MicroRNA or Not MicroRNA?

David Langenberger[1–3], Sebastian Bartschat[1], Jana Hertel[1,2], Steve Hoffmann[3,2], Hakim Tafer[1,2], Peter F. Stadler[1–8]

[1]Bioinformatics Group, Department of Computer Science, and [2]Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany; [3]LIFE – Leipzig Research Center for Civilization Diseases, University of Leipzig, Germany; [4]Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, D-04103 Leipzig, Germany; [5]Fraunhofer Institut für Zelltherapie und Immunologie, Perlickstraße 1, D-04103 Leipzig, Germany; [6]Department of Theoretical Chemistry University of Vienna, Währingerstraße 17, A-1090 Wien, Austria; [7]Center for non-coding RNA in Technology and Health, University of Copenhagen, Grønnegårdsvej 3, DK-1870 Frederiksberg C, Denmark; [8]Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA

Abstract. The avalanche of next generation sequencing data has led to a rapid increase of annotated microRNAs in the last few years. Many of them are specific to individual species or rather narrow clades. A closer inspection of the current version of `miRBase` shows that dozens of entries conflict with other ncRNAs, in particular snoRNAs. With few exceptions, these cases show little similarities to canonical microRNAs, however, and thus they should be considered as mis-annotations.

1 Introduction

MicroRNAs and small nucleolar RNAs are thought of distinct classes of ncRNAs with very different functions. While microRNAs are matured to $\sim$ 20nt sequences that direct post-transcriptional gene silencing, snoRNAs canonically guide, in their complete form, the chemical modification of mostly rRNAs and snRNAs [1]. On the other hand, high-throughput sequencing studies revealed that snoRNAs are a prolific source of sequence fragments of microRNA size [2–5], termed sdRNAs. At least some of these snoRNA-derived small RNAs, similar to microRNAs, interact human Argonaut and affect gene expression [6]. Recently, efficient gene silencing has been demonstrated for 11 small RNAs derived from box C/D sno-miRNA [5]. Similar short RNAs, in a few cases with validated functions in gene silencing, are also produced from most other well-known structured RNAs including Y RNA [4, 7], vault RNAs [8–10], snRNAs [4], and tRNAs [11–14]. Recent work [15], furthermore, cast doubt on the microRNA nature of several short RNA products that likely originate from the 3'-end of matured tRNAs since they include the post-transcriptionally append CCA tail. The large numbers of CCA-tagged reads from nearly all tRNAs, which are abundant in deep sequencing data, supports a tRNA-origin of a few annotated "microRNAs".

Canonical microRNAs are generated from a quite specific processing pathway [16]: a polymerase II transcript, the primary miRNA precursor (pri-miRNA)

is cropped by the DroshaDGCR8 complex, also known as Microprocessor. The resulting pre-microRNA hairpin uses the exportin-5 pathway to reach the cytoplasm, where it is cleaved to generate the mature miRNA. Early reports [17] of pre-microRNAs originating from pol-III transcription have recently been refuted [18]. A survey of human pol-III transcription [19], furthermore, recovered no annotated microRNA except two mis-annotations: a vault RNA (hsa-mir-886) and the Y5 RNA (hsa-mir-1975). Mirtrons, on the other hand, are short introns forming stable hairpin structures [20–23]. Both ends of mirtrons are defined by the splice sites. A related, mirtron-like source of small RNAs requires both splicing and exosome-mediated trimming to extract the pre-microRNA hairpin [24, 25]. In this case only one end of the precursor hairpin is defined by the splicing reaction. The production of small RNAs from these intronic precursors is independent of Drosha [22, 25]. A recent review [16] lists several additional esoteric pathways, including at least two of them independent of both Drosha and Dicer.

The similarity between H/ACA snoRNAs and microRNAs has been noticed in several computational studies. For example, [27] reports twenty miRNA precursors that show significant similarity to H/ACA snoRNAs; five of these (miR-151, miR-605, mir-664, miR-215 and miR-140) even bind to dyskerin, a component of the H/ACA snoRNP. Some microRNAs, furthermore, are known to be predominantly localized in the nucleolus [28] emphasizing their snoRNA-like features. This may suggest that a subset of microRNA precursors may have evolved from snoRNAs [27]. The production of small RNAs from snoRNAs, on the other hand, is independent of Drosha [6, 3, 5], and in some cases Drosha even inhibits sdRNA formation [3], suggesting that the snoRNAs and (canonical) microRNAs are in general clearly distinguished entities.

Here we investigate systematically the conflicts in annotation between microRNAs in `miRBase` [29] and other classes of ncRNAs as defined by a variety of other databases. Since most of the conflicts, not surprisingly, concern overlaps of microRNA and snoRNA assignments, we focus in particular on these cases.

2 Conflicting microRNA Annotation

In order to determine to what extent the microRNA annotation of `miRBase` conflicts with non-coding RNA annotation stored in other databases, we retrieved the mature miR sequences from `miRBase` (v. 16) and compared them against `Rfam` [30] (v. 10.0) using the mapping tool `segemehl` [31]. We found 38 mature miRNAs mapping perfectly to other annotated ncRNAs. Stringently requiring exact hits of sequences from the same species and collapsing overlaps observed in more than one species left 26 examples. In addition, a few previously known cases from the literature have been included in Table 1. Most of the overlaps concern snoRNAs. In some cases, these "mature microRNAs" have length of 24 or larger, i.e., outside the range observed for canonical microRNAs.

It is important in this context to recall the common practice of annotating microRNAs. Experimental evidence is almost always only available for the *mature* microRNA. After mapping the mature sequence to genome, putative precursor hairpins are then assigned based solely on computational secondary

Table 1. Overlap of annotation as microRNA and other ncRNA classes. Classification probabilities for microRNAs and snoRNAs are listed. SVM refers to the analysis described in section 3.

miRBase		PMID	RNA class	RNA-micro	snoReport H/ACA	snoReport C/D	SVM
gga-miR-3528		19891781	SNORA17	**0.99**	**0.94**	0	
cfa-miR-1836		18392026	SNORA20	0	**0.99**	0	
bta-miR-2427		19633723	SNORA25	0	**0.82**	0	
hsa-miR-664	* ¶	20413612	SNORA36	**0.64**	**0.86**	0	0.10
mmu-miR-1940	¶	18849523	SNORA47	–	**0.97**	0	
bta-miR-2311		19633723	SNORA61	–	**0.60**	0	
mdo-miR-1543		17965199	SNORA74	**0.99**	0	0	
hsa-miR-1248	* ¶	18285502	SNORA81	**0.99**	0	0	0.38
hsa-mir-3651		20483914	SNORA84	0	**0.99**	0	0.05
hsa-miR-1291	¶	18285502	SNORA34	**0.81**	**0.97**	0	**0.89**
tgu-miR-2989	§	20360741	SNORD74	0.09	0	**0.96**	
mmu-miR-3096		20413612	SNORD93	0.04	0	**0.99**	
gga-miR-3535		19891781	SNORD20	**0.82**	0	**0.98**	
gga-miR-3538		19891781	SNORD83B	0.18	0	**0.99**	
gga-miR-1454	§	18430245	SNORD100	0.08	0	**0.91**	
hsa-mir-3647		20483914	SNORD111B	0.04	0	**0.99**	0.02
hsa-mir-3653		20483914	SNORD125	0.02	0	**0.99**	0.25
hsa-miR-1201	†		SNORD126	0	0	**0.99**	0.42
miR-1843	* ¶	20413612	SCARNA3b	0.08	0	0	
oan-miR-1354		18463306	snoU85	0.09	0	0	
bta-miR-2424		19633723	SCARNA10	0.01	**0.98**	**0.99**	
mmu-miR-3069		20413612	SCARNA13	0.09	0	**0.96**	
oan-miR-1348	§	18463306	SCARNA15	–	**0.99**	0	
gga-miR-3540	§	19891781	SCARNA15	0.01	**0.99**	0	
bta-miR-1940	§	19633723	SCARNA4	0	**0.76**	**0.91**	
dre-miR-735		16698962	Y RNA [26]	–	0	0	
hsa-mir-1975	†		Y5 RNA [7]	–	0	0	
dre-miR-733		16698962	vault RNA [8]	–	0	0	
mir-866	†		VTRNA2 [8]	–			
mmu-miR-699	†		RNase MRP	–			
hsa-miR-1246	[1]	18285502	U2	0.05	0	0	
hsa-mir-1274		18285502	tRNA-Lys [15]	0	0	0	
hsa-mir-1280		18285502	tRNA-Leu [15]	**0.99**	0	0	
mir-720		16582102	tRNA-Thr [15]	0	0	0	
hsa-mir-1308	†		tRNA-Gly [15]	–			
mmu-miR-1937b		18849523	tRNA-Pro [15]	0	0	**0.78**	
hsa-mir-151			—	**0.99**	0	0	**0.94**
hsa-mir-215			—	**0.99**	0	0	0.33
hsa-mir-140			—	**0.99**	0	0	**0.84**
hsa-mir-605			—	**0.93**	0	0	**0.62**

† indicates miRBase entries that have been removed in the most release(s) because their source has been convicingly identified as another class of ncRNAs. § reported sno-miRs from [5] in human. ¶ also discussed in [27]. * overlap of microRNA and snoRNA annotation in multiple species. [1] the mature hsa-miR-1246 maps both to the U2 snRNA and a degraded hairpin-like structure deriving from a MLT1M ERVL-MaLR repetitive element.

structure predictions of the surrounding genomic DNA sequence. In many of the cases listed in Tab. 1 we observe that the annotated precursor hairpins only partially overlap alternative annotations, while the short RNA may arise from either of the conflicting putative precursors. Crucially, annotations as snoRNAs or other RNA classes are often supported by direct evidence, such as cloning and sequencing or Northern blots, which are lacking for the putative pre-microRNA.

One possibility to distinguish evolutionarily conserved microRNAs from evolutionarily conserved other ncRNAs, say snoRNAs, is to evaluate the patterns of

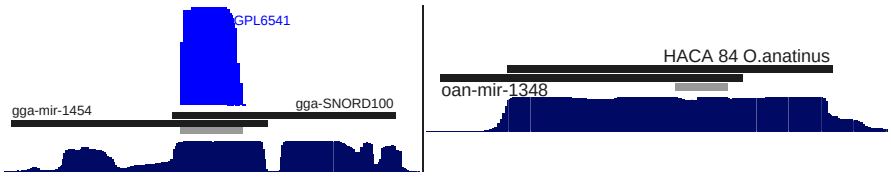

Fig. 1. Examples of putative microRNAs that are most likely mis-annotated. L.h.s.: SNORD100 is a well-conserved box C/D snoRNA, while "mir-1454" would be specific to chicken. R.h.s.: The annotated platypus mir-1348 precursor sequence is located in a putative precursor hairpin whose 5' side is not conserved at all. The, alternative explanation, the 5' side of the 2nd hairpin of the box H/ACA snoRNA SCARNA15, on the other hand, is highly conserved.

sequence conservation. Clearly, this can be conclusive only in those cases where both the mature microRNA and the snoRNA map to unique positions in the genome. Otherwise, the microRNA might, e.g., derive from a paralog of the conserved functional snoRNA locus. A likely example for the latter is tgu-miR-2995, reported as a species-specific microRNA in the zebrafinch. It derives from a degraded paralog of SNORA54 rather than the syntenically conserved, complete, and presumably functional copy of SNORA54. The NAPIL4 gene harbors a single complete copy of SNORA54 in both chicken and zebrafinch. In addition, both tgu-miR-2995 and a part of the SNORA54 sequence also map to different, more 5', intron of the same gene in the zebrafinch only. This mechanism for generating novel microRNAs is consistent with two well established facts: snoRNAs are known to behave like retro-elements in many genomes [32, 33], and several well-document microRNAs arose by exaptation from repetitive elements [34]. It is a possible explanation for the origin of the snoRNA-like microRNAs in [27].

In many cases, however, the putative microRNA is rather poorly conserved and there is little or no conservation for the precursor hairpin, while at the same time the alternative annotation as a snoRNA or other ncRNA features a deep phylogenetic conservation. UCSC Genome Browser representations of two examples are shown in Fig. 1. Although there is clear block of short RNAs for chicken mir-1454, the sequence conservation is extremely poor and there is no signal for a matching miR*. Thus, if mir-1454 is indeed a microRNA, it is almost certainly specific to chicken. On the other hand, SNORD100 is conserved at least across vertebrates. Since there is no paralog of the snoRNA in the chicken genome, it is parsimonious to assume the short reads interpreted as mir-1454 constitute an sdRNA deriving from a box C/D snoRNA precursor. Another example of this kind is SCARNA15, a box H/ACA snoRNA, for which a platypus mir-1348 was annotated, Fig. 1B. The 3'arm of the microRNA containing the annotated mature sequence overlaps the 5'arm of the second hairpin of the H/ACA snoRNA. The stem loop structure of the putative pre-microRNA untypically shows two larger interior loops, while the putative snoRNA shows a perfect double stem loop pattern with perfect conservation of both the H and ACA boxes. Again, the detailed inspection of the locus suggests that it should be considered as a conserved snoRNA rather than a microRNA.

In addition to manual inspection, we applied the class-specific gene finders `RNAmicro` [35] and `snoReport` [36] to assess the overlaps of microRNA and snoRNA annotations of Table 1. The possible classifications are (1) microRNA but not snoRNA, (2) *vice versa*, (3) both classes predicted with high probability and (4) no classification as microRNA or snoRNA at all. As expected, the majority falls into the classes (1) or (2). There are only three candidates for case (3). Neither class is assigned in cases where the putative microRNA precursor hairpin is not conserved in related species so that `RNAmicro` cannot be used, and `snoReport` fails to recognize a box H/ACA or box C/D snoRNA structure.

The main advantage for classifying microRNAs with `RNAmicro` is the use of comparative information. Thus, stem loop structures of annotated microRNAs that look characteristically at a first glance are nevertheless not classified as microRNA if the conservation pattern is not as expected for typical microRNAs. Applying `snoReport` to those sequences (extended if necessary) almost always yields good snoRNA classification. A nice example is the overlapping annotation of mir-1940 and SNORA26 in mouse. While the secondary structure of the annotated miRNA in mouse and related species is a nicely conserved stem loop the underlying conservation pattern is not miRNA-like (constantly high at the mature and mature-star part and low in the hairpin loop region). This is the reason for the low prediction probability ($p = 0.000017$) of the `RNAmicro` SVM. The clear occurrence and conservation of the H and ACA box, their distances to each other and the hairpin-hinge-hairpin-tail secondary structure prediction pattern, however, yields a high classification probability ($p = 0.97$ for box H/ACA snoRNA) of the `snoReport` SVM.

3 Comparative Analysis of H/ACA hairpins

Most box H/ACA snoRNAs consist of two hairpins. We ask here whether hairpins that give rise to large amounts of small reads are more "microRNA-like" than hairpins of H/ACA snoRNAs that are no prolific sources of short RNA products. Hence, we employ an SVM classifier that is trained from two disjoint sets of hairpins: (1) The *bona fide* evolutionary conserved microRNA precursor compiled in [37], which contains neither repeat-derived microRNAs nor lineage-specific ones. (2) A H/ACA hairpin set consisting of those hairpins of H/ACA snoRNAs that show very low levels of short RNA production. To determine small RNA molecules originating from these loci we used mapped reads from different developmental stages of the human brain (GSE18012). Using principal component analysis, we selected the following features for the final SVM classifier: the mean pairing probability of all nucleotides, the number of bound bases, the GC content, the longest paired region, the energy z-score of the precursor and its flanking region, the number of asymmetric bulges of the miRNA arms, as well as the conservation of the arms and the loop of the hairpin. The libSVM classifier was trained and executed in R, using the e1071 package. Repeatedly using 1/3 of the randomly chosen positive and negative loci to trains the SVM resulted in a positive predictive value of 0.94, a sensitivity of 0.88, and a specificity of 0.87.

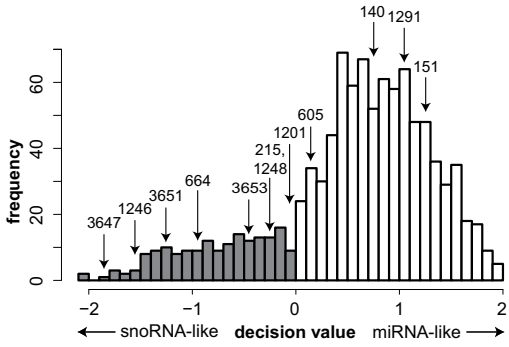

Fig. 2. Histogram of the SVM decision values for all 1,048 miRNAs annotated in `miRBase` v16. The positions of the 12 putatively mis-annotated miRNAs are indicated. Only four of these (mir-605, mir140, mir-1291 and mir-151) are unambiguously classified as miRNAs, of these, only mir-1291 overlaps a known snoRNA. The remaining ones show conservation patterns and structural features similar to snoRNA hairpins.

We apply the SVM to the complete set of microRNAs, including the putative mis-annotations. The classification results are compiled in Tab. 1. Overall, two thirds of the annotated microRNAs with conflicting annotation are not recognized as "microRNA-like" by this approach, supporting our view that these sequence do not constitute true microRNAs.

4 Discussion

In addition to an increasing number of transcripts with multiple processing products and multiple functions, an increasingly diverse universe of small RNAs has been described. Small RNAs are produced by a wide variety of mechanisms, they originate from a broad array of source transcripts, and they exert a broad range of biological functions. This begs the question what exactly should be considered as a microRNA as opposed to the many other types of small RNAs. The most inclusive definition, favored in at least part of the literature, encompasses any short RNA that is incorporated in an Argonaute complex. This point of view has lead to the inclusion in `miRBase` of significant number of small RNAs that are from snoRNAs, snRNAs, tRNAs, and other structured RNAs. We systematically search for such cases and investigated to what extent the ambiguities in the annotation can be decided. We found that short RNAs can often be recognized as products of well-know structured ncRNAs other than microRNAs, leaving also the annotated putative pre-microRNA hairpin doubtful at best.

Although the definition of "microRNA" at first glance may seem to be a purely semantic issue, it has important consequences in practice, since it determines what is included in databases such as `miRBase`. This in turn determines, e.g., what is used in practice as training sets for machine learning approaches. In the case of microRNAs, for which typically the precursor hairpins are utilized, one unknowingly works with contaminated datasets when "microRNAs" are included that are not produced in the canonical way or not all from the annotated data set. The inclusion of mitrons and other non-canonical precursors, for instance, precludes the identification of features associated with Drosha processing. From this point of view, a more stringent curation of microRNAs as well as an explicit annotation of the source of the short RNAs would be highly desirable.

Funding. DFG, FP-7 project QUANTOMICS, European Structure Fund, and the Free State of Saxony.

References

1. Terns, M.P., Terns, R.M.: Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin. Gene Expr. **10** (2002) 17–39
2. Kawaji, H., Nakamura, M., Takahashi1, Y., Sandelin, A., Katayama, S., email, Fukuda, S., Daub, C., Kai, C., Jun Kawai, J., Yasuda, J., Carninci, P., Hayashizaki, Y.: Hidden layers of human small RNAs. BMC Genomics **9** (2008) 157
3. Taft, R.J., Glazov, E.A., Lassmann, T., Hayashizaki, Y., Carninci, P., Mattick, J.S.: Small RNAs derived from snoRNAs. RNA **15** (2009) 1233–1240
4. Langenberger, D., Bermudez-Santana, C., Stadler, P.F., Hoffmann, S.: Identification and classification of small RNAs in transcriptome sequence data. Pac. Symp. Biocomput. **15** (2010) 80–87
5. 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. **39** (2011) 675–686
6. Ender, C., Krek, A., Friedländer, M.R., Beitzinger, M., Weinmann, L., Chen, W., Pfeffer, S., Rajewsky, N., Meister, G.: A human snoRNA with microRNA-like functions. Mol. Cell **32** (2008) 519–528
7. Meiri, E., Levy, A., Benjamin, H., Ben-David, M., Cohen, L., Dov, A., Dromi, N., Elyakim, E., Yerushalmi, N., Zion, O., Lithwick-Yanai, G., Sitbon, E.: Discovery of microRNAs and other small RNAs in solid tumors. Nucleic Acids Res. **38** (2010) 6234–6246
8. Stadler, P.F., Chen, J.J.L., Hackermüller, J., Hoffmann, S., Horn, F., Khaitovich, P., Kretzschmar, A.K., Mosig, A., Prohaska, S.J., Qi, X., Schutt, K., Ullmann, K.: Evolution of vault RNAs. Mol. Biol. Evol. **26** (2009) 1975–1991
9. Persson, H., Kvist, A., Vallon-Christersson, J., Medstrand, P., Borg, A., Rovira, C.: The non-coding RNA of the multidrug resistance-linked vault particle encodes multiple regulatory small RNAs. Nat. Cell Biol. **11** (2009) 1268–1271
10. Mosig, A., Stadler, P.F.: Evolution of vault RNAs. In N.N., ed.: Encyclopedia of Life Sciences. Wiley-Blackwell, Hoboken, NJ (2011) doi: 10.1002/9780470015902.a0022883
11. Lee, Y.S., Shibata, Y., Malhotra, A., Dutta, A.: A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). Genes Dev. **23** (2009) 2639–2649
12. Cole, C., Sobala, A., Lu, C., Thatcher, S.R., Bowman, A., Brown, J.W., Green, P.J., Barton, G.J., Hutvagner, G.: Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. RNA **15** (2009) 2147–2160
13. Haussecker, D., Huang, Y., Lau, A., Parameswaran, P., Fire, A.Z., Kay, M.A.: Human tRNA-derived small RNAs in the global regulation of RNA silencing. RNA **16** 673–695
14. Findeiß, S., Langenberger, D., Stadler, P.F., Hoffmann, S.: Traces of post-transcriptional RNA modifications in deep sequencing data. Biol. Chem. **392** (2011) 305–313
15. Schopman, N.C.T., Heynen, S., Haasnoot, J., Berkhout, B.: A miRNA-tRNA mix-up: tRNA origin of proposed miRNA. RNA Biology **7** (2010) 573–576
16. Miyoshi, K., Miyoshi, T., Siomi, H.: Many ways to generate microRNA-like small RNAs: non-canonical pathways for microRNA production. Mol. Genet. Genomics **284** (2010) 95–103

17. Borchert, G.M., Lanier, W., Davidson, B.L.: RNA polymerase III transcribes human microRNAs. Nat. Struct. Mol. Biol. **13** (2006) 1097–1101
18. Bortolin-Cavaillé, M.L., Dance, M., Weber, M., Cavaillé, J.: C19MC microRNAs are processed from introns of large Pol-II, non-protein-coding transcripts. Nucleic Acids Res. **37** (2009) 3464–3473
19. Canella, D., Praz, V., Reina, J.H., Cousin, P., Hernandez, N.: Defining the RNA polymerase III transcriptome: Genome-wide localization of the RNA polymerase III transcription machinery in human cells. Genome Res. **20** (2010) 710–721
20. Berezikov, E., Chung, W.J., Willis, J., Cuppen, E., Lai, E.C.: Mammalian mirtron genes. Mol. Cell **28** (2007) 328–336
21. Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., Lai, E.C.: The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. Cell **130** (2007) 89–100
22. Ruby, G.J., Jan, C.H., Bartell, D.P.: Intronic microRNA precursors that bypass Drosha processing. Nature **48** (2007) 83–86
23. Chung, W.J., Agius, P., Westholm, J.O., Chen, M., Okamura, K., Robine, Nicolas, Leslie, C.S., Lai, E.C.: Computational and experimental identification of mirtrons in *Drosophila melanogaster* and *Caenorhabditis elegans*. Genome Res. **21** (2011) 286–300
24. Flynt, A.S., Greimann, J.C., Chung, W.J., Lima, C.D., Lai, E.C.: MicroRNA biogenesis via splicing and exosome-mediated trimming in *Drosophila*. Mol. Cell **38** (2010) 900–907
25. Chong, M.M.W., Zhang, G., Cheloufi, S., Neubert, T.A., Hannon, G.J., Littman, D.R.: Canonical and alternate functions of the microRNA biogenesis machinery. Genes Dev. **24** (2010) 1951–1960
26. Mosig, A., Guofeng, M., Stadler, B.M.R., Stadler, P.F.: Evolution of the vertebrate Y RNA cluster. Th. Biosci. **126** (2007) 9–14
27. Scott, M.S., Avolio, F., Ono, M., Lamond, A.I., Barton, G.J.: Human miRNA precursors with box H/ACA snoRNA features. PLoS Comput Biol. **5** (2009) e1000507
28. Politz, J.C., Hogan, E.M., Pederson, T.: MicroRNAs with a nucleolar location. RNA **15** (2009) 1705–1715
29. Griffiths-Jones, S.: The microRNA Registry. Nucleic Acids Res. **32** (2004) D109–11
30. Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., Eddy, S.R.: Rfam: an RNA family database. Nucleic Acids Res. **31** (2003) 439–41
31. Hoffmann, S., Otto, C., Kurtz, S., Sharma, C., Khaitovich, P., Vogel, J., Stadler, P.F., Hackermüller, J.: Fast mapping of short sequences with mismatches, insertions and deletions using index structures. PLoS Comp. Biol. **5** (2009) e1000502
32. Weber, M.J.: Mammalian small nucleolar RNAs are mobile genetic elements. PLoS Genet. **2** (2006) e205
33. Schmitz, J., Zemann, A., Churakov, G., Kuhl, H., Grützner, F., Reinhardt, R., Brosius, J.: Retroposed SNOfall–a mammalian-wide comparison of platypus snoRNAs. Genome Res. **18** (2008) 1005–1010
34. Smalheiser, N.R., Torvik, V.I.: Mammalian microRNAs derived from genomic repeats. Trends Genet. **21** (2005) 322–326
35. Hertel, J., Stadler, P.F.: Hairpins in a haystack: Recognizing microRNA precursors in comparative genomics data. Bioinformatics **22** (2006) e197–e202
36. Hertel, J., Hofacker, I.L., Stadler, P.F.: `snoReport`: Computational identification of snoRNAs with unknown targets. Bioinformatics **24** (2008) 158–164
37. Hertel, J., Lindemeyer, M., Missal, K., Fried, C., Tanzer, A., Flamm, C., Hofacker, I.L., Stadler, P.F., Students of Bioinformatics Computer Labs 2004 & 2005: The expansion of the metazoan microRNA repertoire. BMC Genomics **7** (2006) 15