BMC Genomics

CORRESPONDENCE Open Access

SNOntology: Myriads of novel snornas or just a mirage?

Julia A Makarova and Dmitri A Kramerov*

Abstract

Background: Small nucleolar RNAs (snoRNAs) are a large group of non-coding RNAs (ncRNAs) that mainly guide 2'-O-methylation (C/D RNAs) and pseudouridylation (H/ACA RNAs) of ribosomal RNAs. The pattern of rRNA modifications and the set of snoRNAs that guide these modifications are conserved in vertebrates. Nearly all snoRNA genes in vertebrates are localized in introns of other genes and are processed from pre-mRNAs. Thus, the same promoter is used for the transcription of snoRNAs and host genes.

Results: The series of studies by Dahai Zhu and coworkers on snoRNAs and their genes were critically considered. We present evidence that dozens of species-specific snoRNAs that they described in vertebrates are experimental artifacts resulting from the improper use of Northern hybridization. The snoRNA genes with putative intrinsic promoters that were supposed to be transcribed independently proved to contain numerous substitutions and are, most likely, pseudogenes. In some cases, they are localized within introns of overlooked host genes. Finally, an increased number of snoRNA genes in mammalian genomes described by Zhu and coworkers is also an artifact resulting from two mistakes. First, numerous mammalian snoRNA pseudogenes were considered as genes, whereas most of them are localized outside of host genes and contain substitutions that question their functionality. Second, Zhu and coworkers failed to identify many snoRNA genes in non-mammalian species. As an illustration, we present 1352 C/D snoRNA genes that we have identified and annotated in vertebrates.

Conclusions: Our results demonstrate that conclusions based only on databases with automatically annotated ncRNAs can be erroneous. Special investigations aimed to distinguish true RNA genes from their pseudogenes should be done. Zhu and coworkers, as well as most other groups studying vertebrate snoRNAs, give new names to newly described homologs of human snoRNAs, which significantly complicates comparison between different species. It seems necessary to develop a uniform nomenclature for homologs of human snoRNAs in other vertebrates, e.g., human gene names prefixed with several-letter code denoting the vertebrate species.

Background

Small nucleolar RNAs constitute one of the largest groups of ncRNAs. They guide 2'-O-methylation and pseudouridylation of target RNAs, mainly rRNAs. SnoRNAs are divided into two groups according to the modification type: C/D box snoRNAs guide 2'-O-methylation, while H/ACA box snoRNAs guide pseudouridylation [1,2]. To date, ~200 RNAs of both groups have been described [3]. C/D box snoRNAs contain conserved C (UGAUGA) and D (CUGA) boxes brought together by complementary interactions between the snoRNA termini [4]. In addition, their (often imperfect) copies C' and D' are located internally [5]. Four core proteins bind these boxes, NOP56, NOP58, 15.5 kDa protein, and fibrillarin that catalyzes 2'-O-methylation [6]. Upstream of the D and/or D' box there is an antisense element of 9-20 nucleotides that is complementary to one of the cellular RNAs and is able to interact with it. A nucleotide in the cellular RNA located four nucleotides from the D/D' box in the resulting RNA/RNA duplex is 2'-O-methylated [2,7]. H/ACA box snoRNAs carry boxes H (ANANNA) and ACA (ACA) located at the base of two hairpins. The hairpins contain the antisense elements that are complementary to the target RNAs and are capable to interact with them. Four core proteins bind the H and ACA boxes, NHP2, NOP10, Gar1, and dyskerin; the latter catalyzes

* Correspondence: kramerov@eimb.ru
Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 32 Vavilov St., Moscow 119991, Russia

BioMed Central

© 2011 Makarova and Kramerov; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

pseudouridylation [1,8]. Some C/D and H/ACA RNAs called scaRNAs are localized to Cajal bodies rather than to the nucleolus and guide modification of the snRNAs [9]. According to the new nomenclature accepted for human snoRNAs and scaRNAs, C/D snoRNAs, H/ACA snoRNAs, and scaRNAs are designated as SNORD, SNORA, and SCARNA, respectively [10].

Nearly all snoRNAs and scaRNAs genes in vertebrates are located within introns of other genes called host genes. The small RNAs are processed from pre-mRNAs of host genes [6,11]. Only SNORD3, SNORD13, SNORD118, SCARNA2, and SCARNA17 are transcribed from intrinsic promoters [3]. Most snoRNAs guide rRNA modifications. These modifications are essential for the ribosome function and probably contribute to rRNA folding, maturation, and stability [12,13]. The modification pattern is conserved in vertebrates: most 2'-O-methylation sites are identical between *Xenopus laevis* and human [14]. Homologous snoRNAs in different vertebrate species share the same antisense elements.

Recently, vertebrate snoRNAs have attracted the attention of several research groups [15-18]. In particular, our study of C/D snoRNAs in vertebrates demonstrated a trend towards low copy numbers of C/D snoRNA genes in placental mammals [16]. We have also demonstrated that the set of C/D snoRNAs is well conserved among vertebrates and that species-specific snoRNAs guiding rRNA modifications are extremely rare. Shortly after this publication, Zhu and coworkers reported opposite results [18,19]. Here, we demonstrate that their conclusions are incorrect due to a number of technical errors. We have mainly focused our criticism on their paper in *BMC* Genomics [18]; however, we also considered two other recent publications from the same group which are based on the same erroneous approaches [19,20].

Results

Lineage-specific and species-specific expression patterns of snoRNAs in rhesus monkey are experimental artifacts

Zhang et al. cloned 64 rhesus monkey snoRNAs encoded by 80 genes [18]. All of them were homologs of known human snoRNAs. Expression of these RNAs was tested by Northern hybridization in the muscle of several vertebrate species. Based on the results, Zhang et al. claimed that most of the cloned snoRNAs are not expressed in chicken, and some were not detected even in human and mouse (Table one in Zhang et al. [18]). Stated differently, they claimed lineage- or species-specific expression pattern for most of the cloned snoRNAs (59 out of 64).

This statement is contrary to the following. First, all snoRNAs cloned from rhesus monkey have been previously found in human (which allowed Zhang et al. to identify them) [3]. Second, the pattern of rRNA modifications as well as the set of snoRNAs guiding these modifications are conserved in vertebrates [14-17,21].

The data obtained by Zhang et al. can be interpreted in the following way. The efficiency of Northern hybridization is well known to decrease when a probe contains regions not complementary to the target. Sequence identity between snoRNA homologs from different vertebrate species ranges from ~55 to ~90%. Taxonomically close species have more similar snoRNA homologs. At the same time, different snoRNAs have different similarity levels (Table 1). Accordingly, a hybridization probe for a rhesus snoRNA does not necessarily allow the detection of this snoRNA homologs in other vertebrate species. For instance, we failed to detect SNORD87 RNA in birds using a probe for rat SNORD87, although it readily detected the homologs in different mammals ([22] and our unpublished data). This explains why Zhang et al. could detect only six chicken snoRNAs using rhesus snoRNA sequences as probes (Table one in Zhang et al. [18]). They claim that 58 out of 64 snoRNAs studied are not expressed in chicken; however, 33 of them have been identified by other researchers [17] by cDNA cloning (Additional file 1). Moreover, Zhang et al. reported many snoRNA species as not expressed in chicken [18] but had previously cloned them from chicken [19] (Additional file 1 and see below).

The failure to detect snoRNA expression in human and mouse can be explained similarly. As one would expect, the closer genomic sequences, the more snoRNAs can be detected. Rhesus snoRNA probes detected more snoRNAs in human than in mouse, and more snoRNAs in mouse than in chicken (Table one in Zhang et al. [18]). Note that some snoRNAs whose expression was not detected in mouse (7 out of 17) had been described before (Additional file 1) [23-25]. Due to the same reasons, the attempt of Zhang et al. to detect snoRNAs that were not detected in muscle, in other human and mouse tissues also failed since the same rhesus probes were used.

The cases when snoRNA expression was not detected in human look particularly odd considering that all these snoRNAs have been initially described in human (Additional file 1). Moreover, the names specified,

Table 1 Examples of similarity variation between mammalian and avian snoRNAs

SnoRNA	Human snoRNA identity to	
	mouse snoRNA, %	chicken snoRNA, %
SNORD46	92	61
SNORD87	88	71
SNORA13	82	56

SNORA and SNORD, correspond to the new nomenclature specifically designed for human snoRNAs [10], a fact that alone indicates their expression in human. Thus, the lineage-specific and species-specific expression patterns of rhesus snoRNAs reported by Zhang et al. are experimental artifacts.

Identification of species-specific ncRNAs in chicken results from improper use of Northern hybridization

A similar mistake was made by Zhang et al. in their publication describing chicken snoRNAs [19]. They cloned 125 chicken ncRNAs, mainly snoRNAs, and attempted to detect these RNAs in chicken, mouse, and human tissues by Northern hybridization. Similarly to the results discussed above, positive signal was largely observed in chicken only.

Zhang et al. detected the same snoRNAs in chicken but not in human and mouse [19]; and later, in rhesus, human, and/or mouse but not in chicken [18]. Each time species-specific expression of these snoRNAs was alleged. Examples of such detection experiments are given in Figure 1 and Additional file 2.

Novel chicken ncRNAs are homologs of known human ncRNAs

Zhang et al. reported 35 new ncRNAs in chicken [19]. They claimed that these RNAs (with a single exception) can be detected by Northern hybridization only in chicken, and genes for most of them (28 out of 35) are absent in the genomes of other vertebrates. Table 2 demonstrates that 30 out of 35 so-called "novel" RNAs are homologs of previously described human small RNAs, 27 of which are snoRNAs. In each case, a snoRNA shares the antisense element with a human homolog (Additional file 3). Most of these allegedly new chicken RNAs can be identified by the search systems of the Rfam database of ncRNAs [21] and the snoRNABase of human nucleolar RNAs [3] (Table 2). Moreover, a good fraction of these "novel" chicken RNAs had been cloned by Shao et al. [17], and this fact was acknowledged by Zhang et al. (Table one in Zhang et al.[19]). Shao et al. managed to identify these RNAs as human snoRNA homologs, while Zhang et al. presented them as new RNAs. Thus, most novel ncRNAs described by Zhang et al. in chicken are homologs of well-known human ncRNAs.

Too long antisense elements and wrong target site predictions

Zhang et al. presented sequences of the C/D snoRNAs cloned from rhesus monkey and identified the whole fragments between C and D' boxes, as well as between C' and D boxes as the antisense elements (Additional file one in Zhang et al.[18], one example is given in Figure 2). However, it is known that an antisense element (or a guide sequence) is not a snoRNA fragment between the conserved boxes but rather a specific fragment complementary to the target RNA. In most cases it is not long, usually from 9 to 20 nt [3], which is much shorter than the fragments specified by Zhang et al.

Zhang et al. performed a computer search for the targets of rhesus C/D snoRNAs (Additional file three in Zhang et al.[18]). However, the targets for these snoRNAs were identified long ago, and the methylation of most of them was demonstrated [3]. For instance, SNORD87 RNA can guide modification of G-3723 in 28S rRNA, and this nucleotide is actually 2'-O-methylated [14,22] (Figure 2). With a few exceptions, the targets identified by Zhang et al. do not correspond to the confirmed ones. For example, the nucleotide in rhesus U6 RNA putatively modified by SNORD87 RNP is not methylated in human RNA [3] and, considering the conserved pattern of RNA modifications, is almost surely unmethylated in rhesus monkey (Figure 2). Zhang et al. identified methylation targets in 5S rRNA, whereas it has no 2'-O-methylated nucleotides in eukaryotes [26]. In addition, due to a small size of antisense elements, hundreds of potential targets can be proposed; and presenting some of them without experimental verification of their methylation status is unsubstantiated.

It was shown that a modified base is located four nucleotides upstream of the D/D' box in the C/D snoRNA/target RNA duplex [2,7]. In many cases presented by Zhang et al., e.g., in the putative SNORD87 target in SSU rRNA (Figure 2), a complementary sequence is more than four nucleotides away from the D/D' box, which makes the modification of these putative target RNAs by the proposed snoRNAs impossible.

Numbers of snoRNAs and their gene copies in non-mammalian species is substantially underestimated

Zhang et al. stated that the numbers of snoRNAs and their genes increase from fish, amphibians, and birds to mammals [18]. Instead of a search for the new snoRNA genes, they used ENSEMBL annotations based on the Rfam database [27]. Identification of homologs of the experimentally detected ncRNAs is much more complex compared to protein homologs due to their low sequence similarity. In the case of snoRNAs, the conserved elements (antisense elements and C, C', D, and D' boxes in C/D snoRNAs or H and ACA boxes in H/ACA snoRNAs) comprise a half of the sequence length at most. The similarity level in non-conserved sequences varies between vertebrates and is usually low (Figure 3; Additional file 3). In addition, snoRNA genes in different species can be located within different introns of the same host gene or within different host genes. Thereby,

Figure 1 Controversial results of detection of snoRNAs. Hybridization of RNA isolated from different tissues of rhesus monkey, chicken, human, and mouse with rhesus snoRNA probes (left panel; from Zhang et al., 2010 [18]) and with chicken snoRNA probes (right panel; from Zhang et al., 2009 [19]). Conventional names are framed. The same RNAs are shown side-by-side. Clearly, the hybridization results on the left and on the right are mutually exclusive.

Table 2 Chicken ncRNAs cloned and presented as novel RNAs by Zhang at al [19] are homologs of well-known human ncRNAs

RNA ID[1]	GenBank ID	RNA name	Identifiable by Rfam search	Identifiable by snoRNAbase search	Cloned and properly identified by Shao et al. [17]
GGN11	EU240230	**SNORD102B**[2]	No	**yes**	no
GGN20	EU240238	**SNORD1B**	No	**yes**	**yes** (GGgCD64)
GGN86	EU240302	**SNORD13**	**Yes**	**yes**	no
GGN120	EU240333	**fragment of SNORA84**	**Yes**	**yes**	no
GGN148	EU240352	**SNORD104**	No	**yes**	no
GGN100	EU240315	**SNORD11A**	**Yes**	**yes**	**yes** (GGgCD12A)
GGN71	EU240287	**SNORD127**	**Yes**	**yes**	no
GGN107	EU240321	**SNORD81**	**Yes**	**yes**	**yes** (GGgCD31)
GGN52	EU240268	**SNORD44**	**Yes**	**yes**	**yes** (GGgCD6)
GGN34	EU240252	**SNORD87C**	No	**yes**	**yes** (GGgCD46a)
GGN108	EU240322	**SNORD46A**	No	**yes**	**yes** (GGgCD47a)
GGN80	EU240296	**SNORD62**	No	**yes**	**yes** (GGgCD14)
GGN82	EU240298	**SNORD4**	**Yes**	**yes**	**yes** (GGgCD4)
GGN17	EU240236	**SNORD1A**	No	**yes**	**yes** (GGgCD64)
GGN79	EU240295	**SNORA77**	**Yes**	**yes**	**yes** (GGgACA12)
GGN72	EU240288	**SNORA40**	**Yes**	no	**yes** (GGgACA20)
GGN87	EU240303	**SNORA44**	**Yes**	**yes**	no
GGN58	EU240274	**SNORA17**	**Yes**	**yes**	no
GGN56	EU240272	**SNORA15**	**Yes**	no	no
GGN32	EU240250	**SNORA31B**	**Yes**	no	**yes** (GGgACA38)
GGN123	EU240336	**SNORA4**	No	no	**yes** (GGgACA26)
GGN74	EU240290	**SNORA64**	No	no	**yes** (GGgACA47)
GGN103	EU240318	**U4atac**	**Yes**	no	no
GGN141	EU240348	**SNORA25**	No	**yes**	**yes** (GGgACA11)
GGN67	EU240283	**fragment of SCARNA11**	**Yes**	no	**yes** (GGgACA29)
GGN105	EU240320	**NET3**[3]	No	no	no
GGN68	EU240284	**SNORD97**	No	**yes**	no
GGN46	EU240262	**SNORD43**	No	**yes**	**yes** (GGgCD29)
GGN147	EU240351	**Vault RNA**	**Yes**	no	no
GGN16	EU240235	**fragment of SNORD46B**	No	No	**yes** (GGgCD47b)

[1] According to Zhang et al. [19];listed in the same order as in Table one in [19].

[2] The SNORD102B transcript has a longer antisense element than SNORD102A, and thus can guide the modification the rRNA nucleotide adjacent to that guided by SNORD102A [16].

[3] NET3 RNA is described by us [16] and is specific for vertebrates except placental mammals.

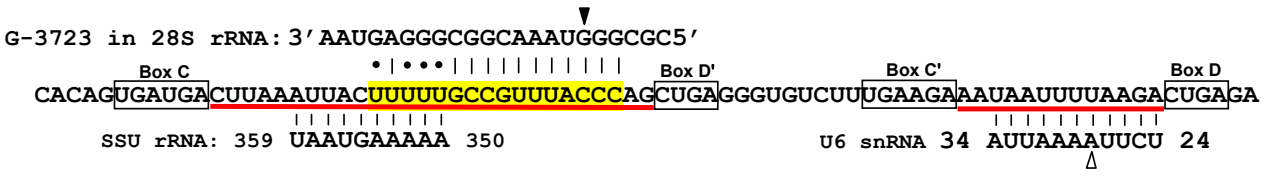


Figure 2 Wrong prediction of snoRNA targets exemplified by rhesus monkey SNORD87 RNA. C, D', C', and D sequences are boxed; the antisense element is marked yellow, and the complementary region in 28S rRNA is shown. The target nucleotide for 2'-O-methylation guided by SNORD87 is indicated by the solid arrowhead. The regions erroneously identified as antisense elements by Zhang et al. [18] are underlined in red. The putative SNORD87 targets identified by Zhang et al. are given below. The only possible SNORD87-guided modification among these targets is indicated by the empty arrowhead. This nucleotide is not methylated in human U6 snRNA.

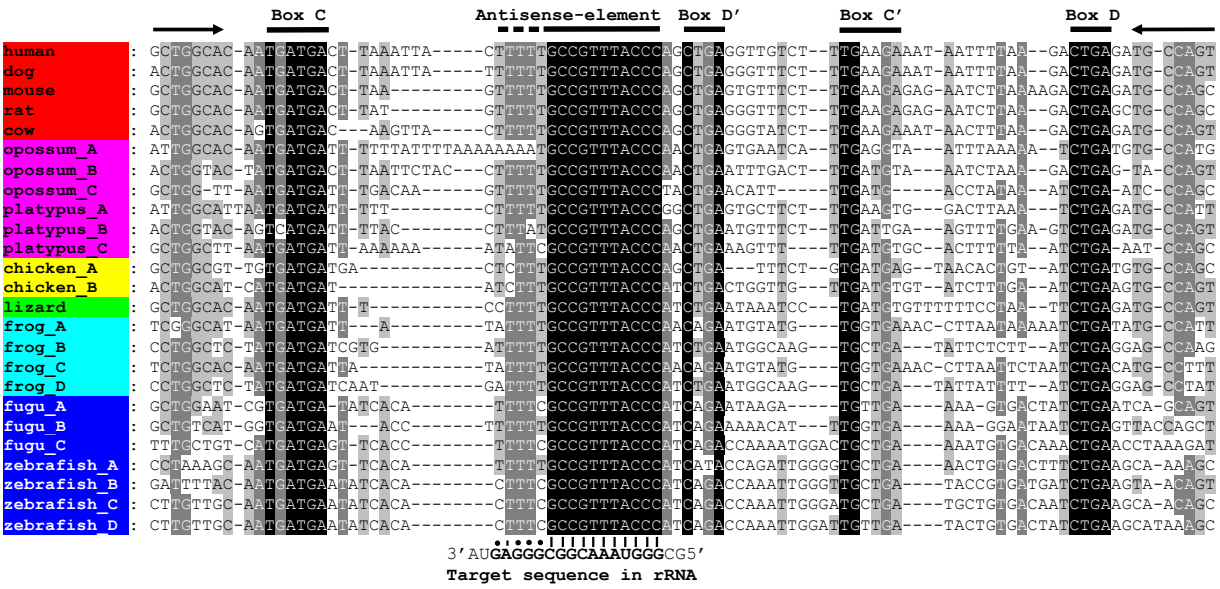


Figure 3 Alignment of SNORD87 RNA genes. Conserved elements are marked with lines above the alignment. A fragment of 28S rRNA complementary to the antisense element in SNORD87 is given below the alignment. The G-T complementarity is marked with dots. SNORD87 sequences are given for the following vertebrates: human (*Homo sapiens*), dog (*Canis familiaris*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), cow (*Bos taurus*), opossum (*Monodelphis domestica*), platypus (*Ornithorhynchus anatinus*), chicken (*Gallus gallus*), lizard (*Anolis carolinensis*), frog (*Xenopus tropicalis*), fugu (*Takifugu rubripes*), and zebrafish (*Danio rerio*).

many snoRNA genes are missing from lists created by annotation programs.

Our study on the numbers of C/D snoRNAs and their genes in representatives of different vertebrate classes [16] yielded results contrary to those obtained by Zhang et al. [18]. Instead of using automatic annotations, we searched for each C/D snoRNA in the vertebrate genomes using the WU BLAST 2.0 algorithm with specifically selected relaxed parameters; and the results of each search were manually inspected [16]. The data obtained and supplemented in this work (1352 C/D snoRNA genes; Figure 4, 5 and Additional file 4) did not reveal any significant increase in the number of C/D snoRNAs in mammals, as compared to other vertebrates. We found that most human snoRNAs have homologs in other vertebrate classes. Moreover, our data demonstrated a trend towards low copy numbers of C/D snoRNA genes in placental mammals. For instance, SNORD87 RNA is encoded by four genes in *Xenopus* and zebrafish each; two genes, in chicken; and by a single gene in human.

Zhang et al. failed to find many snoRNA genes in vertebrates. Figure 6 lists snoRNA genes identified by Zhang et al. (marked gray, according to Figure three in [18]) and missed by them but identified by other researchers (marked red [3,17,21], including our own data (Additional file 5)). The latter portion also includes snoRNAs cloned by Zhang et al. from chicken [19] (even though they claimed the absence of these RNAs in chicken in subsequent paper [18]). A plus sign in Figure 6 indicates genes present in the new release of Rfam (10.0), which shows how severely the conclusions by Zhang et al. depend on the Rfam release used. However, this release still does not contain many snoRNA genes identified in specific snoRNA studies (Figure 6). This particularly applies to the C/D RNA genes described by us (Additional file 4). Thus, studies specifically designed for a search of a particular group of ncRNAs in the whole genomes give much better results than the use of databases with automatically annotated ncRNAs.

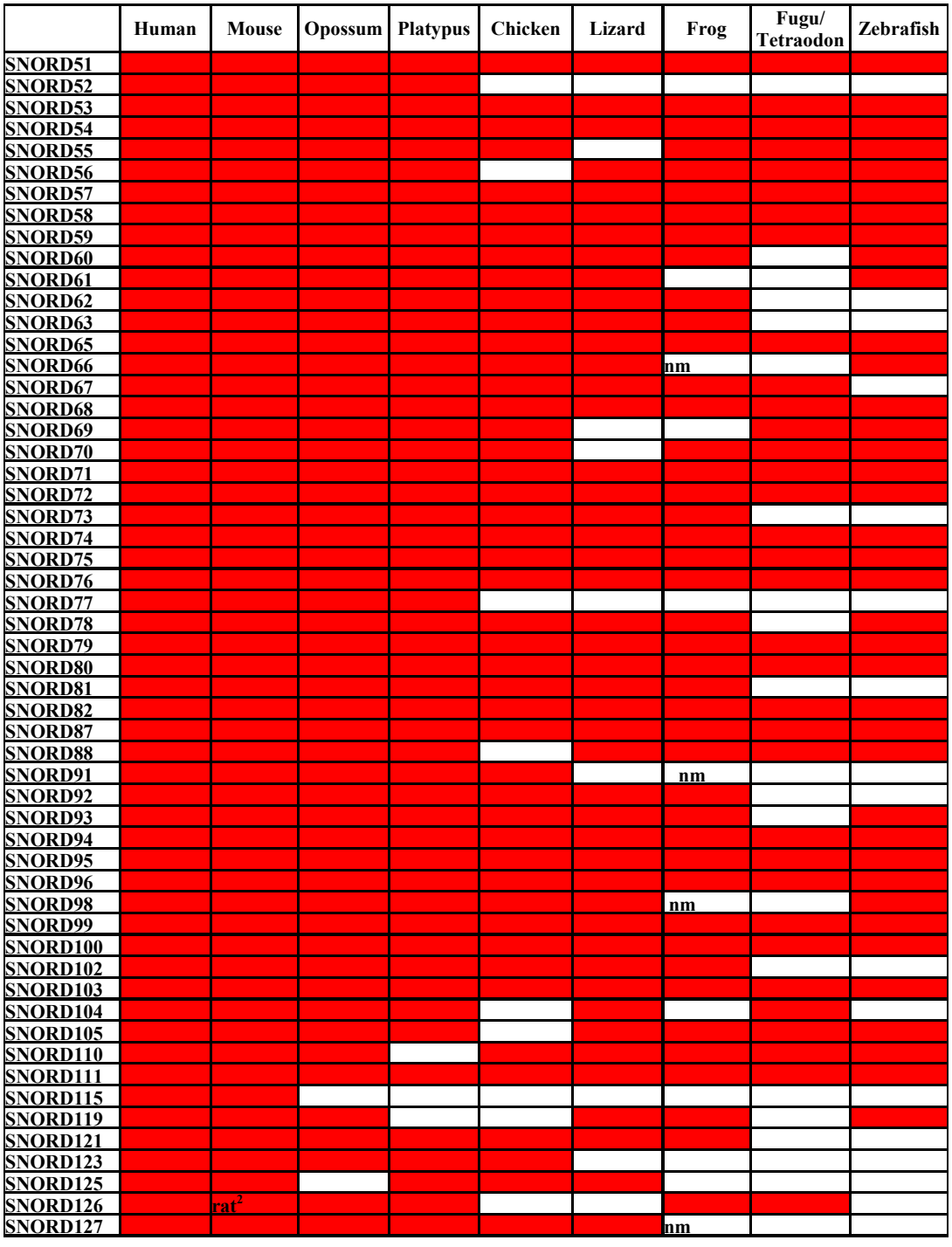
In contrast to the consecutive increase in the number of snoRNAs from fish to mammals alleged by Zhang et al., we found that most mammalian C/D snoRNA genes have homologs in the genomes of other vertebrate classes (Figures 4, 5 and 6). This is not surprising considering that most snoRNAs are involved in rRNA modifications, and that the pattern of rRNA 2'-O-methylation and, likely, pseudouridylation is rather conserved in vertebrates [14]. The cases when some snoRNA gene is not found in a particular species can be attributed to the gaps in the genome sequences (which are abundant in the genomes of vertebrates excluding human and mouse). A minor fraction of snoRNA genes can be missing in some vertebrate classes considering some variations in the pattern of rRNA modifications between vertebrates. For instance, differential rRNA 2'-

Figure 4 Taxonomic distribution of C/D snoRNAs with identified targets[1]. The genes that have been found by us in the genomes assemblies are marked red (Additional File 4). "nm," not methylated site in *Xenopus* [14].1Targets are unknown for SNORD23, SNORD64, SNORD83, SNORD84, SNORD86, SNORD89, SNORD90, SNORD97, SNORD101, SNORD107, SNORD108, SNORD109, SNORD112, SNORD113, SNORD114, SNORD116, SNORD117, and SNORD124. Records SNORD39, SNORD40, SNORD106, SNORD120, and SNORD122 were deleted from the NCBI Nucleotide database. SNORD85 is an isoform of SNORD103. SNORD3, SNORD13, SNORD22, and SNORD118 guide no modifications.

O-methylation between human and frog is observed in 9 out of ~100 sites [14]. It is of interest that about a half of missing snoRNA genes is observed in fishes (Figures 4, 5 and 6), which can point to a specific pattern of their rRNA methylation relative to other vertebrate classes.

Number of mammalian snoRNA genes is substantially overstated

Zhang et al. stated that the number of snoRNA genes steadily increases in the series from fish to mammals, and that there is a burst in their number in mammals [18]. Again, ENSEMBL annotations based on the Rfam



	Human	Mouse	Opossum	Platypus	Chicken	Lizard	Frog	Fugu/ Tetraodon	Zebrafish
SNORD51	■	■	■	■	■	■	■	■	■
SNORD52	■	■	■	■					
SNORD53	■	■	■	■	■	■	■	■	■
SNORD54	■	■	■	■	■	■	■	■	■
SNORD55	■	■	■	■	■		■	■	■
SNORD56	■	■	■	■		■	■	■	■
SNORD57	■	■	■	■	■	■	■	■	■
SNORD58	■	■	■	■	■	■	■	■	■
SNORD59	■	■	■	■	■	■	■	■	■
SNORD60	■	■	■	■	■	■	■		■
SNORD61	■	■	■	■	■	■			■
SNORD62	■	■	■	■	■	■	■		
SNORD63	■	■	■	■	■	■	■		
SNORD65	■	■	■	■	■	■		■	■
SNORD66	■	■	■	■	■	■	nm		■
SNORD67	■	■	■	■	■	■	■	■	
SNORD68	■	■	■	■	■	■	■	■	■
SNORD69	■	■	■	■	■			■	■
SNORD70	■	■	■	■	■		■	■	■
SNORD71	■	■	■	■	■	■	■	■	■
SNORD72	■	■	■	■	■	■	■	■	■
SNORD73	■	■	■	■	■	■	■		
SNORD74	■	■	■	■	■	■	■	■	■
SNORD75	■	■	■	■	■	■	■	■	■
SNORD76	■	■	■	■	■	■	■	■	■
SNORD77	■	■	■	■					
SNORD78	■	■	■	■	■	■	■		■
SNORD79	■	■	■	■	■	■	■	■	■
SNORD80	■	■	■	■	■	■	■	■	■
SNORD81	■	■	■	■	■	■	■		
SNORD82	■	■	■	■	■	■	■	■	■
SNORD87	■	■	■	■	■	■	■	■	■
SNORD88	■	■	■	■		■	■	■	■
SNORD91	■	■	■	■	■		nm		
SNORD92	■	■	■	■	■	■	■		
SNORD93	■	■	■	■	■	■	■		■
SNORD94	■	■	■	■	■	■	■	■	■
SNORD95	■	■	■	■	■	■	■	■	■
SNORD96	■	■	■	■	■	■	■	■	■
SNORD98	■	■	■	■	■	■	nm		■
SNORD99	■	■	■	■	■	■	■	■	■
SNORD100	■	■	■	■	■	■	■	■	■
SNORD102	■	■	■	■	■	■	■		
SNORD103	■	■	■	■	■	■	■	■	■
SNORD104	■	■	■	■		■		■	
SNORD105	■	■	■	■		■	■	■	■
SNORD110	■	■	■		■	■	■	■	■
SNORD111	■	■	■	■	■	■	■	■	■
SNORD115	■	■							
SNORD119	■	■	■			■	■		■
SNORD121	■	■	■	■	■	■	■		
SNORD123	■	■	■	■	■				
SNORD125	■	■		■	■	■			
SNORD126	■	rat[2]	■	■			■	■	
SNORD127	■	■	■	■	■	■	nm		

Figure 5 Taxonomic distribution of C/D snoRNAs with identified targets. *(Continued)*[2]The gene is missing in the mouse genome since the locus is deleted.

database were used rather than their own data. For each ncRNA, Rfam specifies all homologs in different species without specifying if a particular sequence is a gene or a pseudogene. This problem requires detailed examination of both the proper sequence and its genomic environment which is not covered by Rfam. Accordingly, Rfam records do not necessarily represent ncRNA genes, but may represent their pseudogenes as

	Human	Monkey	Mouse	Platypus	Chicken	Frog	Medaka	Zebrafish
SNORA54				+				
SNORD22								
SNORD118					+			
SNORA2								
SNORA62					+			
SNORD94					+			
SNORA5					+			
SNORA23					+			
SNORA73					+			
SNORD15								
SNORA8							+	
SNORA53							+	+
SNORA19							+	+
SNORD67							+	
SNORA71							+	+
SNORA42							+	
SNORD26				+				
SNORD27				+				+
SNORA61								
SNORA24								
SNORA50							+	+
SNORA27							+	+
SNORA28							+	
SNORA66								
SNORA7							+	+
SNORA72								
SNORD45								
SNORA18				+			+	+
SNORA40					+			
SNORD8							+	+
SNORA25					+			
SNORA49								
SNORA41					+			
SNORA17					+		+	+
SNORA31					+		+	+
SNORA13					+		+	+
SNORA36					+	+		+
SNORA70							+	+
SNORA43								
SNORD87				+	+		+	
SNORA76				+	+			
SNORD116								
SNORD17				+	+			
SNORD46							+	+
SNORA14				+	+		+	+
SNORA20					+			
SNORA58					+			
SNORA74				+	+		+	+
SNORA64								
SNORA68				+	+		+	+
SNORA11								
SNORA4				+	+			
SNORA15			+		+		+	

Figure 6 Taxonomic distribution of snoRNA genes cloned from rhesus monkey by Zhang et al. The gene names are listed in the same order as in Figure three of Zhang et al. [18]. The genes detected by Zhang et al. are marked grey, while those not detected by them but available in the open sources (see Additional file 5) are marked red. The latter genes available in Rfam 10.0 are indicated by the plus sign.

well, and this is clearly indicated in the Help section of the database [21]. However, Zhang et al. considered all corresponding Rfam and ENSEMBL entries as snoRNA genes: they reported the identification of 744 snoRNA genes in rhesus monkey, 922 genes in mouse, more than 1000 genes in human, and ~2200 genes in platypus. The problem of snoRNA gene copy numbers in mammals is discussed in several publications by different groups (see review [28] and references therein). All these data agree with each other, as well as with our data [16]: while the number of known mammalian snoRNAs is about 200, the total number of their genes does not exceed ~450 (i.e., some snoRNAs are encoded by single genes, and others are encoded by two, three, or more). This is substantially less than proposed by Zhang et al. Most mammalian-specific snoRNA genes found by them reside in intergenic regions rather than in introns. It is generally accepted that nearly all snoRNA genes of vertebrates are localized in introns of host genes, and only SNORD3 (U3), SNORD118 (U8), SNORD13 (U13), SCARNA2, and SCARNA17 are transcribed from their own promoters. It has been well documented that expression of the intronic snoRNAs requires transcription of the host genes (e.g., review [29] and references therein). That is why any sequence similar to an intronic snoRNA gene outside of introns is most likely a nonfunctional pseudogene. Only full-length copies with intact conserved regions and specific secondary structure can be considered as putative snoRNA genes. In addition, a search for a host gene, which may remain unannotated, should be done. Zhang et al. made no such analysis for the intergenic sequences annotated by ENSEMBL as snoRNA genes. Screening the human genome for snoRNA-like sequences revealed that most of them proved to be nonfunctional retrogenes with substitutions in the conserved regions [16,30]. Clearly, Zhang et al. considered such pseudogenes as snoRNA genes. We have demonstrated that the number of C/D snoRNA pseudogenes is much higher in mammals than in other vertebrates [16]. Therefore, the burst in mammalian snoRNA gene numbers alleged by Zhang et al. most likely represents the burst in the number of their pseudogenes.

Thus, Zhang et al. overestimated the number of snoRNA genes in mammals but underestimated the numbers of snoRNAs and their genes in other vertebrates. This led to a false conclusion that the numbers of snoRNAs and their genes increase in the series from fish to mammals.

Are intronic snoRNA genes indeed transcribed from their own promoters?

SnoRNA pseudogenes with intact conserved regions could, in theory, be functional even when located outside of host gene introns, i.e. in intergenic regions. For that to happen, they should possess their own promoters that would allow independent transcription. Li et al. attempted to find such promoters for intergenic snoRNA-like sequences as well as independent promoters for snoRNA genes located within introns of the host genes [20]. They selected 745 putative human snoRNA genes, 326 of which were located in intergenic regions. This is much a higher number than the generally accepted estimate of the number of snoRNA genes (~450, see above). Again, Li et al. used ENSEMBL annotations, thus, combining snoRNA genes and pseudogenes. The search for snoRNA promoters using the CoreBoost_HM program [31] identified them in 179 out of 745 loci: 155 intronic loci and 24 intergenic ones (Table two in Li et al. [20]).

Based on these results, Li et al. proposed five models of snoRNA transcription. The first model assumes that transcription of a snoRNA and a host gene occurs from a common promoter and is generally accepted. This model describes most of the snoRNAs studied. Other models assume that transcription of a snoRNA gene occurs from an independent promoter.

The second model suggests an intronic snoRNA gene with its own promoter independent of a host gene promoter. This model was exemplified by one of SNORD3 (U3) genes located in an intron of the TEX14 gene on chromosome 17 (Model I, Figure one in Li et al. [20]). However, it is well known that SNORD3 always possesses its own promoter and requires no host gene for its transcription. Therefore, SNORD3 can not be used as an illustration of the proposed model. Moreover, the sequence on chromosome 17 has numerous substitutions in the functional regions and, hence, is a nonfunctional SNORD3 pseudogene (Additional file 6).

The other three models describe snoRNA genes located outside of host genes and putatively transcribed from their own promoters. However, the SNORA75 gene located on the plus strand of chromosome 12 and used for illustrating the third model (Model III, Figure one in Li et al. [20]) is actually a pseudogene with missing 5'-terminus (Additional file 6). Models IV and V are presented in Figure 7. One can see that the snoRNA genes are within introns of overlooked host genes rather than within intergenic regions. Thus, the promoters identified by Li et al. as snoRNA promoters are, in fact, host gene promoters.

Other genes identified by Li et al. as independently transcribed snoRNA genes are presented in Additional file 6. In each case, there is either an unnoticed host gene harboring snoRNA genes in its introns or a snoRNA pseudogene with substitutions questioning its functionality. A few exceptions are SNORA26-like

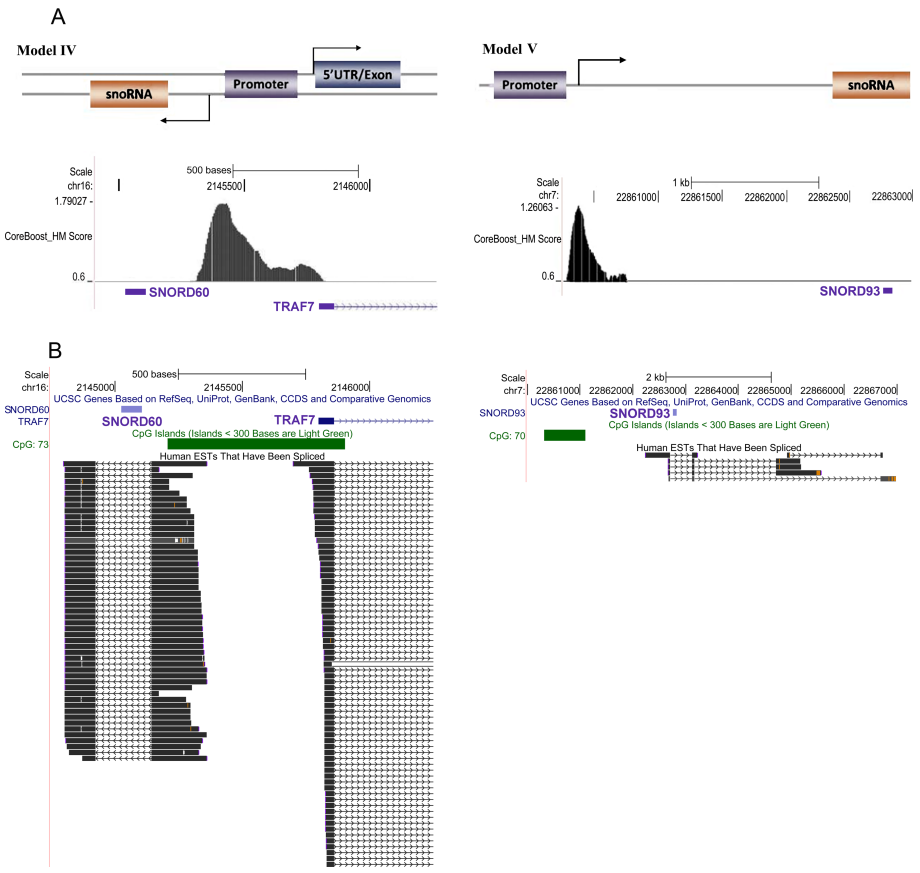


Figure 7 Examples given by Li et al.[20]do not prove models IV and V of independent transcription of snoRNA genes. (a) Models IV and V with the corresponding examples from [20]. (b) Screenshots of UCSC Genome Browser for the loci in panel (a) demonstrating that all snoRNA genes are localized within introns of host genes (EST track). Genomic coordinates for the March 2006 human reference sequence (NCBI Build 36.1) are given.

sequence with intact functional regions and seven SNORD115 genes. However, there are no ESTs confirming independent transcription of these genes, whereas for all independently transcribed human snoRNAs ESTs marking their transcription can be found.

Thus, all examples of snoRNA independent transcription presented by Li et al. (possibly, excluding SNORA26-like sequence and SNORD115 genes) are inadequate.

Discussion

How many snoRNA genes are there?

Studies by Zhu and coworkers attracted our attention since their results were at variance with our data. The main contradiction was the estimated number of snoRNA genes in vertebrates. Our estimation of the number of mammalian C/D snoRNA genes [16] agrees with the data obtained by other groups: the total number of mammalian snoRNA genes known to date does not exceed ~450 (review [28] and references therein). In addition, we have shown a lower number of C/D snoRNA genes guiding rRNA modifications in mammals relative to other vertebrate classes [16]. Conversely, Zhang et al. stated that the number of mammalian snoRNA genes sharply increased to ~1000 compared to other vertebrate classes [18]. Here we demonstrated inadequacy of their techniques, which invalidates their conclusions. In particular, they considered numerous pseudogenes as snoRNA genes in mammals and failed to detect many snoRNA genes in other vertebrate classes.

Northern hybridization has its limitations when used for detection of homologous ncRNAs in vertebrates

Possible existence of species-specific ncRNAs is extremely interesting, and it is being explored by many groups. Zhang et al. reported numerous lineage-specific and species-specific snoRNAs in chicken [19] and in rhesus monkey [18]. Here we demonstrated that their conclusions were based on a systemic error: Zhang et al. detected snoRNA homologs in vertebrate species using a probe for snoRNA of another vertebrate species, while the sequence identity of such homologs can go below

60% (Table 1). Under these conditions, standard Northern hybridization technique can not be used for homologs detection.

Using automatically generated ncRNA databases alone can lead to erroneous conclusions

While application of genomic and EST sequence collections has become routine in bioinformatic studies, using automatic annotations of genes, especially ncRNA genes, requires great caution. For instance, ENSEMBL ncRNA annotations based on the Rfam data are excellent landmarks for genome researchers. However, the rates of false positives and missed genes in these annotations, at least in snoRNA annotations, make their application unacceptable for studies specifically designed to identify new ncRNA genes. For example, Rfam makes no distinction between snoRNA genes and pseudogenes, but Zhang et al. considered all annotated snoRNA sequences as snoRNA genes, which led them to erroneous conclusions [18,20]. In addition, existing automatically generated databases still do not include all ncRNA homologs in different species. Therefore, special studies are needed to prevent underestimation of ncRNA number. E.g., Rfam lacks many snoRNA sequences presented here (Additional file 4) or available in the snoRNABase [3]. Zhang et al. made no attempt to overcome this problem, and, as a result, missed many snoRNA genes in different vertebrates. Thus, relying only on automatic annotations can lead to erroneous conclusions. Actually, most researchers pursue their own way through the genomic thicket to succeed in snoRNA studies [25,32-34].

We especially focused on this issue since at least one more publication reported questionable conclusions concerning vertebrate snoRNAs based on the Rfam and ENSEMBL annotations as well as multispecies whole-genome alignments [35]. Again, the fact that snoRNA genes and pseudogenes are not distinguished in the Rfam entries was not taken into account.

Names of snoRNA homologs need unification

Lots of snoRNAs have been described in different vertebrates to date, which necessitates the unification of their nomenclature. Zhang et al. gave a new name to each chicken homolog of human snoRNA [19]. This practice is not exclusive to Zhang et al. but is common in almost all publications describing snoRNAs in vertebrates apart from human. This was justified during the period of time when novel snoRNAs rather than homologs of known ones were being identified (e.g., [23]). Presently, a convenient nomenclature has been developed for human snoRNAs [10], and identification of novel snoRNAs has become extremely rare. In this context, giving new names to snoRNAs, whose homologs have been identified in other vertebrates, is highly confusing. It gives an erroneous impression that novel snoRNAs have actually been found and confuses the overall picture. For instance, a special investigation should be conducted to understand that the GGgCD37b snoRNA identified in chicken by Shao et al. [17] corresponds to Ggn109 found by Zhang et al. in chicken, too [19], and is a homolog of human SNORD38. The analysis of the whole set of data presented in these papers becomes hardly practicable. Finally, it is very hard to recognize the rare cases of a truly novel RNA identification. A positive practice in the field can be exemplified by the Rfam database specifying all homologs of human snoRNAs by the human RNA name. Since new publications describing snoRNAs in vertebrates can be expected, we propose to develop a nomenclature convention for the homologs. The human snoRNA names can be used with prefixes denoting the vertebrate species, e.g., mmusSNORD87 for the mouse homolog of human SNORD87. We propose to use four-letter prefixes to distinguish species such as *Mus musculus* (mmus) and *Microcebus murinus* (mmur).

Independent transcription of snoRNA genes is an intriguing possibility, but it needs strong support

Recent data indicate that many miRNA genes located within introns of host genes have their own promoters [36]. This interesting and unexpected finding inspires one to test a similar pattern in snoRNAs, nearly all of which are encoded within introns in vertebrates. Noteworthily, no experimental data supporting the hypothesis of intronic snoRNAs transcription from their own promoters are available to date. At the same time, their transcription within the host gene pre-mRNA from the host gene promoter has been well documented dozens of times (e.g., review [29] and references therein). Thus, the idea of transcription of intronic snoRNAs from their own promoters is at variance with our current knowledge about their expression, and identification of such promoters should have solid experimental support. Preliminary bioinformatic analysis can be beneficial, but it should be adequate and thorough, which was not the case with Li et al. [20].

Erroneous data begin to shape our view of ncRNAs

Currently, discovery of the species-specific ncRNAs is generally anticipated that may lead to less critical peer reviewing of publications reporting such RNAs. Here we show that the result can be harmful to the field. Even more importantly, such publications began to misshape our understanding of ncRNAs: one of the papers criticized here [18] has already been cited in a recent review [37].

Vertebrate genomes may actually contain many not yet identified snoRNAs. This idea is supported by the data from several groups [32,33,38]. However, publications like the ones considered here only add confusion

to the problem rather than contribute to the solution. Thus, it is very important to prevent a false start in this exciting field.

Methods

Homologs of human C/D box snoRNA genes in vertebrate genomes were searched as follows. First, homologs of human host genes were found in vertebrate genomes using the Comparative Genomics panel of UCSC Genome Browser at http://genome.ucsc.edu[39]. Then, the introns of the host genes were manually searched for the presence of snoRNA genes. If unsuccessful, snoRNA sequences were searched by WU-BLAST 2.0 http://www.ensembl.org/Multi/blastview with increased sensitivity parameters: high sensitivity (search for distant homologies) was chosen; W (word size for seeding alignments) = 3 and Q (cost of first gap character) = 1 were set. The intronic location of the search hits was checked using the mRNA and EST databases integrated into the UCSC Genome Browser. The hits with intact C, D/D' boxes, and the antisense element, flanked by short inverted repeats and located within introns of host genes were considered as snoRNA genes. Finally, extra copies of snoRNA genes were searched in the host gene introns.

NcRNAs discussed in [18-20] were analyzed using the UCSC Genome Browser and snoRNABase and Rfam databases [3,21]. Pairwise and multiple alignments were generated by Clustal V and Clustal W [40,41]. RNA secondary structures were analyzed using the mfold program [42,43].

Conclusions

Several recent publications reported numerous lineage-specific snoRNAs in vertebrates. However, the myriads of novel snoRNAs are just a mirage. The approaches used allowed no identification of human homologs of these "new" RNA species. Despite substantial sequence variation in snoRNA homologs in different vertebrates, they can be easily identified by the same antisense elements. The conclusion of elevated numbers of snoRNA genes in mammalian genomes relative to other vertebrates also proved erroneous, since no distinction was made between snoRNA genes and pseudogenes and no thorough analysis of recently sequenced genomes of non-mammalian vertebrates was conducted. The reported evidence for the transcription of many snoRNA genes from their own promoters is inconclusive.

Additional material

Additional file 1: NcRNAs whose expression has not been detected by Zhang et al. [**18**]**by Nothern hybridization in chicken, mouse, and human but was detected previously by other authors as well as by Zhang et al.**[**19**]. The order of RNAs is as in Table one from Zhang et al. [18].

Additional file 2: Controversial results of ncRNA detections in chicken and rhesus monkey (14 extra examples). Hybridization of RNA isolated from different tissues of rhesus monkey, chicken, human, and mouse with rhesus snoRNA probes (left panel; from [18]) and with chicken snoRNA probes (right panel; from [19]). The same RNAs are shown side-by-side. Chicken ncRNAs were cloned by Zhang et al. but not identified as homologs of human snoRNAs [19] (shown on the right). The same RNAs are presented in Table 2.

Additional file 3: The majority of chicken ncRNAs cloned and presented as novel RNAs by Zhang at al. [19] are homologs of ncRNAs described previously. Alignments of chicken ncRNAs with the homologs in human or sometimes other vertebrates are shown. GGN sequences are from Zhang et al. [19]. Vault RNA sequence corresponds to the GenBank AF045143 sequence. Other sequences are from snoRNABase [3] and Additional file 4 in this paper. C, D/D', H, ACA, and CAB boxes are underlined; antisense elements are boxed; sequence numbering corresponds to human rRNAs in snoRNABase. In C/D snoRNAs, the nucleotide complementary to the modification site is indicated by the red arrowhead. For the vault RNAs, the secondary structures predicted by mfold [42,43] are shown. The order of ncRNAs is as in Table 2. The SNORD102B transcript has a longer antisense element, and thus can guide the modification of the rRNA nucleotide adjacent to that modified by SNORD102A (marked with black and red arrowheads, respectively) [16].

Additional file 4: Nucleotide sequences of C/D box snoRNA genes in different vertebrate species. Boxes C, D, and D' are shown in gray, and sequences of the antisense elements are highlighted in yellow. The G-T complementarity in the antisense elements or terminal stems is indicated in olive. The 5' and 3' terminal complementary regions forming the stem in snoRNAs are shown in blue. Species-specific complementary substitutions in the antisense elements are marked in pink. Pseudogenes are indicated by Ψ. SNORD115 gene clusters are not listed. They have been found only in eutherian mammals and are available in snoRNABase [3] and UCSC Genome Browser. The following genome assemblies were used: human, March 2006, NCBI Build 36.1; mouse, July 2007, NCBI Build 37; rat, November 2004, version 3.4; dog, May 2005, whole genome shotgun assembly v2.0, cow, October 2007, Baylor release Btau_4.0; horse, January 2007, UCSC version equCab1; opossum, January 2006, monDom4; platypus, March 2007, the v5.0.1 draft assembly; chicken, May 2006, galGal3 version 2.1 draft assembly; lizard, February 2007, Broad Institute AnoCar 1.0; frog, August 2005, whole genome shotgun assembly version 4.1; zebrafish, July 2007, Zv7 assembly; fugu, October 2004, v4.0 whole genome shotgun assembly; tetraodon, February 2004, V7 assembly; stickleback, February 2006, v 1.0 draft assembly; medaka, October 2005, v 1.0 draft assembly.

Additional file 5: SnoRNA genes not found in the genomes of studied species by Zhang et al. [18] but found in the same species by other researchers. Gene names are listed in the same order as in Figure three in [18].

Additional file 6: Nearly all examples of independent transcription of snoRNA genes in Li et al. [20] are erroneous. Screenshots of UCSC Genome Browser (March 2006, NCBI Build 36.1) and nucleotide sequence alignments of snoRNA genes and pseudogenes are shown. The antisense elements are boxed; H, ACA, C, and D/D' sequences are underlined. The nucleotides whose modification is guided by snoRNA are indicated in some cases. SnoRNA genes and pseudogenes (designated as pseudo or Ψ) are listed in the same order as in Tables three, four, and five of Li et al. [20]. The secondary structures were predicted by mfold [42,43].

Acknowledgements

The work was supported by the Molecular and Cellular Biology Program of the Russian Academy of Sciences and the Russian Foundation for Basic Research (project no. 11-04-00439-a).

Response to: **SNOntology: Myriads of Novel SnoRNAs or Just a Mirage?** By Dahai Zhu

Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, 5 Dong Dan San Tiao, 100005, Beijing, China
dhzhu@pumc.edu.cn, dhzhusara@gmail.com

The work presented by Makarova and Kramerov (M&K) examined our previous studies on chicken and monkey snoRNAs, as well as our work on snoRNA promoter analysis [18-20], and raises some questions. We appreciate the attention given to our work. However, although some of the points raised are reasonable, many of the conclusions are based on biased information, misinterpretation of our results, or analysis of inconsistent datasets.

First, many basic concepts on snoRNAs presented in the M&K manuscript are outdated. For example, in the background section, the authors claim that 'To date, ~200 RNAs of both groups have been described', but the reference cited was published in 2006. The current non-coding RNA collection (in Rfam, release version 10.0) includes 519 snoRNA families and a total of 108, 332 snoRNAs [44]. The authors state that "nearly all snoRNAs and scaRNAs genes in vertebrates are located within introns of other genes. In fact, there are only five exceptions". This point also serves as support for the criticisms on our analysis of independently transcribed snoRNAs. However, this statement must be updated, because the reported number of human intergenic snoRNAs has been far exceeded that given by the authors, and some are indeed independently transcribed, even if intronically encoded, as reviewed in [28]. The recently discovered regulatory functions of snoRNAs [45,46] are also overlooked.

The authors criticize our analysis of lineage- or species-specific snoRNAs, and give the following reasons. First, "all snoRNAs cloned from rhesus monkey have been previously found in human"; second, "the pattern of rRNA modifications as well as the set of snoRNAs guiding these modifications are conserved in vertebrates"; and third, "the failure to detect the expression of some snoRNAs is due to the sequence divergence among species". Our answers to these questions follow. In terms of the first statement, as we mentioned in our paper, we indeed identified homologous snoRNA genes or pseudogenes for all the rhesus monkey snoRNAs that we cloned. However, as the human snoRNAs used in our study, as well as those to which M&K refer [16], have been identified by both cloning and computational prediction methods, the presence of a monkey snoRNA homologous sequence in the human genome does not directly indicate that those snoRNAs are expressed in human cells. In terms of the second statement, we do not understand why functional conservation of rRNAs within a large family can be used to support the notion that lineage- or species-specific snoRNAs are absent, especially given the increasing body of evidence indicating the regulatory roles played by snoRNAs in humans [6,7]. In terms of the third statement, it is possible that the lack of detectable signals from some snoRNAs in the chicken is attributable to sequence divergence. However, we speculate that this may not be the major reason as we were able to obtain positive northern blot hybridization signals for some sequences with as low as 12% conservation, but failed to obtain signals for some sequences with 100% conservation. We plan to gather further experimental data using species-specific probes to update our conclusion.

We think that the authors' criticism of our 'novel' chicken ncRNA work is very misleading. In the cited report, we identified 125 chicken ncRNAs including 102 snoRNAs, using a direct cloning method. Compared with the chicken snoRNAs predicted by Rfam, we found 25 snoRNAs that were not reported in chicken, and termed these molecules "novel snoRNA candidates". We also mentioned that 12 of the novel snoRNA candidates that we cloned had also been independently identified by Qu's group [17]. Although the snoRNAs identified by us in chicken have homologs in other vertebrates (Supplemental File 1 of our original work), majority of them have very low levels of sequence similarity as compared to human snoRNAs.

When we conducted the analysis just mentioned, the human snoRNA homologs listed in Table two of M&K were not included in the ENSEMBL and Rfam datasets. Therefore, we could not find human homologs of those snoRNAs. Similarly, the snoRNA homologs listed in Figure six of M&K were also not included in the versions of the ENSEMBL datasets that we used for monkey snoRNA analysis, but are indeed included in the current release. As it is well-known that the human genome annotation is consistently being updated, we think it is inappropriate and misleading to compare results obtained using different datasets.

We admit that our snoRNA target prediction methods may not be perfect; we were aware of this possibility when we conducted our work, but no better snoRNA target prediction software was available at that time. Thus, in our paper, we reported only the comparative conservation of putative snoRNA target sites between human and rhesus monkey. To render comparisons consistent among snoRNAs, we did not refine our predictive results using known targets, because correction in one species may lead to biased results in the conservation analysis. We did emphasize that the target sites that we listed were all putative.

The authors question the accuracy of the numbers of snoRNAs in different species contained in the ENSEMBL and Rfam databases. They have designed a snoRNA prediction tool based on refined sequence similarity search and have identified 1, 352 C/D box snoRNAs in 16 vertebrate species (Additional File five of M&K). Based on that result, they claim that the copy number of C/D box snoRNA genes is lower in mammals than in other vertebrates. We have analyzed the 1, 352 C/D box snoRNAs used in their study (Table 3). To our surprise, only 20 human snoRNAs were included in the list, and the numbers of snoRNAs of other mammals were also very low. However, the current numbers of recorded human C/D box snoRNAs deposited in several major databases range between 230~460 (Table 4), and at least 270 such predictions are supported by EST evidence (Data not shown). Therefore, the number of snoRNAs predicted (by M&K) in vertebrate genomes is obviously far less than the numbers of known snoRNAs supported by experimental evidence.

The authors use SNORD87 as an example to demonstrate the presence of 'a trend towards low copy numbers of C/D snoRNA genes in placental mammals'. However, many opposing examples could be given. One such is the SNORD115 and SNORD116 C/D box snoRNA families which are absent in non-eutherian vertebrate genomes but present as 30~50 tandem repeat copies on human chromosome 15q11-13. Mutations in these snoRNA clusters have been shown to be the cause of autism spectrum disorder and Prader-Willi syndrome [47,48]. However, these clusters were omitted from the M&K analysis.

The authors suggest that the numbers of snoRNAs obtained in our analysis are overestimates, given that some mammalian snoRNAs may be pseudogenes. We mentioned the possible existence of pseudogenes in our original work. However, as we reported (Figure 4A & B of our original paper), the numbers of snoRNAs and snoRNA families can be seen to have increased during evolution even when only intronic snoRNAs are considered. In addition, the expansion of snoRNA pseudogenes could also be considered to reflect snoRNA duplication.

M&K also question our snoRNA promoter prediction results [20]. In that work, we integrated the manual snoRNA dataset of Dieci et al. [28] with the Ensembl dataset (Release 53) [49] to perform promoter predictions for

Table 3 Summary of C/D box snoRNA numbers predicted by M&K in 16 vetebrate genomes (data from additional file five of M&K)

Species	Predicted snoRNA number
stickle-back	1
horse	2
medaka	9
human	20
cow	27
rat	27
fugu	53
dog	64
tetraodon	82
chicken	118
lizard	129
mouse	143
opossum	156
platypus	166
zebrafish	167
frog	188
TOTAL	1352

Table 4 Numbers of C/D box snoRNAs in human genome reported by different groups

Data resourse	Number of C/D box snoRNA	Released year	References
snoRNA-LBME-db	269	2006	[3]
The HUGO Gene Nomenclature Committee	272	2011	[53]
Rfam (release 10.0)	223	2010	[44]
ENSEMBL (release 63)	460 (593)*	2011	[54]
ENSEMBL (release 50)**	387 (502)*	2008	[27]
Reported by M&K	141	2009	[16]
Reported by M&K	20	2011	Additional file four of M&K

* Numbers of C/D box snoRNAs excluding U3 and U13 are given. Copy numbers of those two snoRNA families are shown in brackets.

** The version of ENSEMBL database used in our previous study [18].

human snoRNAs. As a result, we proposed five transcriptional models for human snoRNAs. M&K challenge our models II and III by arguing that several snoRNA loci with putative independent promoters reported in our study might be pseudogenes because of the presence of short sequence deletions or sequence variations. However, their claim of SNORD3 as a pseudogene for the lack of 100% sequence conservation at functional regions is not convincing. As shown in our earlier work [20], the detected DNase I-hypersensitive sites and the Pol II binding site are all located within 500 bp of the predicted TSS of SNORD3, strongly supporting the idea that the SNORD3 locus is transcriptionally active.

Although snoRNAs function mainly as modulators of ribosomal RNAs, snoRNAs may have broader functions than previously appreciated. One possibility is that snoRNAs may serve as precursors of microRNAs and may possess microRNA-like functions [46,50]. Some snoRNAs are known to regulate alternative splicing of their target mRNAs [45,51,52]. Therefore, genomic loci harboring snoRNA variants might have non-canonical functions different from those of typical snoRNAs, although transcriptional activity must be experimentally proven. Moreover, active transcription of pseudogenes actually plays an important role in gene expansion during genome evolution. Overall, it is inadequate and illogical for M&K to point to potential pseudogenes to challenge snoRNA transcription models II and III.

M&K argue that some intergenic snoRNA examples used by us in our snoRNA promoter study were indeed of intronic origin. As illustrated in Figure Four b of M&K, SNORD60 lie in the intronic region of some ESTs, however, many unspliced ESTs were omitted in their figure (Figure 8). Similar cases are SNORD104 and SNORA76 shown in additional file six of M&K. Previous studies have demonstrated that SNORD104 and SNORA76 are independently transcribed [28], which is in agreement with our results. For another example SNORD93, it is located within an intergenic region according to the RefSeq and UCSC gene models (hg18) used in our

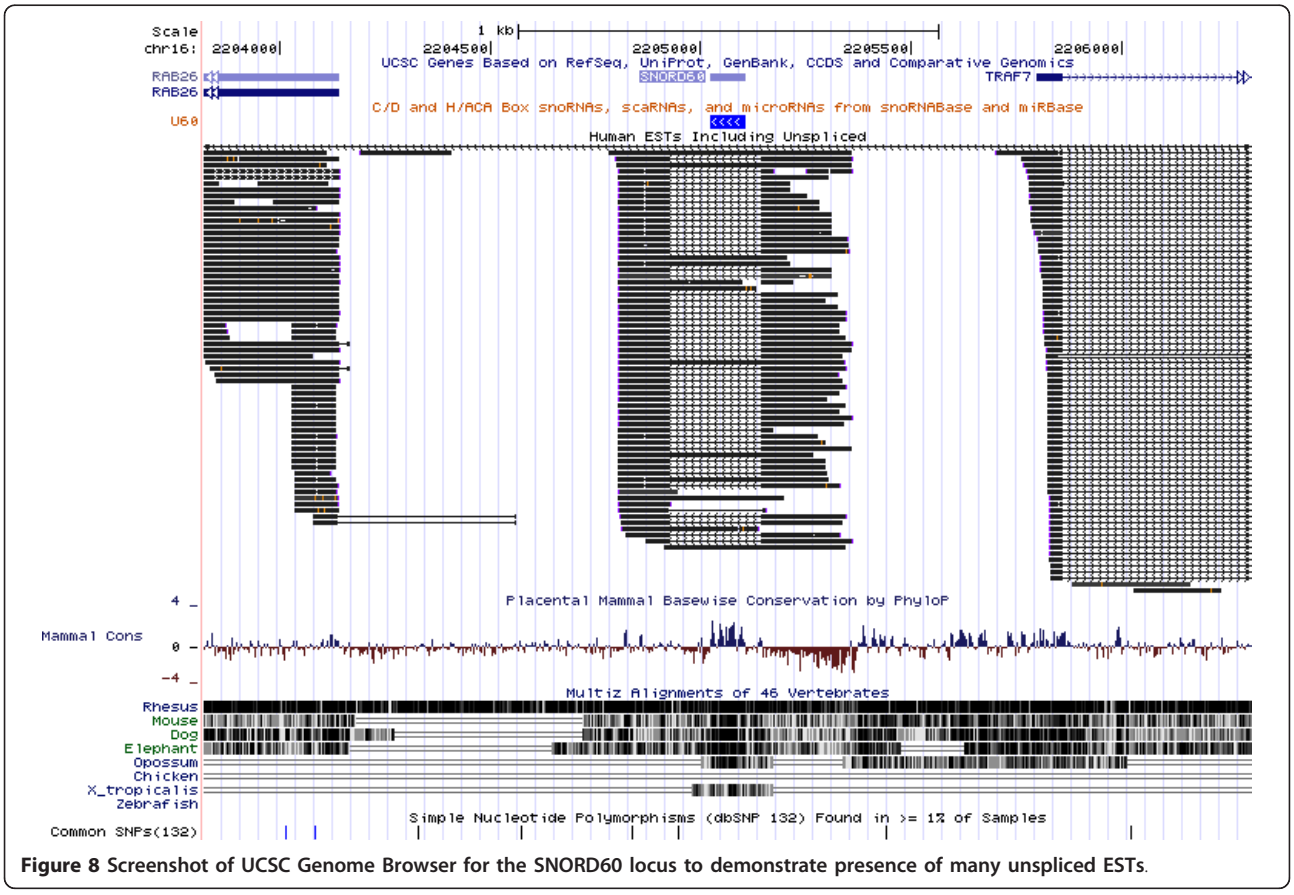


Figure 8 Screenshot of UCSC Genome Browser for the SNORD60 locus to demonstrate presence of many unspliced ESTs.

previous work [20], but was reannotated as an intronic snoRNA in the hg19 release. Such information update should not be classified as analysis errors. In summary, because of the nature of computational prediction work, it is very unlikely that bioinformatic analysis data will ever be error-free. We welcome updated analysis of our data using improved methods and enriched reference sources. However, the work presented in the report by M&K is characterized by the drawing of conclusions based on biased information, and misinterpretation of both their own and our results, which may add more confusions to the field.

Authors' contributions

JM and DK conceived the study. JM carried out all analyses and drafted the manuscript. Both authors read and approved the final manuscript.

Received: 23 March 2011 Accepted: 3 November 2011
Published: 3 November 2011

References

1. Ganot P, Bortolin ML, Kiss T: **Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs.** *Cell* 1997, **89**:799-809.
2. Kiss-Laszlo Z, Henry Y, Bachellerie JP, Caizergues-Ferrer M, Kiss T: **Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs.** *Cell* 1996, **85**:1077-1088.
3. Lestrade L, Weber MJ: **snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs.** *Nucleic Acids Res* 2006, **34**:D158-162.
4. 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**:3747-3757.
5. Kiss-Laszlo Z, Henry Y, Kiss T: **Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA.** *EMBO J* 1998, **17**:797-807.
6. Filipowicz W, Pogacic V: **Biogenesis of small nucleolar ribonucleoproteins.** *Curr Opin Cell Biol* 2002, **14**:319-327.
7. Makarova Iu A, Kramerov DA: **Small nucleolar RNAs.** *Mol Biol (Mosk)* 2007, **41**:246-259.
8. Reichow SL, Hamma T, Ferre-D'Amare AR, Varani G: **The structure and function of small nucleolar ribonucleoproteins.** *Nucleic Acids Res* 2007, **35**:1452-1464.
9. Darzacq X, Jady BE, Verheggen C, Kiss AM, Bertrand E, Kiss T: **Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs.** *EMBO J* 2002, **21**:2746-2756.
10. Bruford EA, Lush MJ, Wright MW, Sneddon TP, Povey S, Birney E: **The HGNC Database in 2008: a resource for the human genome.** *Nucleic Acids Res* 2008, **36**:D445-448.
11. Makarova Iu A, Kramerov DA: **Small nucleolar RNA genes.** *Genetika* 2007, **43**:149-158.
12. Esguerra J, Warringer J, Blomberg A: **Functional importance of individual rRNA 2'-O-ribose methylations revealed by high-resolution phenotyping.** *RNA* 2008, **14**:649-656.
13. Baxter-Roshek JL, Petrov AN, Dinman JD: **Optimization of ribosome structure and function by rRNA base modification.** *PLoS One* 2007, **2**:e174.
14. Maden BE: **The numerous modified nucleotides in eukaryotic ribosomal RNA.** *Prog Nucleic Acid Res Mol Biol* 1990, **39**:241-303.
15. Schmitz J, Zemann A, Churakov G, Kuhl H, Grutzner F, Reinhardt R, Brosius J: **Retroposed SNOfall–a mammalian-wide comparison of platypus snoRNAs.** *Genome Res* 2008, **18**:1005-1010.
16. Makarova JA, Kramerov DA: **Analysis of C/D box snoRNA genes in vertebrates: The number of copies decreases in placental mammals.** *Genomics* 2009, **94**:11-19.
17. Shao P, Yang JH, Zhou H, Guan DG, Qu LH: **Genome-wide analysis of chicken snoRNAs provides unique implications for the evolution of vertebrate snoRNAs.** *BMC Genomics* 2009, **10**:86.
18. Zhang Y, Liu J, Jia C, Li T, Wu R, Wang J, Chen Y, Zou X, Chen R, Wang XJ, Zhu D: **Systematic identification and evolutionary features of rhesus monkey small nucleolar RNAs.** *BMC Genomics* 2010, **11**:61.
19. Zhang Y, Wang J, Huang S, Zhu X, Liu J, Yang N, Song D, Wu R, Deng W, Skogerbo G, Wang XJ, Chen R, Zhu D: **Systematic identification and characterization of chicken (Gallus gallus) ncRNAs.** *Nucleic Acids Res* 2009, **37**:6562-6574.
20. Li T, Zhou X, Wang X, Zhu D, Zhang Y: **Identification and characterization of human snoRNA core promoters.** *Genomics* 2010, **96**:50-56.
21. Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR, Bateman A: **Rfam: updates to the RNA families database.** *Nucleic Acids Res* 2009, **37**:D136-140.
22. Gogolevskaya IK, Makarova JA, Gause LN, Kulichkova VA, Konstantinova IM, Kramerov DA: **U87 RNA, a novel C/D box small nucleolar RNA from mammalian cells.** *Gene* 2002, **292**:199-204.
23. Huttenhofer A, Kiefmann M, Meier-Ewert S, O'Brien J, Lehrach H, Bachellerie JP, Brosius J: **RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse.** *EMBO J* 2001, **20**:2943-2953.
24. Qu LH, Henry Y, Nicoloso M, Michot B, Azum MC, Renalier MH, Caizergues-Ferrer M, Bachellerie JP: **U24, a novel intron-encoded small nucleolar RNA with two 12 nt long, phylogenetically conserved complementarities to 28S rRNA.** *Nucleic Acids Res* 1995, **23**:2669-2676.
25. Schattner P, Barberan-Soler S, Lowe TM: **A computational screen for mammalian pseudouridylation guide H/ACA RNAs.** *RNA* 2006, **12**:15-25.
26. Szymanski M, Barciszewska MZ, Erdmann VA, Barciszewski J: **5S Ribosomal RNA Database.** *Nucleic Acids Res* 2002, **30**:176-178.
27. Flicek P, Aken BL, Ballester B, Beal K, Bragin E, Brent S, Chen Y, Clapham P, Coates G, Fairley S, Fitzgerald S, Fernandez-Banet J, Gordon L, Graf S, Haider S, Hammond M, Howe K, Jenkinson A, Johnson N, Kahari A, Keefe D, Keenan S, Kinsella R, Kokocinski F, Koscielny G, Kulesha E, Lawson D, Longden I, Massingham T, McLaren W, Megy K, Overduin B, Pritchard B, Rios D, Ruffier M, Schuster M, Slater G, Smedley D, Spudich G, Tang YA, Trevanion S, Vilella A, Vogel J, White S, Wilder SP, Zadissa A, Birney E, Cunningham F, Dunham I, Durbin R, Fernandez-Suarez XM, Herrero J, Hubbard TJ, Parker A, Proctor G, Smith J, Searle SM: **Ensembl's 10th year.** *Nucleic Acids Res* 2010, **38**:D557-562.
28. Dieci G, Preti M, Montanini B: **Eukaryotic snoRNAs: a paradigm for gene expression flexibility.** *Genomics* 2009, **94**:83-88.
29. Richard P, Kiss T: **Integrating snoRNP assembly with mRNA biogenesis.** *EMBO Rep* 2006, **7**:590-592.
30. Luo Y, Li S: **Genome-wide analyses of retrogenes derived from the human box H/ACA snoRNAs.** *Nucleic Acids Res* 2007, **35**:559-571.
31. Wang X, Xuan Z, Zhao X, Li Y, Zhang MQ: **High-resolution human core-promoter prediction with CoreBoost_HM.** *Genome Res* 2009, **19**:266-275.
32. Fedorov A, Stombaugh J, Harr MW, Yu S, Nasalean L, Shepelev V: **Computer identification of snoRNA genes using a Mammalian Orthologous Intron Database.** *Nucleic Acids Res* 2005, **33**:4578-4583.
33. Yang JH, Zhang XC, Huang ZP, Zhou H, Huang MB, Zhang S, Chen YQ, Qu LH: **snoSeeker: an advanced computational package for screening of guide and orphan snoRNA genes in the human genome.** *Nucleic Acids Res* 2006, **34**:5112-5123.
34. Hertel J, Hofacker IL, Stadler PF: **SnoReport: computational identification of snoRNAs with unknown targets.** *Bioinformatics* 2008, **24**:158-164.
35. Hoeppner MP, White S, Jeffares DC, Poole AM: **Evolutionarily stable association of intronic snoRNAs and microRNAs with their host genes.** *Genome Biol Evol* 2009, **1**:420-428.
36. Monteys AM, Spengler RM, Wan J, Tecedor L, Lennox KA, Xing Y, Davidson BL: **Structure and activity of putative intronic miRNA promoters.** *RNA* 2010, **16**:495-505.
37. Gardner PP, Bateman A, Poole AM: **SnoPatrol: how many snoRNA genes are there?** *J Biol* 2010, **9**:4.
38. Washietl S, Hofacker IL, Lukasser M, Huttenhofer A, Stadler PF: **Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome.** *Nat Biotechnol* 2005, **23**:1383-1390.
39. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: **The human genome browser at UCSC.** *Genome Res* 2002, **12**:996-1006.
40. Higgins DG, Bleasby AJ, Fuchs R: **CLUSTAL V: improved software for multiple sequence alignment.** *Comput Appl Biosci* 1992, **8**:189-191.
41. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**:4673-4680.
42. Zuker M: **Mfold web server for nucleic acid folding and hybridization prediction.** *Nucleic Acids Res* 2003, **31**:3406-3415.

43. Mathews DH, Sabina J, Zuker M, Turner DH: **Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure.** *J Mol Biol* 1999, **288**:911-940.
44. Gardner PP, Daub J, Tate J, Moore BL, Osuch IH, Griffiths-Jones S, Finn RD, Nawrocki EP, Kolbe DL, Eddy SR, *et al*: **Rfam: Wikipedia, clans and the "decimal" release.** *Nucleic Acids Res* 2011, **39(Database)**:D141-145.
45. Kishore S, Stamm S: **The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C.** *Science* 2006, **311(5758)**:230-232.
46. Ender C, Krek A, Friedlander MR, Beitzinger M, Weinmann L, Chen W, Pfeffer S, Rajewsky N, Meister G: **A human snoRNA with microRNA-like functions.** *Mol Cell* 2008, **32(4)**:519-528.
47. Bolton PF, Veltman MW, Weisblatt E, Holmes JR, Thomas NS, Youings SA, Thompson RJ, Roberts SE, Dennis NR, Browne CE, *et al*: **Chromosome 15q11-13 abnormalities and other medical conditions in individuals with autism spectrum disorders.** *Psychiatr Genet* 2004, **14(3)**:131-137.
48. Cavaille J, Buiting K, Kiefmann M, Lalande M, Brannan CI, Horsthemke B, Bachellerie JP, Brosius J, Huttenhofer A: **Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization.** *Proc Natl Acad Sci USA* 2000, **97(26)**:14311-14316.
49. Hubbard TJ, Aken BL, Ayling S, Ballester B, Beal K, Bragin E, Brent S, Chen Y, Clapham P, Clarke L, *et al*: **Ensembl 2009.** *Nucleic Acids Res* 2009, **37(Database)**:D690-697.
50. Saraiya AA, Wang CC: **snoRNA, a novel precursor of microRNA in Giardia lamblia.** *PLoS Pathog* 2008, **4(11)**:e1000224.
51. Kishore S, Khanna A, Zhang Z, Hui J, Balwierz PJ, Stefan M, Beach C, Nicholls RD, Zavolan M, Stamm S: **The snoRNA MBII-52 (SNORD 115) is processed into smaller RNAs and regulates alternative splicing.** *Hum Mol Genet* 2010, **19(7)**:1153-1164.
52. Bazeley PS, Shepelev V, Talebizadeh Z, Butler MG, Fedorova L, Filatov V, Fedorov A: **snoTARGET shows that human orphan snoRNA targets locate close to alternative splice junctions.** *Gene* 2008, **408(1-2)**:172-179.
53. Eyre TA, Ducluzeau F, Sneddon TP, Povey S, Bruford EA, Lush MJ: **The HUGO Gene Nomenclature Database, 2006 updates.** *Nucleic Acids Res* 2006, **34(Database)**:D319-321.
54. Flicek P, Amode MR, Barrell D, Beal K, Brent S, Chen Y, Clapham P, Coates G, Fairley S, Fitzgerald S, *et al*: **Ensembl 2011.** *Nucleic Acids Res* **39(Database)**: D800-806.

doi:10.1186/1471-2164-12-543
Cite this article as: Makarova and Kramerov: **SNOntology: Myriads of novel snornas or just a mirage?** *BMC Genomics* 2011 **12**:543.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central