

## When one is better than two: RNA with dual functions

Damien Ulveling, Claire Francastel, Florent Hubé

► **To cite this version:**

Damien Ulveling, Claire Francastel, Florent Hubé. When one is better than two: RNA with dual functions. *Biochimie*, Elsevier, 2011, 93 (4), pp.633-644. 10.1016/j.biochi.2010.11.004 . hal-02127323

**HAL Id: hal-02127323**

**<https://hal-univ-paris.archives-ouvertes.fr/hal-02127323>**

Submitted on 16 Dec 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **When one is better than two: RNA with dual functions**

Damien Ulveling, Claire Francastel and Florent Hubé\*

Université Paris Diderot, CNRS UMR7216, Epigenetics and Cell Fate, Paris, France

\* Corresponding author:

Dr. Florent Hubé

UMR7216 - Epigénétique et Destin Cellulaire

Université Paris 7 Diderot

Bâtiment Lamarck - 4ème étage

Case Courrier 7042

35, rue Hélène Brion

75013 PARIS

Tél: +33 1 57 27 89 32

Fax: +33 1 57 27 89 10

E-mail: florent.hube@univ-paris-diderot.fr

**Running head:** RNA with dual functions

**Key words:** bifunctional RNA; mRNA-like RNA; non-coding RNA; gene expression; steroid receptor RNA co-activator; untranslated

## Abstract

The central dogma of biology, until not long ago, held that genetic information stored on DNA molecules was translated into the final protein products through RNA as intermediate molecules. Then, an additional level of complexity in the regulation of genome expression was added, implicating new classes of RNA molecules called non-coding RNA (ncRNA). These ncRNA are also often referred to as functional RNA in that, although they do not contain the capacity to encode proteins, do have a function as RNA molecules. They have been thus far considered as truly non-coding RNA since no ORF long enough to be considered, nor protein, have been associated with them. However, the recent identification and characterization of bifunctional RNA, *i.e.* RNA for which both coding capacity and activity as functional RNA have been reported, suggests that a definite categorization of some RNA molecules is far from being straightforward.

Indeed, several RNA primarily classified as non-protein-coding RNA has been showed to hold coding capacities and associated peptides. Conversely, mRNA, usually regarded as strictly protein-coding, may act as functional RNA molecules. Here, we describe several examples of these bifunctional RNA that have been already characterized from bacteria to mammals. We also extend this concept to fortuitous acquisition of dual function in pathological conditions and to the recently highlighted duality between information carried by a gene and its pseudogenes counterparts.

## 1. Introduction

It is now well admitted that one needs to think beyond the canonical protein-coding function when it comes to study the functionality of the genome. While the protein-coding gene count of eukaryotic genomes is relatively poor and not well correlated with organism complexity, the increasing awareness that most of the human genome may be transcribed, and that the number of genes that do not code for proteins is much larger than expected [1-3], added some levels of sophistication in the way organisms gain their phenotypic variability [4]. Besides expression of protein-coding genes being regulated through alternative splicing and regulation of promoter strength/usage, an additional level in the regulation of genome expression involves RNA molecules called non-coding RNA (ncRNA). ncRNA do not encode for protein and mechanisms whereby they play key roles in the regulations of complex organisms remain largely obscure (For a simplified chronological chart, see Figure 1).

[Figure 1 may be inserted here]

ncRNA are often referred to as functional RNA in that, although they do not hold a capacity to encode proteins, they have a function as RNA molecules. The role of ribosomal RNA (rRNA), transfer RNA (tRNA), and more recently microRNA (miRNA), in protein translation, or small nuclear RNA (snRNA) in splicing of messenger RNA (mRNA), is now well documented. Fascinatingly, other fundamental mechanisms may also depend on new classes of ncRNA. Many long ncRNA (lncRNA), so called to distinguish them from **small interfering RNA** (siRNA) and miRNA, function as RNA scaffolds implicated in chromosome architecture, base-pairing-**dependent** mRNA targeting influencing mRNA turnover or developmental timing of protein expression, and may also directly regulate transcription and alternative splicing [5-14]. Examples include B2-SINE that can directly bind RNA polymerase II to inhibit gene expression under heat-shock conditions [15], or 7SK that

represses transcription elongation through its interaction with the basal transcription factor P-TEFb [16; 17]. Xist and centromeric RNA contribute to assembly of specific chromatin structures at the inactive X in female cells and centromeres respectively, by recruiting histone variants and/or serving as a scaffold for complexes located at these regions [7; 14; 18]. The activity of transcription factors can also be regulated through their sequestration in the cytoplasm of the cells by ncRNA. For example, NRON (non-coding repressor of NFAT) binds to and prevents the nuclear localization, and therefore the action, of the transcriptional activator NFAT [19].

These RNA are thus far considered to be truly non-coding RNA since no **open reading frame** (ORF) long enough to be considered, nor protein, have been associated with them. The class of long non-coding RNA also includes many mRNA-like non-coding RNA (lncRNA), identified in high-throughput studies of the transcriptome, which can undergo splicing, are capped and polyadenylated, but contain little or no protein-coding potential [20-22]. Instead, they may influence cellular activity through specific sequences and/or RNA-folding structures.

Despite accumulating data on so-called non-coding RNA, it is likely that we are only watching the tip of the iceberg. We review here what is currently emerging as an **increasingly fascinating** area in the field, along which many RNA, usually regarded as strictly protein-coding mRNA, may fulfill additional non-coding functions. Conversely, many RNA primarily classified as non-protein-coding RNA hold uncovered coding capacities. We focus on this duality within the “RNA world” that may contribute to increase the spectrum of gene function, **by describing** examples of these bifunctional RNA that have been already characterized from bacteria to mammals. **Attention must also be drawn** to apparently conflicting new functions of mRNA in the control of protein activity by influencing their sub-cellular localization or transport in normal or pathological conditions. Finally, **we extended**

this concept to the recently highlighted duality between information carried by a gene and its pseudogenes counterparts.

## **2. Examples of coding vs. non-coding swing during evolution**

It is not clear why during evolution some genes lost their protein-coding function by mutation or conversely, why non-coding or pseudogenes gained the capacity to encode proteins. Nevertheless, in both cases the result is acquisition of novel functions, mainly due to transposon integration. We review here well-documented examples of such changes, which primarily highlights that genes indeed evolved to acquire new or additional functions, yet, the effector molecule can be a protein or an RNA depending on the organism considered.

### ***2.1 From coding to non-coding: the example of XIC (X-Inactivation Center) genes***

In female mammals, one of the two X chromosomes is transcriptionally silenced during early development to compensate for the double dose of X-linked genes (For recent reviews see [23-25]). In mouse and human, this process of X inactivation is highly **dependent** on a locus located on the X chromosome itself, the X Inactivation Center (XIC) [26; 27]. The XIC genomic region contains several genes of which six produce untranslated RNA (reviewed in [28; 29]), including the two major players Xist and its antisense counterpart Tsix [30-32]. X inactivation is initiated through expression of Xist RNA **from, and coating of**, the inactive X chromosome (Xi), leading to Xi chromosome-wide transcriptional silencing (reviewed in [33]). Tsix acts in *cis* to negatively regulate Xist, through modulation of chromatin over the Xist promoter, thereby affecting the ability of a given chromosome to be inactivated (reviewed in [34]). While the role of Xist and Tsix is well documented, no defined function in X-chromosome inactivation is known for the other neighboring non-coding genes.

As mentioned above, the *Xist* gene is expressed exclusively from the inactive X chromosome as a large processed RNA for which no protein product has been identified. However, sequence comparison between species revealed that evolution of *Xist* gene was accompanied by pseudogenization of a coding gene located within a region of synteny between chicken and human and by integration of mobile elements [35; 36]. Two exons of *Xist* containing frame shift mutations derive from the *LnX3* gene, which in chicken produces an mRNA encoding a protein similar to PDZ domain containing ring finger 1 (ligand of NUMB-protein X/PDZRN (PDZ domain-containing RING finger)). In fact, pseudogenization and invasion by transposable elements appear to have affected other genes in the XIC syntenic locus. The loss of protein-coding function in *LnX3* (to become *Xist*) coincides with the loss of function of several other protein genes (*Fip112*, *Rasl11c*, *UspL* and *Wave4*) to become transcribed but untranslated *Tsx*, *Jpx* and *Ftx* genes [35]). These genes seem to have evolved a little bit further since some of them like *Tsx* are still functionally transcribed and translated in mouse and rat [28] but not in human [37].

It is interesting to note that although pseudogenization led to disruption of the *LnX3* ORF and loss of protein-coding function for the new *Xist* gene, transposon insertions led to the appearance of new functional domains. These domains, like the repeat A in exon 1, then became necessary for the silencing function of *Xist* [38]. As **suggested** earlier [35] this example of how a protein-coding gene lost its protein-coding function by mutation but gained a new function due to transposon integration is not a unique case, but is a more wide-spread phenomenon applying to other ncRNA genes and pseudogenes with new functions.

## ***2.2 From non-coding to coding: transposable elements***

Just like genes losing their protein coding capacity through insertion of mobile elements, the reverse situation exists in which transposable elements (TE) were “exonized”.

In this process, TE insertion into introns was probably followed by mutations that allowed the RNA splicing machinery to recruit part of the inserted transposable element into coding regions as novel exons [39]. Exonization can therefore permit the emergence of new regulatory motifs and new protein functions, without necessarily eliminating the original function of a protein. In that respect, retrotransposons and other classes of transposable elements are now generally considered as significant contributors to genome evolution [40] influencing genome size, function, and structure [41; 42]. Indeed, about half of the human genome is derived from transposable elements (TE), with about 8% of Long Terminal Repeat (LTR) and 33% of non-LTR retro-elements [41; 43]. In addition, TE motifs are found in the coding sequence of 4% of human genes, suggesting that these repeat-forming elements can become exonized (reviewed in [44; 45]). Strikingly, in some cases, distinct exons are exclusively composed of LTR sequences. For example, the 4th, 5th, and 14th exons of interleukin 22 receptor alpha 2 (*IL22RA2*, NM\_052962), SET domain, bifurcated 2 (*SETDB2*, NM\_031915), and methyl-CpG binding domain protein 1 (*MBD1*, NM\_015845) genes, respectively, are entirely formed by LTR sequences [45] and the BCL2/adenovirus E1B 19kDa interacting protein 3 (*BNIP3*, NM\_004052) is 97% made up of sequences from a human retrovirus [46].

### 3. Bifunctional RNA

Bifunctional RNA can be defined as RNA that have the intrinsic ability to produce two distinct functions. A class of bifunctional RNA, on which we will not focus here, produce two different types of ncRNA like a snoRNA and a miRNA, involved in distinct biological pathways [47]. A more remarkable category now contains several examples of **RNA which serve both as intermediate molecules translated into protein and as functional RNA** (Box1). As we will see below, these functions are in essence mechanistically distinct but can be

physiologically redundant, and do not necessarily co-exist within the same cell or at the same period of the life of the cell.

**[BOX 1 may be inserted here]**

### ***3.1 SgrS, a small RNA and its SgrT protein counterpart (Escherichia coli)***

Sugar phosphate stress in *Escherichia coli* is controlled by a 227-nt small RNA (sRNA) SgrS (sugar transport-related small RNA) and SgrR, which encodes the transcriptional activator required for *SgrS* transcription [48]. *SgrS* expression is induced by intracellular accumulation of glucose-6-phosphate (G6P) and is necessary for full recovery from glucose-phosphate stress by inducing the degradation of glucose transporters. SgrS holds two distinct functions that inhibit glucose transport independently. SgrS is first a riboregulator, *i.e.* an RNA functioning as a genetic regulator, a class of RNA widely described in bacteria [49; 50]. Its 3' region is partially complementary to the translation initiation region of mRNA that encode glucose transporters, leading to a base pairing-dependent translational repression and degradation of these target mRNA [106]. In addition, the 5' end of *SgrS* encodes the 43-amino-acid polypeptide SgrT, also implicated in the inhibition of glucose transport, yet by a mechanism independent of the base-pairing activity [51]. It was proposed that SgrT could trigger a rapid response to stress by reducing influx of sugar phosphates without affecting levels or stability of the transport proteins, whereas the riboregulation activity of SgrS could result from an adaptation to prolonged stress, by stopping synthesis of new transporters [51].

SgrS therefore shows features of a bifunctional sRNA that encodes physiologically redundant but mechanistically distinct functions implicated in the same response to stress. Of note, homologs of the transcription factor SgrR and its target sRNA SgrS, primarily identified in *E. coli*, have been identified in a variety of other enteric bacterial species [107].

### 3.2 RNAIII (*Staphylococcus aureus*)

The small RNAIII is the intracellular effector of the quorum-sensing system, a system used by bacteria to coordinate their gene expression according to their local density of population, in the nosocomial pathogen *Staphylococcus aureus*. RNAIII represents a paradigm in the field of regulatory RNA linked to bacterial pathogenesis. Whereas it was originally described as a mRNA coding the  $\delta$ -hemolysin peptide, further analysis revealed that, in addition to the 26 amino acid ORF in its 5' region [54; 55], RNAIII acted as an antisense regulator of virulence and surface protein synthesis through its 3' region [52; 53]. Base pairing of the 3' domain of RNAIII to target mRNA that encode a class of virulence factors, or the transcriptional regulatory protein Rot, occlude ribosome binding site leading to the repression of translation initiation and triggering of endoribonuclease III-mediated hydrolysis [108; 109].

Thus, RNA III of *S. aureus* is a master regulator of Staphylococcal virulence that also meets the definition of a bifunctional sRNA. It acts as a mRNA that encodes the  $\delta$ -hemolysin peptide, which controls the switch between early expression of surface proteins and late expression of several exotoxins, as well as a trans-acting RNA, which modulates the production of *S. aureus* extracellular toxins and enzymes, through a base pairing mechanism on mRNA targets [55; 108-111]. Although precise mechanisms are still unclear, it was proposed that the  $\delta$ -hemolysin protein may act in conjunction with its RNAIII transcript counterpart to regulate the expression of exoprotein genes [55; 112; 113].

### **3.3 tmRNA (*Bacteria*)**

Transfer-messenger RNA (tmRNA), also known as SsrA or 10Sa RNA, is a bifunctional RNA that may look odd in the world of coding/non-coding RNA reviewed here in that it combines the functions of tRNA and mRNA. Trans-translation and the tmRNA system perform translational surveillance to rescue ribosomes stalled on defective mRNA, missing a stop codon for instance, and contribute to the degradation of incompletely synthesized peptides, in all eubacteria and some eukaryotic organelles [56; 57]. tmRNA acts initially as transfer RNA (tRNA), being aminoacylated at its 3'-end, to add alanine to the stalled polypeptide chain. Translation then switches from the defective mRNA to an internal tmRNA short open reading frame (ORF) of 8-35 residues, whereby the mRNA-like domain of tmRNA replaces the nonstop messenger RNA bound to the ribosome [56-60]. The degradation tag encoded by the ORF targets the incompletely synthesized protein for degradation by cellular proteases. The stop codon at the end of the tmRNA ORF provides the missing termination signal, which causes the release of the synthesized polypeptide and the dissociation of ribosomal subunits.

tmRNA dual function primarily serves the quality control of mRNA. A general role for trans-translation in monitoring protein folding and perhaps other co-translational processes has been proposed recently [56-60].

**[Box 2 may be inserted here]**

### **3.4 *enod40*, early nodulin 40 (*Green plants*)**

When leguminous plants enter into symbiosis with N<sub>2</sub>-fixing bacteria, a number of genes are coordinately expressed from both partners. **The peculiar gene *enod40* is rapidly induced and highly expressed in the dividing root cortical cells, leading to nodule formation**

as well as in other non-symbiotic organs [123]. The *enod40* gene is present in all legumes studied so far, and is also found in many non-legume plants. In plants, *enod40* mRNA (about 600 bp) contains two short overlapping ORF leading to translation of two peptides of 10-25 amino acids [124], implicated in the promotion of division of cortical cells, probably through binding to sucrose synthase and regulation of sucrose utilization in nodules. In addition, its core region shows hallmarks of structural RNA, namely a strongly conserved secondary structure topology despite very high sequence diversity. The putative RNA structural elements are more conserved than the peptide sequence, suggesting that the general mechanism of *enod40* action might be determined by the RNA [62]. Deletion of the region spanning the predicted RNA structure, but leaving the sORFs intact, did not affect translation of the mRNA but resulted in a significantly decreased biological activity. The stem-loop structures mediate interaction with the nuclear RNA-binding protein MtRBP1 [61]. One hypothesis is that the RNA itself, through integration into a RNP, will be a major determinant of *enod40* function conferring specificity for the recognition of mRNA targets. The interactions are independent of *enod40* ORFs translation and affects MtRBP1 localization during nodule development, changing from nuclear speckles, involved in splicing and RNA regulatory processes, to cytoplasmic granules, suggesting the implication of *enod40* RNA in the cytoplasmic export of nuclear proteins, independently of a protein targeting motif.

This set of data pointed out the dual function of *enod40* as both a peptide producing and a structured RNA, sharing functions in developmental processes [125].

### ***3.5 Maternal effect protein oskar (Drosophila melanogaster)***

The *Drosophila* maternal effect gene *oskar* encodes the posterior determinant responsible for the establishment of anterior-posterior polarity in *Drosophila melanogaster* and the formation of the abdomen and germ line of the future fly. The establishment of

anteroposterior patterning of the *Drosophila* embryo is determined by the polarization and translation prior to posterior localization of the *oskar* gene product (for review see [126-128]). The site of *oskar* RNA and protein localization within the oocyte determines where in the embryo primordial germ cells form and where the abdomen develops. Upon localization of *oskar* RNA and protein at the posterior pole, Oskar protein is required to maintain localization of *oskar* RNA throughout oogenesis. It was proposed that initially localization of *oskar* RNA permits translation into Oskar protein and that subsequently Oskar protein regulates its own RNA localization through a positive feedback mechanism [66].

Mutations in the two *oskar* isoforms, which are necessary for the recruitment of various factors involved in posterior pole differentiation, result in a classical *oskar* mutant phenotype in which the flies lack germ cells and abdomens [129]. Common to these classical *oskar* alleles is that they all produce significant amounts of *oskar* mRNA. Surprisingly, two new *oskar* mutants that affect *oskar* mRNA abundance showed an even stronger phenotype, with arrest early in oogenesis, which suggests a functional role for *oskar* mRNA at an earlier stage in development [64]. The authors suggested that *oskar* mRNA, and not the Oskar protein, is responsible for the early oogenesis function. In a series of elegant experiments, the authors demonstrated that the early arrest of oogenesis could be complemented by *oskar* nonsense mutant alleles or the 3'-UTR of *oskar* mRNA [64]. Moreover, they showed that expression of the *oskar* 3' untranslated region (3'UTR) is sufficient to rescue the egg-less defect of the RNA null mutant.

It appears then that *oskar* acts as an mRNA during the embryonic stage and a non-coding RNA during early oogenesis, independent of *oskar* coding capacity [64]. Oskar protein serves as a scaffold for the assembly of cytoplasmic structures essential for germ line development [66], the polar granules, at the posterior pole of the oocyte and embryo. Similarly, at earlier stages, *oskar* RNA may play a structural role by providing a similar

scaffold function for assembly of cytoplasmic complexes essential for development of the oocyte [64; 126].

### ***3.6 Transcription factor vegt (Xenopus laevis)***

An important feature of the *Xenopus* oocyte is that it is polarized along the animal/vegetal axis with the animal hemisphere containing the oocyte nucleus, and the vegetal hemisphere containing localized coding and non-coding RNA of maternal origin, involved in the control of endoderm, mesoderm, and germ cell specification during oogenesis. The coding VegT maternal mRNA encodes a transcription factor of the T-box family, related to Brachyury. VegT activates several endoderm-specific transcription factors, essential for the formation of the endodermal germ layer in blastula stages of development, and a group of TGF $\beta$ -related signaling molecules involved in mesoderm induction and required for proper gastrulation movements [70; 72; 73]. Indeed, overexpression of VegT in ectodermal explants generates endoderm in a dose-dependent manner [71]. Likewise, the antisense depletion of VegT transcripts caused abnormal distribution and/or expression of a subset of localized vegetal mRNA and prevents endoderm development [74]. Since then, VegT RNA has been shown to also fulfill a separate structural role in the cytoskeleton network of primordial germ cells. The use of VegT morpholino that inhibits translation of VegT protein without affecting the level of targeted VegT mRNA, did not disperse the vegetally localized mRNA, indicating that a subset of localized mRNA is dependent on VegT mRNA but not on VegT protein [67]. VegT and other maternal ncRNA are anchored at the oocyte vegetal cortex, through integration into a cytoskeleton network of cytoskeleton and actin [68]. Destruction of these RNA results in disruption of the cytoskeleton and defects in the anchoring of other localized RNA [67]. The structural function of VegT mRNA is independent of the VegT

protein, and organization of the cyokeratin cytoskeleton can be rescued by the injection of exogenous VegT mRNA [69].

VegT, like oskar cited above, belongs to a class of “localized” mRNA involved in ensuring a high-level synthesis of protein at the site where they are needed and/or preventing the protein from being synthesized in a region of the cell where it is not needed or could have a harmful effect. However, besides making a protein, as it is assumed from an mRNA, VegT and oskar RNA have an independent function as structural RNA. As a consequence, a normal cell phenotype may require both the function of the protein and the autonomous function of its mRNA (reviewed in [69]).

### ***3.7 A new function for p53 mRNA revealed in stress conditions***

The fascinating tumor suppressor p53 senses a variety of cellular stresses such as DNA damage, oncogenic signals or nutrient deprivation, and promotes cellular arrest and programmed cell death. In normal conditions, p53 protein is persistently degraded by the proteasome (for review see [56; 59; 60; 97-99]). In contrast, upon stress, both half-life of the p53 protein and translation of its mRNA are enhanced. The E3 ubiquitin ligase Mdm2 is the major regulator of p53 tumor suppressor activity. The well-known interaction between Mdm2 and p53 proteins promotes p53 ubiquitination and targets it for degradation by the proteasome. Recently, p53 mRNA this time was shown to interact with Mdm2 protein with the effect of both stimulating p53 synthesis, but, more importantly, preventing Mdm2 from promoting p53 protein ubiquitination and degradation [95; 96]. The authors proposed that p53 mRNA may act as a switch controlling the function of Mdm2 towards p53 [95]. Worth mentioning, the ability of Mdm2 to control both p53 synthesis and degradation are carried by the same genomic sequences in both *mdm2* and *p53*. Hence, mutations within the coding region of this type of mRNA have to be carefully designed since they may affect both the

activity of the encoded protein and the activity of the mRNA as a functional RNA, and lead to results difficult to interpret.

### ***3.8 Steroid Receptor RNA activator (SRA) and SRAP (Mammalian)***

With the aim of discovering new partners of progesterone receptor (PR), with a co-regulatory function, Lanz *et al.* identified a RNA molecule that they named SRA for Steroid Receptor RNA Activator [82]. SRA was an unusual partner since despite the presence of a long ORF no initiating methionines were identifiable. Since then, several studies have confirmed the capacity of SRA RNA to co-activate not only PR, but also many other nuclear receptors such as estrogen (ER $\alpha$  and  $\beta$ ), progesterone, androgen and glucocorticoid receptors, as well as peroxisome proliferator-activated receptor (PPAR), all-trans retinoic acid receptor and many others (For a review, see [77; 78; 86]). More recently, transcriptional activity of other transcription factors including MyoD and GATA-3 has also been shown to be enhanced by SRA RNA [75; 80]. Structural characterization of SRA RNA highlighted that the core region of the messenger, consisting of exons 2 to 5, was necessary and sufficient to increase the ligand-dependent transcriptional activation of target genes by steroid receptors [82; 83]. Specifically, two stem-loop substructures, STR1 and STR7, were involved in ER co-activation.

By analogy to other described ncRNA, SRA transcripts may nucleate the assembly of co-regulators at their target genes and/or favor the correct nuclear localization of the different partners of the complex in response to external stimuli. Alternatively, SRA RNA may serve as a scaffold to stabilize large co-activator complexes. In favor of this hypothesis, SRA RNA has been already described embedded in nucleoprotein complexes containing positive regulators, including the receptor co-activator SRC-1 [82], the RNA helicases p68/p72 [75; 88] as well as the pseudouridine synthase Pus1p and Pus3p [76; 89; 90]. It was also shown to interact with

negative regulators such as the SMRT/HDAC1 Associated Repressor Protein SHARP [83; 87] and the SRA stem-loop interacting RNA binding protein SLIRP [79].

SRA transcripts have been identified in normal human tissues, with a higher expression in liver, skeletal muscle, adrenal and pituitary glands, whereas intermediate expression levels were observed in the placenta, lung, kidney and pancreas [82]. In some pathological cases, increased SRA RNA levels were reported like in breast and ovarian tumors [81; 84; 85]. Interestingly, levels of SRA expression could be characteristic of tumor grade or particular subtypes of lesions among different tumors. Indeed, serous ovarian tumors showed higher levels of SRA than granulosa tumor cells [81].

Searching for transcriptional initiation sites, we identified new SRA RNA isoforms containing an additional exon 1 upstream of the core elements, with the peculiarity that these isoforms exhibited a longer predicted ORF, provided that the first intron was correctly spliced, as well as two ATG in the new exon. These features were not only compatible with the ability to encode a protein, a protein named SRAP was actually identified [80; 91; 93]. Interestingly, among the new isoforms, we also identified isoforms that displayed retention of intron 1 or splicing of exon 3 disrupting the ORF [80; 92; 94]. It appears then, that the *SRA1* gene products consist in fact of two distinct entities: functional RNA with the ability to enhance the activity of transcription factors through structural motifs in their core sequence, and a protein whose function remains to be deciphered. Although SRAP has direct effects on the activity of NR [91], we recently showed that SRAP did not interact with MyoD [80; 92; 94]. However, we demonstrated that SRAP was able to interact with its RNA counterpart, leading to an overall inhibition of MyoD transcriptional activity and myogenic differentiation [80; 92; 94] (Figure 2).

**[Figure 2 may be inserted here]**

Coding and non-coding SRA RNA isoforms co-exist in most of the cells. In pathological conditions, like in breast cancer, we described unbalanced expression levels in favor of non-coding SRA RNA in the three most invasive cell lines (MDA-MB-231, MDA-MB-468, and BT-20) compared to MCF-10A1 breast cells isolated from benign tumors [94]. Likewise, breast tumors with a high progesterone receptor content, *i.e.* a poor prognostic factor, the balance was on the side of non-coding SRA RNA [130], suggesting that a perturbation in the balance between SRA could be associated with tumorigenesis [94]. During myogenic differentiation of human satellite cells from healthy but not from myotonic dystrophy donors, the ratio between non-coding and coding SRA isoforms was also elevated [80; 92; 94].

These accumulating data suggest that the overall effect of SRA on both molecular activities and cellular processes may result from the correct balance between coding and non-coding SRA molecules. This statement may of course hold true for any other bifunctional RNA, and stresses the need to carefully evaluate the nature, *i.e.* coding *vs.* non-coding, of molecules present at a given time or in a given cell population.

Therefore, any attempt to decipher functions of SRA or any RNA with dual functions, must take into account this remarkable bifunctional characteristic consisting of an RNA molecule that potentially encodes several proteins and ncRNA, with opposing functions.

#### **4. New functions for mRNA under pathological conditions**

Certain pathologies arise, not from the production of mutant proteins, but from a deleterious gain-of-function by ncRNA [131; 132]. In these RNA-dominant diseases, the toxic RNA has been implicated in the formation of nuclear foci and associated with degenerative disorders, mainly in brain, heart or skeletal muscle [133]. In addition, the normal function of

mRNA, *i.e.* to serve as an intermediate molecule for protein synthesis, is hijacked in that the modified RNA functions as a functional ncRNA, and thus interferes with physiological pathways. These modifications may include nucleotide triplets repeat expansions in the 3' or 5' untranslated region of the mRNA, like in fragile X mental retardation 1 (FRM1) associated with adult neurodegenerative disorder and dystrophia myotonica-protein kinase (DMPK) transcripts associated with neuromuscular dystrophy (for reviews, see [100; 101]). It appears that the main pathogenic effect of the toxic RNA is to sequester binding proteins and compromise the regulation of alternative splicing. In myotonic dystrophy (DM), for example, the mutation does not directly cause the absence of an essential protein or a dysfunctional protein. The main effect is to induce misdistribution of splicing factors in cells, and the consequence is reversion to a pattern of alternative splicing that is characteristic of immature tissue [134; 135]. For example, the **depletion** by the CGG repeats of RNA-binding proteins such as Sam68, hnRNP-G and MBNL1, which are RNA-binding proteins involved in alternative splicing regulation. Loss of these factors leads to splicing-regulatory dysfunction [134-137]. Transcripts from the mutant DMPK allele accumulate in the nucleus and compromise the regulation of alternative splicing by sequestration of specific RNA-binding proteins such as CUGBP (Elav-like family), Elav-type ribonucleoprotein member 1 and 3 (CUGBP and ETR-3), and MBNL1 in nuclear foci [100; 101; 138].

**Therefore, the repeat expansion characteristic of certain pathologies may fortuitously convert mRNA, with classical function as intermediate in protein synthesis, into functional RNA implicated in protein sequestration, through the acquisition of double-stranded hairpin structures conserved in all types of triplets [139; 140].**

## 5. One gene, one pseudogene, two functions

Pseudogenes are defective genes that have lost their ability to encode proteins. A pseudogene and its related gene often share a common ancestor, but have diverged as separate genetic entities over millions of years. According to recent studies, between 5 to 20% of human pseudogenes, which account for about 17,032 gene units (database build58, NCBI37), retain the ability to be translated (for reviews, see [141; 142]). However, pseudogenes are considered as false genes because of the absence of protein products resulting from deletions, insertions or mutations causing a frame shift or appearance of premature stop codons. There are three main types of pseudogenes, which differ in their origin; processed pseudogenes originate from retro-transposition of mRNA transcripts into the chromosome; duplicated pseudogenes are coming from gene duplication and successive mutations that render it deficient to be transcribed; unitary pseudogenes are genes which are disabled or deactivated without previous gene duplication (for reviews, see [141; 143; 144]).

Whether pseudogenes are nothing more than inactivated copies of protein-coding genes or if they still retain a function is unclear, but some studies formally addressed the question. The most "obvious" mechanisms include the **depletion** of regulatory factors, or the production of antisense transcripts. For example, MKRN1, which encodes the makorin-1 Ring Finger Protein 1, a transcriptional co-regulator of androgen receptor and retinoic acid receptor and an E3 ligase for hTERT [104], and its retro-transposed pseudogene makorin1-p1, both contain receptor sites for transcriptional repressors [102]. Makorin1-p1 competes for the freely available repressor molecules that would otherwise submerge the gene's receptors and severely inhibit its expression. This '**depletion**' of the repressor by the pseudogene therefore favors protein synthesis from the gene [103]. Nitric oxide synthase (NOS) pseudogene includes a region of significant antisense homology to the neuronal NOS (nNOS)-encoding mRNA and acts as a natural antisense regulator of nNOS protein synthesis [145]. A more

novel mechanism involves **depletion** of miRNA otherwise implicated in regulation of expression of the gene counterpart. *PTEN* is a crucial, haplo-insufficient, tumor suppressor gene, **whose** expression is modulated by several miRNA. Transcripts originated from *PTEN* and from its pseudogene counterpart, *PTENP1*, show identical sequences in their 3'UTR regions. Since these regions are targets for miRNA, then transcripts from *PTENP1* compete for miRNA interaction and lead to de-repression of *PTEN* and enhanced tumor-suppressor activity [105]. The same type of regulation was demonstrated for the oncogene *KRAS* by transcripts originated from its pseudogene *KRAS1P* [105].

It is not question here, strictly speaking, of bifunctional RNA since both functions, coding and non-coding, are carried by two distinct genomic entities, although issued from a common ancestral gene. However, we are left here with a coding gene and its non-coding counterpart for which non-coding RNA can arise. As exposed above, these RNA can titrate regulatory factors and miRNA, and because of the probable multiplicity of their target genes, these RNA may affect regulation not only of their coding-gene counterpart but also of any co-expressed gene modulated by the same entities. We believe that it will be important in the future to take into account these pseudo-transcripts as their interference with regulation of gene expression may go far beyond the cited examples.

## **6. ncRNA vs. mRNA: Do we have to make the distinction ?**

In eukaryotes, mRNA undergo several processing steps before **being** translated. This includes addition of a cap structure at the 5' end, splicing of introns from the body of the pre-mRNA, and generation of a 3' end, usually modified by the addition of a poly(A) tail.

Although each of these reactions are biochemically distinct processes, they are interlinked and therefore, influence specificity and efficiency of one another [146]. While polyadenylation

has been implicated mainly in stabilization of the mRNA [147-149], capping is thought to be critical for the recognition and proper attachment of the ribosome, and therefore for translation (for a review [150; 151]). Thus, it is relatively surprising to note that numerous ncRNA are transcribed by polymerase II, processed like conventional mRNA and capped (Figure 3), although it has not been established yet that they are actually capable of producing a protein.

**[FIGURE 3 may be inserted here]**

Conversely, the main criterion which allows for the discrimination between an mRNA or a ncRNA is essentially based on its ability to encode a protein, and thus the length of its ORF, currently established at 100 amino acids (aa) [167]. However, depending of the length of the transcripts, longer ORF could occur by chance. Likewise, 5-10% of described proteins in Swiss-Prot are indeed shorter than 100 aa, even though the majority are variants of proteins longer than 100 aa [168]. Dinger *et al.* described an interesting model to take into account the length of the ORF and the length of the transcripts [122]. Thus, they established a cut-off above which the size of the ORF is not likely to appear randomly. We used their calculation to identify and measure the ORF length in databases containing 940 unique ncRNA sequences which have been identified and manually curated based upon extensive literature reviews [169]. As shown in Figure 4, 326 of the 940 sequences appeared to have a non-random ORF, with a 95% confidence. Even more surprisingly, when using an additional cut-off of 500 aa, which is very unlikely to occur by chance, 49 sequences still presented a coding capacity.

**[FIGURE 4 may be inserted here]**

The reverse is also true; short ORF <100 aa have been described [168; 170-174], and can be as short as 7 aa such as *mccC7*, a microcin heptapeptide involved in inhibition of protein synthesis in *Enterobacteriaceae* [175]. Another example is the already cited

bifunctional RNA enod40, that was first described as a ncRNA because of its very short ORF. Enod40 nonetheless encodes two small proteins of 12 and 24 aa [124].

In essence, whether an mRNA can function as a ncRNA or if a ncRNA has the potential to encode a protein will not be assessed solely based on bioinformatics approaches, but will also require functional evaluation. Because many long ncRNA are likely to carry coding capacities like mRNA, and conversely, mRNA may operate as structural RNA, attempts to classify ncRNA or mRNA in strictly separate classes may become rapidly unrealistic.

## 7. Conclusion

More than a century after the discovery of RNA, and for nearly a decade, bifunctional RNA, *i.e.* RNA carrying both RNA-translatable and RNA-intrinsic functions, emerged as new players in molecular and cellular biology systems.

We have just discussed here the few examples that are well documented and have provided a complementary view of the functional role of RNA in addition to their ability to encode proteins. Even though their number is limited, not fully understood and still necessitates to be closely studied, we are confident that new examples will add to the list quickly. Indeed, in yeast, it was predicted that about 5% of the genes may preserve both non-coding and coding capacity [176; 177], and similar prediction was made in other species [178; 179].

Hence, it becomes increasingly clear that we probably only looked through a keyhole in the last decades. By definition, the majority of ncRNA does not indeed have the ability to be translated into a long protein. However, it will probably prove interesting to test those that fulfill the criteria exposed above. In addition, new uncovered small peptides may also be

translated from ncRNA. It seems also that mRNA can form sub-structures and may be more implicated **than** previously thought in cellular processes, making studies of bifunctionality of RNA the next step in deciphering life complexity. We believe that awareness of bifunctionality of RNA is not only of great conceptual importance, but will be increasingly useful when it comes to decipher RNA function in a manner that discriminates this duality.

## **Acknowledgements**

We are grateful to Dr. Guillaume Velasco for insightful comments on the manuscript (UMR7216, University of Paris Diderot, Paris, France). F.H. was supported by « Association Française contre les Myopathies » and « Association le Cancer du Sein, Parlons-en ! » fellowships.

## Reference List

- [1] J. Berretta, A. Morillon, Pervasive transcription constitutes a new level of eukaryotic genome regulation, *EMBO Rep.* 10 (2009) 973-982.
- [2] M.E. Dinger, P.P. Amaral, T.R. Mercer, J.S. Mattick, Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications, *Brief.Funct.Genomic.Proteomic.* 8 (2009) 407-423.
- [3] A.R. Forrest, P. Carninci, Whole genome transcriptome analysis, *RNA.Biol.* 6 (2009) 107-112.
- [4] J.S. Mattick, Non-coding RNAs: the architects of eukaryotic complexity, *EMBO Rep.* 2 (2001) 986-991.
- [5] J.N. Hutchinson, A.W. Ensminger, C.M. Clemson, C.R. Lynch, J.B. Lawrence, A. Chess, A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains, *BMC.Genomics* 8 (2007) 39.
- [6] V. Tripathi, J.D. Ellis, Z. Shen, D.Y. Song, Q. Pan, A.T. Watt, S.M. Freier, C.F. Bennett, A. Sharma, P.A. Bubulya, B.J. Blencowe, S.G. Prasanth, K.V. Prasanth, The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation, *Mol.Cell* 39 (2010) 925-938.
- [7] F. Ferri, H. Bouzinba-Segard, G. Velasco, F. Hube, C. Francastel, Non-coding murine centromeric transcripts associate with and potentiate Aurora B kinase, *Nucleic Acids Res.* 37 (2009) 5071-5080.
- [8] C. Maison, D. Bailly, A.H. Peters, J.P. Quivy, D. Roche, A. Taddei, M. Lachner, T. Jenuwein, G. Almouzni, Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component, *Nat.Genet.* 30 (2002) 329-334.
- [9] P.A. Latos, D.P. Barlow, Regulation of imprinted expression by macro non-coding RNAs, *RNA.Biol.* 6 (2009) 100-106.
- [10] S. Schoeftner, M.A. Blasco, Chromatin regulation and non-coding RNAs at mammalian telomeres, *Semin.Cell Dev.Biol.* 21 (2010) 186-193.
- [11] M. Matzke, T. Kanno, B. Huettel, L. Daxinger, A.J. Matzke, Targets of RNA-directed DNA methylation, *Curr.Opin.Plant Biol.* 10 (2007) 512-519.
- [12] C.A. Brosnan, O. Voinnet, The long and the short of noncoding RNAs, *Curr.Opin.Cell Biol.* 21 (2009) 416-425.
- [13] K.V. Morris, Long antisense non-coding RNAs function to direct epigenetic complexes that regulate transcription in human cells, *Epigenetics.* 4 (2009) 296-301.

- [14] H. Bouzinba-Segard, A. Guais, C. Francastel, Accumulation of small murine minor satellite transcripts leads to impaired centromeric architecture and function, *Proc.Natl.Acad.Sci.U.S.A* 103 (2006) 8709-8714.
- [15] T.A. Allen, K.S. Von, J.A. Goodrich, J.F. Kugel, The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock, *Nat.Struct.Mol.Biol.* 11 (2004) 816-821.
- [16] V.T. Nguyen, T. Kiss, A.A. Michels, O. Bensaude, 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes, *Nature* 414 (2001) 322-325.
- [17] Z. Yang, Q. Zhu, K. Luo, Q. Zhou, The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription, *Nature* 414 (2001) 317-322.
- [18] S.L. Gilbert, J.R. Pehrson, P.A. Sharp, XIST RNA associates with specific regions of the inactive X chromatin, *J.Biol.Chem.* 275 (2000) 36491-36494.
- [19] A.T. Willingham, A.P. Orth, S. Batalov, E.C. Peters, B.G. Wen, P. za-Blanc, J.B. Hogenesch, P.G. Schultz, A strategy for probing the function of noncoding RNAs finds a repressor of NFAT, *Science* 309 (2005) 1570-1573.
- [20] P. Carninci, Y. Hayashizaki, Noncoding RNA transcription beyond annotated genes, *Curr.Opin.Genet.Dev.* 17 (2007) 139-144.
- [21] S. Griffiths-Jones, Annotating noncoding RNA genes, *Annu.Rev.Genomics Hum.Genet.* 8 (2007) 279-298.
- [22] P. Kapranov, J. Cheng, S. Dike, D.A. Nix, R. Duttagupta, A.T. Willingham, P.F. Stadler, J. Hertel, J. Hackermuller, I.L. Hofacker, I. Bell, E. Cheung, J. Drenkow, E. Dumais, S. Patel, G. Helt, M. Ganesh, S. Ghosh, A. Piccolboni, V. Sementchenko, H. Tammana, T.R. Gingeras, RNA maps reveal new RNA classes and a possible function for pervasive transcription, *Science* 316 (2007) 1484-1488.
- [23] T.S. Barakat, I. Jonkers, K. Monkhorst, J. Gribnau, X-changing information on X inactivation, *Exp.Cell Res.* 316 (2010) 679-687.
- [24] J. Chow, E. Heard, X inactivation and the complexities of silencing a sex chromosome, *Curr.Opin.Cell Biol.* 21 (2009) 359-366.
- [25] M. Leeb, P.A. Steffen, A. Wutz, X chromosome inactivation sparked by non-coding RNAs, *RNA.Biol.* 6 (2009) 94-99.
- [26] N. Brockdorff, A. Ashworth, G.F. Kay, P. Cooper, S. Smith, V.M. McCabe, D.P. Norris, G.D. Penny, D. Patel, S. Rastan, Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome, *Nature* 351 (1991) 329-331.
- [27] C.J. Brown, A. Ballabio, J.L. Rupert, R.G. Lafreniere, M. Grompe, R. Tonlorenzi, H.F. Willard, A gene from the region of the human X inactivation centre is

expressed exclusively from the inactive X chromosome, *Nature* 349 (1991) 38-44.

- [28] C. Chureau, M. Prissette, A. Bourdet, V. Barbe, L. Cattolico, L. Jones, A. Eggen, P. Avner, L. Duret, Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine, *Genome Res.* 12 (2002) 894-908.
- [29] P. Clerc, P. Avner, Multiple elements within the Xic regulate random X inactivation in mice, *Semin.Cell Dev.Biol.* 14 (2003) 85-92.
- [30] G. Borsani, R. Tonlorenzi, M.C. Simmler, L. Dandolo, D. Arnaud, V. Capra, M. Grompe, A. Pizzuti, D. Muzny, C. Lawrence, ., Characterization of a murine gene expressed from the inactive X chromosome, *Nature* 351 (1991) 325-329.
- [31] N. Brockdorff, A. Ashworth, G.F. Kay, V.M. McCabe, D.P. Norris, P.J. Cooper, S. Swift, S. Rastan, The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus, *Cell* 71 (1992) 515-526.
- [32] J.T. Lee, L.S. Davidow, D. Warshawsky, Tsix, a gene antisense to Xist at the X-inactivation centre, *Nat.Genet.* 21 (1999) 400-404.
- [33] H. Marks, J.C. Chow, S. Denissov, K.J. Francoijs, N. Brockdorff, E. Heard, H.G. Stunnenberg, High-resolution analysis of epigenetic changes associated with X inactivation, *Genome Res.* 19 (2009) 1361-1373.
- [34] C. Rougeulle, P. Avner, The role of antisense transcription in the regulation of X-inactivation, *Curr.Top.Dev.Biol.* 63 (2004) 61-89.
- [35] L. Duret, C. Chureau, S. Samain, J. Weissenbach, P. Avner, The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene, *Science* 312 (2006) 1653-1655.
- [36] E.A. Elisaphenko, N.N. Kolesnikov, A.I. Shevchenko, I.B. Rogozin, T.B. Nesterova, N. Brockdorff, S.M. Zakian, A dual origin of the Xist gene from a protein-coding gene and a set of transposable elements, *PLoS.One.* 3 (2008) e2521.
- [37] D.B. Cunningham, D. Segretain, D. Arnaud, U.C. Rogner, P. Avner, The mouse Tsx gene is expressed in Sertoli cells of the adult testis and transiently in premeiotic germ cells during puberty, *Dev.Biol.* 204 (1998) 345-360.
- [38] A. Wutz, T.P. Rasmussen, R. Jaenisch, Chromosomal silencing and localization are mediated by different domains of Xist RNA, *Nat.Genet.* 30 (2002) 167-174.
- [39] A. Nekrutenko, W.H. Li, Transposable elements are found in a large number of human protein-coding genes, *Trends Genet.* 17 (2001) 619-621.
- [40] J. Brosius, RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements, *Gene* 238 (1999) 115-134.
- [41] H.H. Kazazian, Jr., Mobile elements: drivers of genome evolution, *Science* 303 (2004) 1626-1632.

- [42] W. Makalowski, Y. Toda, Modulation of host genes by mammalian transposable elements, *Genome Dyn.* 3 (2007) 163-174.
- [43] A. Jasinska, W.J. Krzyzosiak, Repetitive sequences that shape the human transcriptome, *FEBS Lett.* 567 (2004) 136-141.
- [44] H. Keren, G. Lev-Maor, G. Ast, Alternative splicing and evolution: diversification, exon definition and function, *Nat.Rev.Genet.* 11 (2010) 345-355.
- [45] J. Piriyaopongsa, N. Polavarapu, M. Borodovsky, J. McDonald, Exonization of the LTR transposable elements in human genome, *BMC.Genomics* 8 (2007) 291.
- [46] R.J. Britten, Coding sequences of functioning human genes derived entirely from mobile element sequences, *Proc.Natl.Acad.Sci.U.S.A* 101 (2004) 16825-16830.
- [47] C. Ender, A. Krek, M.R. Friedlander, M. Beitzinger, L. Weinmann, W. Chen, S. Pfeffer, N. Rajewsky, G. Meister, A human snoRNA with microRNA-like functions, *Mol.Cell* 32 (2008) 519-528.
- [48] C.K. Vanderpool, S. Gottesman, Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system, *Mol.Microbiol.* 54 (2004) 1076-1089.
- [49] N. Majdalani, C.K. Vanderpool, S. Gottesman, Bacterial small RNA regulators, *Crit Rev.Biochem.Mol.Biol.* 40 (2005) 93-113.
- [50] G. Storz, J.A. Opdyke, K.M. Wassarman, Regulating bacterial transcription with small RNAs, *Cold Spring Harb.Symp.Quant.Biol.* 71 (2006) 269-273.
- [51] C.S. Wadler, 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.U.S.A* 104 (2007) 20454-20459.
- [52] C. Chevalier, S. Boisset, C. Romilly, B. Masquida, P. Fechter, T. Geissmann, F. Vandenesch, P. Romby, *Staphylococcus aureus* RNAIII binds to two distant regions of *coa* mRNA to arrest translation and promote mRNA degradation, *PLoS.Pathog.* 6 (2010) e1000809.
- [53] R.P. Novick, H.F. Ross, S.J. Projan, J. Kornblum, B. Kreiswirth, S. Moghazeh, Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule, *EMBO J.* 12 (1993) 3967-3975.
- [54] N. Balaban, R.P. Novick, Translation of RNAIII, the *Staphylococcus aureus* agr regulatory RNA molecule, can be activated by a 3'-end deletion, *FEMS Microbiol.Lett.* 133 (1995) 155-161.
- [55] L. Janzon, S. Arvidson, The role of the delta-lysin gene (*hld*) in the regulation of virulence genes by the accessory gene regulator (*agr*) in *Staphylococcus aureus*, *EMBO J.* 9 (1990) 1391-1399.

- [56] D. Dulebohn, J. Choy, T. Sundermeier, N. Okan, A.W. Karzai, Trans-translation: the tmRNA-mediated surveillance mechanism for ribosome rescue, directed protein degradation, and nonstop mRNA decay, *Biochemistry* 46 (2007) 4681-4693.
- [57] R. Gillet, B. Felden, Emerging views on tmRNA-mediated protein tagging and ribosome rescue, *Mol.Microbiol.* 42 (2001) 879-885.
- [58] P.W. Haebel, S. Gutmann, N. Ban, Dial tm for rescue: tmRNA engages ribosomes stalled on defective mRNAs, *Curr.Opin.Struct.Biol.* 14 (2004) 58-65.
- [59] C.S. Hayes, K.C. Keiler, Beyond ribosome rescue: tmRNA and co-translational processes, *FEBS Lett.* 584 (2010) 413-419.
- [60] K.C. Keiler, Biology of trans-translation, *Annu.Rev.Microbiol.* 62 (2008) 133-151.
- [61] A. Campalans, A. Kondorosi, M. Crespi, Enod40, a short open reading frame-containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula*, *Plant Cell* 16 (2004) 1047-1059.
- [62] A.P. Gulyaev, A. Roussis, Identification of conserved secondary structures and expansion segments in enod40 RNAs reveals new enod40 homologues in plants, *Nucleic Acids Res.* 35 (2007) 3144-3152.
- [63] G. Girard, A. Roussis, A.P. Gulyaev, C.W. Pleij, H.P. Spaink, Structural motifs in the RNA encoded by the early nodulation gene enod40 of soybean, *Nucleic Acids Res.* 31 (2003) 5003-5015.
- [64] A. Jenny, O. Hachet, P. Zavorszky, A. Cyrklaff, M.D. Weston, D.S. Johnston, M. Erdelyi, A. Ephrussi, A translation-independent role of oskar RNA in early *Drosophila* oogenesis, *Development* 133 (2006) 2827-2833.
- [65] A. Ephrussi, L.K. Dickinson, R. Lehmann, Oskar organizes the germ plasm and directs localization of the posterior determinant nanos, *Cell* 66 (1991) 37-50.
- [66] C. Rongo, E.R. Gavis, R. Lehmann, Localization of oskar RNA regulates oskar translation and requires Oskar protein, *Development* 121 (1995) 2737-2746.
- [67] J. Heasman, O. Wessely, R. Langland, E.J. Craig, D.S. Kessler, Vegetal localization of maternal mRNAs is disrupted by VegT depletion, *Dev.Biol.* 240 (2001) 377-386.
- [68] M. Kloc, K. Wilk, D. Vargas, Y. Shirato, S. Bilinski, L.D. Etkin, Potential structural role of non-coding and coding RNAs in the organization of the cytoskeleton at the vegetal cortex of *Xenopus* oocytes, *Development* 132 (2005) 3445-3457.
- [69] M. Kloc, Teachings from the egg: new and unexpected functions of RNAs, *Mol.Reprod.Dev.* 76 (2009) 922-932.
- [70] J.A. White, J. Heasman, Maternal control of pattern formation in *Xenopus laevis*, *J.Exp.Zool.B Mol.Dev.Evol.* 310 (2008) 73-84.

- [71] M.E. Horb, G.H. Thomsen, A vegetally localized T-box transcription factor in *Xenopus* eggs specifies mesoderm and endoderm and is essential for embryonic mesoderm formation, *Development* 124 (1997) 1689-1698.
- [72] J.B. Xanthos, M. Kofron, C. Wylie, J. Heasman, Maternal VegT is the initiator of a molecular network specifying endoderm in *Xenopus laevis*, *Development* 128 (2001) 167-180.
- [73] J. Zhang, M.L. King, *Xenopus* VegT RNA is localized to the vegetal cortex during oogenesis and encodes a novel T-box transcription factor involved in mesodermal patterning, *Development* 122 (1996) 4119-4129.
- [74] J. Zhang, D.W. Houston, M.L. King, C. Payne, C. Wylie, J. Heasman, The role of maternal VegT in establishing the primary germ layers in *Xenopus* embryos, *Cell* 94 (1998) 515-524.
- [75] G. Caretti, R.L. Schiltz, F.J. Dilworth, P.M. Di, P. Zhao, V. Ogryzko, F.V. Fuller-Pace, E.P. Hoffman, S.J. Tapscott, V. Sartorelli, The RNA helicases p68/p72 and the noncoding RNA SRA are coregulators of MyoD and skeletal muscle differentiation, *Dev.Cell* 11 (2006) 547-560.
- [76] M. Charette, M.W. Gray, Pseudouridine in RNA: what, where, how, and why, *IUBMB.Life* 49 (2000) 341-351.
- [77] S.M. Colley, K.R. Iyer, P.J. Leedman, The RNA coregulator SRA, its binding proteins and nuclear receptor signaling activity, *IUBMB.Life* 60 (2008) 159-164.
- [78] S.M. Colley, P.J. Leedman, SRA and its binding partners: an expanding role for RNA-binding coregulators in nuclear receptor-mediated gene regulation, *Crit Rev.Biochem.Mol.Biol.* 44 (2009) 25-33.
- [79] E.C. Hatchell, S.M. Colley, D.J. Beveridge, M.R. Epis, L.M. Stuart, K.M. Giles, A.D. Redfern, L.E. Miles, A. Barker, L.M. MacDonald, P.G. Arthur, J.C. Lui, J.L. Golding, R.K. McCulloch, C.B. Metcalf, J.A. Wilce, M.C. Wilce, R.B. Lanz, B.W. O'Malley, P.J. Leedman, SLIRP, a small SRA binding protein, is a nuclear receptor corepressor, *Mol.Cell* 22 (2006) 657-668.
- [80] F. Hube, G. Velasco, J. Rollin, D. Furling, C. Francastel, Steroid receptor RNA activator protein binds to and counteracts SRA RNA-mediated activation of MyoD and muscle differentiation, *Nucleic Acids Res.* (2010) [Epub ahead of print].
- [81] S. Hussein-Fikret, P.J. Fuller, Expression of nuclear receptor coregulators in ovarian stromal and epithelial tumours, *Mol.Cell Endocrinol.* 229 (2005) 149-160.
- [82] R.B. Lanz, N.J. McKenna, S.A. Onate, U. Albrecht, J. Wong, S.Y. Tsai, M.J. Tsai, B.W. O'Malley, A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex, *Cell* 97 (1999) 17-27.
- [83] R.B. Lanz, B. Razani, A.D. Goldberg, B.W. O'Malley, Distinct RNA motifs are important for coactivation of steroid hormone receptors by steroid receptor RNA activator (SRA), *Proc.Natl.Acad.Sci.U.S.A* 99 (2002) 16081-16086.

- [84] R.B. Lanz, S.S. Chua, N. Barron, B.M. Soder, F. DeMayo, B.W. O'Malley, Steroid receptor RNA activator stimulates proliferation as well as apoptosis in vivo, *Mol.Cell Biol.* 23 (2003) 7163-7176.
- [85] E. Leygue, H. Dotzlaw, P.H. Watson, L.C. Murphy, Expression of the steroid receptor RNA activator in human breast tumors, *Cancer Res.* 59 (1999) 4190-4193.
- [86] E. Leygue, Steroid receptor RNA activator (SRA1): unusual bifaceted gene products with suspected relevance to breast cancer, *Nucl.Recept.Signal.* 5 (2007) e006.
- [87] Y. Shi, M. Downes, W. Xie, H.Y. Kao, P. Ordentlich, C.C. Tsai, M. Hon, R.M. Evans, Sharp, an inducible cofactor that integrates nuclear receptor repression and activation, *Genes Dev.* 15 (2001) 1140-1151.
- [88] M. Watanabe, J. Yanagisawa, H. Kitagawa, K. Takeyama, S. Ogawa, Y. Arao, M. Suzawa, Y. Kobayashi, T. Yano, H. Yoshikawa, Y. Masuhiro, S. Kato, A subfamily of RNA-binding DEAD-box proteins acts as an estrogen receptor alpha coactivator through the N-terminal activation domain (AF-1) with an RNA coactivator, SRA, *EMBO J.* 20 (2001) 1341-1352.
- [89] X. Zhao, J.R. Patton, S.L. Davis, B. Florence, S.J. Ames, R.A. Spanjaard, Regulation of nuclear receptor activity by a pseudouridine synthase through posttranscriptional modification of steroid receptor RNA activator, *Mol.Cell* 15 (2004) 549-558.
- [90] X. Zhao, J.R. Patton, S.K. Ghosh, N. Fischel-Ghodsian, L. Shen, R.A. Spanjaard, Pus3p- and Pus1p-dependent pseudouridylation of steroid receptor RNA activator controls a functional switch that regulates nuclear receptor signaling, *Mol.Endocrinol.* 21 (2007) 686-699.
- [91] S. Chooniedass-Kothari, E. Emberley, M.K. Hamedani, S. Troup, X. Wang, A. Czosnek, F. Hube, M. Mutawe, P.H. Watson, E. Leygue, The steroid receptor RNA activator is the first functional RNA encoding a protein, *FEBS Lett.* 566 (2004) 43-47.
- [92] C. Cooper, J. Guo, Y. Yan, S. Chooniedass-Kothari, F. Hube, M.K. Hamedani, L.C. Murphy, Y. Myal, E. Leygue, Increasing the relative expression of endogenous non-coding Steroid Receptor RNA Activator (SRA) in human breast cancer cells using modified oligonucleotides, *Nucleic Acids Res.* 37 (2009) 4518-4531.
- [93] E. Emberley, G.J. Huang, M.K. Hamedani, A. Czosnek, D. Ali, A. Grolla, B. Lu, P.H. Watson, L.C. Murphy, E. Leygue, Identification of new human coding steroid receptor RNA activator isoforms, *Biochem.Biophys.Res.Commun.* 301 (2003) 509-515.
- [94] F. Hube, J. Guo, S. Chooniedass-Kothari, C. Cooper, M.K. Hamedani, A.A. Dibrov, A.A. Blanchard, X. Wang, G. Deng, Y. Myal, E. Leygue, Alternative splicing of the first intron of the steroid receptor RNA activator (SRA) participates in the generation of coding and noncoding RNA isoforms in breast cancer cell lines, *DNA Cell Biol.* 25 (2006) 418-428.

- [95] M.M. Candeias, L. Malbert-Colas, D.J. Powell, C. Daskalogianni, M.M. Maslon, N. Naski, K. Bourougaa, F. Calvo, R. Fahraeus, P53 mRNA controls p53 activity by managing Mdm2 functions, *Nat.Cell Biol.* 10 (2008) 1098-1105.
- [96] N. Naski, M. Gajjar, K. Bourougaa, L. Malbert-Colas, R. Fahraeus, M.M. Candeias, The p53 mRNA-Mdm2 interaction, *Cell Cycle* 8 (2009) 31-34.
- [97] M. Farnebo, V.J. Bykov, K.G. Wiman, The p53 tumor suppressor: a master regulator of diverse cellular processes and therapeutic target in cancer, *Biochem.Biophys.Res.Comm.* 396 (2010) 85-89.
- [98] N.D. Lakin, S.P. Jackson, Regulation of p53 in response to DNA damage, *Oncogene* 18 (1999) 7644-7655.
- [99] M. Schuler, D.R. Green, Mechanisms of p53-dependent apoptosis, *Biochem.Soc.Trans.* 29 (2001) 684-688.
- [100] H. Jiang, A. Mankodi, M.S. Swanson, R.T. Moxley, C.A. Thornton, Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons, *Hum.Mol.Genet.* 13 (2004) 3079-3088.
- [101] L.T. Timchenko, J.W. Miller, N.A. Timchenko, D.R. DeVore, K.V. Datar, L. Lin, R. Roberts, C.T. Caskey, M.S. Swanson, Identification of a (CUG)<sub>n</sub> triplet repeat RNA-binding protein and its expression in myotonic dystrophy, *Nucleic Acids Res.* 24 (1996) 4407-4414.
- [102] S. Hirotsune, N. Yoshida, A. Chen, L. Garrett, F. Sugiyama, S. Takahashi, K. Yagami, A. Wynshaw-Boris, A. Yoshiki, An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene, *Nature* 423 (2003) 91-96.
- [103] Y. Yano, R. Saito, N. Yoshida, A. Yoshiki, A. Wynshaw-Boris, M. Tomita, S. Hirotsune, A new role for expressed pseudogenes as ncRNA: regulation of mRNA stability of its homologous coding gene, *J.Mol.Med.* 82 (2004) 414-422.
- [104] T.A. Gray, L. Hernandez, A.H. Carey, M.A. Schaldach, M.J. Smithwick, K. Rus, J.A. Marshall Graves, C.L. Stewart, R.D. Nicholls, The ancient source of a distinct gene family encoding proteins featuring RING and C(3)H zinc-finger motifs with abundant expression in developing brain and nervous system, *Genomics* 66 (2000) 76-86.
- [105] L. Poliseno, L. Salmena, J. Zhang, B. Carver, W.J. Haveman, P.P. Pandolfi, A coding-independent function of gene and pseudogene mRNAs regulates tumour biology, *Nature* 465 (2010) 1033-1038.
- [106] K. Maki, T. Morita, H. Otaka, H. Aiba, A minimal base-pairing region of a bacterial small RNA SgrS required for translational repression of ptsG mRNA, *Mol.Microbiol.* 76 (2010) 782-792.

- [107] R.S. Horler, C.K. Vanderpool, Homologs of the small RNA SgrS are broadly distributed in enteric bacteria but have diverged in size and sequence, *Nucleic Acids Res.* 37 (2009) 5465-5476.
- [108] E. Geisinger, R.P. Adhikari, R. Jin, H.F. Ross, R.P. Novick, Inhibition of rot translation by RNAIII, a key feature of agr function, *Mol.Microbiol.* 61 (2006) 1038-1048.
- [109] E. Huntzinger, S. Boisset, C. Saveanu, Y. Benito, T. Geissmann, A. Namane, G. Lina, J. Etienne, B. Ehresmann, C. Ehresmann, A. Jacquier, F. Vandenesch, P. Romby, *Staphylococcus aureus* RNAIII and the endoribonuclease III coordinately regulate spa gene expression, *EMBO J.* 24 (2005) 824-835.
- [110] Y. Benito, F.A. Kolb, P. Romby, G. Lina, J. Etienne, F. Vandenesch, Probing the structure of RNAIII, the *Staphylococcus aureus* agr regulatory RNA, and identification of the RNA domain involved in repression of protein A expression, *RNA.* 6 (2000) 668-679.
- [111] 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, P. Romby, *Staphylococcus aureus* RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism, *Genes Dev.* 21 (2007) 1353-1366.
- [112] A. Toledo-Arana, F. Repoila, P. Cossart, Small noncoding RNAs controlling pathogenesis, *Curr.Opin.Microbiol.* 10 (2007) 182-188.
- [113] L.S. Waters, G. Storz, Regulatory RNAs in bacteria, *Cell* 136 (2009) 615-628.
- [114] D.J. SenGupta, B. Zhang, B. Kraemer, P. Pochart, S. Fields, M. Wickens, A three-hybrid system to detect RNA-protein interactions in vivo, *Proc.Natl.Acad.Sci.U.S.A* 93 (1996) 8496-8501.
- [115] A. Dickson, E. Osman, C. Lorson, A Negatively-Acting Bifunctional RNA Increases Survival Motor Neuron in vitro and in vivo, *Hum.Gene Ther.* (2008).
- [116] D.D. Rao, P.B. Maples, N. Senzer, P. Kumar, Z. Wang, B.O. Pappen, Y. Yu, C. Haddock, C. Jay, A.P. Phadke, S. Chen, J. Kuhn, D. Dylewski, S. Scott, D. Monsma, C. Webb, A. Tong, D. Shanahan, J. Nemunaitis, Enhanced target gene knockdown by a bifunctional shRNA: a novel approach of RNA interference, *Cancer Gene Ther.* (2010).
- [117] T.D. Baughan, A. Dickson, E.Y. Osman, C.L. Lorson, Delivery of bifunctional RNAs that target an intronic repressor and increase SMN levels in an animal model of spinal muscular atrophy, *Hum.Mol.Genet.* 18 (2009) 1600-1611.
- [118] M.J. Wood, M.J. Gait, H. Yin, RNA-targeted splice-correction therapy for neuromuscular disease, *Brain* 133 (2010) 957-972.
- [119] J.P. Bachellerie, J. Cavaille, A. Huttenhofer, The expanding snoRNA world, *Biochimie* 84 (2002) 775-790.

- [120] P. Ganot, M. Caizergues-Ferrer, 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 (1997) 941-956.
- [121] T. Kiss, W. Filipowicz, Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns, *Genes Dev.* 9 (1995) 1411-1424.
- [122] M.E. Dinger, K.C. Pang, T.R. Mercer, J.S. Mattick, Differentiating protein-coding and noncoding RNA: challenges and ambiguities, *PLoS.Comput.Biol.* 4 (2008) e1000176.
- [123] C. Charon, A.B. Moreno, F. Bardou, M. Crespi, Non-Protein-Coding RNAs and their Interacting RNA-Binding Proteins in the Plant Cell Nucleus, *Mol.Plant* 3 (2010) 729-739.
- [124] H. Rohrig, J. Schmidt, E. Miklashevichs, J. Schell, M. John, Soybean ENOD40 encodes two peptides that bind to sucrose synthase, *Proc.Natl.Acad.Sci.U.S.A* 99 (2002) 1915-1920.
- [125] C. Sousa, C. Johansson, C. Charon, H. Manyani, C. Sautter, A. Kondorosi, 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 (2001) 354-366.
- [126] J.M. Kugler, P. Lasko, Localization, anchoring and translational control of *oskar*, *gurken*, *bicoid* and *nanos* mRNA during *Drosophila* oogenesis, *Fly.(Austin.)* 3 (2009) 15-28.
- [127] H. Tekotte, I. Davis, Bruno: a double turn-off for Oskar, *Dev.Cell* 10 (2006) 280-281.
- [128] J. Kim-Ha, J.L. Smith, P.M. Macdonald, *oskar* mRNA is localized to the posterior pole of the *Drosophila* oocyte, *Cell* 66 (1991) 23-35.
- [129] R. Lehmann, C. Nusslein-Volhard, Abdominal segmentation, pole cell formation, and embryonic polarity require the localized activity of *oskar*, a maternal gene in *Drosophila*, *Cell* 47 (1986) 141-152.
- [130] S. Chooniedass-Kothari, M.K. Hamedani, S. Troup, F. Hube, E. Leygue, The steroid receptor RNA activator protein is expressed in breast tumor tissues, *Int.J.Cancer* 118 (2006) 1054-1059.
- [131] R. Galvao, L. Mendes-Soares, J. Camara, I. Jaco, M. Carmo-Fonseca, Triplet repeats, RNA secondary structure and toxic gain-of-function models for pathogenesis, *Brain Res.Bull.* 56 (2001) 191-201.
- [132] R.J. Osborne, C.A. Thornton, RNA-dominant diseases, *Hum.Mol.Genet.* 15 Spec No 2 (2006) R162-R169.
- [133] D.G. Arocena, C.K. Iwahashi, N. Won, A. Beilina, A.L. Ludwig, F. Tassone, P.H. Schwartz, P.J. Hagerman, Induction of inclusion formation and disruption of

lamin A/C structure by premutation CGG-repeat RNA in human cultured neural cells, *Hum.Mol.Genet.* 14 (2005) 3661-3671.

- [134] B. Charlet, R.S. Savkur, G. Singh, A.V. Philips, E.A. Grice, T.A. Cooper, Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing, *Mol.Cell* 10 (2002) 45-53.
- [135] T.H. Ho, R.S. Savkur, M.G. Poulos, M.A. Mancini, M.S. Swanson, T.A. Cooper, Colocalization of muscleblind with RNA foci is separable from mis-regulation of alternative splicing in myotonic dystrophy, *J.Cell Sci.* 118 (2005) 2923-2933.
- [136] C. Sellier, F. Rau, Y. Liu, F. Tassone, R.K. Hukema, R. Gattoni, A. Schneider, S. Richard, R. Willemsen, D.J. Elliott, P.J. Hagerman, N. Charlet-Berguerand, Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients, *EMBO J.* 29 (2010) 1248-1261.
- [137] O.A. Sofola, P. Jin, Y. Qin, R. Duan, H. Liu, H.M. de, D.L. Nelson, J. Botas, RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a *Drosophila* model of FXTAS, *Neuron* 55 (2007) 565-571.
- [138] J.W. Miller, C.R. Urbinati, P. Teng-Umnuay, M.G. Stenberg, B.J. Byrne, C.A. Thornton, M.S. Swanson, Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy, *EMBO J.* 19 (2000) 4439-4448.
- [139] M. Napierala, D. Michalowski, M.M. de, W.J. Krzyzosiak, Facile FMR1 mRNA structure regulation by interruptions in CGG repeats, *Nucleic Acids Res.* 33 (2005) 451-463.
- [140] K. Sobczak, M.M. de, G. Michlewski, J. Krol, W.J. Krzyzosiak, RNA structure of trinucleotide repeats associated with human neurological diseases, *Nucleic Acids Res.* 31 (2003) 5469-5482.
- [141] Z. Zhang, M. Gerstein, Large-scale analysis of pseudogenes in the human genome, *Curr.Opin.Genet.Dev.* 14 (2004) 328-335.
- [142] D. Zheng, A. Frankish, R. Baertsch, P. Kapranov, A. Reymond, S.W. Choo, Y. Lu, F. Denoeud, S.E. Antonarakis, M. Snyder, Y. Ruan, C.L. Wei, T.R. Gingeras, R. Guigo, J. Harrow, M.B. Gerstein, Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution, *Genome Res.* 17 (2007) 839-851.
- [143] E.S. Balakirev, F.J. Ayala, Pseudogenes: are they "junk" or functional DNA?, *Annu.Rev.Genet.* 37 (2003) 123-151.
- [144] Z. Zhang, N. Carriero, M. Gerstein, Comparative analysis of processed pseudogenes in the mouse and human genomes, *Trends Genet.* 20 (2004) 62-67.

- [145] S.A. Korneev, J.H. Park, M. O'Shea, Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene, *J.Neurosci.* 19 (1999) 7711-7720.
- [146] N.J. Proudfoot, A. Furger, M.J. Dye, Integrating mRNA processing with transcription, *Cell* 108 (2002) 501-512.
- [147] T. Humphrey, G. Christofori, V. Lucijanic, W. Keller, Cleavage and polyadenylation of messenger RNA precursors in vitro occurs within large and specific 3' processing complexes, *EMBO J.* 6 (1987) 4159-4168.
- [148] M. Sadofsky, J.C. Alwine, Sequences on the 3' side of hexanucleotide AAUAAA affect efficiency of cleavage at the polyadenylation site, *Mol.Cell Biol.* 4 (1984) 1460-1468.
- [149] N.J. Proudfoot, G.G. Brownlee, 3' non-coding region sequences in eukaryotic messenger RNA, *Nature* 263 (1976) 211-214.
- [150] N. Berteaux, N. Aptel, G. Cathala, C. Genton, J. Coll, A. Daccache, N. Spruyt, H. Hondermarck, T. Dugimont, J.J. Cury, T. Forne, E. Adriaenssens, A novel H19 antisense RNA overexpressed in breast cancer contributes to paternal IGF2 expression, *Mol.Cell Biol.* 28 (2008) 6731-6745.
- [151] N. Sonenberg, Cap-binding proteins of eukaryotic messenger RNA: functions in initiation and control of translation, *Prog.Nucleic Acid Res.Mol.Biol.* 35 (1988) 173-207.
- [152] K. Panzitt, M.M. Tschernatsch, C. Guelly, T. Moustafa, M. Stradner, H.M. Strohmaier, C.R. Buck, H. Denk, R. Schroeder, M. Trauner, K. Zatloukal, Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA, *Gastroenterology* 132 (2007) 330-342.
- [153] I. Seim, S.L. Carter, A.C. Herington, L.K. Chopin, Complex organisation and structure of the ghrelin antisense strand gene GHRLOS, a candidate non-coding RNA gene, *BMC.Mol.Biol.* 9 (2008) 95.
- [154] H. Ren, Y. Li, Z. Tang, S. Yang, Y. Mu, W. Cui, H. Ao, L. Du, L. Wang, K. Li, Genomic structure, chromosomal localization and expression profile of a porcine long non-coding RNA isolated from long SAGE libraries, *Anim Genet.* 40 (2009) 499-508.
- [155] E. Memili, Y.K. Hong, D.H. Kim, S.D. Ontiveros, W.M. Strauss, Murine Xist RNA isoforms are different at their 3' ends: a role for differential polyadenylation, *Gene* 266 (2001) 131-137.
- [156] H.R. Cohen, B. Panning, XIST RNA exhibits nuclear retention and exhibits reduced association with the export factor TAP/NXF1, *Chromosoma* 116 (2007) 373-383.
- [157] W. Tam, Identification and characterization of human BIC, a gene on chromosome 21 that encodes a noncoding RNA, *Gene* 274 (2001) 157-167.

- [158] T. Glaser, D. Housman, W.H. Lewis, D. Gerhard, C. Jones, A fine-structure deletion map of human chromosome 11p: analysis of J1 series hybrids, *Somat.Cell Mol.Genet.* 15 (1989) 477-501.
- [159] A.J. Zaugg, T.R. Cech, Analysis of the structure of Tetrahymena nuclear RNAs in vivo: telomerase RNA, the self-splicing rRNA intron, and U2 snRNA, *RNA.* 1 (1995) 363-374.
- [160] A.G. Seto, A.J. Zaugg, S.G. Sobel, S.L. Wolin, T.R. Cech, *Saccharomyces cerevisiae* telomerase is an Sm small nuclear ribonucleoprotein particle, *Nature* 401 (1999) 177-180.
- [161] G.P. Shumyatsky, S.V. Tillib, D.A. Kramerov, B2 RNA and 7SK RNA, RNA polymerase III transcripts, have a cap-like structure at their 5' end, *Nucleic Acids Res.* 18 (1990) 6347-6351.
- [162] K.R. Novak, I. Grbesa, M. Ivkic, M. Katdare, K. Gall-Troselj, Curcumin downregulates H19 gene transcription in tumor cells, *J.Cell Biochem.* 104 (2008) 1781-1792.
- [163] C.I. Seidl, S.H. Stricker, D.P. Barlow, The imprinted Air ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export, *EMBO J.* 25 (2006) 3565-3575.
- [164] C.I. Brannan, E.C. Dees, R.S. Ingram, S.M. Tilghman, The product of the H19 gene may function as an RNA, *Mol.Cell Biol.* 10 (1990) 28-36.
- [165] V. Pachnis, C.I. Brannan, S.M. Tilghman, The structure and expression of a novel gene activated in early mouse embryogenesis, *EMBO J.* 7 (1988) 673-681.
- [166] P. Navarro, S. Pichard, C. Ciaudo, P. Avner, C. Rougeulle, Tsix transcription across the Xist gene alters chromatin conformation without affecting Xist transcription: implications for X-chromosome inactivation, *Genes Dev.* 19 (2005) 1474-1484.
- [167] N. Maeda, T. Kasukawa, R. Oyama, J. Gough, M. Frith, P.G. Engstrom, 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, Y. Hayashizaki, Transcript annotation in FANTOM3: mouse gene catalog based on physical cDNAs, *PLoS.Genet.* 2 (2006) e62.
- [168] M.C. Frith, A.R. Forrest, E. Nourbakhsh, K.C. Pang, C. Kai, J. Kawai, P. Carninci, Y. Hayashizaki, T.L. Bailey, S.M. Grimmond, The abundance of short proteins in the mammalian proteome, *PLoS.Genet.* 2 (2006) e52.
- [169] K.C. Pang, S. Stephen, P.G. Engstrom, K. Tajul-Arifin, W. Chen, C. Wahlestedt, B. Lenhard, Y. Hayashizaki, J.S. Mattick, RNADB--a comprehensive mammalian noncoding RNA database, *Nucleic Acids Res.* 33 (2005) D125-D130.

- [170] J.P. Kastenmayer, L. Ni, A. Chu, L.E. Kitchen, W.C. Au, H. Yang, C.D. Carter, D. Wheeler, R.W. Davis, J.D. Boeke, M.A. Snyder, M.A. Basrai, Functional genomics of genes with small open reading frames (sORFs) in *S. cerevisiae*, *Genome Res.* 16 (2006) 365-373.
- [171] M.I. Galindo, J.I. Pueyo, S. Fouix, S.A. Bishop, J.P. Couso, Peptides encoded by short ORFs control development and define a new eukaryotic gene family, *PLoS.Biol.* 5 (2007) e106.
- [172] T. Kondo, S. Plaza, J. Zanet, E. Benrabah, P. Valenti, Y. Hashimoto, S. Kobayashi, F. Payre, Y. Kageyama, Small peptides switch the transcriptional activity of Shavenbaby during *Drosophila* embryogenesis, *Science* 329 (2010) 336-339.
- [173] T. Ota, Y. Suzuki, T. Nishikawa, T. Otsuki, T. Sugiyama, R. Irie, A. Wakamatsu, K. Hayashi, H. Sato, K. Nagai, K. Kimura, H. Makita, M. Sekine, M. Obayashi, T. Nishi, T. Shibahara, T. Tanaka, S. Ishii, J. Yamamoto, K. Saito, Y. Kawai, Y. Isono, Y. Nakamura, K. Nagahari, K. Murakami, T. Yasuda, T. Iwayanagi, M. Wagatsuma, A. Shiratori, H. Sudo, T. Hosoiri, Y. Kaku, H. Kodaira, H. Kondo, M. Sugawara, M. Takahashi, K. Kanda, T. Yokoi, T. Furuya, E. Kikkawa, Y. Omura, K. Abe, K. Kamihara, N. Katsuta, K. Sato, M. Tanikawa, M. Yamazaki, K. Ninomiya, T. Ishibashi, H. Yamashita, K. Murakawa, K. Fujimori, H. Tanai, M. Kimata, M. Watanabe, S. Hiraoka, Y. Chiba, S. Ishida, Y. Ono, S. Takiguchi, S. Watanabe, M. Yosida, T. Hotuta, J. Kusano, K. Kanehori, A. Takahashi-Fujii, H. Hara, T.O. Tanase, Y. Nomura, S. Togiya, F. Komai, R. Hara, K. Takeuchi, M. Arita, N. Imose, K. Musashino, H. Yuuki, A. Oshima, N. Sasaki, S. Aotsuka, Y. Yoshikawa, H. Matsunawa, T. Ichihara, N. Shiohata, S. Sano, S. Moriya, H. Momiyama, N. Satoh, S. Takami, Y. Terashima, O. Suzuki, S. Nakagawa, A. Senoh, H. Mizoguchi, Y. Goto, F. Shimizu, H. Wakebe, H. Hishigaki, T. Watanabe, A. Sugiyama, M. Takemoto, B. Kawakami, M. Yamazaki, K. Watanabe, A. Kumagai, S. Itakura, Y. Fukuzumi, Y. Fujimori, M. Komiyama, H. Tashiro, A. Tanigami, T. Fujiwara, T. Ono, K. Yamada, Y. Fujii, K. Ozaki, M. Hirao, Y. Ohmori, A. Kawabata, T. Hikiji, N. Kobatake, H. Inagaki, Y. Ikema, S. Okamoto, R. Okitani, T. Kawakami, S. Noguchi, T. Itoh, K. Shigeta, T. Senba, K. Matsumura, Y. Nakajima, T. Mizuno, M. Morinaga, M. Sasaki, T. Togashi, M. Oyama, H. Hata, M. Watanabe, T. Komatsu, J. Mizushima-Sugano, T. Satoh, Y. Shirai, Y. Takahashi, K. Nakagawa, K. Okumura, T. Nagase, N. Nomura, H. Kikuchi, Y. Masuho, R. Yamashita, K. Nakai, T. Yada, Y. Nakamura, O. Ohara, T. Isogai, S. Sugano, Complete sequencing and characterization of 21,243 full-length human cDNAs, *Nat.Genet.* 36 (2004) 40-45.
- [174] M. Oyama, C. Itagaki, H. Hata, Y. Suzuki, T. Izumi, T. Natsume, T. Isobe, S. Sugano, Analysis of small human proteins reveals the translation of upstream open reading frames of mRNAs, *Genome Res.* 14 (2004) 2048-2052.
- [175] J.E. Gonzalez-Pastor, J.L. San Millan, F. Moreno, The smallest known gene, *Nature* 369 (1994) 281.
- [176] S. Steigele, W. Huber, C. Stocsits, P.F. Stadler, K. Nieselt, Comparative analysis of structured RNAs in *S. cerevisiae* indicates a multitude of different functions, *BMC.Biol.* 5 (2007) 25.

- [177] C.D. Warden, S.H. Kim, S.V. Yi, Predicted functional RNAs within coding regions constrain evolutionary rates of yeast proteins, *PLoS.One.* 3 (2008) e1559.
- [178] J.S. Pedersen, G. Bejerano, A. Siepel, K. Rosenbloom, K. Lindblad-Toh, E.S. Lander, J. Kent, W. Miller, D. Haussler, Identification and classification of conserved RNA secondary structures in the human genome, *PLoS.Comput.Biol.* 2 (2006) e33.
- [179] A. Stark, M.F. Lin, P. Kheradpour, J.S. Pedersen, L. Parts, J.W. Carlson, M.A. Crosby, M.D. Rasmussen, S. Roy, A.N. Deoras, J.G. Ruby, J. Brennecke, E. Hodges, A.S. Hinrichs, A. Caspi, B. Paten, S.W. Park, M.V. Han, M.L. Maeder, B.J. Polansky, B.E. Robson, S. Aerts, H.J. van, B. Hassan, D.G. Gilbert, D.A. Eastman, M. Rice, M. Weir, M.W. Hahn, Y. Park, C.N. Dewey, L. Pachter, W.J. Kent, D. Haussler, E.C. Lai, D.P. Bartel, G.J. Hannon, T.C. Kaufman, M.B. Eisen, A.G. Clark, D. Smith, S.E. Celniker, W.M. Gelbart, M. Kellis, Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures, *Nature* 450 (2007) 219-232.

## Figure legends

### Figure 1. A historic view of non-coding RNA.

Indicated are some key dates in the discovery of non-coding RNA, from the first ribosomal RNA (rRNA) to the first description of the bifunctionality of SRA RNA, the founding member of this new category of RNA. SRA was first described as a non-coding RNA in 1999, as part of the steroid receptor coactivator-1 (SRC-1) complex. In 2003, the existence of an SRA-associated protein called SRAP was then reported.

### Figure 2. A model for transcriptional co-regulation of MyoD following changes in the SRA RNA/SRA Protein balance during myogenic differentiation.

(A) SRA1 gene products. SRA1 gene transcription leads to two major classes of isoforms, *i.e.* coding isoforms (blue/red ribbon) that preserve the ability to encode a protein SRAP (red puzzle), and non-coding isoforms that do not have an ATG start codon (blue ribbon) or that retain intron 1 (blue/green ribbon) disrupting the ORF. (B) Relationship between SRA expression, SRA function and myogenic differentiation. The balance between SRA-coding and SRA non-coding isoforms varies during myogenic differentiation in favor of SRA ncRNA. SRAP prevents the SRA RNA-dependent co-activation of MyoD (yellow puzzle) through the interaction of SRAP with its RNA counterpart. RRM, RNA Recognition Motif; STR7, RNA sub-structure; TSS, Transcriptional start sites.

### Figure 3. Schematic representation of mRNA primary and secondary structures, regrouping the classical features of a mRNA. The table on the bottom contains examples of ncRNA that preserve the classical features of a mRNA.

Light blue, the 7 methyl-guanosine (m7G) cap; green, the 5' and 3' untranslated regions (UTR); dark blue, Open Reading Frame (ORF) region; red, the polyadenylated (polyA) tail.

Stem loops represent mRNA secondary structures. Pol II, polymerase II-dependent transcription; Splicing, RNA processed by splicing; Editing, RNA modifications other than capping and polyadenylation.

**Figure 4.** Incidence of open reading frames (ORF) according to transcripts length in a ncRNA Dataset.

A database (from RNADB, <http://research.imb.uq.edu.au/rnadb/>) containing 940 unique ncRNA sequences which have been identified and manually curated based upon extensive literature reviews was used in this study [169]. (A) The longest ORF (**ORF max**, beginning with a methionine and finishing with a stop codon) predicted for each ncRNA sequence from the dataset was plotted against the length of the transcript. Each dot corresponds to a unique sequence from the database. The green curve represent incidences of randomly occurring ORF at 2 standard deviations from the mean [122], using the function  $y = 91 \cdot \ln(x) - 330$ , which has been proposed by **Dinger et al.** to discriminate non-coding from protein-coding transcripts in a transcript-length-dependent manner [122]. Even with these criterion, 326 of the 940 sequences data hold a potential ORF that does not occur by chance, with 95% confidence according to Dinger *et al.* [122]. The red line corresponds to the classical view of discrimination (from Fantom3, [167]) between coding and non-coding RNA, *i.e.* the ORF length with a cutoff of 300 nt (100 aa). (B) Cumulative frequency of sequences plotted against the ORF size. aa, amino acid; nt, nucleotide.