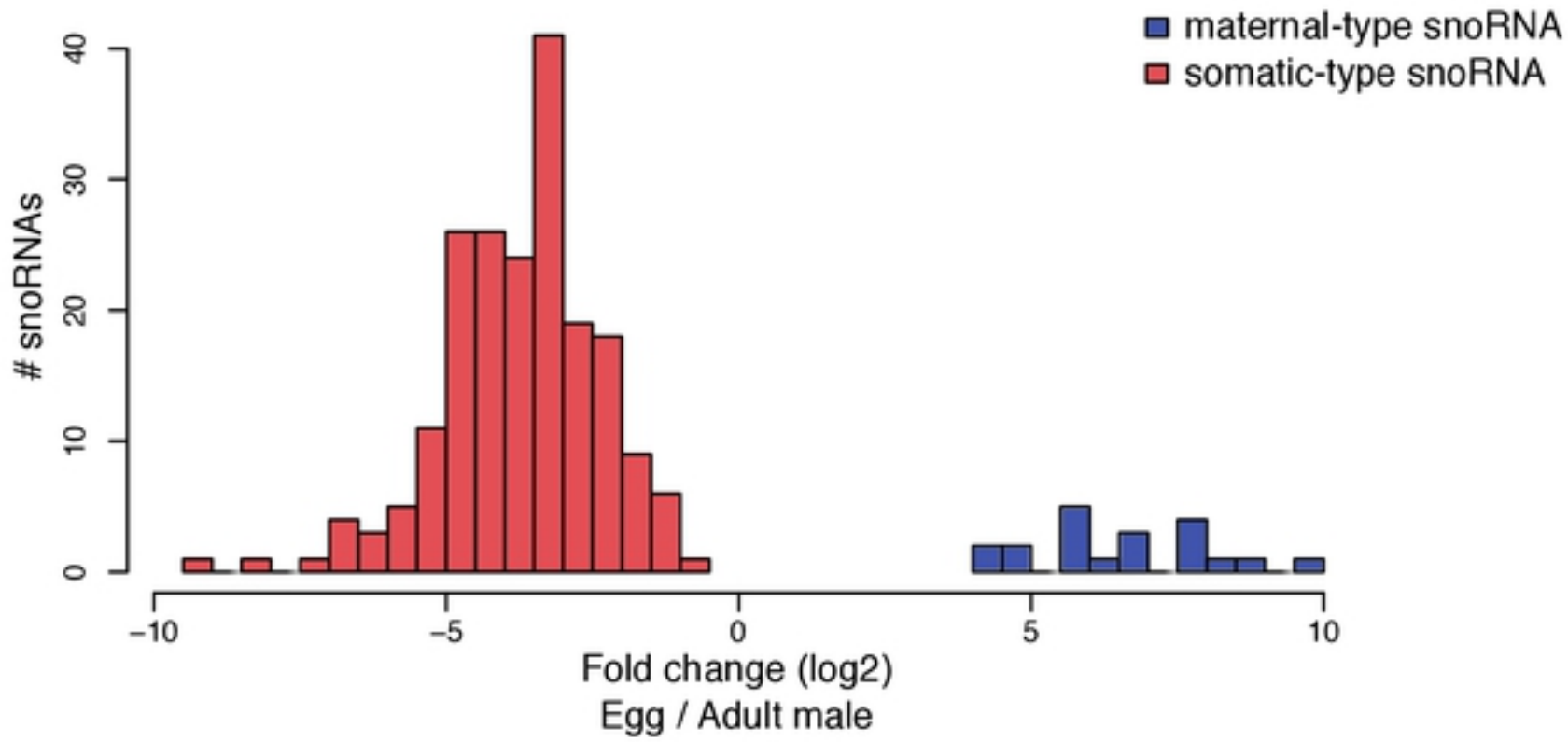

A
U3
U3_k M AAGACTATACTTTCAGGGATCATTTCTATAGTTCTTAACTAGGAAATGTGTTTCTAAAGTGATGAGTCTCCCAAACCACGATGATGATTCTTGATGCTCTTTTCTTAG
U3_a S .....G.....C.....C.....G.T.....A.....G..
U3_k M CGTGAATCAGATGCAGAGAGATGATTCAATTCATGTGCTCTGGATCATTGATGAATGTTCCTCGTTGTGTCTCACAACAGAGGAGGAAAAGAGTATGAGCAGAGTGGC
U3_a S .....A.....G.....A.....T.....A...C.....TC.....T.....T
U8
U8_d M GCATCAGTGAGGTTTATCCTTACCTGTTGTCTTAAAAATAAGACTACAGCACATGATGATTGGATTCTATGAATACGTGATTATATTCTGTATTCTTAAAGGGTTTGCA
U8_e S A.....AC.....AC....GC.....GT....AT.....A.C.....TG.TT....AC....
U8_d M ACACCTTTATGAATCCTGTCTGAC
U8_e S ..C.....C.....
snoZ30
snoZ30_b M CTGCAATGATGAG-TGAAAGAATGCATCTTGACTTGGTATTGTTGCTTTTT-GGCATCA---TTGACCAAGGTTTGATTGAGGTTTAATCTTCTCTCTGAGCTG
snoZ30_a S ...TGC.....C..TG.TTG..--..-....C.....C..T.CC.CC..G.GT.ACTC.....G...C...--AGAC.GT.....A.
B
NGS
U3-DII
maternal-type
somatic-type
proportional expression
E
64 C
H
30% E
70% E
90% E
4-S
12-S
22-S
P-5
P-16
L-pec
P-mou
A FT
A MWB
U8-DII
snoZ30
Egg
64 cells
high
30% epiboly
70% epiboly
90% epiboly
4-somite
12-somite
22-somite
Prim-5
Prim-16
Long-pec
Prot-mouth
Adult FT
Adult MWB
C
qRT-PCR
U3-DII
U8-DII
snoZ30
Egg
A MT
Adult MT
D
RT-PCR-qSeq
U3
U8

Figure3

A



B

Zebrafish snoRNA genes of which specific variants display maternal-type expression during development

snoRNA		Genomic organization (chrom, n)[2]	Normalized read counts[3]		Fold change range Egg / Adult male (log2)	Expression type
Name	Variant[1]		Egg	Adult male		
U3	a	solitary (c1, 1)	1,211	8,370	-2.8	somatic
	b-h	tand rep (c1, 7)	1,452	4	5.4 to 9.0	maternal
	k-t	tand rep (c14, 15)	22,297	7	8.1 to 12.0	maternal
	u-ag	tand rep (c20, 13)	1,318	3	7.0 to 8.8	maternal
U8	a-d	tand rep (c10, 4)	3,947	6	8.5 to 9.5	maternal
	e	solitary (c10, 1)	2,223	24,519	-3.5	somatic
snoZ30	a	intronic (3, 1)	450	8,382	-4.2	somatic
	b	solitary (c5, 1)	10,152	522	4.3	maternal
other		solitary intronic	*variable*	*variable*	-9.1 to -0.4	somatic

Figure2