RNA (1998), *4:*445–454. Cambridge University Press. Printed in the USA.

Human U19 intron-encoded snoRNA is processed from a long primary transcript that possesses little potential for protein coding

MARIE-LINE BORTOLIN and TAMÁS KISS

Laboratoire de Biologie Moléculaire Eucaryote du Centre National de la Recherche Scientifique, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse, Cedex, France

ABSTRACT

While exons were originally defined as coding regions of split eukaryotic genes, introns have long been considered as mainly noncoding "genetic junk." However, recognition that a large number of small nucleolar RNAs (snoRNAs) are processed from introns of pre-mRNAs demonstrated that introns may also code for functional RNAs. Moreover, recent characterization of the mammalian UHG gene that encodes eight box C/D intronic snoRNAs suggested that some genes generate functional RNA products exclusively from their intron regions. In this study, we show that the human U19 box H/ACA snoRNA, which is encoded within the second intron of the U19H gene, represents the only functional RNA product generated from the long U19H primary transcript. Splicing of the U19H transcript, instead of giving rise to a defined RNA, produces a population of diverse U19H RNA molecules. Although the first three exons of the U19H gene are preserved in each processed U19H RNA, the 3′ half of the RNA is generated by a series of apparently random splicing events. Because the U19H RNA possesses limited potential for protein coding and shows a predominant nucleoplasmic localization, we suggest that the sole function of the U19H gene is to express the U19 intronic snoRNA. This suggests that, in marked contrast to our previous dogmatic view, genes generating functionally important RNAs exclusively from their intron regions are probably more frequent than has been anticipated.

Keywords: alternative splicing; intron-encoded snoRNA; noncoding RNA; pseudouridylation guide RNA

INTRODUCTION

In eukaryotic cells, the nucleolar processing of preribosomal RNAs (pre-rRNAs) is assisted by a large number of small nucleolar RNAs (snoRNAs) (Maxwell & Fournier, 1995; Sollner-Webb et al., 1996; Tollervey & Kiss, 1997). Some vertebrate (Kass et al., 1990; Savino & Gerbi, 1990; Peculis & Steitz, 1993; Tycowski et al., 1994) and yeast (Tollervey, 1987; Li et al., 1990; Hughes & Ares, 1991; Morrissey & Tollervey, 1993; Schmitt & Clayton, 1993) snoRNAs play essential roles in the nucleolytic formation of mature 18S, 5.8S, or 25/28S rRNAs. The majority of snoRNAs function as guide RNAs in the posttranscriptional nucleotide modification of rRNAs. Many snoRNAs that share the evolutionarily conserved C and D box sequence motifs and possess long complementarities to rRNAs direct the site-specific 2′-*O*-ribose methylation of rRNAs (Cavaillé et al., 1996; Kiss-László et al., 1996; Tycowski et al., 1996). Another large group of snoRNAs that possess the box H and ACA motifs function in the site-specific synthesis of ribosomal pseudouridines (Ganot et al., 1997b; Ni et al., 1997). Both methylation and pseudouridylation guide snoRNAs to select the correct substrate nucleotide via formation of direct Watson–Crick interactions with the target rRNA sequences.

In yeast, most of the snoRNAs are transcribed from independent genes by using their own transcription control elements (Tollervey & Kiss, 1997). In marked contrast, mammalian snoRNAs, with the exceptions of U3, U8, U13, and 7-2/MRP snoRNAs, are encoded within introns of protein-coding genes (reviewed by Filipowicz & Kiss, 1993; Sollner-Webb, 1993; Maxwell & Fournier, 1995; Tollervey & Kiss, 1997). The intronic snoRNAs are synthesized as part of the parent pre-mRNA and are posttranscriptionally processed from the intron sequences (Leverette et al., 1992; Fragapane et al., 1993; Kiss & Filipowicz, 1993; Tycowski et al., 1993). Most experimental data are consistent with a processing model in which the pre-mRNA hosting an intronic snoRNA first undergoes splicing and the snoRNA is released from the excised and debranched intron by

Reprint requests to: Tamas Kiss, Laboratoire de Biologie Moléculaire Eucaryote du Centre National de la Recherche Scientifique, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse, Cedex, France; e-mail: tamas@ibcg.biotoul.fr.

exonucleolytic activities (Kiss & Filipowicz, 1995; Cavaillé & Bachellerie, 1996; Caffarelli et al., 1996; Watkins et al., 1996). Most of the genes that house intronic snoRNAs encode ribosomal proteins, nucleolar proteins essential for rRNA processing, or cytoplasmic proteins that function in translation (Maxwell & Fournier, 1995; Kiss-László et al., 1996; Nicoloso et al., 1996; Ganot et al., 1997b). Therefore, selection of the host genes of intronic snoRNAs may provide a regulatory function that coordinates the accumulation of factors required for ribosome biogenesis or function (Filipowicz & Kiss, 1993; Mount & Henikoff, 1993; Sollner-Webb, 1993).

Introns of eukaryotic genes have long been considered as noncoding "junk DNA." Consistent with this view, intron sequences vary extensively even between closely related species, and, after splicing of a pre-mRNA, the removed introns are normally rapidly degraded. However, the demonstration that mammalian introns encode more than a hundred stable small RNA molecules that function in the nucleolar processing of rRNAs prompted us to reconsider the original definition of introns as noncoding regions of split genes (Mount & Henikoff, 1993; Sollner-Webb, 1993). Recently, a mammalian gene termed the UHG, which encodes eight box C/D snoRNAs (U22, U25–U31) within its nine introns, has been identified (Tycowski et al., 1996). Surprisingly, the spliced and polyadenylated UHG RNA is poorly conserved between human and mouse, lacks a long open reading frame, and is rapidly degraded in the cytoplasm. It seems, therefore, that the UHG gene does not code for a protein product and that the introns rather than the exons of the UHG gene generate functional RNA products. This notion questions the original definition of an exon as a coding region of a split gene and suggests that, contrary to our long-lived prejudice, exons do not necessarily encode a function, but rather that some genes may generate functional RNAs exclusively from their intron regions.

Here we show that the human U19 box H/ACA snoRNA is processed from the second intron of a long pre-mRNA-like transcript, the U19H precursor RNA. A complex series of alternative splicing events generate several variants of the spliced U19H RNAs. The spliced U19H RNAs, although polyadenylated, have a predominant nucleoplasmic localization and possess limited potential for protein coding. We propose that the U19H precursor RNA is synthesized and spliced only to express the intronic U19 snoRNA.

RESULTS

We have recently characterized a mammalian snoRNA, U19, that belongs to the family of the H and ACA box-containing snoRNAs (Kiss et al., 1996). Most members of this group of snoRNAs function as guide RNAs in the site-specific pseudouridylation of rRNAs. Based on the molecular mechanism that underlies the selection of ribosomal pseudouridines by box H/ACA snoRNAs, the U19 snoRNA was proposed to select the Ψ3744 residue in the central region of the human 28S rRNA (Ganot et al., 1997a). However, closer inspection of U19 RNA revealed another "pseudouridylation pocket" in the 5′-terminal hairpin with the potential to select the Ψ3746 residue in the 28S rRNA (Fig. 1). Interestingly, these two closely spaced pseudouridine residues are the only conserved pseudouridines in all the cytoplasmic large ribosomal subunit (LSU) RNAs examined thus far, including *Escherichia coli*, *Bacillus subtilis*, *Halobacterium halobium*, *Saccharomyces cerevisiae*, fruitfly (*Drosophila melanogaster*), mouse (*Mus musculus*), and human. Moreover, these modified nucleotides were found in maize (*Zea mays*) chloroplast LSU RNA as well (Ofengand et al., 1995; Ofengand & Bakin, 1997). We thus assumed that U19 has an important, if not essential, function in humans. This prompted us to investigate not only the function, but also the expression of this snoRNA.

Our previous preliminary results showed that the U19 snoRNA is encoded by a single-copy gene in the human genome and most likely generated by intron processing (Kiss et al., 1996). Supporting the latter conclusion, the U19 snoRNA was faithfully processed both in vitro from a longer precursor RNA and in vivo from the second intron of the human β-globin pre-mRNA (Kiss & Filipowicz, 1995; Kiss et al., 1996). However, cloning and sequencing of a 157 nt-long cDNA fragment encompassing the exon sequences that flank the U19-containing intron failed to identify a host gene for U19 snoRNA. Because this cDNA contained multiple translation stop signals in all three reading frames, we first tested whether the U19H RNA expressed from the U19 host gene is polyadenylated at all. Human HeLa cellular RNAs were fractionated by oligo(dT) cellulose chromatography and subjected to RNase A/T1 mapping using an antisense RNA probe transcribed from the partial U19H cDNA clone (Fig. 2A). Fractionation of the protected RNA fragments on a denaturing polyacrylamide gel showed that the U19H RNA, compared with total cellular RNAs (Fig. 2A, lane 2), was about 50-fold enriched in the poly(A)$^+$ fraction (Fig. 2A, lane 4) and was largely depleted from the poly(A)$^-$ fraction (Fig. 2A, lane 3) of human cellular RNAs.

Although the spliced RNA product of the U19H gene is apparently polyadenylated, northern blot analyses of HeLa poly(A)$^+$ RNAs, contrary to our repeated efforts, failed to detect a distinct hybridizing band for the U19H RNA. Instead, we observed a weak smear of signals ranging in size from 800 to 1,400 nt (data not shown). Therefore, to learn more about the host gene of U19 snoRNA, we undertook a screening of a human cDNA library using a partial U19H cDNA as a probe. We isolated and sequenced five different cDNAs expressed from the U19H gene. Each cDNA contained the previ-

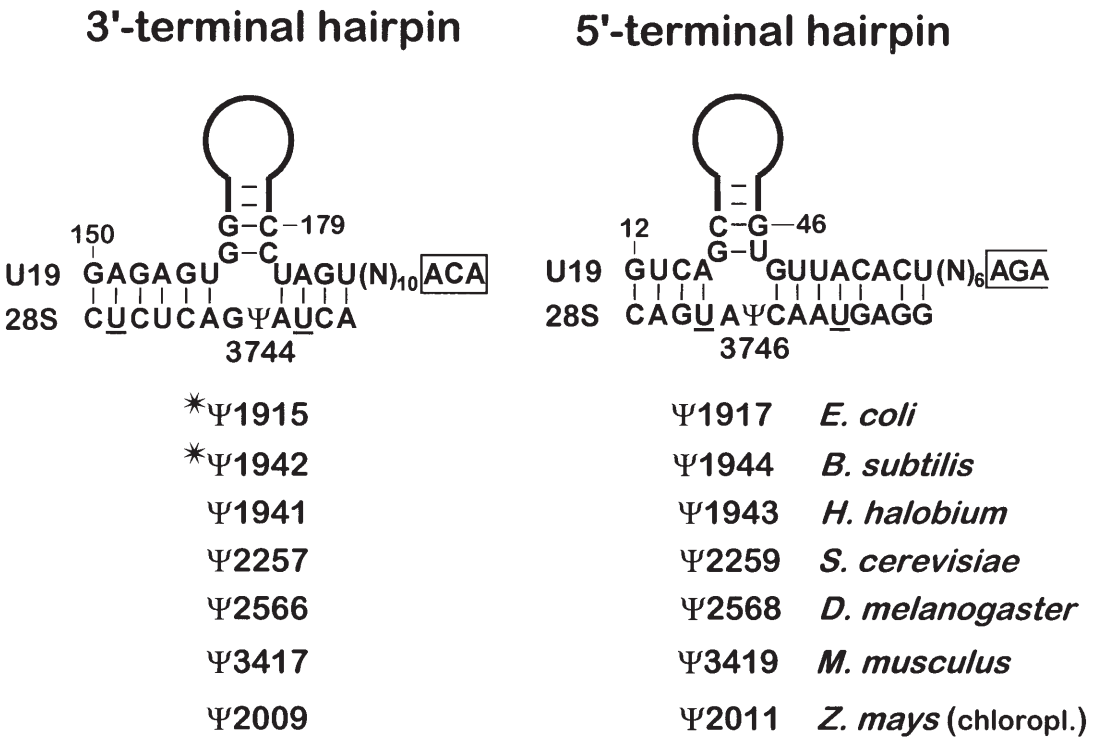

FIGURE 1. Selection of two evolutionarily highly conserved pseudouridine residues in human 28S rRNA by U19 snoRNA. The sequence and secondary structure of the human U19 snoRNA were taken from Kiss et al. (1996). The 3′- and 5′-terminal hairpin domains of the U19 RNA are schematically represented. Selected uridine residues are indicated by Ψ. Other known pseudouridine residues are marked by U. The ACA motif of U19 snoRNA is boxed, whereas the 5′ half of the H box motif is indicated by an open-ended box. For more details about the molecular mechanism underlying the site-specific selection of ribosomal pseudouridines by box H/ACA snoRNAs, see Ganot et al. (1997a). Sequence of human 28S rRNA was taken from the GenBank (accession number U13369). Positions of human pseudouridines as well as locations of pseudouridines equivalent to human Ψ3744 and Ψ3746 residues in *E. coli*, *B. subtilis*, *H. halobium*, *S. cerevisiae*, *D. melanogaster*, and *M. musculus* cytoplasmic LSU RNAs were taken from Ofengand and Bakin (1997). Sequence and numbering of *Z. mays* chloroplast LSU RNAs was taken from Edwards and Kössel (1981), and locations of the pseudouridines were reported by Ofengand and Bakin (1997). Asterisks indicate N^3-methyl Ψs present in *E. coli* and most likely also in *B. subtilis* (Ofengand & Balakin, 1997).

ously identified upstream (E_U) and downstream (E_D) exons, now renamed exons 2 (E2) and 3 (E3), respectively, some common upstream sequences, and, unexpectedly, completely divergent downstream sequences. Primer extension analyses confirmed that the 5′ terminus of the U19H RNAs commences with the common CUGCGCCCU… sequence (Fig. 2B), which is highly reminiscent of the 5′-terminal sequences of vertebrate ribosomal protein mRNAs that start invariantly with a C residue followed by a stretch of 4–13 pyrimidine residues (Meyuhas et al., 1996).

After amplification of a human genomic fragment by using primers specific for the 5′-terminal sequences of the isolated cDNAs and the E2 region, it became evident that four of the five isolated cDNAs share three common 5′-terminal exons (E1–E3) (Fig. 3). One cDNA, cDNA5, that represents about 2–4% of the total processed U19H RNAs as estimated by RNase mappings (data not shown), carries an 88 nt-long intronic fragment that follows the second exon. This alternatively spliced cDNA was also detected by reverse transcription (RT)-PCR (Kiss et al., 1996).

Surprisingly, the third exon (E3) in each cDNA was found to be spliced to different downstream exons. In cDNA1, the E3 is spliced to a 67 nt-long fragment (E4) located 803 nt downstream of the E3 region. About one third of the U19H primary transcripts follow this splicing pathway. RNase mapping revealed a putative 60 nt-long fifth exon in this cDNA, but also showed that only about 25% of the E4 exons are joined to this exon, indicating that the rest of the E4 exons are spliced to other, thus far unidentified, sequences. Because RNase mappings failed to unambiguously detect the further downstream sequences of cDNA1, we concluded that this region of cDNA1 was generated by rare splicing events (data not shown). In about 30% of the total U19H primary transcripts, the E3 region is spliced to downstream sequences of unknown origin resulting in cDNA3.

In cDNA4 that comprises approximately 2% of the total processed U19H RNA population, to our big surprise, the third exon was spliced to sequences that showed 97.5% identity to a 154 nt-long internal fragment of the coding region of the rat matrin 3 cDNA

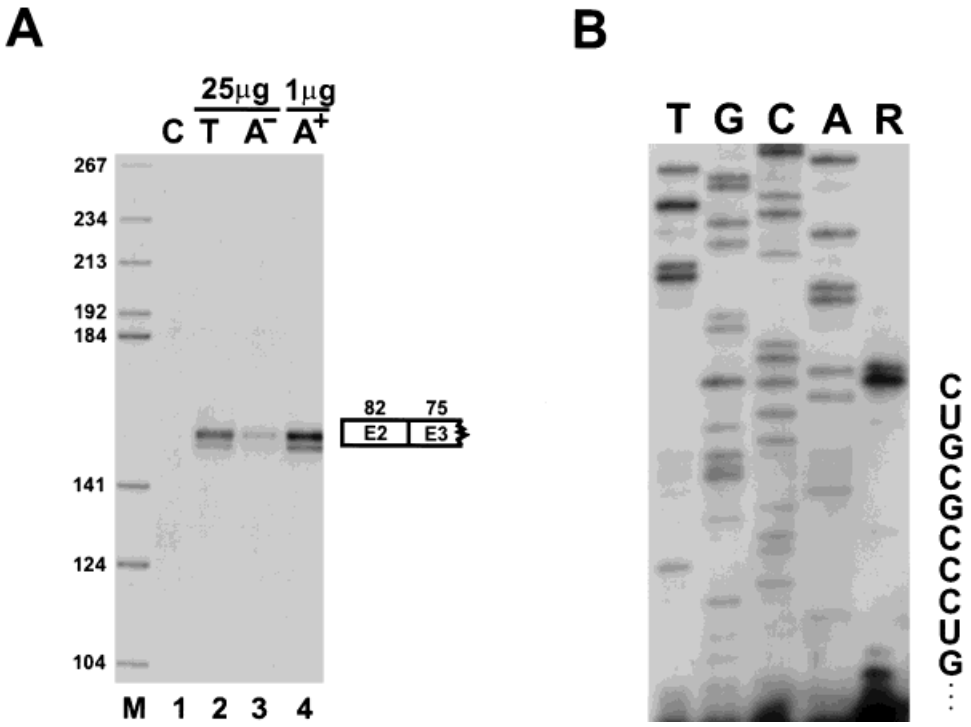

FIGURE 2. Characterization of human U19H RNA. **A:** U19H RNA is polyadenylated. Approximately 25 μg of total RNA (T) isolated from human HeLa cells, 25 μg of poly(A)$^-$ fraction (A$^-$), or 1 μg of poly(A)$^+$ fraction (A$^+$) of HeLa cellular RNAs were mapped by RNase A/T1 protection using an antisense RNA probe complementary to the second and third exons (E2 and E3) of the U19H gene. Protected RNA fragments were separated on a 6% sequencing gel. Lane C shows a control mapping with *E. coli* RNA. Lane M, kinase-labeled DNA size markers (a mixture of *Hae* III-and *Taq* I-digested pBR322). **B:** Determination of the 5′ end of human U19H RNA by primer extension analysis. A terminally labeled oligonucleotide complementary to the U19H RNA from position 23 to 42 was annealed with 6 μg of human HeLa poly(A)$^+$ RNA and extended by AMV reverse transcriptase and the extended products were separated on a 6% sequencing gel (lane R). Lanes T, G, C, and A represent dideoxy sequencing reactions performed on cDNA1. Note that, as a result of a migration artifact, the first G residue of the linker sequence reproducibly comigrates with 5′-terminal C1 residue of the U19H RNA. Sequencing of the U19H genomic DNA or cDNA1 in the opposite direction, however, unambiguously demonstrated that the 5′ terminus of the U19H RNA commences with a C residue.

(from position 912 to 1065; Belgrader et al., 1991). Unfortunately, the corresponding region of the human matrin 3 gene has not yet been characterized. However, the fact that the sequence alterations found in the human sequence, although introducing four amino acid changes, did not alter the protein-coding capacity of the human RNA, suggests that the third exon of cDNA4 is spliced with an internal exon of the authentic human matrin 3 pre-mRNA. To exclude the formal possibility that the cDNA4 derived from ligation of nonlinear DNA fragments during the construction of our cDNA library, we performed an RT-PCR analysis using primers specific for the E2 and matrin sequences (Fig. 4). Cloning and sequencing of the amplified DNA fragment unambiguously demonstrated that the cDNA4 was generated by a real, although rare, splicing event that joined the E3 region of U19H precursor RNA to an internal exon of the human matrin 3 pre-mRNA. Despite the failure in amplification of a genomic fragment spanning the region between the U19H gene and the matrin 3 gene, we believe that transcription initiated from the U19H promoter runs into the matrin gene and that splicing of this long U19H-matrin chimeric transcript results in the cDNA4.

Finally, RNase mappings with a cDNA2-specific probe showed that about 15–20% of the U19H primary transcripts remained unspliced after the third exon, indicating that the 5′ splice site at the end of exon 3 is utilized inefficiently. Taken together, this observation indicates that the primary transcript of the human U19H gene is poorly and randomly spliced. Although the first three 5′-terminal exons are included in all variants of the processed U19H RNA, the 3′ half of the U19H RNA is generated by a series of alternative splicing events that result in a divergent population of U19H RNAs both in sequence and size. We assume that a large number of not yet identified alternatively spliced versions of the U19H RNA exist in human cells. This would explain why northern analyses failed to identify a well-defined RNA product for the U19H gene.

Searching of databases showed that, apart from the Alu elements present in cDNA2 and cDNA3 and the putative matrin gene fragment spliced to the third exon of cDNA4, no sequences identical or similar to our

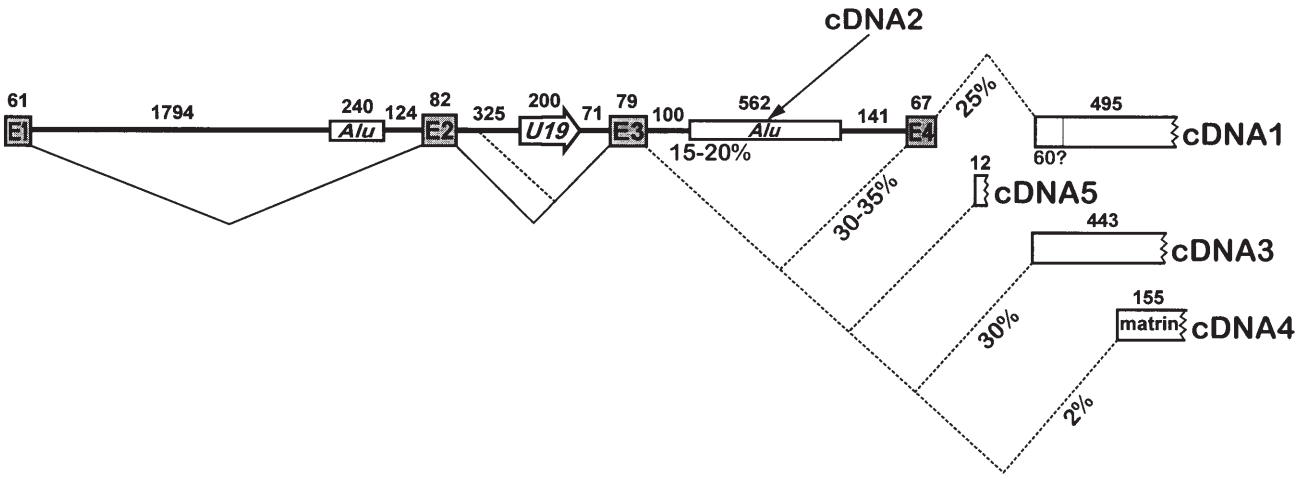

FIGURE 3. Schematic structure of the 5′-terminal part of the human U19H gene. Exons (E1–E4) located on the cloned genomic DNA fragment are represented by shaded boxes. cDNA sequences of unknown genomic origin are indicated by open boxes. Positions of the U19 snoRNA and the *Alu* elements are as shown. Arrow shows the position where the partial sequence of cDNA2 terminates. Alternative splicing events generating cDNA1, cDNA3, cDNA4, and cDNA5 are indicated by dashed lines. Frequency of splicing events determined by RNase A/T1 mappings using specific antisense RNA probes for each cDNA is indicated. For other details, see the text.

cDNAs have been deposited in public databases. Compilation of the putative translation initiator and terminator codons of the U19H cDNAs revealed that these RNAs have very little potential for protein coding (Fig. 5). The putative translation initiator codons in the first three exons that are common to all cDNAs are always followed by multiple translational stop signals in all three reading frames and consequently could initiate the synthesis of only very short oligopeptides. For example, the first AUG codon, which lies within an unfavorable sequence context (CAGAUGC) compared with the consensus RNNAUGR sequence (Kozak, 1996), has the potential to initiate the synthesis of a 14-amino acid oligopeptide. It is noteworthy that the first AUG serves as the unique translation initiation site in about 90% of vertebrate mRNAs, and that normally AUG co-

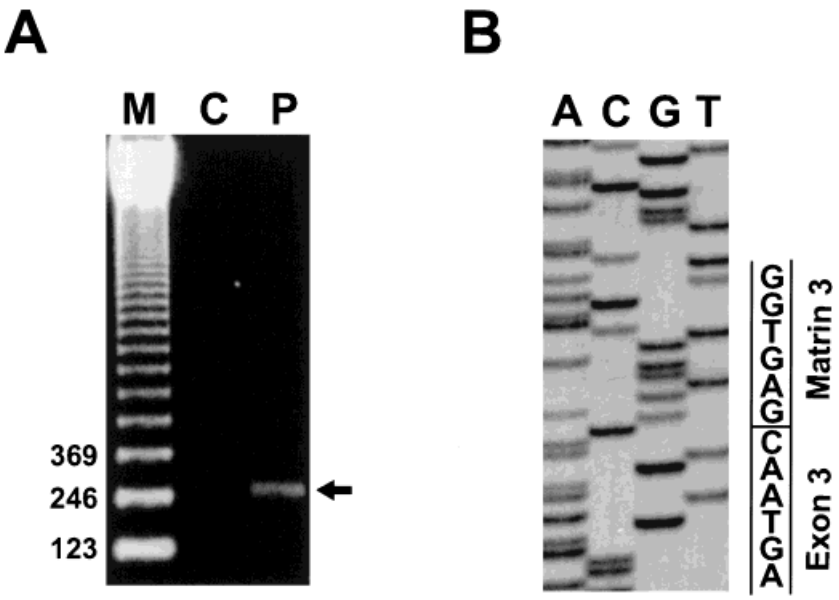

FIGURE 4. Third exon of the U19H gene is spliced to an internal exon of the human matrin 3 gene. **A:** Detection of the chimeric U19H-matrin RNA by RT-PCR. A cDNA fragment of expected size (258 bp) was amplified by using primers specific to the E2 region of U19H RNA and complementary to the human matrin sequences found in cDNA4 (lane P). Lane C represents a control amplification reaction without reverse transcription. **B:** Nucleotide sequence of the junction region of the U19H-matrin chimeric RNA.

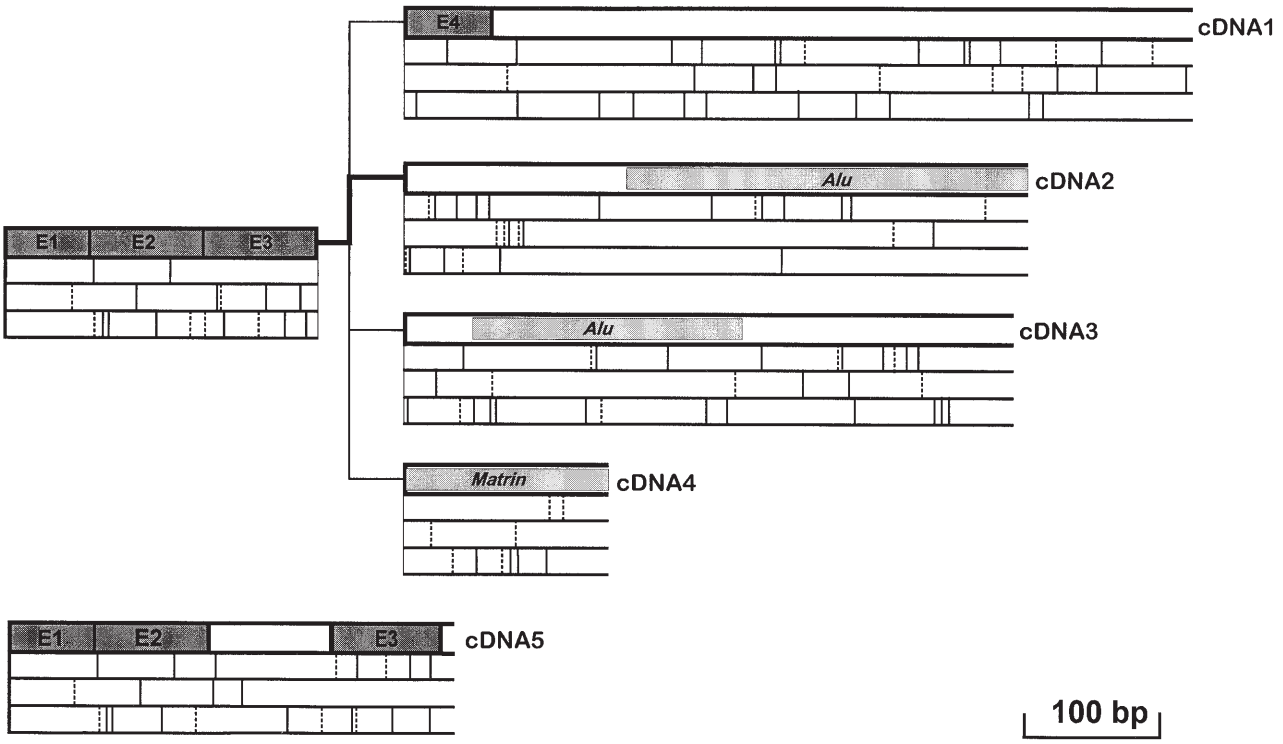

FIGURE 5. Open reading frames of the U19H RNAs. Positions of potential translation start and stop codons in all three reading frames are indicated by dashed or continuous vertical bars, respectively. In cDNA2, the thick line connects contiguous genomic sequences rather than spliced RNA sequences.

dons in the 5′ untranslated regions interfere with translation from downstream open reading frames (Kozak, 1987). The longest open reading frame, present at positions 294–518 in cDNA3, could encode a 74-amino acid polypeptide. However, it seems unlikely that this open reading frame, which encompasses an Alu element, could be translated efficiently, because it is preceded by multiple AUG codons and its putative initiator codon lies within a suboptimal sequence context (AAAUGAC). This suggests that the primary U19H transcript, although it undergoes splicing and polyadenylation, does not code for a protein product.

This conclusion was further corroborated by subcellular localization of the U19H RNA (Fig. 6). Human HeLa cells were fractionated into cytoplasmic, nuclear, nucleoplasmic, and nucleolar fractions. RNAs isolated from each fraction were analyzed by RNase mapping using probes specific for the U19H RNA. As a control, we also tested the subcellular distribution of the human high mobility group protein I (HMG-I) mRNA, the nucleoplasmic U4 snRNA, and the nucleolar U3 snoRNA. To our surprise, the U19H RNA was barely detectable in the cytoplasmic fraction (Fig. 6, lane 4). Instead, it was found mainly in the nuclear (Fig. 6, lane 3) and the nucleoplasmic (Fig. 6, lane 6) fractions, indicating that the processed U19H RNA either stays within the nucleoplasm or that it is rapidly degraded as soon as it has been exported to the cytoplasm. In marked contrast, the HMG-I mRNA was readily detectable in the cytoplasmic fraction (Fig. 6, lane 4) and, as expected, the U4 and U3 RNAs showed a predominant nucleoplasmic and nucleolar localization, respectively. Hence, it seems very unlikely that the U19H RNA serves as a message in cytoplasmic protein synthesis. Instead, we propose that the human U19H primary transcript is synthesized and processed only to express the U19 intronic snoRNA. Our results also suggest that genes generating functional RNAs from their intron regions rather than their exons are more general than has been anticipated.

DISCUSSION

During the last several years, it has become evident that the nucleous of vertebrate cells contains at least 150–200 snoRNAs that are encoded within introns of protein-coding genes (Maxwell & Fournier, 1995; Tollervey & Kiss, 1997). On the basis of their structure and nucleolar function, the intronic snoRNAs fall into two families. One group of snoRNAs that possess the C and D boxes directs the 2′-*O*-ribose methylation of rRNAs, whereas the family of the H and ACA box-containing snoRNAs function in the site-specific synthesis of ribosomal pseudouridines. The U19 snoRNA, which belongs to the H/ACA class of snoRNAs, has the potential to select two evolutionarily highly conserved

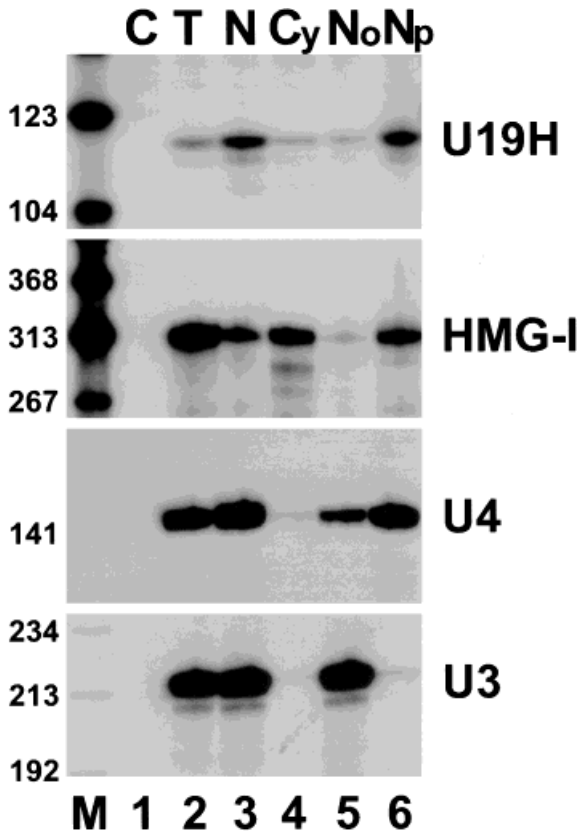

FIGURE 6. Intracellular localization of U19H RNA in human HeLa cells. Equal amounts of RNA samples (1 μg) isolated either from HeLa cells (T) or nuclear (N), cytoplasmic (Cy), nucleolar (No), or nucleoplasmic (Np) fractions of HeLa cells were analyzed by RNase A/T1 mapping using antisense RNA probes as indicated on the right. Protected RNA fragments were separated on a 6% sequencing gel. Lanes C and M are control mappings with *E. coli* RNA and size markers, respectively.

pseudouridine residues in the central domain of the 28S rRNA (Fig. 1). In this study, we have shown that the human U19 snoRNA is processed from the second intron of a long pre-mRNA-like transcript, termed the precursor U19H RNA. Several lines of evidence support the notion that the spliced product of the precursor U19H RNA is a nonprotein-coding, polyadenylated RNA that does not serve as a message in protein synthesis. Although the first three exons of the U19H RNA are conserved, the rest of the RNA is generated by a complex series of apparently random splicing events. The processed U19H RNAs, in spite of being polyadenylated, have limited potential for protein coding. Finally, the U19H RNAs possess a predominant nucleoplasmic rather than cytoplasmic localization. We thus propose that the precursor U19H RNA is synthesized and processed only to express the intron-encoded U19 snoRNA that likely possesses an important function in the nucleolar formation of mature 28S rRNA.

Recently, another mammalian gene (UHG) that encodes eight intronic snoRNAs and an apparently noncoding, polyadenylated RNA has been characterized (Tycowski et al., 1996). The spliced UHG RNA, although it is associated with polysomes, lacks a long open reading frame. Therefore, the authors proposed that the primary UHG transcript serves as a short-lived vehicle for production of intronic snoRNAs. Characterization of the U19H gene not only lends strong support to this assumption, but also indicates that mammalian genes whose sole function is to generate intronic snoRNAs or perhaps other types of intron-born RNAs are more general than has been anticipated. Indeed, additional two candidates of mammalian genes that express intronic snoRNAs but specify no functional mRNAs have been identified recently (P. Pelczar & W. Filipowicz; C.M. Smith & J.A. Steitz, pers. comm.).

Despite the obvious functional similarity, the UHG and U19H genes show striking differences in many aspects. The mouse and human UHG genes are unusually compact; they possess very short exon and intron regions. In marked contrast, the primary transcript of the human U19H gene spans a long region of human genomic DNA. It carries long intron regions, especially at the 3′ half of the RNA, as indicated by our failure in amplification of genomic DNA fragments spanning downstream intron regions that connect the third exon (E3) and the downstream cDNA sequences detected either in cDNA1, cDNA3, or cDNA4. Although the UHG gene encodes eight box C/D snoRNAs, the U19H gene appears to generate only one snoRNA that belongs to the box H/ACA family of snoRNAs. Splicing and polyadenylation of the precursor UHG RNA results in a well-defined mRNA-like product; however, the precursor U19H RNA undergoes multiple alternative splicing events yielding a population of U19H RNAs that possess diverse 3′-terminal regions. Because we did not extend our investigations to the nonconserved 3′ half of the U19H RNA, it remains unclear whether all of the alternatively spliced variants of the U19H RNA utilize a unique polyadenylation signal or whether the 3′ termini of U19 RNAs are formed at various polyadenylation sites. In contrast to the UHG RNA that is associated with polysomes in the cytoplasm, the U19H RNA shows a predominant nucleoplasmic localization. At the moment, it is unclear whether the processed U19H RNA remains within the nucleoplasm or whether it is exported from the nucleus, albeit with low efficiency, and is rapidly degraded in the cytoplasm. Nevertheless, it seems unlikely that a considerable amount of U19H RNA could associate with cytoplasmic polysomes. Consistent with this, treatment of HeLa cells with the protein synthesis inhibitor cycloheximide did not increase the level of the U19H RNA (our unpubl. obs.), indicating that the turnover of the U19H RNA, in contrast to the UHG RNA (Tycowski et al., 1996), is not coupled with association with polysomes.

Several examples of noncoding, polyadenylated RNAs have been implicated in important regulatory functions, such as regulation of X chromosome expression in

mammals and *Drosophila* (Brockdorff et al., 1992; Brown et al., 1992; Meller et al., 1997), regulation of *Xenopus* oogenesis (Kloc & Etkin, 1994) and germ line development (Nakamura et al., 1996), regulation of meiotic DNA synthesis in *Schizosacharomyces pombe* (Watanabe & Yamamoto, 1994), and control of plant growth and development (Crespi et al., 1994). Other noncoding RNAs, *Drosophila* hsr-omega-n and hsr-omega-c RNAs (Hogan et al., 1994), hamster *gadd7* RNA (Hollander et al., 1996), mammalian 7H4, IPW, and H19 RNAs (Brannan et al., 1990; Velleca et al., 1994; Wevrick et al., 1994) are also believed to perform some, although not yet clearly defined, regulatory functions. These "riboregulator" RNAs perform their functions without ever making a protein product. At the moment, we cannot rule out the formal possibility that the U19H RNA also possesses some cellular function. The nucleoplasmic localization of the U19H RNA seems to support this consideration. If the U19H RNA really has a nucleoplasmic function, one can assume that this putative function would be linked to the common 5′-terminal sequences that are preserved in all variants of the HU19 RNA. However, RNase mapping of mouse poly(A)$^+$ RNAs with a probe specific for the 5′-terminal region of human U19H failed to detect protected RNA fragment of significant length, indicating that the U19H RNA shows little or no conservation between human and mouse (data not shown). Computer folding analysis showed that, contrary to the expectation for a functional RNA molecule, the U19H RNA is very loosely structured. These considerations argue against a nucleoplasmic function of U19H RNA and further support the notion that the sole function of the HU19 gene is to express the U19 intronic snoRNA.

The 5′-terminal sequence of the U19H RNA (5′-CUGCGCCCUG-3′) is highly reminiscent of the 5′TOP motifs of ribosomal protein-coding mRNAs, which possess an invariant C residue at their cap site, followed by a stretch of pyrimidine residues (Meyuhas et al., 1996). Although the 5′TOP motif functions as a translational *cis*-regulatory element, it seems unlikely that the 5′TOP-like motif of the U19H RNA has a function in translation. However, the presence of a 5′TOP-like motif in the U19H RNA indicates that the U19H gene and ribosomal protein genes may share similar promoters and their expression may be coordinated at the transcriptional level. This may also provide some hints concerning the origin of human U19H gene. We propose that the U19H gene originally specified, in addition to the U19 snoRNA, a functional mRNA that encoded a protein product likely involved in ribosome biogenesis or function. The ancient U19H gene subsequently lost its protein-coding potential, and its function was adopted by another copy of the gene. To ensure the expression of the U19 intronic snoRNA, however, the U19H gene remained actively transcribed and the primary U19H transcript, at least regarding the second and third exons flanking the U19-containing intron, remained correctly and efficiently spliced, whereas the rest of the RNA was apparently poorly and randomly processed.

In summary, characterization of the human U19H gene strongly supports the assumption that some eukaryotic genes may generate functional RNAs with their introns rather than their exons. This notion, together with the fact that a large number of small nucleolar RNAs are encoded within introns of pre-mRNAs, indicates that the original definition of exon and intron as a coding and noncoding region of split eukaryotic genes, respectively, does not necessarily hold for all genes; some introns may encode a function, whereas some exons apparently do not.

MATERIALS AND METHODS

General procedures

Standard laboratory protocols were used for manipulating DNA, RNA, and oligodeoxynucleotides (Sambrook et al., 1989). The following oligodeoxynucleotides were used in this study: A, GATGACTTGAAAGTAGGGC; B, TCTGAAGGGAG GAAATAGT; C, CTTATCTTGGTGTCACACAG; D, ATAAAG CTTGAAAGTAGGGCATCCTTC; E, CCGGAATTCATTGTCA GGATTCCATTCTG; F, AGCCAGCCTCTGCGCACGTG.

Isolation and characterization of cDNAs

A human HeLa oligo(dT)+random-primed cDNA library constructed in λgt11 (Clontech) was screened using an internally labeled antisense RNA complementary to the second and third exons of the U19H RNA as a probe (Kiss et al., 1996). Screening of about 2×10^6 plaques identified 12 positive recombinant clones that carried five different cDNA fragments. The positive inserts were subcloned into the *Eco*R I site of the pBluescribe vector (Stratagene) and sequenced by the dideoxynucleotide chain-termination method using [α-^{35}S]ATP and T3 or T7 primers. To facilitate sequencing of the internal regions of long inserts, overlapping sets of deletions were created by unidirectional digestion with exonuclease III (Henikoff, 1984). Genomic sequences flanking the U19 snoRNA coding region were cloned by inverse PCR (Kiss et al., 1996).

RNA analysis

HeLa cellular RNA was isolated by the guanidinium thiocyanate/phenol-chloroform extraction method (Goodall et al., 1990). RNAs from the nuclear, cytoplasmic, nucleoplasmic, and nucleolar fractions of HeLa cells (Tyc & Steitz, 1989) were extracted by the hot phenol/SDS method of Steele et al. (1965). RNase A/T1 mapping was performed as described by Goodall et al. (1990). RNA probes were synthesized in vitro using T3 or T7 RNA polymerases, [α-^{32}P]CTP (sp. act. 30–40 Ci/mmol), and properly linearized recombinant plasmids as templates. All probes were purified on a 6% denaturing polyacrylamide gel. To obtain template DNAs for in vitro transcription of U19H-specific antisense RNA probes,

used both for RNase mapping and screening of the human cDNA library, two cDNA fragments of the U19H RNA were cloned by RT-PCR using a cDNA obtained by randomly primed reverse transcription of HeLa poly(A)$^+$ RNAs as a template. In these amplification reactions, oligonucleotides A or B, specific for exon 2, and oligonucleotide C, complementary to exon 3, were used as primers. The amplified fragments were cloned into the *Sma* I site of pBluescribe, linearized with *Eco*R I, and transcribed by T3 RNA polymerase. A recombinant plasmid, pHMG (N. Aulner, D. Jullien, & E. Käs, unpubl. results), carrying the full-length cDNA of the human high mobility group I protein (HMG-I) inserted into the *Kpn* I and *Xba* I sites of the pBluescript vector (Stratagene), was kindly provided by N. Aulner. The pHMG was linearized by *Eco*R I and transcribed by T3 RNA polymerase. RNA probes used for mapping human U3 and U4 snRNAs have been described (Kiss et al., 1996).

A splicing event that joins the third exon (E3) of the U19H gene to an internal exon of the human matrin 3 gene was verified by RT-PCR using primers specific for the second exon of the U19H gene (oligo D) and complementary to the human matrin 3 sequences present in cDNA4 (oligo E). The amplified fragment was cloned into the *Eco*R I and *Hin*d III sites of the pBluescript vector using the PCR-introduced *Eco*R I and *Hin*d III sites and was subjected to sequence analysis. The 5′ terminus of the U19H RNA was determined by primer extension analysis using 6 μg of HeLa poly(A)$^+$ RNAs and 1 pM of terminally labeled oligonucleotide F as a primer (complementary to the U19H RNA from position 23 to 42). The extended DNA products were separated on a 6% sequencing gel.

NOTE ADDED IN PROOF

The accession numbers for the cDNA sequences reported in this paper are as follows: cDNA1 (AJ224166), cDNA2 (AJ224167), cDNA3 (AJ224168), cDNA4 (AJ224169), and cDNA5 (AJ224170).

ACKNOWLEDGMENTS

We are grateful to N. Aulner for providing us with human HMG-I cDNA, Y. de Preval for synthesis of deoxyoligonucleotides, and C. Monod for critical reading of the manuscript. This study was supported by the Centre National de la Recherche Scientifique, and by grants from la Ligue Nationale Contre le Cancer and la Fondation pour la Recherche Médicale.

Received November 19, 1997; returned for revision December 19, 1997; revised manuscript received January 26, 1998

REFERENCES

Belgrader P, Dey R, Berezney R. 1991. Molecular cloning of matrin 3, a 125-kilodalton protein of the nuclear matrix contains an extensive acidic domain. *J Biol Chem 266*:9893–9899.

Brannan CI, Dees EC, Ingram RS, Tilghman SM. 1990. The product of the *H19* gene may function as an RNA. *Mol Cell Biol 10*:28–36.

Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, Swift S, Rastan S. 1992. The product of the mouse *Xist* gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell 71*:515–526.

Brown CJ, Hendrich BD, Rupert JL, Lafrenière RG, Xing Y, Lawrence J, Willard HF. 1992. The human *XIST* gene: Analyses of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell 71*:527–542.

Caffarelli E, Fatica A, Prislei S, De Gregorio E, Fragapane P, Bozzoni I. 1996. Processing of the intron-encoded U16 and U18 snoRNAs: The conserved C and D boxes control both the processing reaction and the stability of the mature snoRNA. *EMBO J 15*:1121–1131.

Cavaillé J, Bachellerie JP. 1996. Processing of fibrillarin-associated snoRNAs from pre-mRNA introns: An exonucleolytic process exclusively directed by common stem-box terminal structure. *Biochimie 78*:443–456.

Cavaillé J, Nicoloso M, Bachellerie JP. 1996. Targeted ribose methylation of RNA in vivo directed by tailored antisense RNA guides. *Nature 383*:732–735.

Crespi MD, Jurkevitch E, Poiret M, d'Aubenton-Carafa Y, Petrovics G, Kondorosi E, Kondorosi A. 1994. *enod40*, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth. *EMBO J 13*:5099–5112.

Edwards K, Kössel H. 1981. The rRNA operon from *Zea mays* chloroplasts: Nucleotide sequence of 23S rDNA and its homology with *E. coli* 23S rDNA. *Nucleic Acids Res 9*:2853–2869.

Filipowicz W, Kiss T. 1993. Structure and function of small nucleolar RNPs. *Mol Biol Rep 18*:149–156.

Fragapane P, Prislei S, Michienzi A, Caffarelli E, Bozzoni I. 1993. A novel small nucleolar RNA (U16) is encoded inside a ribosomal protein intron and originates by processing of the pre-mRNA. *EMBO J 12*:2921–2928.

Ganot P, Bortolin ML, Kiss T. 1997a. Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell 89*:799–809.

Ganot P, Caizergues-Ferrer M, Kiss T. 1997b. The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes & Dev 11*:941–956.

Goodall GJ, Wiebauer K, Filipowicz W. 1990. Analysis of pre-mRNA processing in transfected plant protoplasts. *Methods Enzymol 181*:148–161.

Henikoff S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene 28*:351–359.

Hogan NC, Traverse KL, Sullivan DE, Pardue ML. 1994. The nucleus-limited Hsr-Omega-n transcript is a polyadenylated RNA with a regulated intranuclear turnover. *J Cell Biol 125*:21–30.

Hollander MC, Alamo I, Fornace AJ Jr. 1996. A novel DNA damage-inducable transcript, *gadd7*, inhibits cell growth, but lacks a protein product. *Nucleic Acids Res 24*:1589–1593.

Hughes JMX, Ares M Jr. 1991. Depletion of U3 small nucleolar RNA inhibits cleavage in the 5′ external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA. *EMBO J 10*:4231–4239.

Kass S, Tyc K, Steitz JA, Sollner-Webb B. 1990. The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing. *Cell 60*:897–908.

Kiss-László Z, Henry Y, Bachellerie JP, Caizergues-Ferrer M, Kiss T. 1996. Site-specific ribose methylation of preribosomal RNA: A novel function for small nucleolar RNAs. *Cell 85*:1077–1088.

Kiss T, Bortolin ML, Filipowicz W. 1996. Characterization of the intron-encoded U19 RNA, a mammalian small nucleolar RNA that is not associated with fibrillarin. *Mol Cell Biol 16*:1391–1400.

Kiss T, Filipowicz W. 1993. Small nucleolar RNAs encoded by introns of the human cell cycle regulatory gene *RCC1*. *EMBO J 12*:2913–2920.

Kiss T, Filipowicz W. 1995. Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns. *Genes & Dev 9*:1411–1424.

Kloc M, Etkin LD. 1994. Delocalization of Vg1 mRNA from the vegetal cortex in *Xenopus* oocytes after destruction of Xlsirt RNA. *Science 265*:1101–1103.

Kozak M. 1987. An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res 15*:8125–8148.

Kozak M. 1996. Interpreting cDNA sequences: Some insights from studies on translation. *Mamm Genome 7*:563–574.
Leverette RD, Andrews MT, Maxwell ES. 1992. Mouse U14 snRNA is a processed intron of a cognat *hsc70* heat shock pre-messenger RNA. *Cell 71*:1215–1221.
Li HV, Zagorski J, Fournier MJ. 1990. Depletion of U14 small nuclear RNA (snR128) disrupts production of 18S rRNA in *Saccharomyces cervisiae*. *Mol Cell Biol 10*:1145–1152.
Maxwell ES, Fournier MJ. 1995. The small nucleolar RNAs. *Annu Rev Biochem 35*:897–934.
Meller VH, Wu KH, Roman G, Kuroda MI, Davis RL. 1997. *roX1* RNA paints the X chromosome of male *Drosophila* and is regulated by the dosage compensation system. *Cell 88*:445–457.
Meyuhas O, Avni D, Shama S. 1996. Translational control of ribosomal protein mRNAs in eukaryotes. In: Hershey JWB, Mathews MB, Sonnenberg N, eds. *Translational control*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 363–388.
Morrissey JP, Tollervey D. 1993. Yeast snR30 is a small nucleolar RNA required for 18S rRNA synthesis. *Mol Cell Biol 13*:2469–2477.
Mount S, Henikoff S. 1993. Nested genes take flight. *Curr Biol 3*:372–374.
Nakamura A, Amikura R, Mukai M, Kobayashi S, Lasko PF. 1996. Requirement for a noncoding RNA in *Drosophila* polar granules for germ cell establishment. *Science 247*:2075–2079.
Ni J, Tien AL, Fournier MJ. 1977. Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell 89*:565–573.
Nicoloso M, Qu LH, Michot B, Bachellerie JP. 1996. Intron-encoded, antisense small nucleolar RNAs: The characterization of nine novel species points to their direct role as guides for the 2′-O-ribose methylation of rRNAs. *J Mol Biol 260*:178–195.
Ofengand J, Bakin A. 1997. Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaebacteria, mitochondria and chloroplasts. *J Mol Biol 266*:246–268.
Ofengand J, Bakin A, Wrzesinski J, Nurse K, Lane BG. 1995. The pseudouridine residues of ribosomal RNA. *Biochem Cell Biol 73*:915–924.
Peculis BA, Steitz JA. 1993. Disruption of U8 nucleolar snRNA inhibits 5.8S and 28S rRNA processing in the Xenopus oocyte. *Cell 73*:1233–1245.
Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: A laboratory manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
Savino R, Gerbi SA. 1990. In vivo disruption of *Xenopus* U3 snRNA affects ribosomal RNA processing. *EMBO J 9*:2299–2308.
Schmitt ME, Clayton DA. 1993. Nuclear RNase MRP is required for correct processing of pre-5.8S rRNA in *Saccharomyces cerevisiae*. *Mol Cell Biol 13*:7935–7941.
Sollner-Webb B. 1993. Novel intron-encoded small nucleolar RNAs. *Cell 75*:403–405.
Sollner-Webb B, Tycowski KT, Steitz JA. 1996. Ribosomal RNA processing in eukaryotes. In: Zimmermann RA, Dahlberg AE, eds. *Ribosomal RNA. Structure, evolution, processing, and function in protein biosynthesis*. Boca Raton, Florida: CRC Press. pp 469–490.
Steele WJ, Okamura N, Busch H. 1965. Effects of thioacetamide on the composition and biosynthesis of nucleolar and nuclear ribonucleic acid in rat liver. *J Biol Chem 240*:1742–1749.
Tollervey D. 1987. A yeast small nuclear RNA is required for normal processing of pre-ribosomal RNA. *EMBO J 6*:4169–4175.
Tollervey D, Kiss T. 1997. Function and synthesis of small nucleolar RNAs. *Curr Opin Cell Mol Biol 9*:337–342.
Tyc K, Steitz JA. 1989. U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus. *EMBO J 8*:3113–3119.
Tycowski KT, Shu MD, Steitz JA. 1993. A small nucleolar RNA is processed from an intron of the human gene encoding ribosomal protein S3. *Genes & Dev 6*:1120–1130.
Tycowski KT, Shu MD, Steitz JA. 1994. Requirement for intron-encoded U22 small nucleolar RNA in 18S ribosomal RNA maturation. *Science 266*:1558–1561.
Tycowski KT, Shu MD, Steitz JA. 1996. A mammalian gene with introns instead of exons generating stable RNA products. *Nature 379*:464–466.
Velleca MA, Wallace MC, Merlie JP. 1994. A novel synapse-associated noncoding RNA. *Mol Cell Biol 14*:7095–7104.
Watanabe Y, Yamamoto M. 1994. *S. pombe* mei2^{+} encodes an RNA-binding protein essential for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA species meiRNA. *Cell 78*:487–498.
Watkins NJ, Leverette RD, Xia L, Andrews MT, Maxwell ES. 1996. Elements essential for processing intronic U14 snoRNA are located at the termini of the mature snoRNA sequence and include conserved nucleotide boxes C and D. *RNA 2*:118–133.
Wevrick R, Kerns JA, Francke U. 1994. Identification of a novel paternally expressed gene in the Prader–Willi syndrome region. *Hum Mol Gen 3*:1877–1882.