UNIVERSITÀ DI PARMA

ARCHIVIO DELLA RICERCA

University of Parma Research Repository

Eukaryotic snoRNAs: a paradigm for gene expression flexibility.

This is the peer reviewd version of the followng article:

*Original*
Eukaryotic snoRNAs: a paradigm for gene expression flexibility / DIECI G.; PRETI M; MONTANINI B. - In: GENOMICS. - ISSN 0888-7543. - 94(2009), pp. 83-88.

*Availability:*
This version is available at: 11381/2282968 since: 2017-03-29T14:55:52Z

*Publisher:*

*Published*
DOI:

*Terms of use:*
openAccess

Anyone can freely access the full text of works made available as "Open Access". Works made available

*Publisher copyright*

(Article begins on next page)

20 November 2022

ARTICLE IN PRESS

YGENO-08112; No. of pages: 6; 4C:

Genomics xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Minireview

# Eukaryotic snoRNAs: A paradigm for gene expression flexibility

Giorgio Dieci *, Milena Preti [1], Barbara Montanini

*Dipartimento di Biochimica e Biologia Molecolare, Università degli Studi di Parma, Viale G.P. Usberti 23/A, 43100 Parma, Italy*

ARTICLE INFO

*Article history:*
Received 13 March 2009
Accepted 7 May 2009
Available online xxxx

*Keywords:*
Small nucleolar RNAs
snoRNA genes
ncRNAs
Intron
Promoter
Gene expression
Transcription

ABSTRACT

Small nucleolar RNAs (snoRNAs) are one of the most ancient and numerous families of non-protein-coding RNAs (ncRNAs). The main function of snoRNAs – to guide site-specific rRNA modification – is the same in Archaea and all eukaryotic lineages. In contrast, as revealed by recent genomic and RNomic studies, their genomic organization and expression strategies are the most varied. Seemingly snoRNA coding units have adopted, in the course of evolution, all the possible ways of being transcribed, thus providing a unique paradigm of gene expression flexibility. By focusing on representative fungal, plant and animal genomes, we review here all the documented types of snoRNA gene organization and expression, and we provide a comprehensive account of snoRNA expressional freedom by precisely estimating the frequency, in each genome, of each type of genomic organization. We finally discuss the relevance of snoRNA genomic studies for our general understanding of ncRNA family evolution and expression in eukaryotes.

© 2009 Published by Elsevier Inc.

## Introduction

Small nucleolar (sno) RNAs are a group of untranslated RNA molecules of variable length (80 to 1000 nt in yeast) mostly required for rRNA maturation. SnoRNAs can be divided into two classes which possess distinctive, evolutionarily conserved sequence elements: the box C/D and box H/ACA snoRNAs, that guide by base pairing respectively 2′-O-ribose methylation and pseudouridylation of specific rRNA nucleotides [1,2]. A few snoRNAs from both classes do not function as guide RNAs, but are required for pre-rRNA endonucleolytic processing, a process also involving an abundant and evolutionarily conserved snoRNA that cannot be included in either of the two above classes: the ribonuclease MRP RNA. Furthermore, there is mounting evidence that guide snoRNA targets are not limited to rRNA [3,4]. Apart from conserved sequence signatures, each class of snoRNAs displays a characteristic secondary structure, and interacts with a distinct core set of highly conserved proteins to form the well defined C/D and H/ACA snoRNPs [4]. In contrast, highly variable features of snoRNAs in the different eukaryotes are their genomic location and mode of transcription. The recent explosion of RNomic studies in the most representative eukaryotic systems [5,6] is revealing a striking evolutionary adaptability of snoRNA gene organization. Excellent reviews have accompanied during the last few years the exploration of snoRNA continents in the eukaryotic genomes. Such reviews focus either on the structure, function and targets of the snoRNPs [1–3,7] or on snoRNA expression and function in individual eukaryotic lineages: plant [8], *Drosophila* [9], trypanosomatids [10], and humans [11,12]. The scope of this review is to attempt for the first time a comprehensive account of snoRNA gene expression flexibility, as it unfolds from a comparative inspection of snoRNA gene complements of prototypical eukaryotic genomes.

## Diversity of snoRNA gene location and expression strategies

Many different types of organization of snoRNA coding units have been documented in eukaryotes, each corresponding to a specific mode of transcription. In this section, we will describe the salient features of each type of organization, aside from their frequency in the different genomes.

As outlined in Fig. 1, snoRNA gene organization ranges from independently transcribed genes, endowed with their own promoter elements, to intronic coding units lacking an independent promoter. In both yeast and animals, processing of intron-encoded snoRNAs is largely splicing-dependent; in contrast, the production of plant snoRNAs from introns seems to rely on a splicing-independent process [13]. Moreover, in both contexts (intergenic or intronic), genes can be either single or part of clusters. In the latter case, the generation of individual snoRNAs involves the enzymatic processing of polycistronic precursor RNAs. Such a processing, at least in yeast, appears to involve the same combination of endo- and exoribonucleases required for the maturation of monocistronic pre-snoRNAs [14–16].

### *SnoRNA genes with independent promoters*

All eukaryotic genomes contain a number of snoRNA genes endowed with independent promoters. In yeast and plants, such

* Corresponding author. Fax: +39 0521 905151.
*E-mail address:* giorgio.dieci@unipr.it (G. Dieci).
[1] Present address: Laboratoire de Biologie Moléculaire Eucaryote, Université de Toulouse et CNRS, 118, route de Narbonne, 31062 Toulouse, France.

0888-7543/$ – see front matter © 2009 Published by Elsevier Inc.
doi:10.1016/j.ygeno.2009.05.002

ARTICLE IN PRESS



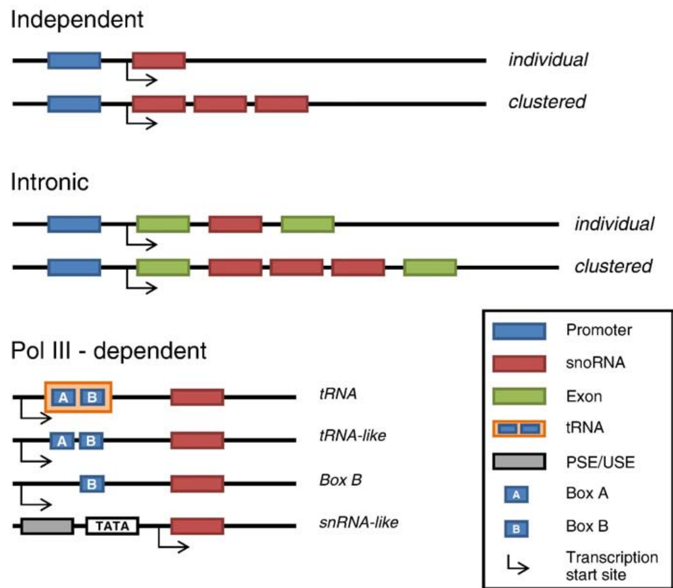


**Fig. 1.** Genomic organization of snoRNA coding units. Schematic representation of the different types of genomic location of snoRNA genes. The snoRNA coding units endowed with independent promoters (top) and those located within introns (middle) are transcribed by RNA polymerase II. Frequently, neighbouring introns of the same host gene contains snoRNA coding units with a one-gene-per-intron distribution. In such cases, the snoRNA coding units have been considered as "intronic individual" (Table 1), even though several different snoRNAs can originate from the same precursor transcript.

promoters direct the synthesis of both monocistronic snoRNA transcripts and of polycistronic precursors. Most independent snoRNA gene promoters are served by RNA polymerase (Pol) II, but a few possess control elements recognized by the Pol III transcription machinery. The somewhat interchangeable character of Pol II and Pol III in snoRNA gene transcription across eukaryotes has long been known from studies of the genes coding for the U3 snoRNA, involved in an essential endonucleolytic step of pre-rRNA processing. In both vertebrates and invertebrates, this gene is transcribed by Pol II from an upstream core promoter containing the proximal sequence element (PSE) typical of small nuclear (sn) RNA genes [17]. Also in yeast, the *U3* gene is transcribed by Pol II, from TATA-containing promoters potentiated by farther upstream elements ([18]; M.P. and G.D., unpublished observations), whereas in plants, and even in the unicellular alga *Chlamydomonas reinhardtii*, it is transcribed by Pol III through an upstream sequence element (USE, functionally analogous to the metazoan PSE), followed by a TATA-like element [19,20]. Another abundant, non-guide snoRNA whose independent promoter has been characterized in several eukaryotes is the MRP RNA, whose gene is transcribed by Pol III (from PSE/USE-TATA promoters) in both metazoa and plants, but not in budding yeast [21,22]. Much less is known about the promoter organization of autonomously transcribed genes coding for guide snoRNAs. They generally appear to be transcribed by Pol II from upstream promoters [23–25], but the nature of these promoters is largely uncharacterized. Some information is available in the case of budding yeast, whose snoRNA gene promoter regions tend to contain TATA boxes and A/T-rich elements, as well as binding sites for general regulatory factors, such as Rap1p and Abf1p [16]. Yeast *SNR52* is an exception, as it is transcribed by Pol III through control regions (*A* and *B box*, typical of tRNA genes) that are located within a transcribed leader sequence [26]. The utilization of upstream Pol III-specific *box A* and *box B* to drive transcription of guide snoRNA genes has also been documented in the land plants *A. thaliana* and *O. sativa*, where a few guide snoRNAs are synthesized as dicistronic tRNA–snoRNA precursors [27], and in the nematode *C. elegans* [24]. At least two *Drosophila* guide snoRNA genes are also independently transcribed by Pol III, possibly through the utilization of *box B* promoter elements [28]. In conclusion, it appears as if autonomous snoRNA gene transcription was achieved rather opportunistically during evolution, through the flexible exploitation of different types of specialized promoters.

### *Intronic snoRNA genes*

Intronic snoRNA coding units have been identified in all eukaryotic genomes. As illustrated by Fig. 1, they can be found either as individual units, following a "one-gene-per-intron" organization, or as clusters of two or more coding units located in the same intron. Such clusters, in turn, can either be made up of homologous snoRNA genes (homo-clusters), likely originating from local tandem duplications, or consist of heterologous snoRNA genes (heteroclusters) that can even contain together box C/D and H/ACA coding units [29]. Large intronic heteroclusters can be composed of duplicated smaller heteroclusters [8]. Frequently, individual snoRNA gene units are found within introns of non-protein-coding genes. Such a location was initially identified in mammals, later in *Drosophila* [9], and is characterized by the presence of several different snoRNA genes within consecutive introns of the same non-protein-coding transcription unit, with a "one-gene-per-intron" distribution [30]. Apparently such transcription units, also referred to as UHG (from the originally identified *U22 host* gene [31]),

**Table 1**
Organization of snoRNA genes in representative eukaryotic genomes[a].

| Organism[b] | snoRNAs | Genes | Independent | | Intronic | | Polymerase III[a] |
|---|---|---|---|---|---|---|---|
| | | | Individual | Clustered | Individual | Clustered | |
| *S. cerevisiae* | 75 | 76 (47 C/D; 29 H/ACA) | 50 (23 C/D; 27 H/ACA) | 17 (C/D) | 8 (6 C/D; 2 H/ACA) | 0 | 1 |
| *S. pombe* | 55[c] | 55 (32 C/D; 24 H/ACA) | 43 (20 C/D; 24 H/ACA) | 8 (C/D) | 4 (C/D) | 0 | 0 |
| *C. elegans* | 161[c] | 161 (96 C/D; 65 H/ACA) | 42 (33 C/D; 9 H/ACA)[d] | 0 | 119 (63 C/D; 56 H/ACA) | 0 | 71[e] |
| *D. melanogaster* | 131 | 227 (111 C/D; 116 H/ACA) | 8 (5 C/D; 3 H/ACA) | 0 | 135 (101 C/D; 34 H/ACA)[f] | 82 (5 C/D; 77 H/ACA) | 2 (H/ACA) |
| *H. sapiens* | 216[g] | 456 (257 C/D; 181 H/ACA) | 42 (15 C/D; 27 H/ACA) | 2 (1 C/D; 1 H/ACA) | 412 (259 C/D; 153 H/ACA) | 0 | 0 |
| *O. sativa* | 140 | 357 (345 C/D; 12 H/ACA) | 76 (C/D) | 174 (169[h] C/D; 5 H/ACA) | 0 | 104 (97 C/D; 7 H/ACA) | 3 (C/D) |
| *A. thaliana* | 155 | 246 (199 C/D; 47 H/ACA) | 57 (42 C/D; 15 H/ACA) | 146 (131 C/D; 15 H/ACA) | 23 (6 C/D; 17 H/ACA) | 6 (C/D) | 14 (C/D) |

[a] As a general note, the inventories of known snoRNA genes are likely to be incomplete at the current time (with the possible exception of yeast). The genes for MRP RNA are not computed in this table.
[b] SnoRNA data for the different organisms are based on the following references: *S. cerevisiae* [34]; *S. pombe* [35,62], *Schizosaccharomyces pombe* GeneDB; *C. elegans* [24,37–39,63,64]; *D. melanogaster* [28,30,65,66]; *H. sapiens* [11,44,45,67,68]; *O. sativa* [27,41,69–72]; *A. thaliana* [27,71,73–79].
[c] In the absence of evidence for significant snoRNA gene redundancy we report here the total number of snoRNA coding units.
[d] Of which 3 exon antisense.
[e] snoRNA coding units reported by [24] to have upstream promoter elements (UM1 or UM2) potentially recognized by Pol III. On the basis of their location, these units have been classified either as "independent" or as "intronic" in this table.
[f] 2 intronic genes are intron-antisense.
[g] At least 216.
[h] 10 snoRNA pseudogenes are not computed in this table.

Please cite this article as: G. Dieci, et al., Eukaryotic snoRNAs: A paradigm for gene expression flexibility, Genomics (2009), doi:10.1016/j.ygeno.2009.05.002

ARTICLE IN PRESS



are devoted to the production of snoRNAs. They could thus, in principle, be classified among the snoRNA genes served by dedicated promoters. Due to their particular location, however, we have preferred to count them among intronic snoRNA coding units in Table 1 (see below). Finally, of particular interest is the identification, in the genome of *C. elegans*, of intronic snoRNA loci showing signs of independent transcription (e.g. the presence of conserved upstream sequences, that in some cases resemble Pol III-specific control regions; [24]). A very similar observation has been made recently in a genome-wide search for human miRNA gene promoter signatures, showing that one third of human intronic miRNA display independent, largely Pol II-specific transcription initiation regions (yet some of them are occupied by Pol III *in vivo* and exhibit Pol III-specific promoter elements) [32].

## Organization of snoRNA genes in representative eukaryotic genomes

SnoRNA-based rRNA processing predates the separation of Archaea and Eukarya [33]. Over the course of evolution, snoRNA coding units spread in eukaryotic genomes through different routes, thus attaining composite and profoundly different organizations in present genomes. In this section, we will outline the salient features of snoRNA gene organization and expression as they emerge from both genomic and transcriptomic studies in model eukaryotes. Such features are comprehensively summarized in Table 1, that provides detailed information on the frequency, in each genome, of each type of genomic organization. Fig. 2 shows a graphical synthesis of such information, illustrating the distinctive features of snoRNA gene organization in the seven genomes analyzed. The data source for Table 1 and Fig. 2 is a set of more detailed, genome-specific tables (Tables S1 to S7) that can be found in Supplementary data online. It can be anticipated that the numbers of snoRNA genes identified in these genomes will increase in the future. We are confident, however, that the general conclusions made possible by the currently available snoRNA inventories will still be valid after their completion.

### *Yeast snoRNA genes*

In both the distantly related yeasts *S. cerevisiae* and *S. pombe*, the vast majority of snoRNA genes are monocistronic and served by independent promoters. Only eight intronic snoRNA coding units have been recognized in *S. cerevisiae* [34], and only four in *S. pombe* [35,36], in agreement with the scarcity of introns in the corresponding genomes. A few polycistronic clusters are found under the control of independent promoters (see [16] as an example), while all the intronic snoRNA units are individual. In *S. pombe* one of them, encoding snR80, is located within the intron of the independently transcribed *snR90* gene, thus providing a unique example of an ncRNA gene that encodes two different types of snoRNAs by both its exon and intron [35].

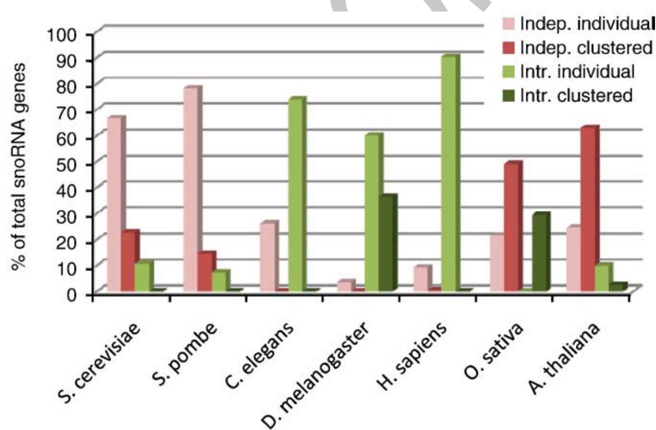


**Fig. 2.** Distinctive snoRNA gene organizations in eukaryotic genomes. The plot, based on Table 1 data, reports the frequency of occurrence of the different types of snoRNA gene organization in each of the genomes indicated on the *x* axis.

### *Nematode snoRNA genes*

The availability of comprehensive inventories of snoRNA genes in the nematode *Caenorhabditis elegans* allows to appreciate a dramatic increase, with respect to yeast, of the number of intron-located snoRNA coding units [24,37–39]. About 75% of *C. elegans* snoRNA genes are indeed embedded within introns of protein-coding genes, with a "one-gene-per-intron" distribution. However, as reminded above, some of them are likely to be transcribed independently from the host gene [24]. This possibility is especially significant in the case of a dozen snoRNA units that are located within introns, but with an antisense orientation (see Table S3 in Supplementary Material). Interestingly, and at variance with the other eukaryotic model genomes from yeast to man, no polycistronic snoRNA coding units (either intergenic or intronic) have been reported in the *C. elegans* genome (even though *Ce173.1*, *Ce173.2* and *Ce173.3* genes appear to be contained within the same intron; see Table S3).

### *Drosophila snoRNA genes*

The most evident feature of snoRNA gene organization in *Drosophila*, as compared to the one in *C. elegans*, is a strong tendency towards intronic integration of snoRNA coding units. Indeed, about 95% of fruit fly snoRNA genes are located within introns. Strikingly, however, 54 out of a total of 217 intronic snoRNA genes map within introns of 8 non-protein-coding host genes (dUHGs). In dUGH introns, the snoRNA coding units have a strictly "one-gene-per-intron" distribution, and almost all code for box C/D snoRNAs, while the snoRNA genes hosted by introns of protein-coding genes are often organized in clusters composed of isoforms of the same snoRNA genes, that prevalently code for box H/ACA snoRNAs [9,30,40]. Up to now, *Drosophila* appears to be unique in such a divergence in genomic organization and expression strategies of the two snoRNA classes [30].

### *Plant snoRNA genes*

A distinctive feature of plant snoRNA genes is their prevailing organization in polycistronic clusters. Clustering seems to be more pronounced in the model monocot *Oryza sativa*: here intronic snoRNA genes, representing ~30% of total snoRNA coding units, are all organized in polycistronic clusters, as are 70% of the snoRNAs transcribed from independent promoters [8,41]. (It should be considered, however, that the inventory of rice box H/ACA snoRNA genes is still largely incomplete.) The situation is different in the case of the model dicot, *Arabidopsis thaliana*. Here, similar to what happens in rice, 75% of independently transcribed snoRNAs derive from polycistronic clusters. At variance with rice, however, *A. thaliana* intronic snoRNA genes (representing ~15% of the total) are mostly unclustered (only 8 clustered out of a total of 33 intronic snoRNA genes; in rice, the 104 intronic clustered snoRNA genes represent the totality of known intronic snoRNA coding units; see Fig. 2). As in *Drosophila*, *C. elegans* and yeast, also in plants (both monocot and dicot) some snoRNA coding units have been adopted by the Pol III transcription apparatus, that transcribes them as dicistronic tRNA–snoRNA transcripts using the internal promoter of an upstream tRNA gene [27]. Pol III does not appear to transcribe snoRNA clusters, as expected on the basis of its proneness to termination at very simple signals ($T_n$ with $n \geq 4$) and thus its exclusive utilization for transcription of very short DNA tracts [42]. Another peculiar feature of snoRNA gene organization in plants is the presence of multiple genes coding

Please cite this article as: G. Dieci, et al., Eukaryotic snoRNAs: A paradigm for gene expression flexibility, Genomics (2009), doi:10.1016/j.ygeno.2009.05.002

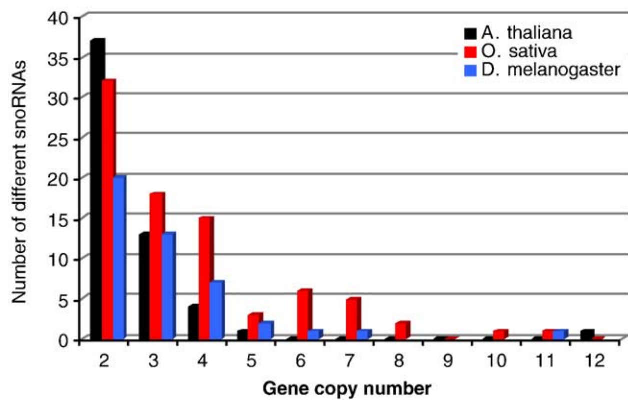


**Fig. 3.** Redundancy of snoRNA genes. The plot reports on the *y* axis the number of different snoRNA species characterized by the gene-copy numbers indicated on the *x* axis.

for identical or almost identical snoRNAs, as a probable consequence of the prevalence of polyploidy in plants [8]. Fig. 3 illustrates the degree of snoRNA gene redundancy in both *Arabidopsis* and rice, as compared to *Drosophila*, in whose genome snoRNA gene redundancy has also been documented. The presence of isocoding snoRNA gene copies might represent a preliminary stage in the evolution of snoRNAs with novel specificities [8]. An extreme example of genome enrichment with multiple copies of the same snoRNA coding units has been recently discovered in platypus (*Ornithorhynchus anatinus*), in whose genome thousands of copies of an H/ACA snoRNA gene were dispersed through a snoRNA-derived retroposon [43].

### *Human snoRNA genes*

In the genome of humans and, more generally, of mammals most of the snoRNA gene complement is intronic. As shown in Table 1 and Fig. 2, more than 90% of human snoRNA genes reside within introns. Interestingly, however, the genes coding for essential snoRNAs involved in pre-rRNA endonucleolytic processing (e.g. U3, U8, U13) are characterized by intergenic location and autonomous transcription (in particular, 7 out of 44 non-intronic snoRNA genes are devoted to the production of the U3 snoRNA). It should also be pointed out that more than half of human non-intronic snoRNA coding units corresponds to retrogenes derived from snoRNA retroposition [44,45]. As such, part of these gene copies might be non-functional. Another remarkable feature of human snoRNA gene organization, emerging from Table 1 and Fig. 2, is the absence of clustering for both independent and intronic snoRNA genes.

## Common functional features of intronic snoRNA host genes

When intronic snoRNAs were first discovered, they were found to be associated with genes coding for proteins involved in nucleolar function, ribosome structure or protein synthesis [46]. As snoRNAs ultimately participate in ribosome biogenesis, such a location appeared as physiologically relevant, having the potential to provide a regulatory link between partners acting in the same biological process [7,46]. As shown by Table 2, the recent genomic data strongly confirm that the tendency of snoRNA units to colonize ribosome-related genes represents a universal feature of snoRNA gene organization in eukaryotes. Importantly, most guide snoRNA-hosting genes in vertebrates belong to the family of TOP (terminal oligopyrimidine) genes, that include translation-related protein genes but also other genes characterized by high-level transcription and growth-dependent regulation ([47]; see also Table S5 in Supplementary data). The universal localization bias of intronic snoRNA genes immediately suggests the possibility of a coordinately regulated expression of snoRNAs and other components involved in the same process, i.e. ribosome biogenesis. Such a co-regulation is apparent when intronic snoRNAs originate through debranching of spliced introns (and this is the case for the majority of intron-nested snoRNAs), while it appears, at least in principle, more complex in cases of intronic snoRNA maturation in which the snoRNA-containing precursor is directly subjected to endonucleolytic cleavage, so that both splicing and cleavage can operate on the same precursor RNA [48]. It should be pointed out, however, that concrete examples of snoRNA and host gene co-regulation in response to stimuli have not been reported (see also below).

## Regulatory and evolutionary implications of snoRNA gene expressional adaptability

In going from yeast to plants and metazoa, the observed trend of snoRNA gene organization and expression is towards a reduction of the number of independent promoters. Such a reduction occurred in two different ways: clustering of snoRNA coding units, that allows for production of polycistronic transcripts and thus of multiple snoRNAs from a single promoter, and colonization of introns, allowing for exploitation of host gene promoters for snoRNA synthesis (see Fig. 2). These two strategies appear to have been at work together in the generation of snoRNA gene complements in plants (intronic snoRNA gene clusters occur frequently in the genomes of both land plants and the unicellular alga *Chlamydomonas reinhardtii* [8,29]) and, uniquely among metazoans, in *Drosophila* [30]. But it can be anticipated that the constant accumulation of new genomic and transcriptomic data will reveal intronic snoRNA clusters in other animal genomes. For example, small putative intronic snoRNA clusters have been detected in *C. elegans* (see above, and Table S3 in Supplementary Material), and in the zebrafish, *Danio rerio* (see http://snoopy.med.miyazaki-u.ac.jp/snorna_db.cgi?mode=code_seq_info&id=Danio_rerio100055). The ability of snoRNA coding units to get free of independent promoters, either by clustering or intronic integration or both, has important regulatory implications. One of them is that there must be no particular need for independent regulation of individual snoRNA genes. Such a property is typical of other ncRNA gene families, for example of tRNA genes, and also of gene sets coding for different protein components of the same cellular machinery, for example r-protein genes. An important difference, however, between such gene families and snoRNA gene complements is that tRNA and r-protein genes are all independently transcribed, and coordinately regulated as gene sets, either by modulation of the machinery acting on them [49,50] or by gene-family-specific transcription regulatory factors [51]. In contrast, the different members of a given snoRNA gene complement can differ profoundly in their way of expression, so that their coordinate regulation is difficult to imagine. Indeed, there are very few reports of regulation of snoRNA gene expression in response to environmental changes or any other stimulus: transcription of a yeast snoRNA gene cluster from its dedicated promoter was reported

**Table 2**
Functional features of intronic snoRNA host genes.

| Organism | % Ribosomal protein genes on | | % Ribosome- and translation-related genes on host genes[a] |
|---|---|---|---|
| | Total genes | Host genes | |
| *S. cerevisiae* | 3.0% | 37.5% | 87.5% |
| *S. pombe* | 2.9% | 50.0% | 75.0% |
| *C. elegans* | 0.4% | 19.6% | 23.4% |
| *D. melanogaster* | 0.6% | 43.3% | 50.8% |
| *H. sapiens* | 0.4% | 22.2% | 26.3% |
| *O. sativa* | 0.5% | 44.4% | 51.9% |
| *A. thaliana* | 2.0% | 21.7% | 52.2% |

[a] Gene Ontology terms for Ribosome- and translation-related genes: GO:0006412 translation; GO:0042254 ribosome biogenesis; GO:0003735 structural constituent of ribosome; GO:0003743 translation initiation factor activity.

UNCORRECTED PROOF

Please cite this article as: G. Dieci, et al., Eukaryotic snoRNAs: A paradigm for gene expression flexibility, Genomics (2009), doi:10.1016/j.ygeno.2009.05.002

to be 2-fold up-regulated in cells grown on glucose with respect to cells grown on glycerol [16]; several *C. elegans* snoRNAs have been found to display developmentally variable expression [24]; and, very recently, the expression of some *Arabidopsis* snoRNAs has been reported to be circadian clock-regulated [52]. In the particular case of yeast, where the ribosome biogenesis pathway and its regulation are relatively well characterized [53], the expression of snoRNAs has never been shown to be co-regulated with the expression of other genes involved in ribosome biogenesis. Even the essential genes coding for U3 and MRP RNA, that have independent promoters in all eukaryotes, have not been reported to be target of specific transcription regulatory pathways. The levels of U3 snoRNA, for example, have recently been reported to be co-regulated with the levels of r-protein mRNAs in *S. pombe* [18], yet this snoRNA appears to be regulated mainly at the post-transcriptional level [54]. What is then the meaning of snoRNA adaptability in transcription, leading to such a widespread lack of regulation at this stage? One reasonable explanation is that only high transcription levels matter for snoRNAs. High-level transcription can be achieved either by strong independent promoters (this seems to be the case in yeast, whose snoRNA genes ranked among the most highly occupied by RNA polymerase II in a genome-wide analysis of Pol II location [55]) or by localization in introns of highly transcribed, housekeeping genes, as is generally the case in all the analyzed eukaryotes (see above). Post-transcriptional regulation strategies could then operate to modulate the levels of the abundant pre-snoRNA transcripts [48,54]. With this respect, the strict coordination existing for the synthesis, assembly and trafficking of C/D and H/ACA snoRNPs should be taken into account as a fundamental aspect of regulation [4].

According to a model for the evolutionary origin of guide snoRNAs, the bulk of snoRNA species of each class (C/D or H/ACA) arose by duplication of an ancestral snoRNA gene [56]. The generation of snoRNA paralogs has been found to proceed with high plasticity in nematodes, both by *cis*-duplication (from one intronic location to a neighboring intron of the same gene) and by *trans*-duplication to distant genomic locations [37]. Retrotransposition can result in *trans*-duplication. Accordingly, human snoRNAs have recently been identified as a new family of retroposons that, when inserting in gene introns in the sense orientation, can be processed into functional snoRNAs that can eventually acquire new specificities [44]. Such a phenomenon must be general, as a high-copy number snoRNA retroposon has recently been revealed in platypus (*Ornithorhynchus anatinus*) [43]. Along this evolutionary scheme, the dissemination of snoRNA gene coding units in genomes would have mainly resulted in retention of gene copies characterized by high-level expression. In intron-rich genomes, the introns of housekeeping genes turned out to be ideally suited as snoRNA gene residence, while in intron-poor genomes, like those of yeasts, snoRNA coding units preceded by strong basal promoters were mainly retained as efficient snoRNA producers [16].

It is instructive to compare the genome organization of snoRNA genes with the one of miRNA coding units. In the human genome, only ~60% of miRNA coding units are located within introns [57], and several important examples of regulatory circuits involving independently transcribed and regulated miRNA genes have been reported (see for example [58]). In plants, too, a relatively small number of miRNA coding units have been found to be intronic [13], and specific regulation of intergenic miRNA genes is amply documented [59]. Thus miRNA genes, that need a much more complex regulation than snoRNA genes, display a less marked tendency to be incorporated within introns and thus to loose the potential of being autonomously regulated. We note, however, that this relatively simple picture has recently been complicated by the discovery that miRNAs can originate from snoRNA precursors [60,61]. A full understanding of snoRNA expression regulation will thus first require the disentanglement of the complex biosynthetic relationships between the increasing number of RNA families that compose the eukaryotic transcriptome [6].

## Acknowledgments

Research in authors' laboratory is supported by the Italian Ministry of University and Research (2007 PRIN Program grant to G.D.) and by Fondazione Cariparma (Research Fellowship to B.M.).

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2009.05.002.

## References

[1] T. Kiss, Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions, Cell 109 (2002) 145–148.
[2] S.L. Reichow, T. Hamma, A.R. Ferre-D'Amare, G. Varani, The structure and function of small nucleolar ribonucleoproteins, Nucleic Acids Res. 35 (2007) 1452–1464.
[3] W. Filipowicz, V. Pogacic, Biogenesis of small nucleolar ribonucleoproteins, Curr. Opin. Cell Biol. 14 (2002) 319–327.
[4] A.G. Matera, R.M. Terns, M.P. Terns, Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs, Nat. Rev., Mol. Cell Biol. 8 (2007) 209–220.
[5] A. Huttenhofer, J. Brosius, J.P. Bachellerie, RNomics: identification and function of small, non-messenger RNAs, Curr. Opin. Chem. Biol. 6 (2002) 835–843.
[6] P.P. Amaral, M.E. Dinger, T.R. Mercer, J.S. Mattick, The eukaryotic genome as an RNA machine, Science 319 (2008) 1787–1789.
[7] J.P. Bachellerie, J. Cavaille, A. Huttenhofer, The expanding snoRNA world, Biochimie 84 (2002) 775–790.
[8] J.W. Brown, M. Echeverria, L.H. Qu, Plant snoRNAs: functional evolution and new modes of gene expression, Trends Plant Sci. 8 (2003) 42–49.
[9] K.T. Tycowski, J.A. Steitz, Non-coding snoRNA host genes in *Drosophila*: expression strategies for modification guide snoRNAs, Eur. J. Cell Biol. 80 (2001) 119–125.
[10] S. Uliel, X.H. Liang, R. Unger, S. Michaeli, Small nucleolar RNAs that guide modification in trypanosomatids: repertoire, targets, genome organisation, and unique functions, Int. J. Parasitol. 34 (2004) 445–454.
[11] L. Lestrade, M.J. Weber, snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs, Nucleic Acids Res. 34 (2006) D158–D162.
[12] T. Kiss, E. Fayet, B.E. Jady, P. Richard, M. Weber, Biogenesis and intranuclear trafficking of human box C/D and H/ACA RNPs, Cold Spring Harbor Symp. Quant. Biol. 71 (2006) 407–417.
[13] J.W. Brown, D.F. Marshall, M. Echeverria, Intronic noncoding RNAs and splicing, Trends Plant Sci. 13 (2008) 335–342.
[14] G. Chanfreau, P. Legrain, A. Jacquier, Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism, J. Mol. Biol. 284 (1998) 975–988.
[15] E. Petfalski, T. Dandekar, Y. Henry, D. Tollervey, Processing of the precursors to small nucleolar RNAs and rRNAs requires common components, Mol. Cell. Biol. 18 (1998) 1181–1189.
[16] L.H. Qu, A. Henras, Y.J. Lu, H. Zhou, W.X. Zhou, Y.Q. Zhu, J. Zhao, Y. Henry, M. Caizergues-Ferrer, J.P. Bachellerie, Seven novel methylation guide small nucleolar RNAs are processed from a common polycistronic transcript by Rat1p and RNase III in yeast, Mol. Cell. Biol. 19 (1999) 1144–1158.
[17] G.W. Jawdekar, R.W. Henry, Transcriptional regulation of human small nuclear RNA genes, Biochim. Biophys. Acta 1779 (2008) 295–305.
[18] S. Nabavi, R.N. Nazar, U3 snoRNA promoter reflects the RNA's function in ribosome biogenesis, Curr. Genet. 54 (2008) 175–184.
[19] T. Kiss, C. Marshallsay, W. Filipowicz, Alteration of the RNA polymerase specificity of U3 snRNA genes during evolution and in vitro, Cell 65 (1991) 517–526.
[20] M. Antal, A. Mougin, M. Kis, E. Boros, G. Steger, G. Jakab, F. Solymosy, C. Branlant, Molecular characterization at the RNA and gene levels of U3 snoRNA from a unicellular green alga, *Chlamydomonas reinhardtii*, Nucleic Acids Res. 28 (2000) 2959–2968.
[21] G. Dieci, G. Fiorino, M. Castelnuovo, M. Teichmann, A. Pagano, The expanding RNA polymerase III transcriptome, Trends Genet. 23 (2007) 614–622.
[22] J.F. Briand, F. Navarro, O. Gadal, P. Thuriaux, Cross talk between tRNA and rRNA synthesis in *Saccharomyces cerevisiae*, Mol. Cell. Biol. 21 (2001) 189–195.
[23] D.J. Leader, G.P. Clark, J. Watters, A.F. Beven, P.J. Shaw, J.W. Brown, Clusters of multiple different small nucleolar RNA genes in plants are expressed as and processed from polycistronic pre-snoRNAs, EMBO J. 16 (1997) 5742–5751.
[24] W. Deng, X. Zhu, G. Skogerbo, Y. Zhao, Z. Fu, Y. Wang, H. He, L. Cai, H. Sun, C. Liu, B. Li, B. Bai, J. Wang, D. Jia, S. Sun, Y. Cui, D. Bu, R. Chen, Organization of the *Caenorhabditis elegans* small non-coding transcriptome: genomic features, biogenesis, and expression, Genome Res. 16 (2006) 20–29.
[25] K.T. Tycowski, A. Aab, J.A. Steitz, Guide RNAs with 5′ caps and novel box C/D snoRNA-like domains for modification of snRNAs in metazoa, Curr. Biol. 14 (2004) 1985–1995.
[26] E. Guffanti, R. Ferrari, M. Preti, M. Forloni, O. Harismendy, O. Lefebvre, G. Dieci, A minimal promoter for TFIIIC-dependent in vitro transcription of snoRNA and tRNA genes by RNA polymerase III, J. Biol. Chem. 281 (2006) 23945–23957.
[27] K. Kruszka, F. Barneche, R. Guyot, J. Ailhas, I. Meneau, S. Schiffer, A. Marchfelder,

UNCORRECTED PROOF

Please cite this article as: G. Dieci, et al., Eukaryotic snoRNAs: A paradigm for gene expression flexibility, Genomics (2009), doi:10.1016/j.ygeno.2009.05.002

ARTICLE IN PRESS



M. Echeverria, Plant dicistronic tRNA–snoRNA genes: a new mode of expression of the small nucleolar RNAs processed by RNase Z, EMBO J. 22 (2003) 621–632.
[28] Y. Isogai, S. Takada, R. Tjian, S. Keles, Novel TRF1/BRF target genes revealed by genome-wide analysis of *Drosophila* Pol III transcription, EMBO J. 26 (2007) 79–89.
[29] C.L. Chen, C.J. Chen, O. Vallon, Z.P. Huang, H. Zhou, L.H. Qu, Genomewide analysis of box C/D and box H/ACA snoRNAs in *Chlamydomonas reinhardtii* reveals an extensive organization into intronic gene clusters, Genetics 179 (2008) 21–30.
[30] Z.P. Huang, H. Zhou, H.L. He, C.L. Chen, D. Liang, L.H. Qu, Genome-wide analyses of two families of snoRNA genes from *Drosophila melanogaster*, demonstrating the extensive utilization of introns for coding of snoRNAs, RNA 11 (2005) 1303–1316.
[31] K.T. Tycowski, M.D. Shu, J.A. Steitz, A mammalian gene with introns instead of exons generating stable RNA products, Nature 379 (1996) 464–466.
[32] F. Ozsolak, L.L. Poling, Z. Wang, H. Liu, X.S. Liu, R.G. Roeder, X. Zhang, J.S. Song, D.E. Fisher, Chromatin structure analyses identify miRNA promoters, Genes Dev. 22 (2008) 3172–3183.
[33] A.D. Omer, T.M. Lowe, A.G. Russell, H. Ebhardt, S.R. Eddy, P.P. Dennis, Homologs of small nucleolar RNAs in Archaea, Science 288 (2000) 517–522.
[34] D. Piekna-Przybylska, W.A. Decatur, M.J. Fournier, New bioinformatic tools for analysis of nucleotide modifications in eukaryotic rRNA, Rna 13 (2007) 305–312.
[35] S.G. Li, H. Zhou, Y.P. Luo, P. Zhang, L.H. Qu, Identification and functional analysis of 20 Box H/ACA small nucleolar RNAs (snoRNAs) from *Schizosaccharomyces pombe*, J. Biol. Chem. 280 (2005) 16446–16455.
[36] V. Wood, R. Gwilliam, M.A. Rajandream, M. Lyne, R. Lyne, A. Stewart, J. Sgouros, N. Peat, J. Hayles, S. Baker, D. Basham, S. Bowman, K. Brooks, D. Brown, S. Brown, T. Chillingworth, C. Churcher, M. Collins, R. Connor, A. Cronin, et al., The genome sequence of *Schizosaccharomyces pombe*, Nature 415 (2002) 871–880.
[37] A. Zemann, A. op de Bekke, M. Kiefmann, J. Brosius, J. Schmitz, Evolution of small nucleolar RNAs in nematodes, Nucleic Acids Res. 34 (2006) 2676–2685.
[38] Z.P. Huang, C.J. Chen, H. Zhou, B.B. Li, L.H. Qu, A combined computational and experimental analysis of two families of snoRNA genes from *Caenorhabditis elegans*, revealing the expression and evolution pattern of snoRNAs in nematodes, Genomics 89 (2007) 490–501.
[39] M. Wachi, T. Ogawa, K. Yokoyama, Y. Hokii, M. Shimoyama, A. Muto, C. Ushida, Isolation of eight novel *Caenorhabditis elegans* small RNAs, Gene 335 (2004) 47–56.
[40] Z.P. Huang, H. Zhou, D. Liang, L.H. Qu, Different expression strategy: multiple intronic gene clusters of box H/ACA snoRNA in *Drosophila melanogaster*, J. Mol. Biol. 341 (2004) 669–683.
[41] D. Liang, H. Zhou, P. Zhang, Y.Q. Chen, X. Chen, C.L. Chen, L.H. Qu, A novel gene organization: intronic snoRNA gene clusters from *Oryza sativa*, Nucleic Acids Res. 30 (2002) 3262–3272.
[42] P. Braglia, R. Percudani, G. Dieci, Sequence context effects on oligo(dT) termination signal recognition by *Saccharomyces cerevisiae* RNA polymerase III, J. Biol. Chem. 280 (2005) 19551–19562.
[43] J. Schmitz, A. Zemann, G. Churakov, H. Kuhl, F. Grützner, R. Reinhardt, J. Brosius, Retroposed SNOfall—a mammalian-wide comparison of platypus snoRNAs, Genome Res. 18 (2008) 1005–1010.
[44] M.J. Weber, Mammalian small nucleolar RNAs are mobile genetic elements, PLoS Genet. 2 (2006) e205.
[45] Y. Luo, S. Li, Genome-wide analyses of retrogenes derived from the human box H/ACA snoRNAs, Nucleic Acids Res. 35 (2007) 559–571.
[46] E.S. Maxwell, M.J. Fournier, The small nucleolar RNAs, Annu. Rev. Biochem. 64 (1995) 897–934.
[47] R. Yamashita, Y. Suzuki, N. Takeuchi, H. Wakaguri, T. Ueda, S. Sugano, K. Nakai, Comprehensive detection of human terminal oligo-pyrimidine (TOP) genes and analysis of their characteristics, Nucleic Acids Res. 36 (2008) 3707–3715.
[48] V. de Turris, G. Di Leva, S. Caldarola, F. Loreni, F. Amaldi, I. Bozzoni, TOP promoter elements control the relative ratio of intron-encoded snoRNA versus spliced mRNA biosynthesis, J. Mol. Biol. 344 (2004) 383–394.
[49] I. Willis, R.D. Moir, Integration of nutritional and stress signaling pathways by Maf1, Trends Biochem. Sci. 32 (2007) 51–53.
[50] A.A. Michels, N. Hernandez, Does Pol I talk to Pol II? Coordination of RNA polymerases in ribosome biogenesis, Genes Dev. 20 (2006) 1982–1985.
[51] Y. Zhao, K.B. McIntosh, D. Rudra, S. Schawalder, D. Shore, J.R. Warner, Fine-structure analysis of ribosomal protein gene transcription, Mol. Cell. Biol. 26 (2006) 4853–4862.
[52] S.P. Hazen, F. Naef, T. Quisel, J.M. Gendron, H. Chen, J.R. Ecker, J.O. Borevitz, S.A. Kay, Exploring the transcriptional landscape of plant circadian rhythms using genome tiling arrays, Genome Biol. 10 (2009) R17.
[53] C.H. Wade, M.A. Umbarger, M.A. McAlear, The budding yeast rRNA and ribosome biosynthesis (RRB) regulon contains over 200 genes, Yeast 23 (2006) 293–306.
[54] S. Nabavi, S. Nellimarla, R.N. Nazar, Post-transcriptional regulation of the U3 small nucleolar RNA, J. Biol. Chem. 283 (2008) 21404–21410.
[55] E.J. Steinmetz, C.L. Warren, J.N. Kuehner, B. Panbehi, A.Z. Ansari, D.A. Brow, Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase, Mol. Cell 24 (2006) 735–746.
[56] D.L. Lafontaine, D. Tollervey, Birth of the snoRNPs: the evolution of the modification-guide snoRNAs, Trends Biochem. Sci. 23 (1998) 383–388.
[57] M. Morlando, M. Ballarino, N. Gromak, F. Pagano, I. Bozzoni, N.J. Proudfoot, Primary microRNA transcripts are processed co-transcriptionally, Nat. Struct. Mol. Biol. 15 (2008) 902–909.
[58] A. Marson, S.S. Levine, M.F. Cole, G.M. Frampton, T. Brambrink, S. Johnstone, M.G. Guenther, W.K. Johnston, M. Wernig, J. Newman, J.M. Calabrese, L.M. Dennis, T.L. Volkert, S. Gupta, J. Love, N. Hannett, P.A. Sharp, D.P. Bartel, R. Jaenisch, R.A. Young, Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells, Cell 134 (2008) 521–533.
[59] M.W. Jones-Rhoades, D.P. Bartel, B. B., MicroRNAS and their regulatory roles in plants, Annu. Rev. Plant. Biol. 57 (2006) 19–53.
[60] C. Ender, A. Krek, M.R. Friedlander, M. Beitzinger, L. Weinmann, W. Chen, S. Pfeffer, N. Rajewsky, G. Meister, A. human snoRNA with microRNA-like functions, Mol. Cell 32 (2008) 519–528.
[61] A.A. Saraiya, C.C. Wang, snoRNA, a novel precursor of microRNA in *Giardia lamblia*, PLoS Pathog. 4 (2008) e1000224.
[62] Y.Z. Bi, L.H. Qu, H. Zhou, Characterization and functional analysis of a novel double-guide C/D box snoRNA in the fission yeast, Biochem. Biophys. Res. Commun. 354 (2007) 302–308.
[63] Y. Sasano, Y. Hokii, K. Inoue, H. Sakamoto, C. Ushida, T. Fujiwara, Distribution of U3 small nucleolar RNA and fibrillarin during early embryogenesis in *Caenorhabditis elegans*, Biochimie 90 (2008) 898–907.
[64] S. Higa, N. Maeda, N. Kenmochi, T. Tanaka, Location of 2(′)-O-methyl nucleotides in 26S rRNA and methylation guide snoRNAs in *Caenorhabditis elegans*, Biochem. Biophys. Res. Commun. 297 (2002) 1344–1349.
[65] G. Yuan, C. Klämbt, J.P. Bachellerie, J. Brosius, A. Hüttenhofer, RNomics in *Drosophila melanogaster*: identification of 66 candidates for novel non-messenger RNAs, Nucleic Acids Res. 31 (2003) 2495–2507.
[66] R.J. Wilson, J.L. Goodman, V.B. Strelets, F. Consortium, FlyBase: integration and improvements to query tools, Nucleic Acids Res. 36 (2008) D588–D593.
[67] J.H. Yang, X.C. Zhang, Z.P. Huang, H. Zhou, M.B. Huang, S. Zhang, Y.Q. Chen, L.H. Qu, snoSeeker: an advanced computational package for screening of guide and orphan snoRNA genes in the human genome, Nucleic Acids Res. 34 (2006) 5112–5123.
[68] S. Nahkuri, R.J. Taft, D.J. Korbie, J.S. Mattick, Molecular evolution of the HBII-52 snoRNA cluster, J. Mol. Biol. 381 (2008) 810–815.
[69] C.L. Chen, D. Liang, H. Zhou, M. Zhuo, Y.Q. Chen, L.H. Qu, The high diversity of snoRNAs in plants: identification and comparative study of 120 snoRNA genes from *Oryza sativa*, Nucleic Acids Res. 31 (2003) 2601–2613.
[70] F. Qu, W. Zhai, H. Chen, L.H. Zhu, T.J. Morris, Cloning, characterization and transient expression of the gene encoding a rice U3 small nuclear RNA, Gene 172 (1996) 217–220.
[71] P. Piccinelli, M.A. Rosenblad, T. Samuelsson, Identification and analysis of ribonuclease P and MRP RNA in a broad range of eukaryotes, Nucleic Acids Res. 33 (2005) 4485–4495.
[72] W. Li, G. Jiang, D. Zeng, Y. Jin, Identification of six new box C/D snoRNA gene clusters from rice, IUBMB Life 59 (2007) 664–674.
[73] F. Barneche, F. Steinmetz, M. Echeverria, Fibrillarin genes encode both a conserved nucleolar protein and a novel small nucleolar RNA involved in ribosomal RNA methylation in *Arabidopsis thaliana*, J. Biol. Chem. 275 (2000) 27212–27220.
[74] F. Barneche, C. Gaspin, R. Guyot, M. Echeverria, Identification of 66 box C/D snoRNAs in *Arabidopsis thaliana*: extensive gene duplications generated multiple isoforms predicting new ribosomal RNA 2′-O-methylation sites, J. Mol. Biol. 311 (2001) 57–73.
[75] J.W. Brown, G.P. Clark, D.J. Leader, C.G. Simpson, T. Lowe, Multiple snoRNA gene clusters from *Arabidopsis*, Rna 7 (2001) 1817–1832.
[76] J.W. Brown, M. Echeverria, L.H. Qu, T.M. Lowe, J.P. Bachellerie, A. Huttenhofer, J.P. Kastenmayer, P.J. Green, P. Shaw, D.F. Marshall, Plant snoRNA database, Nucleic Acids Res. 31 (2003) 432–435.
[77] C. Marker, A. Zemann, T. Terhorst, M. Kiefmann, J.P. Kastenmayer, P. Green, J.P. Bachellerie, J. Brosius, A. Huttenhofer, Experimental RNomics: identification of 140 candidates for small non-messenger RNAs in the plant *Arabidopsis thaliana*, Curr. Biol. 12 (2002) 2002–2013.
[78] C. Marshallsay, T. Kiss, W. Filipowicz, Amplification of plant U3 and U6 snRNA gene sequences using primers specific for an upstream promoter element and conserved intragenic regions, Nucleic Acids Res. 18 (1990) 3459–3466.
[79] L.H. Qu, Q. Meng, H. Zhou, Y.Q. Chen, Identification of 10 novel snoRNA gene clusters from *Arabidopsis thaliana*, Nucleic Acids Res. 29 (2001) 1623–1630.

UNCORRECTED PROOF

Please cite this article as: G. Dieci, et al., Eukaryotic snoRNAs: A paradigm for gene expression flexibility, Genomics (2009), doi:10.1016/j.ygeno.2009.05.002