

Contents lists available at [ScienceDirect](http://ScienceDirect)

# Genomics

journal homepage: [www.elsevier.com/locate/ygeno](http://www.elsevier.com/locate/ygeno)

## Review

# Long intronic noncoding RNA transcription: Expression noise or expression choice?

Rodrigo Louro<sup>1</sup>, Anna S. Smirnova<sup>1</sup>, Sergio Verjovski-Almeida<sup>\*</sup>

Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, 05508-900 São Paulo, SP, Brazil

## ARTICLE INFO

### Article history:

Received 27 September 2008

Accepted 16 November 2008

Available online 18 December 2008

### Keywords:

Noncoding RNA

Intronic transcription

Transcriptional regulation

microRNA

## ABSTRACT

Recently, it was discovered that non-protein-coding RNAs (ncRNAs) represent the majority of the human transcripts. Regulatory role of many classes of ncRNAs is broadly recognized; however, long intronic ncRNAs have received little attention. In the past few years, evidence that intronic regions are key sources of regulatory ncRNAs has first appeared. Here we present an updated vision of the intronic ncRNA world, giving special attention to the long intronic ncRNAs. We summarize aspects of their expression pattern, evolutionary constraints, biogenesis, and responsiveness to physiological stimuli, and postulate their mechanisms of action. Deciphering nature's choice of different types of messages conveyed by ncRNAs will shed light on the RNA-based layer of regulatory processes in eukaryotic cells.

© 2008 Elsevier Inc. All rights reserved.

## Contents

|  |     |
|--|-----|
| Introduction . . . . .   | 291 |
| Features of long intronic ncRNAs expressed from mammalian genomes . . . . .      | 292 |
| Tissue-specific transcription and evolutionary conservation . . . . .            | 292 |
| Expression in subcellular compartments . . . . .                                 | 292 |
| Responsiveness to physiological stimuli . . . . .                                | 292 |
| Expression in cancer . . . . .   | 294 |
| Mechanisms of gene expression regulation by long intronic ncRNAs . . . . .       | 294 |
| Long intronic ncRNAs as precursors of shorter RNAs . . . . .                     | 294 |
| Interaction of ncRNAs with promoter elements and transcription factors . . . . . | 295 |
| Epigenetic control of gene expression . . . . .                                  | 295 |
| Regulation of protein-coding RNA alternative splicing . . . . .                  | 295 |
| Protein-coding RNA stabilization . . . . .                                       | 296 |
| Global gene expression regulation by trans-acting control . . . . .              | 296 |
| Long intronic ncRNA biogenesis . . . . .   | 296 |
| Could it be merely an artifact? . . . . .  | 296 |
| Conclusions and perspectives . . . . .   | 297 |
| Acknowledgments . . . . .  | 297 |
| References . . . . .   | 297 |

## Introduction

Recent improvements in high-throughput gene expression analysis have led to the discovery that noncoding RNAs (ncRNAs) represent by large the majority of the human transcriptional units [1–4]. Uncovering

this pervasive transcription has been followed by characterization of many novel functional regulatory ncRNAs, and therefore an update to the definition of a gene and a review of the molecular biology central dogma seems to be required [5,6]. Many classes of ncRNAs are now extensively studied, and their regulatory role is broadly recognized. The miRNAs represent a classical example of well-known ncRNA molecules that perform regulation at the RNA level [7]. Other classes of short RNAs, such as piRNA and snoRNA, have also attracted wide interest [8–10], and new classes of short RNAs, such as XIST-derived RNAs (xiRNAs), are now being considered [11]. Likewise, long regulatory intergenic ncRNAs such as SRA (SRA1) [12] and HOTAIR [13] have been characterized. On the other hand,

<sup>\*</sup> Corresponding author. Universidade de São Paulo, Instituto de Química, Departamento de Bioquímica, Av. Prof. Lineu Prestes 748 sala 1200, 05508-000 São Paulo, SP, Brazil. Fax: +55 11 3091 2186.

E-mail address: [verjo@iq.usp.br](mailto:verjo@iq.usp.br) (S. Verjovski-Almeida).

<sup>1</sup> These authors have contributed equally to the work.

long ncRNAs transcribed from intronic regions have received less attention from the scientific community, possibly due to the biased perception that all of them might result from immature RNAs, a view that is now changing, as discussed in the present review. Despite the fact that there are a number of reviews concerning long ncRNAs regulating the transcriptional machinery [14–17], none focuses on those transcribed from introns.

In 2005, our group surveyed the main findings of the large-scale expression analysis projects that led to the identification of independently transcribed intronic ncRNAs [18]. In most cases, information about this class of ncRNAs was hidden among high-throughput generated data, concerning different types of unknown transcripts. Using *in silico* approaches based on genomic mapping and clustering of ESTs, together with microarray experiments using combined intron/exon oligoarrays we were able to point to intronic regions as key sources of potentially regulatory ncRNAs [19,20]. The up-to-date catalog of human and mouse intronic transcription described 78,147 and 39,660 long EST contigs, respectively, whose majority should represent novel ncRNAs [20]. In addition, our work revealed that about 81% and 70% of all spliced human and mouse protein-coding genes, respectively, have transcriptionally active introns [20]. The last few years were marked by functional characterization of some interesting intronic ncRNAs, and it has been proposed that a number of steps in the protein-coding RNA processing pathway may be regulated by ncRNAs (Fig. 1), as detailed below. In the present review we discuss the increasing knowledge about the intronic ncRNA world.

### Features of long intronic ncRNAs expressed from mammalian genomes

#### *Tissue-specific transcription and evolutionary conservation*

Unbiased efforts that surveyed the human transcriptional output have shown that novel unannotated transcripts, for which more than 30% reside within intronic regions, appear to be more tissue-specific than well-characterized protein-coding RNAs [1,21]. In addition, two recent papers focusing on intronic large-scale transcription have reported similar results. Tissue-specific expression signatures of 1417 intronic ncRNAs were observed in prostate, kidney and liver, using a gene-oriented combined intron/exon expression array platform [19]. Likewise, using *in situ* hybridization, 182 intronic ncRNAs were associated with specific neuroanatomical regions, cell types, or subcellular compartments of adult mouse brain [22].

Expression of ncRNAs seems to be under diverse levels of evolutionary constraints in mammals. In general, small ncRNAs such as miRNAs seem to be highly conserved, while longer transcripts are less conserved than typical exons [10,23]. *Xist*, responsible for guiding X chromosome inactivation, and *Air*, involved in mouse imprinted gene-silencing at the *Igf2r* locus, are examples of poorly conserved long ncRNAs [23]. Probably the most extreme example that shows lack of primary sequence conservation in long ncRNAs is *HAR1* (*HAR1A*), part of a novel RNA gene that is expressed specifically in Cajal–Retzius neurons in the developing human neocortex [24]. *HAR1* is the acronym for “Human Accelerated Region 1”, and it was identified as one genomic region that evolved rapidly in humans as a result of a search for transcripts that were unique in human biology [24]. In contrast, some long ncRNAs exhibit an unexpected high level of nucleotide sequence conservation (>60%) in mammals, such as *MALAT1* an intergenic long ncRNA involved in cancer and its murine ortholog hepcarcin (*Malat1*) [25], and the *Dnm3os* intronic ncRNA, differentially expressed during embryogenesis [26].

Recent reports have added other exciting evolutionary aspects to the long intronic ncRNAs expression, using cross-species microarray hybridization. A study on the expression of intergenic noncoding sequences in human and chimpanzee showed tissue-specific conservation of expression of ncRNAs in equivalent genomic loci, but no

conservation of the nucleotide sequences [27]. Similar results were obtained for intronic sequences in humans and mice by our group. Using a cross-species microarray hybridization approach, we identified a set of 22 long intronic ncRNAs, expressed from syntenic loci in humans and mice. These long intronic ncRNAs showed very similar patterns of tissue-specific expression in both species in prostate, kidney or liver tissues [20]. Surprisingly, these tissue-specific ncRNAs map to intronic regions that have only short stretches of sequences with evidence of identity conservation [20]. With an *in silico* approach, pyknons were recently described, which by definition are recurrent motifs (>15 bases; >29 copies) present in intergenic and intronic regions with at least one additional copy in an exonic region [28]. In both human and mouse genomes pyknons were found to be over-represented in the introns of protein-coding genes belonging to the same set of biological processes and molecular functions, even though the underlying sequences were not conserved between the two genomes [29]. Another report, concerning 945 ncRNAs (338 intronic ncRNAs) expressed during embryonic stem cell differentiation in mouse also contributed to describe evolutionary constraints that govern ncRNAs expression [30]. The authors found that ncRNAs expressed during differentiation were enriched for predicted RNA secondary structures relative to the genome average [30]. In fact, more than 30,000 structured RNA elements had already been predicted in the human genome, almost 1000 of which are conserved across all vertebrates; interestingly, a third of them are found in introns of known genes [31].

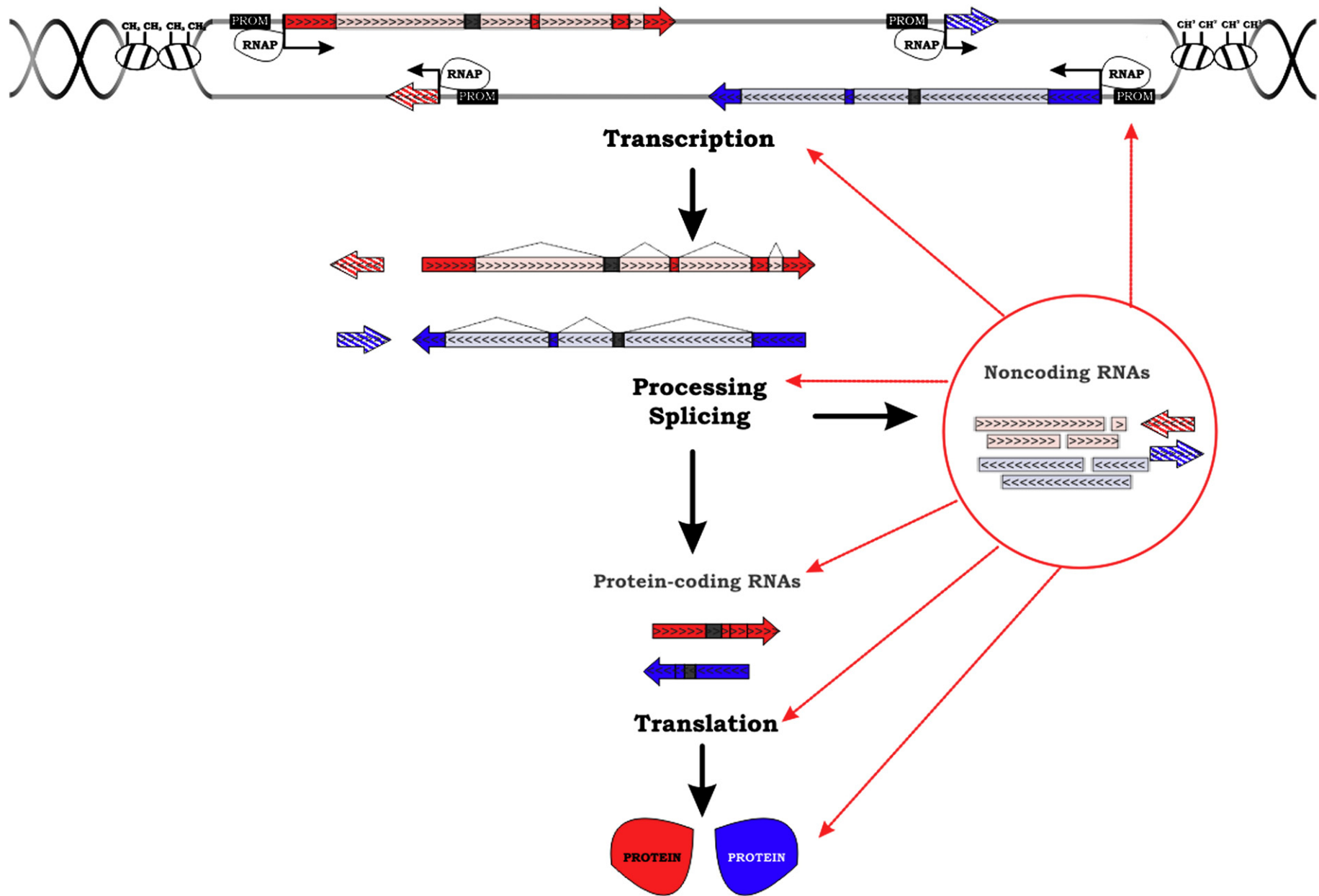
These findings give additional support to the previous suggestion that longer ncRNAs are under the influence of evolutionary constraints different from those of miRNAs and snoRNAs, and that the lack of primary sequence conservation does not necessarily signify an absence of function [23]. Primary sequence conservation among species might not be important for some cis acting mechanisms of regulation, such as the case of transcriptional interference or activation through transcription across regulatory regions [17]. Sequence conservation might be relevant for ncRNAs likely to work in trans, when secondary structure is a requirement for these ncRNAs to bind at RNA-binding protein targets in order to exert their cellular functions. Recently, an intergenic ncRNA named *HOTAIR* has been shown to bind to a polycomb group factor in order to regulate the expression of homeobox genes [13], however no particular secondary structure could be predicted in this case. A systematic search for secondary structure conservation throughout the intronic ncRNA dataset is warranted, and is likely to point to conserved motifs.

#### *Expression in subcellular compartments*

Alongside tissue specificity, intronic ncRNAs transcription seems to be spatially restricted by subcellular expression preferences. Once again, unbiased analyses have already indicated that a significant proportion of unannotated ncRNAs are exclusively detected in nuclear or cytoplasmic cellular extracts [5,32]. Intronic ncRNAs expression seems to be predominantly nuclear; however, some subsets were primarily detected in the cytoplasm, and only a few seem to be equally expressed in both compartments [33].

#### *Responsiveness to physiological stimuli*

To exert their functions on gene-expression control, it is expected that intronic ncRNA expression would be responsive to physiological stimuli. Indeed, cell line treatment with androgen hormone significantly alters the expression of a subset of intronic ncRNAs [34]. Similarly, exposure of two cell lines to retinoic acid altered the expression of ncRNAs that lie downstream from intronic transcription-factor binding sites for Sp1 (*SP1*), c-Myc (*MYC*) and p53 (*TP53*) [35]. Either by direct activation [34], or indirectly via proteinic second messengers [34,35], intronic ncRNA transcription responds to extracellular environment modifications.



**Fig. 1.** Schematic view of the genetic information flow, highlighting the roles played by intronic non-protein-coding RNAs (ncRNAs). Thin arrows (red) represent regulation performed by intronic ncRNAs at different steps of the gene expression pathway. Dark (red and blue) solid boxes and arrows represent constitutive exons of protein-coding transcripts. Solid (grey) boxes represent alternatively spliced exons. Introns are symbolized by light (red or blue) solid boxes. Independently transcribed intronic ncRNAs are shown as dashed (red or blue) arrows. Black solid boxes represent promoter regions (PROM). Transcriptional direction is oriented by arrowheads. Proteins were drawn as roundish forms (red, blue or white). RNAP – RNA polymerase. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

## Expression in cancer

Oncogenes and tumor suppressor genes such as N-myc (*MYCN*) and p53 (*TP53*) were identified a long time ago as having ncRNAs transcribed from their intronic regions [36,37], although a correlation of these ncRNAs to cancer has not been studied. More recently, aberrant expression of intronic ncRNAs has been correlated to cancer development and progression. Expression signatures that distinguish different degrees of neoplastic tissue differentiation in human prostate cancer included a significant number of intronic ncRNAs [38]. Also in renal cell carcinoma, an expression signature which includes intronic ncRNAs allows discrimination between tumor and adjacent normal tissues [39]. Experiments with whole-genome tiling arrays, searching for long, conserved, abundantly expressed ncRNAs, revealed 15 transcripts whose expression was altered in breast and ovarian cancer; at least three of them are intronic [40]. An intergenic long 8.6 kb ncRNA named *MALAT1* has been associated with metastasis in human lung adenocarcinomas [41] and was shown to be overexpressed in five other types of human cancers [25]. It is now being recognized that long intronic ncRNAs are cancer biomarkers [42], and further efforts are underway in our group to use intronic ncRNA expression profiles to foresee metastasis and tumor recurrence in different types of cancer.

## Mechanisms of gene expression regulation by long intronic ncRNAs

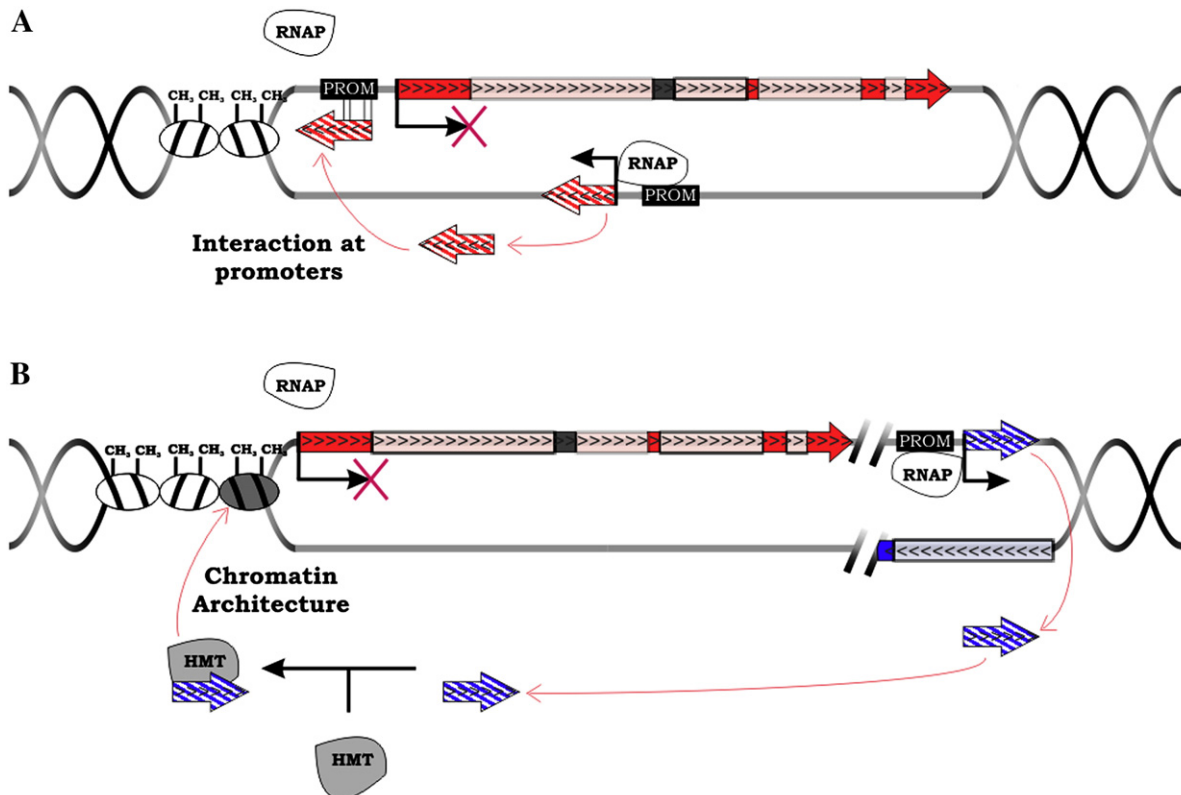
The widespread occurrence, tissue and subcellular expression specificity, evolutionary conservation, environment alteration responsiveness and aberrant expression in human cancers are features that accredit intronic ncRNAs to be mediators of gene expression regulation.

According to current knowledge, major control of gene expression is mediated by protein-coding RNAs. A few large-scale studies showed that certain sets of intronic ncRNAs have the same tissue expression pattern as the corresponding protein-coding genes, whereas others are inversely correlated [19,22,30,43]. These findings point to complex regulatory relationships between intronic ncRNAs and their host loci. It has been postulated that ncRNAs can act transcriptionally or post-transcriptionally; however, mechanisms that underlie such regulatory behaviors still remain to be fully understood. Below we discuss some scenarios for gene expression regulation by long intronic ncRNAs.

### Long intronic ncRNAs as precursors of shorter RNAs

It was initially thought that miRNAs are encoded mostly in intergenic regions, however more recent data revealed that at least one-fourth of mammalian miRNA loci are located within intron regions and are transcribed by RNA polymerase II [44–46]. Some snoRNAs also are encoded within intronic regions [9,47,48]. Interestingly, a recent work showed that both human and mouse intronic miRNAs tend to be present in large introns with 5'-biased position distribution [48], what correlates with the previous observation that most long intronic transcripts are expressed within first introns of the host genes [19]. The 5'-biased positions of miRNA host introns may be necessary for the transcription and regulation of intronic miRNAs to utilize the regulatory signals within the 5'-UTRs of their host genes [48]. We expect that a number of long intronic ncRNAs are processed into smaller ncRNAs to exert their cellular functions, according to previously described mechanisms [7,9].

In this regard, a recent report opens a possible new perspective for functional studies of long ncRNAs [49]. This study described a very



**Fig. 2.** Postulated mechanisms of gene expression control by intronic ncRNAs at the transcriptional level. (A) Transcriptional interference through interaction at promoters. (B) Epigenetic control of gene-expression through modulation of chromatin architecture. Dark (red and blue) solid boxes and arrows represent constitutive exons of protein-coding transcripts. Solid (grey) boxes represent alternatively spliced exons. Introns are symbolized by light (red or blue) solid boxes. Independently transcribed intronic ncRNAs are shown as dashed (red or blue) arrows. Black solid boxes represent promoter regions (PROM). Transcriptional direction is oriented by arrowheads. Proteins are drawn as roundish forms (red, blue, white or grey). RNAP – RNA polymerase. HMT – histone methyl transferase. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

long (~35,000 nt) antisense ncRNA that encompasses the entire *p15* (*CDKN2B*) genomic locus (*p15AS*), spanning the intronic region [49]. In this work the authors proved that just a fragment of *p15AS* that overlaps the first exon of *p15* protein-coding gene is sufficient to control its chromatin architecture [49]. We envisage that functional studies similar to that described for *p15AS* [49] can be conducted by overexpression of only a portion of other long ncRNAs in analysis, irrespective of the fact that a shorter ncRNA is or is not endogenously produced in a cell. However, it remains to be determined if the effects observed when using just a fragment of the ncRNA [49] are similar to those exerted by the full-length ncRNAs, and which regions are essential to their functions. In this context, it should be noted that Dicer (*DICER1*) was not required for the antisense ncRNA to exert its regulatory function [49].

#### Interaction of ncRNAs with promoter elements and transcription factors

Intronic sense and antisense ncRNAs may repress a neighboring protein-coding gene through a phenomenon termed transcriptional interference (Fig. 2A), which prevents initiation complex recruitment or transcriptional elongation [50]. In humans, there is evidence that a partially intronic ncRNA, produced from the genomic locus encoding dihydrofolate reductase (*DHFR*), directly interacts with the major promoter, decreasing the expression of the protein-coding RNA [51]. Another recent work showed that long spliced antisense transcripts, overlapping the promoter of the progesterone receptor gene (*PGR*), are necessary for activation of *PGR* expression [52]. Small duplex RNAs, called antigene RNAs, are complementary to promoter sequences and could activate or repress gene expression, probably through binding these long spliced antisense transcripts [52,53]. It seems that some endogenous miRNAs are also complementary to promoters and, therefore, could be involved in these mechanisms [52].

Several reports have suggested a model in which ncRNAs could function as molecular adaptors, providing increased specificity in gene expression control by guiding the RNA/DNA-binding proteins to promoter regions. The *DHFR* protein-coding RNA repression involves not only the formation of a stable complex between ncRNA and the major promoter, but it is also accompanied by a direct interaction of the ncRNA with the general transcription factor IIB and the dissociation of the pre-initiation complex from the major promoter [51]. Another example is the gene-specific repression of cyclin D1 (*CCND1*) in human cell lines, caused by recruitment of the translocated in liposarcoma (TLS) RNA-binding protein (*FUS*). The TLS protein is directed to the *CCND1* promoter by long (>200 nt) ncRNAs, and tethered to its 5' regulatory region in response to DNA damage signals [54].

Interestingly, taking the full repertoire of human long ncRNAs, a higher proportion of transcripts are mapped to the first introns of protein-coding genes [19,33], a region relatively close to the promoter. They could act through *DHFR*-like mechanisms [51], or be precursors of promoter-complementary miRNAs [52].

#### Epigenetic control of gene expression

Only lately modifications on the chromatin architecture that regulate gene expression – epigenetic regulation – have been linked to ncRNAs (Fig. 2B). The most prominent examples involving ncRNAs include the trimethylation of histone H3 on lysine 27 by Polycomb group proteins (PcG). Transcription of intergenic *XIST* ncRNA from one of the two female X chromosomes is involved in recruiting PcG to normalize the copy number of X chromosomes between male and female cells [55]. Likewise, the expression of homeobox genes that encode key development regulators in the embryo could be repressed by an RNA–protein complex of intergenic *HOTAIR* ncRNA and PcG repressive complex 2 [13].

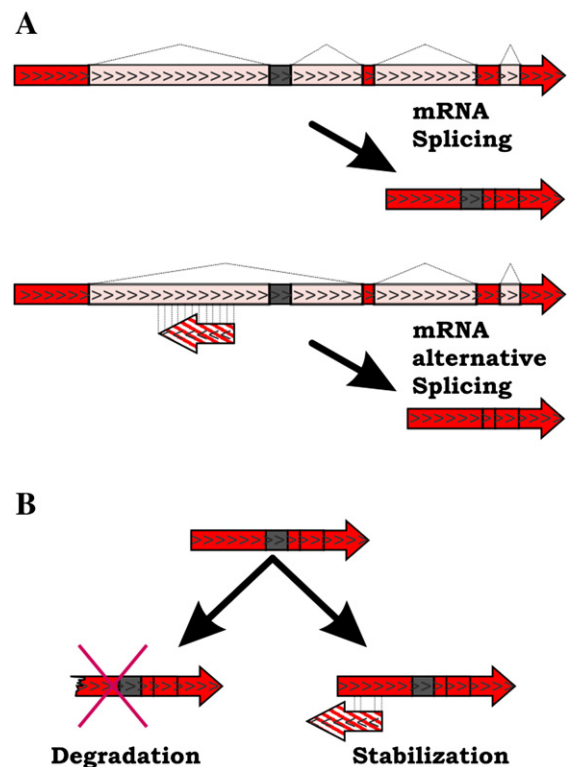
However, to our knowledge, until now there is only one description of a naturally occurring long intronic ncRNA that acts as

an epigenetic modifier. Endogenous long antisense ncRNA *KHPS1* is originated from the CpG island of sphingosine kinase-1 gene (*SPHK1*), and it overlaps a sense regulatory element named tissue-dependent differentially methylated region (T-DMR) [56]. Overexpression of two *KHPS1* fragments caused DNA demethylation of T-DMR. Intriguingly, this intronic transcript is also transported to the cytoplasm, suggesting that the same transcript could be involved in the post-transcriptional regulation mechanisms [56].

#### Regulation of protein-coding RNA alternative splicing

Control of alternative pre-mRNA splicing is another layer of gene expression regulation to which intronic ncRNAs are demonstrated to be related (Fig. 3A). About two decades ago, it had already been recognized that an antisense RNA, transcribed from the first intron of N-myc genomic locus, could modulate RNA splicing by forming RNA–RNA duplexes and preserving a population of N-myc mRNA retaining intron 1 [36]. More recently, it was shown that overexpression of *SAF*, a 1,500 nt partially intronic ncRNA, transcribed from the opposite strand of intron 1 of the human *FAS* gene, caused functionally important alterations in *FAS* alternative splicing [57]. Expression of *SAF* ncRNA did modulate the expression of different Fas protein soluble forms, making cells more resistant to Fas-mediated apoptosis [57].

Alternative splicing events must be strictly regulated to properly encode diverse, but functional proteins [58], and the average frequency of exon-skipping for exons overlapped by or located immediately 3' to the intronic ncRNAs is higher than the average frequency of exon-skipping in the overall set of human RefSeq RNAs



**Fig. 3.** Postulated mechanisms of gene expression control by intronic ncRNAs at post-transcriptional level. (A) Regulation of protein-coding RNA alternative splicing. (B) Protein-coding RNA stabilization. Dark (red) solid boxes and arrows represent constitutive exons of protein-coding transcripts. Solid (grey) boxes represent alternatively spliced exons, and light (red) boxes represent introns. Independently transcribed intronic ncRNAs are shown as dashed (red) arrows. Transcriptional direction is oriented by arrowheads. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

[19]; it is tempting to speculate that this sort of regulatory relation is probably more common than anticipated.

#### Protein-coding RNA stabilization

Another possibility of control by ncRNAs is that intronic ncRNAs could play their regulatory roles via stabilization of protein-coding RNA, transcribed from the same genomic loci (Fig. 3B). Many unstable mammalian protein-coding RNAs contain adenylate- and uridylate-rich (AU-rich) elements that determine their half-life [59]. In B-cell lymphomas that exhibit the 14;18 chromosome translocation, a bc1-2/IgH antisense ncRNA contributes to the upregulation of bcl-2 (*BCL2*) gene expression, probably by masking AU-rich motifs present in the 3' UTR of the bcl-2 protein-coding RNA [59]. Although no direct evidence has been obtained for the involvement of naturally occurring long intronic ncRNAs in such a mechanism, the concordant expression profiles between protein-coding and intronic ncRNAs, observed in genome-wide expression analysis, suggest that a gene expression control by protein-coding RNA stabilization is plausible in normal cells [19,30,34,35].

#### Global gene expression regulation by trans-acting control

Mechanisms of gene expression regulation by intronic ncRNAs, previously discussed, are illustrated through examples related to locus-specific regulation. However, introns of protein-coding genes were postulated to contain signals that could contribute to the complexity of gene expression regulation networks [60]. ncRNAs fulfill the essential conditions for being system connectivity and multi-tasking agents, and this could allow them to be control molecules that mediates gene–gene communication [61]. Only recently some experimental evidence of these early conjectures was reported.

Hill et al. [62] demonstrated that complete human introns are capable of coordinating expression of functionally related genes. They overexpressed three long intronic sequences from the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene in epithelial cells (HeLa), in which *CFTR* is not normally expressed. Surprisingly, they observed that the expression of the *CFTR* introns caused extensive and specific transcriptional changes, affecting mainly genes linked to *CFTR* function. Affected genes were distributed at spatially diverse sites within the genome. Each of three intronic sequences induced unique and highly reproducible changes in gene expression profile. Since these transfected cells do not express the *CFTR* protein-coding transcript, observed effects were certainly caused by the intronic sequences. Because all three intronic sequences do not include any known miRNAs or predicted stem–loop structures, they seem to act in trans as long ncRNA regulatory elements. In conclusion, Hill et al. [62] reported the first evidence that long intronic sequences could coordinate waves of gene expression important for particular cellular processes, functionally related to the protein-coding transcript of the same locus.

Interestingly, a recent work showed that intron-derived miRNA also could be involved in a similar type of control, silencing a cohort of genes that are functionally antagonistic to the host gene itself [63]. Transcriptional activation of apoptosis-associated tyrosine kinase (*AATK*), essential for neuronal differentiation, also generates intronic miR-338 that silences a family of mRNAs, related to the negative regulation of neuronal differentiation [63].

The fact that even non-degraded spliced introns can be selectively exported to the cytoplasm [64] and be involved in global gene expression regulation [62] makes even more intriguing the question about the function of independently transcribed intronic sequences.

Long intronic ncRNAs, or shorter RNAs originated from the former, could regulate global gene expression using mechanisms previously described for mammalian ncRNAs, such as binding to promoters and transcription factors [51], recruiting of Polycomb Repressive Complex 2 [13] or other RNA-binding proteins [12,54]. Interestingly, many of

these mechanisms are associated with transcription repression. A large number of genes up-regulated after overexpression of intronic sequences [62] could be explained by both guiding activation factors and indirect targeting through inhibition of suppressors.

#### Long intronic ncRNA biogenesis

The biogenesis of long intronic ncRNA is poorly understood. An ample involvement of RNA polymerase II (RNAP II) on the transcription of ncRNAs has been presumed because of indirect evidence such as (1) concordant and co-regulated expression profiles of many intronic ncRNAs and their corresponding protein-coding genes [19,30,34,35]; (2) the presence of poly(A+) tail [2,3,5,19,21]; and also (3) the broad contribution of RNAP II-associated transcription factors and physiological stimuli in the transcription of intronic ncRNAs [34,35]. Such observations do not completely exclude the possibility that some intronic transcripts could represent unexpectedly long-lasting lariats that could be processed post-splicing as well.

Long intronic ncRNAs have been extensively identified in poly(A-) RNA populations [2,3,5,19,21] by using approaches that assume the RNAP II-transcript poly(A) tail occurrence, such as oligo-d(T)-primed reverse transcription. The difficulties to perform large-scale measurements of poly(A-) ncRNA expression add a bias towards the detection of poly(A+) intronic ncRNAs; nevertheless, poly(A-) RNAs seem to represent an important fraction of human transcriptional output [5,32], and large-scale sequencing of the poly(A-) RNA fraction [32] will probably open the door for exploring the pattern of expression of this class of RNAs.

Enhancing the complexity of this picture, our group has reported that over 10% of long intronic poly(A+) ncRNAs are up-regulated after treatment with the RNAP II specific inhibitor  $\alpha$ -amanitin [19], whereas only 4% of protein-coding transcripts are up-regulated under the same conditions [19,65]. These findings suggest that some intronic ncRNA and peculiar protein-coding RNAs could be transcribed by another RNA polymerase such as the recently described spRNAP-IV, whose transcriptional output seems to be enhanced by  $\alpha$ -amanitin [65], or also could be transcribed by RNAP III [46,66]. Nonetheless,  $\alpha$ -amanitin cell treatment also evoked mRNA accumulation of some well studied RNAP II transcribed protein-coding genes, such as p53 [67] and TNF-alpha (*TNF*) [68]. A report demonstrated that treatment with  $\alpha$ -amanitin could influence phosphorylation of RNAP II CTD (C-terminal domain) and, through this, induce enhanced elongation of some protein-coding transcripts [69]. This suggests that the previously cited results of apparent up-regulation of noncoding and protein-coding RNAs could represent a less explored effect of  $\alpha$ -amanitin on RNAP II transcription elongation.

Interestingly, two *in silico* identified totally intronic ncRNAs (TIN 52044 and 52045) [19] map to a very long (~35,000 nt) recently described antisense ncRNA that encompasses the entire *p15* genomic locus (*p15AS*) [49]. Such observation means that the present level of coverage of intronic ncRNA transcription probably results in partial identification of many longer messages. Undoubtedly, further studies are warranted to characterize full-length intronic messages and convincingly unveil the biogenesis of long intronic ncRNAs. Systematic search for the presence of characteristic RNAP II-derived attributes in long intronic ncRNAs, a deeper study of poly(A-) RNA populations and the investigation of possible involvement of other polymerases could provide new information that would certainly clarify this subject.

#### Could it be merely an artifact?

Experimental artifacts have been widely discussed as the main contributors in the detection of pervasive intronic transcription. The most commented are contamination of RNA samples with residual genomic DNA (gDNA) or immature unspliced RNAs [4], and annealing of oligo-dT primers to poly(A) repeats instead of to poly(A) tails [70].

In fact, even abundant DNase treatment is not sufficient for complete elimination of contaminant gDNA [4,71]. Controls for gDNA contamination, omitting reverse transcription step, are widely applied for checking sample quality [4,34,38,72]. Even in samples having residual gDNA contamination, mispriming with oligo-dT from adenine-rich DNA regions is not so common due to the fact that the RNA-dependent DNA-polymerase activity of reverse transcriptase is more efficient than the DNA-dependent one [73]. However, mispriming could occur to adenine-rich regions of RNA transcripts, producing a cluster of multiple, shorter cDNAs from longer parental transcripts [74]. Mispriming of RNA evokes the problem of full-length transcript characterization, but is not a source of false positives.

Many groups have detected intronic ncRNAs in the cytoplasmic fraction [5,21,22,33] whose contamination with gDNA and unprocessed mRNA should be minimal. A number of loci show discordant expression profiles between ncRNAs and their associated protein-coding genes [19,22]. Even in cases with apparently concordant expression profiles, sense and antisense transcripts may not be expressed in the exactly same cell or may have different compartmentalization [75]. In addition, many long ncRNAs were confirmed by Northern blot [34,36–38,40,57]. Finally, all the specific features of ncRNA expression, discussed previously, make it improbable that detection of ncRNAs is an artifact originated by random contamination with gDNA, immature transcripts or mispriming.

The early suggestions about participation of intronic sequences in the regulation of gene expression have considered as the main actors the stable intronic sequences derived from pre-mRNAs [60,61], but no independently transcribed products. Currently, there is evidence of independent promoters and transcription factor binding sites within intronic sequences [35,76,77]; nonetheless, so far no particular intronic ncRNA promoter has been characterized in detail. At this point, one cannot exclude the possibility that intronic RNAs, processed post-splicing, also participate in RNA-mediated transactions within the cell [61]. Concordant expression profiles for the majority of intronic ncRNA-protein-coding mRNA pairs [19,22] indicate that at least some of them could be processed from the same pre-mRNA. However, because these ncRNAs are stable and most likely functional, one cannot consider them merely artifacts from unprocessed RNA, but rather intronic ncRNAs produced from a common protein-coding RNA precursor (Fig. 1).

## Conclusions and perspectives

Studies of gene expression in eukaryotes have begun to rediscover the RNA world and its direct relation with structural, physiological, and behavioral complexity of organisms. As listed above, recent studies have provided significant contributions to the characterization of intronic ncRNAs as the largest portion of the human genome transcriptional units, pointed to mechanisms of action during gene expression control and investigated aspects of their biogenesis and responsiveness to physiological stimuli. Such gain of knowledge might provide answers to the molecular basis of eukaryotic complex attributes, which do not seem to rely solely on the number of expressed proteins.

The present scenario argues that evolutionary complexity should be a matter of novel and complex interactions using relatively few and fairly similar proteinic components, whose expression is temporally and spatially controlled by ncRNA networks. Until recently, the complexity level of the genetic programming in higher organisms was underestimated, merely pointing to proteins as the regulatory players of gene expression. An updated view of the molecular biology central dogma seems to be in order, highlighting the role of an RNA-based layer of regulation, and summarizing the possible mechanisms of action of intronic ncRNAs. It is envisaged that a deeper transcript sequencing of different cell and tissue types under physiological and pathological conditions using the next-generation sequencers will probably reveal in the near future the wide diversity and the peculiar

specificities of the human transcriptome. Deciphering nature's choice of different types of messages, conveyed by these ncRNAs, through the identification of their most relevant expression patterns and mechanisms of action, is the greatest challenge lying ahead of us.

## Acknowledgments

Work in the authors' laboratory is supported by a grant and fellowships from Fundação de Amparo a Pesquisa do Estado de São Paulo, and by Conselho Nacional de Desenvolvimento Científico e Tecnológico. We thank Carlos F.M. Menck (Universidade de São Paulo) for critically reading the manuscript.

## References

- [1] E. Birney, et al., Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project, *Nature* 447 (2007) 799–816.
- [2] P. Bertone, et al., Global identification of human transcribed sequences with genome tiling arrays, *Science* 306 (2004) 2242–2246.
- [3] J.L. Rinn, et al., The transcriptional activity of human Chromosome 22, *Genes Dev.* 17 (2003) 529–540.
- [4] J.M. Johnson, S. Edwards, D. Shoemaker, E.E. Schadt, Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments, *Trends Genet.* 21 (2005) 93–102.
- [5] J. Cheng, et al., Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution, *Science* 308 (2005) 1149–1154.
- [6] M.B. Gerstein, et al., What is a gene, post-ENCODE? History and updated definition, *Genome Res.* 17 (2007) 669–681.
- [7] W. Filipowicz, S.N. Bhattacharyya, N. Sonenberg, Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Rev. Genet.* 9 (2008) 102–114.
- [8] A.A. Aravin, G.J. Hannon, J. Brennecke, The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race, *Science* 318 (2007) 761–764.
- [9] T. Hirose, et al., A spliceosomal intron binding protein, IBP160, links position-dependent assembly of intron-encoded box C/D snoRNP to pre-mRNA splicing, *Mol. Cell* 23 (2006) 673–684.
- [10] H. Kawaji, et al., Hidden layers of human small RNAs, *BMC Genomics* 9 (2008) 157.
- [11] Y. Ogawa, B.K. Sun, J.T. Lee, Intersection of the RNA interference and X-inactivation pathways, *Science* 320 (2008) 1336–1341.
- [12] R.B. Lanz, et al., A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex, *Cell* 97 (1999) 17–27.
- [13] J.L. Rinn, et al., Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs, *Cell* 129 (2007) 1311–1323.
- [14] I. Shamovsky, E. Nudler, Gene control by large noncoding RNAs, *Sci. STKE* 2006 (2006) pe40.
- [15] O. Yazgan, J.E. Krebs, Noncoding but nonexpendable: transcriptional regulation by large noncoding RNA in eukaryotes, *Biochem. Cell. Biol.* 85 (2007) 484–496.
- [16] C. Barrandon, B. Spiluttini, O. Bensaude, Non-coding RNAs regulating the transcriptional machinery, *Biol. Cell* 100 (2008) 83–95.
- [17] P. Carninci, J. Yasuda, Y. Hayashizaki, Multifaceted mammalian transcriptome, *Curr. Opin. Cell Biol.* 20 (2008) 274–280.
- [18] E.M. Reis, R. Louro, H.I. Nakaya, S. Verjovski-Almeida, An antisense RNA gets intronic, *Omics* 9 (2005) 2–12.
- [19] H.I. Nakaya, et al., 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 (2007) R43.
- [20] R. Louro, T. El-Jundi, H.I. Nakaya, E.M. Reis, S. Verjovski-Almeida, Conserved tissue expression signatures of intronic noncoding RNAs transcribed from human and mouse loci, *Genomics* 92 (2008) 18–25.
- [21] D. Kampa, et al., Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22, *Genome Res.* 14 (2004) 331–342.
- [22] T.R. Mercer, M.E. Dinger, S.M. Sunkin, M.F. Mehler, J.S. Mattick, Specific expression of long noncoding RNAs in the mouse brain, *Proc. Natl. Acad. Sci. U. S. A.* 105 (2008) 716–721.
- [23] K.C. Pang, M.C. Frith, J.S. Mattick, Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function, *Trends Genet.* 22 (2006) 1–5.
- [24] K.S. Pollard, et al., An RNA gene expressed during cortical development evolved rapidly in humans, *Nature* 443 (2006) 167–172.
- [25] R. Lin, S. Maeda, C. Liu, M. Karin, T.S. Edgington, A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas, *Oncogene* 26 (2007) 851–858.
- [26] D.A. Loebel, B. Tsoi, N. Wong, P.P. Tam, A conserved noncoding intronic transcript at the mouse Dnm3 locus, *Genomics* 85 (2005) 782–789.
- [27] P. Khaitovich, et al., Functionality of intergenic transcription: an evolutionary comparison, *PLoS Genet.* 2 (2006) e171.
- [28] I. Rigoutsos, et al., Short blocks from the noncoding parts of the human genome have instances within nearly all known genes and relate to biological processes, *Proc. Natl. Acad. Sci. U. S. A.* 103 (2006) 6605–6610.
- [29] A. Tsirigos, I. Rigoutsos, Human and mouse introns are linked to the same processes and functions through each genome's most frequent non-conserved motifs, *Nucleic Acids Res.* 36 (2008) 3484–3493.

- [30] M.E. Dinger, et al., Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation, *Genome Res.* 18 (2008) 1433–1445.
- [31] S. Washietl, L.L. Hofacker, M. Lukasser, A. Huttenhofer, P.F. Stadler, Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome, *Nat. Biotechnol.* 23 (2005) 1383–1390.
- [32] Q. Wu, et al., Poly A<sup>+</sup> transcripts expressed in HeLa cells, *PLoS ONE* 3 (2008) e2803.
- [33] P. Kapranov, et al., RNA maps reveal new RNA classes and a possible function for pervasive transcription, *Science* 316 (2007) 1484–1488.
- [34] R. Louro, et al., Androgen responsive intronic non-coding RNAs, *BMC Biol.* 5 (2007) 4.
- [35] S. Cawley, et al., Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs, *Cell* 116 (2004) 499–509.
- [36] G.W. Krystal, B.C. Armstrong, J.F. Battey, N-myc mRNA forms an RNA–RNA duplex with endogenous antisense transcripts, *Mol. Cell. Biol.* 10 (1990) 4180–4191.
- [37] S. Khochbin, J.J. Lawrence, An antisense RNA involved in p53 mRNA maturation in murine erythroleukemia cells induced to differentiate, *EMBO J.* 8 (1989) 4107–4114.
- [38] E.M. Reis, et al., Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer, *Oncogene* 23 (2004) 6684–6692.
- [39] G.C. Brito, et al., Identification of protein-coding and intronic noncoding RNAs down-regulated in clear cell renal carcinoma, *Mol. Carcinog.* 47 (2008) 757–767.
- [40] D.S. Perez, et al., Long, abundantly expressed non-coding transcripts are altered in cancer, *Hum. Mol. Genet.* 17 (2008) 642–655.
- [41] P. Ji, et al., MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer, *Oncogene* 22 (2003) 8031–8041.
- [42] M. Mallardo, P. Poltronieri, O.F. D'Urso, Non-protein coding RNA biomarkers and differential expression in cancers: a review, *J. Exp. Clin. Cancer Res.* 27 (2008) 19.
- [43] S. Katayama, et al., Antisense transcription in the mammalian transcriptome, *Science* 309 (2005) 1564–1566.
- [44] A. Rodriguez, S. Griffiths-Jones, J.L. Ashurst, A. Bradley, Identification of mammalian microRNA host genes and transcription units, *Genome Res.* 14 (2004) 1902–1910.
- [45] Y.K. Kim, V.N. Kim, Processing of intronic microRNAs, *EMBO J.* 26 (2007) 775–783.
- [46] S.C. Li, P. Tang, W.C. Lin, Intronic microRNA: discovery and biological implications, *DNA Cell Biol.* 26 (2007) 195–207.
- [47] W. Filipowicz, V. Pogacic, Biogenesis of small nucleolar ribonucleoproteins, *Curr. Opin. Cell Biol.* 14 (2002) 319–327.
- [48] H. Zhou, K. Lin, Excess of microRNAs in large and very 5' biased introns, *Biochem. Biophys. Res. Commun.* 368 (2008) 709–715.
- [49] W. Yu, et al., Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA, *Nature* 451 (2008) 202–206.
- [50] A. Mazo, J.W. Hodgson, S. Petruk, Y. Sedkov, H.W. Brock, Transcriptional interference: an unexpected layer of complexity in gene regulation, *J. Cell. Sci.* 120 (2007) 2755–2761.
- [51] I. Martianov, A. Ramadass, A. Serra Barros, N. Chow, A. Akoulitchev, Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript, *Nature* 445 (2007) 666–670.
- [52] J.C. Schwartz, et al., Antisense transcripts are targets for activating small RNAs, *Nat. Struct. Mol. Biol.* 15 (2008) 842–848.
- [53] B.A. Janowski, et al., Activating gene expression in mammalian cells with promoter-targeted duplex RNAs, *Nat. Chem. Biol.* 3 (2007) 166–173.
- [54] X. Wang, et al., Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription, *Nature* 454 (2008) 126–130.
- [55] K. Plath, et al., Role of histone H3 lysine 27 methylation in X inactivation, *Science* 300 (2003) 131–135.
- [56] T. Imamura, et al., Non-coding RNA directed DNA demethylation of Sphk1 CpG island, *Biochem. Biophys. Res. Commun.* 322 (2004) 593–600.
- [57] M.D. Yan, et al., Identification and characterization of a novel gene Saf transcribed from the opposite strand of Fas, *Hum. Mol. Genet.* 14 (2005) 1465–1474.
- [58] C.W. Smith, J. Valcarcel, Alternative pre-mRNA splicing: the logic of combinatorial control, *Trends Biochem. Sci.* 25 (2000) 381–388.
- [59] A. Bevilacqua, M.C. Ceriani, S. Capaccioli, A. Nicolini, Post-transcriptional regulation of gene expression by degradation of messenger RNAs, *J. Cell. Physiol.* 195 (2003) 356–372.
- [60] J.S. Mattick, Introns: evolution and function, *Curr. Opin. Genet. Dev.* 4 (1994) 823–831.
- [61] J.S. Mattick, M.J. Gagen, The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms, *Mol. Biol. Evol.* 18 (2001) 1611–1630.
- [62] A.E. Hill, et al., Micro-RNA-like effects of complete intronic sequences, *Front. Biosci.* 11 (2006) 1998–2006.
- [63] S. Barik, An intronic microRNA silences genes that are functionally antagonistic to its host gene, *Nucleic Acids Res.* 36 (2008) 5232–5241.
- [64] J.Q. Clement, L. Qian, N. Kaplinsky, M.F. Wilkinson, The stability and fate of a spliced intron from vertebrate cells, *RNA* 5 (1999) 206–220.
- [65] J.E. Kravchenko, I.B. Rogozin, E.V. Koonin, P.M. Chumakov, Transcription of mammalian messenger RNAs by a nuclear RNA polymerase of mitochondrial origin, *Nature* 436 (2005) 735–739.
- [66] G.M. Borchert, W. Lanier, B.L. Davidson, RNA polymerase III transcribes human microRNAs, *Nat. Struct. Mol. Biol.* 13 (2006) 1097–1101.
- [67] M. Yamaizumi, T. Sugano, U.v.-induced nuclear accumulation of p53 is evoked through DNA damage of actively transcribed genes independent of the cell cycle, *Oncogene* 9 (1994) 2775–2784.
- [68] M. Leist, et al., Tumor necrosis factor-induced apoptosis during the poisoning of mice with hepatotoxins, *Gastroenterology* 112 (1997) 923–934.
- [69] C. Casse, F. Giannoni, V.T. Nguyen, M.F. Dubois, O. Bensaude, The transcriptional inhibitors, actinomycin D and alpha-amanitin, activate the HIV-1 promoter and favor phosphorylation of the RNA polymerase II C-terminal domain, *J. Biol. Chem.* 274 (1999) 16097–16106.
- [70] J.N. Hutchinson, et al., A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains, *BMC Genomics* 8 (2007) 39.
- [71] G.J. Hurteau, S.D. Spivack, mRNA-specific reverse transcription-polymerase chain reaction from human tissue extracts, *Anal. Biochem.* 307 (2002) 304–315.
- [72] J. Shendure, G.M. Church, Computational discovery of sense–antisense transcription in the human and mouse genomes, *Genome Biol.* 3 (2002) RESEARCH0044.
- [73] H.E. Huber, J.M. McCoy, J.S. Seehra, C.C. Richardson, Human immunodeficiency virus 1 reverse transcriptase. Template binding, processivity, strand displacement synthesis, and template switching, *J. Biol. Chem.* 264 (1989) 4669–4678.
- [74] M. Furuno, et al., Clusters of internally primed transcripts reveal novel long noncoding RNAs, *PLoS Genet.* 2 (2006) e37.
- [75] C. Vanhee-Brossollet, C. Vaquero, Do natural antisense transcripts make sense in eukaryotes? *Gene* 211 (1998) 1–9.
- [76] R. Martone, et al., Distribution of NF-kappaB-binding sites across human chromosome 22, *Proc. Natl. Acad. Sci. U. S. A.* 100 (2003) 12247–12252.
- [77] G. Euskirchen, et al., CREB binds to multiple loci on human chromosome 22, *Mol. Cell. Biol.* 24 (2004) 3804–3814.