

Available online at www.sciencedirect.com





# Different Expression Strategy: Multiple Intronic Gene Clusters of Box H/ACA snoRNA in Drosophila melanogaster

# Zhan-Peng Huang, Hui Zhou, Dan Liang and Liang-Hu Qu\*

Key Laboratory of Gene Engineering of the Ministry of Education, Biotechnology Research Center, Zhongshan University, Guangzhou 510275, People's Republic of China The high degree of rRNA pseudouridylation in Drosophila melanogaster provides a good model for studying the genomic organization, structural and functional diversity of box H/ACA small nucleolar RNAs (snoRNAs). Accounting for both conserved sequence motifs and secondary structures, we have developed a computer-assisted method for box H/ACA snoRNA searching. Ten snoRNA clusters containing 42 box H/ACA snoRNAs were identified from D. melanogaster. Strikingly, they are located in the introns of eight protein-coding genes. In contrast to the mode of one snoRNA per intron so far observed in all animals, our results demonstrate for the first time a novel polycistronic organization that implies a different expression strategy for a box H/ACA snoRNA gene when compared to box C/D snoRNAs in D. melanogaster. Mutiple isoforms of the box H/ACA snoRNAs, from which most clusters are made up, were observed in D. melanogaster. The degree of sequence similarity between the isoforms varies from 99% to 70%, implying duplication events in different periods and a trend of enlarging the intronic snoRNA clusters. The variation in the functional elements of the isoforms could lead to partial alternation of snoRNA's function in loss or gain of rRNA complementary sequences and probably contributes to the great diversity of rRNA pseudouridylation in *D. melanogaster*.

© 2004 Elsevier Ltd. All rights reserved.

\*Corresponding author

*Keywords:* box H/ACA snoRNA; intronic snoRNA gene cluster; rRNA pseudouridylation; *Drosophila melanogaster* 

## Introduction

In eukaryotes, a myriad of small nucleolar RNAs (snoRNAs) are enriched in the nucleoli in the form of small nucleolar ribonucleoprotein particles (snoRNPs).<sup>1,2</sup> Except for RNase MRP, all snoRNAs fall into two major families, i.e. box C/D and box H/ACA snoRNAs, on the basis of common sequence motifs and structural features.<sup>3</sup> A large number of snoRNAs characterized so far are box C/D snoRNAs that share two conserved motifs, the 5' end box C (RUGAUGA) and the 3' end box D (CUGA); whereas the box H/ACA snoRNAs exhibit a common hairpin-hinge-hairpin-tail secondary structure with the H (ANANNA) motif

in the hinge region and an ACA triplet that is always three nucleotides from the 3' end of the molecule.<sup>4</sup> Several snoRNAs, such as U3, snR30 and RNase MRP, are required for specific cleavage of pre-rRNAs.<sup>1</sup> However, the majority of box C/D snoRNAs function as guides for site-specific 2'-Oribose methylation and most box H/ACA snoRNAs as guides for pseudouridylation in the posttranscriptional processing of diverse RNAs, such as rRNAs,<sup>5,6</sup> snRNAs<sup>1,7</sup> and even tRNA.<sup>8</sup>

During the last decade, research on snoRNAs has provided new, often quite unexpected, information about their genomic locations that has challenged some established principles on eukaryotic gene organization and expression. So far as we know, the genomic organization of snoRNAs has shown greatest diversity among eukaryotes. In the budding yeast *Sacharomyces cerevisiae*, most snoRNAs are independently transcribed as a monocistronic unit from their own promoters.<sup>9</sup>

Abbreviations used: snoRNA, small nucleolar RNA; snoRNPs, small nucleolar ribonucleoprotein particles. E-mail address of the corresponding author: lsbrc04@zsu.edu.cn

snoRNA	Len (nt)	Chr	rRNA target		Location	Host gene	
SHORWAY	Leff (fit)	Ciu	5' Hairpin	3' Hairpin	Location	Tiost gene	
Ψ28S-1135a	148	Х		28 S Ψ1135	2nd Intron	snoRNA host gene I	
Ψ28S-1135b	145			28 S Ψ1135		(ribosomal protein S5 gene)	
Ψ28S-1135c	143			28 S Ψ1135			
Ψ28S-1135d	142		28 S Ψ483	28 S Ψ1135			
Ψ28S-1135e	151			28 S Ψ1135			
Ψ28S-1135f	151			28 S Ψ1135			
Ψ18S-1854a	139		18 S Ψ1854	18 S Ψ1937			
Ψ18S-1854b	139		18 S Ψ1854	18 S Ψ1937			
Ψ18S-1854c	140		18 S Ψ1854	18 S Ψ1937			
Ψ28S-2876(snoR825)	140			28 S U2876	3rd Intron		
Ψ28S-1192a	140		28 S Ψ1192	28 S Ψ2533	4th Intron		
Ψ28S-1192b	138		28 S Ψ1192				
Ψ28S-1192c	134		28 S Ψ1192				
Ψ28S-1192d	135		28 S Ψ1192				
Ψ28S-2626	142	3R		28 S Ψ2626	3rd Intron	snoRNA host gene II	
Ψ18S-1377a(snoR328)	139		18 S Ψ1279	18 S Ψ1377		(ribosomal protein S7 gene)	
Ψ18S-1377b	138		18 S Ψ1279	18 S Ψ1377			
Ψ18S-1377c	137		18 S Ψ1279	18 S Ψ1377			
Ψ18S-1377d	136		18 S Ψ1279	18 S Ψ1377			
Ψ18S-1377e	137		18 S Ψ1279	18 S Ψ1377			
Ψ28S-2149(snoR143)	144		28 S U2149				
Ψ28S-3316a	132	2R	28 S Ψ3316		4th Intron	snoRNA host gene III	
Ψ28S-3316b	141		28 S Ψ3316			(Dom gene)	
$\Psi$ 18S-841a(snoR66)	144		18 S Ψ841				
Ψ28S-3378	150		28 S Ψ3378		6th Intron		
Ψ28S-3316c	133		28 S Ψ3316		7th Intron		
Ψ28S-3316d	132		28 S Ψ3316	28 S U584			
Ψ28S-3316e	132		28 S ¥3316	28 S U584			
Ψ18S-841b	138		18 S Ψ841				
Ψ18S-841c	133		18 S Ψ841				
Ψ18S-841d	138		18 S Ψ841				
Ψ28S-1232	141	Х		28 S U1232	1st Intron	snoRNA host gene IV	
$\Psi$ 28S-1060	150		28 S Ψ1060		2nd Intron	(ribosomal protein L17 gene)	
Ψ28S-3436a(snoR708)	141			28 S Ψ3436			
Ψ28S-3436b(snoR75)	143			28 S Ψ3436			
Ψ28S-1837a(snoR14)	137	3L	28 S Ψ1837	28 S Ψ3801	4th Intron	snoRNA host gene V	
Ψ28S-1837b	137		28 S Ψ1837	28 S Ψ2938		(proliferation-associated protein gene)	
Ψ28S-1837c	139		28 S Ψ1837	28 S Ψ2938			
Ψ18S-1397	141		18 S Ψ1397		6th Intron		
Ψ18S-1347a	137	2R	18 S Ψ1347	28 S Ψ1313	2nd Intron	snoRNA host gene VI	
Ψ18S-1347b	143		$18 \text{ S} \Psi 1347$			(poly(A)-binding protein gene)	
$\Psi$ 18S-1347c(snoR203)	146		18 S Ψ1347	28 S Ψ1313			
Ψ28S-3327a	140	3L	28 S Ψ3327	18 S U1920	4th Intron	snoRNA host gene VII	
Ψ28S-3327b	140		28 S Ψ3327	18 S U1920		(ribosomal protein S4 gene)	
Ψ28S-3327c(snoR586)	140		28 S Ψ3327	18 S U1920	5th Intron	/	
Ψ28S-2719(snoR734)	145	Х	28 S Ψ2719		2nd Intron	snoRNA host gene VIII	
Ψ18S-531	152		18 S Ψ531	$28~\mathrm{S}~\Psi 1850$		(ribosomal protein L22 gene)	
Ψ185-531	152		18 S Ψ531	28 S Ψ1850		(ribosomal protein L22 gene)	

Table 1. The 47 box H/ACA snoRNAs in the introns of eight host genes from D. melanogaster

All snoRNAs are named after the coordination of their rRNA pseudouridylation sites, and the isoforms of each snoRNA are denoted by a, b, c, etc. snoRNAs previously identified by Yuan *et al.*<sup>20</sup> are indicated by their names in parentheses.  $\Psi$  represents rRNA pseudouridine sites that are conserved in *S. cerevisiae* and/or mammals,<sup>22</sup> and the known rRNA pseudouridine bases determined previously<sup>19</sup> or mapped here. U denotes a predicted pseudouridine site that has not been confirmed experimentally.

However, the vast majority of animal snoRNAs, as well as a small fraction of yeast snoRNAs, are nested within introns of protein-coding genes,<sup>10,11</sup> demonstrating an alternative pathway of maturation of snoRNAs and showing the important role of the introns in the expression of genetic information. This observation has been further emphasized by the characterization of some non-coding RNA genes that are exclusively transcribed for the expression of intronic snoRNAs.<sup>11,12</sup>

Another genomic organization is the snoRNA gene cluster that was first discovered in plants<sup>13,14</sup> and then, intronic snoRNA clusters were found in prevalence in the rice genome.<sup>15,16</sup> Although the snoRNA cluster has also been found in yeasts<sup>9,17</sup> and protozoa,<sup>18</sup> this polycistronic organization of snoRNA has never been reported in metazoa.

The degree of rRNA pseudouridylation in *Drosophila melanogaster* is the highest among all

### snoRNA host gene I

	CCGCATCGTC Ggtaaacatg			ccaattcca		Exon2
	ggccagagaa tttaatgtcc			tagaatgggc		ψ28S -1135a
aacttaaaac		aaccattata	aagcatatgc	ggtaccttgg	tttaaactga	300
ttccgaagca	taactctcaa	tctagtc agggactaat		agatttaatt		ψ28S -1135b
tagaatagaa	agcaaattct	ggtteteetg	gaacttaaag		attataga gg	ψ28S
actcatagga	aatgccagtt aaaggtatta attcg <mark>aca</mark> tt	actatccttg				-1135c
acctcaatgc	tagttgaagg		gcatattagt		tgagtct	ψ28S −1135d
	gttaattatc					
aactgcagcg	gaaag ttgatctctg	aaagcattca	5000 00 00 00 00 00 00 00 00 00 00 00 00	cgt	ccaaaactaa	ψ18S -1854a
	cctcttaagg					10014
	attctcagaa tactgagcat	tc	agtecacete	and the second s	THE OWNER AND ADDRESS OF THE OWNER AND ADDRESS OF	ψ28S -1135e
and the second	ttgaagatat				and the second se	1200
3.721 35	tattcaccaa	141123	cgtccaa	aactaaaact		ψ18S -1854b
	ggcctcaaaa					1380
tcagaaaccg	tcagtc	cacctcaatg				₩28S -1135f
	aaaaatggac		iyaaaliyit	LLAACLALCC	LUYLYCCLYA	11001
agatataage	cactteetge	tctcctggaa	attaaaacaa		aagcattatt	1500
caccaagggt	atcattatcc	gtccaaaact	aaaactgcag	aagtt cgttgatccc	tacagcgttg	ψ18S
caccaagggt tggtattcaa	atcattatc	gtccaaaact ataaacccca	aaaactgcag tgattagcta tca <mark>aca</mark> gca	aagtt cgttgatccc	tacagcgttg gggcatggcc	
caccaagggt tggtattcaa tcaaaaatgc gagtggatca TGGTCAGCGC	atcattatc c ttggacc <u>aca</u> aaccttaata agtgatagct CATCATCAAC	gtccaaaact ataaacccca aagcaataaa tactcatttc TCGGGACCCC	aaaactgcag tgattagcta tca <u>aca</u> gca a gattttttt GTGAGGACTC	aagtt cgttgatccc ttcctcttaa aaattctact tgtgacctct CACCCGTATT	tacagcgttg gggcatggcc ttggtgcttc ttgcagATCC GGACGTGCCG	ψ18S -1854c 1680 Exon3
caccaagggt tggtattcaa tcaaaaatgc gagtggatca TGGTCAGCGC GTACCGTCCG tcttctgaag	atcattatc c ttggacc <u>aca</u> aaccttaata agtgatagct CATCATCAAC TCGCCAGGCC atttgccttc	gtccaaaact ataaacccca aagcaataaa tactcatttc TCGGGACCCC GTCGATGTGT cagatgcatg	aaaactgcag tgattagcta tcaacagca a gattttttt GTGAGGACTC CGCCCCTGCG tgctcacg gt	aagtt cgttgatccc ttcctcttaa aaattctact tgtgacctct CACCGTATT TCGCGTCAAC ggccggtagc	tacagcgttg gggcatggcc ttggtgcttc ttgcagATCC GGACGTGCCG CAGgtgggtt aatggcttgt	ψ18S -1854c 1680 Exon3 1860 ψ28S
caccaagggt tggtattcaa tcaaaaatgc gagtggatca TGGTCAGCGC GTACCGTCCG tcttctgaag ggttcgttta	atcattatc c ttggacc <u>aca</u> aaccttaata agtgatagct CATCATCAAC TCGCCAGGCC	gtccaaaact ataaacccca aagcaataaa tactcatttc TCGGGACCCC GTCGATGTGT cagatgcatg gcttataaca	aaaactgcag tgattagcta tca <u>aca</u> gca a gattttttt GTGAGGACTC CGCCCCTGCG tgctcacg gt cttggccaga	aagtt cgttgatccc ttcctcttaa aaattctact tgtgacctct CACCGTATT TCGCGTCAAC ggccggtagc acgaaaatct	tacagcgttg gggcatggcc ttggtgcttc ttgcagATCC GGACGTGCCG CAGgtgggtt aatggcttgt atctaatatc	₩18S -1854c 1680 Exon3 1860
caccaaggt tggtattcaa tcaaaaatgc gagtggatca TGGTCAGCGC GTACCGTCCG tcttctgaag ggttcgttta cgtagcgtca aattgcataa GCTGCTGTGC	atcattatc c ttggacc <u>aca</u> aaccttaata agtgatagct CATCATCAAC TCGCCAGGCC atttgccttc gagactgctt taatgtgtag cacactttca ACTGGAGCTC	gtccaaaact ataaacccca aagcaataaa tactcatttc TCGGGACCCC GTCGATGTGT cagatgcatg gcttataaca tatagccgac ( 70nt ) GTGAGGCTGC	aaaactgcag tgattagcta tcaacagca a gattttttt GTGAGGACTC CGCCCTGCG tgctcacg gt cttggccaga atttggcgta tttttccgta CTTCAGGAAC	aagtt cgttgatccc ttcctcttaa aaattctact tgtgacctct cACCCGTATT TCGCGTCAAC ggccggtagc acgaaatct tgaaacgaag tatttccttt ATCAAGACCA	tacagcgttg gggcatggcc ttggtgcttc ttgcagATCC GGACGTGCCG CAGgtgggtt actaatat agacatta tc agGCTATCTG TCGCCGAGTG	ψ18S -1854c 1680 Exon3 1860 ψ285 -2876 2040 Exon4
caccaaggt tggtattcaa tcaaaaatgc gagtggatca TGGTCAGCGC GTACCGTCCG tcttctgaag ggttcgttta cgtagcgtca aattgcataa GCTGCTGTGC	atcattatc c ttggacc <u>aca</u> aaccttaata agtgatagct CATCATCAAC TCGCCAGGCC atttgccttc gagactgctt taatgtgtag cacactttca ACTGGAGCTC GAGCTGATCA gggg	gtccaaaact ataaacccca aagcaataaa tactcatttc TCGGGACCCC GTCGATGTGT cagatgcatg gcttataaca tatagccgac ( 70nt ) GTGAGGCTGC	aaaactgcag tgattagcta tcaacagca a gattttttt GTGAGGACTC CGCCCCTGCG tgctcacg gt cttggccaga atttggcgta tttttccgta CTTCAGGAAC GGtgggtaaa	aagtt cgttgatccc ttcctcttaa aaattctact tgtgacctct CACCGTATT TCGCGTCAAC ggccggtagc acgaaaatct tgaaacgaag tatttccttt ATCAAGACCA	tacagcgttg gggcatggcc ttggtgcttc ttgcagATCC GGACGTGCCG CAGgtgggtt aatggcttgt atctaatatc agacatta tc agGCTATCTG TCGCCGAGTG ggcgagatca	ψ18S -1854c 1680 Εχοη3 1860 ψ28S -2876 2040
caccaagggt tggtattcaa tcaaaaatgc gagtggatca TGGTCAGCGC GTACCGTCCG tcttctgaag ggttcgttta cgtagcgtca aattgcataa GCTGCTGTGC CCTGGCTGAT aagtgggaat cgcttgtaga	atcattatc c ttggacc <u>aca</u> aaccttaata agtgatagct CATCATCAAC TCGCCAGGCC atttgccttc gagactgctt taatgtgtag cacactttca ACTGGAGCTC GAGCTGATCA gggg	gtccaaaact ataaacccca aagcaataaa tactcatttc TCGGGACCCC GTCGATGTGT cagatgcatg gcttataaca tatagccgac ( 70nt ) GTGAGGCTGC ACGCTGCTAA actctaatgc aaccacgttc	aaaactgcag tgattagcta tcaacagca a gattttttt GTGAGGACTC CGCCCTGCG tgctcacg gt cttggccaga atttggcgta tttttccgta CTTCAGGAAC GGtgggtaaa ttggcattgt ggcactcgga acag	aagtt cgttgatccc tteetettaa aaattetaet tgtgaeetet cACCGTATT TCGCGTCAAC ggeeggtage acgaaaatet tgaaacgaag tattteettt ATCAAGACCA ttgggettgt actaecaaaa	tacagcgttg gggcatggcc ttggtgcttc ttgcagATCC GGACGTGCCG CAGgtgggtt aatggcttgt atctaatatc agGCTATCTG TCGCCGAGTG ggcgagatca tgcctggcct	ψ185 -1854c 1680 Exon3 1860 ψ285 -2876 2040 Exon4 2280
caccaagggt tggtattcaa tcaaaatgc gagtggatca TGGTCAGCGC GTACCGTCCG tcttctgaag ggttcgttta cgtagcgtca aattgcataa GCTGCTGTGC CCTGGCTGAT aagtgggaat cgaaatagct attctgcatc	atcattatc c ttggacc <u>aca</u> aaccttaata agtgatagct CATCATCAAC TCGCCAGGCC atttgccttc gagactgctt taatgtgtag cacactttca ACTGGAGCTC GAGCTGATCA gggg tggtct cc <u>agatcg</u> ag cttttettcc	gtccaaaact atajaacccca aagcaataaa tactcatttc TCGGGACCCC GTCGATGTGT cagatgcatg gcttataaca tatagccgac ( 70nt ) GTGAGGCTGC ACGCTGCTAA actctaatgc aaccacgttc ttctagtgac tatagtagtct	aaaactgcag tgattagcta tcaacagca a gattttttt GTGAGGACTC CGCCCTGCG tgctcacg gt cttggccaga atttggcgta tttttccgta CTTCAGGAAC GGtgggtaaa ttggcattgt ggcactcgga acag ctagtc	aagtt cgttgatccc ttcctcttaa aaattctact tgggacctct cACCGTATT TCGCGTCAAC ggccggtagc acgaaaatct tggaacgaag tatttccttt ATCAAGACCA ttgggcttgt actaccaaaa ttaacccaag atttcaac ta	tacagcgttg gggcatggcc ttggtgcttc ttgcagATCC GGACGTGCCG CAGgtgggtt aatggcttgt atctaatatc agGCTATCTG TCGCCGAGTG ggcgagatca tgcctggcct aattgagctt ggccaatcta ctccaaatct	<pre>ψ18S -1854c 1680 Exon3 1860 ψ28S -2876 2040 Exon4 2280 ψ28S</pre>
caccaagggt tggtattcaa tcaaaatgc gagtggatca TGGTCAGCGC GTACCGTCCG tcttctgaag ggttcgttta cgtagcgtca aattgcataa GCTGCTGTGC CCTGGCTGAT aagtgggaat cgcttgtaga cgaaatagct attctgcatc attctgcatc	atcattatc c ttggacc <u>aca</u> aaccttaata agtgatagct CATCATCAAC TCGCCAGGCC atttgccttc gagactgctt taatgtgtag cacactttca ACTGGAGCTC GAGCTGATCA gggg tggtct cc <u>agatcg</u> ag ctttcttcc	gtccaaaact atajaacccca aagcaataaa tactcattc TCGGGACCCC GTCGATGTGT cagatgcatg gcttataaca tatagccgac ( 70nt ) GTGAGGCTGC ACGCTGCTAA actctaatgc aaccacgttc ttctagtg <u>ac</u> tatgtagtct gaccacttaa	aaaactgcag tgattagcta tcaacagca a gattttttt GTGAGGACTC CGCCCTGCG tgCtcacg gt cttggccaga atttggcgta tttttccgta CTTCAGGAAC GGtgggtaaa ttggcattgt ggcactcgga acag ctagtc	aagtt cgttgatccc ttcctcttaa aaattctact tgtgacctct cACCGTATT TCGCGTCAAC ggccggtagc acgaaaatct tgaaacgaag tatttccttt ATCAAGACCA ttgggcttgt actaccaaga atttcaac ta	tacagcgttg gggcatggcc ttggtgcttc ttgcagATCC GGACGTGCCG CAGgtgggtt aatggcttgt atctaatatc agGCTATCTG TCGCCGAGTG ggcgagatca tgcctggcct aattgagctt ggccaatcta ctccaaatct	<ul> <li>ψ 18S</li> <li>-1854c</li> <li>1680</li> <li>Exon3</li> <li>1860</li> <li>ψ 28S</li> <li>-2876</li> <li>2040</li> <li>Exon4</li> <li>2280</li> <li>ψ 28S</li> <li>-1192a</li> <li>ψ 28S</li> </ul>
caccaagggt tggtattcaa tcaaaatgc gagtggatca TGGTCAGCGC GTACCGTCCG tcttctgaag ggttogttta cgtagcgtca aattgcataa GCTGCTGTGC CCTGGCTGAT aagtgggaat egettgtaga cgaaatagct attctgcatc cttattatcc acactt ttat	atcattatc c ttggacc <u>aca</u> aaccttaata agtgatagct CATCATCAAC TCGCCAGGCC atttgcttc gagactgctt taatgtgtag cacacttca ACTGGAGCTC GAGCTGGATCC gggg tggtct cc <u>agatcg</u> ag ctttcttcc	gtccaaaact atajaacccca aagcaataaa tactcattc TCGGGACCCC GTCGATGTGT cagatgcatg gcttataaca tatagccgac ( 70nt ) GTGAGGCTGC ACGCTGCTAA actctaatgc aaccacgttc ttctagtgac tatgtagtct gaccacttaa ttaggaactg	aaaactgcag tgattagcta tcaacagca a gattttttt GTGAGGACTC CGCCCTGCG tgctcacg gt cttggccaga atttggcgta tttttccgta CTTCAGGAAC GGtgggtaaa ttggcattgt ggcactcgga acag ctagtc cgcttttgga attcgattag attcgattag	aagtt cgttgatccc ttcctcttaa aaattctact tgtgacctct cACCGTAAT TCGCGTCAAC ggccggtagc acgaaaatct tggaacgaag tatttccttt ATCAAGACCA ttgggettgt actaccaaaa ttaacccaaaa ttaacccaaga atttcaac ta ccaacgagaa gatagcgg gg ccaaacaat	tacagcgttg gggcatggcc ttggtgcttc cGACGTGCCG CAGgtgggtt aatggcttgt atctaatatc agGCTATCTG TCGCCGAGTG ggcgagatca tgcctggct aattgagett attgacetg ggcgagatca ggccaatcta ctccaaatct acgtaaaaga	<ul> <li>ψ 18S</li> <li>-1854c</li> <li>1680</li> <li>Exon3</li> <li>1860</li> <li>ψ 28S</li> <li>-2876</li> <li>2040</li> <li>Exon4</li> <li>2280</li> <li>ψ 28S</li> <li>-1192a</li> <li>ψ 28S</li> </ul>
caccaagggt tggtattcaa tcaaaatgc gagtggatca TGGTCAGCGC GTACCGTCCG tcttctgaag ggttcgttta cgtagcgtca aattgcataa GCTGCTGTGC CCTGGCTGAT aagtgggat cgcttgtaga cgaaatagct attctgcatc cttattatcc acactt ttat	atcattatc c ttggacc <u>aca</u> aaccttaata agtgatagct CATCATCAAC TCGCCAGGCC atttgccttc gagactgctt taatgtgtag Cacactttca ACTGGAGCTC GAGCTGATCA gggg tggtct cclagatcgag cttttcttcc atggtttatt ttggctagt gaattagcat	gtccaaaact atajaacccca aagcaataaa tactcattc TCGGGACCCC GTCGATGTGT cagatgcatg gcttataaca tatagccgac ( 70nt ) GTGAGGCTGC ACGCTGCTAA actctaatgc aaccacgttc ttctagtgac tatgtagtct gaccacttaa ttaggaactg tatgtagtct caatatcagt	aaaactgcag tgattagcta tcaacagca a gattttttt GTGAGGACTC CGCCCTGCG tgctcacg gt cttggcaga atttggcgta tttttccgta CTTCAGGAAC GGtgggtaaa ttggcattgt ggcactcgga acag ctagtc cgcttttgga attcgattag attcgattag	aagtt cgttgatccc ttcctcttaa aaattctact tgtgacctct cACCGTATT TCGCGTCAAC ggccggtagc acgaaaatct tgaaacgaag tatttccttt ATCAAGACCA ttggcttgt actaccaaga atttcaac ta cc <u>agatca</u> aa gatagcgg gg cc <u>agaaca</u> at	tacagcgttg gggcatggcc ttggtgcttc cGACGTGCCG CAGgtgggtt aatggcttgt atctaatatc agGCTATCTG TCGCCGAGTG ggcgagatca tgcctggct aattgagett attgacetg ggcgagatca ggccaatcta ctccaaatct acgtaaaaga	<ul> <li>ψ18S</li> <li>-1854c</li> <li>1680</li> <li>Exon3</li> <li>1860</li> <li>ψ28S</li> <li>-2876</li> <li>2040</li> <li>Exon4</li> <li>2280</li> <li>ψ28S</li> <li>-1192a</li> <li>ψ28S</li> <li>-1192b</li> <li>ψ28S</li> </ul>
caccaagggt tggtattcaa tcaaaatgc gagtggatca TGGTCAGCGC GTACCGTCCG ctttetgaag ggttegttta cgtagegtca aattgcataa GCTGCTGTGC CCTGGCTGAT aagtgggaat cgaaataget attetgeata tctattatec accodtt ttat attetgeata tattatectt ta gcaaacta	atcattatc c ttggacc <u>aca</u> aaccttaata agtgatagct CATCATCAAC TGGCCAGGCC atttgccttc gagactgctt taatgtgtag Cacactttca ACTGGAGCTC GAGCTGATCA gggg tggtct cdgatcgag cttttcttcc atggtttatt ttgtgctagt gaattagcat gtggtttaat gtggtttaat	gtccaaaact atajaacccca aagcaataaa tactcattc TCGGGACCCC GTCGATGTGT cagatgcatg gcttataaca tatagccgac ( 70nt ) GTGAGGCTGC ACGCTGCTAA actctaatgc aaccacgttc ttctagtgactg tatgtagtct gaccacttaa ttaggaactg tatgtagtct caatatcagt ggtatataga ctcgcttttg	aaaactgcag tgattagcta tcaacagca a gattttttt GTGAGGACTC CGCCCTGCG tgctcacg gt cttggccaga atttggcgta tttttccgta CTTCAGGAAC GGtgggtaaa ttggcattgt ggcactcgga acag ctagtc cgcttttgga ctgtcctatg attcgattag acagtatag	aagtt cgttgatccc ttcctcttaa aaattotact gggacgtagc acgaaaatct tggaacgaag tattccttt ATCAAGACCA ttgggettgt actaccaaaa ttaacccaaga attcaac gatagcgg gg cd <u>agaacg</u> aa gggcggaacgaa gatagcgg	tacagcgttg gggcatggcc ttgcagATCC GGACGTGCCG CAGgtgggtt aatggcttgt atctaatatc agGCTATCTG TCGCCGAGTG ggcgagatca tgcctggct aattgagctt ggccaatcta ctccaatct acgtaaaaga gtccaatcta caccatcctt aaaagaacat	<ul> <li>ψ18S</li> <li>-1854c</li> <li>1680</li> <li>Exon3</li> <li>1860</li> <li>ψ28S</li> <li>-2876</li> <li>2040</li> <li>Exon4</li> <li>2280</li> <li>ψ28S</li> <li>-1192a</li> <li>ψ28S</li> <li>-1192b</li> <li>ψ28S</li> <li>-1192c</li> </ul>

Figure 1 (gene I) (legend on p.675)

AAGCGCATCC	GCGTCAAGCT	GGACGGCTCC	CAGCTGGTCA	AGGTGCACCT	GGACAAGAAC	Exon3
CAGCAGACCA	CCATTGAACA	CAAAgtaagc	ataagcatcc	tccacctcaa	agaagagtgc	
atcgtgtgtg	ttagcag					
		and the second s	atatttgggc	and the state of the		$\Psi 28S$
			gccaattact	cgttattgca	ggcattagaa	-2626
gccattgcgg	tcctgcatct	ccaaatattg	gccacagtt			
			g	gtcttttgat	aca tcgttct	ψ18S
tgtgtttttc	gctggctgtc	cttccaaagc	cagtgattgt	caaacagaac		-1377a
ggcagcttat	cactccacaa	tgttgagccg	agttttcacg	gttactgcat	tactttccgg	
ctcgccacat	ca					
	ttacacac	tcaaatcgat	acttgaaacc	atttaaacag	tagaatgaaa	480
agtcaaataa	gggggataac	agatggatag	aaagattata	tgtttaaatg	tacatagaac	540
tttcta					1010	
tcgt	ttttgtgtta	tttgctggaa	gttacatcca	aagccagtga	ttgtcaaaca	$\Psi 18S$
gaacgtaaat	taatgcggct	atcactccac	aatgttgagc	cggctttcca	gggctactgc	-1377b
attacttttc	ggctcgccac	actc				
		tcgtag	aatcttttaa	ataccttggt	ttctacca	
					cc	$\Psi 18S$
			ggcctagcga			-1377c
		caatggtggg	tctactttac	caggactgca	tcattactaa	
aggattgcta	Contraction of the second					
	gaata	tccacgagaa	tgctaattgt	The second se		100
A	100		24	and the second se	cttgtgttat	ψ18S
			gtcaaacaga	and the second se	the second s	-1377d
atcactccac	aatgttgagc	cgactctcaa	cggctactgc	attactatat	gcgccacatt	
C				- 1		
atgtcaatc	aacacttagc	ccttgaatac	tttggtttcc	the second s	****	Ψ18S
ttotagoogt	tattapaata	apagastast	22222222222		ttgtgttatt	Ψ185 -1377e
			caaacagaac			-13776
actocacaat	grigggiola	Cullgueagg	acagcatcat	LactLaayya	LUGCALACALYC	
a	togogttogt	astatassas	cacacacgat	googtttgog		
CCLLACALL	Layayttact	Callydada	Cacacacyac	gcacttyay	tgccgcctat	Ψ28S
ttacctctca	agegeegettt	agtttcctgg	ggcctcggga	cttcctatat	and the second se	-2149
			tgcggtattt			2110
ggtgctgtac		acacoccoge	cycyycutte	cocycoyddy	caaraa ccac	
55590090 <u>40</u>		ataagttaat	tectetgate	gtgtaaacta	atacottete	1440
tatttatctc		•	TCGGTCTACA		•	Exon4
	3		AATGTCTAGA			

snoRNA host gene II

Figure 1 (geneII) (legend on p.675)

organisms,19 which provides a good model for studying the diversity of box H/ACA snoRNAs and their genomic organization. A recent study of experimental RNomics in D. melanogaster has identified 20 box H/ACA snoRNAs that were predicted for about 30 pseudouridylation sites in rRNAs and snRNAs.<sup>201</sup> However, it seems that about 100 pseudouridylation sites are located in D. melanogaster rRNAs, and more than half of the box H/ACA snoRNAs remain to be found in this organism. Here, we searched for box H/ACA snoRNAs in D. melanogaster with a computerassisted method, and subsequently identified ten novel box H/ACA snoRNAs with their numerous isoforms. Interestingly, a novel polycistronic organization of the box H/ACA snoRNAs was discovered. The results demonstrate the utilization of different strategies for the expression of two classes of snoRNA genes in D. melanogaster.

### Results

# Identification of ten intronic box H/ACA snoRNA gene clusters from *D. melanogaster*

Accounting for the conserved secondary struc-

tures, we have developed a computer-assisted method for box H/ACA snoRNA searching. The program searches for the candidates, ranging in size from 120 to 160 nt, that exhibit the hairpin-hinge-hairpin-tail structure with a box H (ANANNA) in the hinge region and box ACA (ACA or AUA) in the tail, and display an rRNA complementarity of at least 9 nt in the internal loop of the hairpin domain. The analytical approach was applied to all the intron sequences of eight protein-coding genes of D. melanogaster, from which ten box H/ACA snoRNA genes had been identified.<sup>20</sup> To our surprise, in addition to known snoRNAs, many new box H/ACA snoRNA candidates were also discovered in these introns. After sequence analyses, a total of 47 snoRNA coding regions corresponding to 19 different box H/ACA snoRNAs were identified from 15 introns of the host genes, respectively (Table 1 and Figure 1). Among the 19 snoRNAs, ten novel snoRNAs with 24 isoforms were first identified here. Notably, ten intronic clusters containing 42 snoRNA variants were characteristic of the genomic organization of the box H/ACA snoRNAs in *D. melanogaster.* The clusters are mainly formed by isoforms of the same snoRNA species, while the gene content of the clusters varies largely from

#### snoRNA host gene III

TCCACCAAGC	TTCCCGCTGC	CGTCGTCCAG	CTAACGCAAC	AAGgtaaact	catcttgttc	Exon4
ccagtcgacc	cctttgacct	( 1030nt )	cattagtact	agactagttt	tatg	
					tcgcac	Ψ28S
	tttctcgtcg					-3316a
	gttgcaagtg	atacggccat	tttttgtggc	agatatccat	aacgctgact	
acaagt						
gatc	tcggcgg					11.11.11.11.11.11.11.11.11.11.11.11.11.
No. 2		cacctaaaag				ψ28S
	gatgtgtgcc			aagtgatcta	gatccgtgac	-3316b
agaatcagat	ccaattaaag	ctgggctcat				1.1.10
			ta	gtatctaaac	ttacaaatcg	1440
gccggtacac						₩18S
		actetgtgte				Ψ185 -841a
	cttgctagaa			ccatgcggca	actatettea	-641a
geeeagatge	cgcattgaat	Legggeerda		aagtttgcat	apptttappp	1620
cototttoat	ttagcaaccg	attegeaaat			a second second of the second second second	1820
	taagGTGGCA					Exon5-6
	GTTGCTACTT					EX0113-C
ctaagettge		AAGIIIIGGA	CHOHOHOGCG	agraatgatt	gealacycly	
ceaugerege		caaaaatcag	taatcatoco	attoctaaac	tagcatogco	ψ28S
actacoctto	ggtgctalaaa					-3378
	ctagaatgct				ougugourgo	0010
caucocagoo		-9-99900009	ccccdrd <mark>aca</mark>		acgcctttta	4260
cacqaatqaa	tttataaagt	actcttaact	ttctacaqGC			Exon7
	( 130nt )					
	gtttatttt			5 5		
	-		aaaagcgttt	tettatcage	ttgataaatg	Ψ28S
ctcttgaaaa	tcgtagatgt					-3316c
	caatccgttc					
			attat	aagtettaca	a	
					tgcaaatgc	ψ18S
actetgtgte	acattgaatt	tacctacatg	tgaagacaat	ttgttagttt	gctatatgaa	-841b
agagatgtga	cttatccatg	caattagtgc	cgtaaaaggt	tgattgcatt	gaaatcgcat	
tttacacta						
t	cgaaaattaa					
			tgcacctaaa			ψ28S
	ttggaaatcg				gcaagtgatt	-3316d
ttagtteete	taactcaaat	ccatttacat	tgtgaacata			
				gtcctcagg	and the second se	
					gg	Ψ18S
SU SUTATION STORES	ctgtgtcaca					-841c
	gtattatcca	tgtatttaaa	gccgtatcag	ggtgattaca	ttgaaatcgc	
attttacact						
	ttattaatc	aattattttc	BURNESS AND			¥ 28S
	h a h h a a a a h h	and a such all a		aaagcgtttt		Ψ285 −3316e
	tcttgaaaat				tagcaagtga	-33166
ttttagttcc	cttaactcta	atccgtttac	attgtgaaca			5280
ttostattt	gttcctactt			gatecat	aagtaaaatg	5280
clargeret	guidelacti		actctgtgtc	acattaasaa	tagettegtg	Ψ18s
tantanannt	ttgttagttt	and the second s				-841d
and the second second second second	tattcacatt	And a second		conaccedaty	eggeegagte	0410
secusedate	caccoudtt	gaaaaogeeee	Contraction of the second second second	cctcttattt	aagttateet	5460
tattttcaaa	ataaatgcgc	agtaattaga				Exon8
	CTACTTCGAC					Discino
1001000110	0.110110010	00010111111	551100100hh		0.01 M 101 M 10 171	

Figure 1 (gene III) (legend on p.675)

two to nine snoRNAs. Probably because of modest intron sizes in the fly, the clusters appear very compact. Intergenic spacers between the snoRNA genes are usually 20–40 nt but in some cases the short ones are less than ten nucleotides and a long spacer of 700 nt is also observed in host gene VIII. Among the eight host genes, five are ribosomal protein genes. All introns of snoRNA hosts possess standard boundary signals, i.e. GU at the 5' end and AG at the 3' end. The distance from the last snoRNA genes in the clusters to the 3' end of the introns is at least 55 nt, slightly different from the 71–80 nt that is important for the processing of intronic snoRNAs in mammals.<sup>21</sup> On the other hand, the sequences between the first snoRNA gene in the cluster and the 5' end of the intron are relatively short, many are less than 50 nt.

### Positive detection of the ten novel H/ ACA snoRNAs

Ten oligonucleotides were designed and synthesized according to the coding regions of ten corresponding snoRNAs, i.e.  $\Psi$ 28S-2626,  $\Psi$ 28S-3316,  $\Psi$ 28S-1060,  $\Psi$ 28S-1135,  $\Psi$ 18S-531,  $\Psi$ 18S-1854,  $\Psi$ 28S-1192,  $\Psi$ 18S-1397,  $\Psi$ 28S-3378 and  $\Psi$ 28S-1232, which were newly identified in this analysis. As shown in Figure 2A, all the target snoRNAs were positively detected by Northern blotting with the labeled probes. In each case, a unique and strong

### snoRNA host gene IV

TGATTTCCAG	CACTAGAACT	GCCATTTCCT	TTCTTTTTCG	TTCCACGTTT	CCGgtgagtg	Exon1
catgtgccta	gaaagttatt	ttccgaaaat	( 310nt )	catgtggtgt	tgtcattgtg	
tctgtgtgtg						
	gtgtgcc	tatgcgcgaa	cattgtttct	caaaagtatt	taaactatcc	$\Psi 28S$
aaataaagtg	gtaactcaat					-1232
cggaatccgt	gcagaatttt	tgttgtttac	agga			
			gcacac	acatgcgcac	tcccttacgc	600
acgcacacac	acacatagga	( 460nt )	attgcccatg	tttcttgcag	ATATTGAAAA	Exon2
ATGGGCCGTT	ACTCACGCGA	GTCAGACAAC	GTGGCCAAGT	CGTGCAAGGC	GCGCGGGGCCC	
AATCTACGTG	TGCACTTCAA	Ggtgagaatc	ctgcccagaa	atccgaggaa	cctgggcctt	1230
cttgggcatc	aaccagt					
	ccg	cagettaaat	cagtgaatgc	cagccgattg	tgcagcatct	$\Psi 28S$
cgcttttgca	ttgctgctag	aagaatcete	caatttggac	gagttcatca	gctgagtgcc	-1060
aagteecaaa	actctggacg	ccctcttcta	aaatctcgcc	aacaacc		
				aat	tateccaatg	1410
aatccagtat	tt					
			ggcggttgct			ψ28S
cgatacgtgt	gcccagagcg	aatccaaagc	cttgctaaaa	cagcggcctg	ctgctttcaa	-3436a
tgagctagtc	tctacctgta	gcctggcaca	gat			
				an earlier and the second second second	aaatccatca	1590
cggaatccac	tagatcaagc	( 130nt )	atttatgttc	cccctgc		
					ggcacagttt	$\Psi 28S$
	ggttatcctc					-3436b
cccagccatt	gtaaaacagc	ggcgtgttgg	ccctcagtgc	tegegtetet	acctgaacct	
tggcacaaag						
			actatgattt			1950
	aactgatgaa					Exon3
CAGGCCATCA	AGCGCATGCC	CCTGCGCCGT	GCCCAGCGTT	ACCTGAAGGC	CGTCATCGAC	
snoRNA h	lost gene	V				
	ost gene		CCTCGCCGCC	TACTGGGCTG	TCCAGGCTGC	Exon4
GAAGATCAGT	_	CCGATGTCAT				Exon4
GAAGATCAGT	GGTCGCCAGG	CCGATGTCAT				Exon4
GAAGATCAGT CTTACGTCTG gttaatc	GGTCGCCAGG	CCGATGTCAT GCGCCAATgt	gagtcetece	ttacttctag	gtaateetee	Exon4 ¥28S
GAAGATCAGT CTTACGTCTG gttaatc cct	GGTCGCCAGG CTCAAGTCCG	CCGATGTCAT GCGCCAATgt gattgtctgc	gagtcetece cgcgattete	ttacttctag cagcgactga	gtaateetee acateteaac	
GAAGATCAGT CTTACGTCTG gttaatc cct acttgcaa <mark>ag</mark>	GGTCGCCAGG CTCAAGTCCG gcaagaaacg	CCGATGTCAT GCGCCAATgt gattgtctgc gcagctggta	gagtcetece cgcgattete	ttacttctag cagcgactga	gtaateetee acateteaac	₩28S
GAAGATCAGT CTTACGTCTG gttaatc cct acttgcaa <mark>ag</mark>	GGTCGCCAGG CTCAAGTCCG gcaagaaacg atcagctgtg	CCGATGTCAT GCGCCAATgt gattgtctgc gcagctggta	gagtcctccc cgcgattctc attgccctgg ct	ttacttctag cagcgactga cctattattc	gtaateetee acateteaac aggaetggag	¥28S −1837a
GAAGATCAGT CTTACGTCTG gttaatc cct acttgcaa <u>ag</u> gcttcttgtc	GGTCGCCAGG CTCAAGTCCG gcaagaaacg atcagctgtg agttgtccac	CCGATGTCAT GCGCCAATgt gattgtctgc gcagctggta aagg ttattt	gagtcctccc cgcgattctc attgccctgg ct tctgcagg	ttacttctag cagcgactga cctattattc caacggattg	gtaateetee acateteaae aggaetggag actgegetea	ψ28S -1837a ψ28S
GAAGATCAGT CTTACGTCTG gttaatc ect acttgcaa <u>ag</u> gcttcttgtc aactctgaca	GGTCGCCAGG CTCAAGTCCG gcaagaaacg atcagetgtg agttgtcc <u>ac</u> cagatcagct	CCGATGTCAT GCGCCAATgt gattgtetge gcagetggta agg ttattt caacacetge	gagtceteee egegattete attgeeetgg et tetgeagg ggatagaaa	ttacttetag cagegaetga cetattatte caaeggattg tgtgteaatt	gtaateetee acateteaae aggaetggag actgegetea	¥28S −1837a
GAAGATCAGT CTTACGTCTG gttaatc ect acttgcaa <u>ag</u> gcttcttgtc aactctgaca	GGTCGCCAGG CTCAAGTCCG gcaagaaacg atcagctgtg agttgtccac	CCGATGTCAT GCGCCAATgt gattgtetge gcagetggta agg ttattt caacacetge	gagtceteee egegattete attgeeetgg et tetgeagg ggatagaaa	ttacttetag cagegaetga cetattatte caaeggattg tgtgteaatt tee <u>aea</u> tet	gtaateetee acateteaac aggaetggag actgegetea tegtgaaetg	ψ28S -1837a ψ28S -1837b
GAAGATCAGT CTTACGTCTG gttaatc cct acttgcaa <u>ag</u> gcttcttgtc aactctgaca aacaagttca	GGTCGCCAGG CTCAAGTCCG gcaagaaacg atcagctgtg agttgtccac cagatcagct ttccatagaa	CCGATGTCAT GCGCCAATgt gattgtetge gcagetggta agg ttattt caacacetge	gagtceteee egegattete attgeeetgg et tetgeagg ggatagaaa	ttacttetag cagegaetga cetattatte caaeggattg tgtgteaatt tee <u>aea</u> tet	gtaateetee acateteaae aggaetggag actgegetea	ψ28S -1837a ψ28S
GAAGATCAGT CTTACGTCTG gttaatc ect acttgcaa <u>ag</u> gcttcttgtc aactctgaca	GGTCGCCAGG CTCAAGTCCG gcaagaaacg atcagctgtg agttgtccac cagatcagct ttccatagaa attgtaa	CCGATGTCAT GCGCCAATgt gattgtctgc gcagctggta aagg ttattt caacacctgc gtgttcggtc	gagtcetcee cgcgattete attgecetgg ct tetgeagg gg <u>atagaa</u> ac tttaaatttg	ttacttctag cagogactga cotattattc caacggattg tgtgtcaatt tccacatct c	gtaatcctcc acatctcaac aggactggag actgcgctca tcgtgaactg cagtttatag	₩285 -1837a ₩285 -1837b 420
GAAGATCAGT CTTACGTCTG gttaatc cct acttgcaa <u>ag</u> gcttcttgtc aactctgaca aacaagttca atatgtcgga	GGTCGCCAGG CTCAAGTCCG gcaagaaacg atcagctgtg agttgtccac cagatcagct ttccatagaa attgtaa tct	CCGATGTCAT GCGCCAATgt gattgtctgc gcagctggta aagg ttattt caacacctgc gtgttcggtc gcaggcaacg	gagtectee egegattete attgeeetgg et tetgeagg gg <u>atagaa</u> e tttaaattg gattgtetge	ttacttctag cagcgactga cotattattc caacggattg tgtgtcaatt tccacetc c tgccttaact	gtaateetee acateteaac aggaetggag actgegetea tegtgaaetg cagtttatag egtggeteag	ψ28S -1837a ψ28S -1837b 420 ψ28S
GAAGATCAGT CTTACGTCTG gttaatc cct acttgcaa <u>ag</u> gcttcttgtc aactctgaca aacaagttca atatgtcgga cacagctcaa	GGTCGCCAGG CTCAAGTCCG gcaagaaacg atcagctgtg agttgtccac cagatcagct ttccatagaa attgtaa tct cgtctgcag	CCGATGTCAT GCGCCAATgt gattgtctgc gcagctggta aagg ttattt caacacctgc gtgttcggtc gcaggcaacg gatcaacagt	gagtectee cgcgattete attgeeetgg ct tetgeagg gg <u>atagaa</u> ac tttaaatttg gattgtetge gtegatteg	ttacttctag cagcgactga cotattattc caacggattg tgtgtcaatt tccacetc c tgccttaact	gtaateetee acateteaac aggaetggag actgegetea tegtgaaetg cagtttatag egtggeteag	₩285 -1837a ₩285 -1837b 420
GAAGATCAGT CTTACGTCTG gttaatc cct acttgcaa <u>ag</u> gcttcttgtc aactctgaca aacaagttca atatgtcgga cacagctcaa	GGTCGCCAGG CTCAAGTCCG gcaagaaacg atcagctgtg agttgtccac cagatcagct ttccatagaa attgtaa tct	CCGATGTCAT GCGCCAATgt gattgtctgc gcagctggta aagg ttattt caacacctgc gtgttcggtc gcaggcaacg gatcaacagt	gagteetee cgegattete attgeeetgg ct tetgeagg gg <u>atagaa</u> ac tttaaattg gattgtetge gtegatteg acaate	ttacttctag cagcgactga cetattatte caacggattg tgtgtcaatt tco <u>aca</u> tct c tgccttaact tgaactgaac	gtaateetee acateteaac aggaetggag actgegetea tegtgaaetg cagtttatag egtggeteag aagtttagat	ψ285 -1837a ψ285 -1837b 420 ψ285 -1837c
GAAGATCAGT CTTACGTCTG gttaatc cct acttgcaa <u>ag</u> gcttcttgtc aactctgaca aacaagttca atatgtcgga cacagctcaa acttgaaatg	GGTCGCCAGG CTCAAGTCCG actagctgtg agttgtccac cagatcagct ttccatagaa attgtaa ttgtaa tcgtctgcaga	CCGATGTCAT GCGCCAATgt gattgtotgc gcagctggta aagg ttattt caacacetgc gtgttcggtc gatcaacagt aaagttgtcc	gagtcetcee cgcgattete attgecetgg ct tetgeagg gg <u>atagaa</u> ac tttaaatttg gattgtetge gtegatteg <u>acaate</u> gcaa	ttacttctag cagcgactga cctattattc caacggattg tgtgtcaatt tccacgtatt cc tgccttaact tgaactgaac	gtaatectee acateteaae aggaetggag actgegetea tegtgaaetg cagtttatag egtggeteag aagtttagat gateagttat	<ul> <li>₩ 285</li> <li>-1837a</li> <li>₩ 285</li> <li>-1837b</li> <li>420</li> <li>₩ 285</li> <li>-1837c</li> <li>600</li> </ul>
GAAGATCAGT CTTACGTCTG gttaatc cct acttgcaa <u>ag</u> gcttcttgtc aactctgaca aacaagttca atatgtcgga cacagctcaa acttgaaatg tgttattttg	GGTCGCCAGG CTCAAGTCCG gcaagaaacg atcagctgtg agttgtccac cagatcagct ttccatagaa attgtaa ttcgtctgcaga ttcggtcttt ( 130nt )	CCGATGTCAT GCGCCAATgt gattgtotgc gcagctggta aagg ttattt caacacctgc gtgttcggtc gcaggcaacg gatcaacagt aaagttgtcc tacagAACTA	gagtcetcee cgegattete attgecetgg ct tetgeagg gg <u>tagaa</u> ac tttaaatttg gattgtetge gtegatteg gcegatteg gcegatteg gcea creecteAcc	ttacttctag cagcgactga cctattattc caacggattg tgtgtcaatt tccacgtatt cc tgccttaact tgaactgaac	gtaateetee acateteaae aggaetggag actgegetea tegtgaaetg cagtttatag egtggeteag aagtttagat gateagttat AACAAATEAG	ψ285 -1837a ψ285 -1837b 420 ψ285 -1837c
GAAGATCAGT CTTACGTCTG gttaatc cct acttgcaaag gcttcttgtc aactctgaca aacaagttca atatgtcgga cacagctcaa acttgaaatg tgttattttg CGAGTCGTAT	GGTCGCCAGG CTCAAGTCCG gcaagaaacg atcagctgtg agttgtccac cagatcagct ttccatagaa attgtaa ttc cgtctgcaga ttcggtott ( 130nt ) AAGTGCAAGC	CCGATGTCAT GCGCCAATgt gattgtotgc gcagctggta aagg ttattt caacacctgc gtgttcggtc gcaggcaacg gatcaacagt aaagttgtcc tacagAACTA ( 610nt )	gagtectee attgeeetgg et tetgeagg gg <u>tagaa</u> ac tttaaattg gattgtetge gtegatteg gcaa creecteAcc gGAGTTGTTG	ttacttctag caacgactga cotattattc caacggattg tgtgtcaatt tccacatct c tgccttaact tgaactgaac	gtaatcctce acatctcaac aggactggag actgcgctca tcgtgaactg cagtttatag cgtggctcag aagtttagat gatcagttat AACAAATCAG CCACAAGATG	<ul> <li>₩ 285</li> <li>-1837a</li> <li>₩ 285</li> <li>-1837b</li> <li>420</li> <li>₩ 285</li> <li>-1837c</li> <li>600</li> <li>Exon5-6</li> </ul>
GAAGATCAGT CTTACGTCTG gttaatc oct acttgcaaag gcttcttgtc aactctgaca aacaagttca atatgtcgga cacagctcaa acttgaaatg tgttattttg cGAGTCGTAT ATTGAGCCCT	GGTCGCCAGG CTCAAGTCCG gcaagaaacg atcagctgtg agttgtccac cagatcagct ttccatagaa attgtaa ttcgtctgcaga ttcggtcttt ( 130nt ) AAGTGCAAGC TCCAAGTGCT	CCGATGTCAT GCGCCAATgt gattgtctgc gcagctggta aagg ttattt caacacctgc gtgttcggtc gcaggcaacg <u>gatca</u> acagt aaagttgtcc tacagAACTA ( 610nt ) GTACGAGAAG	gagtectee attgeeetgg et tetgeagg gg <u>tagaa</u> ac tttaaatttg gattgtetge gtegatteg geaa creecteAcc gGAGTTGTTG	ttacttctag caacgactga cotattattc caacggattg tgtgtcaatt tccacatct c tgccttaact tgaactgaac	gtaatcctce acatctcaac aggactggag actgcgctca tcgtgaactg cagtttatag cgtggctcag aagtttagat gatcagttat AACAAATCAG CCACAAGATG	<ul> <li>₩ 28S</li> <li>-1837a</li> <li>₩ 28S</li> <li>-1837b</li> <li>420</li> <li>₩ 28S</li> <li>-1837c</li> <li>600</li> </ul>
GAAGATCAGT CTTACGTCTG gttaatc oct acttgcaaag gcttcttgtc aactctgaca aacaagttca atatgtcgga cacagctcaa acttgaaatg tgttattttg cGAGTCGTAT ATTGAGCCCT	GGTCGCCAGG CTCAAGTCCG gcaagaaacg atcagctgtg agttgtccac cagatcagct ttccatagaa attgtaa ttc cgtctgcaga ttcggtott ( 130nt ) AAGTGCAAGC	CCGATGTCAT GCGCCAATgt gattgtctgc gcagctggta aagg ttattt caacacctgc gtgttcggtc gcaggcaacg gatcaacagt aaagttgtcc tacagAACTA ( 610nt ) GTACGAGAAG g	gagtectee cgegattete attgeeetgg ct tetgeagg ggatagaae tttaaatttg gattgtetge gtegatteg gecaa creecteae gecaa creecteae GaGaTTGTTG GCATgtaagt	ttacttctag caacgactga cetattattc caacggattg tgtgtcaatt tccactct c tgccttaact tgaactgaac	gtaatcctcc acatctcaac aggactggag actgcgctca tcgtgaactg cagtttatag cgtggctcag aagtttagat gatcagttat AACAAATCAG cCACAAGATG ttattataa	<pre>₩285 -1837a ₩285 -1837b 420 ₩285 -1837c 600 Exon5-6 1500</pre>
GAAGATCAGT CTTACGTCTG gttaatc cct acttgcaaag gcttottgtc aactctgaca aacaagttca atatgtcgga cacagctcaa acttgaaatg tgttattttg CGAGTCGTAT ATTGAGCCCT tcctattccc	GGTCGCCAGG CTCAAGTCCG gcaagaaacg agttgtccac cagatcagct ttccatagaa attgtaa ttcgtaga ttcggtcttt ( 130nt ) AAGTGCAAGC TCCAAGTGCT tattatgcga	CCGATGTCAT GCGCCAATgt gattgtotgc gcagctggta aagg ttattt caacacotgc gtgttcggtc gatcajacagt aaagttgtcc tacagAACTA ( 610nt ) GTACGAGAAG g ttggcagaa	gagtectee cgegattete attgeedgg ct tetgeagg gg <u>atagaaac</u> tttaaattg gatgtetge gtegatteg gacaate geaa creerCreAce GGAGTGETG CCATgtaagt	ttacttctag cagcgactga cetattatte caacggattg tgtgtcaatt tco <u>aca</u> tct c tgccttaact tgaactgaac tgataatgcc GATGCAGTGC AGTGCGTCGG gtgatgcata	gtaatectee acateteaac aggaetggag actgegetea tegtgaaetg cagtttatag cgtggeteag aagtttagat gateagttat AACAAATEAG CCACAAGATG ttattataa aecttegggt	<pre>₩285 -1837a ₩285 -1837b 420 ₩285 -1837c 600 Exon5-6 1500 ₩185</pre>
GAAGATCAGT CTTACGTCTG gttaatc cct acttgcaaag gcttcttgtc aactotgaca aacaagttca atatgtcgga cacagctcaa acttgaaatg tgttatttg cGAGTCGTAT ATTGAGCCCT tcctattccc	GGTCGCCAGG CTCAAGTCCG gcaagaaacg attagctgtg agttgtccac cagatcagct ttccatagaa attgtaa ttgtaa ttcgtctgcag ttcggtcttt ( 130nt ) AAGTGCAAGC TCCAAGTGCT tattatgcga gccagacatt	CCGATGTCAT GCGCCAATgt gattgtotgc gcagctggta aagg ttattt caacacctgc gtgttcggtc gatcaacagt aaagttgtcc tacagAACTA ( 610nt ) GTACGAGAAG g ttggcagaa ttgccagaac	gagtcetcee cgegattete attgecetgg ct tctgeagg gg <u>atagaa</u> ac tttaaatttg gattgtetge gtegatteg gaca creercace gGAGTGTTG cCATgtaagt cttaatteeg aaatteeagg	ttacttctag cagcgactga cctattattc caacggattg tgtgtcaatt tccacgattg cc tgccttaact tgaactgaac	gtaatectee acateteaac aggaetggag actgegetea tegtgaaetg cagtttatag cgtggeteag aagtttagat gateagttat AACAAATEAG CCACAAGATG ttattataa aecttegggt	<pre>₩285 -1837a ₩285 -1837b 420 ₩285 -1837c 600 Exon5-6 1500</pre>
GAAGATCAGT CTTACGTCTG gttaatc cct acttgcaaag gcttcttgtc aactotgaca aacaagttca atatgtcgga cacagctcaa acttgaaatg tgttatttg cGAGTCGTAT ATTGAGCCCT tcctattccc	GGTCGCCAGG CTCAAGTCCG gcaagaaacg agttgtccac cagatcagct ttccatagaa attgtaa ttcgtaga ttcggtcttt ( 130nt ) AAGTGCAAGC TCCAAGTGCT tattatgcga	CCGATGTCAT GCGCCAATgt gattgtotgc gcagctggta aagg ttattt caacacctgc gtgttcggtc gatcaacagt aaagttgtcc tacagAACTA ( 610nt ) GTACGAGAAG g ttggcagaa ttgccagaac	gagtcetcee cgegattete attgecetgg ct tctgeagg gg <u>atagaa</u> ac tttaaatttg gattgtetge gtegatteg gaca creercace gGAGTGTTG cCATgtaagt cttaatteeg aaatteeagg	ttacttctag caacgactga cotattattc caacggattg tgtgtcaatt tccacatct cgacttaatc tgaactgaac	gtaatcetce acatetcaac aggactggag actgegetca tegtggactg cagtttatag egtggetcag aagtttagat gateagttat AACAAATCAG CCACAAGATG ttattataa acettegggt	<ul> <li>₩ 285</li> <li>-1837a</li> <li>₩ 285</li> <li>-1837b</li> <li>420</li> <li>₩ 285</li> <li>-1837c</li> <li>600</li> <li>Exon5-6</li> <li>1500</li> <li>₩ 185</li> <li>-1397</li> </ul>
GAAGATCAGT CTTACGTCTG gttaatc cct acttgcaaag gcttcttgtc aactctgaca aacaagttca atatgtcgga cacagctcaa acttgaaatg tgttattttg CGAGTCGTAT ATTGAGCCCT tcctattccc gctaagtgcg gcctcagtgt	GGTCGCCAGG CTCAAGTCCG gcaagaaacg atcagctgtg agttgtccac cagatcagct ttccatagaa attgtaa ttcgtctgcag <u>a</u> ttcggtcttt ( 130nt ) AAGTGCAAGC TCCAAGTGCT tattatgcga gccagacatt ggcctctgtc	CCGATGTCAT GCGCCAATgt gattgtotgc gcagctggta aagg ttattt caacacctgc gtgttcggtc gcaggcaacg gatcaacagt aaagttgtcc tacagAACTA ( 610nt ) GTACGAGAAG g ttgccagaa ttgccagaac	gagtectee attgeeetgg ct tetgeagg gg <u>tagaa</u> ac tttaaattg gatgtetge gtegatteg acaate gcaa creercAce GAGTGTGTG CCATgtaagt ettaatteegg aatteeagg actgecaca	ttacttctag caacgactga cotattattc caacggattg tgtgtcaatt tccactct c tgccttaact tgaactgaac	gtaatcetce acatctcaac aggactggag actgcgotca tcgtgaactg cagtttatag cgtggctcag aagtttagat gatcagttat AACAAATCAG CCACAAGATG ttattattaa acettcgggt ttcagcagtt	<ul> <li>₩ 285</li> <li>-1837a</li> <li>₩ 285</li> <li>-1837b</li> <li>420</li> <li>₩ 285</li> <li>-1837c</li> <li>600</li> <li>Exon5-6</li> <li>1500</li> <li>₩ 185</li> <li>-1397</li> <li>1680</li> </ul>
GAAGATCAGT CTTACGTCTG gttaatc oct acttgcaaag gcttcttgtc aactagtcga atatgtcgga cacagctcaa acttgaaatg tgttatttg CGAGTCGTAT ATTGAGCCCT tcctattccc gctaagtgcg gcctcagtgt	GGTCGCCAGG CTCAAGTCCG gcaagaaacg atcagctgtg agttgtccac cagatcagct ttccatagaa attgtaa ttcgtctgcaga ttcggtcttt ( 130nt ) AAGTGCAAGC TCCAAGTGCT tattatgcga gccagacatt ggcctctgtc tacttattaa	CCGATGTCAT GCGCCAATgt gattgtotgc gcagctggta aagg ttattt caacacctgc gtgttcggtc gcaggcaacg gatcaacagt aaagttgtcc tacagAACTA ( 610nt ) GTACGAGAAG g ttgcdgagaa ttgcdgagaa ttgcdgagaac tgacatggc agttaattta	gagtectee attgeeetgg ct tetgeagg gg <u>atagaaac</u> tttaaatttg ggtggtggatteg geogatteg geogatteg geogatteg acaate ggaagactaagg actgeogag actgeogag actgeogag	ttacttctag caacgactga cotattattc caacggattg tgtgtcaatt tcoacgtct c tgccttaact tgaactgaac dgactgaacg agrocAGTCCG AGTCCAGTCCG gtgatgcata gacctgtcgtc tc gtatccaa tagtttgagc	gtaatcetce acatetcaac aggactggag actgegetea tegtgaaetg cagtttatag gateagtta AACAAATCAG CCACAAGATG ttattattaa acettegggt tteageagtt tecagtat	<ul> <li>₩ 285</li> <li>-1837a</li> <li>₩ 285</li> <li>-1837b</li> <li>420</li> <li>₩ 285</li> <li>-1837c</li> <li>600</li> <li>Exon5-6</li> <li>1500</li> <li>₩ 185</li> <li>-1397</li> <li>1680</li> <li>1740</li> </ul>
GAAGATCAGT CTTACGTCTG gttaatc cct acttgcaaag gcttottgtc aactctgaca aacaagttca atatgtcgga cacagctcaa acttgaaatg CGAGTCGTAT ATTGAGCCCT tcctattccc gctaagtgcg gcctcagtgt tgtttctta aagttcgaat	GGTCGCCAGG CTCAAGTCCG gcaagaaacg atcagctgtg agttgtccac cagatcagct ttccatagaa attgtaa ttcgtctgcag <u>a</u> ttcggtcttt ( 130nt ) AAGTGCAAGC TCCAAGTGCT tattatgcga gccagacatt ggcctctgtc	CCGATGTCAT GCGCCAATgt gatgtotgc gcagctggta alagg ttattt caacacotgc gtgttcggtc ggtgtcggtc gatcalacagt aaagttgtcc tacagAACTA ( 610nt ) GTACGAGAAG g ttggcagaa ttggcagaa ttggcagaa tgacatggc agttattta agtaatagc	gagtcetcee cgeqattete attgeeetgg ct tetgeagg gg <u>atagaa</u> ac tttaaatttg gatgtetge gtegatteg gacaate geaa creecrece GGAGTIGTIG GGAGTIGTIG CCATgtaagt cttaatteeg aaatteeagg actgte <u>acaa</u> gagactaaac aagttaaaca	ttacttctag caacgactga cetattattc caacggattg tgtgtcaatt tco <u>aca</u> tct c tgacettaact tgaactgaac dgacegaca agacetggtac cattgtcgtc tc gtatccaa tagttcaa tagttcaa	gtaatectee acateteaac aggaetggag cagtgaaetg cagtttatag cgtggeteag aagtttagat gateagttat AACAAATCAG CCACAAGATG ttattataa acettegggt tteageagtt tecageagtt	<ul> <li>₩285</li> <li>-1837a</li> <li>₩285</li> <li>-1837b</li> <li>420</li> <li>₩285</li> <li>-1837c</li> <li>600</li> <li>Exon5-6</li> <li>1500</li> <li>₩185</li> <li>-1397</li> <li>1680</li> </ul>

Figure 1 (genes IV and V) (legend opposite)

band was revealed under stringent conditions of hybridization and the size of each snoRNA was satisfactorily detected as expected. The 5' end of the snoRNAs was further mapped by reverse transcription assays with the same labeled oligonucleotide. In most cases, a major cDNA product was obtained for each snoRNA (Figure 2B). The 5' end of each snoRNA is essentially homogenous and is at the position predicted. Particularly in the case of  $\Psi$ 28S-1135, the mapping assay revealed two bands that correspond in fact to two isoforms varying in length as expected for 151 nt and 142 nt, respectively.

To investigate the expression of different isoforms in the same intron, four oligonucleotides specific to \Perp28S-1192a, \Perp28S-1192b, \Perp28S-1192c and  $\Psi$ 28S-1192d were designed and used to identify each isoform in total cellular RNA. The results of reverse transcription demonstrated that the polycistronic precursor from host gene IV was processed to release correctly four individual snoRNAs as expected (Figure 2C). In some cases, higher molecular mass bands corresponding to the polycistronic precursor were observed in the reverse transcription of snoRNAs (Figure 2D).

#### snoRNA host gene VI

			and a second second	1999 199 2020 2020 20 2020	220 5225
TTCTCGTCTG CTGGTCCAGT G					Exon2
TCGTTGGGCT ATGCCTACGT C	AACTTCCAG	CAGCCAGCCG	ATGgtgagta	atccccgcaa	
aagtagtgta ( 1090nt ) c	gattcatat	ct			
		cgcaactc	ctacccattg	gttggtcaga	Ψ18S
tttatctctt tttgacacga c	taagattca	atgttgccag	agaaagcaat	gcaacttacc	-1347a
atgtagatgg ctgatctatg c	catccacag	aagtagaagc	ataacaaac		
			a	ctaaataata	1380
catcatatga ttagactcga a	ggattatgc	cctttcttat	ctaagcttct		
5 5 55	56			agcaacttet	Ψ18S
acccattggt tggtcagatt t	aacccattt	tgacacgact	aagattcata	gttgccagag	-1347b
aaagcaatgc gacttaccat g	ctgatgggt	tttgcatacc	atctgcataa	caaacactca	
atgtatcaca tga	, , , , , , , , , , , , , , , , , , , ,	-			
	ataagattg	tccttgtatt	tagacatct		
			_	gcaactccta	<b>W18S</b>
cccattggtt ggtcagattc a	accetttt	gacacgacta			-1347c
aaagatcaga cttgacttac c				and the second se	
tgaagagtca cattg		333		5-5-5-5-5-	
	aagcccctt	caacttataa	tgatcttgat	atacccaaat	1800
cgtgtcggta gtaaccaaat g					Exon3
ACCATGAACT TTGACCTGGT T					Linono
ACCATOAACT TTOACCTOOT T	COCANCIANO	CCCATICOCA	TIAIOIOOIC	ICAOCOTOAT	
and Data has the second	слт				
snoRNA host gene	VШ				
CGTCTGGTCT ACGACGTGAA G				CGAGGAGGCC	Exon4
AAGgtgagtt atctagtctg (	190nt )	taattaactt	aaatatcgaa		
				agggetttgt	ψ28S
cgaagaccgt tttgcgtgta g					-3327a
tttttgatac gacggtctct g	attcgacaa	atcccagttg	attcagtaac	ttttacgtgc	
aattacaaaa					

gaaaccaagg agtataaagc attatattgt aattctg

	gaaaccaagg	aylalaaayc	accalactyc	aductury		
				tgg	gctacgtcga	$\Psi 28S$
agaccgattt	ccgcttgttg	ccatgtgtct	attggttcaa	atcgaagccc	aaagcaattt	-33271
ttgatacgac	ggtctctgat	tcggcattcc	acagtcgtct	gagtaacttt	tacgtgcaat	
tacaatg						
caa	ttaaatgaaa	gtggtagtgc	aattgtgtag	aaataaaacg	tcttaaaatt	660
aacccacatt	ttetttgett	cttttgcagT	ACAAGTTGTG	CAAGGTCAAG	AAGACCCAGC	Exon5
IGGGAGCCAA	GGGAGTTCCT	TTCCTGGTTA	CACACGACGG	TCGCACCATC	CGCTACCCGG	
ATCCCCTGAT	CCACGCCAAC	GATTCCGTGC	AGGTGGACAT	TGCCTCTGGC	AAGATCACCG	840
ACTACATTAA	GTTCGATTCT	Ggtaagcatc	cgctatcgta	gatcatcgtg	ttca	
					ttggct	Ψ28S
ggttcgaaga	ccacaaaacg	ctcctcgaaa	gttgcgtggt	tcaaaaacaa	taagccaaaa	-33270
gcaatttttg	tcacgacggt	ctctgatgcg	gcattccaaa	gtcgcttcag	taacttttac	
gtgcaatgac	agtc					
	gcccaa	tattattaac	caagtcgtat	gattccaact	aactggtata	1080
tttctccgtc	tccacagGCA	ACCTCTGCAT	GATCACCGGA	GGCAGGAATT	TGGGACGTGT	Exon6

#### snoRNA host gene VII

CGACTGCACC AACATTGCTG AGGATAGCAT CATGGATGTG GCCGACTTCg taagtacatg Exon2 gggtacacag ( 190nt ) gtct taccge accaagettt ceactgeeet ceaagttaac ψ 28S caggggctct ggataaaaag ttgtggtt<mark>at agtg</mark>agttca ctccattttg cgataaactg -2719ccageggaet tteeegtgee geeggeggtt aacaaettet gga<mark>aca</mark>aeg 420 c taaatgtaga tgatagtcat ( 670nt ) ataaatctat tttaca ggcc acgccgattt gccgtcagag ψ18S agtgagettt atgegteget etgatggeag tgtaeteagt gge<mark>agagea</mark>a aaaattggte -531 actettettt atttgggete tetgttetae ceaetegaet ettageetta acaaagaaag agacattt 1320 ag cccagettae egetageagg agetgeecea aagateeage eeagteeaag

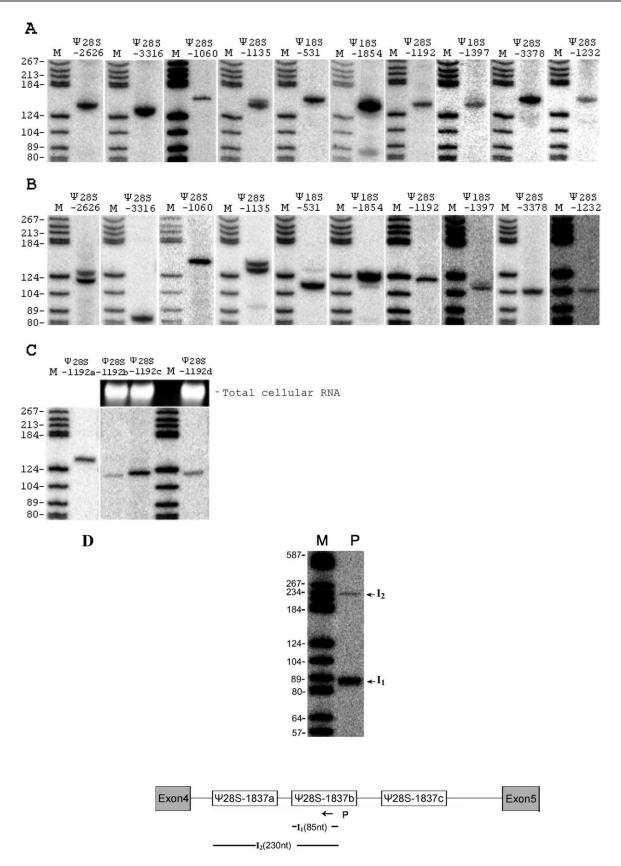
ctgaactaac coggaatacc ctototatto cgocogoagG AGAAGTACAT CAAGGCCCGC CTTAAGGTCA ACGGCAAGGT GAACAACCTG GGCAACAACG TCACCTTCGA GCGCTCCAAG

### Analysis of pseudouridylation sites predicted by the novel box H/ACA snoRNAs

In addition to nine snoRNAs whose functions have been described,<sup>20</sup> ten novel box H/ACA snoRNAs were predicted to guide 15 pseudouridylation sites (Figure 3) according to the relationship between the structure and function of this snoRNA gene family.<sup>4</sup> Among the 15 pseudouridylation sites, seven, i.e.  $\Psi$ 1060,  $\Psi$ 1135,  $\Psi$ 1192, **Figure 1.** The sequences of ten intronic box H/ACA snoRNA gene clusters from *D. melanogaster*. Exons are in capital letters; introns are in lower case. The coding regions for snoRNAs are shaded, and box H/ACA are boxed.

Ψ2533, Ψ2626, Ψ3316 and Ψ3378 in 28 S rRNA have been experimentally verified by Ofengand *et al.*<sup>19</sup> In addition, two pseudouridylation sites, Ψ531 and Ψ1397 in 18 S rRNA, are conserved in the yeast *Saccharomyces cerevisiae* and/or mammals.<sup>22</sup> Furthermore, the isoforms of nine known snoRNAs were further analyzed and predicted to guide four novel pseudouridylation sites based on their functional elements that are absent from other isoforms (Figure 3B).

Exon3



**Figure 2**. Positive detection and mapping of 5' end of novel snoRNAs. A, Northern blot analyses. Aliquots of 30  $\mu$ g of total cellular RNA were separated in each lane and hybridized with the labeled oligonucleotide probes described in Materials and Methods. Lane M, molecular mass markers (pBR322 digested with HaeIII and 5'-end-labeled with [ $\gamma$ -<sup>32</sup>P]ATP). B, Reverse transcription analyses of ten novel snoRNAs with the same primers as Northern blots. C, Reverse transcription analyses of the expression of four isoforms  $\Psi$ 28S-1192 in the intron 4 of snoRNA host gene I by primers specific to each isoform. These primers are complementary to the divergent regions of 3' end of each

CMC-alkali-treated *D. menalogaster* RNA was applied to determine the new rRNA pseudouridylation sites. Six sites, ¥1854, ¥1937 in 18 S rRNA and \$\Psi\_483\$, \$\Psi\_1837\$, \$\Psi\_1850\$, \$\Psi\_2938\$ in \$28\$ S rRNA, were precisely mapped (Figure 4). However, U1232 in 28 S rRNA, a pseudouridylation site predicted by  $\Psi$ 28S-1232, was found to be unmodified while two known pseudouridylation sites,  $\Psi$ 1227,  $\Psi$ 1240, were clearly shown in the same assay (Figure 4E). In addition, two pseudouridylations,  $\Psi$ 1347 in 18 S rRNA and  $\Psi$ 3801 in 28 S rRNA, predicted by Yuan et al.<sup>20</sup> were also determined in the primer extension. For some unknown reason, the extension of primer Dm28S584, which was applied to analyze  $\Psi$ 584 in 28 S rRNA, failed to show any results (data not shown).

# Large number of box H/ACA snoRNA isoforms in *D. menalogaster*

Multiple isoforms of the box H/ACA snoRNA genes were identified from the D. menalogaster genome. Of the 19 different box H/ACA snoRNA genes, ten have at least one variant (Table 1). For example,  $\Psi$ 18S-1377 has five isoforms in the same intron, while five isoforms of  $\Psi$ 28S-3316 are distributed in the two introns of host gene III. Evidently, multiple isoforms of one snoRNA gene are frequently found within one intron, suggesting the important role of local duplications for the cluster formation. The sequence alignment of six  $\Psi$ 28S-1135 isoforms and three  $\Psi$ 18S-1854 isoforms clearly demonstrates the duplication of the two snoRNA genes, especially in the three regions. They all consist of different isoforms of  $\Psi$ 28S-1135 and  $\Psi$ 18S-1854, in the second intron of host gene I (Figure 5). The degree of sequence similarity between the isoforms varies from 99% to 70%, implying the duplication events in different periods and a trend of enlarging the intronic snoRNA clusters.

In many cases, although mutations including insertions or deletions had occurred frequently, the secondary structures of the isoforms remained unchanged and the isoforms kept in common at least one functional element. Interestingly, the accumulation of mutation in the isoforms would lead to partial alternation of snoRNA's function in loss or gain of rRNA complementary sequences. For example, the variations in the 5'-hairpin of 28 S¥1135d and 3'-hairpin of 28 S¥1192a result in novel complementary sequences for 28 S rRNA at U483 and U2533 pseudouridylation, respectively. On the other hand, the sequence variation among

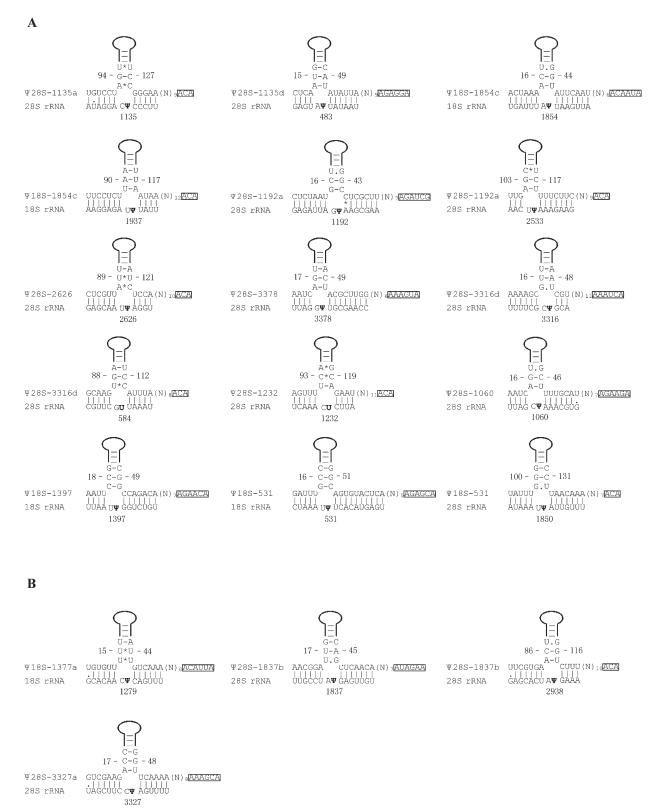
the isoforms can also cause the loss or change of the guide sequence for rRNA pseudouridylation, such as the 3'-hairpin of  $18 \text{ S}\Psi 1347\text{b}$  and  $28 \text{ S}\Psi 1837a$  (see Table 1).

### Discussion

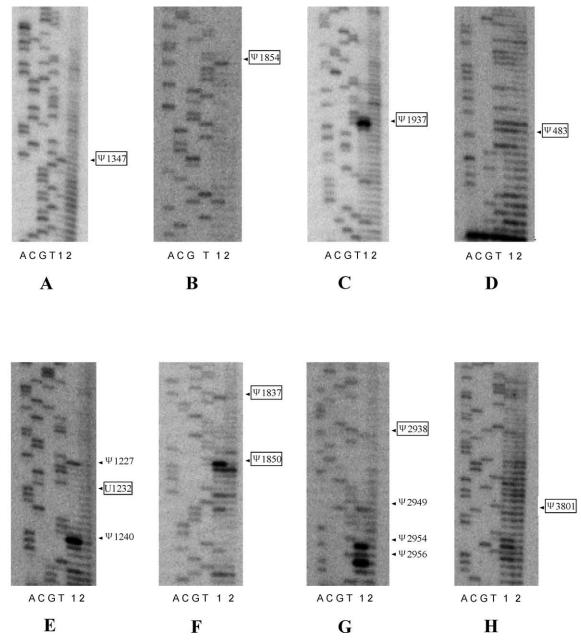
Ten intronic snoRNA gene clusters were identified from *D. melanogaster* in this study, demonstrating for the first time a novel polycistronic organization of snoRNA genes in animals. In fact, all intronic snoRNA gene clusters identified so far from plants consist of homogeneous box C/D snoRNA genes with few exceptions, such as the rice hsp70 gene where two box H/ACA snoRNA genes are mixed with four other box C/D snoRNA genes in the first intron of the host.<sup>15</sup> Here, we show homogeneous box H/ACA snoRNA gene clusters without any intervening box C/D snoRNA. Therefore, the intronic box H/ACA snoRNA gene clusters are first demonstrated not only in *D. melanogaster* but also in eukaryotes.

Up to now, more than 30 box C/D snoRNA gene variants including three methylation guides for snRNA and five orphan guides with unknown target have been identified from D. melanogaster (GenBank accession number U40615 for Z1 and AJ010684 for Z5 sno(RNA)).<sup>20,23</sup> Similar to mammals, all the methylation guide snoRNAs characterized so far in D. melanogaster are intronencoded and arranged strictly in the mode of one snoRNA per intron. For example, DUHG1, a human UHG-like non-coding RNA gene in D. melanogaster, encodes 16 box C/D snoRNAs in their 16 introns, respectively (Figure 6), representing an outstanding gene organization and expression strategy for the C/D snoRNA gene family in *D. melanogaster.*<sup>24</sup> Interestingly, we have shown that box H/ACA snoRNAs in D. melanogaster possess a different genomic organization, that is, intronic gene clusters. A typical case is snoRNA host gene I, in which 14 box H/ACA snoRNA genes are nested within three introns of a ribosomal protein gene RpS5. Among 47 snoRNAs analyzed here, 42 were found in the intronic clusters. The high proportion of box H/ACA snoRNAs in the intronic clusters suggests that this kind of gene organization may be prevalent in the D. melanogaster genome. The polycistronic snoRNAs in an intron imply a processing mechanism that involves both endonucleolytic and exonucleolytic cleavages, substantially different from processing a singleton of intronic snoRNA

isoform, respectively. The sequence of the primer for  $\Psi$ 28S-1192a is completely different from the three others, and there are at least six nucleotide differences among the primers for  $\Psi$ 28S-1192b,  $\Psi$ 28S-1192c and  $\Psi$ 28S-1192d. D, Detection of a polycistronic snoRNA precursor in the reverse transcription of total RNA. Lane P, reverse transcription with primer P $\Psi$ 28S-1837b, which is specific antisense to the 3' end of  $\Psi$ 28S-1837b. Lane M, molecular mass markers. I<sub>1</sub> and I<sub>2</sub> indicate cDNA bands from mature  $\Psi$ 28S-1837b and the precursor containing two box H/ACA snoRNAs, respectively, as illustrated below.



**Figure 3.** Predicted pseudouridylation guide duplex between snoRNAs and rRNAs. The snoRNA sequences in a 5' to 3' orientation are shown in the upper strands, while rRNA sequences in a 3' to 5' orientation are shown in the lower strands. The two sequence motifs are boxed and the upper parts of the hairpins are represented by continuous lines. The positions of pseudouridine bases are indicated by numbers.  $\Psi$  represents rRNA pseudouridine sites that are conserved in *S. cerevisiae* and/or mammals,<sup>22</sup> and the known rRNA pseudouridine bases determined previously<sup>19</sup> or mapped here. U denotes a predicted pseudouridine site that has not been confirmed experimentally. Only one isoform is shown if the snoRNA has more isoforms. A, Pseudouridylation sites predicted by novel snoRNAs. B, Four novel sites predicted from isoforms of snoRNAs identified previously.



**Figure 4**. Mapping of rRNA pseudouridylation sites predicted by the novel snoRNAs. Lane 1, reverse transcription with CMC-treated total RNA; lane 2, reverse transcription reaction with CMC-untreated total RNA. Lanes A, C, G and T, the rDNA sequence ladder; positions of pseudouridine residues are indicated by arrows. Pseudouridylation sites boxed are predicted in this work, and those without boxes were previously identified by Ofengand and Bakin.<sup>19</sup>

(Figure 6).<sup>14,17</sup> The adoption of different organizations and expression strategies for the snoRNA genes may reflect intrinsic differences in gene regulation and functional evolution between the two classes of snoRNAs in *D. melanogaster*. It is worth noting that most intronic snoRNAs are produced by a splicing-dependent processing pathway involving exonucleolytic trimming of the debranched lariat.<sup>25</sup> However, another minor, splicing-independent mode for intron-encoded box C/D snoRNAs has been found in yeast and mammals,<sup>26–28</sup> suggesting the complexity of intronic snoRNA biogenesis in eukaryotes. In Figure 6, we describe a splicing-

dependent pathway for snoRNA release from host introns, but the possibility that the pre-mRNA is a substrate of endonucleolytic cleavage, particularly, when the splicing efficiency is reduced in the organism, cannot be ruled out.

Although endonucleolytic activity is absolutely necessary for processing the polycistronic transcript of snoRNAs, diverse endonuclease(s) may be involved in the maturation of the snoRNA. Endonuclease RNase III that recognizes a strong potential secondary structure has been proved to play a key role for processing the polycistronic transcript of snoRNAs in the budding yeast.<sup>17,29</sup>

	-1135a	-1135b-113		-1135e	-1135f -		
		11552 115					
Exc	on2						xon3
				1	ł		
			1	1	1	1	
			і т	· i Π	i I	1	
в					Ψ <b>28S</b> -	1135	
		5 G S N	31	+			
I II		cctattctca					50
1000							50
Ш	aagacaatag	-aagtg	tcttt-ac	aa-GT-	T	TA**AG	58
Ι	<b>333303000000</b>	GCCATACTGA	CCAMAACCAA	3 3 3 m C C 3 C m 3		<b>7 7 7 00000000</b>	110
П			GCATAACCAA	AAAIGGACIA	AIAICAAIIG	AAA11G1111	110
ш		CGCTAAG	T****	**	CACCAC-	TAAC	112
m	G	C GCIAAG	1	IA A	GAC GAG	IAA G	112
Ι	AACTATCCTT	GTGCCTGAAG	ATATAAGCCA	CTTCCTGCTC	TCCTGGAAAT	ТААААСААСА	170
П		T					170
Ш	TA	T	*	C	T-	GT	171
	+			4	Ψ <b>18S</b> -'	1854	
Ι	CTAaagttaa	gcattattca	ccaagggtat	cattatcCGT	ССААААСТАА	AACTGCAGCG	230
II							230
III	Tg-a**-g	aa-gc	tt-ca-ag	g-g			229
Ι		CAGCGTTGTG					290
Π		G					290
III	G		-C	GG	T	G	289
T	0.000				00000	¥	0.15
I		GCATGGCCTC					347
II					*	AA-	346

Ψ18S -1854a

Ψ28S Ψ28S Ψ28S

Ψ28S

Ψ18S

-1854b

Ψ28S Ψ18S

Figure 5. Duplication of snoRNA-coding region in the second intron of snoRNA host gene I. A, snoRNA genes are represented by boxes. Three duplication regions for two snoRNA genes are indicated. B, Sequence alignment of the duplication regions. snoRNA genes are in upper-case letters and indicated by arrowheads. Nucleotide identities are denoted by hyphens and those absent from either sequence by asterisks.

----- ----\*----- -----AA-

Recently, dicistronic tsnoRNA, a heterogeneous cluster consisting of a box C/D snoRNA and tRNA, was found in plants.<sup>30</sup> Dicistronic precursors transcribed from the cluster were processed by endonuclease RNase Z that specifically recognizes 3' ends of tRNAs, instead of intergenic sequences which were less than ten nucleotides between the two RNAs.<sup>30</sup> Intergenic spacers between the box H/ACA snoRNA genes in D. melanogaster are remarkably short and rich in A and T, so they can hardly form secondary structures similar to those that are involved in RNase

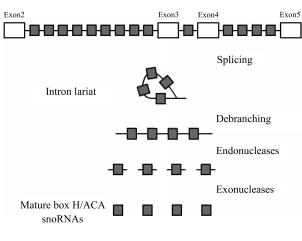
Ш

III cleavage. Furthermore, the analysis of sequences reveals no conserved element among 33 intergenic spacers from the snoRNA genes. It is very likely that the processing signals recognized by endonuclease are not in the intergenic spacer regions but in box H/ACA snoRNAs, especially in the high structure of the snoRNAs. Endonuclease that is involved in the processing of polycistronic transcripts of box H/ACA snoRNAs may contribute to the amplification of intronic clusters of this snoRNA family and the high diversity of rRNA pseudouridylation in *D. melanogaster*.

345

Α

Ψ28S



RpS5 gene (snoRNA host gene I) pre-mRNA

dUHG1 pre-mRNA

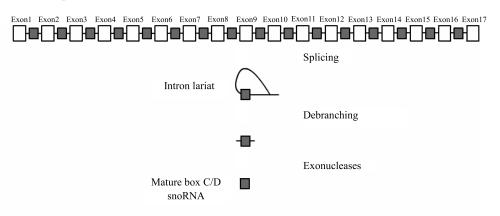


Figure 6. Different gene organizations and expression strategies for the two classes of snoRNAs in D. melanogaster.

### **Materials and Methods**

# Computational search for box H/ACA in the *D. melanogaster* genomic database

Complete sequences of eight *D. melanogaster* genes, CG8922, CG1883, CG9696, CG3203, CG10576, CG5119, CG11276 and CG7434, were obtained from the Flybase.<sup>31</sup> All the intron sequences in the genes were analyzed by using our computer program that takes into account both the sequence motifs and secondary structures in box H/ACA snoRNAs. The search for isoforms of the snoRNAs in the *D. melanogaster* genome was performed by using BLAST<sup>32</sup> and FASTA<sup>33</sup> programs. Sequence alignments and comparison of snoRNAs were performed by using Clustal X 1.8 and DNAstar packages.

### **RNA extraction and analyses**

Fresh wild-type *D. melanogaster* larva were cultured and collected for the RNA extraction. Total cellular RNA was isolated and purified according to the method of guanidine thiocyanate/phenol/chloroform.<sup>34</sup> An aliquot of 30  $\mu$ g of total RNAs was analyzed by electrophoresis on 8% (w/v) acrylamide/7 M urea gels. Electrotransfer onto nylon membrane (Hybond-N + ; Amersham) was followed by UV irradiation for five minutes. Hybridization with 5'-labeled probes was performed as described.<sup>23</sup> Reverse transcription was carried out in a 20  $\mu$ l reaction mixture containing 20  $\mu$ g of *D. melanogaster* total RNA and 20 ng of 5'-labeled primer in the presence of 500  $\mu$ M dNTPs. After denaturation at 65 °C for five minutes, the mixture was cooled to 42 °C for ten minutes, and then 200 units of MMLV reverse transcriptase (Promega) was added and incubated at 42 °C for 60 minutes. The cDNA synthesized by reverse transcription was analyzed by electrophoresis on 8% acrylamide/7 M urea gels.

# Mapping of ribose pseudouridylation by CMC-primer extension method

Mapping of *D. melanogaster* ribosomal pseudouridines was performed essentially as described by Bakin and Ofengand.<sup>35</sup> The *D. melanogaster* 28 S and 18 S rDNA

were amplified by PCR with the primer pairs Dm28SL1/ Dm28SR1, Dm28SL2/Dm28SR2 and Dm18SL/Dm18SR, then cloned into the SmaI site of plasmid pUC18. An rDNA sequence ladder was prepared with the same primer used for rRNA pseudouridylation mapping and run in parallel with the reverse transcription reaction as a molecular mass marker.

#### Oligodeoxynucleotides

Oligonucleotides were synthesized and purified by Sangon Co. (Shanghai, China). The sequences of oligonucleotide probes and primers used for Northern blotting and reverse transcription were as follows: 5'-GCAGGACCGCAATGGCTTCT-3' (P\P28S-2626); 5'-GCT ATTGTGAGGATTGATTTTGC-3' (PΨ28S-3316); 5'-GCG GGGCGTC-3' (PΨ28S-1060); AGATTTTAGAAGA 5'-TTAAATTCCAGGAGAGCGGG-3' (PΨ28S-1135); 5'-TAGAACAGAGAGCCCAAATAAAGAA-3' (PΨ18S-531); 5'-TTTATTAAGGTTGCAT TTTTGAGG-3' (PΨ18S-1854); 5'-TTTCTCGTGTGCATTAGACTGAT-3' (PΨ28S-1192c); 5'-CCACACTGAGGCAACTGCTGAA-3' (PY18S-1397); 5'-GCTCTGGTTTAGTATGC GGC-3' (P\28S-3378); 5'-CGAAATCCATTGAAAACTGTCTAA-3' (P\v28S-1232); 5'-TTCACGAAATTGACACAGTTTCTA-3' (PΨ28S-1837b). The following oligonucleotides were used for identification of \U28S-1192 isoforms: 5'-GTCACTAGAAGGAA GAAA AGAGC-3' (P\P28S-1192a); 5'-GTGTGCATAGG ACA GTTAAGTGG-3' (P\v28S-1192b); 5'-TTTCTCGTGT GC ATTA GACTGAT-3' (P\P28S-1192c); 5'-CTCATGTGC AT TGAACTAAGCAAT-3' (PΨ28S-1192d). Primers used for rRNA pseudouridylation mapping were as follows: 5'-AACCAGACAAATCGCTCCAC-3' (Dm18S1347); 5'-AA ACAACCGTAACACGCAAGG-3' (Dm18S1854); 5'-TGA TCCTTC CGCAGGTTCACC-3' (Dm18S1937); 5'-CAAT GTCCTTATATGGAAAAAATGC-3' (Dm28S483); 5'-CAC TGTAATCATATAAATCTATCAGCACTT-3' (Dm28S584); 5'-GATCT TCATATCAAGAAAGTTAAGGTTC-3' (Dm28S1232); 5'-TTTATGGTCGTTCCTGTTG CC-3′ (Dm28S1850); 5'-CGTTTTATTAAAGAATTTGTTTGCG-3' (Dm28S2938); 5'-ACCACTTACAACACCTTGCCTG-3' (Dm28S3801). Primers used for PCR of D. melanogaster 28 S and 18 S rDNA were: 5'-GGTTATGTTATTATTCTT CGTTGGTT-3' (Dm28SL1); 5'-AATTCGCTTTGTTTATAT AGTTAGGC-3' (Dm28SR1); 5'-GCCTAACTA TATAAAC AAAGCGAATT-3' (Dm28SL2); 5'-AATTGATGACGAG CTGTTTGG-3' (Dm28 SR2); 5'-TGGTTGATCCTGCCAG TAGTTAT-3' (Dm18SL); 5'-AACCCATCTTCGTTTTA TTTTGA-3' (Dm18SR). The primers and probes used in reverse transcription and rDNA sequencing were 5'end-labeled with  $[\gamma^{-32}P]ATP$  (Yahui Co.) and submitted to purification according to standard laboratory proto-cols as described.<sup>36</sup>

#### Database accession codes

All snoRNA gene sequences identified in this study have been deposited in the EMBL database under accession numbers from AJ629193 to AJ629216 and from AJ629256 to AJ629278.

### Acknowledgements

We are grateful to Xiao-Hong Chen for technical assistance and Quan-Sheng Du for *D. melanogaster* 

culture. We also thank Professor Mohsen Ghadessy for revising the English text. This research was supported by the National Natural Science Foundation of China (key project 30230200, 30170216), and by a Fund from the Ministry of Education of China to L.H.Q.

### References

- 1. Kiss, T. (2002). Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell*, **109**, 145–148.
- Filipowicz, W. & Pogaèiæ, V. (2002). Biogenesis of small nucleolar ribonucleoproteins. *Curr. Opin. Cell. Biol.* 14, 319–327.
- Balakin, A. G., Smith, L. & Fournier, M. J. (1996). The RNA world of the nucleolus: two major families of small nucleolar RNAs defined by different box elements with related functions. *Cell*, 86, 823–834.
- Ganot, P., Bortolin, M. L. & Kiss, T. (1997). Related site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell*, 89, 799–809.
- Smith, C. M. & Steitz, J. A. (1997). Sno storm in the nucleolus: new roles for myriad small RNPs. *Cell*, 89, 669–672.
- Bachellerie, J. P., Cavaille, J. & Qu, L. H. (2000). Nucleotide modifications of eukaryotic rRNAs: the world of small nucleolar RNA guides revisited. In *The Ribosome: Structure, Function, Antibiotics and Cellular Interactions* (Garrett, R. A., Douthwaite, S., Liljas, A., Matheson, A., Moore, P. B. & Noller, H., eds), pp. 191–203, ASM Press, Washington, DC.
- Tycowski, K. T., You, Z. H., Graham, P. J. & Steitz, J. A. (1998). Modification of U6 spliceosomal RNA is guided by other small RNAs. *Mol. Cell*, 2, 629–638.
- Clouet d'Orval, B., Bortolin, M. L., Gaspin, C. & Bachellerie, J. P. (2001). Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNATrp intron guides the formation of two ribosemethylated nucleosides in the mature tRNATrp. *Nucl. Acids Res.* 29, 4518–4529.
- 9. Lowe, T. M. & Eddy, S. R. (1999). A computational screen for methylation guide snoRNAs in yeast. *Science*, **283**, 1168–1171.
- Maxwell, E. S. & Fournier, M. J. (1995). The small nucleolar RNAs. *Annu. Rev. Biochem.* 35, 897–934.
- Bachellerie, J. P., Cavaille, J. & Huttenhofer, A. (2002). The expanding snoRNA world. *Biochimie*, 84, 775–790.
- Tycowski, K. T., Shu, M. D. & Steitz, J. A. (1996). A mammalian gene with introns instead of exons generating stable RNA products. *Nature*, **379**, 464–466.
- Leader, D. J., Sanders, J. F., Waugh, R., Shaw, P. & Brown, J. W. (1994). Molecular characterisation of plant U14 small nucleolar RNA genes: closely linked genes are transcribed as polycistronic U14 transcripts. *Nucl. Acids Res.* 22, 5196–5203.
- Brown, J. W., Echeverria, M. & Qu, L. H. (2003). Plant snoRNAs: functional evolution and new modes of gene expression. *Trends Plant Sci.* 8, 42–49.
- Liang, D., Zhou, H., Zhang, P., Chen, Y. Q., Chen, X., Chen, C. L. & Qu, L. H. (2002). A novel gene organization: intronic snoRNA gene clusters from *Oryza* sativa. Nucl. Acids Res. **30**, 3262–3272.
- 16. Chen, C. L., Liang, D., Zhou, H., Zhou, M., Chen,

Y. Q. & Qu, L. H. (2003). The high diversity of snoRNAs in plants: identification and comparative study of 120 snoRNA genes from *Oryza sativa*. *Nucl. Acids Res.* **31**, 2601–2613.

- Qu, L. H., Henras, A., Lu, Y. J., Zhou, H., Zhou, W. X., Zhu, Y. Q. *et al.* (1999). Seven novel methylation guide small nucleolar RNAs are processed from a common polycistronic transcript by Rat1p and RNase III in yeast. *Mol. Cell. Biol.* **19**, 1144–1158.
- Dunbar, D. A., Chen, A. A., Wormsley, S. & Baserga, S. J. (2000). The genes for small nucleolar RNAs in *Trypanosoma brucei* are organized in clusters and are transcribed as a polycistronic RNA. *Nucl. Acids Res.* 28, 2855–2861.
- 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.
- Yuan, G., Klambt, C., Bachellerie, J. P., Brosius, J. & Huttenhofer, A. (2003). RNomics in *Drosophila melanogaster*: identification of 66 candidates for novel non-messenger RNAs. *Nucl. Acids Res.* 31, 2495–2507.
- Hirose, T. & Steitz, J. A. (2001). Position within the host intron is critical for efficient processing of box C/D snoRNAs in mammalian cells. *Proc. Natl Acad. Sci. USA*, 98, 12914–12919.
- Ofengand, J., Bakin, A., Wrzesinski, J., Nurse, K. & Lane, B. G. (1995). The pseudouridine residues of ribosomal RNA. *Biochem. Cell. Biol.* 73, 915–924.
- Zhou, H., Chen, Y. Q., Du, Y. P. & Qu, L. H. (2002). The Schizosaccharomyces pombe mgU6-47 gene is required for 2'-O-methylation of U6 snRNA at A41. Nucl. Acids Res. 30, 894–902.
- Tycowski, K. T. & Steitz, J. A. (2001). Non-coding snoRNA host genes in *Drosophila*: expression strategies for modification guide snoRNAs. *Eur. J. Cell. Biol.* 80, 119–125.
- Tollervey, D. & Kiss, T. (1997). Function and synthesis of small nucleolar RNAs. *Curr. Opin. Cell. Biol.* 9, 337–342.
- 26. Villa, T., Ceradini, F., Presutti, C. & Bozzoni, I. (1998). Processing of the intron-encoded U18 small

nucleolar RNA in the yeast *Saccharomyces cerevisiae* relies on both exo- and endonucleolytic activities. *Mol. Cell. Biol.* **18**, 3367–3383.

- 27. Cavaille, J., Vitali, P., Basyuk, E., Huttenhofer, A. & Bachellerie, J.-P. (2001). A novel brain-specific box C/D small nucleolar RNA processed from tandemly-repeated introns of a non-coding RNA gene in rat. J. Biol. Chem. 276, 26374–26383.
- Hirose, T., Shu, M.-D. & Steitz, J. A. (2003). Splicingdependent and -independent modes of assembly for intron-encoded box C/D snoRNPs in mammalian cells. *Mol. Cell*, **12**, 113–123.
- Chanfreau, G., Legrain, P. & Jacquier, A. (1998). Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism. *J. Mol. Biol.* 284, 975–988.
- Kruszka, K., Barneche, F., Guyot, R., Ailhas, J., Meneau, I., Schiffer, S. *et al.* (2003). Plant dicistronic tRNA-snoRNA genes: a new mode of expression of the small nucleolar RNAs processed by RNase Z. *EMBO J.* 22, 621–632.
- The FlyBase Consortium (2003). The FlyBase database of the *Drosophila* genome projects and community literature. *Nucl. Acids Res.* 31, 172–175. http:// flybase.org/.
- flybase.org/.
  32. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
- Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA*, 85, 2444–2448.
- Chomczynski, P. & Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162, 732–735.
- 35. Bakin, A. & Ofengand, J. (1993). Four newly located pseudouridylate residues in *Escherichia coli* 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. *Biochemistry*, **32**, 9754–9762.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edit., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

### Edited by M. Yaniv

(Received 5 March 2004; received in revised form 24 May 2004; accepted 14 June 2004)