

# RNomics in *Drosophila melanogaster*: identification of 66 candidates for novel non-messenger RNAs

Guozhong Yuan, Christian Klämbt<sup>1</sup>, Jean-Pierre Bachellerie<sup>2</sup>, Jürgen Brosius and Alexander Hüttenhofer\*

Institute for Experimental Pathology (ZMBE), <sup>1</sup>Institute for Zoology, Universität Münster, D-48149 Münster, Germany and <sup>2</sup>Laboratoire de Biologie Moléculaire Eucaryote du CNRS, Université Paul-Sabatier, 118 route de Narbonne, F-31062 Toulouse cedex 04, France

Received February 19, 2003; Revised and Accepted March 17, 2003

## ABSTRACT

By generating a specialised cDNA library from four different developmental stages of *Drosophila melanogaster*, we have identified 66 candidates for small non-messenger RNAs (snmRNAs) and have confirmed their expression by northern blot analysis. Thirteen of them were expressed at certain stages of *D.melanogaster* development, only. Thirty-five species belong to the class of small nucleolar RNAs (snoRNAs), divided into 15 members from the C/D subclass and 20 members from the H/ACA subclass, which mostly guide 2'-O-methylation and pseudouridylation, respectively, of rRNA and snRNAs. These also include two outstanding C/D snoRNAs, U3 and U14, both functioning as pre-rRNA chaperones. Surprisingly, the sequence of the *Drosophila* U14 snoRNA reflects a major change of function of this snoRNA in Diptera relative to yeast and vertebrates. Among the 22 snmRNAs lacking known sequence and structure motifs, five were located in intergenic regions, two in introns, five in untranslated regions of mRNAs, eight were derived from open reading frames, and two were transcribed opposite to an intron. Interestingly, detection of two RNA species from this group implies that certain snmRNA species are processed from alternatively spliced pre-mRNAs. Surprisingly, a few snmRNA sequences could not be found on the published *D.melanogaster* genome, which might suggest that more snmRNA genes (as well as mRNAs) are hidden in unsequenced regions of the genome.

## INTRODUCTION

The genome of *Drosophila melanogaster* exhibits a size of ~180 Mb, a third of which is centric heterochromatin. The 120 Mb of euchromatin sequence maps to two large autosomes and the X chromosome, while the remaining fourth

chromosome contains only ~1 Mb of euchromatin. The heterochromatin consists of short repeated elements, occasionally interrupted by transposable elements, and tandem arrays of rRNA genes (1–3).

The euchromatin portion of the *D.melanogaster* genome has been sequenced recently and was predicted to contain 14 113 mRNAs coding for proteins (1,4). However, besides mRNAs, genomes also encode untranslated RNAs, the smaller ones designated as snmRNAs (small non-messenger RNAs), that function at the level of RNA [for reviews see Mattick (5), Eddy (6) and Hüttenhofer *et al.* (7)]. In addition to tRNAs and small nuclear RNAs (snRNAs), abundant classes of snmRNAs include small nucleolar RNAs (snoRNAs) and microRNAs. Despite their key involvement in a wide range of fundamental cellular processes—such as splicing, modification and processing of RNA, or regulation of translation—the identification of snmRNAs, until recently, was greatly neglected in most genome projects. This is mainly due to the fact that their computational annotation is more difficult to achieve than in the case of mRNAs exhibiting an open reading frame (ORF). No snmRNA genes were annotated in the first report on the identification of the genomic sequence of *D.melanogaster* (1) and even up till now only a few of them have been annotated, apart from tRNAs and snRNAs. For example, from the abundant class of snoRNAs, directing modification within ribosomal or spliceosomal RNAs, only four species have been identified so far, designated as snoRNAs U85 (8), Z30 (GenBank accession no. AJ007735), Z1 (GenBank accession no. U46015) and H1 (9), respectively. In addition, sequences of 16 members from the novel class of miRNAs have been reported recently (10).

We therefore undertook an experimental approach, designated as experimental RNomics, in order to identify candidates for snmRNAs in the fruitfly *D.melanogaster*, which would complement the annotation of protein-coding genes within the genome of this model organism (1). Previously, our approach has been applied successfully to different model organisms such as the mouse *Mus musculus*, the plant *Arabidopsis thaliana* and the archaeon *Archaeoglobus fulgidus* (11–14). By generating a cDNA library encoding snmRNAs from four different developmental stages of

\*To whom correspondence should be addressed at present address: Institute for Molecular Biology, Department of Functional Genomics, Fritz-Pregl-Str. 3, A-6020 Austria. Tel: +43 512 507 3630; Fax: +43 512 507 9880; Email: alexander.huettnerhofer@uibk.ac.at  
Correspondence may also be addressed to Jean-Pierre Bachellerie. Tel: +33 5 61 33 59 34; Fax: +33 5 61 33 58 86; Email: bachel@ibcg.biotoul.fr

**Table 1.** Oligonucleotides (in 5' to 3' orientation) used for reverse transcription of RNAs, PCR amplification and sequencing of cDNA clones

	Sequence
NotI-Adaptor	TCG CGA GCG GCC GCG GGG GGG GGG GG
M13 fsp	GCT ATT ACG CCA GCT GGC GAA AGG GGG ATG TG
M13 rsp	CCC CAG GCT TTA CAC TTT ATG CTT CCG GCT CG
Sequencing primers	
M13 rev	AGC GGA TAA CAA TTT CAC ACA GG
Oligonucleotides used for selective hybridisation	
28s-1	CGCTCCATACACTGCATCTCAC
28s-2	GCTTCTTTATGGTCGTTTCCTGTTG
28s-3	TTCAAATCAAAAATACATAAATGCATCGTT
28s-4	TGTAAGTACGCGCGGCATCAGGTG
28s-5	TAAGCTTCAGTGGATCGCAGTATGG
28s-6	CCGCCACAAGCCAGTTATCCCTATG
18s-1	TAATTACGCGCCTGCTGCCTTCC
18s-2	CGCGCGTGCAGCCAGGACATCT
18s-3	TCCGGAGACCTCACTAAATAATTCAA
18s-4	ACCTTGTTACGACTTTTACTTCTCT
18s-5	ACTCCTGGTGGTGCCTTCCGTCAA
18s-6	ATTTGTCCATTTAAGAAGCTAGTGTCT
5.8s-1	CGACCCATGAGCCGAGTGATCCA
5.8s-2	ATGGACTGCGATATGCGTTCAA
5s-1	GATGTATTCAGCGTGGTATGGTC
5s-2	GTGTTCCCAAGCGGTCCCCATCTA
tRNA-1	CCATACCGGGAATTGAACC
tRNA-2	GCGTGACAGGCGGGGATACT
U1	ATTACACGCACGAGTATTACATT
U2	GAAATCCCGCGGTTACTGCAATA
Oligonucleotides used for Southern blot hybridisation	
Dm-173	CTGCGCGCCATCGCCAAACGAGGCTA
Dm-463	CCATGGCTAGATCACTCGGTTTC
Dm-558	GCGACAGGCAGGACGTTTCCT
Dm-637	CGCCGAAGCGATCAACTTGCA
Dm-682	CTCTTACGGAAGGACGGATC
Dm-705	TGGAGACGCCGGTACTGCC
Dm-782	CCGTGAATCTCCTCCAAGCTG
Oligonucleotides used for PCR amplification of genomic DNA	
Dm-463 upper	CAGCGAGCTTAAGCCGATAGG
Dm-463 lower	CCATGGCTAGATCACTCGGTTTC
Dm-558 upper	TCTTTTACCTTCACTCGTCGTT
Dm-558 lower	GCGACAGGCAGGACGTTTCCT
Dm-637 upper	CCGTGATGGAGGCGGCCGG
Dm-637 lower	CGCCGAAGCGATCAACTTGCA
Dm-682 upper	GATTATGATGCTGCGTCTGC
Dm-682 lower	CTCTTACGGAAGGACGGATC
Dm-705 upper	GGCTTGCCCCTTGGCTATG
Dm-705 lower	TGGAGACGCCGGTACTGCC
Dm-782 upper	GCAGCAGGAGCTGGTGA
Dm-782 lower	CCGTGAATCTCCTCCAAGCTG

*D.melanogaster*, we have identified 111 novel candidates for snmRNAs. By northern blot analysis, we confirmed the presence of a majority of these snmRNA candidates in *D.melanogaster* and investigated their abundance and temporal expression patterns.

## MATERIALS AND METHODS

### Generation of a cDNA library encoding snmRNAs from *D.melanogaster*

Total RNA from four developmental stages of *D.melanogaster*, the embryo, larva, pupa and the adult stage, was prepared by the TRIzol method (Gibco-BRL). Subsequently, total RNA from the four developmental stages (200 µg) was size fractionated on a denaturing 8% polyacrylamide gel (7 M

urea, 1× TBE buffer). RNAs in the size range of ~50 to ~500 nt were excised from the gel, passively eluted and ethanol precipitated. Subsequently, 5 µg of RNA was tailed with CTP using poly(A) polymerase, as described previously (15). RNA was reverse transcribed into cDNA using primer NotI-Adaptor (see Table 1) and cloned into pSPORT 1 vector employing the GIBCO Superscript™ system (Gibco-BRL). cDNAs were amplified by PCR using primers M13 fsp and M13 rsp (Table 1). PCR products were spotted by robots in high density arrays on filters (16), performed at the Resource Center of the German Human Genome Project (Berlin, Germany).

### DNA sequencing and sequence analysis

We sequenced cDNA clones using the M13 rsp reverse primer and the BigDye terminator cycle sequencing reaction kit (PE

Applied Biosystems) on an ABI Prism 3700 (Perkin Elmer) sequencer. We analysed sequences with the LASERGENE sequence analysis program package. After exclusion of the most abundant known snmRNAs by filter hybridisation screening (see below), about 5000 cDNA clones were analysed by sequencing and compared with each other using the Lasergene Seqman II program package to identify identical clones. Following a BLASTN search from the GenBank database (NCBI), all sequences which had not been annotated in GenBank previously were treated as potential candidates for novel snmRNAs.

#### Filter hybridisation and identification of clones encoding novel snmRNAs

For exclusion of the most abundant, known, small RNA species, we end-labelled oligonucleotides (see Table 1) derived from these sequences with [ $\gamma$ - $^{32}$ P]ATP and T4 polynucleotide kinase and hybridised oligonucleotides to DNA arrays which were spotted on filters (see above). We performed hybridisation in 0.5 M sodium phosphate pH 7.2, 7% SDS, 1 mM EDTA at 53°C for 12 h. Filters were washed (twice, at room temperature for 15 min in 40 mM sodium phosphate buffer pH 7.2, 0.1 % SDS), exposed to a phosphor-imaging screen and analysed by computer-aided treatment of hybridisation signals (Raytest, Germany), according to Maier *et al.* (17).

#### Northern blot analysis

Total RNA from *D.melanogaster* was separated on 8% denaturing polyacrylamide gels (7 M urea, 1× TBE buffer) and transferred onto nylon membranes (Qiabrane Nylon Plus, Qiagen, Germany) using the Biorad semi-dry blotting apparatus (Trans-blot SD, Biorad, Germany). After immobilising RNAs using the STRATAGENE cross-linker, we pre-hybridised nylon membranes for 1 h in 1 M sodium phosphate buffer pH 6.2, 7% SDS. Oligonucleotides complementary to potential novel RNA species were end-labelled with [ $\gamma$ - $^{32}$ P]ATP and T4 polynucleotide kinase; hybridisation was carried out at 58°C in 1 M sodium phosphate buffer pH 6.2, 7% SDS for 12 h. We washed blots twice at room temperature in 2× SSC buffer (20 mM sodium phosphate pH 7.4, 0.3 M NaCl, 2 mM EDTA), 0.1% SDS for 15 min and subsequently at 58°C in 0.1× SSC, 0.5% SDS for 1 min. Membranes were exposed to Kodak MS-1 film for 35 min to 5 days.

#### Oligonucleotides

Oligonucleotides used for reverse transcription of RNAs, PC amplification and sequencing of cDNA clones are listed in Table 1.

#### Folding prediction of snmRNA structures

Secondary structures of snmRNA candidates were predicted using the mfold version 3.0 by Zuker and Turner (Department of Mathematical Sciences, 331 Amos Eaton Hall, Rensselaer Polytechnic Institute, Troy, NY 12180-3590).

#### Analysis of snoRNA sequences

*Box C/D snoRNAs.* Presumptive antisense elements at least 9 nt extending from position -2 upstream from box D (or box D') were searched for any portion of the *D.melanogaster* rRNAs or spliceosomal U snRNAs, using the DNAMAN

program. Matches including not more than one mismatched pair (18) were selected, as listed in Table 2, and the positions presumably targeted for 2'-O-methylation (i.e. paired to the fifth nucleotide upstream from box D or D' in each RNA duplex) were examined further by reference to the location of 2'-O-ribose methylations mapped in eukaryal rRNAs or snRNAs (19,20). In the snoRNA 5' half, presumptive antisense elements were examined upstream from all possible box D' motifs (i.e. with maximally 1 nt deviation from the canonical CUGA) upstream from a box C' motif (5'UGA-UGA3' with maximally 2 nt deviations). For a few snoRNA sequences with severe 5' truncations, analysis of the snoRNA 5' half could generally be achieved by examination of the predicted complete snoRNA sequence found in the *D.melanogaster* genome. In such cases, 5' ends of the snoRNA sequences were generally identified through the presence of the hallmark 5'-3' terminal stem (of 4-5 bp)-box C/D structure.

*H/ACA snoRNA.* After delineation of the H box motif, ANANNA, in the central region of the snoRNA sequence (21), the upstream and downstream portions of each snoRNA sequence were folded separately using the Zuker program (see above). Stable structures displaying the stem-internal-loop-stem organisation typical of each H/ACA snoRNA basic domain were examined further. Those in which the uppermost nucleotide in the 3' strand of the large internal loop is located 13-16 nt upstream from box H or ACA were searched for the presence of a potential bipartite antisense element at the appropriate location within the large internal loop (22). Presumptive uridine targets were identified by the presence of a bipartite duplex (of at least 9 bp) made up by two stems (of at least 4 bp) flanking the targeted uridine and the adjacent downstream nucleotide. Matches with rRNA or snRNA sequences corresponding to canonical guide duplexes (shown in Fig. 6) were examined further by reference to known sites of pseudouridylation in these RNAs in various eukaryotic organisms.

## RESULTS AND DISCUSSION

### Construction and analysis of a cDNA library encoding snmRNAs from *D.melanogaster*

We generated a cDNA library from the fruitfly *D.melanogaster* encoding snmRNAs sized from ~50 to 500 nt. Total RNA was prepared from four developmental stages of *D.melanogaster*, the embryo, pupa, larva and the adult. The RNA was size fractionated on a denaturing 8% polyacrylamide gel, tailed with CTP, reverse transcribed into cDNA and cloned into plasmid pSPORT1 (see Materials and Methods). To exclude the most abundant known snmRNAs, we spotted about 30 000 clones on filters in high-density arrays. Subsequently, we hybridised them to labelled oligonucleotide probes derived from known abundant snmRNAs or large rRNAs to reduce their frequent degradation products. By this approach, we could increase the number of cDNAs encoding potential candidates for snmRNAs from 6.3% (before selection) to 20.5% (after selection; Fig. 1). Subsequently, 5300 cDNA clones exhibiting the lowest hybridisation scores were sequenced. The resulting 111 unique sequences that did not

**Table 2.** Compilation of novel expressed RNA sequences from Class I from a *D.melanogaster* cDNA library derived from RNAs sized 50–500 nt

ERNS	cDNA clones	cDNA (nt)	Northern blot (nt)	Abund.	Ch.	Location	Comments	Devel. R	Acc. No.
Group 1: C/D box snoRNAs							Modification		
rRNA chaperones									
Dm-396	1	51	80	++	2L	IR	U14 snoRNA, two gene copies	+	AJ543993
Dm-818	143	173	140, 200	+++	2R	IR	U3 snoRNA, two genomic copies in opposite orientation related to Dm-830	–	AJ543994
Dm-830	27	168	170, 250	+++	X	IR	U3 snoRNA, related to Dm-818	–	AJ543995
Target: rRNAs									
Dm-122	1	68	70	++	3R	Intron	Am1374/18S	–	AJ543996
Dm-229	10	140	140	+	2R	Intron	Um1906/18S; Um1970/18S	–	AJ543997
Dm-442	1	46	110	++	3R	Intron	Am2564/28S	+	AJ543998
Dm-700	1	49	88	+	X	Intron	Cm2933/28S	–	AJ543999
Dm-737	1	72	80	+	2L	Intron	Universally conserved Cm1813/18S	+	AJ544000
Dm-797	1	47	75	++	2R	Intron	Am1322/28S snoRNAZ1 in GenBank	+	AJ544001
Target: snRNAs									
Dm-755	3	111	140	++	3L	Intron	Um42/U5	+	AJ544002
Without known target									
Dm-185	5	55	125, 150	+++	2R	IR		–	AJ544003
Dm-284	2	67	75	++	X	Intron		–	AJ544004
Dm-291	2	74	70, 105	+++	3R	Intron		–	AJ544005
Dm-461	3	102	130	++	2R	IR		–	AJ544006
Dm-684	1	78	155	+	3L	Intron		–	AJ544007
Group 2: H/ACA box snoRNAs									
Target: rRNAs									
Dm-3	1	16	147	++	3L	Intron	Ψ2562/28S; Ψ2622/18S	–	AJ544008
Dm-14	4	108	130	++	3L	Intron	U3801/28S	–	AJ544009
Dm-66	1	140	157	+	2R	Intron	ψ841/18S	–	AJ544010
Dm-72	5	86	162	+++	2L	Intron	Ψ2648/28S	–	AJ544011
Dm-75	1	38	141	++	3L	Intron	Ψ3436/28S	–	AJ544012
Dm-143	6	123	110, 151	+++	3R	Intron	U2149/28S	–	AJ544013
Dm-203	2	53	145	+	2R	Intron	U1347/18S; Ψ1313/28S	+	AJ544014
Dm-227	1	29	140	++	2R	IR	ψ1205/18S	–	AJ544015
Dm-269	2	183	410	+	3L	Intron	Ψ2566/28S, 3 stem snoRNA, universally conserved	–	AJ544016
Dm-328	2	68	130	+	3R	Intron	U176/18S; ψ1377/18S	–	AJ544017
Dm-535	2	43	141	++	2R	Intron	U1930/28S; U760/18S	–	AJ544018
Dm-639	2	128	145	++	2R	Intron	Ψ1816–1818/18S snoRNAH1 in GenBank	–	AJ544019
Dm-644	1	81	170	+++	X	IR	U1353/18S; Ψ3342/28S	–	AJ544020
Dm-708	1	53	147	++	X	Intron	ψ1205/18S	–	AJ544021
Dm-734	2	133	145	+	1	Intron	Ψ2719/28S; ψ210/18S	–	AJ544022
Dm-783	1	54	140	+	3R	Intron	ψ1275/18S	–	AJ544023
Dm-825	1	34	135	+	X	Intron	U2876/28S	–	AJ544024
Target: rRNAs + snRNAs									
Dm-165	5	53	143	+++	3R	Intron	U59/U4; Ψ3186/28S	–	AJ544025
Dm-586	2	80	141	+	3L	Intron	U1920/18S; U1505/18S; U703/28S; U22/U1	–	AJ544026
Without known target									
Dm-50	5	160	220	++	3R	Intron		+	AJ544027
Group 3: H/ACA-snoRNA-like									
Dm-314	2	64	130	++	2R	IR	H/ACA-snoRNA-like structure	–	AJ544028
Dm-660	1	96	165	+	2R	3'-UTR	H/ACA-snoRNA-like structure	–	AJ544029

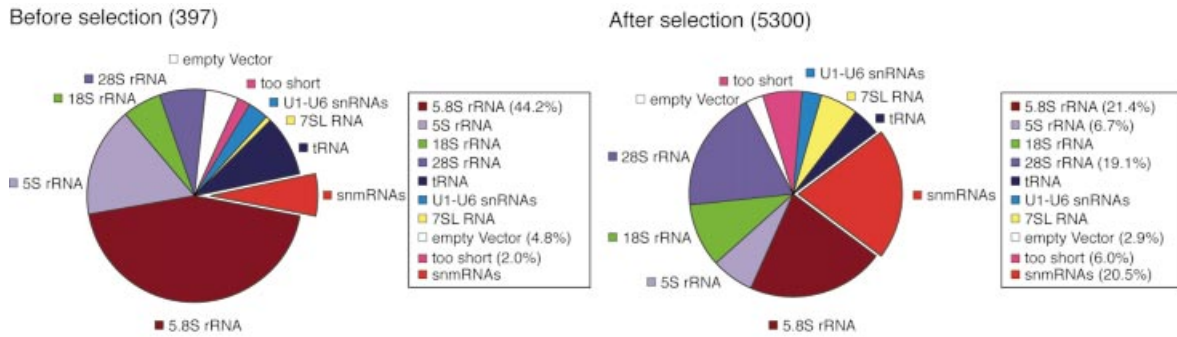
ERNS, expressed RNA sequences (Dm-); cDNA clones, number of independent cDNA clones identified from each RNA species; cDNA (nt), length of cDNA encoding an snmRNA candidate, as assessed by sequencing; Northern blot (nt), length of RNA as assessed by northern blot analysis; Abund., relative abundance of snmRNA as estimated by northern blot analysis; Ch., chromosomal location of snmRNA gene (L and R refer to the right or left arm of respective chromosome; X, X-chromosome); Location, location of snmRNA gene in intron, IR (intergenic region) or UTR (untranslated region); Comments, comments related to the proposed function of snmRNA; in the case of Class I/Groups 1 and 2 snmRNAs, modification refers to predicted modified nucleotides within rRNAs or snRNAs, ψ (lower case) indicates pseudouridines not experimentally tested in *D.melanogaster*, but confirmed in yeast and/or vertebrates, Ψ (bold, upper case) indicates pseudouridines experimentally verified in *D.melanogaster*, and U refers to a targeted uridine not shown to be pseudouridylated in yeast, vertebrates or *D.melanogaster*; Devel. R., developmentally regulated snmRNAs are indicated by '+'; Acc. No., accession number of the snmRNA sequence in the DDBJ/EMBL/GenBank database.

map to a previously annotated gene were considered as potential candidates for novel snmRNAs.

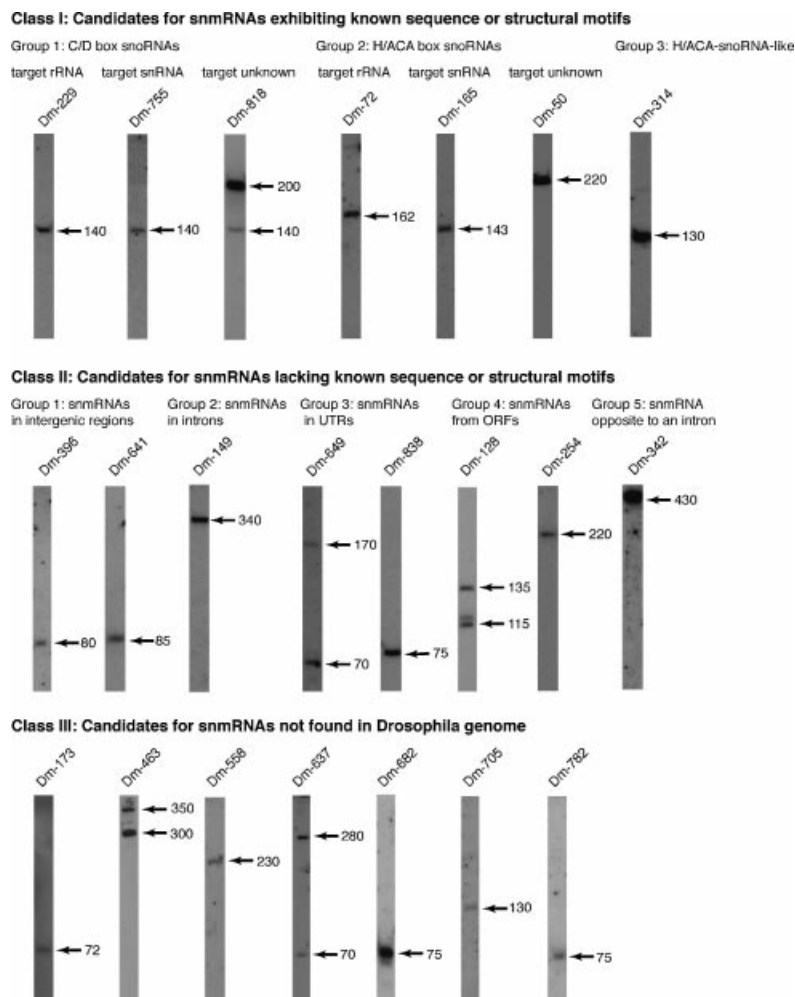
### Expression analysis of novel candidates for snmRNAs

To confirm and analyse their expression, we performed northern blots on all 111 candidates for snmRNAs (for selected examples see Fig. 2). By this method, the expression

of 66 snmRNA species could be confirmed, while expression of the remaining 45 snmRNA candidates was not detected. From the 66 snmRNAs confirmed by northern blotting, we investigated whether snmRNAs were expressed preferentially at certain stages of fly development (e.g. the embryo, larva, pupa and adult stage). In fact, we could detect 13 snmRNA species, whose expression appeared developmentally



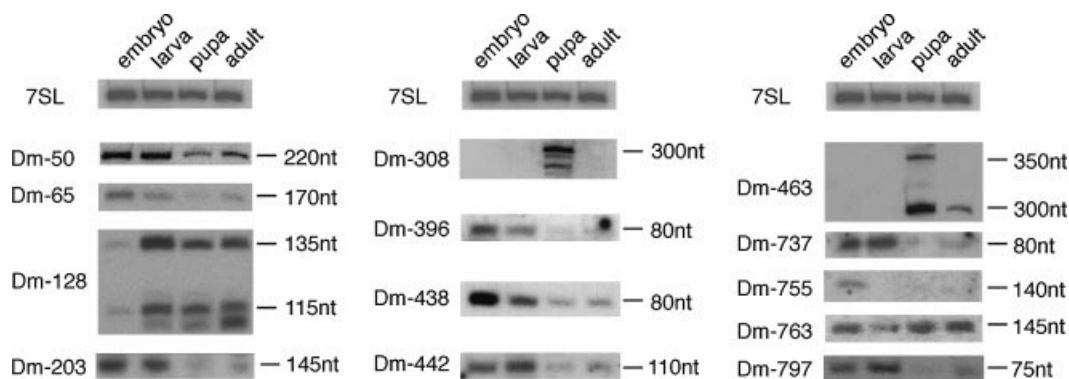
**Figure 1.** Sequence analysis of cDNA clones from the *D.melanogaster* cDNA library before (left) and after (right) selective hybridisation, amounting to 397 and 5300 randomly chosen cDNA clones, respectively. cDNA clones representing different RNA species or categories are shown as a percentage of total clones. The sector denoted snmRNAs identifies candidates for novel snmRNAs. 'Too short' refers to cDNA clones with an insert length below 12 nt which precluded analysis by BLASTN database search.



**Figure 2.** A selection of expressed candidates for novel snmRNAs in *D.melanogaster*, as deduced by northern blot analysis (designated as Class I, II and III snmRNAs). Clone names for each snmRNA are indicated above each lane; sizes of RNAs, as estimated by comparison with an internal RNA marker, are indicated by arrows on the right.

regulated (Fig. 3). In some instances, snmRNAs were expressed almost exclusively at one or two stages, while the expression of others decreased from the embryo to the adult stage (Fig. 3). Interestingly, some RNA species from the class of snoRNAs also showed developmentally regulated

expression, which is, to date, unprecedented (see Table 2). In plants, another class of snmRNAs, designated as microRNAs, has recently been shown to be developmentally regulated and suggested to regulate expression of protein-coding genes involved in plant development (23). It remains to be tested



**Figure 3.** Northern blot analysis showing developmentally regulated expression of Class I, II or III snmRNA candidates; the clone designation is indicated on the left of each panel, the estimated size of the snmRNA is indicated on the right; the four selected developmental stages are: embryo, larva, pupa and adult. As an internal control, ubiquitous expression of 7SL RNA is shown at the top.

whether developmentally regulated snmRNA species identified in our screen serve a similar function in *D.melanogaster*.

In most cases, as observed for other libraries, sizes of RNAs as determined by northern blot analysis were somewhat larger compared with the sizes of the respective cDNAs. This can be attributed to the library construction strategy (see Materials and Methods), which interferes with cloning of the very 5' ends of novel snmRNA species, or to the fact that RNA structure or modification impeded a complete conversion of the snmRNA into cDNA. Hence, the sequences resemble ESTs (expressed sequence tags) of mRNAs, and we therefore designated them as ERNS (expressed RNA sequence) in our tables (Tables 2–4 and Table S1 available as Supplementary Material).

We grouped candidates for novel snmRNAs that could be identified as small stable RNA species on northern blots based on known sequence or structure motifs (Table 2, Class I snmRNAs) or the lack of these (Table 3, Class II snmRNAs; see Fig. 4). In Table 4 (Class III snmRNAs), we list all snmRNAs from *D.melanogaster* for which no corresponding sequence could be found in the reported *D.melanogaster* genome but for which a northern blot signal could be obtained. In Table S1, we list all snmRNA candidates which lack expression as assessed by northern blot analysis (Class IV snmRNAs). For these candidates, we cannot exclude the possibility that they represent spurious low-level transcription products from the *D.melanogaster* genome, which lack (a) cellular function(s). Hence, they will be not discussed in detail, but sequence information on these clones can be accessed in the Supplementary Material, Results and Discussion, Table S1A and B.

#### Class I: snmRNAs exhibiting known sequence and structure motifs

A large fraction of snmRNAs in Eukarya identified so far corresponds to members of the two expanding subclasses of snoRNAs, termed C/D and H/ACA snoRNAs, that guide 2'-O-methylation and pseudouridylation, respectively, in rRNAs and snRNAs (24). From our library, a total of 37 cDNA sequences could be assigned to either snoRNA subclass, based on the presence of hallmark sequence and structural motifs (Table 2; for northern blot analysis of selected examples see Fig. 2).

**Group I: C/D box snoRNAs.** Three snmRNAs in this group correspond to two outstanding C/D snoRNAs, U3 and U14, which act as chaperones for pre-rRNA folding rather than methylation guides [reviewed in Gerbi *et al.* (25)]. These *Drosophila* snmRNAs exhibit a very low level of overall sequence conservation with previously reported U3 and U14 snoRNAs from other species, and hence could not be identified by BlastN analyses. Dm-830 (encoded by two identical genomic copies 2030 bp apart in opposite orientation) and closely related Dm-818 (on a different chromosome) are both encoded in intergenic regions, unlike other bona fide *Drosophila* modification guide snoRNAs (see below). Dm-818 and Dm-830 both display all U3 structural hallmarks, with conserved boxes GAC, A, A', B, C, C' and D found at the expected location within a two-domain secondary structure typical of U3 snoRNA (Fig. 5A). Remarkably, the portion of the large 3' domain extending from the 11 bp central stem folds into a single hairpin supported by compensatory changes among the two *Drosophila* U3 isoforms, whereas the corresponding U3 region in vertebrates and yeast folds into two and three hairpins, respectively (26,27). In contrast, in *Tetrahymena* U3 snoRNA, no hairpin is observed. Thus, U3 snoRNA in *D.melanogaster* reflects an intermediate case between vertebrates and *Tetrahymena* U3 snoRNA. U3 RNA is part of a large RNP including at least 28 different proteins, the small subunit processome required for nucleolar processing of pre-18S rRNA (28). The processome complexity reflects multiple functions of U3 which apparently serves as a chaperone for pre-rRNA folding and is also important for cleavage of several sites within pre-rRNA through series of dynamic base-paired interactions (25–27,29). Although details of U3 functions in the processome remain largely elusive, there appear to be differences among yeast and vertebrates in the function of various U3 structural elements for rRNA processing.

Dm-396, encoded by two closely linked genomic copies in an intergenic region, corresponds to a U14 snoRNA-like RNA. In yeast and vertebrates, U14 is unique among known C/D snoRNAs due to its dual functions of pre-rRNA chaperone and methylation guide which are both mediated by two conserved antisense elements to 18S rRNA, located in its 5' and 3' halves, respectively (30). Domain A, the 13 nt long complementarity mediating the essential chaperone function

**Table 3.** Compilation of novel expressed RNA sequences from Class II from a *D.melanogaster* cDNA library derived from RNAs sized 50–500 nt

ERNS	cDNA clones	cDNA (nt)	Northern blot (nt)	Abund.	Ch.	Location	Comments	Devel. R	Acc. No.
Group 1: snmRNAs in intergenic regions									
Dm-158	1	157	170	+	2L			–	AJ544030
Dm-331	1	112	75, 445	+	3R			–	AJ544031
Dm-400	2	52	70	+++	X			–	AJ544032
Dm-430	1	36	440	++	2L			–	AJ544033
Dm-641	1	143	85	++	3L			–	AJ544034
Group 2: snmRNAs in introns									
Dm-157	1	43	70	+	3R		tRNA-like snmRNA	–	AJ544036
Dm-184	1	35	80	+	2R			–	AJ544037
Group 3: snmRNAs in UTRs									
Dm-204	1	35	450	+++	X	3'-UTR	CG32638	–	AJ544038
Dm-359	1	34	75	+	2L	3'-UTR	Paxillin-derived LIM-only protein	–	AJ544039
Dm-438	1	64	80	++	3L	3'-UTR	Klarsicht protein	+	AJ544040
Dm-649	1	59	170, 70	++	3R	3'-UTR	CG5663	–	AJ544041
Dm-838	1	73	75	++	3L	3'-UTR	GH15083	–	AJ544042
Group 4: snmRNAs from ORFs									
Dm-65	1	69	170	++	3L		CG8588 (ORF: 1959 nt) generated by alternative splicing	+	AJ544043
Dm-128	1	63	115, 135	+++	2L		CG31649 (ORF: 12072 nt)	+	AJ544044
Dm-254	2	31	220	++	2L		Ef2b (ORF: 2233 nt) overlapping ORF by 160 nt	–	AJ544045
Dm-308	1	78	300	+	3L		CG14998 (ORF: 2485 nt) generated by alternative splicing, expressed only in pupa	+	AJ544046
Dm-357	1	77	170	+++	3R		Engrailed nuclear homeoprotein-regulated gene (ORF: 1827 nt) with canonical H, ACA boxes	–	AJ544047
Dm-419	2	74	151	++	3R		CG4115 (ORF: 603 nt) overlapping by 88 nt	–	AJ544048
Dm-763	2	95	145	++	X		CG8051 (ORF: 1849 nt)	+	AJ544049
Dm-765	2	46	135	++	2L		CG31649 (ORF: 11789 nt)	–	AJ544050
Group 5: snmRNA opposite to introns									
Dm-149	5	67	340	++	3L		Opposite to intron 1 of CG1079	–	AJ544035
Dm-342	4	232	430	++	3R		Opposite to intron 2 of CG10365	–	AJ544051

For an explanation of the column headings, see footnotes to Table 2. For snmRNAs derived from coding regions, the name or accession number of the host gene is indicated; ORF refers to the length of the predicted or confirmed open reading frame (in nucleotides).

**Table 4.** Compilation of novel expressed RNA sequences from Class III from a *D.melanogaster* cDNA library derived from RNAs sized 50–500 nt which are not found in the annotated genomic sequence of *D.melanogaster*

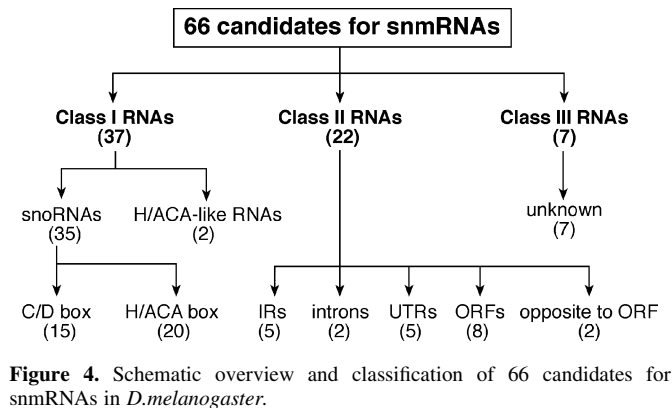
ERNS	cDNA clones	cDNA (nt)	Northern blot (nt)	Abund.	PCR	Southern blot	Devel. R	Acc. No.
Dm-173	2	31	72	+	ND	++	–	AJ544052
Dm-463	4	129	300, 350	++	++	+	+	AJ544053
Dm-558	2	63	230	+	–	–	–	AJ544054
Dm-637	1	66	70, 280	+	–	–	–	AJ544055
Dm-682	1	116	75	++	+	–	–	AJ544056
Dm-705	1	78	130	+	++	+	–	AJ544057
Dm-782	1	110	75	+	–	–	–	AJ544058

ERNS, expressed RNA sequences (Dm-); cDNA clones, number of independent cDNA clones identified from each RNA species; cDNA (nt), length of cDNA encoding a snmRNA candidate, as assessed by sequencing; Northern blot (nt), length of RNA as assessed by northern blot analysis; Abund., relative abundance of snmRNA as estimated by northern blot analysis; PCR, relative abundance of a DNA fragment (exhibiting the expected size) generated by PCR from genomic DNA of *D.melanogaster* by using oligonucleotide primers from the snmRNA candidate sequence; Southern blot, intensity of hybridisation signal generated by Southern blotting of genomic DNA from *D.melanogaster* using an end-labelled oligonucleotide derived from the respective snmRNA sequence; Devel. R., developmentally regulated snmRNAs are indicated by '+' (see also Fig. 3); Acc. No., accession number of snmRNA sequence in the DDBJ/EMBL/GenBank database.

of yeast and vertebrate U14, is perfectly conserved in Dm-396. However, Dm-396 does not contain domain B, the 14 nt long 3' antisense element of yeast and vertebrate U14, which guides formation of Cm462 in 18S rRNA (human coordinates). Dm-396 does not exhibit any potential 3' antisense element for an alternative rRNA or snRNA target. Conversely, we have identified by genomic search in both *D.melanogaster* and *Anopheles gambiae* a typical, intron-encoded C/D snoRNA able to direct Cm462 in their respective 18S rRNA (data not shown). This strongly suggests that the dual chaperone and

guide functions of the single U14 snoRNA in yeast *Saccharomyces cerevisiae* and vertebrates are mediated by two distinct snoRNA species in Diptera. Clearly, it has been taken over in Diptera by a distinct snoRNA, as confirmed by analysis of the *A.gambiae* genome (data not shown). Dm-396 provides the opportunity to assess better the separate chaperone function of this snoRNA in the biogenesis of small ribosomal subunits in the *Drosophila* context.

Except for U3, U14 and a few additional specimens, C/D snoRNAs generally function as methylation guides (24,31). In

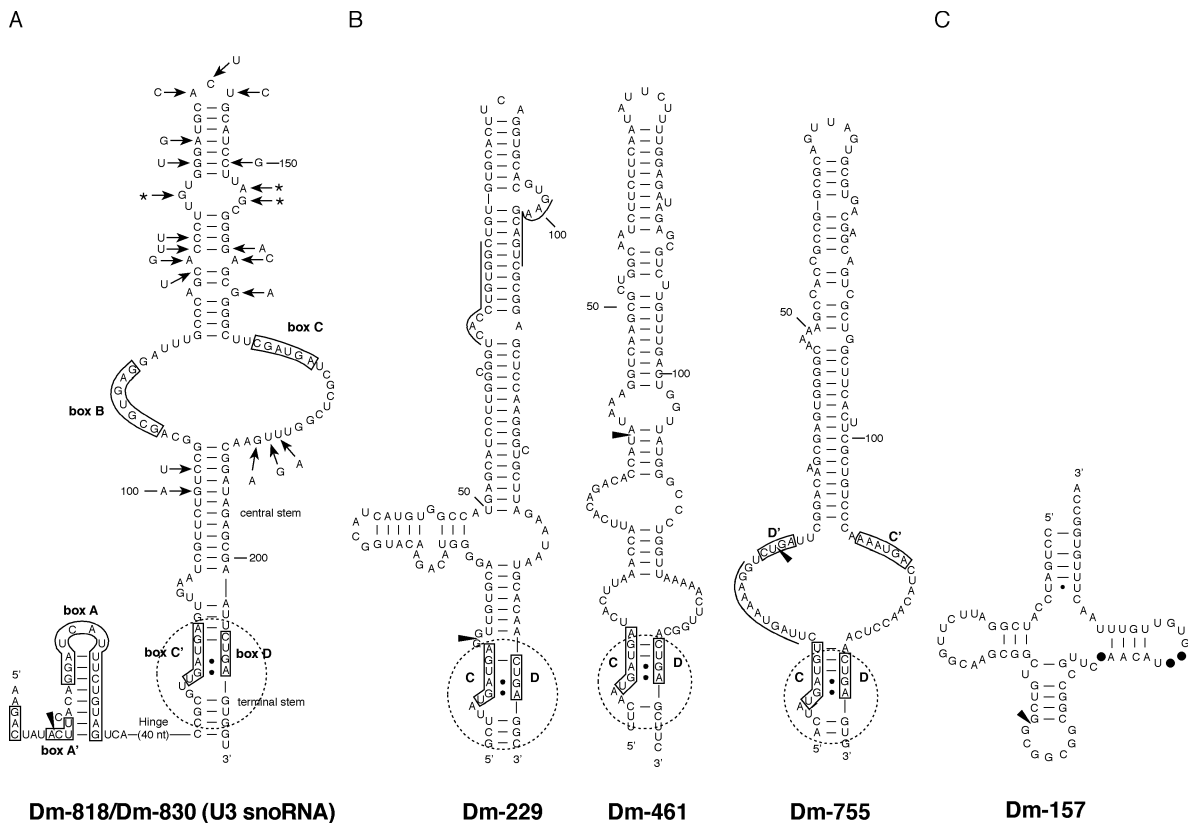


**Figure 4.** Schematic overview and classification of 66 candidates for snmRNAs in *D.melanogaster*.

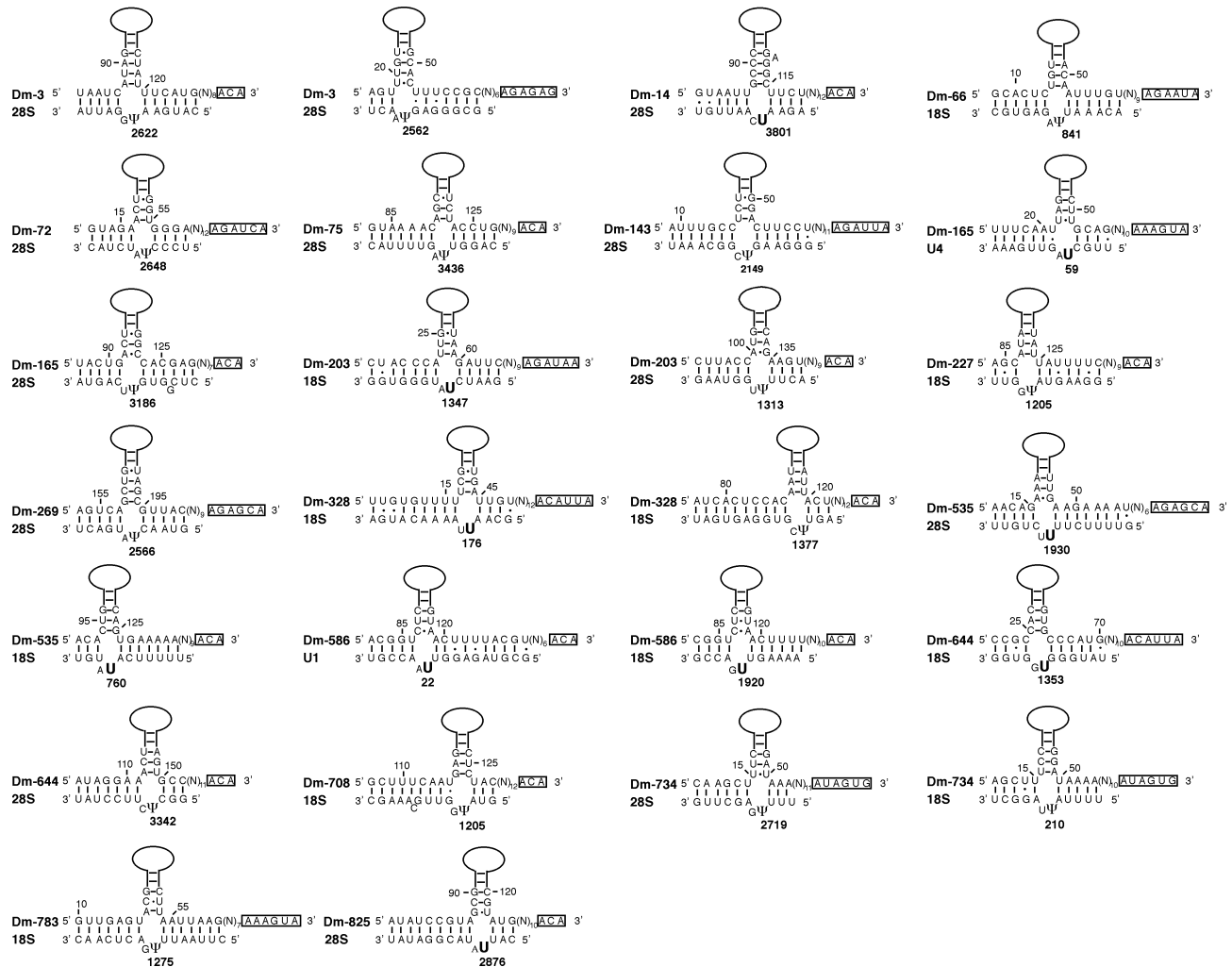
our screen, we have also identified six novel C/D snoRNAs predicted to guide 2'-O-methylations in rRNAs or snRNAs (Table 2, Group 1). All of them are intron encoded, as is the case with vertebrate modification guide snoRNAs. Collectively, they can target a total of six rRNA nucleotides and a single snRNA nucleotide in U5 snRNA. Although rRNA 2'-O-methylations have not been tested experimentally in *Drosophila*, two of the predicted sites, presumably targeted by Dm-442 and Dm-737, correspond to conserved rRNA 2'-O-methylations, which are guided in vertebrates by C/D

snoRNAs U40 and U43, respectively (32). However, sequence similarity of Dm-442 and Dm-737 to vertebrate U40 and U43 is restricted to their rRNA antisense elements only. While Dm-442 and its vertebrate homologue U40 are encoded in introns of the same host gene in the *Drosophila* and vertebrate genomes, Dm-737 and vertebrate U43 are hosted by different ribosomal protein genes.

Dm-755 is able to direct formation of Um42 in U5 snRNA, a phylogenetically conserved methylation, which has been verified experimentally in *D.melanogaster* (33,34). In vertebrates, this methylation is directed by U87 and its variant U88, which both localise to the Cajal bodies, not the nucleolus, like the subset of RNA guides directing modifications of RNA polymerase II-transcribed snRNAs (35). Vertebrate U87 and U88 both belong to a subclass of extended, composite C/D-H/ACA guide RNAs (36). Interestingly, while the D'-C' box interval of Dm-755 is dramatically expanded as compared with typical C/D snoRNAs in vertebrates, yeast or plants, it lacks the two central H/ACA domains present in vertebrate U87 and U88 and can fold instead into a single hairpin encompassing ~80 nt (Fig. 5B). Dm-229, which can direct two 18S rRNA methylations through two separate 5' antisense elements, corresponds to another unusually long C/D specimen devoid of H/ACA hallmarks exhibiting an atypical single hairpin of ~80 nt in its central region (Fig. 5B).



**Figure 5.** Secondary structure predictions of selected snmRNA candidates from *D.melanogaster*. The 5' end of cDNA clones encoding respective snmRNAs is indicated by an arrowhead. (A) Structure of *Drosophila* C/D snoRNA U3: box motifs are shown as previously delineated for yeast *S.cerevisiae* and amphibian *X.laevis* U3 snoRNAs (25). The sequence shown is for Dm-830 (differences in the Dm-818 isoform are denoted by arrows). (B) Additional C/D snoRNAs with long hairpins: the C and D motifs are boxed, and the K-turn structural motifs are delineated by a broken circle. For Dm-229 and Dm-755, antisense elements to 18S rRNA and U5 snRNA, respectively, are overlined. (C) Dm-157, an intron-encoded tRNA-like snmRNA; canonical tRNA nucleotides missing in Dm-157 are indicated by filled circles.



**Figure 6.** Potential base pairing interactions between novel H/ACA snoRNAs from *D.melanogaster* and rRNAs or snRNAs. The snoRNA sequences in a 5' to 3' orientation are shown in the upper strands, with the two H and ACA motifs boxed and the apical part of the snoRNA 5' or 3' hairpin domains schematised by a solid line. The two complementarities to the RNA target are always found within the large internal loop of one (or both) of their hairpin domains, invariably abutting its apex-proximal stem. The positions of pseudouridines are indicated by numbers. Known pseudouridines in *D.melanogaster* and other species, predicted to be directed by the respective snoRNA, are denoted by Ψ. U indicates a uridine at the expected target position in the canonical, bipartite guide RNA duplex that has not been previously reported to be pseudouridylated.

Two *D.melanogaster* genomic loci encoding clusters of candidate methylation guide C/D snoRNAs have been identified by a genome search (37). The larger one, spanning ~4 kb of DNA, contains a total of 16 predicted C/D snoRNA genes encoding homologues of vertebrate U27, U29, U31 and U76 and yeast snR38, while the smaller one contains a single gene copy for homologues of vertebrate U15 and U25, respectively, as well as two copies of a U14 homologue. Although all 20 snoRNA sequences, located in introns of two non-protein-coding host genes, are likely to correspond to functional snoRNAs, these remain to be verified experimentally. Surprisingly, in our library, only one of them was represented, the U14 homologue mentioned above.

Finally, we have detected five additional sequences with C/D snoRNA hallmarks but devoid of typical antisense elements to rRNA or snRNAs. Three of them, intron-encoded Dm-284, Dm-291 and Dm-684, might target as yet unknown cellular RNAs, as proposed for an expanding number of

orphan C/D and H/ACA snoRNAs recently identified in vertebrates and plants (7,24,35). Conversely, Dm-461, which exhibits a very large single-hairpin structure extending from a K-turn motif (Fig. 5B), and Dm-185 both map within intergenic regions of the genome, probably encoded by independent snoRNA genes, suggesting they might have functions more complex than, or even entirely distinct from bona fide methylation guides.

**Group 2: H/ACA snoRNAs.** Twenty snmRNA candidates with signals in northern hybridisation analyses were assigned to the H/ACA snoRNA group (Table 2). Among them was snoRNA H1, the sole previously reported *D.melanogaster* snoRNA of this type (9). All but one of the novel *Drosophila* H/ACA snoRNAs are predicted to direct the isomerisation of at least one rRNA or snRNA uridine, based on their ability to form a canonical, bipartite guide duplex of at least 9 bp around the uridine to be modified (see Fig. 6). In each case, bipartite

antisense elements are invariably found in the upper part of the internal loop of one of the large hairpin domains, and the target uridine always positioned 13–16 nt from the H or ACA motif, as previously observed in other Eukarya or in Archaea [(22); for reviews see Hüttenhofer *et al.* (7), Bachellerie *et al.* (24) and Kiss (35)].

Altogether, the 19 novel H/ACA snoRNAs are predicted to direct 26 rRNA pseudouridylations (13 in 18S rRNA and 13 in 28S rRNA) and two snRNA pseudouridylations (U1 and U4 snRNAs). Several of the presumptive rRNA pseudouridines have been verified experimentally in *D.melanogaster* 28S rRNA (38), while a substantial fraction of those untested so far in *Drosophila* have been characterised in corresponding positions of *S.cerevisiae* and/or vertebrates (Table 2). Except for two of them, Dm-227 and Dm-644, which map within intergenic and unannotated regions, respectively, all the novel *D.melanogaster* H/ACA snoRNAs are intron encoded. This is similar to the situation in vertebrates and unlike the majority of both types of modification guide snoRNAs in *S.cerevisiae* or *A.thaliana* (14,39). The only snmRNA specimen in this group, for which we could not identify any rRNA or snRNA target, Dm-50, is also intron encoded. Intriguingly, in contrast to all but one guide for rRNA or snRNA pseudouridylation, expression of Dm-50 is developmentally regulated. Dm-50 is, however, hosted in a ubiquitously expressed gene, *rpL3*, suggesting that developmental regulation of its expression takes place at the post-transcriptional level.

**Group 3: H/ACA-like snoRNAs.** Two snmRNAs, Dm-314 and Dm-660, mapping to an intergenic region and a 3'-untranslated region (UTR), respectively, both exhibit ACA- and H-like motifs located within a two-domain secondary structure reminiscent of a typical H/ACA folding (21). They do not show any complementarity to ribosomal or spliceosomal RNAs.

#### **Class II: snmRNAs lacking known sequence and structure motifs**

In our screen, we identified a total of 22 snmRNA candidates confirmed by northern signals, which did not exhibit any known sequence or structure motifs which would allow a functional assignment of these snmRNAs (Table 3; for northern blot analysis of selected examples, see Fig. 2). Hence, these snmRNAs were grouped according to their location on the genome, e.g. genes located in intergenic regions (Group 1), in introns (Group 2), in UTRs of (predicted) mRNAs (Group 3), derived from ORFs (Group 4) and opposite to introns (Group 5).

In addition, we have tested the possibility that these snmRNAs might represent mRNAs encoding small ORFs, which have escaped identification by bioinformatical approaches. Therefore, we have analysed all snmRNA sequences for the presence of an ORF, e.g. a start codon as well as a stop codon and a ribosome-binding site element (RBS), namely the Kozak sequence (40). Thereby, we extrapolated the sequence of snmRNAs to the size identified by northern blot analysis, since in most cases our cDNAs lacked mature 5' ends (see above).

From these analyses, on average we observed reading frames within these snmRNAs well below 30 amino acids. The majority of these sequences lacked a stop codon,

however, and in all cases a RBS sequence (e.g. a Kozak sequence, see above) preceding the AUG start codon was missing. Hence, we propose that most, if not all, of our sequences represent snmRNAs rather than mRNAs exhibiting small ORFs.

**Group 1: snmRNAs in intergenic regions.** In general, snmRNA genes located in intergenic regions are the most obvious candidates for bona fide non-messenger RNAs since their location usually does not interfere with the expression of other (protein-coding) genes. Indeed searches for snmRNAs applying bioinformatical methods focus almost exclusively on intergenic regions of genomic sequences. From this group of snmRNAs, we have identified five candidates, one of which is encoded by the X chromosome.

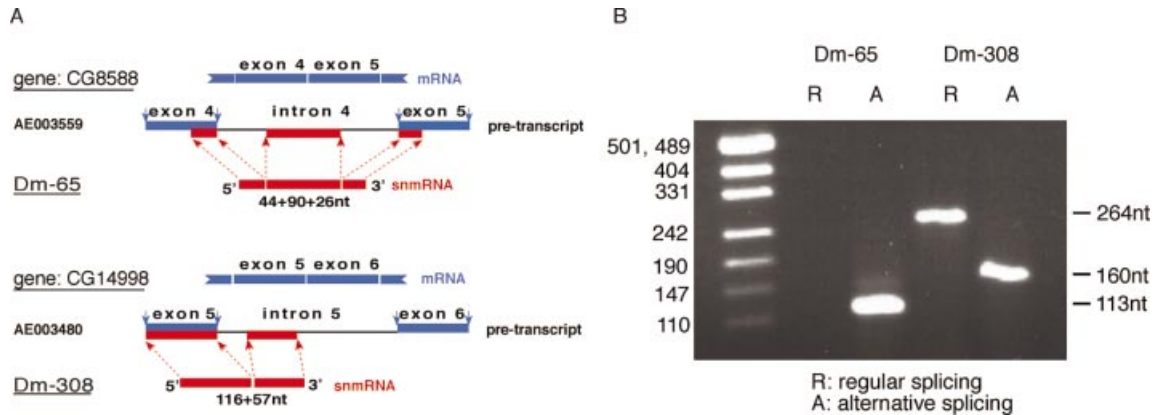
**Group 2: snmRNAs in introns.** Introns are excised from pre-mRNAs by the splicing machinery and subsequently are rapidly degraded. A prominent exception to this rule is snoRNAs (see above) which, in vertebrates, are processed from introns by exo- and endonucleolytic cleavage (31). Also, in some rare cases, eukaryal tRNAs can be found to be encoded within introns (Sean Eddy, personal communication).

We have identified two potential candidates for snmRNAs whose genes are located in introns. One of them, Dm-184, apparently does not belong to either the class of snoRNAs or to the class of tRNAs. The other snmRNA, clone Dm-157, exhibits some tRNA-like features such as a CCA-end and typical secondary structure pattern, but lacks universally conserved features of canonical tRNAs, such as, for example, the GT $\Psi$ C sequence in the T-loop (Fig. 5C). It remains to be tested whether the snmRNA serves some tRNA-like functions *in vivo*.

**Group 3: snmRNAs in UTRs.** The 5'- or 3'-UTRs of mRNAs often contain *cis*-acting regulatory sequences, which are able to influence the expression of the respective ORF of an mRNA. For example, 5'- or 3'-UTRs can serve as binding sites for proteins which in turn regulate translation or stability of ferritin or transferrin receptor mRNAs (41). In addition, the 3'-UTR of developmentally regulated mRNAs is proposed to bind antisense elements of developmentally regulated microRNAs [for a review see Hüttenhofer *et al.* (7)]. Thereby, translation of these mRNAs is repressed by an as yet unknown mechanism.

In addition to *cis*-acting UTR elements of mRNAs, some might be processed from the respective mRNAs, thereby exerting their function *in trans*. Consistent with such a model, five stable snmRNA species, two of them developmentally regulated, have been identified in our screen which are derived exclusively from the 3'-UTRs of protein-coding genes.

**Group 4: snmRNAs from ORFs.** snmRNAs which are derived from ORFs might represent mRNAs or degradation products of these. Therefore, extreme care was taken not to include any false positives in our tables: first, as a prerequisite to being assigned to this group, the length of candidate snmRNA genes had to be considerably smaller than that of the respective predicted mRNA; secondly, as for all other snmRNAs from that class, the expression of these candidates had to be confirmed by northern blot analysis. Still, snmRNAs from



**Figure 7.** Processing of Dm-65 (top) or Dm-308 snmRNA candidates from pre-mRNAs by alternative splicing. (A) Exons of predicted ORFs (1) are indicated by blue bars; sequences of snmRNA candidates are indicated by red bars. The length of the bi- or tri-partite snmRNA sequences (in nt) is indicated below each snmRNA. (B) RT-PCR amplification of the predicted (R) and alternatively spliced (A) mRNAs for Dm-65 and Dm-308 snmRNAs. Sizes of expected DNA fragments are indicated on the right; on the left, a DNA size marker is shown.

ORFs might represent small stable degradation products of mRNAs. Alternatively, the ORF from which they are derived might be wrongly annotated and hence encode an snmRNA instead of an mRNA. In line with that notion, six out of the eight snmRNAs from that group are encoded by hypothetical, unverified ORFs.

Interestingly, two snmRNAs from that group, Dm-65 and Dm-308, are derived from exon as well as non-adjacent intron portions of the respective pre-mRNAs, but lacking the intervening sequences (Fig. 7A). Upon closer inspection of genomic sequences, we observed that the presumptive intron portion in each bi- or tripartite snmRNA sequence was likely to correspond to an alternative, not annotated exon (Fig. 7A). Detection of clones Dm-65 and Dm-308 might therefore reflect utilisation of alternative splice sites (both 5' and 3' splice sites, or only a 3' splice site, respectively) followed by processing of the spliced product into a small stable RNA. Thus, by alternative splicing, two different RNA species, exhibiting different functions, could be processed from one pre-mRNA; an mRNA being translated into a protein as well as an snmRNA functioning at the RNA level.

To test this possibility, we investigated for both clones by RT-PCR which of the two predicted spliced mRNAs were present in total RNA from *D.melanogaster* (Fig. 7B). In fact, from Dm-308, we observed both the regular and the alternatively spliced RNA (labelled 'R' and 'A' in Fig. 7B), while from clone Dm-65 only the alternatively spliced RNA was detected. Interestingly, both snmRNAs are developmentally regulated: Dm-308 is only expressed at the pupa stage of fly development, while Dm-65 is expressed mainly in the embryo (see Fig. 3).

**Group 5: snmRNA opposite to an intron.** Two snmRNAs (Dm-149 and Dm-342) were detected which map entirely to introns of hypothetical pre-mRNAs. However, unlike specimens from Group 4, these snmRNAs are transcribed in the opposite orientation relative to the presumptive protein-coding gene. Five independent cDNA clones were identified for Dm-149, while four were found for Dm-342. One possible function of these snmRNA candidates might be that the antisense RNA regulates splicing of the respective intron by

base pairing to the sense strand in the pre-mRNA or that they act via an RNAi-like mechanism.

### Class III: snmRNA sequences not found in the *D.melanogaster* genome

Surprisingly, we have identified seven snmRNA clones exhibiting significant expression levels, as assessed by northern blot analysis, but whose sequences were not found within the reported *D.melanogaster* genome (Table 4; for northern blot analysis of selected examples see Fig. 2). Since only 120 Mb of euchromatin have been sequenced out of the 180 Mb *D.melanogaster* genome thus far (1), the possibility remains that genes encoding these snmRNAs are located in unsequenced regions of the genome. To investigate this possibility, we performed Southern blot and PCR analysis on genomic DNA from *D.melanogaster*. The analysis was hampered in some cases, however, by the small size of cDNA clones encoding snmRNAs, which did not allow for optimal oligonucleotide primer design.

However, by both Southern and/or PCR approaches, we could verify the presence of four of the seven clones (Dm-173, Dm-463, Dm-682 and Dm-705) on the genome of *D.melanogaster*. This might indicate that more snmRNAs—and even mRNA genes—are still hidden within the unsequenced portions of the *D.melanogaster* genome, mainly comprised of heterochromatin (1). As for the three remaining snmRNAs (Dm-558, Dm-637 and Dm-782) whose presence on the genome could not be confirmed by Southern and/or PCR analysis, this merely might be due to technical reasons (see above). Alternatively, these snmRNAs might be derived from organisms (such as bacteria) associated with *D.melanogaster*. However, by a BLASTN database search, we failed to identify the presence of these sequences within other reported genomes. In addition, the high abundance of these snmRNAs species in the total RNA population, as assessed by northern blot analysis, is inconsistent with a microbial contamination.

### Class IV: snmRNAs lacking northern blot signals

We have also identified 45 RNA species whose expression could not be confirmed by northern blot analysis (see

Table S1). Accordingly, we cannot exclude the possibility that they might be derived from spurious transcription products of the genome rather than represent small stable RNA molecules. In any case, these sequences might be useful since they represent ESTs of the *D.melanogaster* genome. Sequence information on these clones can be accessed in Table S1.

### Conclusion

By generating a cDNA library encoding snmRNAs from the fruitfly *D.melanogaster*, we have identified 66 candidates for novel snmRNA species, thereby considerably extending our knowledge on snmRNAs in *D.melanogaster*. To exclude the possibility that certain snmRNAs are expressed at certain stages of *D.melanogaster* development only, and hence would escape detection by our experimental RNomics approach, we generated the specialised library from total RNA of four stages: the embryo, pupa, larva and adult. By northern blot analysis, we confirmed expression of the 66 snmRNA species; among these, 13 snmRNAs are expressed preferentially or exclusively at certain developmental stages.

From the 66 snmRNAs, 35 belong to the category of snoRNAs. Based on sequence and structure motifs, 15 can be assigned to the subclass of C/D box snoRNAs, while 20 belong to the subclass of H/ACA snoRNAs. Importantly, for the first time, from the subclass of C/D snoRNAs we were able to identify U3 and U14 snoRNAs in *D.melanogaster*. U3, the most abundant snoRNA, is ubiquitous in Eukarya. In addition, the U3 structure exhibits marked differences between yeast and vertebrates as well as between *Drosophila* and vertebrates. Identification of the *Drosophila* specimen should therefore set the stage for better understanding of U3 functions, through experimental approaches exclusively provided by this model organism. Like U3, C/D snoRNA U14 has a chaperone function for pre-rRNA folding (30). In yeast as well as in vertebrates, U14 is remarkable by its additional guide function for an 18S rRNA 2'-O-methylation (32). Unexpectedly, our detection of a *Drosophila* U14-like snoRNA, Dm-396, lacking the potential for guiding the 18S rRNA methylation shows that this additional function is not conserved throughout Metazoa.

In the present study, the number of *Drosophila* C/D snoRNAs assigned to this family, whether or not they are associated with a predicted RNA target, remains significantly lower than in previous RNomics analyses in mouse or *A.thaliana* (12,14). This could reflect a particularly low level of 2'-O-methylations in *Drosophila* rRNA (and snRNAs), unlike what is observed for *D.melanogaster* rRNA pseudo-uridylation (38). Alternatively, methylation guide snoRNAs might be under-represented in our cDNA library, possibly due to unusual structures or nucleotide modifications impeding reverse transcriptase progression. In addition to the identification of 35 snmRNAs from the class of snoRNAs, we also detected two species which showed some resemblance to this group and hence were termed snoRNA-like.

The developmentally regulated expression observed for several of the *D.melanogaster* snoRNAs predicted to guide rRNA or snRNA modifications (Table 2) deserves comment. In mammals, an increasing number of snoRNAs, mostly of the C/D family, exhibit a tissue-specific expression pattern, being expressed mainly in the brain (42,43). However, none of them seems able to direct rRNA or snRNA modifications, in contrast to the developmentally regulated *Drosophila*

specimens. Our finding therefore points for the first time to the tantalising possibility that a subset of rRNA and snRNA modifications might be developmentally regulated.

From the 66 novel snmRNA candidates, 22 were grouped according to their location relative to other genes; they lacked sequence or structure motifs which would have allowed assignment to a known snmRNA class. Two snmRNAs from this group, Dm-65 and Dm-308, are especially noteworthy since they are derived from exon and intron portions of predicted ORFs by alternative splicing. Thus, an snmRNA as well as an mRNA could be processed from a single pre-m/snmRNA transcript by regular and alternative splicing.

Surprisingly, we could identify seven snmRNA candidates whose expression could be confirmed by northern blot analysis, but whose sequences could not be found on the annotated genomic sequence of *D.melanogaster*. From four of these, we could confirm their *D.melanogaster* descent by PCR and Southern blot analysis. Thus, our data imply that more genes (encoded by snmRNAs as well as mRNAs) might be hidden in the as yet unsequenced portions of the *D.melanogaster* genome.

### SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

### NOTE ADDED IN PROOF

After submission of our manuscript we detected that clone Dm-637 (a Class III snmRNA) had already been identified previously as 7SL RNA from yeast; in addition, clone Dm-173 (also a Class III snmRNA, not found on the *D.melanogaster* genome previously) has now been annotated in the most recent release of the *D.melanogaster* genome: interestingly, the snmRNA is complementary to a splice site of the ubiquinol-cytochrome-c reductase pre-mRNA. We thank Tom Jones and Sean Eddy (Washington University, St Louis, USA) for pointing these findings out to us.

### ACKNOWLEDGEMENTS

This work was supported by the German Human Genome Project through the BMBF (01KW9966) to A.H. and J.B., an IZKF grant (Teilprojekt IKF3 G6, Münster) to A.H., by laboratory funds from the Centre National de la Recherche Scientifique and Université Paul Sabatier, Toulouse, and by a grant from the Toulouse Genopole to J.-P.B.

### REFERENCES

- Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F. *et al.* (2000) The genome sequence of *Drosophila melanogaster*. *Science*, **287**, 2185–2195.
- Celniker,S.E., Wheeler,D.A., Kronmiller,B., Carlson,J.W., Halpern,A., Patel,S. Adams,M., Champe,M., Dugan,S.P., Frise,E. *et al.* (2002) Finishing a whole-genome shotgun: release 3 of the *Drosophila melanogaster* euchromatic genome sequence. *Genome Biol.*, **3**, RESEARCH0079-9.
- Hoskins,R.A., Smith,C.D., Carlson,J.W., Carvalho,A.B., Halpern,A., Kaminker,J.S. Kennedy,C., Mungall,C.J., Sullivan,B.A., Sutton,G.G. *et al.* (2002) Heterochromatic sequences in a *Drosophila* whole-genome shotgun assembly. *Genome Biol.*, **3**, RESEARCH0085-5.

4. Stapleton, M., Liao, G., Brokstein, P., Hong, L., Carninci, P., Shiraki, T., Hayashizaki, Y., Champe, M., Pacleb, J., Wan, K. *et al.* (2002) The *Drosophila* gene collection: identification of putative full-length cDNAs for 70% of *D.melanogaster* genes. *Genome Res.*, **12**, 1294–1300.
5. Mattick, J.S. (2001) Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep.*, **2**, 986–991.
6. Eddy, S.R. (2001) Non-coding RNA genes and the modern RNA world. *Nature Rev. Genet.*, **2**, 919–929.
7. Hüttenhofer, A., Brosius, J. and Bachellerie, J.P. (2002) RNomics: identification and function of small, non-messenger RNAs. *Curr. Opin. Chem. Biol.*, **6**, 835–843.
8. Jady, B.E. and Kiss, T. (2001) A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. *EMBO J.*, **20**, 541–551.
9. Giordano, E., Peluso, I., Senger, S. and Furia, M. (1999) *minifty*, a *Drosophila* gene required for ribosome biogenesis. *J. Cell Biol.*, **144**, 1123–1133.
10. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. and Tuschl, T. (2001) Identification of novel genes coding for small expressed RNAs. *Science*, **294**, 853–858.
11. Filipowicz, W. (2000) Imprinted expression of small nucleolar RNAs in brain: time for RNomics. *Proc. Natl Acad. Sci. USA*, **97**, 14035–14037.
12. Hüttenhofer, A., Kiefmann, M., Meier-Ewert, S., O'Brien, J., Lehrach, H., Bachellerie, J.-P. and Brosius, J. (2001) RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J.*, **20**, 2943–2953.
13. Tang, T.H., Bachellerie, J.P., Rozhdestvensky, T., Bortolin, M.L., Huber, H., Drungowski, M., Elge, T., Brosius, J. and Hüttenhofer, A. (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl Acad. Sci. USA*, **99**, 7536–7541.
14. Marker, C., Zemann, A., Terhorst, T., Kiefmann, M., Kastenmayer, J.P., Green, P., Bachellerie, J.P., Brosius, J. and Hüttenhofer, A. (2002) Experimental RNomics. Identification of 140 candidates for small non-messenger RNAs in the plant *Arabidopsis thaliana*. *Curr. Biol.*, **12**, 2002–2013.
15. DeChiara, T.M. and Brosius, J. (1987) Neural BC1 RNA: cDNA clones reveal nonrepetitive sequence content. *Proc. Natl Acad. Sci. USA*, **84**, 2624–2628.
16. Schmitt, A.O., Herwig, R., Meier-Ewert, S. and Lehrach, H. (1999) High density cDNA grids for hybridisation fingerprinting experiments. In Innis, M.A., Gelfand, D.H. and Sninsky, J.J. (eds), *PCR Applications: Protocols for Functional Genomics*. Academic Press, San Diego, CA, pp. 457–472.
17. Maier, E., Meier-Ewert, S., Ahmadi, A.R., Curtis, J. and Lehrach, H. (1994) Application of robotic technology to automated sequence fingerprint analysis by oligonucleotide hybridisation. *J. Biotechnol.*, **35**, 191–203.
18. Cavaille, J. and Bachellerie, J.P. (1998) SnoRNA-guided ribose methylation of rRNA: structural features of the guide RNA duplex influencing the extent of the reaction. *Nucleic Acids Res.*, **26**, 1576–1587.
19. Maden, B.E. (1990) The numerous modified nucleotides in eukaryotic ribosomal RNA. *Prog. Nucleic Acid Res. Mol. Biol.*, **39**, 241–303.
20. Massenet, S., Mouglin, A. and Branlant, C. (1998) Posttranscriptional modification in the U snRNAs. In Grosjean, H. and Benne, R. (eds), *Modification and Editing of RNA: The Alteration of RNA Structure and Function*. ASM Press, Washington, DC, pp. 201–228.
21. Ganot, P., Caizergues-Ferrer, M. and Kiss, T. (1997) 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.
22. Ganot, P., Bortolin, M.L. and Kiss, T. (1997) Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell*, **89**, 799–809.
23. Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B. and Bartel, D.P. (2002) MicroRNAs in plants. *Genes Dev.*, **16**, 1616–1626.
24. Bachellerie, J.P., Cavaille, J. and Hüttenhofer, A. (2002) The expanding snoRNA world. *Biochimie*, **84**, 775–790.
25. Gerbi, S.A., Borovjagin, A.V., Ezrokhi, M. and Lange, T.S. (2001) The ribosome. *Cold Spring Harbor Symp. Quant. Biol.*, **58**, pp. 575–590.
26. Borovjagin, A.V. and Gerbi, S.A. (2001) *Xenopus* U3 snoRNA GAC-box A' and box A sequences play distinct functional roles in rRNA processing. *Mol. Cell. Biol.*, **21**, 6210–6221.
27. Samarsky, D.A. and Fournier, M.J. (1998) Functional mapping of the U3 small nucleolar RNA from the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **18**, 3431–3444.
28. Dragon, F., Gallagher, J.E., Compagnone-Post, P.A., Mitchell, B.M., Porwancher, K.A., Wehner, K.A., Wormsley, S., Settlage, R.E., Shabanowitz, J., Osheim, Y. *et al.* (2002) A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature*, **417**, 967–970.
29. Borovjagin, A.V. and Gerbi, S.A. (1999) U3 small nucleolar RNA is essential for cleavage at sites 1, 2 and 3 in pre-rRNA and determines which rRNA processing pathway is taken in *Xenopus* oocytes. *J. Mol. Biol.*, **286**, 1347–1363.
30. Liang, W.Q., Clark, J.A. and Fournier, M.J. (1997) The rRNA-processing function of the yeast U14 small nucleolar RNA can be rescued by a conserved RNA helicase-like protein. *Mol. Cell. Biol.*, **17**, 4124–4132.
31. Filipowicz, W. and Pogacic, V. (2002) Biogenesis of small nucleolar ribonucleoproteins. *Curr. Opin. Cell Biol.*, **14**, 319–327.
32. Bachellerie, J.P. and Cavaille, J. (1997) Guiding ribose methylation of rRNA. *Trends Biochem. Sci.*, **22**, 257–261.
33. Myslinski, E., Branlant, C., Wieben, E.D. and Pederson, T. (1984) The small nuclear RNAs of *Drosophila*. *J. Mol. Biol.*, **180**, 927–945.
34. Szukalek, A., Myslinski, E., Mouglin, A., Luhrmann, R. and Branlant, C. (1995) Phylogenetic conservation of modified nucleotides in the terminal loop 1 of the spliceosomal U5 snRNA. *Biochimie*, **77**, 16–21.
35. Kiss, T. (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell*, **109**, 145–148.
36. Darzacq, X., Jady, B.E., Verheggen, C., Kiss, A.M., Bertrand, E. and Kiss, T. (2002) Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *EMBO J.*, **21**, 2746–2756.
37. Tycowski, K.T. and Steitz, J.A. (2001) Non-coding snoRNA host genes in *Drosophila*: expression strategies for modification guide snoRNAs. *Eur. J. Cell Biol.*, **80**, 119–125.
38. Ofengand, J. and Bakin, A. (1997) Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaeobacteria, mitochondria and chloroplasts. *J. Mol. Biol.*, **266**, 246–268.
39. Samarsky, D.A. and Fournier, M.J. (1999) A comprehensive database for the small nucleolar RNAs from *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **27**, 161–164.
40. Kozak, M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.*, **15**, 8125–8148.
41. Klausner, R.D. and Harford, J.B. (1989) *cis-trans* models for post-transcriptional gene regulation. *Science*, **246**, 870–872.
42. Cavaille, J., Buiting, K., Kiefmann, M., Lalande, M., Brannan, C.I., Horsthemke, B., Bachellerie, J.P., Brosius, J. and Hüttenhofer, A. (2000) Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Natl Acad. Sci. USA*, **97**, 14311–14316.
43. Cavaille, J., Vitali, P., Basyuk, E., Hüttenhofer, A. and Bachellerie, J.P. (2001) A novel brain-specific box C/D small nucleolar RNA processed from tandemly repeated introns of a noncoding RNA gene in rats. *J. Biol. Chem.*, **276**, 26374–26383.