

# CHAPTER 1

---

## NON-CODING RNAs

---

Alexander Donath<sup>a</sup>, Sven Findeiß<sup>a</sup>, Jana Hertel<sup>a</sup>, Manja Marz<sup>a</sup>, Wolfgang Otto<sup>a</sup>, Christine Schulz<sup>c</sup>, Peter F Stadler<sup>a,b,c,d</sup>, and Stefan Wirth<sup>a</sup>

<sup>a</sup>Bioinformatics Group, Department of Computer Science; and Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstrasse 16-18, D-01407, Leipzig, Germany

<sup>b</sup>Institute for Theoretical Chemistry, University of Vienna, Währingerstraße 17, A-1090 Wien, Austria

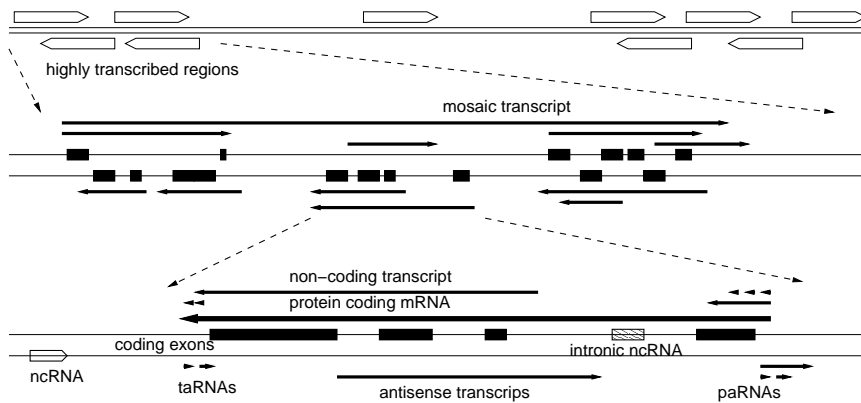
<sup>c</sup>Fraunhofer Institute for Cell Therapy und Immunology, Perlickstrasse 1, D-04103 Leipzig, Germany

<sup>d</sup>Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA

### 1.1 INTRODUCTION

The advent of high-throughput techniques that allow comprehensive unbiased studies of transcription has led to a dramatic change in our understanding of genome organization. A decade ago, the genome was seen as a linear arrangement of separated individual genes which are predominantly protein-coding, with a small set of ancient non-coding “house-keeping” RNAs such as tRNA and rRNA dating all the way back to an RNA-World. However, in contrast to this simple views more recent studies reveal a much more complex genomic picture. The ENCODE Pilot Project [239], the mouse cDNA project FANTOM [151], and a series of other large scale transcriptome studies, e.g. [202], leave no doubt that the mammalian transcriptome is characterized by a complex mosaic of overlapping, bi-directional transcripts and a plethora of non-protein coding transcripts arising from the same locus, Fig. 1.1.

This newly discovered complexity is not unique to mammals. Similar high-throughput studies in invertebrate animals [152, 93] and plants [135] demonstrate the generality of the mammalian genome organization among higher eukaryotes. Even the yeasts *Saccharomyces cerevisiae* and *Saccharomyces pombe*, whose genomes have been considered to be well understood, are surprising us with a much richer repertoire of transcripts than previ-



**Figure 1.1** Sketch of the post-ENCODE view of a mammalian transcriptome (adapted from [113]). Highly transcribed regions consist of a complex mosaic of overlapping transcripts (arrows) in both reading-directions. These transcripts link together the locations of several protein coding genes (coding exon indicated by black rectangles). Conversely, multiple transcription products, many of which are non-coding, are processed from the same locus as a protein coding mRNA.

ously thought [92, 49, 166, 266]. Even in bacteria, an unexpected complexity of regulatory RNAs was discovered in recent years [80].

Given the importance and ubiquity of non-coding RNAs (ncRNAs) and RNA-based mechanisms in all extant lifeforms, it is surprising that we still know relatively little about the evolutionary history of most RNA classes, although a series of systematic studies have greatly improved our understanding since our first attempt at a comprehensive review of this topic [21].

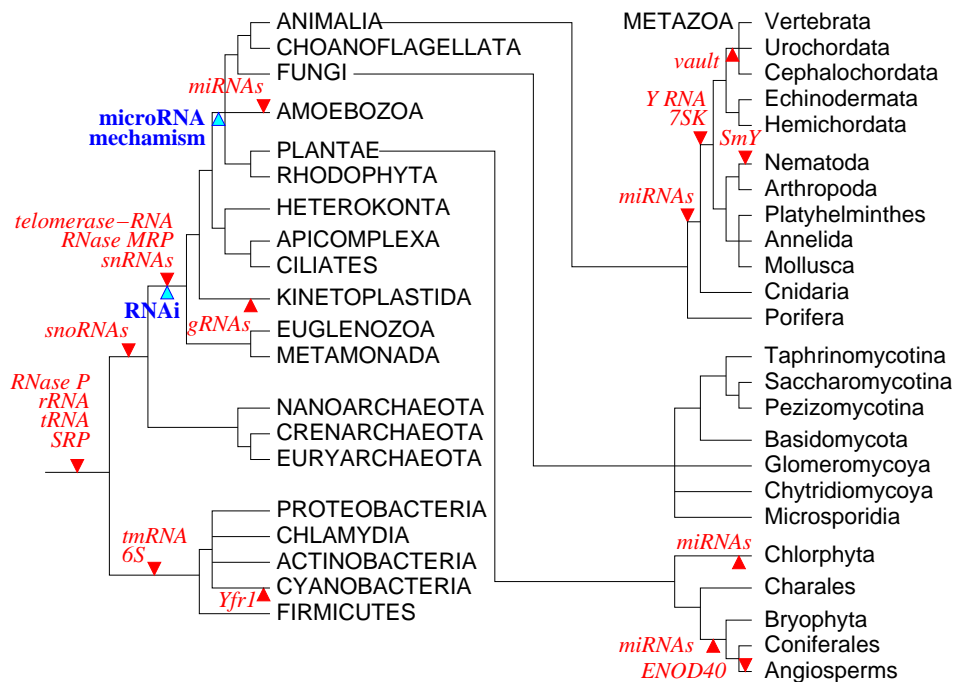
There are strong reasons to conclude that the Last Universal Common Ancestor (LUCA) was preceded by simpler life forms that were based primarily on RNA. In this RNA-World scenario [77, 76], the translation of RNA into proteins, and the usage of DNA [72] as an information storage device are later innovations. The wide range of catalytic activities that can be realized by relatively small ribozymes [218, 227] as well as the usage of RNA catalysis at crucial points of the information metabolism of modern cells provides further support for the RNA-World hypothesis. Multiple ancient ncRNAs are involved in translation: the

**Table 1.1** Selected experimental surveys for ncRNAs. We only list, without claim to completeness, extensive studies for RNAs larger than those associated with the RNA machinery and smaller than typical mRNAs.

Organism	ncRNAs		Ref.	Organism	ncRNAs		Ref.
	(a)	(b)			(a)	(b)	
Caenorhabditis elegans	161	31	[55]	Dictyostelium discoideum	20	16	[8]
Aspergillus fumigatus	30	15	[110]	Giardia intestinalis	30	26	[40]
Plasmodium falciparum	41	6	[34]	Caulobacter crescentus	3	27	[129]
Sulfolobus solfataricus	22	23	[279]	Sulfolobus solfataricus	31	33	[235]

(a) ncRNAs for which at least membership in known class such as snoRNAs could be established

(b) ncRNAs without annotation

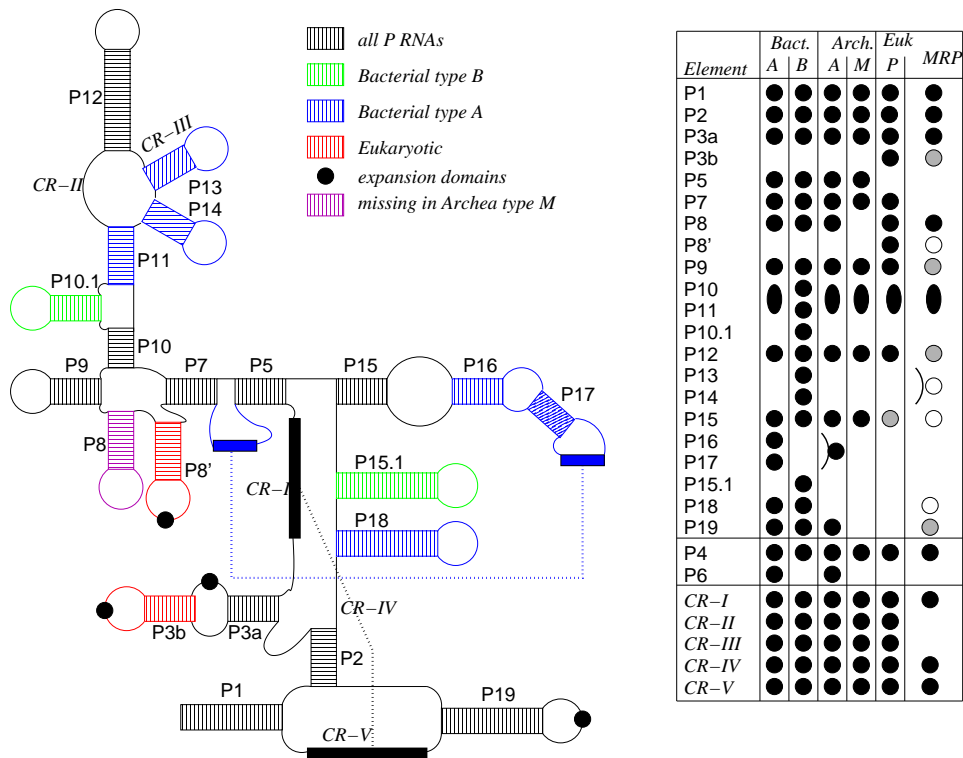


**Figure 1.2** Origins of major ncRNA families. The origin of ncRNA families is marked leading to the last common ancestor of the known representative. For details on RNAi and microRNAs we refer to [→chapter \\*\\*\\*](#). The microRNA families of (a) eumetazoa (animals except sponges), (b) the slime mold *Dictyostelium*, (c) embryophyta (land plants), and (d) the green algae *Chlamydomonas* are non-homologous. In addition, the putative origin of the RNAi mechanism and the microRNA pathways is indicated.

ribosome itself is an RNA machine [168], tRNAs perform a major part of the decoding on the messenger RNAs, and RNase P, another ribozyme, is involved in processing of primary tRNA transcripts. The signal recognition particle, another ribonucleoparticle (RNP), also interacts with the ribosome and organizes the transport of secretory proteins to their target locations. For a discussion of rRNAs and tRNAs we refer to [→chapter \\*\\*\\*](#).

On the other hand, most functional ncRNAs do not date back to the LUCA but are the result of later innovations. Some crucial “housekeeping” functions involve domain-specific ncRNAs. Eukaryotes, for instance have invented the splicing machinery involving several small spliceosomal RNAs (snRNAs), while eubacteria use tmRNA to free stalled ribosomes and the 6S RNA as a common transcriptional regulator. The invention of the RNAi machinery in eukaryotes and the subsequent evolution of microRNAs in plants and animals is discussed in detail in [→chapter \\*\\*\\*](#).

The innovation of ncRNAs is an ongoing process. In fact, most experimental surveys for ncRNAs, Table 1.1, report lineage-specific elements without detectable homologies in other species. An overview of evolutionary older ncRNA families is compiled in Fig. 1.2 without claim to completeness. Many RNA classes, however, such as Y RNAs and vault RNAs, and most eubacterial ncRNA families have not been studied in sufficient detail to date their origin with certainty. Some of them thus might have originated earlier than shown.



**Figure 1.3** Schematic drawing of the consensus structures of P and MRP RNAs. Adapted from [155, 254, 272, 283]. The table indicates the distribution of structural elements. Black circles indicate conserved elements, stems indicated in gray are present in known sequences, open circles refer to elements that are sometimes present.

## 1.2 ANCIENT RNAs

### 1.2.1 RNase P and RNase MRP RNA

RNases P and MRP are ribonucleoprotein complexes that act as endoribonucleases in tRNA and rRNA processing, respectively. Their RNA subunits are evolutionarily related and are involved in the catalytic activity of the enzymes. While it has long been known that RNase P RNA is a ribozyme in bacteria and several archaea, it was demonstrated only recently that eukaryotic RNase P RNA also exhibits ribozyme activity [270]. The main function of RNase P is the generation of the mature 5' ends of tRNAs. See [254] for a recent review of RNase P. In contrast, RNase MRP is eukaryote-specific. It processes nuclear precursor rRNA (cleaving the  $A_3$  site and leading to the maturation of the 5' end of 5.8S rRNA), generates RNA primers for mitochondrial DNA replication, and is involved in the degradation of certain mRNAs.

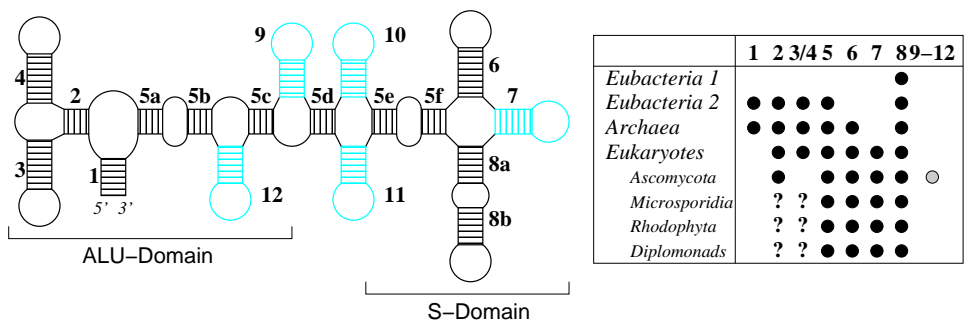
The phylogenetic distribution of P RNA clearly indicates that it dates back to the Last Universal Common Ancestor [193]. MRP RNA can be traced to the most basal eukaryotes [193] and apparently was part of the rRNA processing cascade of the Eukaryotic Ancestor [272]. The high similarity of P and MRP RNA secondary structures [45] and similarity of

the protein contents and interactions of RNase P and MRP [9, 254] suggest that P and MRP RNAs are paralogs.

RNase P RNA is found almost ubiquitously. Interestingly, so far only MRP RNA has been found in plants including green algae, and red algae [193]. Whether the ancestral P RNA has been lost in these clades or possibly replaced by MRP RNA is unclear. It is also possible that the P RNA sequences are derived from each other that they have escaped detection so far. Despite the highly conserved core structures, P and MRP RNAs can exhibit dramatic variations in size, which mostly arise from large insertions in several “expansion domains” [193, 112]. In eukaryotes, additional P RNAs are often encoded in organelle genomes. Chloroplast P RNA is structurally similar to bacterial type A [52] and exhibits ribozyme activity [135]. Mitochondrial P RNAs, in particular those of fungi, are highly derived and exhibit only a small subset of the conserved structural elements shown in Fig. 1.3, mostly P1, P4, and P18 [217]. Despite its core function in tRNA processing, RNase P appears to be absent in the archaeon *Nanoarchaeum equitans*. Instead, placement of its tRNA gene promoters allows the synthesis of leaderless tRNAs [201].

### 1.2.2 Signal Recognition Particle RNA

The signal recognition particle (SRP) is a ribonucleoprotein that interacts with the ribosome during the synthesis and translocation of secretory proteins. SRP is responsible for the co-translational targeting of proteins that contain signal peptides to membranes including the prokaryotic plasma membrane and the endoplasmic reticulum. SRP recognizes first the signal sequence of the nascent polypeptide and then the SRP receptor in the membrane.



**Figure 1.4** Standard nomenclature of the helices of SRP RNA and their phylogenetic distribution. A subgroup of Eubacteria which includes proteobacteria have a drastically reduced SRP RNA. Adapted from [284, 5].

The assembled SRP consists of two structurally and functionally distinct domains. The small domain, which consists of the Alu-domain of the SRP RNA and the proteins SRP9 and SRP14, modulates the elongation of the secretory protein. The larger S domain, which consists of the S-domain of the SRP RNA and several specific proteins, captures the signal peptide. Eukaryotic SRP RNA is also known as 7SL RNA.

SRP RNAs are highly conserved across large phylogenetic distances and fit a generalized structure, sketched in Fig. 1.4, that can be used to define a standard nomenclature for the individual structural features [284]. Eukaryotes, archaea, firmicutes and several other early-branching bacteria such as *Thermatoga maritima* have large SRPs containing both domains. In contrast, most Gram-negative bacteria as well as all known organellar SRPs

have a reduced SRP RNA consisting of helix 8 only [5]. Since firmicutes are among the most early branching bacterial phyla, the SRP RNA have mostly likely lost the Alu-domain in higher eubacteria.

It is interesting to note that the SRP of trypanosomatids contains a second, tRNA-like, RNA component [142].

### 1.2.3 snoRNAs

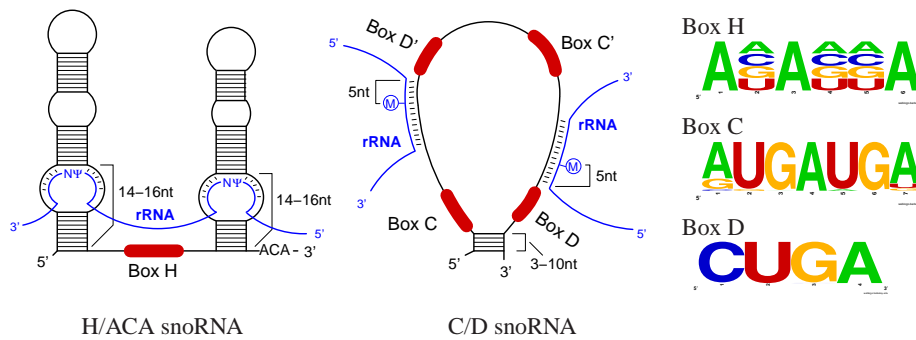
Small nucleolar RNAs (snoRNAs) represent one of the most abundant classes of ncRNAs. They act as guides for single nucleotide modification in nascent ribosomal RNAs and other RNAs in the nucleolus of eukaryotic species [54]. While no targets are present or known for so-called *orphan snoRNAs*, most guide snoRNAs target ribosomal RNAs (rRNAs) or small nuclear RNAs (snRNAs). Recently, there has been a report that several snoRNAs may also target tRNAs and other snoRNAs [282]. The orphan snoRNAs have been implicated in modulating alternative splicing [119, 17].

Based on conserved sequence elements and secondary structure, snoRNAs are divided into two major classes designated as C/D and H/ACA snoRNAs. Both differ in their characteristic secondary structure and in the sequence motifs that gave them their names: C/D snoRNAs exhibit the “boxes” C and D with consensus sequence (A)UGAUGA and CUGA respectively. They frequently contain a second copy of these two boxes, usually designated C' and D', in the region between the C and D box. The H box of the H/ACA snoRNAs has the consensus sequence ANANNA. The ACA motif located at the 3' end of the molecule is highly conserved. Believed to be protein-binding elements, the sequence boxes are located in unpaired regions, Fig. 1.5.

The characteristic secondary structures in Fig. 1.5 of snoRNAs are inferred mostly from phylogenetic comparisons. Thermodynamic predictions using e.g. *mfold* or the Vienna RNA Package may differ considerably from the functional structure, indicating that the functional RNA structure within the small nucleolar ribonucleoprotein (snoRNP) particles is changed by the protein components, which bring the binding sites in the correct formation and thus help to bind the target [75, 213]. The poor predictability of their secondary structures is a major obstacle for computational snoRNA gene finding, see [95] and the references therein.

The two major classes direct distinct chemical modifications. Most C/D snoRNAs determine specific target nucleotides for methylation of the 2'-hydroxyl position of the sugar, while most of the H/ACA snoRNAs specify uridines that are converted into pseudouridine by rotation of the base. These modifications apparently influence the rRNA structure and are functionally essential [128, 54]. A few exceptional snoRNAs of both classes are also involved in pre-rRNA processing but do not lead to chemical modifications. Instead, they mediate structural changes of the rRNA to establish the correct conformation endonuclease cleavage [10]. In human and yeast these are U3, U17 (also known as sn30), U22, and U14, as well as vertebrate U8 and yeast snR10.

A subgroup of snoRNAs (only U3, U8 and U13 in human with several additional families in nematodes) share features of snRNAs, such as a post-transcriptionally hypermethylated 2,2,7-trimethylguanosine (TMG) cap at their 5' end [109]. Another small subgroup of both C/D and H/ACA snoRNAs carries an additional localization signal for the the Cajal body [108]. These small Cajal body-specific RNAs (scaRNAs) target snRNAs [48] and are retained in the Cajal bodies (or yeast nuclear bodies, respectively). These sequences may be chimeric composites containing both a C/D and an H/ACA domain for recruitment of both classes of snoRNPs; an example is U85 [61]. There are a number of scaRNAs known



**Figure 1.5** Schematic representation of H/ACA and C/D snoRNA structures and their characteristic consensus sequence motifs (boxes). H/ACA snoRNAs (left) fold into a double stem loop structure with the Box H between both stems and the ACA motif immediately following after the 2<sup>nd</sup> stem. The target is bound to both stem-loop-structures inside their symmetrical interior loops. Neither the *U* that is modified nor the following nucleotide are bound to the snoRNA: instead, they are centered within the interior loop.

The functional structure of C/D snoRNAs is probably stabilized by proteins bound to the boxes, since there is only a short (4-10nt) stem that encloses box C and D while the region between the two boxes is mainly unpaired. The target RNA is bound upstream of box D so that the 5th nucleotide is methylated. Many C/D snoRNAs have a second copy of both boxes between box C and D (namely C' and D'), with a second functional binding site upstream of box D'.

that modify RNA pol II-specific U1, U2, U4 and U5 snRNAs. For example, the human U85 scaRNA directs 2'-*O*-methylation of the C45 and pseudouridylation of the U46 residues in the U5 spliceosomal snRNA [48, 61].

The snoRNA-based RNA modification system is found in both Eukaryotes and Archaea [236, 237, 204]. However, little is known about the evolution of individual snoRNA families. Even the characteristic sequence motifs of snoRNAs are modified in several basal eukaryotic lineages. In *Leishmania major*, the ACA box is systematically replaced by an AGA motif [138], while other snoRNA molecules in the primitive eukaryote *Gardia lamblia* show mutations in the D box of C/D-like snoRNAs [276]. Due to their poor sequence conservation it is a non-trivial problem to establish the homology of snoRNAs over large evolutionary distances, e.g. between a mammalian and a yeast sequence. A plausible assumption is that homologous snoRNAs target homologous target positions. Based on this assumption C/D snoRNAs in the *Drosophila melanogaster* genome were successfully identified [1]. Additionally, a table of human/yeast correspondences can be found in the snoRNABase [133]. On much shorter timescales, the investigation of snoRNA evolution in nematodes [282] and in mammals [215] has shown that snoRNAs, like many other ncRNAs, evolve by duplication events and by coevolution with their targets. As with duplicated proteins [71], three different fates for duplicated snoRNAs were observed: (i) inactivation and eventual loss of one of the paralogs, (ii) maintenance of function and the same target in both paralogs, and (iii) divergence of the target site of one of the paralogs. This may provide a mechanism by which the extant diversity of snoRNAs may have evolved from a single ancestral C/D and H/ACA snoRNA.

SnoRNAs are found in two different genomic contexts: within introns of protein coding genes and in independent transcription units. In both cases, the snoRNA is initially processed as pre-snoRNA that requires further maturation [263]. Almost all snoRNAs in



vertebrates are localized within introns of “house-keeping” genes, in particular of genes for ribosomal proteins. However, in several cases it is also known that the host-gene is non-coding, see e.g. [14]. Outside the Metazoa, gene organization is more diverse. In yeast, most snoRNAs are processed from independent mono-, di-, or polycistronic RNA transcripts [251]. Such polycistronic clusters of multiple different snoRNA genes are also common in higher plants [26] and Kinetoplastida [139]. The biosynthesis of box C/D snoRNAs seems to be related to the position of the gene within the intron and optimal distances between snoRNA coding sequence and conserved intron elements (e.g., 3'-end and branch point) in mammals and yeast [97, 251]. In contrast, it has been demonstrated that H/ACA snoRNAs originate in a splicing dependent manner but they do not possess any preferential localization with respect to intronic splice sites [206].

Duplicated snoRNA paralogs are often inserted into different positions in the same gene (*cis*-duplication). However, also duplicated snoRNAs in distant genes or even other chromosomes (*trans*-duplication) [282, 215] are reported. Finally, snoRNA paralogs may be moved to different chromosomal locations through a duplication of the host genes. Consistent with these mechanisms, duplications and “jumps” of the snoRNA gene from an intron in the ancestral host-gene to an intron of another gene have been observed in vertebrates [21]. Intergenic copies may also remain functional: The polymerase-III transcribed C/D snoRNA in yeast (snRN52) [88] may have arose from a paralog that by chance came under the control of a polymerase-III promoter after duplication.

### 1.3 DOMAIN-SPECIFIC RNAs

#### 1.3.1 Telomerase RNA

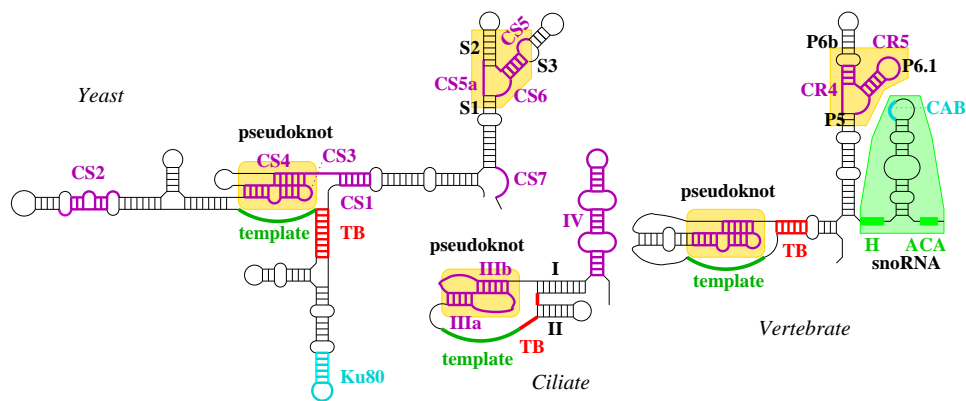
In contrast to the circular genomes of prokaryotes, eukaryotes have linear chromosomes. Special mechanisms are necessary to replicate the chromosome ends, the telomers. In almost all species investigated to-date, a telomerase enzyme maintains telomere length by adding G-rich telomeric repeats to the ends eukaryotic chromosomes. Telomerase thus dates back to the origin of eukaryotes. Notable exceptions are diptera including *Anopheles* and *Drosophila*, which use retrotransposons or unequal recombination instead of a telomerase enzyme.

The core telomerase enzyme consists of two components: an essential RNA component, which serves as template for the repeat sequence, and the catalytic protein component telomerase reverse transcriptase (TERT). The RNA component varies dramatically in sequence composition and size. Although dozens of telomerase RNAs (usually called TER in vertebrates and TLC-1 in yeasts) have been cloned and sequenced, the known examples are restricted to four narrow phylogenetic groups: vertebrates, yeasts, ciliates, and plasmodia. The protein component TERT on the other hand is known in a much wider range of eukaryotes [196].

Yeast telomerase RNAs appear to be even less well conserved: In [245], only seven short sequence motifs are reported within more than 1.2kb transcripts of *Kluyveromyces* species, and of these only a few are partially conserved in *Saccharomyces*. In fact, *Saccharomyces* and *Kluyveromyces* TLC-1 genes cannot be aligned with each other by standard alignment programs. The same is true for the recently discovered TLC gene of *Schizosaccharomyces pombe* [132, 262].

The small ciliate TER genes include a pseudoknot domain that contains an unusual triple-helical segment with an AUU base triplet. This domain is also shared by the ver-





**Figure 1.6** Telomerase RNA structures of yeast and human share the topology of the pseudoknot region and a functionally important junction region. The template and its boundary element (TB) are highlighted. The yeast structure is a consensus of *Saccharomyces* [47, 281] and *Kluyveromyces* structures [27]. The Ku80 binding domain is specific for *Saccharomyces*. Vertebrate telomerases have a snoRNA domain [165] at their 3' end. This domain carries a Cajal-body localization signal (CAB) [108], which is present in all vertebrates except teleosts [274]. Black regions may vary dramatically in length.

tebrate and yeast telomerase RNAs [246]. Whether such a structure is also present in the computationally predicted TER genes of plasmodia [34] is not yet known.

Although there is a common core structure of all these telomerase RNAs [37], and despite their length of several hundred to almost 2000nt, these RNAs remain a worst case scenario for homology search. Indeed, a survey of vertebrate telomerase RNAs [39] shows dramatic sequence variation with only a few, short, well-conserved sequence patterns separated by regions of highly variable length. The recent discovery of the TER genes of teleost fishes [274] highlights the variability of this molecule, which has acquired several lineage-specific domains, such as the snoRNA domain in vertebrates and the Ku-80 binding domain in budding yeast, see Figure 1.6.

### 1.3.2 Spliceosomal snRNAs

In eukaryotes, introns of protein-coding mRNA and mRNA-like ncRNAs are spliced out of the primary transcript by the spliceosome, a large ribonucleoprotein (RNP) complex which consists of up to 200 proteins and five small non-coding RNAs [181]. Mounting evidence suggests that these snRNAs exert crucial catalytic functions in the splicing process [248]. Spliceosomal splicing is one of four distinct mechanisms, see Tab. 1.2 for details.

The spliceosomal machinery itself may be present in three distinct variants in eukaryotic cells. The dominant form is the *major spliceosome* which contains the snRNAs U1, U2, U4, U5 and U6 and removes introns delimited by the canonical donor-acceptor pair GT-AT (as well as some AT-AC and GC-AG introns). A recent report on the expression of a U5 snRNA candidate in *Giardia* [40], a protozoan with few introns, suggests that the spliceosome and its snRNA date back to the Eukaryote ancestor. In general, snRNAs are subject to concerted evolution if they are present in multiple copies. Nevertheless, there is evidence for differential regulation of paralogous snRNA genes in several lineages [57, 38, 156].

**Table 1.2** Splicing Mechanisms. Three major mechanisms, (A), (B), and (C) can be distinguished [146]. Group I [91] and group II [69] (which include the group III introns) are self-splicing. However, Group II introns also share several characteristic traits, including the lariat intermediate, with spliceosomal introns and might share a common origin. The splicing of eukaryotic tRNAs and all archaeal introns uses specific splicing endonucleases, reviewed in [30]. The spliceosomal machinery does not distinguish between protein coding mRNAs and mRNA-like ncRNAs.

Domain	(A) group I	(B) group II	(B) spliceosomal	(C) endonuclease
Bacteria	+	+	–	–
Archaea	–	–	–	tRNA, rRNA, mRNA
Eukaryota	+	+	“mRNA”	tRNA

About 1 in 10 000 protein coding genes is spliced by the *minor spliceosome* [188] which is composed of the snRNAs U11, U12, U4atac, U5 and U6atac and acts on AT-AC (rarely GT-AG) [221] introns. The snRNAs U11, U12, U4atac, and U6atac take on the roles of U1, U2, U4, and U6. Whereas, both U6 and U6atac are polymerase-III transcripts, all other spliceosomal snRNAs are transcribed by polymerase-II. Interestingly, the minor spliceosome can also act outside the nucleus and has a function in the control of cell proliferation [123]. Functional and structural differences between the two types of spliceosomes are reviewed in [267]. The snRNAs themselves are not only part of the spliceosomes but are also involved in transcriptional regulation [127].

The third type of splicing is *spliced-leader-trans-splicing*. Here a “miniexon” derived from the non-coding spliced-leader RNA (SL RNA) is attached to the 5’ end of each protein-coding exon [185, 198, 90]. The corresponding spliceosomal complex contains the snRNAs U2, U4, U5, and U6, as well as an SL RNA [90].

The minor spliceosome is present in most eukaryotic lineages and traces back to an origin early in eukaryotic evolution [44, 143, 211]. Although it appears to have been lost in many lineages, most metazoa have a minor spliceosome, with the notable exception of nematodes such as *Caenorhabditis elegans* [188] and certain cnidaria [50, 156]. Within fungi, minor spliceosomes have been reported only for zygomycota and some chytridiomycota. Minor spliceosomes are also reported in oomycetes (Heterokonta) and streptophyta [50]. Whereas, Euglenozoa and Alveolata do not seem to have minor spliceosomes.

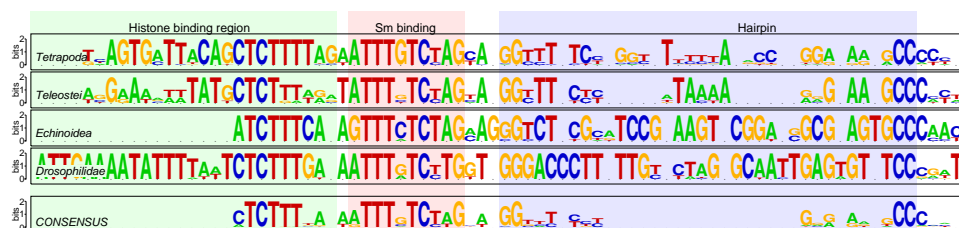
The evolutionary origin of SL-trans-splicing is unclear. It has been described in tunicates, nematodes, platyhelminthes, cnidarians, kinetoplastids [90], rotifera [198] and dinoflagellates [140]. In contrast SL RNAs are absent in vertebrates, insects, plants, and yeasts [198]. Due to the rapid evolution and the small size of SL RNAs it is hard to determine whether examples from different phyla are true homologs or not. Thus, two competing hypotheses are discussed in the literature: (i) ancient trans-splicing and SL RNAs have been lost in multiple lineages and (ii) the mechanism has evolved independently as a variant of spliceosomal *cis*-splicing in multiple lineages.

In nematodes polycistronic pre-mRNAs are trans-spliced into two or even more [191] distinct SL RNAs which provide the 5’ acceptor site for the first (SL1) and all subsequent (SL2) mRNA sequences. This leads to the formation of discrete monocistronic mRNAs that start with either the SL1 or the SL2 sequences [19]. Trans-splicing to the SL1 acceptor

requires no specific signal. In contrast, to *cis*-splicing or SL1-trans-splicing the attachment of SL2 appears to be linked with polyadenylation of the preceding mRNA in the polycistron [194].

### 1.3.3 U7 snRNA

Replication-dependent histone genes are the only known eukaryotic protein-coding mRNAs that are not polyadenylated, ending instead in a conserved stem-loop sequence, see [158] for a recent review. The processing of the 3' end of these histone genes is performed by the U7 snRNP. The U7 snRNA is the smallest RNA polymerase-II transcript known to-date, with a length ranging from 57nt (sea urchin) to 70nt (fruit-flies). Its expression level of only a few hundred copies per cell in mammals is at least three orders of magnitude smaller than the abundance of other snRNAs.



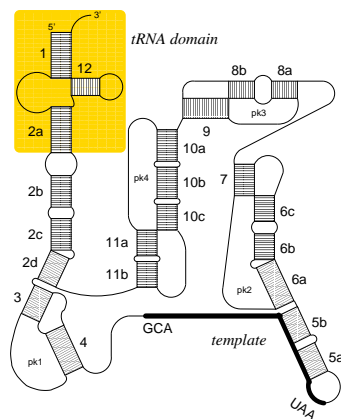
**Figure 1.7** Aligned sequence logos of the consensus sequences of U7 snRNAs from tetrapods, teleosts, sea urchins, and flies. Adapted from [157].

The U7 snRNP-dependent mode of histone end processing has long been believed to be a metazoan innovation [13, 158]. Anyway, the phylogenetic distribution of the U7 snRNP proteins, provide evidence for an origin of this mechanism early in Eukaryote evolution [51]. Nevertheless, U7 snRNAs so far have been reported only for Deuterostomes and Drosophilids [157, 51]. A detailed analysis shows that each of its three major domains, the histone binding region, the Sm-binding sequence, and the 3' stem-loop structure, exhibits substantial variation in both sequence and structural details, see Figure 1.7.

### 1.3.4 tmRNA

The transfer-messenger RNA (tmRNA), also known as 10Sa RNA or SsrA, is part of a complex that acts as a unique translation quality control and ribosome rescue system in all eubacteria and some eukaryotic organelles [6, 83]. “Nonstop” mRNAs, which due to processing errors lack appropriate termination signals, cannot promote release factor binding and thus lead to stalled ribosomes attached to the nascent incomplete polypeptide. tmRNA rescues these ribosomes and provides the template for a peptide-tag that then causes the rapid degradation of the incomplete translation product. In a typical *E. coli* bacterium, there are approximately 700 tmRNAs per cell; that is, one tmRNA for every 10-20 ribosomes [169].

For recent detailed reviews of tmRNA structure and function we refer to [170, 58]. In brief, tmRNA combines the functions of a tRNA and an mRNA, a dichotomy that is also reflected by its structure, Fig. 1.8. Its tRNA-like domain binds to the protein SmpB and is then aminoacetylated by alanyl-tRNA synthetase. A quaternary complex that in addition



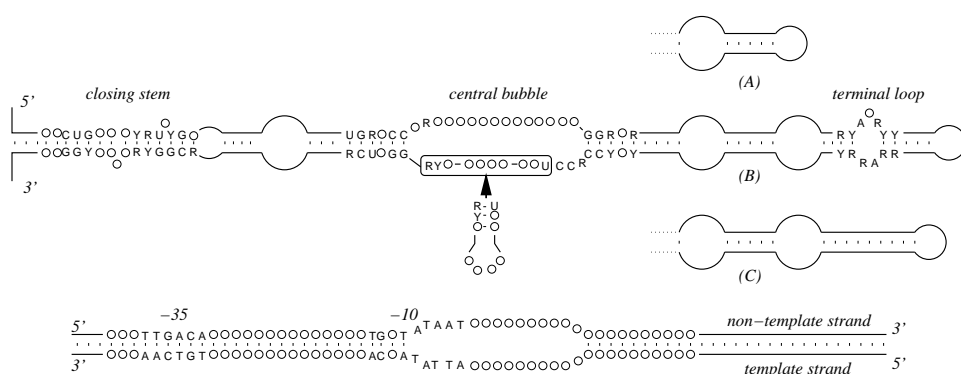
**Figure 1.8** Canonical tmRNA structure. The 5' and 3' ends form a tRNA-like domain (shaded) which includes an acceptor stem, a D-loop without stem, and a T-arm [122]. Instead of the anticodon stem-loop that is typical for a normal tRNA, however, tmRNA has a long stem that connects the tRNA-domains with the rest of the molecule. The mRNA-like domain contains the template ORF and is terminated by a hairpin with the stop codon in its loop. The mRNA-like and tRNA-like domains are connected by a series of four pseudoknots of unknown function. (Adapted from [5])

contains elongation factor EF-Tu recognizes ribosomes stalled at the 3' end of a nonstop mRNA and like tRNA enters the ribosome's A site. The nascent polypeptide is transferred to the alanyl-tmRNA, which now switches to its mRNA-like mode of action by translocating to the P site of the ribosome where it places a TAG codon in the ribosome's mRNA channel. This leads to the release of the defective mRNA and its subsequent selective degradation by RNase R. The ribosome continues translation with the tmRNA ORF as a surrogate template and terminates at a tmRNA-encoded stop codon, thereby releasing the nascent protein with the 11-amino acid degradation tag, which contains epitopes for ubiquitin proteases.

Genes for tmRNA and the accompanying SmpB protein so far have been found in each of the completely sequenced eubacterial genomes, see e.g. [83], although tmRNA is not crucial for survival in some bacterial species, see e.g. [182, 22]. In *E. coli*, tmRNA is transcribed as a 457nt precursor and then cleaved to obtain the 363nt mature molecule. In some species, circularly permuted genes produce split tmRNAs, which still share the ancestral domain organization [219]. There are tmRNA genes in some eubacterial-like organelles, in particular in the chloroplasts of diatoms (e.g., *Thalassiosira pseudonana*) and red algae (e.g., *Cyanophora paradoxa*) [78, 6]. Reduced tmRNAs, which lack the mRNA-like region, were identified in the mitochondrial genomes of a few protozoans including the jakobid *Reclinomonas americana* [106]. So far neither an archaeal nor a nuclear Eukaryotic tmRNA have been found, although there is a nuclear-encoded SmpB gene with an organelle import signal in heterokonta with organellar tmRNAs [107].

### 1.3.5 6S RNA

The 6S RNA of *E. coli* is one of the first known bacterial small RNAs. Nevertheless, its function in transcriptional regulation was only recently elucidated, reviewed in [269, 259]. It binds specifically to the bacterial RNA-Polymerase (RNAP) holoenzyme and selectively inhibits  $\sigma^{70}$ -dependent transcription. During the exponential growth almost all housekeeping genes are  $\sigma^{70}$  dependent regulated [150]. 6S RNA seems to mimic the open  $\sigma^{70}$ -dependent promoter complex [33]. This hypothesis is supported by the fact that mutation within the "central bubble" of the 6S RNA abrogated  $\sigma^{70}$ -holoenzyme binding. Therefore, the conserved secondary structure of the RNA molecule is critical for its function. Since the 6S RNA concentration increases 10-fold from the exponential to the stationary phase [261] it stops transcription of the "housekeeping" genes and stores the RNAP-holoenzyme. This RNAP-6S complex is present during late stationary phase, when NTP concentrations are



**Figure 1.9** 6S RNA structure. The central bubble is delimited by stable, GC-rich stems in all known 6S RNAs. It consists of three domains, known as the 'closing stem', 'central bubble' and 'terminal loop'. Three major variants of the terminal loop have been identified (A)  $\alpha$ -proteobacteria, (B)  $\gamma$ -proteobacteria, cyanobacteria, and firmicutes, (C)  $\beta$ -proteobacteria,  $\delta$ -proteobacteria, and spirochetes. For comparison, the open promoter complex is shown. Adapted from [16].

low. As soon as new NTPs become available to the RNAP-6S complex, the 6S RNA serves as a template for the transcription of the 14-20nt pRNAs. This results in the formation of an unstable 6S RNA-pRNA complex and the release of the RNAP-holoenzyme [260].

The secondary structure of 6S RNA is discussed in detail in [16]. It consists of three domains (Fig. 1.9) and shows close similarities with an open  $\sigma^{70}$  promoter.

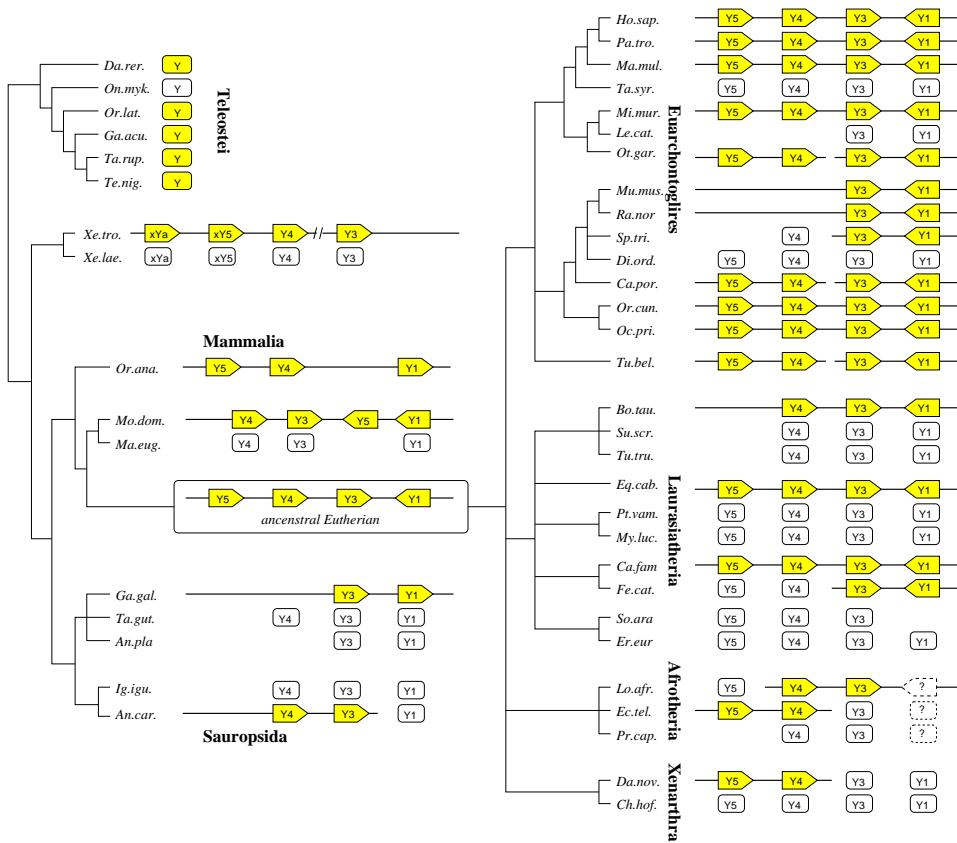
While 6S is a single-copy gene in most bacterial phyla, there are two differentially regulated copies in Gram-positive bacteria [16]. Interestingly, two alternative structural conformations have been reported for cyanobacterial 6S RNAs [11]. In *E. coli*, the *ssrS1* gene forms a bi-cistronic message combining the 6S RNA (5') and the open reading frame of *ygfA* (3'). A tandem promoter is responsible for regulating 6S RNA expression, producing either a short,  $\sigma^{70}$  dependent transcript or a longer version that uses either  $\sigma^{70}$  or  $\sigma^S$ . Different RNases process these transcripts [118]. Additionally, the repression by several regulatory proteins is reported [178]. A similar gene structure was observed in *Prochlorococcus*, although the short transcript does not seem to be processed in this case [11].

The broad phylogenetic distribution of 6S RNA across all bacterial phyla emphasizes the universality of its function. Nevertheless, 6S RNA-null cells show only a minor phenotype effect in bacteria such as *E. coli* and *Synechococcus* [243].

## 1.4 CONSERVED ncRNAs WITH LIMITED DISTRIBUTION

### 1.4.1 Y RNAs

Y RNAs were discovered as the RNA component of the Ro RNP particle. They form a small family of short RNA polymerase-III transcripts with characteristic secondary structure [68, 238]. The function of Y RNAs remains elusive. A direct role of Y RNAs for DNA replication was demonstrated in [41], and Y5 was recently implicated in 5S rRNA quality control [99].



**Figure 1.10** Evolution of the vertebrate Y RNA locus. With the exception of *Xenopus*, the functional Y RNA genes are located in a single cluster in all sufficiently assembled genomes (symbols with arrows on a line marking an uninterrupted piece of genomic DNA). For most species, only short scaffolds or shotgun traces are available (white symbols without direction). Updated from [172].

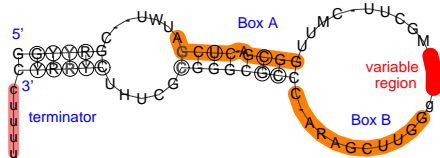
Y RNAs have so far been reported only in vertebrates, the nematode *Caenorhabditis elegans* [249] and the prokaryote *Deinococcus radiodurans* [39]. At this point it remains unclear whether the latter is homologous to the animal Y RNAs and if so, whether Y RNAs date back to the LUCA.

The amniota exhibit a single cluster of Y RNAs whose evolution has been traced in detail [172, 190]. Interestingly, the orientation of the Y1 gene is inverted in eutheria. Loss of members of the Y RNA family seems to be a fairly frequent phenomenon in several families. In both rat and mouse, there is no trace of either Y5 or Y4, while in the closely related squirrel genome Y4 is still present, see Figure 1.10.

In primates, Y RNAs are the founders of a family of about 1000 pseudogenes that constitute a class of L1-dependent non-autonomous retroelements [189, 190]. In contrast, in almost all other species (with the notable exception of the guinea pig *Cavia procellus*) there are only a few Y-RNA derived pseudogenes.

### 1.4.2 Vault RNAs

Vaults are large ribonucleoprotein particles that are ubiquitous in eukaryotic cells with characteristic barrel-like shape and a still poorly understood function in multi-drug resistance [250]. Vault RNAs are short (80-150nt) polymerase-III transcripts comprising about 5% of the mammalian vault complex.

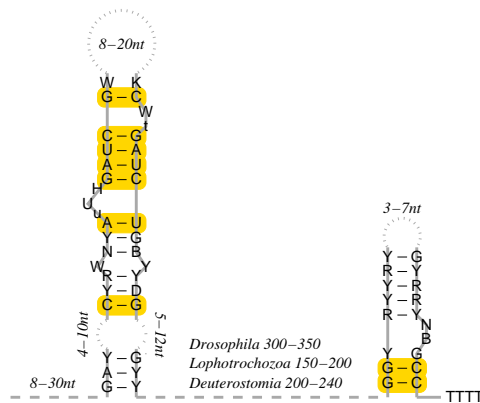


**Figure 1.11** Vault RNAs contain two internal polymerase-III promoter sequences, Box A and Box B, and a typical terminator sequence at the 3'-end. The terminal stem is conserved among all known examples. Circles indicate positions with compensatory mutations. The variable regions have a length of 20 to more than 100 nucleotides. Adapted from [171].

So far, vault RNAs have been studied experimentally in frogs and a few mammals. They exhibit little sequence conservation beyond their Box A and Box B internal promoter elements. Nevertheless, computational analyses have identified vault RNAs in most vertebrates [171]. Similar to Y RNAs, mammalian vault RNAs are organized in a small cluster typically comprising two or three (in primates) paralogs (unpublished data).

### 1.4.3 7SK RNA

The 7SK snRNA is one of the most highly abundant ncRNAs in vertebrate cells [258]. Due to its abundance it has been known since the 1960s. Its function as a transcriptional regulator, however, has only recently been discovered: 7SK mediates the inhibition of transcription elongation factor P-TEFb, a critical regulator of RNA polymerase-II transcription which stimulates the elongation phase, see [62, 125] and the references therein.



**Figure 1.12** Like telomerase RNA, 7SK RNA is highly variable in both sequence and structure. Only two stem-loop structures, towards the 5' and the 3' end, are conserved. The intervening sequence is highly variable in both length and structure. The 5'-terminal hairpin is responsible for HEXIM1 and P-TEFb binding [62], the 3'-stem also interacts with P-TEFb.

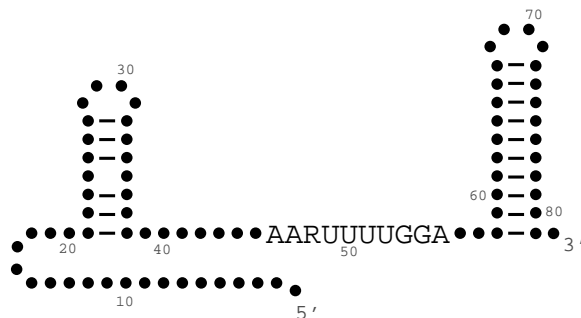
The polymerase-III transcript, with a length of about 330nt, is highly conserved in vertebrates [85]. In contrast to the nearly perfect sequence and structure conservation in jawed vertebrates [258, 62], however, the 7SK RNA from the lamprey homolog differs in more than 30% of its nucleotide positions from its mammalian counterpart. Recently, the molecule was regarded as a vertebrate innovation because searches for invertebrate homologs remained unsuccessful despite considerable efforts. A combination of computational and experi-



mental approaches, however, uncovered 7SK homologs in basal deuterostomes and several lophotrochozoans and revealed its evolutionary flexibility [82], Fig. 1.12. A genome-wide survey for well-conserved type-3 polymerase-III promoter structures in *Drosophila* finally lead to the discovery of arthropod 7SK RNAs [81]. The 7SK RNA hence was present already in the Protostome-Deuterostome Ancestor.

#### 1.4.4 Sm Y RNA

The analysis of the cis- and trans-spliceosomes in *Ascaris lumbricoides* [154] lead to the discovery of two snRNA-like small RNAs that exhibit a canonical Sm protein-binding site. In *Caenorhabditis elegans*, this novel class of RNAs contains at least 12 closely related genes [55, 93, 147].



**Figure 1.13** Consensus structure of the *Caenorhabditis elegans* SmY RNAs. The terminal hairpin and the consensus pattern AARUUUUGGA form a canonical Sm binding site.

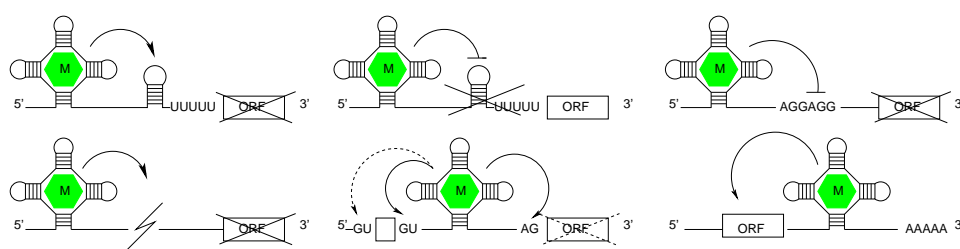
The SmY RNAs occur in a complex with the spliced-leader RNA (SL RNA) formed by direct base-pairing reminiscent of the interactions between spliceosomal RNAs. This strongly suggests a direct involvement in the trans-splicing machinery, either as a direct component of the trans-spliceosome that is required for function, or as a chaperone for the SL snRNP that prevents inappropriate splicing. Although all sequenced nematode genomes contain SmY homologs, no SmY RNAs were found outside the phylum Nematoda (unpublished data).

#### 1.4.5 Bacterial RNAs

A plethora of evolutionarily unrelated *small RNAs* (sRNAs) have been described across the eubacteria, most of which have a more or less limited phylogenetic distribution. It is at present nearly impossible to offer a comprehensive survey or even a reasonably complete list. A series of reviews including [4, 226, 252] cover much of this terrain. However, many novel and almost poorly understood ncRNAs continue to be discovered at a staggering rate.

The Sm-like protein *Hfq* plays a special role in small RNA based regulation of bacterial gene expression, reviewed in [24]. The RNA chaperone *Hfq* is present in half of all Gram-positive as well as Gram-negative bacteria and at least one archaeon, *Methanococcus Jannaschii*. Existence of an *hfq* gene, however, does not imply the presence of a functional *Hfq* protein or its involvement in sRNA-based regulation. The 70-110 amino acid protein forms homohexamers and binds sRNA as well as mRNA (via its proximal site) and also poly(A) tails. In *E. coli*, approximately one fourth of the known sRNAs bind to *Hfq*. They also interact with proteins that are involved in mediated decay of specific mRNAs.

In the following we can only briefly introduce a few subjectively selected examples of small bacterial RNAs.



**Figure 1.14** Mechanisms of riboswitch control. Most switches are located in the 5' UTR (top), acting either by forming a transcription terminator, blocking the anti-terminator hairpin, or sequestering the Shine-Dalgarno sequence. Less common riboswitches (below) act by cleaving (gmlS), causing alternative splicing (TPP), or stabilizing the mRNA.

**OLE RNA.** The “ornate, large and extremophilic” (OLE) RNA with a length of approximately 610nt is highly conserved in both sequence and secondary structure. OLE is found predominantly in extremophilic Gram-positive eubacteria. A functional link between OLE RNA and putative membrane protein including BH2780 has been suggested [200].

**GcvB.** Computer-based detection of GcvB showed the existence of the sequence within many enterobacteria as well as *Pasteurellaceae* and *Vibrionaceae*. Further investigation in *Salmonella enteria* revealed that it regulates at least seven mRNAs. The GcvB RNA represses its targets by binding to highly conserved regions. Examples include a 29nt G-U rich region and the Shine-Dalgarno sequence of *dpA* and *oppA*. GcvB RNA also binds upstream of the Shine-Dalgarno sequence, preventing the binding of the 30S ribosome [220].

**Yfr1.** Cyanobacterial functional RNAs (Yfr) [12, 176] were discovered by computer aided searches. The Yfr2-5 appear to be related due to a highly conserved sequence pattern ('GGAAACA' x2) within the loop region of a hairpin. Yfr7 exhibits a faint sequence similarity to 6S RNA. The approximately 60nt short Yfr1 RNA folds into a structure containing an ultra-conserved sequence surrounded by two stem-loops. This element is found in all cyanobacteria lineages, where it appears in very high copy numbers and with a long half-life of 60min, emphasizing its functional importance. Yfr1 seems to be involved in growth and stress responses, but its regulatory mechanism remains largely unknown.

**RsmY and RsmZ.** A complex network of ncRNAs and proteins controls pathogenesis in *Pseudomonas aeruginosa* [241]. Homologous networks of GacS-GacA-RsmY-Z-RsmA are known in *E. coli* and many other bacterial species. The GacS-GacA proteins up-regulate the expression of RsmY and RsmZ. The sensor proteins LadS and RetS seem to up- as well as down-regulate GacA respectively. RsmY and RsmZ bind to the translation regulatory protein RsmA. The sequestration of this molecule results in expression of survival and virulence associated mRNAs such as the elements of the type-III-secretion complex.

**Riboswitches.** These primarily regulatory elements of protein-coding mRNAs are mostly found in bacteria. We include them here because in the “off” state the riboswitch sequence itself is often transcribed. Almost all riboswitches are concatenations of two components: a very well conserved binding domain acting as a highly specific *aptamer* that senses its metabolite, and a comparatively variable *expression platform* which changes its structure

**Table 1.3** Classification of verified riboswitches by their function, and primary ligand as well as candidate riboswitches are listed. The SAM and PreQ<sub>1</sub> groups utilize different structures to bind the same ligand. Where known, the mode of action is indicated by  $\square$ (transcription control) and/or  $\boxtimes$ (translation control). Data are taken compiled from [23, 15, 46, 203, 162, 255].

	Riboswitch	ligand	mRNA-location	expression regulation	
Cleavage	glmS	glucose-amine-6-phosphate	5' UTR	-	$\boxtimes$
	TPP (THI-box)	thiamin pyrophosphate	intron in 5'UTR	-	$\boxtimes$
ON/OFF	TPP (THI-box)	thiamin pyrophosphate	3'UTR, 5'UTR	+/-	$\square$
	Purine	adenine, guanine and 2'-deoxyguanosine	5'UTR	+/-	$\square$ $\boxtimes$
	Lysine (L-box)	lysine	5'UTR	-	$\square$
	Glycin	glycine	5'UTR	+	$\square$
	Cobalmin (B12-element)	adenosylcobalamin	5'UTR	-	$\boxtimes$
	FMN (RFN-element)	flavin mononucleotide	5'UTR	-	$\square$ $\boxtimes$
	Mg <sup>2+</sup>	Mg <sup>2+</sup>	5'UTR	-	$\square$
	PreQ <sub>1</sub> -I	pre-queuosine <sub>1</sub>	5'UTR	-	$\boxtimes$
	PreQ <sub>1</sub> -II				
	SAM-I (S-box leader)				
	SAM-II (SAM-alpha)				
	SAM-III (S <sub>MK</sub> )	S-adenosylmethionine	5'UTR	-	$\square$ $\boxtimes$
	SAM-IV	S-adenosylhomocystein, S-adenosylcysteine	5'UTR	+/-	$\square$ $\boxtimes$
	SAH				
Candidates	Moco	molybdenum cofactor	5'UTR	-	$\square$ $\boxtimes$
	Tuco	tungst cofactor			
	SAM-V				
	yybP-ykoY (SraF)				
	ykkC-yxkD				

upon binding of the metabolite. One can distinguish two functional principles: in kinetically controlled riboswitches, metabolite binding competes with RNA folding; in thermodynamically controlled ones, the binding energy of the metabolite is sufficient to cause structural changes [46].

The expression platform can employ two principal modes of action, Fig. 1.14. It can regulate translation by inhibiting translational initiation, e.g. by blocking the ribosomal binding site. Alternatively, riboswitches can control transcription. Depending on whether the aptamer is loaded with the metabolite or not, a terminator hairpin is formed by the expression platform that pre maturely terminates the transcription of the mRNA. In this case, the “raw” riboswitch RNA is produced as an unstable product. The TPP riboswitch is mechanistically more complex [15] and employs various expression platforms. It can be found in both 5' and 3' UTRs; in the latter case it usually acts in conjunction with an upstream

ORF. Riboswitches may act as activators or repressors depending on genomic context [36] by switching ON→OFF or OFF→ON with increasing substrate concentration. A more complex regulatory logic can be implemented by placing two or more riboswitches which sense the same or different ligands in series. For instance, a tandem riboswitch architecture in *Bacillus clausii* implements a logical NOR gate. This and further examples are reviewed in [229].

A unique case is the *gmlS* riboswitch in certain Gram-positive bacteria. Instead of instigating a conformational change, it stimulates a self-cleaving ribozyme activity that acts on the *GmlS* mRNA [120, 43].

Riboswitches are common in bacteria, although their phylogenetic distribution and genomic frequency is quite variable [116, 255]. A few riboswitches have also been detected in archaea, fungi and plants [228, 15, 23]. A TPP riboswitch in *Neurospora crassa*, for example, regulates the expression of mRNAs by controlling alternative splicing.

#### 1.4.6 A Zoo of Diverse Examples

The reasonably well understood ncRNA classes have been complemented during the last few years by many other examples which are less well-described. In the following paragraphs we can only briefly touch upon a few of them.

**Guide RNAs.** Mitochondrial mRNAs of some protozoa need to undergo a post-transcriptional editing process before they can be translated. Kinetoplastids of the trypanosomatid group possess two types of mitochondrial DNA molecules: Maxicircles bear protein and ribosomal RNA genes. Minicircles specify guide RNAs (gRNAs), with a typical length of about 50nt that mediate uridine insertion/deletion RNA editing. Following the hybridization of the 5'-anchor region of a gRNA to the 3' end of its target mRNA, an U insertion and deletion is directed by sequential base pairing. The enzyme cascade involves cleavage of the mRNA, U insertion by a TUTase, and re-ligation, see [2] and the references therein. With a few exceptions, gRNAs are encoded in the short variable regions located between highly conserved sequence blocks on the minicircles. A computational approach based on this observation predicted about 100 gRNA candidates in *Trypanosoma cruzi* [240], consistent with recent experimental results for *Trypanosoma brucei* [148, 149].

***Dictyostelium discoideum.*** In addition to the usual repertoire of eukaryotic ncRNA, there are two related classes of ncRNAs that share the sequence motif CCUUACAGCAA [8]. One of them, the class I RNAs, is organized in a few genomic clusters [173, 96]. Several novel families of ncRNAs have been identified based on a computational screen and subsequent experimental verification [131].

***Plasmodium falciparum.*** The ncRNAs of this malaria parasite have been studied rather extensively. Its notable scarcity of identifiable transcription factors led to speculation that this organism may be unusually reliant on chromatin modifications as a mechanism for regulating gene expression. The centromeres of *Plasmodium falciparum* contain transcriptionally active promoters and produce non-coding transcripts with a length of 75-175nt that are retained in the nucleus and appear to associate with the centromeres [66]. A large number of other ncRNAs without homologs outside apicomplexa have been found using computational techniques and microarrays [34, 174].

*TelRNAs.* A recent study [216] showed that mammalian telomeric repeats are transcribed from the C-rich strand by polymerase-II. The transcripts contain UUAGGG repeats and are polyadenylated. Transcription is regulated depending on developmental status, telomere length, cellular stress, tumour stage and chromatin structure. *In vitro*, TelRNAs block telomerase activity suggesting an active role of TelRNAs in regulating telomerase *in vivo*. Similar transcripts were recently reported in *Leishmania donovani* [214], suggesting that TelRNAs might be evolutionarily ancient.

*Promoter- and Termini-Associated RNAs.* A high resolution tiling array map of the human transcriptome implied a novel role for some unannotated RNAs as primary transcripts for the production of short RNAs, and identified three novel classes of RNAs [113]: Both “promoter-associated small RNAs” (pasRNAs) and “termini-associated small RNAs” (tasRNAs) are syntenically conserved between human and mouse. The presence of pasRNAs appears to be associated with a small increase in the expression of the corresponding protein coding gene. Short sequence reads arising from human bi-directional promoters, which might be pasRNAs or at least related to them, are also reported in [115]. Longer “promoter-associated long RNAs” (palRNAs) with a length of a few hundred nucleotides are also abundant throughout the human genome [113]. A detailed study of a palRNA associated with the promoter of EF1a suggests a function in epigenetic gene silencing [87].

*RNA Polymerase-III transcripts.* Genome-wide surveys have recently uncovered a plethora of novel, hitherto unclassified transcripts are produced by RNA polymerase-III [105, 184], reviewed in [56]. In *Drosophila* many of these transcripts exhibit well-conserved secondary structures [209]. The snaR-A RNA, on the other hand, is present in human and chimpanzee only and appears to have undergone accelerated evolution [187].

*TIN RNAs.* A survey of human mRNA and EST public databases revealed more than 55,000 totally intronic non-coding (TIN) RNAs transcribed from the introns of the majority of RefSeq genes [177]. Surprisingly, RNA polymerase-II inhibition resulted in increased expression of a fraction of intronic RNAs in cell cultures. This suggests that the recently discovered spRNAP-IV, an RNA polymerase of mitochondrial origin [124], might be responsible for these transcripts. Functional importance of some TIN RNAs is supported by conserved expression patterns between human and mouse [144].

*IGS RNAs.* The intergenic spacers (IGS) that separate individual ribosomal rDNA genes contain polymerase-I promoters that cause the transcription of 150-300nt ncRNA complementary to rRNA promoter regions. These IGS RNAs help to establish and maintain a specific heterochromatin configuration in a subset of rRNA promoters [160].

*Ciliates.* *Tetrahymena* and other ciliates use an RNA-based mechanism for directing their genome-wide DNA rearrangements [167, 278]. The “scanning RNAs” appear to be closely related to the RNAi pathway.

## 1.5 ncRNAs FROM REPEATS AND PSEUDOGENES

Repetitive elements account for about half of mammalian genomes. In the past, these sequences were often considered as “junk DNA”, i.e., devoid of cellular function [18]. We are only beginning to understand that this is probably very far from the truth: For

example, two distinct short polymerase-III transcribed SINE elements, B2 in mouse and Alu in human, have been recognized as negative regulators of polymerase-II transcription [3, 64, 65, 153]. Both act in a similar way by directly binding to polymerase-II. The Alu RNA arose from the fusion of the 5' and 3' ends of 7SL RNA and later on evolved by a head to tail fusion of two related Alu sequences into a dimeric structure. It can still bind some of the SRP proteins. The resulting Alu RNP complex has been found to down regulate translation initiation [89]. In *Leishmania infantum*, conserved tandem head-to-tail subtelomeric repeats are expressed in a stage-specific manner [59]. The expression of telomeric repeats has been briefly described in the previous section.

Repetitive DNA elements as well as pseudogenes are an important source of novel ncRNA genes. In some cases, protein coding genes lose their coding capacity and become functional as ncRNA. Probably the best-known example is *Xist* [60], see section 1.6.

A different mechanism, “exaptation” [25], starts with the reactivation of retrotransposed pseudogenes, which are either by chance integrated into a locus that provides promoter sequences, or integrated into a locus where a promoter happens to be generated by mutations after integration. An example of an exapted gene is the mRNA-like ncRNA *Makorin1-pl* [277]: Both the functional protein-coding gene *Makorin1* and the pseudogene *Makorin1-pl* possesses the same cis-acting destabilizing elements in their 5' region. The pseudogene stabilizes its functional paralog by competing for the the mRNA degradation apparatus. In some cases, pseudogenes may also be processed to generate small RNAs, including microRNAs and piRNAs, see →chapter \*\*\* for examples. In addition, several snoRNAs seem to belong to repetitive families in the mouse genome [102].

Another source of pseudogenes are small stable RNAs such as tRNAs, 7SK RNAs or even snoRNAs [145]. Retrotransposition produces large numbers of such pseudogenes which may give rise to novel ncRNAs. The BC1 RNA, for instance, which is specific to rodent neurons, shares 80% sequence similarity with its progenitor tRNA<sup>Ala</sup>. It folds into a stable stem/loop rather than into a cloverleaf structure [210]. BC200, another brain-specific transcript, is specific to anthropoidea [126] and exapted from a retrotransposed ancient Alu monomer. Although unrelated evolutionary, BC1 RNA and BC200 RNA share the same expression pattern and exert analogous functions [280]. A related primate lineage contains the analogous, also Alu-derived, ncRNA G22 [117]. Two closely related RNAs are exapted from rodent SINEs. The 94nt 4.5SH RNA and the 101-108nt 4.5S<sub>I</sub> RNA are exapted from rodent B1 and B2 elements, respectively [79]. The function of these RNAs remains unknown although it was shown that 4.5SH is bound by mouse nucleolin [98].

## 1.6 mRNA-LIKE ncRNAs

A rapidly growing class of ncRNAs looks like protein-coding messenger RNAs in many respects. These mRNA-like ncRNAs (mlncRNAs) are transcribed by polymerase-II, polyadenylated at their 3' end, capped with 7-methylguanosine at the 5' end, and typically spliced. These transcripts are the main target of systematic full-length cDNA cloning, reviewed in [32]. While huge numbers of mlncRNAs were found in both animals [103, 32, 231] and plants [264, 212], next to nothing is known about most of them. There is, however, mounting evidence that many of them are associated with diseases [232], Tab. 1.4. A recent study based on *in situ* hybridization data from the Allen Brain Atlas identified 849 ncRNAs expressed in mouse, of which most were specifically associated with particular neuroanatomical regions, cell types, or subcellular compartments [161]. This kind of tight regulation is at least indicative of specific functionalities.



**Table 1.4** mlncRNAs associated with human diseases. Altered expression in disease/disorder: ↑ up-regulated, ↓ down-regulated.

ncRNA	Disease/disorder	Ref.
Altered Expression Levels in Cancer		
PCGEM1	↑ prostate cancer	[225]
DD3/PCA3	↑ prostate cancer	[28]
MALAT-1	NSCLC, endometrial sarcoma, hepatocellular carcinoma	[275, 141]
OCC-1	↑ colon carcinoma	[192]
NCRMS	↑ alveolar rhabdomyosarcoma	[35]
BCMS/DLEU1	B-cell neoplasia	[271]
H19	↑ liver and breast cancer	[73, 159]
NC612	prostate cancer	[222]
HULC	↑ hepatocellular carcinoma	[186]
HIS-1	↑ myeloid leukemia	[136]
BIC	accumulates B cell lymphoma & leukemia	[63, 233]
SRA	isoform in breast cancer	[130]
TRNG10	various cancers	[207]
U50HG	at chromosomal breakpoint in B-cell lymphoma	[234]
PEG8/IGF2AS	fetal tumors	[183]
Neurological diseases/disorders		
SZ-1/PSZA11q14	↓ schizophrenia	[197]
DISC2	schizophrenia and bipolar affective disorder	[163, 42]
IPW	Prader-Willi syndrome	[265]
SCA8	Spinocerebellar ataxia type 8	[175]
Miscellaneous diseases/disorders		
DGCR5	disrupted in DiGeorge syndrome	[230]
MIAT	risk of myocardial infarction	[104]
22k48	deletion in Dg George syndrome	[195]
LIT1	Romano-Ward, Jervell, Lange-Nielsen & Beckwith-Wiedemann	[100, 180]
BR514	congenital developmental abnormalities	[86]

Recent data strongly suggest that mlncRNAs do not form a homogenous class with respect to function and processing. A large subclass are natural antisense transcripts (NATs), which are implicated in the expression regulation of their protein-coding counterparts in both animals and plants [114, 137].

A significant fraction of transcripts, which probably included many of the mlncRNAs is processed into short RNAs [113]. Only a small number of such examples is reasonably well understood at present however. The most prominent mlncRNAs that function as carriers of other functional ncRNAs are the non-coding host-genes of snoRNAs [244, 53] and primary microRNA precursors [94] including H19 [29] and BIC [233]. In [31], it is shown that a co-expressed pair of a sense and antisense transcript of the phosphate transporter gene *Slc34a2a* is specifically processed into small RNAs. We refer to →chapter \*\*\* for more details on miRNAs, siRNAs, and their relatives.

A small subclass of mlncRNAs is predominantly present in the nucleus. A recent screen [101], identified only four such genes: the two well-conserved genes NEAT1 and MALAT-



1, NTT, and *Xist*, which is well-studied in X-chromosome inactivation [101, 179]. The neuronally expressed mouse transcript *Gomafu* also seems to belong to this class [223].

From a functional point of view one can distinguish transcripts involved in dosage compensation, imprinting events, stress signals, and regulators of gene expression. Mechanistic details, however, remain to be elucidated. In the following we briefly touch upon a few of the better-understood representatives.

**Dosage compensation.** Many species with sex chromosomes, including mammals and flies, need to equalize the expression levels of (in this case) the X chromosome genes in the different sexes. In *Drosophila* this is achieved by X chromosome up regulation in XY cells with the help of the lncRNAs *rox1* and *rox2* [273], while mammals inactivate one of the two X chromosomes, reviewed in [101, 179]. The X inactivation proceeds at an early developmental stage in females and is regulated by different factors, including a region of chromosome X the so-called X inactivation center (XIC). The *Xist* (X inactive specific transcript) gene, a nuclear 19.3 kb transcript exclusively expressed from the XIC of the inactive X chromosome, is necessary and sufficient for the inactivation of the X chromosome. So far, the exact mechanism is unclear but it is assumed that only the transcription of *Xist* could be enough to change the chromatin structure to allow the binding of different silencing factors.

**Imprinting.** The 2.3kb *H19* transcript is the first and probably best characterized autosomally imprinted gene. It is located in a cluster of imprinted genes (11p15 in human) containing also the *IGF2* gene. While *H19* is expressed exclusively from the maternal allele, *IGF2* expression is limited to the paternal allele. *H19* RNA is highly expressed in placental, embryonic and most foetal tissues, but after birth the expression is suppressed in nearly all tissues. Recently identified as a microRNA precursor [29], it appears to play a role in development and differentiation. Both loss and overexpression of *H19* is associated with different cancers.

**Stress Response.** Both prokaryotes and eukaryotes induce a set of heat shock genes to counteract environmental stress, reviewed in [7]. The heat shock response not only causes a widespread inhibition of transcription, but also a blockade of splicing and other post-transcriptional processing. Usually, heat shock genes code for proteins. In *Drosophila*, however, the major site of transcription after temperature induced stress is the *hsw* locus, which produces a ncRNA, see [111] for a recent review. The *hsw* RNA is constitutively expressed nearly ubiquitously and its transcription level can be rapidly expressed in response to stress signals. There seem to be three distinct isoforms: both the full length (10kb) *hsw1* and the 7-8kb *hsw2* RNA, which is obtained by alternative polyadenylation, accumulate in the nucleus. In contrast, the spliced 1.2-1.3kb *hsw3* RNA is cytoplasmic. The *hsw* RNAs contain a short translatable reading frame in several *Drosophila* species; however, the corresponding peptides have not been detected. The large RNAs appear to act as “organizers” for the sequestering components of the mRNA processing machinery: Together with diverse hnRNPs, the nuclear *hsw* RNAs are localized in subnuclear compartments. These *ω*-speckles are believed to act as dynamic storage for RNA-processing and related proteins.

Mammals do not have an *hsw* homolog. Recently heat-induced ncRNAs, transcribed from satellite III repetitive sequence, have been described in human cells. Therefore, the polyadenylated satellite III transcripts are functional analogs of the *hsw* [111].

**Transcriptional Regulators.** The 2.7kb *Evf-1* transcript and its 3.8kb splice variant *Evf-2* originate from the *Dlx-5/6* bi-gene cluster and overlap an ultraconserved region. The *Dlx* genes are homeodomain transcription factors with crucial functions in differentiation and migration of neuronal cells as well as craniofacial and limb patterning during development. The *Evf-2* RNA specifically interacts with the Dlx-2 protein, forming a stable complex in the nucleus. The *Evf-2/Dlx-2* interaction increases the transcriptional activity of the *Dlx-5/6* enhancer region in a target and homeodomain-specific manner. Most likely, the *Evf-2/Dlx-2* complex stabilizes the binding between the Dlx-2 homeodomain protein and the *Dlx5/6* enhancer sequence [67, 70, 121].

A few more examples that are at least partially understood are described in some detail in recent reviews [199, 212, 231].

## 1.7 RNAs WITH DUAL FUNCTIONS

The complex mosaic of transcripts outlined in the introduction implies that protein-coding and non-coding transcripts frequently overlap, in different reading directions or even in the same direction. In several cases, distinct types of functional products are produced from the same primary transcript, the best-known example being snoRNAs that are frequently processed from introns of genes for ribosomal proteins. An extreme example of this type is *mfl*, the locus for the pseudouridine synthase minifly/Nop60b of *Drosophila melanogaster*. It not only encodes alternative splice forms that can be polyadenylated at different downstream poly(A) sites but also contains within its introns a cluster comprising four isoforms of a C/D box snoRNA and two highly related copies of a small ncRNA genes of unknown function. The alternative 3' ends allow *mfl* not only to produce two distinct protein subforms, but also to differentially release different ncRNAs [205]. Coding and non-coding information can be packed even more tightly, however, by superimposing it on the same sequence. Outside RNA viruses, a few examples are known in both eubacteria and eukaryotes.

**RNAIII.** With a length of 514nt, the *Staphylococcus*-specific RNAIII is one of the largest regulatory RNAs in bacteria. As an intracellular effector of the quorum-sensing system, it is a key regulator in virulence gene expression [241]. A 14 stem-loop regulatory structure and the  $\delta$ -hemolysin peptide, a short ORF close to the 5' end are encoded within one genomic locus. RNAIII exerts its function by binding to at least three target mRNAs, *hla*, *spa* and *rot*, and alters the expression levels of the corresponding proteins by modifying the accessibility of the Shine-Dalgarno sequence [20, 241].

**SgrS.** The 227nt SgrS RNA is expressed in *Escherichia coli* during glucose-phosphate stress by downregulating the translation of the glucose transporter in an Hfq-dependent manner. The 5' region contains a 43nt ORF, *sgrT*, which is well conserved and translated under stress conditions [253]. So far, SgrS RNAs have been described for various enterobacteria.

**SRA/SRAP.** In human, the *steroid receptor RNA activator* modulates transcriptional activity of steroid receptors as an RNA molecule. On the other hand, it encodes a protein that is highly conserved among chordata. A recent review [134] lists 13 SRA variants, apparently arising from alternative transcription start sites and alternative splicing. They all share exons 2-5 which encode the functional secondary structure core. Some of these

isoforms also encode the SRAP protein, others lack the translation start. It appears that the main function of SRA RNA is to organize the various protein components in the SRA-RNP complexes which contain both transcription factors and positive or negative regulators of nuclear receptor activity.

*Enod40.* The plant gene *enod40* participates in the regulation of symbiotic interactions between leguminous plants and bacteria or fungi, and it has been implicated in the development also of non-symbiotic plants [224]. Its molecular mechanisms remain unclear, but both short peptides and well-conserved RNA secondary structure appear to play a role. A recent computational study [84] demonstrated a well-conserved structural core that is conserved across angiosperms, and the presence of highly variable expansion domains reminiscent of the patterns observed in many other functional ncRNAs. Legumes often contain more than one *enod40* gene.

The analysis of transcript structures in the ENCODE regions shows that overlapping arrangements of coding and non-coding transcripts and spliceforms are the rule rather than the exceptions. At this point, however, the relevance of non-coding RNAs arising from protein coding loci remains unclear.

The recent discovery of the short *tarsal-less* peptide translated from only 33nt-long ORFs of a *Drosophila* transcript previously classified as non-coding [74] might indicate that many other “mlncRNAs” in fact code for short peptides. The peptides of *enod40* and *tarsal-less* are highly conserved over long evolutionary time. So far, no additional examples have been reported (apart from uORFs of protein coding mRNAs).

## 1.8 CONCLUDING REMARKS

We have attempted in this chapter to give a comprehensive overview of the inventory of non-protein-coding RNAs across all domains of life, excluding viral RNAs, viroid and satellite DNAs, ribozymes, and regulatory RNA elements of mRNAs. Of course, in a few pages such an endeavor is bound to remain incomplete and subjective. New classes, mechanisms, and functions of ncRNAs being discovered almost every week. Therefore, this chapter will likely be even more incomplete, and half-way outdated, when it reaches the reader in printed form. After all, a series of computational studies has provided quite convincing evidence for tens of thousands of unclassified RNAs whose secondary structure is under stabilizing selection [164, 209, 242, 247, 256, 257]. Structure-based clustering [268, 208] furthermore strongly suggests that several new classes of ncRNAs with characteristic secondary structures are still lurking in eukaryotic genomes.

In order to keep the list of references at reasonable length, we had to give priority to reviews over the reviewed original works, and to give preference to most recent publications over classical papers, hence this chapter does not attempt to review the history of the discovery of the “Modern RNA-World” over the last decade.

*Acknowledgements.* This work was supported in part by the German *DFG* under the auspices of SPP-1258 “Sensory and Regulatory RNAs in Prokaryotes”, SPP-1174 “Metazoan Deep Phylogeny”, and the *Graduierten-kolleg Wissensrepräsentation*, by the European Union through the 6th framework program projects EMBIO <http://www-embio.ch.cam.ac.uk/> and SYNLET <http://synlet.izbi.uni-leipzig.de/>. We thank Claudia S. Copeland for editing the manuscript for english language and clarity.

## REFERENCES

1. M. C. Accardo, E. Giordano, S. Riccardo, F. A. Digilio, G. Iazzetti, R. A. Calogero, and M. Furia. A computational search for box C/D snoRNA genes in the *Drosophila melanogaster* genome. *Bioinformatics*, 20:3293–3301, 2004.
2. V. S. Alatortsev, J. Cruz-Reyes, A. G. Zhelonkina, and B. Sollner-Webb. *Trypanosoma brucei* rna editing: coupled cycles of U deletion reveal processive activity of the editing complex. *Mol Cell Biol*, 28:2437–2445, 2008.
3. T. A. Allen, S. Von Kaenel, J. A. Goodrich, and J. Kugel. The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nat Struct Mol Biol*, 11:816–821, 2004.
4. S. Altuvia. Identification of bacterial small non-coding RNAs: experimental approaches. *Curr Opin Microbiol*, 3:257–261, 2007.
5. E. S. Andersen, M. A. Rosenblad, N. Larsen, J. C. Westergaard, J. Burs, I. K. Wower, J. Wower, J. Gorodkin, T. Samuelsson, and C. Zwieb. The tmRDB and SRPDB resources. *Nucleic Acids Res*, 34:D163–D168, 2006.
6. E. V. Armbrust, J. A. Berges, C. Bowler, B. R. Green, D. Martinez, N. H. Putnam, S. Zhou, A. E. Allen, K. E. Apt, M. Bechner, M. A. Brzezinski, B. K. Chaal, A. Chiovitti, A. K. Davis, M. S. Demarest, J. C. Detter, T. Glavina, D. Goodstein, M. Z. Hadi, U. Hellsten, M. Hildebrand, B. D. Jenkins, J. Jurka, V. V. Kapitonov, N. Kröger, W. W. Y. Lau, T. W. Lane, F. W. Larimer, J. C. Lippmeier, S. Lucas, M. Medina, A. Montsant, M. Obornik, M. S. Parker, B. Palenik, G. J. Pazour, P. M. Richardson, T. A. Rynearson, M. A. Saito, D. C. Schwartz, K. Thamtrakoln, K. Valentin, A. Vardi, F. P. Wilkerson, and D. S. Rokhsar. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science*, 306:79–86, 2004.
7. R. Arya, M. Mallik, and S. C. Lakhota. Heat shock genes — integrating cell survival and death. *J Biosci*, 32:595–610, 2007.
8. A. Aspegren, A. Hinas, P. Larsson, A. Larsson, and F. Söderbom. Novel non-coding RNAs in *Dictyostelium discoideum* and their expression during development. *Nucleic Acids Res*, 32:4646–4656, 2004.
9. T. V. Aspinall, J. M. B. Gordon, H. J. Bennet, P. Karahalios, J.-P. Bukowski, S. C. Walker, D. R. Engelke, and J. M. Avis. Interactions between subunits of *Saccharomyces cerevisiae* RNase MRP support a conserved eukaryotic RNase P/MRP architecture. *Nucleic Acids Res*, 35:6439–6450, 2007.
10. V. Atzorn, P. Fragapane, and T. Kiss. U17/snR30 is a ubiquitous snoRNA with two conserved sequence motifs essential ribosome assembly. *Cell*, 72:443–457, 2004.
11. I. M. Axmann, J. Holtzendorff, B. Voss, P. Kensch, and W. R. Hess. Two distinct types of 6S RNA in *Prochlorococcus*. *Gene*, 406:69–78, 2007.
12. I. M. Axmann, P. Kensch, J. Vogel, S. Kohl, H. Herzel, and W. R. Hess. Identification of cyanobacterial non-coding RNAs by comparative genome analysis. *Genome Biol*, 6:R73, 2005.
13. T. N. Azzouz and D. Schümperli. Evolutionary conservation of the U7 small nuclear ribonucleoprotein in *Drosophila melanogaster*. *RNA*, 9:1532–1541, 2003.
14. J.-P. Bachellerie, J. Cavaillé, and A. Hüttenhofer. The expanding snoRNA world. *Biochimie*, 84:775–790, 2002.
15. J. E. Barrick and R. R. Breaker. The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol*, 8:R239, 2007.
16. J. E. Barrick, N. Sudarsan, Z. Weinberg, W. L. Ruzzo, and R. R. Breaker. 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. *RNA*, 11:774–784, 2005.

17. P. S. Bazeley, V. Shepelev, Z. Talebizadeh, M. G. Butler, L. Fedorova, V. Filatov, and A. Fedorov. snoTARGET shows that human orphan snoRNA targets locate close to alternative splice junctions. *Gene*, 408:172–179, 2008.
18. V. P. Belancio, D. J. Hedges, and P. Deininger. Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. *Genome Res*, 18:343–358, 2008.
19. T. Blumenthal. Trans-splicing and polycistronic transcription in *Caenorhabditis elegans*. *Trends Genet*, 11:132–136, 1995.
20. S. Boisset, T. Geissmann, E. Huntzinger, P. Fechter, N. Bendridi, M. Possedko, C. Chevalier, A. C. Helfer, Y. Benito, A. Jacquier, C. Gaspin, F. Vandenesch, and P. Romby. *Staphylococcus aureus* RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. *Genes Dev*, 21:1353–1366, 2007.
21. A. F. Bompfünnewerer, C. Flamm, C. Fried, G. Fritsch, I. L. Hofacker, J. Lehmann, K. Missal, A. Mosig, B. Müller, S. J. Prohaska, B. M. R. Stadler, P. F. Stadler, A. Tanzer, S. Washietl, and C. Witwer. Evolutionary patterns of non-coding rnas. *Th Biosci*, 123:301–369, 2005.
22. S. Braud, C. Lavire, A. Bellier, and P. Mazodier. Effect of SsrA (tmRNA) tagging system on translational regulation in *Streptomyces*. *Arch Microbiol*, 184:343–352, 2006.
23. R. R. Breaker. Complex riboswitches. *Science*, 319:1795–1797, 2008.
24. R. G. Brennan and T. M. Link. Hfq structure, function and ligand binding. *Curr Opin Microbiol*, 10:125–133, 2007.
25. J. Brosius. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene*, 238:115–134, 1999.
26. J. W. S. Brown, M. Echeverria, and L.-H. Qu. Plant snoRNAs: functional evolution and new modes of gene expression. *Trends Plant Sci*, 8:42–49, 2003.
27. Y. Brown, M. Abraham, S. Pearl, M. M. Kabaha, E. Elboher, and Y. Tzfati. A critical three-way junction is conserved in budding yeast and vertebrate telomerase RNAs. *Nucleic Acids Res*, 35:6280–6289, 2007.
28. M. J. Bussemakers, A. van Bokhoven, G. W. Verhaegh, F. P. Smit, H. F. Karthaus, J. A. Schalken, F. M. Debruyne, N. Ru, and W. B. Isaacs. DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Res*, 59:5975–5979, 1999.
29. X. Cai and B. R. Cullen. The imprinted H19 noncoding RNA is a primary microRNA precursor. *RNA*, 13:313–316, 2007.
30. K. Calvin and H. Li. Rna-splicing endonuclease structure and function. *Cell Mol Life Sci*, 65:1176–1185, 2008.
31. M. Carlile, P. Nalbant, K. Preston-Fayers, G. S. McHaffie, and A. Werner. Processing of naturally occurring sense/antisense transcripts of the vertebrate Slc34a gene into short rnas. *Physiol Genomics*, 2008. Doi:10.1152/physiolgenomics.00004.2008.
32. P. Carninci. Constructing the landscape of the mammalian transcriptome. *J Exp Biol*, 210:1497–1506, 2007.
33. A. T. Cavanagh, A. D. Klocko, X. Liu, and K. M. Wassarman. Promoter specificity for 6S RNA regulation of transcription is determined by core promoter sequences and competition for region 4.2 of sigma70. *Mol Microbiol*, 67:1242–1256, 2008.
34. K. Chakrabarti, M. Pearson, L. Grate, T. Sterne-Weiler, J. Deans, and M. Donohue, J P and Ares Jr. Structural RNAs of known and unknown function identified in malaria parasites by comparative genomics and RNA analysis. *RNA*, 13:1923–1939, 2007.
35. A. S. Chan, P. S. Thorner, J. A. Squire, and M. Zielenska. Identification of a novel gene NCRMS on chromosome 12q21 with differential expression between rhabdomyosarcoma subtypes. *Oncogene*, 21:3029–3037, 2002.

36. M. T. Cheah, A. Wachter, N. Sudarsan, and R. R. Breaker. Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature*, 447:497–500, 2007.
37. J.-L. Chen and C. W. Greider. An emerging consensus for telomerase RNA structure. *Proc Natl Acad Sci USA*, 101:14683–14684, 2004.
38. L. Chen, D. J. Lullo, E. Ma, S. E. Celniker, D. C. Rio, and J. A. Doudna. Identification and analysis of U5 snRNA variants in *Drosophila*. *RNA*, 11:1473–1477, 2005.
39. X. Chen, A. M. Quinn, and S. L. Wolin. Ro ribonucleoproteins contribute to the resistance of *Deinococcus radiodurans* to ultraviolet resistance. *Genes Dev*, 14:777–782, 2000.
40. X. Chen, T. S. Rozhdestvensky, L. J. Collins, J. Schmitz, and D. Penny. Combined experimental and computational approach to identify non-protein-coding RNAs in the deep-branching eukaryote *Giardia intestinalis*. *Nucleic Acids Res*, 35:4619–4628, 2007.
41. C. P. Christov, T. J. Gardiner, D. Szüts, and T. Krude. Functional requirement of noncoding Y RNAs for human chromosomal DNA replication. *Mol Cell Biol*, 26:6993–7004, 2006.
42. J. E. Chubb, N. J. Bradshaw, D. C. Soares, D. J. Porteous, and J. K. Millar. The DISC locus in psychiatric illness. *Mol Psychiatry*, 13:36–64, 2008.
43. J. C. Cochrane, S. V. Lipchok, and S. A. Strobel. Structural investigation of the GlmS ribozyme bound to its catalytic cofactor. *Chem Biol*, 14:97–105, 2007.
44. L. Collins and D. Penny. Complex spliceosomal organization ancestral to extant eukaryotes. *Mol Biol Evol*, 22:1053–1066, 2005.
45. L. J. Collins, V. Moulton, and D. Penny. Use of RNA secondary structure for studying the evolution of RNase P and RNase MRP. *J Mol Evol*, 51:194–204, 2000.
46. R. L. Coppins, K. B. Hall, and E. A. Groisman. The intricate world of riboswitches. *Curr Opin Microbiol*, 10:176–181, 2007.
47. A. T. Dandjinou, N. Lévesque, S. Larose, J.-F. Lucier, S. A. Elela, and R. J. Wellinger. A phylogenetically based secondary structure for the yeast telomerase RNA. *Curr Biol*, 14:1148–1158, 2004.
48. X. Darzacq, B. E. Jány, C. Verheggen, A. M. Kiss, E. Bertrand, and T. Kiss. Cajal body-specific small nuclear RNAs: a novel class of 2'-o-methylation and pseudouridylation guide RNAs. *The EMBO journal*, 21:2746–2756, 2002.
49. L. David, W. Huber, M. Granovskaia, J. Toedling, C. J. Palm, L. Bofkin, T. Jones, R. W. Davis, and L. M. Steinmetz. A high-resolution map of transcription in the yeast genome. *Proc Natl Acad Sci USA*, 103:5320–5325, 2006.
50. M. Dávila López, M. Alm Rosenblad, and T. Samuelsson. Computational screen for spliceosomal RNA genes aids in defining the phylogenetic distribution of major and minor spliceosomal components. *Nucleic Acids Res*, 36:001–3010, 2008.
51. M. Dávila López and T. Samuelsson. Early evolution of histone mRNA 3' end processing. *RNA*, 14:1–10, 2008.
52. J. de la Cruz and A. Vioque. A structural and functional study of plastid RNAs homologous to catalytic bacterial RNase P RNA. *Gene*, 321:47–56, 2003.
53. T. de los Santos, J. Schweizer, C. A. Rees, and U. Francke. Small evolutionarily conserved RNA, resembling C/D box small nucleolar RNA, is transcribed from PWCR1, a novel imprinted gene in the Prader-Willi deletion region, which is highly expressed in brain. *Am J Hum Genet*, 67:1067–1082, 2000.
54. W. A. Decatur, X. H. Liang, D. Piekna-Przybylska, and M. J. Fournier. Identifying effects of snoRNA-guided modifications on the synthesis and function of the yeast ribosome. *Methods Enzymol*, 425:283–316, 2007.



55. W. Deng, X. Zhu, G. Skogerbø, Y. Zhao, Z. Fu, Y. Wang, L. He, Housheng Cai, H. Sun, C. Liu, B. L. Li, B. Bai, J. Wang, Y. Cui, D. Jai, Y. Wang, D. Du, and R. Chen. Organisation of the *Caenorhabditis elegans* small noncoding transcriptome: genomic features, biogenesis and expression. *Genome Res*, 16:30–36, 2006.
56. G. Dieci, G. Fiorino, M. Castelnovo, M. Teichmann, and A. Pagano. The expanding RNA polymerase III transcriptome. *Trends Genet*, 23:614–622, 2007.
57. A. M. Domitrovich and G. R. Kunkel. Multiple, dispersed human U6 small nuclear RNA genes with varied transcriptional efficiencies. *Nucleic Acids Res*, 31:2344–2352, 2003.
58. D. Dulebohn, J. Choy, T. Sundermeier, N. Okan, and A. W. Karzai. Trans-translation: the tmRNA-mediated surveillance mechanism for ribosome rescue, directed protein degradation, and nonstop mRNA decay. *Biochemistry*, 46:4681–4693, 2007.
59. C. Dumas, C. Chow, M. Müller, and B. Papadopoulou. A novel class of developmentally regulated noncoding RNAs in *Leishmania*. *Eukaryotic Cell*, 5:2033–2046, 2006.
60. L. Duret, C. Chureau, S. Samain, J. Weissenbach, and P. Avner. The *Xist* RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science*, 312:1653–1655, 2006.
61. J. B. E. and T. Kiss. A small nucleolar guide RNA functions in both 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. *EMBO*, 20:541–551, 2001.
62. S. Egloff, E. Van Herreweghe, and T. Kiss. Regulation of polymerase II transcription by 7SK snRNA: two distinct RNA elements direct P-TEFb and HEXIM1 binding. *Mol Cell Biol*, 26:630–642, 2006.
63. P. S. Eis, W. Tam, L. Sun, A. Chadburn, Z. Li, M. F. Gomez, E. Lund, and J. E. Dahlberg. Accumulation of miR-155 and BIC RNA in human B cell lymphomas. *Proc Natl Acad Sci U S A*, 102:3627–3632, 2005.
64. C. A. Espinoza, T. A. Allen, A. R. Hieb, J. F. Kugel, and J. A. Goodrich. B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nat Struct Mol Biol*, 11:822–829, 2004.
65. C. A. Espinoza, J. A. Goodrich, and J. F. Kugel. Characterization of the structure, function, and mechanism of B2 RNA, an ncRNA repressor of RNA polymerase II transcription. *RNA*, 13:583–596, 2007.
66. L. F. L. Sonbuchner, S. A. Kyes, C. Epp, and K. W. Deitsch. Nuclear non-coding RNAs are transcribed from the centromeres of *Plasmodium falciparum* and are associated with centromeric chromatin. *J Biol Chem*, 283:5692–5698, 2008.
67. A. Faedo, Q. J. C., P. Stoney, J. E. Long, C. Dye, M. Zollo, J. Rubenstein, D. Price, and A. Bulfone. Identification and characterization of a novel transcript down-regulated in *dlx1/dlx2* and up-regulated in *pax6* mutant telencephalon. *Dev Dyn*, 231:614–620, 2004.
68. A. D. Farris, G. Koelsch, G. J. Pruijn, W. J. van Venrooij, and J. B. Harley. Conserved features of Y RNAs revealed by automated phylogenetic secondary structure analysis. *Nucl Ac Res*, 27:1070–8, 1999.
69. O. Fedorova and N. Zingler. Group II introns: structure, folding and splicing mechanism. *Biol Chem*, 388:665–678, 2007.
70. J. Feng, C. Bi, B. S. Clark, R. Mady, P. Shah, and J. D. Kohtz. The *Evf-2* noncoding RNA is transcribed from the *Dlx-5/6* ultraconserved region and functions as a *Dlx-2* transcriptional coactivator. *Genes Dev*, 20:1470–1484, 2006.
71. A. Force, M. Lynch, F. B. Pickett, A. Amores, Y.-I. Yan, and J. Postlethwait. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, 151:1531–1545, 1999.
72. S. J. Freeland, R. D. Knight, and L. F. Landweber. Do proteins predate DNA? *Science*, 286:690–692, 1999.
73. A. Gabory, M. A. Ripoché, T. Yoshimizu, and L. Dandolo. The H19 gene: regulation and function of a non-coding RNA. *Cytogenet Genome Res*, 113:188–193, 2006.



74. M. I. Galindo, J. I. Pueyo, S. Fouix, S. A. Bishop, and J. P. Couso. Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS Biol*, 5:e106, 2007.
75. P. Ganot, M. Caizergues-Ferrer, and T. Kiss. 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, 1997.
76. R. F. Gesteland and J. F. Atkins, editors. *The RNA World*. Cold Spring Harbor Laboratory Press, Plainview, NY, 1993.
77. W. Gilbert. The RNA world. *Nature*, 319:618, 1986.
78. O. Gimple and A. Schön. *In vitro* and *in vivo* processing of cyanelle tmRNA by RNase P. *Biol Chem*, 382:1421–1429, 2001.
79. I. K. Gogolevskaya, A. P. Koval, and D. A. Kramerov. Evolutionary history of 4.5SH RNA. *Mol Biol Evol*, 22:1546–1554, 2005.
80. S. Gottesman. The small RNA regulators of *Escherichia coli*: roles and mechanisms. *Annu Rev Microbiol*, 58:303–328, 2004.
81. A. Gruber, C. Kilgus, A. Mosig, I. L. Hofacker, W. Hennig, and P. F. Stadler. Arthropod 7SK rna. *Mol Biol Evol*, 2008. In press.
82. A. R. Gruber, D. Koper-Emde, M. Marz, H. Tafer, S. Bernhart, G. Obernosterer, A. Mosig, I. L. Hofacker, P. F. Stadler, and B.-J. Benecke. Invertebrate 7SK snRNAs. *J Mol Evol*, 107-115:66, 2008.
83. P. Gueneau de Novoa and K. P. Williams. The tmRNA website: reductive evolution of tmRNA in plastids and other endosymbionts. *Nucleic Acids Res*, 32:D104–D108, 2004.
84. A. P. Gultyaev and A. Roussis. Identification of conserved secondary structures and expansion segments in enod40 RNAs reveals new enod40 homologues in plants. *Nucleic Acids Res*, 35:3144–3152, 2007.
85. H.-C. Gürsoy, D. Koper, and B.-J. Benecke. The vertebrate 7S K RNA separates hagfish (*Myxine glutinosa*) and lamprey (*Lampetra fluviatilis*). *J Mol Evol*, 50:456–464, 2000.
86. S. Haider, R. Matsumoto, N. Kurosawa, K. Wakui, Y. Fukushima, and M. Isobe. Molecular characterization of a novel translocation t(5;14)(q21;q32) in a patient with congenital abnormalities. *J Hum Genet*, 51:335–340, 2006.
87. J. Han, D. Kim, and K. V. Morris. Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cell. *Proc Natl Acad Sci USA*, 104:12422–12427, 2007.
88. O. Harismendy, C. G. Gendrel, P. Soularue, X. Gidrol, A. Sentenac, M. Werner, and O. Lefebvre. Genome-wide location of yeast RNA polymerase III transcription machinery. *EMBO J*, 22:4738–4747, 2003.
89. J. Häslner and K. Strub. Alu elements as regulators of gene expression. *Nucleic Acids Res*, 34:5491–5497, 2006.
90. K. E. Hastings. SL trans-splicing: easy come or easy go? *Trends Genet*, 21:240–247, 2005.
91. D. M. Haugen P, Simon and B. D. The natural history of group I introns. *Trends Genet*, 21:111–119, 2005.
92. M. Havalio, E. Y. Levanon, G. Lerman, M. Kupiec, and E. Eisenberg. Evidence for abundant transcription of non-coding regions in the sac charomyces cerevisiae genome. *BMC Genomics*, 6:93, 2005.
93. H. He, J. Wang, T. Liu, X. S. Liu, T. Li, Y. Wang, Z. Qian, H. Zheng, X. Zhu, T. Wu, B. Shi, W. Deng, W. Zhou, G. Skogerbø, and R. Chen. Mapping the *C. elegans* noncoding transcriptome with a whole-genome tiling microarray. *Genome Res*, 17:1471–1477, 2007.
94. S. He, H. Su, C. Liu, G. Skogerbø, H. He, D. He, X. Zhu, T. Liu, Y. Zhao, and R. Chen. MicroRNA-encoding long non-coding RNAs. *BMC Genomics*, 21:236, 2008.

95. J. Hertel, I. L. Hofacker, and P. F. Stadler. snoReport: Computational identification of snoRNAs with unknown targets. *Bioinformatics*, 24:158–164, 2008.
96. A. Hinas and F. Söderbom. Treasure hunt in an amoeba: non-coding RNAs in *Dictyostelium discoideum*. *Curr Genet*, 51:141–159, 2007.
97. T. Hirose and J. A. Steitz. 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, 2001.
98. Y. Hirose and F. Harada. Mouse nucleolin binds to 4.5S RNAh, a small noncoding RNA. *Biochem Biophys Res Commun*, 365:62–68, 2008.
99. J. R. Hogg and K. Collins. Human Y5 rna specializes a Ro ribonucleoprotein for 5S ribosomal RNA quality control. *Genes Dev*, 21:3067–3072, 2007.
100. S. Horike, K. Mitsuya, M. Meguro, N. Kotobuki, A. Kashiwagi, T. Notsu, T. C. Schulz, Y. Shirayoshi, and M. Oshimura. Targeted disruption of the human LIT1 locus defines a putative imprinting control element playing an essential role in Beckwith-Wiedemann syndrome. *Hum Mol Genet*, 9:2075–2083, 2000.
101. J. N. Hutchinson, A. W. Ensminger, C. M. Clemson, C. R. Lynch, J. B. Lawrence, and A. Chess. A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics*, 8:39, 2007.
102. A. Hüttenhofer, M. Kiefmann, S. Meier-Ewert, J. O'Brien, H. Lehrach, J. P. Bachellerie, and J. Brosius. RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J*, 20:2943–2953, 2001.
103. S. Inagaki, K. Numata, T. Kondo, M. Tomita, K. Yasuda, A. Kanai, and Y. Kageyama. Identification and expression analysis of putative mRNA-like non-coding RNA in *Drosophila*. *Genes Cells*, 10:1163–1173, 2005.
104. N. Ishii, K. Ozaki, H. Sato, H. Mizuno, S. Saito, A. Takahashi, Y. Miyamoto, S. Ikegawa, N. Kamatani, M. Hori, S. Saito, Y. Nakamura, and T. Tanaka. Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction. *J Hum Genet*, 51:1087–1099, 2006.
105. Y. Isogai, S. Takada, R. Tjian, and S. Keles. Novel TRF1/BRF target genes revealed by genome-wide analysis of *Drosophila* Pol III transcription. *EMBO J*, 26:79–89, 2007.
106. Y. Jacob, E. Seif, P.-O. Paquet, and B. F. Lang. Loss of the mRNA-like region in mitochondrial tmRNAs of jakobids. *RNA*, 10:605–614, 2004.
107. Y. Jacob, S. M. Sharkady, K. Bhardwaj, A. Sanda, and K. P. Williams. Function of the SmpB tail in transfer-messenger RNA translation revealed by a nucleus-encoded form. *J Biol Chem*, 280:5503–5509, 2005.
108. B. E. Jány, E. Bertrand, and T. Kiss. Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body specific localization signal. *J Cell Biol*, 164:647–652, 2004.
109. D. Jia, L. Cai, H. He, G. Skogerbø, T. Li, M. N. Aftab, and R. Chen. Systematic identification of non-coding RNA 2,2,7-trimethylguanosine cap structures in *Caenorhabditis elegans*. *BMC Mol Biol*, 8:86, 2007.
110. C. Jöchl, M. Rederstorff, J. Hertel, P. F. Stadler, I. L. Hofacker, M. Schrettl, H. Haas, and A. Hüttenhofer. Small ncRNA transcriptome analysis from *Aspergillus fumigatus* suggests a novel mechanism for regulation of protein-synthesis. *Nucleic Acids Res*, 36:2677–2689, 2008.
111. C. Jolly and S. C. Lakhota. Human sat III and *Drosophila hsromega* transcripts: a common paradigm for regulation of nuclear RNA processing in stressed cells. *Nucleic Acids Res*, 34:5508–5514, 2006.
112. R. Kachouri, V. Stribinskis, Y. Zhu, K. S. Ramos, E. Westhof, and Y. Li. A surprisingly large RNase P RNA in *Candida glabrata*. *RNA*, 11:1064–1072, 2005.

113. P. Kapranov, J. Cheng, S. Dike, D. Nix, R. Duttagupta, A. T. Willingham, P. F. Stadler, J. Hertel, J. Hackermüller, I. L. Hofacker, I. Bell, E. Cheung, J. Drenkow, E. Dumais, S. Patel, G. Helt, G. Madhavan, A. Piccolboni, V. Sementchenko, H. Tammana, and T. R. Gingeras. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*, 316:1484–1488, 2007.
114. S. Katayama, Y. Tomaru, T. Kasukawa, K. Waki, M. Nakanishi, M. Nakamura, H. Nishida, C. C. Yap, M. Suzuki, J. Kawai, H. Suzuki, P. Carninci, Y. Hayashizaki, C. Wells, M. Frith, T. Ravasi, K. C. Pang, J. Hallinan, J. Mattick, D. A. Hume, L. Lipovich, S. Batalov, P. G. Engström, Y. Mizuno, M. A. Faghihi, A. Sandelin, A. Chalk, S. Mottagui-Tabar, Z. Liang, B. Lenhard, C. Wahlestedt, and RIKEN Genome Exploration Research Group; Genome Science Group (Genome Network Project Core Group); FANTOM Consortium. Antisense transcription in the mammalian transcriptome. *Science*, 309:1564–1566, 2005.
115. H. Kawaji, M. Nakamura, Y. Takahashi, A. Sandelin, S. Katayama, S. Fukuda, C. O. Daub, C. Kai, J. Kawai, J. Yasuda, P. Carninci, and Y. Hayashizaki. Hidden layers of human small rnas. *BMC Genomics*, 9:157, 2008.
116. M. D. Kazanov, A. G. Vitreschak, and M. S. Gelfand. Abundance and functional diversity of riboswitches in microbial communities. *BMC Genomics*, 8:347, 2007.
117. T. Khanam, T. S. Rozhdestvensky, M. Bundman, C. R. Galiveti, S. Handel, V. Sukonina, U. Jordan, J. Brosius, and B. V. Skryabin. Two primate-specific small non-protein-coding RNAs in transgenic mice: neuronal expression, subcellular localization and binding partners. *Nucleic Acids Res*, 35:529–539, 2007.
118. K.-s. Kim and Y. Lee. Regulation of 6S RNA biogenesis by switching utilization of both sigma factors and endoribonucleases. *Nucleic Acids Res*, 32:6057–6068, 2004.
119. S. Kishore and S. Stamm. Regulation of alternative splicing by snoRNAs. *Cold Spring Harb Symp Quant Biol*, 71:329–334, 2006.
120. D. J. Klein and A. R. Ferré-D’Amaré. Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science*, 313:1752–1756, 2006.
121. J. Kohtz and G. Fishell. Developmental regulation of EVF-1, a novel non-coding RNA transcribed upstream of the mouse *Dlx6* gene. *Gene Expr Patterns*, 4:407–412, 2004.
122. Y. Komine, M. Kitabatake, T. Yokogawa, K. Nishikawa, and H. Inokuchi. A tRNA-like structure is present in 10Sa RNA, a small stable RNA from *Escherichia coli*. *Proc Natl Acad Sci U S A*, 91:9223–9227, 1994.
123. H. König, N. Matter, R. Bader, W. Thiele, and F. Müller. Splicing segregation: the minor spliceosome acts outside the nucleus and controls cell proliferation. *Cell*, 131:718–729, 2007.
124. J. Kravchenko, I. B. Rogozin, E. V. Koonin, and P. M. Chumakov. Transcription of mammalian mRNAs by a novel nuclear RNA polymerase of mitochondrial origin. *Nature*, 436:735–739, 2005.
125. B. J. Krueger, C. Jeronimo, B. B. Roy, A. Bouchard, C. Barrandon, S. A. Byers, C. E. Searcey, J. J. Cooper, O. Bensaude, É. A. Cohen, B. Colombe, and D. H. Price. LARP7 is a stable component of the 7SK snRNP while P-TEFb, HEXIM1 and hnRNP A1 are reversibly associated. *Nucleic Acids Res*, 2008.
126. V. Y. Kuryshv, B. V. Skryabin, J. Kremerskothen, J. Jurka, and J. Brosius. Birth of a gene: locus of neuronal BC200 snmRNA in three prosimians and human BC200 pseudogenes as archives of change in the Anthropeidea lineage. *J Mol Biol*, 309:1049–66, 2001.
127. K. Y. Kwek, S. Murphy, A. Furger, B. Thomas, W. O’Gorman, H. Kimura, N. J. Proudfoot, and A. Akoulitchev. U1 snRNA associates with TFIIF and regulates transcriptional initiation. *Nat Struct Biol*, 9:800–805, 2002.
128. D. L. Lafontaine and D. Tollervey. Ribosomal rna. *Encyclopedia of life sciences*, 2001.

129. S. G. Landt, E. Abeliuk, P. T. McGrath, J. A. Lesley, H. H. McAdams, and L. Shapiro. Small non-coding RNAs in *Caulobacter crescentus*. *Mol Microbiol*, 2008.
130. R. B. Lanz, N. J. McKenna, S. A. Onate, U. Albrecht, J. Wong, S. Y. Tsai, M. J. Tsai, and B. W. O'Malley. A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell*, 97:17–27, 1999.
131. P. Larsson, A. Hinas, D. H. Ardell, L. Kirsebom, A. Virtanen, and F. Sjölander. *De novo* search for non-coding RNA genes in the AT-rich genome of *Dictyostelium discoideum*: Performance of Markov-dependent genome feature scoring. *Genome Res*, 18:888–899, 2008.
132. J. Leonardi, J. A. Box, J. T. Bunch, and P. Baumann. TER1, the RNA subunit of fission yeast telomerase. *Nat Struct Mol Biol*, 15:26–33, 2008.
133. L. Lestrade and M. J. Weber. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res*, 34:D158–D162, 2008.  
URL <http://www-snoRNA.biotoul.fr/>
134. E. Leygue. Steroid receptor RNA activator (SRA1): unusual bifaceted gene products with suspected relevance to breast cancer. *Nucl Recep Signaling*, 5:e006, 2007.
135. D. Li, D. K. Willkomm, A. Schön, and R. K. Hartmann. Rnase P of the *Cyanophora paradoxa* cyanelle: a plastid ribozyme. *Biochimie*, 89:1528–1538, 2007.
136. J. Li, D. P. Witte, T. V. Dyke, and D. S. Askew. Expression of the putative proto-oncogene His-1 in normal and neoplastic tissues. *Am J Pathol*, 150:1297–1305, 1997.
137. L. Li, X. Wang, R. Sasidharan, V. Stolc, W. Deng, H. He, J. Korbel, X. Chen, W. Tongprasit, P. Ronald, R. Chen, M. Gerstein, and X. Wang Deng. Global identification and characterization of transcriptionally active regions in the rice genome. *PLoS ONE*, 2:e294, 2007.
138. X. Liang, A. Hury, E. Hoze, S. Uliel, I. Myslyuk, A. Apatoff, R. Unger, and S. Michaeli. Genome-wide analysis of C/D and H/ACA-like small nucleolar rnas in *Leishmania major* indicates conservation among trypanosomatids in the repertoire and in their rRNA targets. *Eukaryotic Cell*, 6:361–377, 2006.
139. X.-h. Liang, A. Hury, E. Hoze, S. Uliel, I. Myslyuk, A. Apatoff, R. Unger, and S. Michaeli. Genome-wide analysis of C/D and H/ACA-like small nucleolar RNAs in *Leishmania major* indicates conservation among Trypanosomatids in the repertoire and in their rRNA targets. *Eukaryotic Cell*, 6:361–377, 2007.
140. K. B. Lidie and F. M. van Dolah. Spliced leader RNA-mediated trans-splicing in a dinoflagellate, *Karenia brevis*. *J Eukaryot Microbiol*, 54:427–435, 2007.
141. R. Lin, S. Maeda, C. Liu, M. Karin, and T. S. Edgington. A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas. *Oncogene*, 26:851–858, 2007.
142. L. Liu, H. Ben-Shlomo, Y. X. Hu, M. Z. Stern, I. Goncharov, Y. Zhang, and S. Michaeli. The trypanosomatid signal recognition particle consists of two RNA molecules, a 7SL RNA homolog and a novel tRNA-like molecule. *J Biol Chem*, 278:18271–18280, 2003.
143. Z. J. Lorković, R. Lehner, C. Forstner, and A. Barta. Evolutionary conservation of minor U12-type spliceosome between plants and humans. *RNA*, 11:1095–1107, 2005.
144. R. Louro, T. El-Jundi, H. I. Nakaya, E. M. Reis, and S. Verjovski-Almeida. Conserved tissue expression signatures of intronic noncoding RNAs transcribed from human and mouse loci. *Genomics*, 2008. Doi:10.1016/j.ygeno.2008.03.013.
145. Y. Luo and S. Li. Genome-wide analyses of retrogenes derived from the human box H/ACA snornas. *Nucleic Acids Res*, 35:559–571, 2007.
146. J. Lykke-Andersen, C. Aagaard, M. Semionenkov, and R. A. Garrett. Archaeal introns: splicing, intercellular mobility and evolution. *Trends Biochem Sci*, 22:326–331, 1997.

147. M. MacMorris, M. Kumar, E. Lasda, A. Larsen, B. Kraemer, and T. Blumenthal. A novel family of *C. elegans* snRNPs contains proteins associated with trans-splicing. *RNA*, 13:511–520, 2007.
148. M. J. Madej, J. D. Alfonzo, and A. Hüttenhofer. Small ncRNA transcriptome analysis from kinetoplast mitochondria of *Leishmania tarentolae*. *Nucleic Acids Res*, 35:1544–1554, 2007.
149. M. J. Madej, M. Niemann, A. Hüttenhofer, and H. U. Gäring. Identification of novel guide RNAs from the mitochondria of *Trypanosoma brucei*. *RNA Biol*, 5, 2008. <http://www.landesbioscience.com/journals/rnabiology/article/6043>.
150. H. Maeda, N. Fujita, and A. Ishihama. Competition among seven *Escherichia coli* sigma subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Res*, 28:3497–3503, 2000.
151. N. Maeda, T. Kasukawa, R. Oyama, J. Gough, M. Frith, P. G. Engström, B. Lenhard, R. N. Aturaliya, S. Batalov, K. W. Beisel, C. J. Bult, C. F. Fletcher, A. R. Forrest, M. Furuno, D. Hill, M. Itoh, M. Kanamori-Katayama, S. Katayama, M. Katoh, T. Kawashima, J. Quackenbush, T. Ravasi, B. Z. Ring, K. Shibata, K. Sugiura, Y. Takenaka, R. D. Teasdale, C. A. Wells, Y. Zhu, C. Kai, J. Kawai, D. A. Hume, P. Carninci, and Y. Hayashizaki. Transcript annotation in FANTOM3: Mouse gene catalog based on physical cdnas. *PLoS Genetics*, 2:e62, 2006. Doi:10.1371/journal.pgen.0020062.
152. J. R. Manak, S. Dike, V. Sementchenko, P. Kapranov, F. Biemar, J. Long, J. Cheng, I. Bell, S. Ghosh, A. Piccolboni, and T. R. Gingeras. Biological function of unannotated transcription during the early development of *Drosophila melanogaster*. *Nat Genet*, 38:1151–1158, 2006.
153. P. D. Mariner, R. D. Walters, C. A. Espinoza, L. F. Drullinger, S. D. Wagner, J. F. Kugel, and J. A. Goodrich. Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol Cell*, 29:499–509, 2008.
154. P. A. Maroney, M. Yu, Y. T. Jankowska, and T. W. Nilsen. Direct analysis of nematode cis- and trans-spliceosomes: a functional role for U5 snrna in spliced leader addition trans-splicing and the identification of novel Sm snRNPs. *RNA*, 2:735–745, 1996.
155. S. M. Marquez, J. K. Harris, S. T. Kelley, J. W. Brown, S. C. Dawson, E. C. Roberts, and N. R. Pace. Structural implications of novel diversity in eucaryal RNase P RNA. *RNA*, 11:739–751, 2005.
156. M. Marz, T. Kirsten, and P. F. Stadler. Evolution of spliceosomal snrna genes in metazoan animals. 2008. Submitted.
157. M. Marz, A. Mosig, B. M. R. Stadler, and P. F. Stadler. U7 snRNAs: A computational survey. *Geno Prot Bioinf*, 5:187–195, 2007.
158. W. F. Marzluff. Metazoan replication-dependent histone mRNAs: a distinct set of RNA polymerase II transcripts. *Curr Opin Cell Biol*, 17:274–280, 2005.
159. I. J. Matouk, N. DeGroot, S. Mezan, S. Ayes, R. Abu-lail, A. Hochberg, and E. Galun. The H19 non-coding RNA is essential for human tumor growth. *PLoS ONE*, 2:e845, 2007.
160. C. Mayer, K. M. Schmitz, J. Li, I. Grummt, and R. Santoro. Intergenic transcripts regulate the epigenetic state of rRNA genes. *Mol Cell*, 22:351–361, 2006.
161. T. R. Mercer, M. E. Dinger, S. M. Sunkin, M. F. Mehler, and J. S. Mattick. Specific expression of long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci USA*, 105:716–721, 2008.
162. M. M. Meyer, A. Roth, S. M. Chervin, G. A. Garcia, and R. R. Breaker. Confirmation of a second natural preQ1 aptamer class in Streptococcaceae bacteria. *RNA*, 14:685–695, 2008.
163. J. K. Millar, R. James, N. J. Brandon, and P. A. Thomson. DISC1 and DISC2: discovering and dissecting molecular mechanisms underlying psychiatric illness. *Ann Med*, 36:367–378, 2004.
164. K. Missal, D. Rose, and P. F. Stadler. Non-coding RNAs in *Ciona intestinalis*. *Bioinformatics*, 21 S2:i77–i78, 2005.

165. J. R. Mitchell, J. Cheng, and C. K. A box H/ACA small nucleolar RNA-like domain at the human telomerase 3' end. *Mol Cell Biol*, 19:567–576, 1999.
166. F. Miura, N. Kawaguchi, J. Sese, A. Toyoda, M. Hattori, S. Morishita, and T. Ito. A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. *Proc Natl Acad Sci USA*, 103:17846–17851, 2006.
167. K. Mochizuki, N. A. Fine, T. Fujisawa, and M. A. Gorovsky. Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. *Cell*, 110:689–699, 2002.
168. P. B. Moore and T. A. Steitz. The involvement of RNA in ribosome function. *Nature*, 418:229–235, 2002.
169. S. D. Moore and R. T. Sauer. Ribosome rescue: *tmRNA* tagging activity and capacity in *Escherichia coli*. *Mol Microbiol*, 58:456–466, 2005.
170. S. D. Moore and R. T. Sauer. The tmRNA system for translational surveillance and ribosome rescue. *Annu Rev Biochem*, 76:101–124, 2007.
171. A. Mosig, J. L. Chen, and P. F. Stadler. Homology search with fragmented nucleic acid sequence patterns. In R. Giancarlo and S. Hannenhalli, editors, *Algorithms in Bioinformatics (WABI 2007)*, volume 4645 of *Lecture Notes in Computer Science*, 335–345. Springer Verlag, Berlin, Heidelberg, 2007.
172. A. Mosig, M. Guofeng, B. M. R. Stadler, and P. F. Stadler. Evolution of the vertebrate Y RNA cluster. *Th Biosci*, 126:9–14, 2007.
173. A. Mosig, K. Sameith, and P. F. Stadler. *fragrep*: Efficient search for fragmented patterns in genomic sequences. *Geno Prot Bioinfo*, 4:56–60, 2005.
174. T. Mourier, C. Carret, S. Kyes, Z. Christodoulou, P. P. Gardner, D. C. Jeffares, R. Pinches, B. Barrell, M. Berriman, S. Griffiths-Jones, A. Ivens, C. Newbold, and A. Pain. Genome-wide discovery and verification of novel structured RNAs in *Plasmodium falciparum*. *Genome Res*, 18:281–292, 2008.
175. M. Mutsuddi, C. M. Marshall, K. A. Benzow, M. D. Koob, and I. Rebay. The spinocerebellar ataxia 8 noncoding RNA causes neurodegeneration and associates with staufen in *Drosophila*. *Curr Biol*, 14:302–308, 2004.
176. T. Nakamura, K. Naito, N. Yokota, C. Sugita, and M. Sugita. A cyanobacterial non-coding RNA, Yfr1, is required for growth under multiple stress conditions. *Plant Cell Physiol*, 48:1309–1318, 2007.
177. H. I. Nakaya, R. Amaral, P. P. Louro, A. Lopes, A. A. Fachel, Y. B. Moreira, T. A. El-Jundi, A. M. da Silva, E. M. Reis, and S. Verjovski-Almeida. Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. *Genome Biol*, 8:R43, 2007.
178. T. Neusser, N. Gildehaus, R. Wurm, and R. Wagner. Studies on the expression of 6S RNA from *E. coli*: involvement of regulators important for stress and growth adaptation. *Biol Chem*, 389:285–297, 2008.
179. K. Ng, D. Pullirsch, M. Leeb, and A. Wutz. *Xist* and the order of silencing. *EMBO Rep*, 8:34–39, 2007.
180. E. L. Niemitz, M. R. DeBaun, J. Fallon, K. Murakami, H. Kugoh, M. Oshimura, and A. P. Feinberg. Microdeletion of LIT1 in familial Beckwith-Wiedemann syndrome. *Am J Hum Genet*, 75:844–849, 2004.
181. T. W. Nilsen. The spliceosome: the most complex macromolecular machine in the cell? *Bioessays*, 25:1147–1149, 2003.
182. N. A. Okan, J. Bliska, and A. W. Karzai. A role for the SmpB-SsrA system in *Yersinia pseudotuberculosis* pathogenesis. *PLoS Pathog*, 2:e6, 2006.



183. T. Okutsu, Y. Kuroiwa, F. Kagitani, M. Kai, K. Aisaka, O. Tsutsumi, Y. Kaneko, K. Yokomori, M. A. Surani, T. Kohda, T. Kaneko-Ishino, and F. Ishino. Expression and imprinting status of human PEG8/IGF2AS, a paternally expressed antisense transcript from the IGF2 locus, in Wilms' tumors. *J Biochem*, 127:475–483, 2000.
184. A. Pagano, M. Castelnovo, F. Tortelli, R. Ferrari, G. Dieci, and C. R. New small nuclear RNA gene-like transcriptional units as sources of regulatory transcripts. *PLoS Genet*, 3:e1, 2007.
185. Z. Palfi, B. Schimanski, A. Günzl, S. Lücke, and A. Bindereif. U1 small nuclear RNP from *Trypanosoma brucei*: a minimal u1 snrna with unusual protein components. *Nucleic Acids Res*, 33:2493–2503, 2005.
186. K. Panzitt, M. M. Tschernatsch, C. Guelly, T. Moustafa, M. Stradner, H. M. Strohmaier, C. R. Buck, H. Denk, R. Schroeder, M. Trauner, and K. Zatloukal. Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA. *Gastroenterology*, 132:330–342, 2007.
187. A. M. Parrott and M. B. Mathews. Novel rapidly evolving hominid RNAs bind nuclear factor 90 and display tissue-restricted distribution. *Nucleic Acids Res*, 35:6249–6258, 2007.
188. A. A. Patel and J. A. Steitz. Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol*, 4:960–970, 2003.
189. J. Perreault, J.-F. Noël, F. Brière, B. Cousineau, J.-F. Lucier, J.-P. Perreault, and G. Boire. Retroseudogenes derived from human Ro/SS-A autoantigen-associated hY RNAs. *Nucl Acids Res*, 33:2032–2041, 2005.
190. J. Perreault, J.-P. Perreault, and G. Boire. The Ro associated Y RNAs in metazoans: evolution and diversification. *Mol Biol Evol*, 24:1678–1689, 2007.
191. J. Pettitt, B. Müller, I. Stansfield, and B. Connolly. Spliced leader trans-splicing in the nematode *Trichinella spiralis* uses highly polymorphic, noncanonical spliced leaders. *RNA*, 14:760–770, 2008.
192. L. Pibouin, J. Villaudy, D. Ferbus, M. Muleris, M.-T. Prospéri, Y. Remvikos, and G. Goubin. Cloning of the mRNA of overexpression in colon carcinoma-1: a sequence overexpressed in a subset of colon carcinomas. *Cancer Genet Cytogenet*, 133:55–60, 2002.
193. P. Piccinelli, M. A. Rosenblad, and T. Samuelsson. Identification and analysis of ribonuclease P and MRP RNA in a broad range of eukaryotes. *Nucleic Acids Res*, 33:4485–4495, 2005.
194. V. Pirrotta. Trans-splicing in *Drosophila*. *Bioessays*, 24:988–991, 2002.
195. A. Pizzuti, G. Novelli, A. Ratti, F. Amati, R. Bordoni, P. Mandich, E. Bellone, E. Conti, M. Bengala, A. Mari, V. Silani, and B. Dallapiccola. Isolation and characterization of a novel transcript embedded within HIRA, a gene deleted in DiGeorge syndrome. *Mol Genet Metab*, 67:227–235, 1999.
196. J. D. Podlevsky, C. J. Bley, R. V. Omana, X. Qi, and J. J. Chen. The telomerase database. *Nucleic Acids Res*, 36:D339–D343, 2008.
197. O. O. Poleskaya, V. Haroutunian, K. L. Davis, I. Hernandez, and B. P. Sokolov. Novel putative nonprotein-coding RNA gene from 11q14 displays decreased expression in brains of patients with schizophrenia. *J Neurosci Res*, 74:111–122, 2003.
198. N. N. Pouchkina-Stantcheva and A. Tunnacliffe. Spliced leader RNA-mediated trans-splicing in phylum *Rotifera*. *Mol Biol Evol*, 22:1482–1489, 2005.
199. K. V. Prasanth and D. L. Spector. Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. *Genes Dev*, 21:11–42, 2007.
200. E. Puerta-Fernandez, J. E. Barrick, A. Roth, and R. R. Breaker. Identification of a large non-coding RNA in extremophilic eubacteria. *Proc Natl Acad Sci U S A*, 103:19490–19495, 2006.
201. L. Randau, I. Schröder, and D. Söll. Life without RNase P. *Nature*, 453:120–123, 2008.



202. T. Ravasi, H. Suzuki, K. C. Pang, S. Katayama, M. Furuno, R. Okunishi, S. Fukuda, K. Ru, M. C. Frith, M. M. Gongora, S. M. Grimmond, D. A. Hume, Y. Hayashizaki, and J. S. Mattick. Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res*, 16:11–19, 2006.
203. E. E. Regulski and R. R. Breaker. In-line probing analysis of riboswitches. *Methods Mol Biol*, 419:53–67, 2008.
204. S. L. Reichow, T. Hamma, A. R. Ferré-D’Amaré, and G. Varani. The structure and function of small nucleolar ribonucleoproteins. *Nucleic Acids Res*, 35:1452–1464, 2007.
205. S. Riccardo, G. Tortoriello, E. Giordano, M. Turano, and F. M. The coding/non-coding overlapping architecture of the gene encoding the *Drosophila* pseudouridine synthase. *BMC Mol Biol*, 8:15, 2007.
206. P. Richard, M. A. Kiss, X. Darzacq, and T. Kiss. Cotranscriptional recognition of human intronic box H/ACA snoRNAs occurs in a splicing-independent manner. *Mol Cell Biol*, 26:2540–2549, 2006.
207. T. Roberts, O. Chernova, and J. K. Cowell. NB4S, a member of the TBC1 domain family of genes, is truncated as a result of a constitutional t(1;10)(p22;q21) chromosome translocation in a patient with stage 4S neuroblastoma. *Hum Mol Genet*, 7:1169–1178, 1998.
208. D. Rose, J. Jöris, J. Hackermüller, K. Reiche, Q. Li, and P. F. Stadler. Duplicated RNA genes in teleost fish genomes. *J Bioinf Comp Biol*, 2008. Submitted.
209. D. R. Rose, J. Hackermüller, S. Washietl, S. Findeiß, K. Reiche, J. Hertel, P. F. Stadler, and S. J. Prohaska. Computational RNomics of drosophilids. *BMC Genomics*, 8:406, 2007.
210. T. S. Rozhdstvensky, A. M. Kopylov, J. Brosius, and A. Hüttenhofer. Neuronal BC1 RNA structure: evolutionary conversion of a tRNA(Ala) domain into an extended stem-loop structure. *RNA*, 7:722–730, 2001.
211. A. G. Russell, J. M. Charette, D. F. Spencer, and M. W. Gray. An early evolutionary origin for the minor spliceosome. *Nature*, 443:863–866, 2006.
212. L. A. Rymarquis, J. P. Kastenmayer, A. G. Hüttenhofer, and P. J. Green. Diamonds in the rough: mRNA-like non-coding rnas. *Trends Plant Sci*, 2008. Doi:10.1016/j.tplants.2008.02.009.
213. D. A. Samarsky, M. J. Fournier, R. H. Singer, and E. Bertrand. The snoRNA box C/D motif directs nucleolar targeting and also couples snoRNA synthesis and localization. *EMBO*, 17:3747–3757, 1998.
214. A. Saxena, T. Lahav, N. Holland, G. Aggarwal, A. Anupama, Y. Huang, H. Volpin, P. J. Myler, and D. Zilberstein. Analysis of the *Leishmania donovani* transcriptome reveals an ordered progression of transient and permanent changes in gene expression during differentiation. *Mol Biochem Parasitol*, 152:53–65, 2007.
215. J. Schmitz, A. Zemann, G. Churakov, H. Kuhl, F. Grützner, R. Reinhardt, and J. Brosius. Retroposed SNOfall a mammalian-wide comparison of platypus snoRNAs. *Genome Res*, 18:1005–1010, 2008.
216. S. Schoeftner and M. A. Blasco. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nature Cell Biol*, 10:228–236, 2008.
217. E. R. Seif, L. Forget, N. C. Martin, and B. F. Lang. Mitochondrial rnae p rnas in ascomycete fungi: lineage-specific variations in RNA secondary structure. *RNA*, 9:1073–1083, 2003.
218. A. Serganov and D. J. Patel. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat Rev Genet*, 8:776–790, 2007.
219. S. M. Sharkady and K. P. Williams. A third lineage with two-piece tmrna. *Nucleic Acids Res*, 32:4531–4538, 2004.

220. C. M. Sharma, F. Darfeuille, T. H. Plantinga, and J. Vogel. A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev*, 21:2804–2817, 2007.
221. N. Sheth, X. Roca, M. L. Hastings, T. Roeder, A. R. Krainer, and R. Sachidanandam. Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res*, 34:3955–3967, 2006.
222. A. P. M. Silva, A. C. M. Salim, A. Bulgarelli, E. de Souza, Jorge Estefano S Osório, O. L. Caballero, C. Iseli, B. J. Stevenson, C. V. Jongeneel, S. J. de Souza, A. J. G. Simpson, and A. A. Camargo. Identification of 9 novel transcripts and two RGSL genes within the hereditary prostate cancer region (HPC1) at 1q25. *Gene*, 310:49–57, 2003.
223. M. Sone, T. Hayashi, H. Tarui, K. Agata, M. Takeichi, and S. Nakagawa. The mrna-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons. *J Cell Sci*, 120:2498–2506, 2007.
224. C. Sousa, C. Johansson, C. Charon, H. Manyani, C. Sautter, A. Kondorosi, and M. Crespi. Translational and structural requirements of the early nodulin gene *enod40*, a short-open reading frame-containing RNA, for elicitation of a cell-specific growth response in the alfalfa root cortex. *Mol Cell Biol*, 21:354–366, 2001.
225. V. Srikantan, Z. Zou, G. Petrovics, L. Xu, M. Augustus, L. Davis, J. R. Livezey, T. Connell, I. A. Sesterhenn, K. Yoshino, G. S. Buzard, F. K. Mostofi, D. G. McLeod, J. W. Moul, and S. Srivastava. PCGEM1, a prostate-specific gene, is overexpressed in prostate cancer. *Proc Natl Acad Sci U S A*, 97:12216–12221, 2000.
226. G. Storz, J. A. Opdyke, and K. M. Wassarman. Regulating bacterial transcription with small RNAs. *Cold Spring Harb Symp Quant Biol*, 71:269–273, 2006.
227. S. A. Strobel and J. C. Cochrane. Rna catalysis: ribozymes, ribosomes, and riboswitches. *Curr Opin Chem Biol*, 11:636–643, 2007.
228. N. Sudarsan, J. E. Barrick, and R. R. Breaker. Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA*, 9:644–647, 2003.
229. N. Sudarsan, M. C. Hammond, K. F. Block, R. Welz, J. E. Barrick, A. Roth, and R. R. Breaker. Tandem riboswitch architectures exhibit complex gene control functions. *Science*, 314:300–304, 2006.
230. H. F. Sutherland, R. Wadey, J. M. McKie, C. Taylor, U. Atif, K. A. Johnstone, S. Halford, U. J. Kim, J. Goodship, A. Baldini, and P. J. Scambler. Identification of a novel transcript disrupted by a balanced translocation associated with DiGeorge syndrome. *Am J Hum Genet*, 59:23–31, 1996.
231. M. Széll, Z. Bata-Csörgo, and L. Kemény. The enigmatic world of mRNA-like ncRNAs: their role in human evolution and in human diseases. *Semin Cancer Biol*, 18:141–148, 2008.
232. M. Szymanski, M. Z. Barciszewska, V. A. Erdmann, and J. Barciszewski. A new frontier for molecular medicine: noncoding RNAs. *Biochim Biophys Acta*, 1756:65–75, 2005.
233. W. Tam and J. E. Dahlberg. miR-155/BIC as an oncogenic microRNA. *Genes Chromosomes Cancer*, 45:211–212, 2006.
234. R. Tanaka, H. Satoh, M. Moriyama, K. Satoh, Y. Morishita, S. Yoshida, T. Watanabe, Y. Nakamura, and S. Mori. Intronic U50 small-nucleolar-RNA (snoRNA) host gene of no protein-coding potential is mapped at the chromosome breakpoint t(3;6)(q27;q15), of human B-cell lymphoma. *Genes Cells*, 5:277–287, 2000.
235. T.-H. Tang, N. Polacek, M. Zywicki, H. Huber, K. Brugger, R. Garrett, J. P. Bachelierie, and A. Hüttenhofer. Identification of novel non-coding rnas as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. *Mol Microbiol*, 55:469–481, 2005.

236. T. H. Tang, T. S. Rozhdestvensky, B. C. d'Orval, M. L. Bortolin, H. Huber, B. Charpentier, C. Branlant, J. P. Bachellerie, J. Brosius, and A. Hüttenhofer. RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing. *Nucleic Acids Res*, 30:921–930, 2002.
237. M. P. Terns and R. M. Terns. Small nucleolar RNAs: Versatile trans-acting molecules of ancient evolutionary origin. *Gene Expr*, 10:17–39, 2002.
238. S. W. M. Teunissen, M. J. M. Kruithof, A. D. Farris, J. B. Harley, W. J. van Venrooij, and G. J. M. Pruijn. Conserved features of Y RNAs: a comparison of experimentally derived secondary structures. *Nucl Acids Res*, 28:610–619, 2000.
239. The ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447:799–816, 2007.
240. S. Thomas, L. I. T. Martinez, S. J. Westenberger, and N. R. Sturm. A population study of the minicircles in *Trypanosoma cruzi*: predicting guide RNAs in the absence of empirical rna editing. *BMC Genomics*, 8:133, 2007.
241. A. Toledo-Arana, F. Repoila, and P. Cossart. Small noncoding RNAs controlling pathogenesis. *Curr Opin Microbiol*, 10:182–188, 2007.
242. E. Torarinsson, M. Sawera, J. Havgaard, M. Fredholm, and J. Gorodkin. Thousands of corresponding human and mouse genomic regions unalignable in primary sequence contain common RNA structure. *Genome Res*, 16:885–889, 2006. Erratum *Genome Res* 16:1439 (2006).
243. A. E. Trotochaud and K. M. Wassarman. 6S RNA regulation of pspF transcription leads to altered cell survival at high pH. *J Bacteriol*, 188:3936–3943, 2006.
244. K. T. Tycowski and J. A. Steitz. Non-coding snoRNA host genes in *Drosophila*: expression strategies for modification guide snoRNAs. *Eur J Cell Biol*, 80:119–125, 2001.
245. Y. Tzfati, Z. Knight, J. Roy, and E. H. Blackburn. A novel pseudoknot element is essential for the action of a yeast telomerase. *Genes & Dev*, 17:1779–1788, 2003.
246. N. B. Ulyanov, K. Shefer, T. L. James, and Y. Tzfati. Pseudoknot structures with conserved base triples in telomerase RNAs of ciliates. *Nucleic Acids Res*, 35:6150–6160, 2007.
247. A. V. Uzilov, J. M. Keegan, and D. H. Mathews. Detection of non-coding RNAs on the basis of predicted secondary structure formation free energy change. *BMC Bioinformatics*, 7:173, 2006.
248. S. Valadkhan, A. Mohammadi, C. Wachtel, and J. L. Manley. Protein-free spliceosomal snRNAs catalyze a reaction that resembles the first step of splicing. *RNA*, 13:2300–2311, 2007.
249. D. J. van Horn, D. Eisenberg, C. A. O'Brien, and S. L. Wolin. *Caenorhabditis elegans* embryos contain only one major species of Ro RNP. *RNA*, 1:293–303, 1995.
250. A. van Zon, M. H. Mossink, R. J. Scheper, P. Sonneveld, and E. A. C. Wiemer. The vault complex. *Cell Mol Life Sci*, 60:1828–1837, 2003.
251. S. Vincenti, V. D. Chiara, I. Bozzoni, and C. Presutti. The position of yeast snorna-coding regions within host introns is essential for their biosynthesis and for efficient splicing of the host pre-mRNA. *RNA*, 13:138–150, 2007.
252. J. Vogel and C. M. Sharma. How to find small non-coding RNAs in bacteria. *Biol Chem*, 386:1219–1238, 2005.
253. C. S. Wadler and C. K. Vanderpool. A dual function for a bacterial small RNA: *SgrS* performs base pairing-dependent regulation and encodes a functional polypeptide. *Proc Natl Acad Sci USA*, 104:20454–20459, 2007.
254. S. C. Walker and D. R. Engelke. Ribonuclease P: the evolution of an ancient RNA enzyme. *Crit Rev Biochem Mol Biol*, 41:77–102, 2006.
255. J. X. Wang and R. R. Breaker. Riboswitches that sense S-adenosylmethionine and S-adenosylhomocysteine. *Biochem Cell Biol*, 86:157–168, 2008.

256. S. Washietl, I. L. Hofacker, M. Lukasser, A. Hüttenhofer, and P. F. Stadler. Mapping of conserved RNA secondary structures predicts thousands of functional non-coding RNAs in the human genome. *Nature Biotech*, 23:1383–1390, 2005.
257. S. Washietl, J. S. Pedersen, J. O. Korb, A. Gruber, J. Hackermüller, J. Hertel, M. Lindemeyer, K. Reiche, C. Stocsits, A. Tanzer, C. Ucla, C. Wyss, S. E. Antonarakis, F. Denoeud, J. Lagarde, J. Drenkow, P. Kapranov, T. R. Gingeras, R. Guigó, M. Snyder, M. B. Gerstein, A. Reymond, I. L. Hofacker, and P. F. Stadler. Structured RNAs in the ENCODE selected regions of the human genome. *Gen Res*, 17:852–864, 2007.
258. D. A. Wassarman and J. A. Steitz. Structural analyses of the 7SK ribonucleoprotein (RNP), the most abundant human small RNP of unknown function. *Mol Cell Biol*, 11:3432–3445, 1991.
259. K. M. Wassarman. 6S RNA: a small RNA regulator of transcription. *Curr Opin Microbiol*, 10:164–168, 2007.
260. K. M. Wassarman and R. M. Saecker. Synthesis-mediated release of a small RNA inhibitor of RNA polymerase. *Science*, 314:1601–1603, 2006.
261. K. M. Wassarman and G. Storz. 6S RNA regulates *E. coli* RNA polymerase activity. *Cell*, 101:613–623, 2000.
262. C. J. Webb and V. A. Zakian. Identification and characterization of the *Schizosaccharomyces pombe* TER1 telomerase RNA. *Nat Struct Mol Biol*, 15:34–42, 2008.
263. L. B. Weinstein and J. A. Steitz. Guided tours: from precursor snoRNA to functional snoRNP. *Curr Opin Cell Biol*, 11:378–384, 1999.
264. J. Wen, B. J. Parker, and G. F. Weiller. *In Silico* identification and characterization of mRNA-like noncoding transcripts in *Medicago truncatula*. *In Silico Biol*, 7:485–505, 2007.
265. R. Wevrick, J. A. Kerns, and U. Francke. Identification of a novel paternally expressed gene in the Prader-Willi syndrome region. *Hum Mol Genet*, 3:1877–1882, 1994.
266. B. T. Wilhelm, S. Marguerat, S. Watt, F. Schubert, V. Wood, I. Goodhead, C. J. Penkett, J. Rogers, and J. Bähler. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature*, 2008. Doi:10.1038/nature07002.
267. C. L. Will and R. Lüthmann. Splicing of a rare class of introns by the U12-dependent spliceosome. *Biol Chem*, 386:713–724, 2005.
268. S. Will, K. Missal, I. L. Hofacker, P. F. Stadler, and R. Backofen. Inferring non-coding RNA families and classes by means of genome-scale structure-based clustering. *PLoS Comp Biol*, 3:e65, 2007.
269. D. K. Willkomm and R. K. Hartmann. 6S RNA — an ancient regulator of bacterial RNA polymerase rediscovered. *Biol Chem*, 386:1273–1277, 2005.
270. D. K. Willkomm and R. K. Hartmann. An important piece of the rna p jigsaw solved. *Trends Biochem Sci*, 32:247–250, 2007.
271. S. Wolf, D. Mertens, C. Schaffner, C. Korz, H. Döhner, S. Stilgenbauer, and P. Lichter. B-cell neoplasia associated gene with multiple splicing (BCMS): the candidate B-CLL gene on 13q14 comprises more than 560 kb covering all critical regions. *Hum Mol Genet*, 10:1275–1285, 2001.
272. M. D. Woodhams, P. F. Stadler, D. Penny, and L. J. Collins. RNase MRP and the RNA processing cascade in the eukaryotic ancestor. *BMC Evol Biol*, 7:S13, 2007.
273. A. Wutz. RNAs templating chromatin structure for dosage compensation in animals. *Bioessays*, 25:434–442, 2003.
274. M. Xie, A. Mosig, X. Qi, Y. Li, P. F. Stadler, and J. J.-L. Chen. Size variation and structural conservation of vertebrate telomerase RNA. *J Biol Chem*, 283:2049–2059, 2008.
275. K. Yamada, J. Kano, H. Tsunoda, H. Yoshikawa, C. Okubo, T. Ishiyama, and M. Noguchi. Phenotypic characterization of endometrial stromal sarcoma of the uterus. *Cancer Sci*, 97:106–112, 2006.

276. C.-Y. Yang, H. Zhou, J. Luo, and L.-H. Qu. Identification of 20 snorna-like rnas from the primitive eukaryote *Gardia lamblia*. *Biochemical and Biophysical Research Communication*, 328:1224–1231, 2005.
277. Y. Yano, R. Saito, N. Yoshida, A. Yoshiki, A. Wynshaw-Boris, M. Tomita, and S. Hirotsune. A new role for expressed pseudogenes as ncRNA: regulation of mrna stability of its homologous coding gene. *J Mol Med*, 82:414–422, 2004.
278. M. C. Yao, P. Fuller, and X. Xi. Programmed DNA deletion as an RNA-guided system of genome defense. *Science*, 300:1517–1518, 2003.
279. M. A. Zago, P. P. Dennis, and A. D. Omer. The expanding world of small rnas in the hyperthermophilic archaeon *sulfolobus solfataricus*. *Mol Microbiol*, 55:1812–1828, 2005.
280. F. Zalfa, S. Adinolfi, I. Napoli, E. Kühn-Hölsken, H. Urlaub, T. Achsel, A. Pastore, and C. Bagni. Fragile X mental retardation protein (FMRP) binds specifically to the brain cytoplasmic RNAs BC1/BC200 via a novel RNA-binding motif. *J Biol Chem*, 280:33403–33410, 2005.
281. D. C. Zappulla and T. R. Cech. Yeast telomerase RNA: a flexible scaffold for protein subunits. *Proc Natl Acad Sci USA*, 101:10024–10029, 2004.
282. A. Zemann, A. op de Bekke, M. Kiefmann, J. Brosius, and J. Schmitz. Evolution of small nucleolar RNAs in nematodes. *Nucleic Acids Res*, 34:2676–2685, 2006.
283. Y. Zhu, D. K. Pulkunnat, and Y. Li. Deciphering RNA structural diversity and systematic phylogeny from microbial metagenomes. *Nucleic Acids Res*, 35:2283–2294, 2007.
284. C. Zwieb, R. W. van Nues, M. A. Rosenblad, J. D. Brown, and T. Samuelson. A nomenclature for all signal recognition particle RNAs. *RNA*, 11:7–13, 2005.