Matching of Soulmates: Coevolution of snoRNAs and Their Targets

Stephanie Kehr,*1 Sebastian Bartschat,1 Hakim Tafer,1 Peter F. Stadler,1,2,3,4,5,6 and Jana Hertel1

1Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, University of Leipzig, Leipzig, Germany
2Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany
3RNomics Group, Fraunhofer Institute for Cell Therapy and Immunology, Leipzig, Germany
4Institute for Theoretical Chemistry, University of Vienna, Vienna, Austria
5Center for Non-Coding RNA in Technology and Health, University of Copenhagen, Frederiksberg, Denmark
6The Santa Fe Institute

*Corresponding author: E-mail: steffi@bioinf.uni-leipzig.de.

Associate editor: Ryan Hernandez

Abstract

Ribosomal and small nuclear RNAs (snRNAs) comprise numerous modified nucleotides. The modification patterns are retained during evolution, making it even possible to project them from yeast onto human. The stringent conservation of modification sites and the slow evolution of rRNAs and snRNAs contradicts the rapid evolution of small nucleolar RNA (snoRNA) sequences. To explain this discrepancy, we investigated the coevolution of snoRNAs and their targeted sites throughout vertebrates. To measure and evaluate the conservation of RNA-RNA interactions, we defined the interaction conservation index (ICI). It combines the quality of individual interaction with the scope of its conservation in a set of species and serves as an efficient measure to evaluate the conservation of the interaction of snoRNA and target. We show that functions of homologous snoRNAs are evolutionarily stable, thus, members of the same snoRNA family guide equivalent modifications. The conservation of snoRNA sequences is high at target binding regions while the remaining sequence varies significantly. In addition to elucidating principles of correlated evolution, we were able, with the help of the ICI measure, to assign functions to previously orphan snoRNAs and to associate snoRNAs as partners to known chemical modifications unassigned to a given snoRNA. Furthermore, we used predictions of snoRNA functions in conjunction with sequence conservation to identify distant homologies. Because of the high overall entropy of snoRNA sequences, such relationships are hard to detect by means of sequence homology search methods alone.

Key words: target prediction, RNA-RNA interaction, ICI.

Introduction

Small nucleolar RNAs (snoRNAs) compose a special class of noncoding RNA genes responsible for guiding protein complexes to specific positions in other RNA molecules. Those snoRNPs introduce chemical modifications of single residues in the target RNA. Precise targeting is achieved by means of sequence complementarity, while the protein components stabilize the snoRNPs and catalyze the chemical reactions (Bachellerie et al. 2002).

There are two main types of snoRNAs: box H/ACA and box C/D snoRNAs. While the first guide pseudouridylation, the latter direct 2’-O-ribose methylation of any kind of nucleotide. Both modifications are introduced concurrently or immediately after transcription of the RNA operon, before cleavage of the 45S rRNA (primary transcription product of the rRNA operons). These modifications are essential for maturation of rRNAs. The special snoRNAs U3, U8, U14, U17, and U22 direct cleavage steps of the 45S rRNA rather than chemical modifications (Atzorn et al. 2004).

Small Cajal body-specific RNAs (scRNAs) constitute a third type of snoRNA. These snoRNAs guide the modification of small nuclear RNAs (snRNAs) in the Cajal body of eukaryotic cells (Darzacq et al. 2002). Some scRNAs combine features of both snoRNA classes, see Marz et al. (2011) for more details on these atypical RNAs.

Beyond these regular functions, there exist snoRNAs with different tasks. The brain specific box C/D snoRNA SNORD115 (HBII-52), for example, has been reported to affect editing and/or alternative splicing of mRNAs in human and mouse (Kishore and Stamm 2006; Doe et al. 2009; Kishore et al. 2010; Soneo et al. 2010). A box H/ACA snoRNA domain has been detected at the 3’-end of vertebrate telomerase RNAs (Mitchell et al. 1999). Taft et al. (2009) and Ender et al. (2008) proved that some snoRNAs may be processed into smaller RNAs, so-called sdRNAs, that act in a microRNA-like fashion in posttranscriptional gene silencing. In fact, most box H/ACA snoRNA snoRNAs, but not box C/D snoRNA snoRNA, are substrates for Dicer (Langenberger et al. 2012). Despite this diversity of snoRNA function, we limit our attention here to the common modification tasks, pseudouridylation, and 2’-O-ribose methylation.
The positions of chemical modifications in rRNAs and snRNAs are evolutionary highly conserved. For example, all modified residues of the large subunit of rRNA (LSU) in human can be aligned to the respectively modified residues of LSU in mouse (Ofengand and Bakin 1997). Many of these sites are even conserved in yeast (Saccharomyces cerevisiae) (Maden 1986, 1996). A correspondence of the (predicted) target sites was used successfully to identify box C/D snoRNAs in the fruitfly genome (Accardo et al. 2004). Most known modified sites in rRNAs and snRNAs (101 of 113 pseudouridines and 112 of 130 methylations) correspond to at least one known snoRNA (Lestrade and Weber 2006).

snoRNAs form characteristic secondary structures and encode antisense elements (ASEs) at specific positions. These are responsible for the recognition of the target (see fig. 1). In case of box C/D snoRNAs, a target sequence can directly be inferred from the ASE, adjacent to the 5′-end of the box D sequence element(s). The snoRNA/target duplex has a length of 7–20 nt, no bulges, and may only be interrupted by a few mismatches (Chen et al. 2007). The nucleotide in the target RNA that pairs with the fifth base of the ASE is methylated. For box H/ACA snoRNAs, the situation is more complex. Here the ASE, which has a total length of 6–20 nt, is bipartite and located in an interior loop of the hairpin(s), the pseudouridylation pocket. The modified uridine is anchored, unpaired, underneath the upper stem, central to the two parts of the ASE (Canot et al. 1997). Again, only a few mismatches but no bulges are tolerated. Efficient software tools to predict targets based on these principles are PLEXY (Kehr et al. 2011) for box C/D snoRNAs and RNAsnoop (Tafer et al. 2010) for box H/ACA snoRNAs. Nevertheless, there are still about 40 orphan snoRNA families without known or predicted target sites (Bachellerie et al. 2002).

The stringent conservation of modification sites as well as the slow evolution of rRNAs and snRNAs are at odds with the rapid evolution of snoRNA sequences (Weber 2006; Luo and Li 2007; Schmitz et al. 2008; Hoeppner and Poole 2012). To better understand this apparent contradiction, we investigate here the coevolution of snoRNAs with their target sites across vertebrates in detail. Thus, we clarify...
whether snoRNAs are stable interaction partners for a certain modification or whether a changeover of the RNA guide is the common pattern in vertebrates.

New Approaches
To study conservation of interaction, we started by computing targets for each individual snoRNA sequence (see Materials and Methods), and subsequently, we evaluated their conservation in other vertebrate species. To formally investigate the conservation, we develop here the interaction conservation index (ICI).

Interaction Conservation Index
A target \( t \) specifies a particular column in the alignments of the target sequences. A snoRNA family \( s \) is defined by homology, that is, it may contain more than one paralog. We denote by \( X(t, s, k) \) the set of all snoRNAs from family \( s \) in species \( k \) that are predicted to target \( t \) in species \( k \). Furthermore, we write \( S(t, k) \) for the set of snoRNA families predicted to target \( t \) in species \( k \), that is, \( S(t, k) = \{ s | X(t, s, k) \neq \emptyset \} \). Similarly, \( O(t, s) = \{ k | X(t, s, k) \neq \emptyset \} \) denotes the set of species in which family \( s \) has a representative that targets \( t \), and \( T(s, k) = \{ t | X(t, s, k) \neq \emptyset \} \) is the set of targets of the snoRNA family \( s \) in species \( k \).

The interaction is scored at the level of families

\[
e(t, s, k) = \min_{x \in X(t, s, k)} E_{\text{mfe}}(x, y_{t,k}),
\]

where \( E_{\text{mfe}}(x, y_{t,k}) \) is the minimum free energy (mfe) of the interaction between the snoRNA \( x \) and the sequence \( y_{t,k} \) centered at the target \( t \) in species \( k \), computed as minimum free energy of the interaction given by PLEXY or RNAsnoop. Hence, if we have more than one paralog in the same family \( s \), we consider only the one with the best interaction energy.

The average predicted interaction energy for target \( t \) in species \( k \) is \( \bar{e}(t, k) = \sum_{s \in S(t, k)} e(t, s, k) / |S(t, k)| \), whereas the average interaction energy of snoRNA \( s \) with all its putative targets \( t \) in species \( k \) is \( \bar{e}(s, k) = \sum_{t \in T(s, k)} e(t, s, k) / |T(s, k)| \). Averaging over all species in which the interaction is predicted, we introduce the two normalized parameters

\[
\text{ICImod}(t, s, k) = e(t, s, k) / \bar{e}(t, k),
\]

\[
\text{ICIno}(t, s, k) = e(t, s, k) / \bar{e}(s, k),
\]

For \( k \notin O(t, s) \), these energy-based scores are not defined because no member of family \( s \) interacts with target \( t \) in species \( k \).

To summarize these data over all species, we define the ICIs

\[
\text{ICImod}(t, s) = \sum_{k \in O(t, s)} e(t, s, k) / \bar{e}(t, k),
\]

\[
\text{ICIno}(t, s) = \sum_{k \in O(t, s)} e(t, s, k) / \bar{e}(s, k),
\]

for the modification (target) and the snoRNA, respectively. Both scores measure how much better \( s \) fits the target \( t \) compared with the predicted alternatives for which an interaction is also feasible. Large values of ICIs \( t(s, s) \geq 1 \) suggest that \( t \) is consistently a target of snoRNA family \( s \). The parameter ICImod \( t(s, s) \) emphasizes the conservation of the modification site, while ICIno \( t(s, s) \) emphasizes the conservation of the snoRNAs ASE.

Results
Target Prediction
RNAsnoop returned 59 known human box H/ACA snoRNA interactions although it has been trained on yeast only. Deactivation of the yeast model and scoring predictions based on interaction energy only, yielded 86 of the 112 known interactions, including all those predicted with the yeast model. Surprisingly, the recovery of known human box C/D snoRNA targets performed better without the use of accessibility information. While considering the internal structure of the target RNA, 103 of the 115 known interactions were recovered, and neglect of accessibility information recovered 111. This is in agreement with the observation that accessibility around methylated site does, in contrast to pseudouridylated residues, not significantly differ from the average accessibility of nucleotides in the ribosomal RNAs (data not shown). Subsequently, predicted pseudouridylated and methylated positions were mapped to the corresponding columns of the target RNA alignments.

Conservation of Known Modifications
Vertebrate rRNAs have a highly conserved modification pattern and a high level of sequence conservation in the vicinity of the modified nucleotides. By contrast, snoRNAs are famous for their overall high-sequence entropy. Nevertheless, this study confirms that in general the same modifications are guided by the same snoRNA families (Hoepner and Poole 2012). In fact, together with the functional sequence boxes, the ASEs form the regions of strongest sequence conservation within the snoRNAs, disclosing coevolution of snoRNAs and their targets. Furthermore, the interactions are maintained by compensatory mutations within the snoRNA sequences preserving the base pairing.

Known Human Interactions
First, we investigated conservation of the known human interactions (data retrieved from snoRNA-LBME-db [snoRNAbase, https://www.snonma.biotoul.fr/], last accessed November 7, 2013]). Therefore, we evaluated the individual target predictions from all investigated species using the new ICIs (see Materials and Methods). We recovered 87% of the interactions as conserved within vertebrate species with known snoRNA and target RNA sequences. In 18S rRNA, all 35 reported human interactions with box C/D snoRNAs are conserved at least in Eutheria (fig. 2). For 18S rRNA and box H/ACA snoRNAs, we find 31 of the 39 reported human interactions to be evolutionarily conserved at least in mammals. Five VPs in 18S rRNA are reported to have two matching box H/ACA guides. We predicted conserved function of both guides only in one case. For the remaining doubly guided modifications, only a single conserved snoRNA guide (supplementary fig. S12) was identified.
Figure 2 shows the conservation pattern of interactions between box C/D snoRNA guides and methylated sites in the 18S rRNA. The header row lists the abbreviated species names of investigated vertebrates. The subsequent rows provide detailed information about the interactions of a modification site (first column) and certain snoRNAs (2nd last column). For each target site, ≤3 snoRNAs are shown. These are ordered by their IC$_{m_2}^0$ scores (last column). The color intensity of each field is correlated to the predicted minimum free energy of the individual interaction. A field is crossed out if the snoRNA (from lower left to upper right) or rRNA sequence (from upper left to lower right) is not available for the species. For empty white fields, no interaction was predicted. The rows are ordered according to the range of conservation, that is, interactions with stable partners in all vertebrates appear in higher rows than interactions conserved only within mammals. In general, we observe that once an snoRNA family has occurred, its function is conserved. This results in three main groups. The first group of snoRNAs emerged at the root of vertebrates and accordingly its function is conserved in all Vertebrata, the second group of interactions appeared at within Teleostomi, and the third main group arose in Amniota. Supplementary figures S1.2–S1.6 analogously show conservation of box C/D snoRNA-, box H/ACA snoRNA-, and scaRNA- interactions with 18S rRNA, 28S rRNA, 5.8 rRNA, and snRNAs, respectively.

**Fig. 2.** Interaction conservation of box C/D snoRNA and targets in 18S rRNA. For each target ≤3, snoRNA families are displayed in different colors. Interaction energies $e(t, s, k)$ determine the saturation of the color. ASEs of both D and the D’ box are considered. Crossed-out fields indicate a missing part of the rRNA (from upper left to lower right) and a missing snoRNA (from lower left to upper right). Empty (white) fields indicate that no interaction was predicted. For details see text.
For LSU (28S and 5.8S rRNAs), we found 61 of 62 human box C/D snoRNA interactions and 47 of 54 box H/ACA snoRNA interactions conserved in vertebrates (supplementary figs. S1.3, S1.4, and S1.5). In LSU for one of two doubly guided 2%-O-ribose methylations and three of four /C9s, only a single guide was found to be conserved. One position (alignment site: 28S-4754, corresponding to human 28S-3797) is reported as methylated and pseudouridylated. Interestingly, high ICI scores agree with both interactions.

In snRNAs, we recover 16 of 18 box C/D snoRNA interactions and 17 of the 20 box H/ACA snoRNA interactions (supplementary fig. S1.6). Two /C9s and two methylations are reported to have two interacting scaRNAs. We recovered both guides for the methylated residues but only one guiding scaRNA for both pseudouridylations.

ICI Scores
The distributions for known human interactions are given in the left column of figure 3. The median $ICImod$ values for box C/D snoRNAs interactions with the small subunit (SSU), large subunit (LSU), and snRNAs are 1.04, 0.81, and 0.85, respectively. In analogy, median $ICIsno$ values are 1.64, 1.14, and 1.13. Box H/ACA snoRNAs interacting with SSU, LSU, and snRNAs yield median $ICImod$ scores of 1.07, 0.73, and 0.75, respectively. Here, median $ICIsno$ values are 1.08, 0.69, and 0.89 (table 1).

We missed only one of 98 interaction between box C/D snoRNAs and ribosomal RNAs and 2 of 18 with snRNAs. Because of a more complicated interaction structure of box H/ACA snoRNAs and target RNAs, RNAnoop is less sensitive. Missed interactions (15 of 93) explain the zero peaks in box H/ACA snoRNA concerning ICI curves (solid lines).

For all interactions with targets on 18S rRNA, the scores are >1, displaying high conservation of the interactions. Because of lower alignment quality and therefore putative unaligned modification sites, we obtain lower average ICI scores for 28S rRNA interactions. This emphasizes the importance of high-quality alignments for reliable statements on targets and target conservation.

For many target sites, more than one snoRNA family is predicted to have an appropriate ASE to interact. Comparing the $ICImod$ scores for these families show that there is almost always a single dominating family whose $ICImod$ scores can clearly be discriminated from those of alternative predictions, see figure 3 right part. The $ICIsno$ scores show a similar behavior.

SNORD68 and 18S-484
SNORD68 and 18S-484 is a good example for a conserved interaction. The region containing the methylated uridine at 18S-484 (corresponding alignment position to 18S-428 in the human sequence) and the D'-ASE of SNORD68 comprises

Table 1. Median ICI Values for the Interactions Listed in snoRNAbase.

<table>
<thead>
<tr>
<th>Score</th>
<th>rRNA</th>
<th>All</th>
<th>C/D</th>
<th>H/ACA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ICImod$</td>
<td>18S</td>
<td>1.05</td>
<td>1.04</td>
<td>1.07</td>
</tr>
<tr>
<td></td>
<td>28S</td>
<td>0.76</td>
<td>0.81</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>snRNAs</td>
<td>0.81</td>
<td>0.85</td>
<td>0.75</td>
</tr>
<tr>
<td>$ICIsno$</td>
<td>18S</td>
<td>1.23</td>
<td>1.64</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td>28S</td>
<td>0.90</td>
<td>1.14</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>snRNAs</td>
<td>0.91</td>
<td>1.13</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Fig. 3. $ICImod$ and $ICIsno$ scores. The plots separately show the values of both types of ICI score (sno and mod) and both types of snoRNA (C/D and H/ACA) on the putative targets 18S rRNA, 28S rRNA, and snRNAs. The left part shows the density distribution of the ICI scores. The right part shows a comparison of the best three snoRNA families according to ICI scores for each target site.
11 nt complementarity. The two interacting sequence segments are almost completely conserved from human to sea lamprey (figs. 2 and 4). Two alignment columns show mutations on the rRNA side, but these mutated nucleotides can still form G-U pairs with the corresponding snoRNA nucleotides. The snoRNA region not involved in the interaction display lower sequence conservation. The ICI reflects the good conservation of the interaction: \( IC_{\text{mod}}(18S - 484, \text{SNORD68}_D') = 1.03 \). The average individual mfe \( \epsilon(t, s, k) \) of \(-9.76\) and \(-10.2\) imply low stability and weak conservation of these interactions, although the snoRNA families are present in Eutheria and Amniota, respectively. Furthermore, SNORD68 is a doubly guiding box C/D snoRNA. The ASE upstream of the D-box is complementary to alignment site 28S-3267. This interaction is well conserved in vertebrates. An ICI value of 0.94 and an average mfe of \(-20.01\) indicate that this interaction is very stable.

**Special Cases**

We observed several modifications for which at some point in evolution a second snoRNA guide occurs and is retained. A feasible explanation may be differential snoRNA expression so that a back up of the modification under different cellular conditions is necessary. We found several examples supporting this hypothesis. Pseudouridylation of site 28S-5501 (corresponding to human 28S-6491) has been reported to be guided by SNORA10. With the snoStrip, pipeline homologs have been identified in suprapiamates, carnivores, cow, and horse. For four of five species, where snoRNA and LSU sequences are available, the interaction is predicted with \( IC_{\text{mod}} = 0.73 \). As additional guide matching 28S-5501, we found SNORA63. The interaction is conserved throughout vertebrates with \( IC_{\text{mod}} = 0.81 \). The two snoRNAs are encoded in introns (sense direction) of RPS2 (ribosomal

---

**Fig. 4.** Conservation of the interaction between the region upstream of D'-box of snoRNA family SNORD68 (right side) and the region around the 2'-O-methylated uridine at alignment column 484 (left side). Target RNA segment and ASE are separated by &. The methylated residue is marked with M in the last row. The first row provides information about the structure of the RNA-RNA interaction. Intermolecular base pairs are indicated by arrow brackets. The blue box marks the D'-box. Red and green columns highlight conservation of the RNA-RNA interaction. Completely conserved base pairs are shown in red. Green columns mark base pairs with compensatory mutations. Lighter colors indicate loss of base pairs in individual species. The gray bars at the bottom correspond to the degree of sequence conservation computed by RNAallfold (Bernhart et al. 2008). mfe for the interactions \( \epsilon(t, s, k) \) as predicted by PLEXY are provided in the last column. In *O. garnettii* no interaction can occur due to additional mutations in the core region; therefore, it is marked with asterisk.
protein 2) and EIF4A2 (eukaryotic translation initiation factor), respectively. The expression of these proteins in human is antagonistically up- and downregulated in fetus and adult state as well as in blood and umbilical cord tissue (according to data from the Gene Expression Atlas (Kapushesky et al. 2012) E-GEOD-6236 (Goh et al. 2007), http://www.ebi.ac.uk/gxa/experiment/E-GEOD-6236/, last accessed November 7, 2013; see fig. 5). As a consequence, the expression patterns of these host genes ensure that at least one of the two snoRNAs is present in the different tissues and for different developmental stages.

Another example of redundant guides is the modification of site 18S-761 (corresponding to human 18S-683) by SNORD19 and SNORD69. Both interactions are conserved in amniotes with ICI values of 1.41 and 0.7, respectively. Both snoRNAs reside in introns of guanine nucleotide binding protein-like 3 (GNL3). This protein comprises different isoforms. All transcripts include the intron encoding SNORD19 but at least one isoform ends in front of the intron hosting SNORD69. Thus, SNORD69 expression is regulated by alternative splicing of GNL3. Additional data are compiled in supplementary table S2. A comprehensive analysis of host gene expression patterns, however, will be presented elsewhere.

Another exception from the general pattern of canonical snoRNA functionality is a changeover of the snoRNA guide for a single modification. Consider, for example, the methylation of 18S-1457 (corresponding to sequence position 1328 in human and 1286 in chicken). The site matches the 3′-ASE of the mammalian-specific SNORD32 family. In aves, on the other hand, the modification is addressed by the 5′-ASE of bird-specific snoRNA family GGgCD25. We discovered this interesting behavior by high-scoring ICIs of 1.1 and 1.08 for the same modification. Zemann et al. (2006) reported similar observations in nematodes.

We used the ICI score to investigate conservation of the few published experimentally verified interactions between an snoRNA and its guided modification.

**Human Experimentally Verified Interactions**

Xiao et al. (2009) tested 16 predicted interactions of box H/ACA snoRNAs andΨs in human rRNAs. Although 12 have been verified, four predictions have been rejected in their study. We measured the conservation throughout vertebrates of all 16 interactions using our ICI score (see table 2). High ICI values agree with the experimentally verified interactions. Two of the four negative results are not conserved at all so that no ICI could be computed. Of the remaining two, one cannot be conclusively resolved by our method, leaving a single case where our predictions disagree with the experimental results.

Because of high sequence similarity, snoRNA families SNORA50 and SNORA76 as well as SNORA42 and SNORA80, respectively, have been merged into one family each in our snoRNA data set by snoStrip, denoted SNORA50-76 and SNORA42-80 in following (see supplementary material S3.1 for the alignments). In terms of guiding potential, we treat snoRNA families as entities and do not distinguish between guiding potential of single paralogous sequences. This is not problematic for family SNORA50-76, where both sequences have been verified to guide 18S-46 (human 18S-34) and 18S-118 (human 18S-105) in human by Xiao et al. (2009). In the second case, however, SNORA42 and SNORA80 both have been verified to guide 18S-122 (human 18S-109), but only SNORA80 interacts with 18S-634 (human 18S-572) while SNORA42 does not. The resolution of our analysis correctly predicts that family SNORA42-80 targets 18S-572.

The confirmed interaction between SNORA24 and modification corresponding to 18S-961 (human 18S-863) yields a very low value of $IC_{mod} = 0.38$. Nevertheless, SNORA24 homologs are predicted to interact with the verified target in elephant, opossum, armadillo, wallaby, kangaroo rat, black

![Fig. 5. Expression of SNORA10 and SNORA63 host genes EIF4A2 and RPS2 is shown at different developmental stages (left) and in different organism parts (right). Arrows above the bars indicate down and up regulation, respectively. Figures are taken from the Gene Expression Atlas.](https://academic.oup.com/mbe/article-abstract/31/2/455/998592/1245-969832)

**Table 2. Comparison of $IC_{mod}$ Values with Experimentally Tested Interactions between Box H/ACA snoRNAs and Ribosomal RNAs.**

<table>
<thead>
<tr>
<th>Alignment</th>
<th>Human</th>
<th>Guiding Families</th>
<th>Verified</th>
<th>$IC_{mod}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>18S-46</td>
<td>18S-34</td>
<td>SNORA50,SNORA76</td>
<td>+, +</td>
<td>1.47</td>
</tr>
<tr>
<td>18S-118</td>
<td>18S-105</td>
<td>SNORA50,SNORA76</td>
<td>+, +</td>
<td>1.33</td>
</tr>
<tr>
<td>18S-122</td>
<td>18S-109</td>
<td>SNORA80,SNORA42</td>
<td>+, +</td>
<td>1.47</td>
</tr>
<tr>
<td>18S-634</td>
<td>18S-572</td>
<td>SNORA80,SNORA42</td>
<td>+, −</td>
<td>1.11</td>
</tr>
<tr>
<td>18S-673</td>
<td>18S-609</td>
<td>SNORA24</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>18S-908</td>
<td>18S-815</td>
<td>SNORA28</td>
<td>+</td>
<td>0.93</td>
</tr>
<tr>
<td>18S-961</td>
<td>18S-863</td>
<td>SNORA24,SNORA19</td>
<td>+, −</td>
<td>0.38, 0.86</td>
</tr>
<tr>
<td>18S-964</td>
<td>18S-866</td>
<td>SNORA28,SNORA19</td>
<td>+, −</td>
<td>0.89, −</td>
</tr>
<tr>
<td>28S-4573</td>
<td>28S-3618</td>
<td>SNORA19</td>
<td>+</td>
<td>0.76</td>
</tr>
<tr>
<td>28S-4665</td>
<td>28S-3709</td>
<td>SNORA19</td>
<td>+</td>
<td>0.58</td>
</tr>
</tbody>
</table>

Note.—The SNORA50 and SNORA76 as well as SNORA80 and SNORA42 are paralogous and hence members of the same family in our survey. +, interaction verified; −, interaction rejected.
rat, marmoset, zebra finch, turkey, chicken, lizard, and coelacanth with an average interaction energy of $-26.96$. SNORA19, however, is the best scoring interaction for the modification at 18S-961 (18S-863 in publication). Although this interaction seems to be conserved throughout vertebrates, it has been rejected by Xiao et al. (2009). In the remaining two cases (18S-673 [human 18S-609] and SNORA24 and 18S-964 [human 18S-866] and SNORA19), our analysis agrees with the published negative experimental results.

**Zebra Fish snoRNAs**

Zebra fish snoRNAs that are essential during embryonal development have recently been identified by Higa-Nakamine et al. (2012). Three methylated residues in the rRNAs are guided by snoRNA families SNORD44, SNORD78, and SNORD26, respectively. Our method identified all three interactions as conserved within vertebrates. The interaction between SNORD44 and site 18S-180 (18S-163 in publication) yields a high $IC_{ICImod}$ score of 1.04. SNORD78 guides 2'-O-ribose methylation of the guanine 28S-526 (28S-3745 in publication) in vertebrates with $IC_{ICImod} = 0.85$. We recover SNORD26 as a conserved guide for modification of adenosine at alignment position 28S-939 (28S-398 in publication) with $IC_{ICImod} = 0.62$. The lower value is probably explained by the lower quality of this alignment.

**The Two Most Conserved Ψs**

The two most conserved Ψs are modified by SNORA74 (U19) at least in vertebrates. The modifications at alignment positions 28S-4697 and 28S-4699 (corresponding to human residues 28S-3741 and 28S-3743) are conserved even in bacteria. There, these modifications are produced by the specialized pseudouridine synthase RluD (Ofengand 2002; Ofengand and Bakin 1997). Both Ψs are located in the decoding center, a central region of the ribosome contacting the SSU and the passing tRNAs.

SNORA74 has an exceptional three-hairpin structure conserved from yeast to human (Badis et al. 2003). The computed $IC_{ICImod}$ values of 0.95 and 0.92 confirm the complementarity of corresponding alignment sites and ASEs in the 5'- and 3'-hairpins in all vertebrates (see supplementary material S1.4 and S3.1).

**New snoRNAs for Known Human Modifications**

Although most 2'-O-ribose methylations and pseudouridylation in human rRNAs can already be assigned to snoRNA guides, some cases do not match with a given snoRNA. We suggest that snoRNA families with above average $IC_{ICImod}$ values are most likely the unrecognized conserved snoRNA guides for these modifications. Thus, we predicted appropriate snoRNA guides for 8 of 21 rRNA sites and 2 of 10 unaccounted for U2 snoRNA modifications. The results are summarized in table 3 and supplementary figure S4.1.1 and illustrated in supplementary figures S4.1.2 and S4.1.3. For the rest of the modifications unassigned to known snoRNAs, we suggest that the vertebrate genomes still harbor a few undiscovered snoRNA families containing matching ASEs.

<table>
<thead>
<tr>
<th>Alignment</th>
<th>Human</th>
<th>Type</th>
<th>snoRNA</th>
<th>ASE</th>
<th>$IC_{ICImod}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>18S-526</td>
<td>18S-468</td>
<td>Am</td>
<td>SNORD83</td>
<td>D</td>
<td>1.16</td>
</tr>
<tr>
<td>18S-756</td>
<td>18S-681</td>
<td>Ψ</td>
<td>SNORD55</td>
<td>HP2</td>
<td>1.14</td>
</tr>
<tr>
<td>18S-890</td>
<td>18S-797</td>
<td>Ψ</td>
<td>GgCD20</td>
<td>D</td>
<td>0.98</td>
</tr>
<tr>
<td>18S-1026</td>
<td>18S-918</td>
<td>Ψ</td>
<td>SNORA61</td>
<td>HP1</td>
<td>1.13</td>
</tr>
<tr>
<td>18S-1513</td>
<td>18S-1383</td>
<td>Am</td>
<td>SNORD30</td>
<td>D'</td>
<td>1.1</td>
</tr>
<tr>
<td>28S-2722</td>
<td>28S-1849</td>
<td>Ψ</td>
<td>SNORA51</td>
<td>HP1</td>
<td>1.09</td>
</tr>
<tr>
<td>28S-4824</td>
<td>28S-3863</td>
<td>Ψ</td>
<td>SNORA48</td>
<td>HP2</td>
<td>0.86</td>
</tr>
<tr>
<td>28S-5263</td>
<td>28S-4266</td>
<td>Ψ</td>
<td>SNORA78</td>
<td>HP1</td>
<td>0.83</td>
</tr>
<tr>
<td>U2-23</td>
<td>U2-15</td>
<td>Ψ</td>
<td>GGoA7</td>
<td>HP2</td>
<td>0.83</td>
</tr>
<tr>
<td>U2-79</td>
<td>U2-47</td>
<td>Um</td>
<td>GgCD76</td>
<td>D</td>
<td>0.9</td>
</tr>
</tbody>
</table>

**SSU rRNA**

For three methylations and two Ψs in 18S with previously unassigned snoRNAs, we predict appropriate guides (supplementary fig. S4.1.2). The 3' hairpin of SNORA55 putatively guides pseudouridylation of 18S-756 (human 18S-681) in tetrapods and the 5' hairpin of SNORA61 pseudouridylation of 18S-1026 (human 18S-918) in tetrapods and coelacanth. The interactions have $IC_{ICImod}$ scores of 1.14 and 1.13, respectively. We suppose that these hairpins are double guides, as both hairpins have alternative targets listed in snorNambase. These interactions are also widely conserved in amniotes and in tetrapods and coelacanth with $IC_{ICImod}$ values of 1.24 and 1.05, respectively. We observe a double guiding function of a target binding region not only in box H/ACA snoRNAs but also within the box C/D snoRNA family SNORD30. The reported target to the ASE adjacent to the D'-box is 28S-4761 (28S-3804 in human). We support this interaction by conservation and suggest additional conserved guiding potential for the unassigned methylation of the cytosine corresponding to alignment column 18S-1513 (18S-1383 in human). For orphan methylation of the cytosine at 18S-890 (18S-797 in human), we detected conserved complementarity ($IC_{ICImod} = 0.98$) from coelacanth throughout tetrapods to chicken snoRNA GgCD20. For this snoRNA family, no vertebrate homologs were known previously to the snoStrap search. At last, we identified an interaction between residue 18S-526 (18S-468 in human) and the orphan snoRNA family SNORD83. This interaction is conserved in Teleostomi and has not been reported earlier.

**LSU rRNA**

We find three box H/ACA snoRNAs with $IC_{ICImod}$ value above average as putative guides for modified nucleotides with unknown guide in 28S (supplementary fig. S4.1.3). The first hairpin of SNORA78 is able to interact with the pseudouridylated residue at 28S-5263 (28S-4266 in human), while the other hairpin of this snoRNA is known to target position 28S-5339 (28S-4331 in human). The fact that box H/ACA snoRNAs often guide nearby pseudouridylation enhances our prediction. The two other box H/ACA snoRNAs...
interacting with the previously unknown guided Ψs 28S-2722 (28S-1849 in human) and 28S-4824 (28S-3863 in human) are SNORAS51 and SNORA84, respectively. Both families are reported as orphan snoRNAs, so far. SNORA51 is identified as a homolog of orphan chicken snoRNA GGoACA9 during snoStrip run, and our predicted interaction also turned out to be conserved in chicken.

**U2 snRNA**

U2 snRNA is pseudouridylated at position 23 (U2-15 in human). We predict an interaction of this orphan Ψ with snoRNAs that belong to chicken-annotated orphan GGoACA7 that is conserved throughout vertebrates with $\text{ICl}_{\text{mod}} = 0.83$ (supplementary fig. S4.14). A high-scoring snoRNA family for unassigned 2'-O-ribose modification at site 79 (corresponding to human 47) is chicken snoRNA GGGCD76. The chicken interaction has already been predicted by Shao et al. (2009) without respect to any conservation issues. With $\text{ICl}_{\text{mod}} = 0.9$, this interaction is conserved in vertebrates since coelacanth.

**Functions for Orphan snoRNAs**

The function of 41 human snoRNAs is still unknown. We used our $\text{ICl}_{\text{neo}}$ score to identify conserved complementarity of these orphan snoRNAs to rRNAs or snRNAs. In table 4, we summarize all predictions where at least one of the ICI scores is positive, as well as those where the predicted target site is known to be modified. The interactions are illustrated in supplementary materials S4.2.2 and S4.2.3.

**Orphan Guides for Modifications without Matching snoRNA**

We identified four orphan snoRNAs as conserved guides for known modifications with previously unknown guides (GGgCD20 and 18S-890, SNORD83 and 18S-468, SNORAS51 and 28S-2722, and SNORA84 and 28S-4824). These interactions were already described in the previous section.

**Orphan snoRNA targets $m^5$C**

Orphan SNORD109 has two paralogs in human encoded in introns of the paternally expressed SNURF-SNRNP locus. SNORD109 is expressed in brain and kidney and at lower levels in lung and muscle (Runte et al. 2001). Surprisingly, our analysis revealed that ten nucleotides conserved complementarity of the D'-ASE to target site 28S-5424 (corresponding to human 28S-4414) with $\text{ICl}_{\text{mod}}$ and $\text{ICl}_{\text{neo}}$ values of 0.9 and 0.93, respectively (supplementary figs. S3.2 and S4.2). This nucleotide is reported as 5-methylcytidine ($m^5$C) in 3D Ribosomal Modification Maps Database (http://people.biochem.umass.edu/fournierlab/3dmodmap/humlsu2dframes.php, last accessed November 7, 2013). The methylation of the nucleobase instead of the associated ribose is a chemical modification not normally associated with snoRNAs. It should be kept in mind, however, that also other SNURF-SNRNP-snoRNA complexes, such as SNORD115, have been shown to exhibit noncanonical behavior.

**Orphan snoRNAs and Intricate Structures**

Furthermore, we predict complementarities between stretches of ribosomal RNA without reported modifications and orphan snoRNAs. A careful examination of how differential expression of snoRNAs affects rRNA modification has not been carried out so far (Xue and Barna 2012). Hence, we cannot exclude the possibility of undetected modifications occurring only under certain conditions.

High-scoring interactions are predicted for SNORD116 and 18S-1286 (human position 18S-1162) and SNORD86 and 18S-1345 (human 18S-1219). The putatively targeted nucleotides are located next to pseudoknotted RNA stretches (according to 3D Ribosomal Modification Maps Database, http://people.biochem.umass.edu/fournierlab/3dmodmap/humlsu2dframes.php, last accessed November 7, 2013). Because modifications are capable of stabilizing intricate structures, our predictions are plausible.

**Orphan snoRNA and Extensively Modified Regions**

High scores of $\text{ICl}_{\text{mod}} = 0.88$ and $\text{ICl}_{\text{neo}} = 1.77$ are computed for the ASE upstream of the D-box of SNORD125 as a guide for modification of 18S-1623 (human 18S-1440). It is located in a functional region of the SSU, close to rRNA binding sites, where other reported modifications are proximal.

For SNORA49, we predict targets for both hairpins. The putative modifications are 28S-3725 and 28S-3729 (human 28S-2826 and 28S-2830). They are located in a small helical structure with multiple reported modifications (http://people.biochem.umass.edu/fournierlab/3dmodmap/humlsu2dframes.php, last accessed November 7, 2013).

A complete list of predictions of rRNA and snRNA targets for the orphan snoRNAs is provided in the supplementary table S4.2.1. The remaining snoRNAs that lack complementarity to rRNAs and snoRNAs might have alternative functions, for example, cleavage of the primary rRNA transcript, targeting mRNA and altering their alternative splicing, or translational control in a miRNA-like fashion after being processed into smaller fragment (sdRNAs).

---

**Table 4. Predictions for Orphan snoRNA Families.**

<table>
<thead>
<tr>
<th>snoRNA</th>
<th>ASE</th>
<th>Alignment</th>
<th>Human</th>
<th>Type</th>
<th>$\text{ICl}_{\text{mod}}$</th>
<th>$\text{ICl}_{\text{neo}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGGCD20</td>
<td>D</td>
<td>18S-890</td>
<td>18S-797</td>
<td>Cm*</td>
<td>0.97</td>
<td>1.49</td>
</tr>
<tr>
<td>SNORD109</td>
<td>D'</td>
<td>28S-5424</td>
<td>28S-4414</td>
<td>m5C</td>
<td>0.9</td>
<td>0.93</td>
</tr>
<tr>
<td>SNORD116</td>
<td>D</td>
<td>18S-1286</td>
<td>18S-1162</td>
<td>C</td>
<td>1.09</td>
<td>1.21</td>
</tr>
<tr>
<td>SNORD125</td>
<td>D</td>
<td>18S-1623</td>
<td>18S-1440</td>
<td>C</td>
<td>0.88</td>
<td>1.77</td>
</tr>
<tr>
<td>SNORD83</td>
<td>D</td>
<td>18S-526</td>
<td>18S-468</td>
<td>Am*</td>
<td>1.16</td>
<td>1.86</td>
</tr>
<tr>
<td>SNORD86</td>
<td>D</td>
<td>18S-1345</td>
<td>18S-1219</td>
<td>C</td>
<td>1.1</td>
<td>1.64</td>
</tr>
<tr>
<td>SNORA49</td>
<td>HP1</td>
<td>28S-3725</td>
<td>28S-2826</td>
<td>U</td>
<td>0.75</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>HP2</td>
<td>28S-3729</td>
<td>28S-2830</td>
<td>U</td>
<td>0.9</td>
<td>0.87</td>
</tr>
<tr>
<td>SNORAS51</td>
<td>HP1</td>
<td>28S-2722</td>
<td>28S-1849</td>
<td>Ψ*</td>
<td>1.1</td>
<td>1.07</td>
</tr>
<tr>
<td></td>
<td>HP2</td>
<td>28S-4824</td>
<td>28S-3863</td>
<td>Ψ*</td>
<td>0.86</td>
<td>0.72</td>
</tr>
</tbody>
</table>

Note.—Known modifications are marked by asterisk. HP1 and HP2 stand for ASE in 5'-hairpin or 3'-hairpin of the box H/A/C snoRNA, respectively.
Identification of Distant Homologs

Early studies into snoRNAs frequently used homologous target sites as an argument for the homology of the snoRNAs themselves. The chicken snoRNAs GGgCD3, GGgCD4, GGgCD14, GGgCD24, GGgCD29, GGgCD63, GGgCD64, and GGgCD66 reported by Shao et al. (2009), all members of the box C/D class, may serve as a good example for the validity of this approach. According to the Basic Local Alignment Search Tool (BLAST)-based homology search procedure implemented in snoStrip, they are specific to the avian lineage. Target prediction with PLEXY and their ICI_mod scores identified them as conserved guides for methylations of the target alignment positions 18S-129, 18S-134, 18S-653, 18S-1415, 18S-1892, 28S-5348, 28S-5371, and 28S-5474 (the coordinates refer to the homologous nucleotides in the human RNAs). These positions are targeted by human snoRNA families SNORD42, SNORD4, SNORD62, SNORD110, SNORD43, SNORD60, SNORD1, and SNORD69, respectively. Homologs of these families were readily identified by snoStrip in other mammals but not in sauropsids. They can be combined into common alignments, and a detailed inspection shows that they are indeed homologs. (Alignments are provided in supplementary material S3.2.)

Although most platypus snoRNA sequences from the study of Schmitz et al. (2008) could be merged to mammalian families during the snoStrip search procedure, some appeared as species specific. Analogous to the avian snoRNA families, platypus snoRNA sequences Oa1759, Oa2916, and Oa2126 could also be identified as homologs of the vertebrate snoRNA families SNORD110, SNORD96, and SNORD4 respectively.

In case of box H/ACA snoRNAs, the inference of an evolutionary origin from functional homology was possible only for SNORA64 and GGgACA47, both guiding pseudouridylation of 28S-6029 (28S-4975 in human).

Particularly divergent sequences were observed for the family containing human SNORD62, chicken GGgCD14, and platypus bOaCD62i. The alignment reveals a large deletion in aves lineage in the 3'-part of the snoRNA. The selective pressure is focused on the ASE downstream of the prime box to retain complementarity to the target in the SSU. The aves lineage has 12 nucleotides perfectly complementary to the region around the modified adenine 653 (18S-590 in human), while in mammals the interaction is with 14 nucleotides longer but comprises a mismatch at the eighth position. (Alignment and detailed figure of interaction is provided on the supplementary page S3.2.)

Alignments of avian, platypus, and mammalian families can readily be combined by manual inspection and editing according to the list of correspondences in table 5, demonstrating that these families are indeed homologs.

Discussion

Starting from known snoRNA sequences reported in chicken, platypus, and human, we compiled a data set of 723 snoRNAs categorized into 259 snoRNA families analyzing 47 vertebrate species. Target predictions were computed for all of them, and the known and predicted interactions were studied in various aspects.

We introduced the ICI to measure snoRNA and target coevolution. Our ICI score combines thermodynamic stability of the RNA-RNA duplex with its evolutionary preservation. Evaluation of this measure on all known human interactions recovered ~87% (scaRNAs-snRNAs), 83% (box H/ACA snoRNAs-rRNAs), and nearly 100% (box C/D snoRNAs-rRNAs) as evolutionary conserved (at least) in Eutheria. The benefit of our ICI measure is further supported by consistently high values for experimentally verified interactions. The lack of published negative experimental results makes the estimation of false-positive rates infeasible, because we cannot exclude the functionality of high-scoring interactions.

Surprisingly, complete 28S rRNA sequences are not available for many of the species investigated here. Results concerning snoRNA guides targeting the LSU therefore remain incomplete. Ribosomal RNA operons exist in tandem arrays of different sequence variants. Because of their repetitive nature, they are not included in genome assemblies. However, specific mutations in rRNAs may disable or enable snoRNA interactions. This may expand the complex system of snoRNA families with various slightly different paralogs modifying "special" nucleotides in slightly different rRNAs. Such variations may contribute to the biogenesis of specialized ribosomes, highly adaptive to certain cellular demands (Xue and Barna 2012). When rRNA operons are fully included in genome assemblies, studies like this can be completed.

There have been assumptions that snoRNAs frequently change targets and have lineage-specific functions (Zemann et al. 2006), while other studies assume conserved functions for snoRNA families (Hoeppner and Poole 2012). For a number of reasons, previous studies were unable to investigate snoRNA function over time in such detail. They were based on a more limited set of species, and/or snoRNA sequences, were limited only to box C/D snoRNA interactions, and had no formal method to evaluate conservation of targets. Using the ICI, we traced back the evolutionary history of all snoRNA families and observed stable partnerships between snoRNAs and their associated target sites throughout vertebrates. This is at odds with the stringent conservation of modification sites and the slow evolution of rRNAs and snRNAs on the one hand, while snoRNAs show high variation in their sequences on the other. Closer inspection of the snoRNA sequences revealed high selective pressure on the ASEs preserving the complementarity to the target. Redundant guides and less common changeovers of guides could be resolved using our evaluation method. In individual cases, we found that redundant guides might be processed from host genes with anticorrelated expression profiles, suggesting how such redundant snoRNAs could evolutionarily be maintained. This is in agreement with the hypothesis of Hoeppner and Poole (2012), who suggest constrained drift during the evolution of snoRNAs: An ongoing mobility of snoRNA genes with occupation of genomic locations that maintain the adequate expression pattern.
Hoeppner and Poole (2012) also suggest an artifactual split into separate snoRNA families through RNA family database (RFAM) covariance models. We agree with that idea and were even able to use conserved targets, identified by high ICI values to unravel unrecognized distant homologies.

We applied the ICI to all putative modifications and all snoRNAs and could add new edges to the network of snoRNAs and the network of interactions. We merged snoRNA families, reported in distantly related organisms, based on sequence similarity and function. We further assigned guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unassigned guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unassigned guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designated putative guides to 10 of the 31 modifications with so-far unknown snoRNAs in rRNA and U2 snRNA, designed...
(Cannone et al. 2002) have been used as initial structural constraints. The initial structure constraint for human 5.8S rRNA has been predicted using RNAfold (Hofacker et al. 1994).

Unfortunately, rRNA operons are often excluded from genome assemblies. At the same time, in particular, LSU rRNAs have been rarely cloned and sequenced in independent studies, for example, for phylogenetic purposes. As a consequence, we have better coverage of snoRNAs than of rRNAs in many of the recently sequenced vertebrates. Although we have detected a majority of SSU RNAs (44/47), for LSU, only 17 of 47 nearly full-length sequences could be used for comparative analyses of modification sites. A suitable 5.8 rRNA sequence could be found in 41 species. Nevertheless, the phylogenetic range of the sequences in all three alignments spans the vertebrates from lamprey to human. All alignments are provided in the supplementary material S6.1, S6.2, and S6.3.

Spliceosomal RNAs (U1, U2, U4, U5, U6, U11, U12, U4atac, and U6atac) have been taken from Marz et al. (2008), where manually curated alignments are provided.

Experimentally verified positions of chemical modifications within rRNAs and snoRNAs have been collected from Maden (1986, 1996), Ofengand and Bakin (1997), the SSU rRNA Modification Database (http://library.med.utah.edu/SSUmods/, last accessed November 7, 2013) (McCloskey and Rozenski 2005), and the RNA Modification Database (http://mods.rna.albany.edu/, last accessed November 7, 2013) (Cantara et al. 2011). Modifications in U2 snRNA have been taken from Dönmez et al. (2004) and Yu et al. (1998). Predicted and verified interactions between snoRNAs and their targets have been collected from sRNABase (Lestrade and Weber 2006) and from the literature (Badis et al. 2003; Xiao et al. 2009; Higa-Nakamine et al. 2012).

Mapping between positions in the human rRNA and snoRNA sequences and their corresponding positions in the alignments has been realized with the BioPerl packages AlignIO and SimpleAlign (Stajich et al. 2002). A table of correspondences putatively modified positions in rRNAs and snoRNAs is provided in the supplementary table S6.4.

Target Prediction

For each single snoRNA sequence of an snoRNA family, an independent target prediction was performed. The internal structure of the target RNA determines the accessibility of a prospective binding region and thus has a large influence on thermodynamics and kinetics of hybridization. These effects can be captured at the level of secondary structures by modified RNA folding algorithms such as RNAup (Mu¨ckstein et al. 2006) or intaRNA (Busch et al. 2008). A table of correspondences putatively modified positions in rRNAs and snoRNAs is provided in the supplementary table S6.4. Incorporating such accessibility information into RNAse snoop (Tafer et al. 2010) from Vienna RNA Package 1.7 improved the prediction of box H/ACA snoRNAs targets. Target prediction for box C/D snoRNAs was performed with PLEXY (Kehr et al. 2011). (The exact parameters for the program calls are given in the supplemental material S7.)

Acknowledgments

This work was supported in part by the European Union under the auspices of the FP-7 QUANTOMICS project and by the Deutsche Forschungsgemeinschaft under the auspices of SPP1174 “Deep Metazoan Phylogeny” STA850/2-3 project.

References


Coevolution of SnoRNAs and Their Targets · doi:10.1093/molbev/mst209


