RNomics: a computational search for box C/D snoRNA genes in the *D.melanogaster* genome

Maria Carmela Accardo [(1)], Ennio Giordano [(1)], Sara Riccardo[(1)], Filomena Anna Digilio[(2)], Giovanni Iazzetti [(1)], Raffaele A. Calogero [3)] and Maria Furia [(1)]

(1) Dept. of Genetics, General and Molecular Biology, University of Naples Federico II, Via Mezzocannone 8, 80134, Napoli, Italy
mfuria@unina.it

(2) Institute of Protein Biochemistry, Consiglio Nazionale delle Ricerche, 80131, Napoli, Italy
digilio@igb.cnr.it

(3) Dept of Clinical and Biological Sciences , University of Torino, Regione Gonzole 10, 10043 Orbassano (TO), Italy
raffaele.calogero@unito.it

Keywords. Genomics, non-coding RNA, snoRNA, Drosophila

Introduction

Genes producing functional RNAs rather than protein products form a large and variegated class in all genomes, from bacteria to mammals. In higher organisms. non-coding RNA (ncRNA) appears to dominate the whole genomic output, and is not surprising that the range of known RNA-induced phenomena is rapidly expanding. The central importance of RNA signaling to eukaryotic cell has become apparent in the last few years, when a large bulk of evidence has pointed out novel roles for ncRNA molecules in both genetic and epigenetic processes. The family of nc-RNA genes comprises many small nucleolar RNAs (snoRNAs) that guide the maturation or post-transcriptional modification of target RNA molecules. Most snoRNAs fall into two classes called box C/D and box H/ACA snoRNAs, with each class defined by the presence of common sequence motifs and common associated proteins [1]. A few snoRNAs in either class are required for definite pre-rRNA cleavages and essential for viability, whereas most are responsible for the 2'-O-ribose methylation (C/D) or pseudouridylation (H/ACA) of target RNA molecules respectively. The C/D class guides site-specific 2'-O-ribose methylation by base-pairing of the 10-21 nt-long sequence positioned upstream from a D (or an internal D') box to the target RNA, with the nucleotide positioned 5 base pairs (bp) upstream from the D/D' box selected for methylation [2]. Although most of the C/D and H/ACA box snoRNAs are involved in modifications of ribosomal RNA (rRNA), other types of RNA molecules, as tRNAs, snRNAs, and possibly mRNAs, might be recognised as targets. Despite the importance of their functional roles, most of snoRNAs have not yet been identified even in organisms whose genome has been completely sequenced.

Results

We have performed a computational scanning of the Dm genome for C/D snoRNA genes, followed by experimental validation of the putative candidates. Since rRNA methylation in Drosophila has not yet been determined experimentally, 18S and 28S rRNA sequences from Dm were aligned with those of S. cerevisiae (Sc) rRNAs. We defined as putative Drosophila rRNA methylation sites those experimentally defined in yeast [3], [4], and present in regions conserved between Dm and Sc rRNAs. The SNOSCAN program [4] was used to identify snoRNA genes in the Dm genome. The putative Dm methylation sites were considered reliable if they allowed identification of the corresponding Sc snoRNA genes with a score higher than 20. Out of the 16 Sc available methylation sites (http://rna.wustl.edu/snoRNAdb/Sc/Sc-snos-bysite.html) on the small ribosomal subunit (SSU) we defined as "reliable" 10 sites on Dm 18S rRNA, and on Dm 28S rRNA 34 sites out of the 39 available on the Sc large ribosomal subunit (LSU). We identified 44 putative candidate Dm snoRNA genes, all having a SNOSCAN score higher than 20, which is the default parameter defined by Lowe and Eddy (1999). We were able to confirm the expression of 26 snoRNA genes (4 on 18S rRNA and 22 on 28S rRNA) out of the 44

candidates identified in our analysis. In our validation assay, a panel of total RNA samples extracted from various stages of Dm development was analyzed by Northern blot using each specific probe. We cannot at present say whether assays more sensitive than Northern blots would validate more of our predicted candidates. Most of the genes identified were constitutively expressed during Dm development, as expected for snoRNA molecules targeting rRNA, and are arranged in clusters and present in multiple copies, confirming that in invertebrates, as in vertebrates, polycistronic organization is common. All the snoRNA genes in each cluster were usually arranged in a head-to-tail fashion and closely linked. Tandem gene duplication events generate functional redundancy and can establish sequence variability allowing the generation of new snoRNAs for selection. Consistent with this assumption, the identified clusters presented both copies in which the antisense motifs were perfectly conserved among tandemly repeated snoRNA coding units (occasionally with polymorphism substantially restricted to sequences immediately downstream from the C or D' boxes), and divergent copies displaying significant nucleotide changes within the antisense motifs and hence unable to target Dm rRNA. Cluster arrangement, indeed, proved useful in identifying ncRNA genes not specifically targeted by our screening. Inspection of flanking sequences, in fact, occasionally revealed the presence of genes encoding either snmRNAs or snoRNAs of the H/ACA class. In 6 cases the snoRNA genes were located in apparently intergenic regions of the Dm genome and may represent new examples of genes devoid of protein-coding potential, thus broadening the repertoire of Dm genes that produce solely ncRNAs. Organization of the Dm genome was found to be more variegated than previously suspected, with snoRNA genes nested in both the introns and exons protein-coding genes. Intriguingly, in 5 cases the snoRNA was in fact encoded within an exon of a protein-coding gene, and in one case even with polarity opposite to that of the host gene. This type of arrangement has not been previously described for any snoRNA coding unit, and may reflect a still uncharacterized mechanism of snoRNA biogenesis based on the alternative production of the overlapping mRNA/snoRNA molecules. Thought certainly not exhaustive, our analysis reveals that snoRNA genes can largely be missed by cDNA cloning procedures, either because their relative low abundance or due to other factors potentially affecting their representation in cDNA libraries. cDNA cloning and computational screens proved to be complementary, and presumably both types of approaches will be necessary to identify the full complement of snoRNA genes hidden in the genome.

References

[1] Bachellerie JP, Cavaillè J, Hüttenhofer A. (2002) The expanding snoRNA world. *Biochimie*, 84, 775-790

[2] Nicoloso M, Qu LH, Michot B, Bachellerie JP. (1996) Intron-encoded, antisense small nucleolar RNAs: the characterization of nine novel species points to their direct role as guides for the 2'-O-ribose methylation of rRNAs. *J Mol Biol*.12, 178-95

[3] Kiss-Laszlo Z, Henry Y, Bachellerie JP, Caizergues-Ferrer M, Kiss T. (1996) Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell*., 28, 1077-88.

[4]Lowe TM, Eddy SR. (1999) A computational screen for methylation guide snoRNAs in yeast. *Science*. 19, 1168-71.