Regulation by RNA

Maciej Szymański and Jan Barciszewski
Institute of Bioorganic Chemistry, Polish Academy of Sciences, 61–704 Poznań, Poland

In recent years, noncoding RNAs (ncRNAs) have been shown to constitute key elements implicated in a number of regulatory mechanisms in the cell. They are present in bacteria and eukaryotes. The ncRNAs are involved in regulation of expression at both transcriptional and posttranscriptional levels, by mediating chromatin modifications, modulating transcription factor activity, and influencing mRNA stability, processing, and translation. Noncoding RNAs play a key role in genetic imprinting, dosage compensation of X-chromosome-linked genes, and many processes of differentiation and development.

KEY WORDS: Noncoding RNA, Riboregulators, Genetic imprinting, microRNAs, Development, Differentiation.

I. Introduction

One of the key properties of living organisms is the ability to adapt to changing conditions. Various types of adaptation can be observed on all levels of organization of living matter. They include interactions of simple unicellular organisms with their environment as well as development and differentiation of more complex life forms. The consequences of external stimulation or intracellular signals are observed on the cellular level as changes of patterns of gene expression. According to the ''central dogma of molecular biology'' (Fig. 1) genetic information encoded in the nucleotide sequence of DNA is first transcribed into messenger RNA (mRNA) and after processing is translated on the ribosomes to produce proteins. Each of these steps—transcription, processing, and translation—can serve as a potential checkpoint for regulatory mechanisms, which can alter or modulate gene expression.

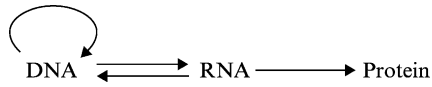

FIG. 1 The flow of genetic information according to the central dogma of molecular biology.

The first RNA molecules identified, mRNAs, transfer RNAs (tRNAs), and ribosomal RNAs (rRNAs), are directly involved in protein biosynthesis. They provide a link between the major carrier of genetic information (DNA) and its expressed form (proteins). Although RNA has been regarded as a "poor cousin of DNA" and is thought to play only an accessory role in the cell, the discoveries of new classes of RNAs, RNA-based catalysis (Kruger *et al.*, 1982; Guerrier-Takada *et al.*, 1983; Doudna and Cech, 2002) and, more recently, crystal structures of ribosomes (Moore and Steitz, 2002; Ramakrishnan, 2002), clearly demonstrate that various types of RNAs can be regarded as key players in the cell. On the other hand, the results of several eukaryotic genome sequencing projects indicate that the complexity of organisms cannot be defined by the number of proteins encoded by the genomic DNA. The estimated number of protein-coding genes in the human genome is approximately only two times higher than that in nematodes (Venter *et al.*, 2001; Lander *et al.*, 2001). Note that the higher the organization of an organism, the smaller portion of its genome actually encodes proteins. In yeast, about 57% of genomic DNA contains open reading frames. In *Caenorhabditis elegans* and *Drosophila* that number drops to 27% and 13%, respectively (Adams *et al.*, 2000; The *C. elegans* Sequencing Consortium, 1998). In humans, only about 1.5–2% of nuclear DNA codes for proteins (Venter *et al.*, 2001). Moreover, current assessments of the number of human genes within the range of 30,000–40,000 do not fit the estimates from expressed sequence tag (EST) cluster analysis, which gives a value approximately twice as high (Zhuo *et al.*, 2001). It is therefore assumed that the lion's share of transcriptional output from the human genome constitutes RNAs that do not encode proteins. The noncoding RNAs can account for 97–98% of all transcribed sequences (Mattick, 2001, 2003).

In the past decade much attention has been focused on a novel class of RNA molecules, whose job goes beyond the earlier recognized functions in transmission, processing, and decoding of genetic information in protein biosynthesis. The regulatory noncoding RNAs, or riboregulators, can fill in the gap in our understanding of complex mechanisms underlying many physiological processes.

Thus, the whole pool of RNAs produced in the cell can be roughly divided into two major groups: protein-coding mRNAs and noncoding RNAs. Among the latter, there are RNAs that play essential roles in the correct functioning of the cell. These housekeeping RNAs are usually constitutively

expressed and are required for the viability of the cell. They include RNAs performing crucial roles in protein biosynthesis (rRNAs, tRNAs), processing and modifications of precursor RNAs (small nuclear RNAs, snRNAs; small nucleolar RNAs, snoRNAs; RNase P RNAs; guide RNAs, gRNAs), synthesis of telomeres (telomerase RNA), quality control of translation (tmRNA), or components of other ribonucleoprotein complexes (4.5S RNA; vault RNAs, vRNAs). Because of their key roles in cellular metabolism, housekeeping RNAs are generally well conserved in the course of evolution and can be found throughout and across all kingdoms. In eukaryotes, the housekeeping noncoding RNA genes are transcribed by RNA polymerases I and III.

The second group of noncoding RNAs includes riboregulators or transcripts with documented or suspected regulatory functions both in prokaryotes as well as in eukaryotes. Most of these RNAs were identified as transcripts of genes activated in response to environmental conditions, developmental signals, or cell- or tissue-specific transcripts. In many cases, a precise role for these RNAs is not known. There are, however, data that clearly demonstrate that in certain processes an expression of noncoding RNA genes is indispensable. In eukaryotes, most of the regulatory RNAs show several features typical of mRNAs. They are transcribed by RNA polymerase II, capped, and polyadenylated. Their primary transcripts are often alternatively spliced. The only difference is lack of substantial open reading frames and thus protein-coding ability. In contrast to housekeeping RNAs, specific regulatory RNAs are poorly conserved, which suggests that they often function only in smaller phylogenetic groups. The size of regulatory RNAs varies from ~20 nt in the case of micro-RNAs (miRNAs) to over 10-kb-long transcripts in mammals.

Speaking of non-protein-coding RNAs it is necessary also to take into account introns, cleaved out from precursor mRNAs during splicing (Mattick and Gagen, 2001). Analysis of the protein-coding genes in the human genome revealed that on average only 5% of the primary transcript of a gene accounts for the translated open reading frame (Venter *et al.*, 2001; Lander *et al.*, 2001). The first impression after the discovery of split genes in eukaryotes was that the intervening sequences, removed during maturation, are nonfunctional and that in evolution they provide means for the generation of novel genes via the process of exon shuffling (Roy *et al.*, 2002). In some cases, introns are processed and produce functional RNAs. A new hypothesis suggests that introns may constitute an important element of regulatory networks (Mattick, 1994, 2001, 2003; Mattick and Gagen, 2001). According to this view, the role of introns may be to provide information to the genetic network about the gene expression status. This information can modulate other elements of the system and influence its behavior.

Noncoding regulatory RNAs have been implicated in a number of processes, whose role is to modulate gene expression (Szymanski and Barciszewski, 2002). Some of them are involved in control of chromatin structure and consequently its transcriptional activity. This includes regulation of dosage of sex chromosome-linked genes and genetic imprinting. A number of noncoding RNAs influence translation as well as RNA localization and processing. RNAs with specific protein-binding sites can modulate the activity of enzymes and transcription factors.

Advances in RNA research in the past several years clearly indicate that noncoding RNAs play a prominent role in regulation of gene expression, controlling its every aspect. Each year the number of new noncoding transcripts and processes in which they are involved is growing (Szymanski *et al*., 2003).

The chemical and biological properties of RNA make it very well suited for the role of an intracellular signaling molecule. First, in contrast to proteins, production of functional RNAs requires much less energy and time. RNA molecules are also less stable than proteins and they can be more easily degraded by cellular enzymes. Polynucleotide chains of RNA can adopt a variety of higher order structures, which can constitute binding sites for proteins or small molecules. In many cases, the activity of noncoding RNAs depends on interactions with other RNA molecules. This can be accomplished via simple base pairing with a complementary sequence within a target RNA. For this task, the protein would require a complicated specific RNA-binding domain.

II. Noncoding RNAs in Dosage Compensation and Sex Determination

In organisms employing a system of distinct sex chromosomes, X and Y, for sex determination, there is the need to equalize the dosage of X-linked genes in male (XY) and female (XX) cells. This goal can be accomplished in different ways. In *Drosophila*, transcription from a single X chromosome in male cells is approximately 2-fold higher than transcription from each of the two female Xs. In nematodes, transcription from both X chromosomes in XX cells is down-regulated and in mammals, one of the female X chromosomes becomes transcriptionally inactive. Although, these mechanisms seem to be very diverse, they share one common feature. All of them involve modifications of X chromosome chromatin structure. This in turn leads to altered transcriptional activity. In *Drosophila* and mammals, the mechanisms underlying dosage compensation involve specific noncoding RNAs, whose

expression is crucial for either up-regulation of X-linked genes or X chromosome inactivation (Kelley and Kuroda, 2000).

A. *roX* RNAs in *Drosophila*

In *Drosophila*, dosage compensation is accomplished by a 2-fold increase of transcription from a single X chromosome in male cells. Factors responsible for the up-regulation of transcription were identified as proteins encoded by *male-specific lethal* (*msl*) *msl-1, msl-2*, and *msl-3*, *maleless (mle)*, and *males absent on first* (*mof*) genes. Mutations in these genes result in male-specific lethality of larvae and their products are collectively termed MSL proteins. These five proteins form a dosage compensation complex (DCC, compensasome) (Kindel and Amrein, 2003). The DCC was found to specifically associate with multiple sites on the male X chromosome. The complex formation and its activity strictly depend on the presence of all MSL proteins (Meller *et al*., 2000). A characteristic feature of the up-regulated X chromosome is specific acetylation of histone H4 at lysine 16 (H4Ac16) (Turner *et al*., 1992). The enzyme responsible for this modification is aMOF protein with an activity of histone acetyltransferase. It is expressed in both sexes, but associates with the X chromosome only in males and only as a part of the MSL complex. Another protein enriched on the male X chromosome in a pattern similar to MSL proteins is a histone H3 kinase JIL-1. It is not a component of the compensasome, yet it is highly possible that it may play a role in chromatin remodeling associated with dosage compensation (Jin *et al*., 1999, 2000).

In addition to the MSL complex proteins, dosage compensation in *Drosophila* depends on expression of two noncoding RNAs, *roX1* and *roX2* (*R*NA *o*n *X*) 3.7 and 1.3 kb long (Franke and Baker, 1999). The two *roX* genes are localized on the X chromosome and their male-specific expression requires the presence of all MSL proteins (Amrein and Axel, 1997). The genes encoding *roX* RNAs and MSL proteins show the same expression patterns in male embryos and their products are found to associate with male X chromosomes from early embryonic stages on. Combined results of immunostaining assay and *in situ* hybridization demonstrated that *roX* RNAs and MSL proteins show the same localization on the X chromosome (Meller *et al*., 2000). The critical role of *roX* RNAs was determined by an analysis of *roX* gene deletions. Mutations affecting expression of either *roX1* or *roX2* RNA did not influence binding of MSL proteins to the male X chromosome, but double mutations deleting both *roX1* and *roX2* genes were lethal and no MSL proteins associated with X chromosome were observed. That lethality, caused by deletion of both *roX* genes, excludes the possibility

that there is a gene encoding another RNA that could be a substitute for *roX1* or *roX2* RNAs. This observation also suggested that the two RNAs are functionally redundant and that the presence of at least one of them is sufficient for dosage compensation to proceed (Franke and Baker, 1999). The functional equivalence of both *roX* RNAs would also suggest some degree of conservation on the level of nucleotide sequence and/or secondary structure. The two RNAs share only about a 30-nt-long stretch of highly similar nucleotides (Franke and Baker, 1999). In spite of this redundancy in function, there are some differences between these two RNAs. Only *roX2* transcripts are able to move to all chromatin entry sites in *msl3* mutants and there are fly lines lacking functional *roX1*, but none lacking *roX2*. On the other hand, the lethal phenotype resulting from deletion of both *roX* genes can be rescued by expression of a transgene encoding either of them (Meller and Rattner, 2002).

The *roX* genes performed two distinct and separable functions in dosage compensation. First, *roX* RNAs constitute indispensable elements of the nucleoprotein complexes responsible for chromatin modifications. Second, the genes themselves provide strong chromatin entry sites for the MSL complex, possibly to ensure rapid recruitment of the MSL proteins for *roX* RNA binding. This could be important for stabilization of the otherwise labile transcripts and their nuclear localization (Meller *et al.*, 2000). It has been proposed that all chromatin entry sites on the X chromosome evolved from once functional *roX* genes that were inactivated due to selective pressure against retaining multiple transcripts with a redundant function (Meller and Rattner, 2002). Fully assembled MSL–*roX* RNA complexes can spread along the entire X chromosome. Experiments with transgenic expression of the *roX2* gene inserted into the autosome demonstrated that the DCC can be assembled at this site and bidirectionally spread in *cis* from the chromatin entry site into flanking chromatin (Meller *et al.*, 2000). On the other hand, constitutive expression from autosome-inserted transgenes of *roX1* or *roX2* in females showed that the RNAs cannot bind X chromosome by itself. In the absence of the full complement of MSL proteins, they are retained at the sites of their synthesis. The stability of both *roX* RNAs strongly depends on the presence of MSL complex proteins (Meller *et al.*, 2000). During the assembly of the DCC, the most important roles are played by MSL1 and MSL2 proteins, which form a core of the complex and are responsible for X chromosome binding (Meller *et al.*, 2000). It was proposed that the MSL proteins bind nascent *roX* transcripts and the rate of this process, which depends on the availability of MSL proteins, determines whether the complex immediately spreads to flanking chromatin or dissociates from the site of assembly (Y. Park *et al.*, 2002). In the next step MLE protein is added, which allows binding of MSL3 and MOF (Akhtar *et al.*, 2000). It has been demonstrated that MOF binding to the DCC depends on the presence of RNA. This protein shows general RNA-binding properties, yet with an affinity much

higher than for DNA. RNA-binding activity is associated with a chromodomain region, frequently found in chromatin-binding proteins. The MOF chromodomain can bind *roX* RNA *in vivo* and may be responsible for the integration of this protein into DCC (Akhtar *et al.*, 2000).

Interestingly, a small fraction of *roX* double mutant males (~5%) can survive. In those cases, the MSL1/MSL2 complex can be assembled in the absence of *roX* RNAs but its localization to the X chromosome is affected and most of the DCCs become available for binding to autosomal sites where their activity results in elevated transcriptional activity. That is supported by the observation that MSL2, which in the presence of *roX* RNAs specifically binds to X chromosome in roX^{-} males, is found to associate with autosomes. It can be concluded that the primary role of *roX* RNAs is to ensure the specificity of MSL complex interaction with the X chromosome (Meller *et al.*, 2000). MSL2 is the only truly male-specific component of the MSL complex. Its expression in females is lethal, unless both of the *roX* genes are mutated. All other MSL proteins are transcribed and effectively translated in both sexes. However, in contrast to males, they are not essential for females and their deletion has no harmful effect. This suggests that they are specifically dedicated to dosage compensation.

The availability of a genome sequence of mosquito, *Anopheles gambiae*, allowed the question of the conservation of the dosage compensation mechanism in insects to be addressed. Orthologs of the five protein components of the compensasome have been identified, but there were no sequences showing similarity to either of the *Drosophila roX* genes (Zdobnov *et al.*, 2002). It is possible that the RNA-coding genes are too divergent to be detected using sequence similarity searches.

B. X Chromosome Inactivation

X chromosome inactivation was first recognized as a mechanism for dosage compensation in mammals by Mary Lyon, over 40 years ago (Lyon, 1961). This process involves transcriptional silencing of all but one of the X chromosomes in females. As a result, all male and female cells possess only one active X chromosome, which ensures equal levels of transcription from the X-linked genes. In female cells, an inactive X chromosome can be observed microscopically as a condensed body (Barr body) at the periphery of the nucleoplasm. In eutherians, the future inactive chromosome is chosen at random and each of the X chromosomes, regardless of its parental origin, has an equal chance of becoming active (Xa) or inactive (Xi). In fact, X inactivation follows an "$n-1$" rule, by which all except one X chromosome in the cell are subject to silencing. In marsupials, and some extraembryonic tissues of rodents, the inactivation process depends on gametic imprinting,

which results in preferential silencing of the paternal X chromosome (Goto and Takagi, 2000; Graves, 1996).

At the beginning of the development of female embryos, both X chromosomes are transcriptionally active. In mouse embryonic tissues, the inactivation process takes place at the late blastocyst stage and is complete at the beginning of the gastrulation stage. The election of the future inactive X is made independently in each cell. The silencing results in stable inactivation of one X chromosome and the inactive state is maintained throughout all subsequent cell divisions. Consequently, all the progeny of a given cell have the same Xi. The only natural case of X reactivation was observed in XX primordial germ cells, when they enter the genital ridge (Nesterova *et al.*, 2002).

The inactive X chromosome shows the properties characteristic of the constitutive heterochromatin such as condensation in the interphase, methylation of the CpG islands, hypoacetylation of histone H4 of inactivated genes (Gilbert and Sharp, 1999), and the presence of the histone macroH2A (Mermoud *et al.*, 1999). The inactive X chromosome is also replicated late in S phase (Brockdorff, 2002). X chromosome inactivation depends on activity of the X-inactivation center (human *XIC*, mouse *Xic*). Its existence was inferred from results of analysis of chromosomal translocations involving the X chromosome. In humans, *XIC* was narrowed down to a Xq13 region on the proximal long arm of the X chromosome (Fig. 2). Its presence was shown to be essential and sufficient for the initiation of X chromosome inactivation (Brown *et al.*, 1991b).

1. *Xist* RNA

In 1991, human *XIST* (Brown *et al.*, 1991a) and mouse *Xist* (Borsani *et al.*, 1991; Brockdorff *et al.*, 1991) (*X-i*nactive *s*pecific *t*ranscript) genes were identified. They were expressed exclusively from inactive X chromosomes and mapped to the *XIC/Xic* regions (Brown *et al.*, 1991b). *Xist* gene transcription produces a spliced and polyadenylated RNA that does not contain any significant open reading frames, does not associate with polysomes, and

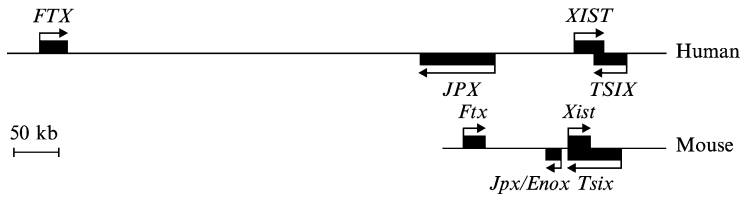

FIG. 2 A comparison of the *XIST/Xist* regions on human and mouse X chromosomes showing genes for noncoding RNAs.

shows nuclear localization exclusively (Brockdorff *et al.*, 1992). Two lines of evidence suggest that *Xist* RNA plays a pivotal role in the X-inactivation process. First, targeted deletion of the *Xist* gene renders the deficient X chromosomes unable to undergo inactivation (Newall *et al.*, 2001). Second, expression of *Xist* RNA from an autosome during embryonic stem cells differentiation initiates inactivation of the chromosome carrying the transgene (Lee and Jaenisch, 1997).

An intron–exon organization of *Xist* genes was analyzed in detail in mouse, human, and bovine (Chureau *et al.*, 2002). Of the eight exons identified in mouse *Xist*, seven are present in all three species. They show relatively low sequence identity of 60–70%, which is comparable to the conservation of untranslated regions between human and mouse orthologous protein-coding genes. There are alternative polyadenylation sites and the primary transcripts are subject to alternative splicing. The longest mouse and human mature *Xist* RNAs are 17.9 and 19.3 kb, respectively.

Although the overall conservation of the *Xist* RNA nucleotide sequence is low, there are several regions of tandem repeats that are shared between species. A long conserved hairpin structure was identified within exon 4, but its functional significance is not known, because deletions of the 3′-terminal fragment of the gene or the exon 4 alone does not affect the inactivation process (Brockdorff, 2002). Of special importance is a region located within intron 1, close to the 5′-end of the molecule. This portion of the molecule contains A-repeats, which are indispensable for the silencing process (Beletskii *et al.*, 2001; Wutz *et al.*, 2002). Individual A-repeats are predicted to form two short stem–loop structures. The A-repeat region is not functional by itself and it has to be located within *Xist* RNA that is able to associate with and spread along the X chromosome. *In vitro* studies demonstrated that A-repeats serve as binding sites for heteronuclear ribonucleoprotein (hnRNP) C1/C2, but a functional significance of these interactions has not been established due to the generally low specificity of C1/C2 hnRNP interactions with RNA (Brown and Baldry, 1996; Wutz *et al.*, 2002; Wutz, 2003).

A crucial role of the A-repeats in silencing is particularly interesting in the context of earlier studies that showed that they are included in only one-third of *Xist* transcripts initiated from the P_1 promoter (Johnston *et al.*, 1998). The remaining two-thirds would therefore retain X chromosome localization but would be unable to mediate silencing. These observations seem to support a model that assumes the existence of a limited number of *Xist* RNA-binding sites on the X chromosome. They would form nucleation centers for the recruitment of the silencing complexes, from which they would subsequently spread along the whole X chromosome. This model could also explain the banded pattern of *Xist* RNA localization on X chromosomes in rodents (Duthie *et al.*, 1999).

Whereas the function in inactivation can be ascribed to a specific region within *Xist* RNA, X chromosome localization depends on a number of sequence elements that are spread along its length. They are poorly conserved, functionally redundant, and do not seem to possess any common structural motif (Nesterova *et al.*, 2001).

Although almost a decade passed since the discovery of *Xist* RNA, the fine details of its role in the inactivation process remain unresolved. One of the early models of *Xist* action suggested that the locus plays the role of a chromatin-organizing region. The transcriptional activity of the gene was proposed to initiate chromatin remodeling that would spread along the entire chromosome (Brockdorff *et al.*, 1992). Experimental evidence suggests, however, that *Xist* RNA plays a more active role in the process of X inactivation. Following the up-regulation of the *Xist* gene on a future inactive X chromosome, the spliced transcripts associate and spread along the chromosome from which they have been produced. This "RNA painting" is proposed to constitute a major factor in recruitment of silencing factors responsible for the establishment of inactive chromatin conformation in a whole chromosome. *Xist* RNA does not show an exclusive specificity for X chromosome chromatin, since it can also be found to be associated with autosomes carrying *Xist* transgenes.

One of the features of the inactive X chromosome is the presence of a specific histone H2A variant, macroH2A1.2, characterized by the presence at its C-terminus of a large, globular, nonhistone domain. The inactive X chromosomes in somatic cells show considerably higher levels of this histone than their active counterparts, which suggests a role of macroH2A1.2 in X inactivation. In undifferentiated embryonic stem cells, prior to X inactivation, macroH2A1.2 accumulates around the centrosomal region in both XX and XY cells. During differentiation and X inactivation in XX cells, macroH2A1.2 disappears from the centrosome and becomes detectable in the inactive X chromosome. The timing of the association of macroH2A1.2 with the inactive X suggests that it may be important for the establishment and to some extent maintenance of the inactive state (Rasmussen *et al.*, 2000). Experimental data suggest that *Xist* RNA is a factor responsible for the recruitment of macroH2A1.2 to the inactivated X chromosome and this process depends on its 3′-terminal domain crucial for chromosomal localization (Wutz *et al.*, 2002). The expression of *Xist* is a prerequisite for deposition of macroH2A1.2, and a conditional deletion of *Xist* interferes with this process. Moreover, *Xist* RNA is present in the chromatin immunoprecipitated with antibodies against macroH2A1.2.

2. *Tsix* RNA

Tsix was initially described as an 40-kb-long intronless RNA initiated 15 kb downstream from the *Xist* gene and found to be dispensable for the silencing step (Lee *et al.*, 1999). It was later demonstrated that the primary transcript is

further processed producing a 2.7-kb transcript originating from a promoter located close to exon 2, which is associated with a CpG island differentially methylated on active and inactive X chromosomes. Another 4.7-kb variant has been identified to be transcribed from a minor promoter 28 kb downstream of *Xist*, which suggested that *Tsix* in fact starts 13 kb further downstream than originally reported (Sado *et al*., 2001). Interestingly, only 30–60% of *Tsix* RNA is spliced, which suggests that the spliced and unspliced forms may perform different functions (Shibata and Lee, 2003).

The *Tsix* gene is transcribed into a nuclear noncoding RNA antisense to *Xist*. The gene organization, expression pattern, and partial complementarity suggested that *Tsix* acts as an *Xist* antagonist and repressor. *Tsix* expression is opposite to that of *Xist*. Before the initiation step, in the XX cells *Tsix* RNA is produced at low levels from both X chromosomes, yet it occurs in 10- to 100-fold excess over *Xist* (Shibata and Lee, 2003). The onset of inactivation and increased transcription from *Xist* parallel turning off *Tsix* transcription on the future inactive X chromosome (Lee *et al*., 1999).

Mutations of the *Tsix* promoter regions demonstrated that *Tsix* RNA operates on a choice level and *Tsix*-deficient X chromosomes are preferentially inactivated. On the other hand, high expression levels of *Tsix* driven by a constitutive elongation factor 1 α (EF1α) promoter on one of the X chromosomes render it resistant to *Xist* RNA accumulation and, consequently, inactivation (Stavropoulos *et al*., 2001). Although the precise mechanism of *Tsix* action is not known, several possible models have been proposed. Because *Tsix* RNA is in part complementary to *Xist* RNA, the most obvious explanation would be duplex formation and blocking of the active, possibly protein-binding, sites on the latter. Alternatively, RNA–RNA interactions could be responsible for decreased *Xist* RNA stability (Luikenhuis *et al*., 2001; Shibata and Lee, 2003).

The inactivation process relies on both *cis*-acting elements and *trans*-acting factors. A computational analysis of the *Tsix* promoter/*DXPas34* region revealed a cluster of binding sites for the CTCF protein, identified earlier as a transcription factor and chromatin insulator. CTCF was proposed to work together with *Tsix* RNA in the choice step. The model assumes that binding of CTCF to one of the X chromosomes marks it as future active by preventing *Xist* transcription. The variability in the differentially methylated region (DMR) could serve as a discriminator allowing CTCF binding to only one *Tsix* allele. The suppression of *Xist* could be achieved either by CTCF-mediated activation of *Tsix* transcription or by blocking the access to a putative *Xist* enhancer located further downstream. The deletion of an array of CTCF-binding sites results in a nonrandom inactivation of the mutated X. Using a gel retardation assay CTCF was shown to bind *Tsix* fragments *in vitro* and the chromatin immunoprecipitation with anti-CTCF antibodies demonstrated *Tsix* binding *in vivo*. On the future inactive X

chromosome, CTCF does not bind *Tsix* probably due to the methylation of the CTCF array (Chao *et al*., 2002).

TSIX was also identified in humans. Like the murine gene, *TSIX* is expressed only in embryo-derived cells. Its transcription initiates ~27 kb downstream from the 3′-end of *Xist* and proceeds from the opposite strand giving rise to an RNA that is in part antisense to *XIST* (Migeon *et al*., 2001). An analysis of the complete *Xic* regions from three mammalian species revealed that unlike *Xist*, *Tsix* is not conserved (Chureau *et al*., 2002). A detailed comparison of *Tsix* expression patterns in humans and mouse revealed that the two genes can not be regarded as functional equivalents (Migeon *et al*., 2002). In human cells *TSIX* RNA is produced exclusively from the inactive X chromosome and its expression does not repress expression of *XIST*. There is no differentially methylated CpG island that is present in murine *Tsix* and the transcript does not overlap the entire length of *XIST*, covering only exons 5 through 8. These observations imply that human *TSIX* may be a defective gene whose primary function, still performed in rodents, was lost (Migeon *et al*., 2002).

3. Other Noncoding RNAs from X Chromosome

A detailed comparative analysis of the complete nucleotide sequences of the *Xic* regions from three mammalian species (mouse, human, and bovine) demonstrated that in addition to *Xist* and *Tsix*, there are three novel genes whose products may function as noncoding RNAs (Chureau *et al*., 2002; Johnston *et al*., 2002).

Ftx is a gene identified in mouse *Xic* approximately 140 kb upstream from *Xist*. It consists of seven exons that can be alternatively spliced into at least four different variants. A human counterpart was identified about 600 kb upstream from *XIST*, but it shows different intron–exon organization. In both mouse and humans, the 5′-regions of *Ftx* are well conserved and contain CpG islands at positions corresponding to cDNAs start sites. Both genes are transcribed in the opposite orientation relative to *Xist/XIST* genes. No conserved long open reading frames (ORFs) were identified, which suggests that the product is functional noncoding RNA with an as yet unidentified function. Expression of *Ftx* was detected in several tissue types (Chureau *et al*., 2002).

Jpx was identified based on the presence of a conserved CpG island upstream of *Xist*. Corresponding ESTs in all three species confirmed that the gene is in fact expressed. The *Jpx* genes consist of three exons. In mouse *Jpx* is expressed only from the active X chromosome and its primary transcript undergoes alternative splicing producing two variants (315 and 562 nt long). Sequence comparison revealed that the first mouse exon corresponds to the second human exon. The last two exons in mouse and bovine sequences

and the third exon in humans match various repeat elements. The product of this gene is probably a noncoding RNA, since no conserved ORFs were found (Chureau *et al.*, 2002). Another group described a transcript coming from the same localization in mouse. The gene was called *Enox* (*e*xpressed *n*eighbor *o*f *X*ist) and was found to escape X inactivation and is expressed from both active and inactive X chromosomes. *Enox* was shown to consist of five exons whose alternative splicing and polyadenylation produce several variants of noncoding RNA. In exons 1, 2, and 3 and the 3′-portion of exon 5 simple repetitive sequences have been identified. The function of *Enox* RNA is not known. Interestingly, the chromosomal environment of *Enox* does not fit the criteria for escape from inactivation, because the region upstream is rich in LINEs thought to be a nucleation site for the inactivation process. It has been proposed that the close proximity of *Xist* may partially interfere with the establishment and maintenance of heterochromatin allowing expression of *Enox* from the inactive X (Johnston *et al.*, 2002).

C. Male Hypermethylated (MHM) Region and Sex Determination in Birds

In birds, sex determination and differentiation depend on sex chromosomes Z and W. Males possess two Z chromosomes, whereas females are determined by ZW karyotype. In contrast to mammals or *Drosophila*, the mechanism underlying these processes is still largely unknown. One of the genes proposed to play a role in sex differentiation in birds is a homolog of the human *DMRT1* (*d*oublesex and *m*ab-3-*r*elated *t*ranscription factor) implicated in testis differentiation. In chicken, *DMRT1* was mapped to the Z chromosome and its transcription level was shown to differ between males (ZZ) and females (ZW) (Raymond *et al.*, 1999). Elevated expression of *DMRT1*, resulting from active up-regulation of transcription, was found to correlate with testis development in males and in sex-reversed females (Smith *et al.*, 2003).

A male hypermethylated (MHM) region was identified on the short arm of the chicken Z chromosome in a vicinity of the *DMRT1* gene. The CpG islands within this 460-kb-long fragment are differentially methylated between males and females. Its methylation status is established shortly after fertilization in the early stages of embryonal development and depends on the presence or absence of the W chromosome. The MHM region consists of a tandemly repeated 2.2-kb-long sequence and it is transcribed producing heterogeneous RNAs. The longest transcripts are approximately 9.5 kb and the majority of them are not polyadenylated. The expression was shown to be female specific and to depend on methylation status. Upon treatment with a demethylating agent, 5-azacytidine, the male MHM became transcriptionally

active, producing high-molecular-weight, heterogeneous RNA as observed in female cells. MHM transcripts were shown to accumulate at or very close to the site of transcription and close to the *DMRT1* locus. The methylation status and consequently the transcriptional activity of the MHM region are proposed to be under the control of the yet unidentified W-linked gene. MHM transcripts may hence play the role of RNA repressor similar to that of *Xist* RNA (Teranishi *et al.*, 2001).

III. Noncoding RNAs from Imprinted Genes

One of the most intriguing genetic phenomena in mammals is genetic imprinting. This term is used to describe the parent-of-origin effect on gene expression in the offspring. In a diploid cell, most of the genes are inherited in two copies, one from each parent. The two alleles are theoretically equivalent and can perform the same functions. Genomic imprinting is a process whereby the expression of an allele depends on whether it comes from the mother or father (Bartolomei and Tilghman, 1997). Abnormalities in imprinted genes have been implicated in several developmental and neurobehavioral disorders. They are well documented in Angelman syndrome, Prader–Willi syndrome, and Beckwith–Wiedemann syndrome, and are suspected to play a role in several other neurobehavioral disorders, including autism, bipolar affective disorder, and schizophrenia (Hanel and Wevrick, 2001; Murphy *et al.*, 2001; Nicholls, 2000).

In most cases, the mammalian imprinted genes are present in clusters often associated with imprinted noncoding RNA genes. Expression of the noncoding RNA from one of the paternal alleles often correlates with repression of the linked protein-coding gene on the same allele. It has been postulated that the noncoding RNAs may be required for silencing, but in most cases the mechanisms are unknown. The noncoding transcripts from imprinted loci are often significantly different between species. The differences are not only limited to nucleotide sequences but also involve intron–exon organization.

A. Imprinted Noncoding RNA Genes at Human Chromosome 15q11–q13

Defects in imprinted genes located within the 15q11–q13 region (Fig. 3) are associated with Angelman syndrome (AS) and Prader–Willi syndrome (PWS). This region, also called the PWS region, contains 11 paternally expressed genes, and one maternally expressed gene. The imprinted genes are regulated by a bipartite imprinting center (PWS-IC and AS-IC) located

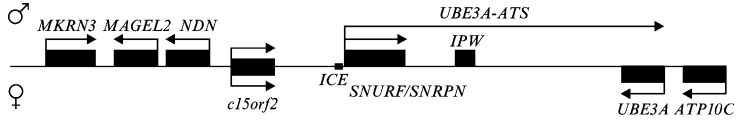

FIG. 3 Prader–Willi and Angelman syndromes region on the human chromosome 15q11–q13. Paternally expressed *IPW* and *SNURF-SNRPN–UBE3A-ATS* noncoding transcripts are responsible for repression of maternally expressed *UBE3A* in brain. The introns of *UBE3A-ATS* harbor a number of snoRNAs.

upstream of the *SNRPN* gene. In mouse, orthologous genes are located in the syntenic region of chromosome 7 and the PWS-IC is functionally conserved in mouse.

PWS is caused by genomic alterations that inactivate multiple, paternally expressed genes in the PWS region. They may result from a maternal uniparental disomy for chromosome 15, deletion of the paternally inherited 15q11–q13 region, paternally inherited balanced translocations, or imprinting mutations resulting in the silencing of paternally expressed genes. One of the genes, whose expression is inhibited in PWS patients, is *IPW* (*i*mprinted in *P*rader–*W*illi syndrome). This gene is located approximately 180 kb from the imprinting control element and encodes 2.3-kb spliced and polyadenylated RNA. The *IPW* RNA does not seem to encode any protein and probably functions as an RNA (Wevrick *et al.*, 1994). Like its human counterpart, murine *Ipw* was found to be expressed exclusively from the paternal allele. The primary transcript undergoes alternative splicing, producing several variants. Nucleotide sequence comparison of *Ipw* and *IPW* showed 79% similarity in a 319-nt-long region. The mouse and human genes show quantitative differences in expression in particular tissues. Expression of human *IPW* shows the same low level expression in all tissues. Mouse *Ipw* RNAs were detected in heart, skin fibroblasts, and liver, yet the transcription levels are very low. High-level expression was observed in brain (Wevrick and Francke, 1997).

As is associated with disruption of maternal expression of a single gene, *UBE3A*, within 15q11–q13. The genetic defects that may lead to the disorder include paternal uniparental disomy for chromosome 15, deletion of the maternally inherited 15q11–q13 region, imprinting mutations disrupting the maternal pattern of gene expression, or mutations in maternally derived *UBE3A*. Interestingly, unlike paternally expressed genes within the PWS region, in most tissues *UBE3A* shows biparental expression. An imprinted, maternally specific pattern of transcription can be observed only in certain brain cells. This gene is located in a position adjacent to a cluster of paternally expressed genes that is known to be positively regulated by the Prader–Willi syndrome imprinting center (PWS-IC), which has been also shown to influence the expression of *UBE3A*. Both in human and mouse an

imprinted, paternally specific expression of antisense transcripts form the *UBE3A/Ube3a* genes was observed. The deletion of PWS-IC results in biparental expression of *UBE3A* and repression of the paternal antisense transcription. It seems that the inhibition of the paternal *UBE3A* gene is an indirect result of the expression of a paternally expressed antisense RNA (UBE3A-AS), which is initiated at the imprinting center. The precise role of antisense transcription is not fully understood. One possible explanation is that it is an indicator of altered chromatin structure that negatively regulates the paternal *Ube3a* allele (Chamberlain and Brannan, 2001). The antisense transcript spans over 460 kb, covering previously identified exons of the *IPW* gene, and serves as a host for a number of snoRNAs, most of which are encoded within the introns. It has been proposed that these RNAs may be directly linked to the etiology of PWS (Runte *et al.*, 2001).

B. Imprinted Noncoding RNA Genes at Human Chromosome 7q32

In the Russel–Silver syndrome, 10% of patients show maternal uniparental disomy of chromosome 7, which suggests that it contains imprinted genes that may be involved in the disorder. *MEST/PEG1* (*me*soderm-*s*pecific *t*ranscript, *p*aternally *e*xpressed *g*ene *1*) was mapped to chromosome 7q32 (Fig. 4). It is expressed in fetal tissues of mouse and humans. There are two isoforms with different first exons. Isoform 1 shows imprinted expression in human lymphoblastoid cells while no imprinting was observed for isoform 2, transcribed from the more upstream promoter (Kosaki *et al.*, 2000).

MESTIT1 (*MEST i*ntronic *t*ranscript *1*), an antisense 4.2-kb-long RNA from *MEST*, was identified as a novel imprinted transcript. Its expression was shown to be limited to the paternal allele in fibroblasts and all analyzed fetal tissues. The transcription start site of *MESTIT1* is located within the intron of isoforms 2 of *MEST*. The transcript is composed of at least two exons (Nakabayashi *et al.*, 2002). A shorter 2.4-kb version of MESTIT1 RNA called PEG1-AS is expressed in human testis and mature spermatozoa (Li *et al.*, 2002). The nucleotide sequence comparison between humans and mouse revealed very low identity (~50%), which may suggest that

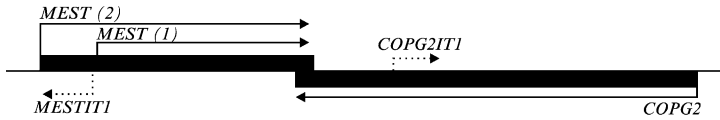

FIG. 4 Imprinted antisense transcripts at human chromosome 7q32. The two antisense transcripts from the *MEST1* and *COPG2* genes are shown as dotted lines.

transcription from the particular promoter is important and not the contents of the message.

Another gene located on human 7q32 is *COPG2* encoding the coatomer protein complex gamma 2 subunit. It is transcribed in an opposite orientation relative to *MEST*, and the 3′-UTRs of the two genes overlap. The antisense transcript originating from intron 20 of this gene *COPG2IT1* (*COPG2 i*ntronic *t*ranscript *1*) is paternally expressed in fetal tissues (Yamasaki *et al.*, 2000). In mouse, the corresponding *Copg2* and *Copg2as2 (Mit1/Lb9)* genes show reciprocally imprinted expression from maternal and paternal alleles, respectively (Lee *et al.*, 2000).

C. Imprinted Noncoding RNA Genes at Human Chromosome 11p15

The human chromosome 11p15 contains 13 maternally and 4 paternally expressed genes (Fig. 5). It consists of two distinct and independently regulated imprinted domains. The telomeric domain includes the *IGF, H19*, and *ASCL2* genes, while the centromeric domain includes the *TSSC3, TSSC5, CDKNIC*, and *KvLQT1* genes. Genetic and epigenetic abnormalities within this region were identified in Beckwith–Wiedeman syndrome (BWS) and in several human cancers.

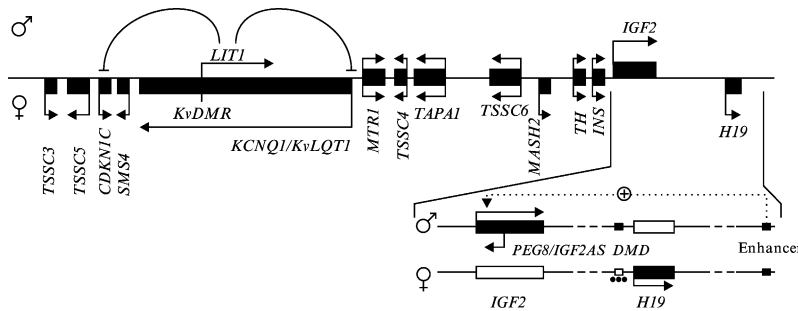

FIG. 5 A cluster of imprinted genes on the human chromosome 11p15. Expression from maternal and paternal alleles is represented by arrows below or above, respectively. Paternally expressed *LIT1* RNA initiated at a differentially methylated CpG island *(KvDMR)* is responsible for repression of both *KvLQT1* and an upstream gene for cyclin-dependent kinase inhibitor 1C ($CDKN1C/p57^{Kip2}$). Another differentially methylated region (DMD) serves as an imprinting control element for the *IGF2–H19* domain. Methylation of the paternal copy prevents binding of the CTCF insulator proteins (black dots) and formation of the chromatin boundary. This allows access to a downstream enhancer and expression of the paternal copy of the *IGF2* gene. Unmethylated maternal DMD binds CTCF proteins stimulating expression of the *H19* gene and preventing expression of *IGF2*.

1. *H19* RNA

H19 was the first imprinted gene whose product was recognized as a non-coding RNA. *H19* was identified as a gene showing coordinated expression with α-fetoprotein (Verona and Bartolomei, 2003). In humans, the gene was mapped to a cluster of imprinted genes on chromosome 11p15, which corresponds to the syntenic region on mouse chromosome 7. The gene shows an evolutionarily conserved localization downstream from the paternally expressed *IGF2* (*i*nsulin-like *g*rowth *f*actor 2) gene (Ohlsson *et al.*, 1993). This chromosomal region has been associated with tumor suppressor activity and expression patterns of *H19* RNA in several cancer cell types were demonstrated to differ from neighboring nonmalignant cells. Moreover, in certain tumors, expression of a transfected copy of the *H19* gene can restore normal cell cycle control in some transformed cell lines suppressing cell proliferation, clonogenicity, and tumorigenicity (Juan *et al.*, 2000).

H19 RNA is normally expressed at high levels in many embryonic tissues. The transcription is limited exclusively to the maternal allele and, except for skeletal muscles, it is extinguished shortly after birth. Sequence comparison of several mammalian *H19* genes revealed only short highly conserved regions. There are no conserved long ORFs, which supports the concept that there is no protein product that is important for *H19* activity (Juan *et al.*, 2000), even though 5′-truncated human *H19* transcripts were shown to be able to direct translation of a 26-kDa polypeptide (Joubel *et al.*, 1996). Computational predictions resulted in a common secondary structure model based on the analysis of *H19* nucleotide sequences from several mammalian species. It consists of 17 helical regions including pseudoknots with conserved sequences often located close to the helices within hairpin loops and linkers (Juan *et al.*, 2000). Another argument for a functional role of *H19* RNA comes from the determination of the rate of evolution between mouse and rat sequences, which shows signs of stabilizing selection (Hurst and Smith, 1999).

H19 and *IGF2* genes show a reciprocal pattern of imprinting. The opposite expression patterns of *H19* and *Igf2* led to the hypothesis that the two genes compete for a common enhancer, located downstream from the *H19* gene, and that transcription of *H19* RNA itself would be sufficient for markedly reduced transcription of *IGF2* (Ripoche *et al.*, 1997). Imprinted maternal-only expression and a reciprocal imprinting of its neighboring *Igf2* gene are consequences of differential methylation of the *H19* promoter and the DMR. On the maternal allele, the DMR and the *H19* promoter are unmethylated, which allows for *H19* RNA expression and binding of the CTCF protein to the DMR. CTCF is an enhancer-blocking protein that inhibits the

access of *Igf2* to the enhancer elements located downstream from the *H19* transcription start site. It also performs the role of a positive transcription factor for the expression of *H19* (Schoenherr *et al*., 2003). Thus, although it clearly plays an important role within the cell, *H19* RNA cannot be viewed as a molecule responsible for imprinting of other genes, and its deletion does not change the status of a linked *Igf2*. It has also been proposed that *H19* RNA may regulate *IGF2* expression on a posttranscriptional level, because *IGF2* mRNA incorporation into polysomes decreases with increased expression *H19* (Li *et al*., 1998).

There is a dispute concerning the significance of *H19* RNA. A functional role is suggested by many pediatric cancers that are associated with a loss of *H19* transcript (DeBaun *et al*., 2002). On the other hand *H19* is not necessary for normal development in mouse (Ripoche *et al*., 1997) and the overgrowth associated with the deletion may be due to altered regulation of the *Igf2* gene and its overexpression (Leighton *et al*., 1995).

2. *LIT1* RNA

Although Beckwith–Wiedemann syndrome (BWS) can arise from chromosome rearrangements, paternal uniparental disomy, or parental duplication of 11p15.5, the most common source of BWS is loss of imprinting at the *IGF2* gene without changes in methylation and expression of *H19* usually observed in Wilms' tumors (Smilinich *et al*., 1999). Thus, loss of imprinting at *IGF2* can be associated with two imprinting control regions, one that is *H19* dependent and the second independent. This notion was also supported by earlier observations that deletion of *H19* in mouse affects the imprinting status of *Igf2* and *Ins2*, but not *Mash2, Kvlqt1*, and $p57^{Kip2}$ (Caspary *et al*., 1998).

KvLQT1 is a 325-kb-long gene whose mutations in the protein coding region are associated with Romano–Ward, Jervell, and Lange–Nielsen syndromes (Chouabe *et al*., 1997). The expression of *KvLQT1* and its mouse counterpart *Kvlqt1* is regulated by genomic imprinting in a developmental and tissue-specific manner. A differentially methylated CpG island containing two direct repeat sequences was localized within intron 10 of *KvLQT1*. This region, termed *KvDMR1*, was demonstrated to be methylated on the maternal (expressed) *KvLQT1* allele. An antisense *LIT1* RNA (*l*ong QT *i*ntronic *t*ranscript *1*, *KvLQT1-AS/KCNQ1OT1*) is produced from the paternal allele in most human tissues (Mitsuya *et al*., 1999). Targeted deletion of the *LIT1* CpG island and lack of its expression were found to activate expression of normally silent paternal alleles of multiple imprinted genes including *KvLQT1* and *CDKN1C* ($p57^{Kip2}$). Yet it has no effect on expression of *H19* (Horike *et al*., 2000).

3. *PEG8/IGF2AS*

An antisense, paternally expressed RNA, *PEG8/IGF2AS* (*p*aternally *e*xpressed *g*ene *8*/*i*nsulin-like *g*rowth *f*actor *2* *a*nti*s*ense) was identified within the *IGF2* locus based on a significantly increased expression in all analyzed Wilms' tumor samples and several other fetal tumors. The human sequence has an ORF capable of encoding a 273–amino acid-long basic protein. This ORF is not conserved in a mouse counterpart and no protein product has been identified in a human sample, which indicate that they may act as noncoding RNAs (Okutsu *et al*., 2000).

D. Imprinting of *Dlk1–Gtl2* Domain

A cluster of imprinted genes similar to *Igf2–H19* has been found on distal mouse chromosome 12, based on the observations of phenotype changes in cases involving uniparental chromosomal duplications in mouse and humans. The identified region contains two genes, *Dlk1* and *Gtl2*, which show genomic organization very similar to that observed in the case of *Igf2* and *H19*. Interestingly, these genes have also been implicated in the regulation of prenatal growth (Takada *et al*., 2000). As in the *Igf2–H19* case, one of these genes produces a maternally expressed noncoding RNA *(Gtl2)* and is located downstream from the gene encoding fetal growth factor *(Dlk1)* (Schmidt *et al*., 2000). In humans, a corresponding cluster *DLK1/GTL2* was identified on chromosome 14q32. The maternally expressed *GTL2* (*g*ene *t*rap *l*ocus *2*) encodes a noncoding RNA. *DLK1* is a paternally expressed gene encoding a transmembrane protein with a six epidermal growth factor repeat motif. Promoter regions of both *DLK1* and *GTL2* are localized within CpG islands whose methylation pattern is similar to that observed for the *IGF2/H19* cluster (Fig. 5) (Wylie *et al*., 2000).

Comparison of the *Dlk1–Gtl2* domain in humans, mouse, and sheep allowed us to identify 20 conserved sequence elements in all three species. These elements were clustered within the *Dlk1* and upstream of *Gtl2*. A new transcript, with paternal-only transcription, was identified downstream from the *Dlk1*. The transcript originating from *Dlk1* exon 5 is about 3 kb long and does not seem to encode any protein. Its transcription was confirmed in all three species (Paulsen *et al*., 2001).

It also seems that the regulation of imprinting within the *Dlk1–Gtl2* domain depends on similar elements such as the imprinting of *Igf2–H19*. In the region upstream of *Gtl2* there are two consensus binding sites for the CTCF insulator protein, which suggests a similar mechanism of regulation by differential methylation of the DMR and noncoding RNA promoter. The term juxtappositioned reciprocally imprinted genes (JRIG) was suggested to

describe clusters of genes showing an organization and imprinting pattern similar to that of *Igf2–H19* or *Dlk1–Gtl2*. The cluster is composed of a pair of reciprocally imprinted genes, one protein coding and one for noncoding RNA with intervening CTCF binding sites and enhancers downstream from the noncoding gene (Paulsen *et al.*, 2001).

E. Imprinted Noncoding RNA Genes at Human Chromosome 20q13

A gene encoding an α-subunit of the stimulatory G-protein $G_s\alpha$ (*GNAS1*) is located on human chromosome 20q13 (Fig. 6) (Hayward and Bonthron, 2000). Its mouse ortholog, *Gnas*, was mapped to a distal chromosome 2 (Peters *et al.*, 1999). The gene shows a highly complex imprinted expression pattern and its null mutations result in a severe hormone resistance syndrome pseudohypoparathyroidism type 1a (PHP 1a), whose transmission suggested that *GNAS1* is an imprinted gene. Analysis of expression patterns in several fetal tissues showed a biparental origin of $G_s\alpha$ transcripts. Subsequent studies showed a complex organization of the locus and led to discovery of two additional proteins *XLαs* and *NESP55* (*55*-kDa *n*euro*e*ndocrine *s*ecretory *p*rotein). These proteins are translated from two alternatively spliced mRNAs initiated from two novel upstream exons located 35 kb (*XLαs*) and 49 kb (*NESP55*) upstream of *GNAS1* exon 1. These new exons are spliced onto *GNAS1* exon 2. The upstream exons are located within differentially methylated regions and they show monoallelic reciprocally imprinted expression. *XLαs* is expressed from the unmethylated paternal allele, whereas *NESP55* is expressed from the maternal copy of the gene (Hayward and Bonthron, 2000).

This reciprocal imprinting of the two closely located promoters suggested that some regulatory interactions are responsible for the specific expression pattern. The antisense transcript, *GNAS1AS*, starting in the region between the *XLαs* and *NESP55* exons, has been identified. It consists of five exons, four of which are present within the region between the *XLαs* and *NESP55*

FIG. 6 *GNAS1* locus. on human chromosome 20q13. Reciprocally imprinted exons of *NESP55* and *XLαs* are expressed from paternal and maternal alleles, respectively. An antisense, paternally expressed transcript *(GNAS1AS)* originates upstream of *XLαs* and traverses the *NESP55* exon.

exons. The fifth exon is located about 19 kb upstream of *NESP55*. The primary transcript is alternatively spliced, and the longest ORF would code for a 97-amino acid-long protein. The putative protein shows no homology to other known proteins. The expression of antisense transcript was demonstrated to be also imprinted and it is expressed from the paternal chromosome only (Hayward and Bonthron, 2000).

An organization similar to the *Gnas* locus was found in mouse (Wroe *et al.*, 2000). The *Nesp* and *Gnasxl* (corresponding to human *NESP55* and *XLαs*) start sites are associated with differentially methylated regions in paternal and maternal copies, respectively. Methylation analyses did not reveal that any parent-specific methylation was detected for the Gnas/GNAS1 promoters, although in some tissues the genes show maternal-specific methylation (Hayward *et al.*, 2001; Yu *et al.*, 1998). The regulation of monoallelic expression of *NESP55/Nesp* and *XLαs/Gnasxl* seems to involve noncoding transcript antisense and reciprocally imprinted to *NESP55/Nesp*. *GNAS1AS/Nespas* RNA expression could repress the *NESP55/Nesp* by promoter occlusion, localized heterochromatization, or competition for the shared transcription factors. The amount of the *Nespas* transcripts is much lower than that of *Nesp* mRNA suggesting differences in transcription rate or stability. Yet it is still able to negatively regulate *Nesp* expression (Wroe *et al.*, 2000).

F. Air RNA

Imprinting of the maternally expressed *Igf2r* (*i*nsulin-like *g*rowth *f*actor type-*2* *r*eceptor) gene is controlled by a 3.7-kb-long imprinting control element (ICE) known as Region2. Region2 is located within intron 2 of *Igf2r* and contains a maternally methylated CpG island (Wutz *et al.*, 1997). This CpG island is a promoter region for the *Air* (antisense *Igf2r* RNA) gene whose transcript overlaps the silenced paternal *Igf2r* promoter in an antisense orientation. The product of *Air* transcription is a 108-kb-long, unspliced, rich in repetitive elements RNA specifically expressed from the paternal allele (Sleutels *et al.*, 2002).

Region 2 also influences expression of two other imprinted genes, *Slc22a2* and *Slc22a3*, located 110 and 155 kb downstream from the *Igf2r* gene, respectively. Both of these genes are maternally expressed in placenta and deletion of the Region2 results in derepression of the paternal alleles of both genes (Zwart *et al.*, 2001) although their promoters are not overlapped by the *Air* RNA (Fig. 7). This demonstrates that the Region2-ICE is a bidirectional silencer on the paternal allele and directs transcriptional silencing of a cluster of genes spanning ~400 kb. The silencing activity depends on the unmethylated CpG island and the transcription of *Air* RNA. Expression of the three maternal genes correlates with Region2 methylation and repression of

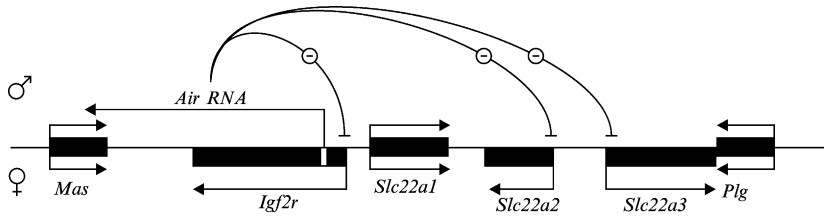

FIG. 7 An imprinted region of mouse chromosome 17. A paternal allele of the maternally expressed *Igf2r* gene is repressed by an antisense transcript (*Air* RNA) originating from the differentially methylated CpG island within intron 2 of *Igf2r* (Region2)—white box. *Air* RNA expression also influences imprinted expression of two other maternally expressed genes, *Slc22a2* and *Slc22a3*. The downstream *Mas1* gene, overlapped by *Air* transcript, is not imprinted and shows biallelic expression.

Air translation. Interestingly expression of *Mas1*, a gene overlapped by a 3′-portion of the *Air* RNA, shows biallelic expression (Lyle *et al.*, 2000).

The importance of Air RNA was demonstrated using a mutated *Air* allele *(Air-T)* obtained by insertion of the polyadenylation signal into the sequence of *Air* RNA, reducing its length to 4%, without disrupting the Region2 function. Mice with a paternally inherited *Air*-T allele showed a phenotype similar to that caused by targeted deletion of the Region2 ICE resulting in loss of transcription of Air RNA and loss of imprinting of the *Igf2r* and consequently its biallelic expression. The mutated *Air* allele was correctly expressed from the paternal allele, but the silencing effect on *Igf2r, Slc22a2*, and *Slc22a3* genes was lost. This was due to loss of methylation within the paternal promoter of *Igf2r*. Interestingly, *Air* RNA influences not only the *Igf2r* expression with which it overlaps, but also two other genes located upstream from its start site. A suggested mechanism of *Air* action involves two steps. At first, noncoding RNA expression results in silencing of the overlapping protein-coding gene by promoter occlusion or *cis*-acting RNA interference. This could result in an induction of the silent chromatin state that would spread and shut off flanking genes. Alternatively, the RNA could recruit repressor proteins to the gene cluster in a manner similar to that observed in X chromosome inactivation in mammalian females (Sleutels *et al.*, 2002). A similar situation is observed in the case of the *LIT1* promoter, whose deletion results in reactivation of genes located both downstream and upstream (Horike *et al.*, 2000).

IV. Noncoding Transcripts from Intergenic Regions

There are several documented cases of transcription activity of the intergenic regions within developmentally regulated gene clusters. Although the function of the transcripts is not fully understood, it seems that they play an

important role in the coordination of gene expression. Intergenic transcripts were identified in the *Drosophila* bithorax complex, mammalian β-globins, and a cluster encoding interleukin (IL)-4 and IL-13 in a subset of T helper cells.

A. Bithorax Complex

In *Drosophila*, the homeotic genes encoded by the bithorax complex are involved in specifying the segmentation of the embryo and determining the body plan (Lewis, 1978). The correct spatial and temporal expression of the three protein-coding genes Ultrabithorax *(Ubx)*, abdominal-A *(abd-A)*, and Abdominal-B *(Abd-B)*, each producing a number of alternatively spliced variants, is crucial for correct development of thoracic and abdominal segments. The expression pattern of *abd-A* and *Abd-B* depends on an array of *cis*-regulatory elements located within the 100-kb intergenic region between the two genes (Sanchez-Herrero and Akam, 1989). Within this region seven genetically defined infraabdominal *(iab)* domains have been identified and their mutations are associated with developmental defects affecting abdominal segments 2 through 8. At the early stages of embryonic development the *iab* regions show transcriptional activity (Sanchez-Herrero and Akam, 1989). Detailed analysis of transcripts originating from the *iab-4* region revealed two 1.7-kb- and 2.0-kb-long polyadenylated noncoding RNAs (Cumberledge *et al*., 1990). The two RNAs are transcribed in the direction opposite to *abd-A*. The two RNAs originate from primary transcripts terminated at alternative polyadenylation sites, from which the same 4.8-kb intron is excised during maturation. Analysis of expression at different stages of development revealed that they are transcribed exclusively in the embryo and become detectable in stage 5 embryos. The transcripts accumulate between 14% and 40% of the egg length from the posterior corresponding to the localization of the primordia for parasegments 8–14 (Cumberledge *et al*., 1990).

A systematic examination of the distribution of the intergenic transcripts from the *iab* regions by *in situ* hybridization revealed that they show highly specific localization along the anteroposterior axis of the blastoderm embryo. The early RNAs originating from each of the *iab* regions are localized within strictly defined borders. The anterior limits of expression depend on the localization of particular *iab* regions relative to the *abd-A* and *Abd-B* genes, while the posterior limits of transcription for all *iab* regions roughly correspond to the posterior limit of *Abd-B* transcription. This colinearity between transcription sites and positions of the *iab* regions within the intergenic region is analogous to expression patterns of the protein-coding genes of the bithorax complex (Fig. 8A). During later stages of development, the

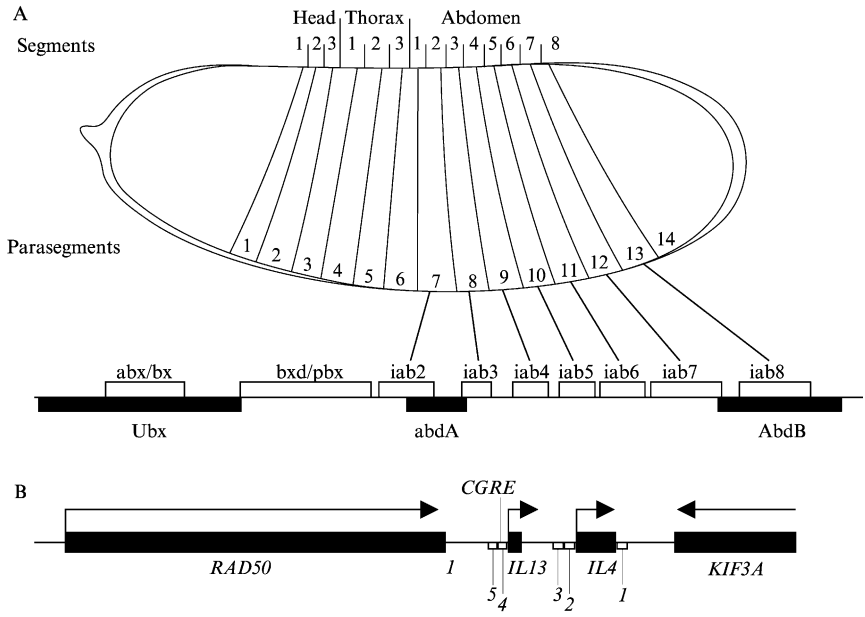

FIG. 8 Noncoding RNAs from intergenic regions. (A) The bithorax complex. The protein-coding genes are shown as black boxes. Regions whose mutations were demonstrated to affect development of particular segments of the *Drosophila* body are represented by white rectangles. During embryo development the iab regions 2 through 8 show localized expression, whose anterior limits roughly correspond to parasegment boundaries. (B) Organization of the IL-4– IL-13 cluster. The intergenic regions IG1–5, for which translation was observed, are shown as white boxes.

transcripts can be detected only in the two most posterior segments of the abdomen. The localization of transcripts within cells appears to be restricted to the nucleus (Bae *et al.*, 2002).

The transcription from the *iab* regions was proposed to play a role in the activation of the *cis*-regulatory elements by interfering with a *Polycomb*-repressing complex, responsible for silencing of homeotic genes (Bender and Fitzgerald, 2002; Hogga and Karch, 2002). Their expression is strictly programmed and the transcripts do not cross the regions' borders. The two antisense transcripts from the *iab-4* and *iab-6* regions (*iab-4as* and *iab-6as*) have been proposed to play a role in preventing the spread of the sense transcript from one *iab* region to another (Bae *et al.*, 2002).

Intergenic transcription within the bithorax complex is not limited only to the *iab* regions. Transcriptional activity was also reported for the bithoraxoid *(bxd)* region (Lipshitz *et al.*, 1987). Early transcripts of 1.1–1.3 kb are alternatively spliced from a 26-kb precursor and appear to be nonprotein

coding. A late 0.8-kb transcript can be translated to produce a 101-amino acid-long protein. Because the *bxd* region also contains *Polycomb*-group response elements (Hodgson *et al*., 2001), the early noncoding transcripts may also perform the function of silencing repressors.

B. IL-4/IL-13 Gene Cluster

A subset of T-helper cells (Th2) is involved in cell-mediated immune responses. They produce IL-4, IL-5, IL-9, IL-10, and IL-13, which induce proliferation and differentiation of B cells into plasma cells that produce and secrete antibodies (Dong and Flavell, 2000). The activation of Th2 cells leads to the regulated expression of the *IL-4* and *IL-13* genes located in tandem on human chromosome 5 and a syntenic region of chromosome 11 in mouse (Frazer *et al*., 1997). This cluster is flanked by two constitutively expressed genes *RAD50* and *KIF3A* encoding double-stranded break repair protein and microtubule motor protein, respectively (Fig. 8B). Experiments with transgenic mice suggested that in addition to the earlier identified *cis*-acting elements in the proximal promoter of the IL-4 gene there must exist additional factors responsible for its regulation. An analysis of transcription products from this region in $CD4^+$ T cells revealed 135- to 266-nt-long, polyadenylated transcripts from all three (i.e., *RAD50-IL13, IL13-IL4*, and *IL4-KIF3A*) intergenic regions (Fig. 8B). They are constitutively transcribed, even in the absence of activity of the interleukin genes. This implies that they are coming from independent transcription units. In contrast to mRNAs from the neighboring protein-coding genes, the intergenic RNAs are not detectable in the cytoplasm and show nuclear localization exclusively. In HeLa cells, which do not express IL-4 and IL-13 genes, only one intergenic transcript (IG1) has been identified (Rogan *et al*., 1999). Thus, it seems that the transcription of IG2–IG5 preceding interleukin genes is associated with the potential for their expression.

One possible explanation for the role of the intergenic transcripts is that they are a result of the activity of a protein complex responsible for chromatin remodeling (Takemoto *et al*., 2000). The differentiation of Th2 cells was found to be associated with hyperacetylation of histone H3 and hypomethylation of the CpG islands (Yamashita *et al*., 2002). A key factor associated with Th2 cell differentiation linked to chromatin structure remodeling is the GATA3 transcription factor. This protein was shown to be necessary and sufficient for an induction of Th2-specific interleukin production and secretion (Zheng and Flavell, 1997). A conserved GATA3 response element (CGRE) located upstream from IL-13 contains four binding sites for GATA3 and one for a cAMP-responsive element-binding protein (CBP) (Fig. 8B). The latter possess an intrinsic histone acetyltransferease activity.

The CGRE element was proposed to provide a site for the assembly of chromatin remodeling complexes consisting of CBP and RNA polymerase II. It is possible that the observed intergenic transcripts are simply a result of its activity (Yamashita *et al.*, 2002). However, their involvement in some other processes cannot be excluded.

C. β-Globin Locus

In humans, the 70-kb-long β-globin locus consists of five erythroid-specific genes: embryonal (ε), fetal ($^{G}\gamma$ and $^{A}\gamma$), and adult (δ and β), whose expression is under the control of the β-LCR (*l*ocus *c*ontrol *r*egion). The order of genes within the cluster reflects the order in which they are expressed during development.

Analysis of nascent transcripts from the β-globin gene cluster revealed that the protein-coding regions are not the only ones that are active. Both the LCR and intergenic regions produce specific noncoding RNAs. The LCR transcription terminates approximately 0.4 kb upstream of the ε-globin gene and its products are restricted to the nucleus (Ashe *et al.*, 1997). The LCR region contains multiple binding sites for erythroid-specific transcription factors GATA-1, NF-E2, and EKLF within DNase I hypersensitive sites (Routledge and Proudfoot, 2002). The intergenic transcripts originate from the same strand as globin mRNAs and are also retained within the nucleus (Ashe *et al.*, 1997).

The precise role of transcription of the LCR and intergenic regions is not known, but it is specifically restricted to erythroid cells. Interestingly, a transient expression of globin genes in nonerythroid cells can induce transcription from the intergenic regions without activating protein-coding domains (Ashe *et al.*, 1997). It has been proposed that transcription is required for the establishment and maintenance of an open chromatin conformation within the locus, which would in turn facilitate its expression (Plant *et al.*, 2001). An alternative role for the transcripts has been proposed in recruiting *trans*-acting factors and RNA polymerase II to the promoter sites (Tuan *et al.*, 1992). These two explanations are not mutually exclusive. It is possible that, as in the case of the IL-13–IL-4 cluster, the transcription factors binding recruits the chromatin remodeling complex.

V. Noncoding RNAs in Posttranscriptional Gene Regulation

RNA molecules by their ability to form double-stranded duplexes with other complementary RNA sequences offer an excellent mechanism of gene expression regulation on the posttranscriptional level. RNAs antisense to mRNA

were used as research tools for gene inactivation for many years and the discovery of several natural antisense RNAs prompted speculations about their role in the regulation of gene expression. In fact, in recent years we realized that sense–antisense RNA instructions play a very important role in the regulation of expression of many genes in both prokaryotes and eukaryotes.

A. Bacterial Posttranscriptional RNA Regulators

The first regulatory RNAs involved in plasmid copy number control in bacteria were identified at the beginning of the 1980s (Stougaard *et al.*, 1981; Tomizawa *et al.*, 1981). The two RNAs, RNAI and CopA, encoded by ColE1 and R1 plasmids, respectively, represent *cis*-encoded RNA regulators, because they originate from the same locus as their target RNAs, but they are transcribed in the opposite direction. In bacteria, most of the antisense regulatory RNAs are *trans*-encoded, or they originate from loci different than their respective target RNAs.

Several *trans*-encoded regulatory RNAs were isolated from various bacteria (Wagner and Vogel, 2003). OxyS RNA is a 109-nt RNA responsible for regulation of expression of about 40 genes as a part of a defense system against oxidative damage. Oxidative stress results in rapid accumulation of high levels of OxyS RNA (Altuvia *et al.*, 1997). One of the affected genes is *fhlA*, whose mRNA forms a complex with OxyS RNA. The nucleotide sequences complementary to OxyS within *fhlA* mRNA are located near the ribosome-binding site and their interactions with OxyS block translation (Altuvia *et al.*, 1998; Argaman and Altuvia, 2000). OxyS is also a negative regulator of *rpoS* (stress σ factor of RNA polymerase) expression. In this case it appears that the regulatory RNA competes for Hfq protein, which is required for *rpoS* mRNA translation (Zhang *et al.*, 1998). Another regulatory RNA involved in *rpoS* expression is DsrA, an 87-nt-long RNA induced and stabilized in low temperature and implicated in regulation of a number of genes in *Escherichia coli* (Sledjeski and Gottesman, 1995). Translational activation of the *rpoS* message by DsrA RNA depends on direct RNA:RNA interactions between the 5′-untranslated region of *rpoS* mRNA and the 5′-portion of DsrA. This interaction disrupts a secondary structure within the *rpoS* mRNA that serves as a *cis*-acting inhibitor of translation (Majdalani *et al.*, 1998). Interestingly, *rpoS* is also activated by osmotic shock, when there is no accompanying increase in DsrA. The secondary structure occluding the ribosome-binding site is targeted by another riboregulator—RprA (Majdalani *et al.*, 2002). A similar *cis*-acting RNA structure that prevents translation is observed in regulation of α-toxin expression by RNAIII in *Staphylococcus aureus* (Morfeldt *et al.*, 1995). On the other hand, DsrA RNA

is a negative regulator, responsible for repression of translation of H-NS mRNA, encoding a global transcription factor. This is accomplished by complementary duplex formation with the 5′- and 3′-portions of the H-NS mRNA ORF (Fig. 9) (Lease and Belfort, 2000a,b). There is a possibility that the DsrA-mediated regulation affects several other genes, which show regions of sequence complementarity to the regulatory RNA. Taking into

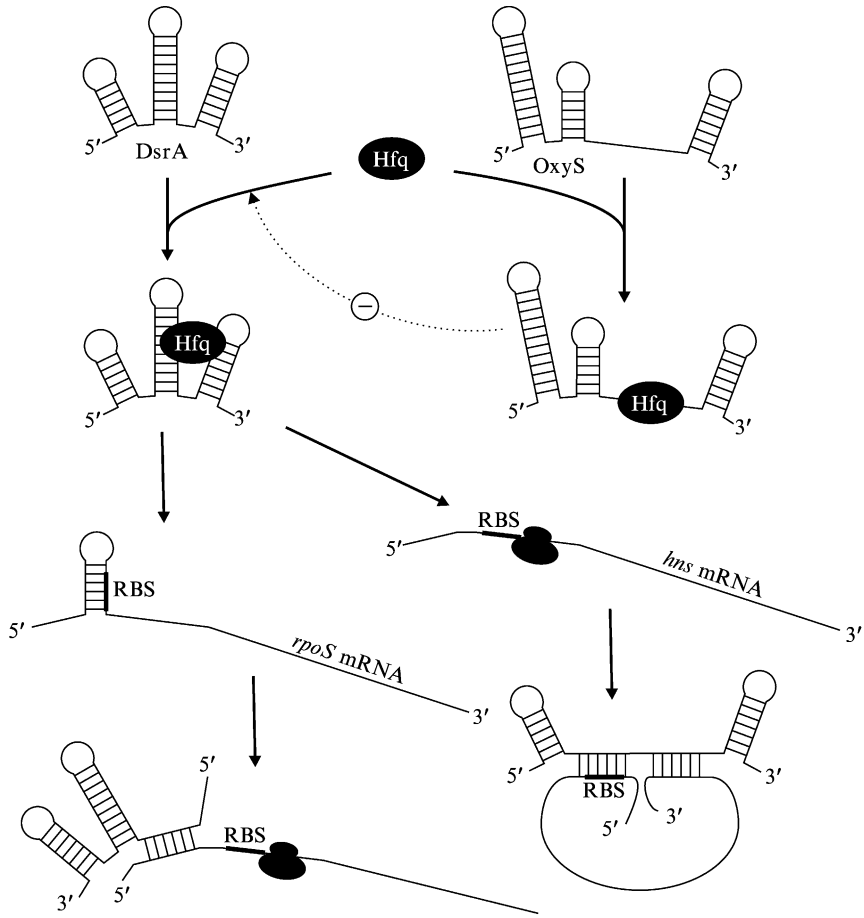

FIG. 9 Regulation of two transcription factors, RpoS and H-NS, expression in bacteria by noncoding RNAs. Translation of *rpoS* mRNA is suppressed by a secondary structure within the 5′-UTR occluding the ribosome-binding site (RBS). An Hfq protein facilitates base pairing between DsrA RNA and *rpoS* mRNA, and *hns* mRNA. DsrA–*rpoS* mRNA interactions open the RBS and allow initiation of translation. Base pairing between DsrA and *hns* mRNA blocks expression of H-NS by masking the RBS and increasing *hns* mRNA turnover. OxyS RNA antagonizes DsrA action competing for Hfq protein.

account the two aforementioned cases, the localization of the potential target sequence relative to the translation initiation would determine if the DsrA would have an inhibitory or stimulatory effect. (Lease *et al*., 1998). MicF RNA is a stress response regulatory RNA, whose expression is induced by various stress conditions (Delihas and Forst, 2001). MicF RNA binding to its target, a translation initiation region of the mRNA encoding the outer membrane porin, OmpF, results in inhibition of translation.

In addition to the regulatory effect on monocistronic mRNAs, bacterial regulators were shown to affect coordinated expression of the genes within single operons. Spot42 RNA was identified for the first time three decades ago, but only recently a function of this RNA has been found. Spot42 RNA is responsible for discoordinate expression of cistrons encoded by the galactose operon *(galETKM)* (Møller *et al*., 2002). A base pairing of Spot42 RNA with the *galK* ribosome-binding site results in inhibition of GalK translation and the decrease of the GalK/GalET ratio with as yet unknown consequences. A sequence complementarity suggests that another possible target may be the *sucC* cistron of the *sucABCD* operon for TCA cycle enzymes. A similar role may be played by a RyhB RNA in regulation of the *sdhD* gene within the *sdhCDAB* operon (Masse and Gottesman, 2002).

RNA–RNA interactions between the regulatory ncRNAs and target mRNAs are facilitated by the Hfq protein (Brescia and Sledjeski, 2003). It plays a crucial role in DsrA-dependent repression of *hns* and stimulation of *rpoS* translation (Sledjeski *et al*., 2001) and in interactions between OxyS RNA and *fhlH* mRNA (Zhang *et al*., 2002). The Hfq protein was shown to perform a variety of roles and seems to be one of the key elements in RNA-dependent regulation in bacteria.

The number of regulatory RNAs and the genes they are affecting in bacteria is constantly growing. The regulation is not limited to sense–antisense RNA interactions, but also may involve specific protein binding and modulation of their functions as in the case of 6S RNA (Wassarman and Storz, 2000).

B. Micro-RNAs

Micro-RNAs (miRNAs) constitute a class consisting of the smallest (20–28 nt long), functional RNA molecules identified to date (Moss, 2003). The first two miRNAs, *lin-4* and *let-7*, were identified in *C. elegans* during studies on heterochronic mutations that affect the timing and sequence of events in postembryonic development (Lee *et al*., 1993; Reinhart *et al*., 2000). They were first called small temporal RNAs (stRNAs), which reflected their function as temporally regulated developmental switches. In the past 2 years, intensive research led to the discovery of hundreds of new miRNAs both in animals and in plants (Moss, 2003).

The best studied miRNA is *lin-4* of the nematode *C. elegans*. A region of the genome, whose mutation resulted in abnormal development, was narrowed down to an ~700-bp fragment, which was able to rescue the mutant phenotype (Lee *et al*., 1993). Interestingly, the isolated piece of DNA did not seem to encode a protein since no ORF could be identified. Northern blot analysis revealed the presence of two RNAs with lengths 22 nt and 61 nt, called lin-4S and lin-4L, respectively. Because the point mutation responsible for the lin-4 mutant phenotype falls within the lin-4S region, it was assumed that this unexpectedly short RNA is an active product of the *lin-4* locus. Subsequent nuclease mapping showed that the lin-4S RNA is in fact processed from the lin-4L precursor molecule. The role of lin-4 RNA was shown to be that of a translational repressor of at least two genes, *lin-14* and *lin-28* (Wightman *et al*., 1993; Moss *et al*., 1997). The *lin-14* gene is expressed at the early stages of *C. elegans* development (Ruvkun *et al*., 1989) and the timing of its repression correlates with the accumulation of *lin-4* transcripts at the end of the first larval stage (Feinbaum and Ambros, 1999). The translational repression depends on the presence of seven short sequence elements, with partial complementarity to lin-4 RNA, within the 1.6-kb 3′-untranslated region (3′-UTR) of *lin-14* mRNA (Ha *et al*., 1996). These elements are sufficient for lin-4 RNA-dependent repression and the mutants missing the regulatory sequences within the *lin-14* 3′-UTR show constitutive expression of the LIN-14 protein, which results in a developmental phenotype identical to *lin-4* mutants (Wightman *et al*., 1993; Ha *et al*., 1996). In the case of the *lin-28* gene there is only one 15-nt-long regulatory element within the 3′-UTR, whose deletion results in a phenotype identical to that observed in the absence of *lin-4* expression (Moss *et al*., 1997).

Although in both *lin-14* and *lin-28*, the *lin-4* RNA-responsive regulatory elements are located within the 3′-UTRs, it seems that in each case the mechanism of repression may be different and depend on the nucleotide sequence context. Obviously, the involvement of the 3′-UTRs in *lin-4*-mediated suppression suggested that the regulation takes place on a post-transcriptional level. This is also consistent with the observation that the mRNA levels remain unaltered, while there is a marked decrease in protein production (Olsen and Ambros, 1999; Seggerson *et al*., 2002). In the case of *lin-14* repression the presence of bulged cytosine in the RNA–RNA duplex is observed in four of seven elements complementary to *lin-4* RNA. This single nucleotide bulge is critical for *lin-4* activity (Lee *et al*., 1993; Ha *et al*., 1996). On the other hand, *lin-4* RNA binding to the regulatory sequence in *lin-28* 3′-UTR does not produce such a bulge, and there are possibly other factors responsible for down-regulation of *lin-28* (Moss *et al*., 1997; Seggerson *et al*., 2002).

The second *C. elegans* miRNA gene whose mutants show developmental defects is *let-7* (Reinhart *et al*., 2000). The gene showed all the features that

were observed earlier for *lin-4*. The isolated DNA did not code for protein. A small, 21-nt-long RNA, processed from a longer ~70-nt-long precursor, was found in large quantities by Northern blot analysis. The expression of *let-7* RNA is developmentally regulated and was proposed to act in a manner similar to *lin-4*. In fact, it has been demonstrated that the 3′-UTR of the *lin-41* gene contains two sites with partial complementarity to *let-7* RNA, whose presence is responsible for repression of translational (Slack *et al*., 2000). A systematic survey revealed that *let-7* RNA is present and almost totally conserved in all bilaterally symmetrical animal groups (Pasquinelli *et al*., 2000). Specific expression patterns of *let-7* in humans and *Drosophila* suggest that as in the nematodes the RNA may be involved in the regulation of development and/or differentiation.

Several systematic studies led to the discovery of a large number of new miRNA species in several animal and plant species. These results clearly indicate that micro-RNAs constitute a large class of noncoding RNAs, which is very common in eukaryotes. A number of new miRNA genes were identified in *Caenorhabditis* (Lau *et al*., 2001; Lee and Ambros, 2001), *Drosophila* (Lagos-Quintana *et al*., 2001), humans (Lagos-Quintana *et al*., 2001, 2003; Mourelatos *et al*., 2002), and mouse (Lagos-Quintana *et al*., 2002, 2003). A common feature of all animal micro-RNAs is their size of 20–25 nucleotides and the fact that that they are processed from one side of stem–loop precursor molecules. The number of known examples is growing, yet in most cases there are few clues as to the function of particular miRNAs, leaving a lot of questions unanswered. It is not clear, for example, whether the miRNA species conserved in evolution such as *let-7* or miR-1 play similar roles in different organisms.

The micro-RNAs are not limited to the animal world. Cloning efforts led to identification of over 100 miRNAs from *Arabidopsis thaliana* and several homologs in the rice genome (Reinhart *et al*., 2002; Llave *et al*., 2002a). The lengths of plant miRNAs were between 16 and 25 nucleotides, with species 21–24 nt long constituting a majority. Most of the genes encoding these RNAs were found within the intergenic regions. They are most likely to constitute independent transcription units, as there is no correlation with the polarity of surrounding genes. A second group of miRNAs originates from coding sequences or introns of protein genes and transposon-like elements (Llave *et al*., 2002a). A common feature of plant and animal micro-RNAs is their biogenesis. In both cases, they are processed from precursors, capable of forming long, imperfectly paired stems. As in animals, plant miRNA expression is developmentally regulated and tissue specific (Reinhart *et al*., 2002; Llave *et al*., 2002a), which indicates possible involvement in development and differentiation. It has also been proposed that some of the micro-RNAs can serve not only as local intracellular, but also as

systemic signals (Llave *et al.*, 2002a), based on the ability of RNAs to travel long distances through plasmodesmata (Wu *et al.*, 2002).

Studies on micro-RNAs showed other parallels with the phenomenon of RNA interference (RNAi) whereby sequence-specific gene silencing is achieved in response to double-stranded RNA (dsRNA) (Hannon, 2002) (Fig. 10). This kind of response to dsRNA, also known as posttranscriptional gene silencing (PTGS) in plants (Escobar and Dandekar, 2003) or quelling in fungi, is evolutionarily conserved in eukaryotes. A mechanism of RNAi involves generation of short interfering RNA (siRNA) molecules from exogenous or endogenous dsRNA by Dicer, a member of the RNase III family (Bernstein *et al.*, 2001). The same enzyme is responsible for maturation of precursors of miRNAs in animals (Ketting *et al.*, 2001; Hutvagner *et al.*, 2001). Its plant homologue, CARPEL FACTORY/SHORT INTEGUMENTUM (CAF/SINI), was identified in plants as a factor involved in flower development (Jacobsen *et al.*, 1999) and then as a component of the miRNA maturation pathway (W. Park *et al.*, 2002; Reinhart *et al.*, 2002). The active RNA molecules generated by Dicer are similar in size (~21 nt),

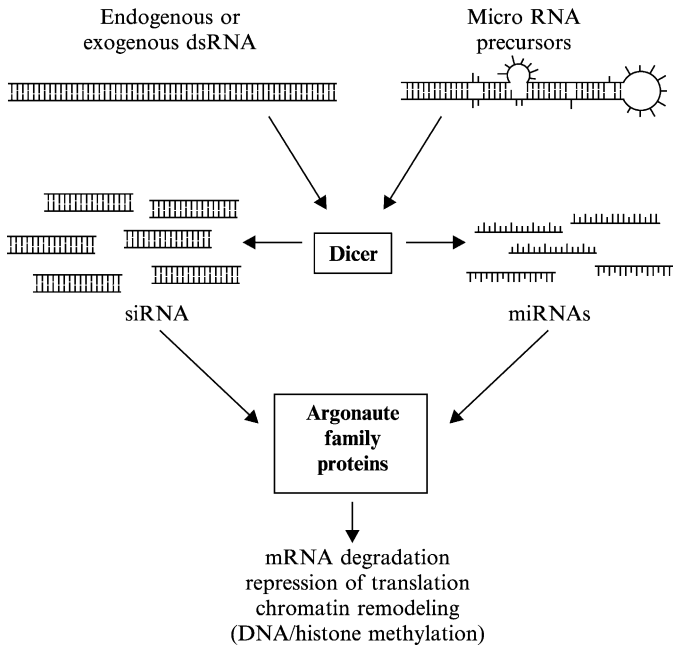

FIG. 10 Micro-RNAs and RNA interference (RNAi) pathways. Single-stranded miRNAs and double-stranded siRNAs are produced by Dicer endonuclease. The Argonaute family proteins direct the active RNAs to various pathways associated with RNA stability, translation, and chromatin structure.

although the miRNAs show greater variation, especially in plants. The presence of enzymes associated with miRNA biogenesis and RNAi in virtually all eukaryotes suggests that these mechanisms are very old and they may predate the origin of eukaryotic cells. In addition to the defense against transposable elements and viruses, these mechanisms offer very flexible and highly specific means of gene expression regulation on both transcriptional and posttranscriptional levels (Brantl, 2002; Hannon, 2002).

Apart from the nature of active RNAs, single-stranded miRNAs vs. double-stranded siRNAs, it has often been emphasized that the difference between the two is in the way they influence gene expression. Micro-RNAs were assumed to operate as translation inhibitors, while siRNAs were thought to direct RNA degradation. The picture was complicated by the finding that in plants miRNAs can act in a way analogous to that of siRNAs. One of the miRNAs identified in *A. thaliana* (miR-179/miRNA 39) shows perfect sequence complementarity with portions of mRNAs encoding three members (*SCL6-II, SCL6-III*, and *SCL6-IV*) of the *Scarecrow-like (SCL)* family of putative transcription factors (Llave *et al*., 2002b). In rice, there are four *SCL* genes that show similar complementarity (Reinhart *et al*., 2002). In *A. thaliana*, it has been demonstrated that miRNA 39 regulates the expression patterns of these genes by mediating RNA-induced cleavage of mRNA by a mechanism analogous to that observed in RNAi (Llave *et al*., 2002a). This finding suggests that there are at least two possible ways in which micro-RNAs can affect expression of message on a posttranscriptional level: either by inhibition of the translation process or by directing targeted mRNA breakdown. The distinction can be made, based on the nature of complementary interactions between the miRNA and its target. StRNAs such as *lin-4* or *let-7* do not form perfectly paired duplexes with the target mRNA regions. On the other hand, the RNAi mechanism involving RISC (RNA-induced silencing complex) requires an exact match between the interfering RNA and target sequence where the short RNA molecule performs the role of a guide within the multiple turnover enzymatic complex.

The mechanisms of translation inhibition are also not fully understood. It seems, therefore, that the regulatory effect of miRNA on the expression of a particular mRNA does not depend solely on a simple base-pair complementarity, but requires additional structural constraints and possibly interactions with other factors. In *Drosophila*, two sequence elements, Brd box and K box, found in 3′-UTRs of developmentally regulated genes, were proposed to mediate miRNA responsiveness (Lai, 2002). Regulatory mechanisms involving both miRNAs and siRNAs require participation of proteins belonging to the Argonaute family (Carmell *et al*., 2002; Dostie *et al*., 2003; Zilberman *et al*., 2003). These highly specialized proteins were shown to be indispensable for processing of precursors of miRNAs and seem to

determine the fate of various Dicer products, directing them to different regulatory pathways (Carmell *et al*., 2002).

Yet another breakthrough in the elucidation of RNAi and miRNA significance was made by finding that small RNA molecules, generated by enzymes involved in RNAi, were involved in transcriptional silencing of heterochromatic DNA (Stevenson and Jarvis, 2003). This finding came from identification of ~20-nt-long dsRNAs homologous to centromeric repeats, and the observation that silent transgenes in heterochromatic centromeric regions can be activated in yeast mutants lacking proteins involved in RNAi (Reinhart and Bartel, 2002; Volpe *et al*., 2002). The mechanism of heterochromatin formation involves methylation of the lysine residue at position 9 in histone H3 (Hall *et al*., 2002). This bears some similarity to X chromosome inactivation where such a modification of histones is crucial for the initiation and spreading of the silencing process (Heard *et al*., 2001). Small RNAs can also be involved in gene silencing employing a mechanism of RNA-directed DNA methylation (Matzke *et al*., 2003).

Recently, a small region of chromosome 13q14 frequently deleted in chronic lymphocytic leukemias was demonstrated to harbor two micro-RNA genes *miR15* and *miR16*. It is not known if these miRNAs play some role in $CD5^+$ B cell differentiation. Arginyl-tRNA synthetase (ArgRS) mRNA was identified as a putative target for *miR16*. The expression of ArgRS were found to correlate with levels of *miR16*, suggesting a posttranscriptional regulatory mechanism. However, the significance of this finding requires further investigation (Calin *et al*., 2002).

C. Natural Antisense Transcripts

In eukaryotes, there is a growing number of documented cases of natural antisense transcripts, produced from the opposite strand of the protein-coding genes. Expression of the antisense RNAs reduces the levels of expression of a sense gene. In most cases, the mechanism of regulation is unknown, but it is assumed that the antisense transcripts form duplexes with mRNAs, which prevents translation. The RNA–RNA duplexes can also be used as substrates for Dicer to produce siRNAs and further trigger RNAi pathways leading to mRNA degradation and/or gene silencing on the chromatin level.

A proliferating cell nuclear antigen (PCNA) is a nuclear protein that acts as a cofactor of DNA polymerase δ and is involved in DNA repair. The expression of the *PCNA* gene is regulated on transcriptional and posttranscriptional levels and can be induced by stimulation with serum or growth factors. An analysis of the 5′-portion of the gene revealed that its cell cycle regulation depends on elements located within its first intron. The putative cell cycle-regulated promoter directs transcription of a short unspliced noncoding RNA,

which covers a fragment of the first intron, first exon, and ~100 bp of the 5′-UTR of the *PCNA* mRNA. This antisense RNA was shown to be transcribed in all normal and cancer cells, but the levels of expression in normal cells are higher. Unlike the *PCNA*, the antisense transcript expression is not cell cycle stage dependent. It has been suggested that the expression of the *PCNA* gene depends on the ratio of sense to antisense RNAs (Tommasi and Pfeifer, 1999).

A developmentally regulated antisense transcript was found to be expressed from the *Msx1* locus in mammals. *Msx1* is a homeobox gene required for craniofacial skeleton formation. It has been shown that the antisense transcript (*Msx1*-AS) overlaps 1 kb of the sense mRNA spanning approximately half of intron 1 and the entire exon 2. *Msx1*-AS can suppress expression of the Msx1 protein, a transcription factor responsible for inhibition of cell differentiation. This suggested that *Msx1*-AS RNA levels may be responsible for the regulation of the transition from cell proliferation to terminal differentiation (Blin-Wakkach *et al.*, 2001). Antisense, polyadenylated transcripts have also been implicated in the regulation of another homeodomain transcription factor, EMX2, involved in the central nervous system and urogenital development (Noonan *et al.*, 2003).

In *Neurospora crassa*, the *frequency (frq)* gene encodes a component of the circadian clock. It has been shown that there are two antisense transcripts, which almost totally overlap the *frq* gene. The antisense RNAs show a cyclic pattern of expression, but 180° out of phase with the sense gene. Like the *frq*, the antisense transcripts are induced by light and their role was shown to prevent strong resetting of the clock in response to light (Kramer *et al.*, 2003).

The number of naturally occurring antisense transcripts from protein coding genes is constantly growing. In addition to the earlier discussed mechanisms involving RNAi pathways or simple inhibition of translation by base pairing with complementary regions of a matching mRNA, antisense RNAs can affect gene expression on other levels. The expression of *N-myc* depends on an antisense RNA whose binding to the mRNA is likely to suppress splicing (Krystal *et al.*, 1990). An antisense RNA complementary to intron 1 of the *p53* gene was proposed to interfere with export of the *p53* mRNA from the nucleus (Khochbin *et al.*, 1992).

VI. Other Regulatory Functions of Noncoding RNAs

A. Noncoding RNAs as Modulators of Proteins Activity

One of the mechanisms by which noncoding RNAs can influence cellular processes is by a modulation of the activity of proteins. Noncoding RNAs of viral, prokaryotic, and eukaryotic origin have been demonstrated to

influence the activity of proteins in various ways. RNA binding to proteins can change their activity either by interfering with active sites or by changing their conformation. High selectivity of RNA binding can therefore provide means for very specific regulatory interactions. The regulatory RNAs that influence protein activity were shown as inhibitors of enzymes and as transcriptional regulators.

1. EBER1 RNA

An RNA-dependent protein kinase (PKR) is an interferon-induced element of the cellular defense system against viral infection. The enzymatic activity of the protein depends on binding of RNA duplexes over a 24 bp length that induces its autophosphorylation and dimerization. The enzyme is responsible for the inactivation of the eukaryotic initiation factor 2α (eIF2α) and consequently shutting off the cellular translational apparatus. As a countermeasure against PKR action, some viruses, including adenovirus and Epstein–Barr virus, developed specific constitutive noncoding transcripts whose role is to bind and repress PKR activity. Epstein–Barr virus (EBV), a human B cell lymphotropic herpesvirus, constitutively expresses two noncoding RNA molecules EBER1 and EBER2, which are 167 nt and 172 nt long, respectively. Both RNAs are highly structured and they have been identified in both cytoplasm and nucleus of the EBV-infected cells. One of the possible roles for EBER RNAs is inhibition of PRK activity (Sharp *et al.*, 1993). EBER1 was demonstrated to bind PKR, ribosomal protein L22, and the La antigen (Glickman *et al.*, 1988; Toczyski *et al.*, 1994). Interestingly, all of these host proteins interact with different structural domains of EBER1, which suggests a highly specific nature of these associations (Vuyisich *et al.*, 2002). The EBER RNAs were also found to play a key role in the maintenance of malignant phenotypes of Burkitt's lymphoma cells (Nanbo and Takada, 2002).

2. SRA RNA

The nuclear receptors of steroid hormones are ligand-inducible transcription factors responsible for coordinated expression of a number of genes involved in metabolism, development, and reproduction. The receptors constitute a family of evolutionarily related proteins. Almost all of them are associated with the activation function AF2 in the carboxyl-terminus of the ligand-binding domain, which is crucial for hormone-dependent transactivation (Lanz and Rusconi, 1994). The amino-terminal modulatory domain contains a transactivation function (AF1) governing target gene specificity (Tora *et al.*, 1988). Activated nuclear receptors mediate establishment and stabilize the preinitiation complex at the promoter of the target gene. In addition to the

components of the preinitiation complex, a number of proteins that associate with activated nuclear receptors have been isolated. A common feature of these proteins, collectively named nuclear receptors coactivators, is their ability to increase transactivation without changes in basal transcriptional activity (McKenna and O'Malley, 2002). One of the transcriptional coactivators identified was a noncoding RNA transcript called steroid receptor activator (SRA) RNA (Lanz *et al*., 1999). The sensitivity to antisense oligodeoxynucleotide insensitivity to nonsense mutations within a potential reading frame suggested that these transcripts perform their role as functional RNA molecules. Attempts to produce protein in an *in vitro* translation system were also unsuccessful. The RNA was shown to exist in several cell type-specific splicing variants (isoforms) with a common core region and variable flanking sequences. The primary transcript was shown to be produced from an independent gene located on human chromosome 5q31 (Lanz *et al*., 2002).

SRA RNA was demonstrated to form a complex with the steroid receptor coactivator-1 (SRC-1). SRA shows strong coactivator activity with receptors for progestins, estrogens, androgens, and glucocorticoids (Lanz *et al*., 1999). The SRC-1–SRA RNA interaction was proposed to be mediated by a subfamily of DEAD-box RNA-binding proteins, p72/p68, which were shown to directly bind SRA RNA (Watanabe *et al*., 2001). Secondary structure predictions and covariation analysis revealed five structural elements within the SRA RNA core sequence, which are important for its function as a coactivator. The mutations, altering secondary structures, but not the predicted amino acid sequence of a putative ORF were shown to reduce transcriptional coactivation by SRA RNA (Lanz *et al*., 2002). SRA RNA can also bind a hormone-induced transcriptional repressor, SHARP. The interactions are mediated by a specific RNA-binding domain RRM, whose deletion abolishes the activity. Thus, one of the effects of SHARP activity might be competition with the nuclear receptors for SRA RNA, resulting in down-regulation of respective genes (Shi *et al*., 2001).

3. 7SK RNA

A positive transcription elongation factor (P-TEFb) consists of a CDK9/cyclin T1 heterodimer. Its function as a transcriptional activator depends on phosphorylation of RNA polymerase II, resulting in the formation of processive elongation complexes. P-TEFb is a Tat cofactor stimulating human immunodeficiency virus type 1 (HIV-1) transcription by interactions with the Tat and TAR structure. An evolutionarily conserved small nuclear 7SK RNA was identified as a specific P-TEFb cofactor (Yang *et al*., 2001; Nguyen *et al*., 2001).

7SK RNA was found as a component of ribonucleoprotein particles containing P-TEFb in HeLa cells. Quantitative analysis suggested that

approximately half of cellular P-TEFb is present as a stable complex with 7SK RNA. The transcriptional activity of P-TEFb was shown to be inhibited by the association with 7SK RNA. The inhibitory effect of 7SK RNA on P-TEFb activity is the result of suppression of kinase activity of CDK9. Another effect was observed in the HIV-1 transcription assay, which demonstrated that 7SK RNA-bound P-TEFb cannot form compexes with the viral promoter. The complexes of 7SK RNA with P-TEFb can be readily dissolved by actinomycin D treatment or UV irradiation. These factors are also known to significantly increase HIV-1 transcription, suggesting a crucial role of 7SK RNA and P-TEFb in the regulation of this process.

Another transcription factor, TFIIH, responsible for the initiation of transcription by RNA polymerase II, was also found to be associated with RNA, whose presence enhances its activity. The RNA involved turned out to be U1 snRNA, which plays a critical role in pre-mRNA splicing (Kwek *et al.*, 2002).

The discovery of RNAs that play a role in transcription activation was not totally unexpected. In HIV RNA, the TAR sequence is responsible for binding TAT protein, which recruits cyclin-dependent protein kinases stimulating transcription elongation (Jones, 1997). An *in vitro* selection approach resulted in isolation of RNA molecules, which show activity of transcriptional activators in yeast. The randomized 10-nt sequence was predicted to form a loop at the end of an invariable stem. The selected RNAs show little variability, which suggested that they may interact with one protein involved in transcription, yet the protein has not been identified so far (Saha *et al.*, 2003).

4. 6S RNA

Bacterial 6S RNA is another abundant RNA species whose function remained a mystery for over 30 years. No aberrant phenotype was associated with either overexpression or null mutations in *E. coli*. This RNA was shown to be responsible for inhibition of the activity of RNA polymerase in stationary phase by forming a stable and highly specific complex with the σ^{70}-RNA polymerase holoenzyme. Although the precise function of this association is not known, it has been proposed that 6S RNA may be responsible for altering promoter utilization in the stationary phase or providing a storage particle for the σ^{70}-holoenzyme during starvation (Wassarman and Storz, 2000).

B. Tissue-Specific and Developmentally Regulated ncRNAs

Most of the regulatory noncoding RNAs show highly specific patterns of expression. Here we would like to review several groups of noncoding RNAs whose expression is tissue specific or related to development and stress conditions. In most cases, these RNAs do not belong to any distinct class,

and their functions largely remain unknown. Their expression is often restricted to very specialized cell types or is in response to certain environmental conditions, which suggests they play important roles in the cell.

1. Noncoding RNAs in the Nervous Tissue

Nervous tissue has been a very rich source of novel noncoding RNAs. In the nervous tissue of primates and rodents specific RNA polymerase III transcripts were identified (Martignetti and Brosius, 1993, 1995). The primates BC200 RNA and rodents BC1 RNA, although similar in size and biogenesis, represent two distinct classes of RNAs. They are about 200 and 150 nucleotides long, respectively, but the size varies depending on the species. BC1- and BC200-encoding genes originated by retroposition of $tRNA^{Ala}$ (BC1) (Kim *et al*., 1994) and a monomeric Alu repeat (BC200) (Martignetti and Brosius, 1995). In both cases the 5′-parts come from the ancestral gene, followed by the A-rich central domain and unique 3′-terminal domain. Both BC200 and BC1 RNAs associate with proteins forming ribonucleoprotein particles 11.4S and 8.7S, respectively (Cheng *et al*., 1996; Kobayashi *et al*., 1991). The central A-rich region facilitates interactions with the poly(A)-binding protein (Muddashetty *et al*., 2002). Both RNAs show similar expression patterns, and in addition to the nervous system, they have been found in testes of rodents and primates (Kuryshev *et al*., 2001; Rozhdestvensky *et al*., 2001). In the nervous system, BC1 and BC200 RNA expression can be detected in neurons but is missing in glial cells. In neurons, the RNAs are found in cell bodies, but also in dendrites (Tiedge *et al*., 1991, 1993). It has been proposed that the function of these RNAs is to provide translational control of gene expression in subcellular domains (Wang *et al*., 2002). It has been demonstrated that BC1 can bind a protein involved in fragile X syndrome (FMRP), which acts as a translational repressor in synapses. The role of RNA in a proposed mechanism is to ensure a specificity of recognition of target mRNA by binding to a complementary sequence within a 3′-UTR (Zalfa *et al*., 2003).

Brain-specific expression was also demonstrated for certain snoRNAs in mouse and humans. Because most of the snoRNAs perform housekeeping functions in modification of rRNAs, snRNAs, and some tRNAs it was interesting to find their tissue-specific localization. Most of them show imprinted patterns of expression, and all of them are encoded within introns of host genes and their functions remain unknown (Cavaillé *et al*., 2000; Bachellerie and Cavaillé, 2003). One of the C/D box snoRNAs encoded within the *PWS* locus shows sequence complementarity to the serotonin receptor 5-HT2C mRNA, which suggested its involvement in 2′-O-methylation of mRNA. This could provide a mechanism for regulation of expression of target mRNA by inhibition of adenosine to inosine editing (Burns *et al*., 1997; Yi-Brunozzi *et al*., 1999).

Two genes, *DISC1* and *DISC2*, were identified as being disrupted by a translocation that segregates with schizophrenia. *DISC1* encodes a large protein that bears some similarity to other proteins, which are known to function in the nervous system. *DISC2* is apparently a gene whose product is a noncoding RNA gene overlapping and transcribed in an opposite direction relative to *DISC1* (Millar *et al*., 2000). Several, abundant *DISC2* transcripts ranging in size from 2.5 to over 9.5 kb were identified in heart, and there is evidence that they are also produced in some fetal tissues. It was proposed that the alterations resulting from the translocations may be responsible for production of aberrant, truncated DISC1 protein or deregulation of expression, which in turn causes schizophrenia. The possible mechanisms of antisense RNA-mediated regulation involve regulation of transcription, processing, or export from the nucleus as well as mRNA stability and translation repression (Millar *et al*., 2000).

UM 9(5)h and UM 9(5)p were identified as novel noncoding transcripts in humans and porcine. Although their expression was detectable in other tissues, the highest steady-state levels of these transcripts were found in adult cerebellum (Michel *et al*., 2002).

Brain-specific noncoding RNAs was also identified in honeybee. *Ks-1* is a gene whose expression was shown to be limited to certain regions of the bee central nervous system, especially in the small-type Kenyon cells of the mushroom bodies. The RNA is 17 kb long and it does not appear to encode any protein. The *Ks-1* transcripts show nuclear localization, where they appear as scattered spots, which would suggest association with some nuclear structures. No *Ks-1* homologs were identified in the *Drosophila* genome and the gene seems to be restricted to a small phylogenetic group, but the existence of related RNA species in other insects cannot be excluded (Sawata *et al*., 2002).

2. Cancer-Associated Transcripts

Several noncoding transcripts have been demonstrated to be overexpressed in certain tumor cell lines. Differential screening between the colon carcinoma cell line (TC7) and normal mucosa revealed a new gene that in three of eight cases shows highly elevated levels of expression in tumor cells. The *OCC-1* (*o*verexpressed in *c*olon *c*arcinoma-*1*) gene is located on human chromosome 12q24.1 and was shown to encode two 1.2- and 1.3-kb-long ncRNAs that differ in 5′- and 3′-terminal parts. The gene is also expressed in normal kidney, skeletal muscles, and pancreas. An absence or very low expression of these RNAs in normal mucosa suggests that there might exist a causal link between OCC-1 RNA and carcinogenesis (Pibouin *et al*., 2002).

One of the subtypes of rhabdomyosarcoma (RMS) is associated with increased expression of the noncoding RNA called *NCRMS* (noncoding

RNA in RMS). The *NCRMS* gene, consisting of at least 11 alternatively spliced exons, was mapped to human chromosome 12q21. In its vicinity there are other genes associated with muscle development, myogenic regulators *Myf5* and *Myf6* and a growth factor *Igf2*. This localization, as well as similar patterns of *NCRMS* expression in neuroblastoma and synovial sarcoma, suggests that these tumors may have a common etiology that involves deregulation of gene expression in a larger chromosomal region (Chan *et al.*, 2002).

In prostate cancer, differential display analyses revealed two noncoding RNA genes specifically overexpressed in malignant prostate cells. The *DD3* gene, mapped to chromosome 9q21–22, shows significantly increased expression in prostate tumors when compared to nonmalignant cells. This pattern of expression was observed in over 90% of analyzed samples. Its expression was shown to be prostate specific and there was no indication that it is expressed in other tissues (Bussemakers *et al.*, 1999). *PCGEM1* is another prostate-specific gene that shows overexpression in androgen receptor-positive cells. Tumor-associated, androgen-dependent expression of *PCGEM1* implies that this gene can play an important role in the origins of prostate cancer (Srikantan *et al.*, 2000).

Genes localized on the long arm of human chromosome 7 have been implicated in autistic disorder and several cancers. A gene disrupted by chromosomal translocation in an autistic patient *(RAY7)* (Vincent *et al.*, 2000) was also described as a tumor suppressor *(ST7)* (Zenklusen *et al.*, 2001). Several mutations in *RAY1/ST7* were identified as associated with some cases of breast cancer and colon carcinomas (Zenklusen *et al.*, 2001). *RAY1/ST7* constitutes a complex transcription unit, producing many potential transcripts. The locus has the potential to produce four RNAs without significant protein-coding capacity. Two transcripts are produced on the sense (ST7OT3, ST7OT4) and two on the antisense (ST7OT1, ST7OT2) strand. The antisense transcripts were proposed to play a role in a regulation of the sense gene via RNA–RNA interactions (Vincent *et al.*, 2002).

B cell chronic lymphocytic leukemias (B-CLL) are frequently associated with deletions within the chromosomal region 13q14.3. This common form of leukemia in adults is characterized by a progressive accumulation of $CD5^{+}$B-lymphocytes accompanied by immunodeficiency and autoimmunity. The deletions on 13q14 are observed in over 50% of the B-CLL cases and in over 60% of cases of mantle cell lymphoma. The nucleotide sequence of a large fragment of 13q14 revealed the presence of a very large gene called *BCMS* (*B-c*ell neoplasia-associated gene with *m*ultiple *s*plicing). Its chromosomal localization makes it a good candidate for a tumor suppressor gene, associated with B-CLL pathogenesis. The gene spans over 560 kb and is split into at least 50 exons. The primary transcript undergoes alternative splicing producing a large number of variants with tissue-specific distribution. None

of the splicing variants contains significant ORF, which strongly suggests that they function as noncoding RNAs (Wolf *et al*., 2001). The functions of these transcripts and their role in tumor suppression remain open questions. Interestingly, the B-CLL associated with the small deletion in this region were also shown to involve two miRNA genes (Calin *et al*., 2002).

3. Stress Response Noncoding RNAs

Some of the noncoding transcripts were found to be associated with response to various stress conditions. In hamster fibroblasts, treatment with a minimally toxic dose of hydrogen peroxide induces expression of noncoding RNA species *adapt 15*, *adapt 33*, and *gadd7* (Crawford *et al*., 1996; Wang *et al*., 1996). The precise role of these transcripts in protection of the cells against effects of oxidative stress is not known (Crawford and Davies, 2003). In *Tetrahymena thermophila* a small cytpolasmic RNA designated G8 RNA was found to be induced in response to heat shock. Mutants unable to express functional G8 RNA showed normal heat shock response, but were deficient in establishing a thermotolerance (Fung *et al*., 1995).

One of the best studied stress induced RNAs is an *hsr*ω (*h*eat *s*hock *R*NA omega) in *Drosophila*. That gene shows expression in virtually all cell types in response to heat shock. Unlike most of the heat shock-induced genes it does not encode protein and its functional products are untranslated RNAs (Lakhotia, 2003). The gene was found in all *Drosophila* species and although there is little sequence similarity, in all cases an intron–exon organization is preserved. Alternative polyadenylation sites result in production of two different hsrω RNAs. A 10- to 15-kb-long (depending on species) hsrω-n is a nuclear transcript, which in addition to the two exons and an intron contains a 3′-domain consisting of 5–10 kb of tandem repeats. The shorter 2-kb transcript, terminated at the first polyadenylation site, is a precursor of a 1.2-kb cytoplasmic form hsrω-c (Garbe *et al*., 1986; Lakhotia, 2003). Inhibition of transcription or translation results in an increase of stability of otherwise labile nuclear and cytoplasmic transcripts, respectively (Bendena *et al*., 1989).

The *hsr*ω gene was demonstrated to be crucial for proper development and viability of the flies (Lakhotia, 2003). The short, cytoplasmic RNA was found to associate with ribosomes to translate a short, poorly conserved ORF present approximately 120 nt from the 5′-end in all *Drosophila* species (Fini *et al*., 1989). It has been suggested that the translation serves as a test for protein biosynthesis machinery and that it somehow is linked to hsrω-c degradation. If the test fails the accumulation of hsrω-c follows (Lakhotia, 1989). The long hsrω-n form was shown to be present in nucleoplasm in a form of speckles (omega speckles) containing various hnRNP proteins (Prasanth *et al*., 2000). The speckles seem to be a storage form of hnRNPs.

The correlation between the levels of hsrω-n and the number of omega speckles suggested that the RNA plays the role of an organizer molecule. It also seems to play the role of a chaperone for hnRNPs (Lakhotia, 2003).

4. Other Noncoding RNAs

A number of noncoding RNAs have been shown to be tissue-specific transcripts. Their expression is either limited to specialized cell types or is induced by factors associated with development and differentiation. In most cases, the details of the role that RNA plays in the cell are unknown, yet their unique patterns of expression suggest some regulatory functions.

Protease-activated receptors (PARs) mediate a majority of thrombin effects essential for vascular integrity. Activation of PAR-1 by thrombin stimulates phosphatidylinositol-specific phospholipase C and when coupled to G_i activates the RAS/MAP kinase pathway. Expression of the PAR-1 gene was shown to be stimulated, through an unknown mechanism, by certain hormones and growth factors and it was found to be associated with atherosclerosis and certain breast carcinomas (Madamanchi *et al.*, 2002). A positive regulatory element located 4.1 kb upstream from the PAR-1 gene was shown to be a promoter region of a gene for a novel ~400-nt-long noncoding RNA, ncR-uPAR (noncoding RNA upstream of the PAR-1 gene). The RNA was shown to be transcribed by Pol II and polyadenylated. NcR-uPAR RNA shows high similarity with an Alu-like sequence. This RNA can specifically bind nuclear proteins, and it has been hypothesized that it may be somehow involved in the regulation of expression of PAR-1 during development (Madamanchi *et al.*, 2002).

In mouse, a noncoding RNA gene (G90) has been identified. The polyadenylated 1.5-kb transcript has no ORF larger than 248 bp. G90 RNA was shown to be abundant in small intestine. Lower levels of transcription were observed in large intestine, testis, and kidney (Krause *et al.*, 1999).

NTT (*n*oncoding *t*ranscript in *T* cells) is a gene encoding a 17-kb noncoding, polyadenylated nuclear transcript expressed exclusively in a subset of activated human $CD4^+$ T cells. The RNA is not spliced, and it apparently does not encode any protein. The expression of *NTT* is not imprinted and both alleles are active. The gene is localized on human chromosome 6q23–q24 near the interferon-γ receptor (IFN-γR) gene. It has been proposed that *NTT* expression may be somehow linked to the regulation of other T cell-specific genes (Liu *et al.*, 1997).

Another developmentally regulated RNA was identified in mouse, as a specific transcript produced in response to bone morphogenetic proteins/osteogenic proteins (BMP/OP), whose activity is crucial for bone and cartilage differentiation. BORG RNA (*B*MP/*O*P-*r*esponsive gene) is transcribed in BMP-responsive cells and its expression is stimulated by BMP-2 or OP-1

treatment. The RNA is 2.8 kb long, spliced, and polyadenylated. Highly specific response to the growth factors suggested that this RNA plays a role in the differentiation process, but the mechanism of its action is not known (Takeda *et al.*, 1998).

In rats, estrogen and progesterone treatment results in alterations in gene expression in mammary glands. One of the genes, whose expression was shown to increase upon hormone treatment, is G.B7. The primary transcript of that undergoes alternative splicing to produce nocoding RNAs with an unknown function. A matching human sequence was mapped to chromosome 2q33 in a region including a breakpoint associated with several human tumors (Ginger *et al.*, 2001).

VII. Searching for ncRNA Genes in Genomic Sequences

The completion of more and more genomes from a variety of prokaryotic and eukaryotic organisms creates the problem of identifying their regions that constitute genes, i.e., undergo transcription to produce functional RNAs. Most of the computational methods for gene finding, as well as experimental methods for construction of expression libraries, are biased against finding noncoding RNAs (Schattner, 2003). This stems partly from the traditional underestimation of the role of RNA in the cell. The quest for genes was for many years equivalent to the search for protein-coding sequences.

A number of features of protein-coding genes that are used in gene-finding methods cannot be generally applied for noncoding RNAs. As a rule, from noncoding RNA searches one has to exclude open reading frames as well as statistical parameters related to them such as codon usage. Promoter sequences and termination signals should occur in every gene, regardless of protein-coding properties, yet noncoding RNAs in eukaryotes can be transcribed by all three RNA polymerases. Polyadenylation signals and splicing donor and acceptor sites narrow the search to mRNA-like ncRNAs. The problem of definition can be even more complicated by the assumption that the noncoding RNAs can be encoded within introns of other genes. Simple similarity search and comparison with other known genes also has limited value because functional RNAs would tend to preserve secondary structure rather than nucleotide sequence. This feature can be used for identification of homologous genes.

The most efficient approach in the identification of noncoding RNA genes is specialization. Gene-finding programs designed for detection of one specific class of RNA take into account conserved sequence elements, secondary structure features, or both. This methodology resulted in several programs

designed for tRNA searches such as tRNAscan (Fichant and Burks, 1991), Pol3Scan (Pavesi *et al.*, 1994), and COVE (Eddy and Durbin, 1994). The combination of the speed of tRNAscan and Pol3Scan with the sensitivity of COVE resulted in tRNAscan-SE, which is now routinely used for identification of tRNA genes (Lowe and Eddy, 1997). Another program, snoscan, was designed to search for snoRNAs involved in 2′-O-methylation. It was used to identify novel snoRNAs in yeast (Lowe and Eddy, 1999) as well as in *Archaea* (Omer *et al.*, 2000). A similar approach based on the observation that microRNAs are processed from precursors that form stem–loop secondary structures was employed in search for miRNA genes in vertebrates. The candidate miRNAs were selected from intergenic regions that satisfy structural constraints and show some degree of conservation among vertebrates (Lim *et al.*, 2003)

Other programs can be used to search for RNAs based on the presence of specific primary or secondary structure motifs, provided as descriptors or multiple sequence alignments (Gautheret *et al.*, 1990; Winker *et al.*, 1990; Macke *et al.*, 2001. These methods, however, like the very specialized programs aimed at one class of RNAs, require some prior knowledge about the nature of the RNAs that being are looked for.

A separate problem is detection of new RNA-coding genes, where we do not have any structural or sequence homologues. Some methods were based on that assumption that in contrast to the random sequences with the same base composition, the sequences of functional RNAs should possess more stable secondary structures (Le *et al.*, 1988). Screening of several genomic sequences using a modified approach with the application of stochastic context-free grammar led to the conclusion that predicted secondary structures of random and real sequences are indistinguishable (Rivas and Eddy, 2000). A number of methods uses base composition, or G + C contents. This approach, together with comparative genomics and experimental verification, was used to identify new noncoding RNAs in *M. jannaschii* and *P. furiosus* (Klein *et al.*, 2002). A combination of base composition statistics, nucleotide sequence signatures, and secondary structure predictions was used to define significant signals that would differentiate RNA- and protein-coding genes from the genomic background. Sequence features and structural elements extracted from known RNAs were then used for screening of the eubacterial and archaeal genomes. The results showed that the RNA-coding sequences do, in fact, contain information that can be used for accurate gene finding. Searching the *E. coli* genome yielded 370 potential noncoding RNAs, some of which were confirmed by experiments (Carter *et al.*, 2001). The search for RNA polymerase III promoters and analysis of expression from the sequence gaps between the predicted ORFs in yeast led to identification of novel noncoding RNA transcripts as well as RNAs containing small open reading frames (Olivas *et al.*, 1997). Several new

ncRNAs have been identified using a similar method based on the presence of promoter and teminator sequences in *E. coli* (Argaman *et al*., 2001)

The availability of genomic sequences offers another way of identifying ncRNAs using comparative analysis. Comparative genomics may prove helpful in cases in which analysis of individual genomes fails. This approach is based on the assumption that in genes encoding homologous RNAs compensatory mutations that preserve secondary structure should be observed (Rivas and Eddy, 2001). This method, applied to genomes of *E. coli* and five other enterobacteria, resulted in detection of 275 candidates for ncRNA genes (Rivas *et al*., 2001). Subsequent experiments confirmed that some of these sequences are in fact functional ncRNA-coding genes (Argaman *et al*., 2001; Wassarman *et al*., 2001).

VIII. Concluding Remarks

In recent years it has been demonstrated that RNA molecules, in addition to the initially identified functions in protein biosynthesis, play a key role in the regulation of many cellular processes. It appears that regulatory mechanisms involving noncoding RNAs can be found in virtually every step of transmission of genetic information (Fig. 11). They can influence the transcriptional status of chromatin, participating in its remodeling and modifications. The activity of protein transcription activators can be modulated by RNA cofactors. Noncoding RNAs may be responsible for regulation of splicing and editing of primary transcripts as well as their stability. A number of ncRNAs were demonstrated to directly influence translation both in prokaryotes and in eukaryotes.

At the beginning of the 1980s the discovery of catalytic RNAs came as a surprise. Before, it was dogmatically assumed that only proteins can act as biocatalysts. It is evident that proteins are much better suited for this task, and protein enzymes are more efficient than their RNA-based counterparts. In a transition from the RNA world to the DNA/protein world, most of the catalytic functions were taken over by proteins, which in addition to being better catalysts are more resistant to hydrolysis than RNA. Thus, present day catalytic RNAs appear to be living molecular fossils from the prebiotic RNA world.

The regulatory RNAs on the other hand seem to be a relatively recent evolutionary addition, most certainly not related to the primordial all-RNA self-replicating systems. Most probably, they evolved in response to the need for tight control of expression of certain genes at the posttranscriptional level. This mode of action of noncoding RNAs prevails in bacteria. New regulatory RNAs evolved with the emergence of eukaryotes and further

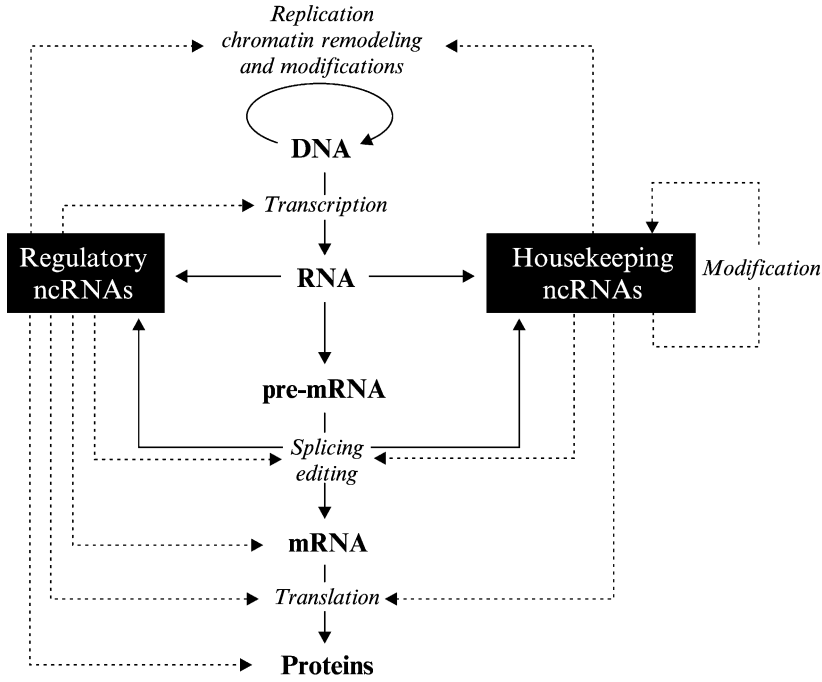

FIG. 11 Activities of noncoding RNAs in the flow of genetic information. Noncoding RNAs control every aspect of gene expression from chromatin structure to mRNA stability.

complication of molecular mechanisms related to development, differentiation, cell specialization, and generally higher organization of organisms.

From genome comparisons it is obvious that the organization of organisms does not depend solely on the information encoded within protein-coding genes. The involvement of many noncoding RNAs in developmental processes suggests that they may constitute key regulatory elements responsible for differences between various life forms (Mattick, 2001).

One question that is always asked when speaking about noncoding regulatory RNAs is how many of them are there in the genomes? The estimated numbers of functional ncRNAs are within the range of 50–200 in bacteria and several hundred to thousands in *C. elegans* (Storz, 2002). A systematic study in *Giardia lamblia* revealed that approximately one-fifth of polyadenylated transcripts represent noncoding (“sterile”) antisense RNAs with unknown (if any) functions (Elmendorf *et al.*, 2001). Even more dramatic are the results of analysis of the full-length mouse cDNAs, where one-third (11,665) of the analyzed “transcriptional units” represented novel noncoding RNAs (Okazaki *et al.*, 2002). This clearly demonstrates that ncRNAs

constitute a substantial fraction of the whole transcriptional output from the genome, but the biological significance of the majority of these RNAs remains to be elucidated.

RNA-based regulation does not necessarily require a separate regulatory RNA molecule. It has been known for many years that the untranslated portions of mRNAs may be involved in regulation of their expression. In bacteria, it has also been demonstrated that the 5′-UTRs can play the role of molecular thermometers (Hoe and Goguen, 1993; Johansson *et al.*, 2002) or small molecule sensors (Mironov *et al.*, 2002; Winkler *et al.*, 2002). The RNA structure changes, depending on the temperature or the presence of small ligands, are responsible for activation or repression of translation or termination of transcription, without participation of any other factors.

From a number of papers published on noncoding RNAs and their varied functions it is evident that it has become one of the most exciting fields in molecular biology. This fascination is further reinforced by the realization that genomics and proteomics alone are not enough to describe living systems.

Acknowledgments

This work was supported by financing from the Polish State Committee for Scientific Research.

References

Adams, M. D., Celniker, S. E., Holt, R. A., *et al.* (2000). The genome sequence of *Drosophila melanogaster*. *Science* **287,** 2185–2195.

Akhtar, A., Zink, D., and Becker, P. B. (2000). Chromodomains are protein-RNA interaction modules. *Nature* **407,** 405–409.

Altuvia, S., Weinstein-Fischer, D., Zhang, A., *et al.* (1997). A small, stable RNA induced by oxidative stress: Role as a pleiotropic regulator and antimutator. *Cell* **90,** 43–53.

Altuvia, S., Zhang, A., Argaman, L., *et al.* (1998). The *Escherichia coli* OxyS regulatory RNA represses *fhlA* translation by blocking ribosome binding. *EMBO J.* **17,** 6069–6075.

Amrein, H., and Axel, R. (1997). Genes expressed in neurons of adult male *Drosophila*. *Cell* **88,** 459–469.

Andersen, J., Delihas, N., Ikenaka, K., *et al.* (1987). The isolation and characterization of RNA coded by the micF gene in Escherichia coli. *Nucleic Acids Res.* **15,** 2089–2101.

Argaman, L., and Altuvia, S. (2000). *fhlA* repression by OxyS RNA: Kissing complex formation at two sites results in a stable antisense-target RNA complex. *J. Mol. Biol.* **300,** 1101–1112.

Argaman, L., Hershberg, R., Vogel, J., *et al.* (2001). Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr. Biol.* **11,** 941–950.

Ashe, H. L., Monks, J., Wijgerde, M., *et al.* (1997). Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev.* **11,** 2494–2509.

Bachellerie, J.-P., and Cavaillé, J. (2003). New frontiers for the snoRNA world. *In* "Noncoding RNAs: Molecular Biology and Molecular Medicine" (J. Barciszewski and V. A. Erdmann, Eds.), pp. 172–192. Kluwer Academic/Plenum Publishers, Dordrecht, The Netherlands.

Bae, E., Calhoun, V. C., Levine, M., *et al.* (2002). Characterization of the intergenic RNA profile at abdominal-A and Abdominal-B in the *Drosophila* bithorax complex. *Proc. Natl. Acad. Sci. USA* **99,** 16847–16852.

Bartolomei, M. S., and Tilghman, S. M. (1997). Genomic imprinting in mammals. *Annu. Rev. Genet.* **31,** 493–525.

Beletskii, A., Hong, Y. K., Pehrson, J., *et al.* (2001). PNA interference mapping demonstrates functional domains in the noncoding RNA *Xist*. *Proc. Natl. Acad. Sci. USA* **98,** 9215–9220.

Bendena, W. J., Garbe, J. C., Traverse, K. L., *et al.* (1989). Multiple inducers of the *Drosophila* heat shock locus 93D (hsr omega): Inducer-specific patterns of the three transcripts. *J. Cell Biol.* **108,** 2017–2028.

Bender, W., and Fitzgerald, D. P. (2002). Transcription activates repressed domains in the *Drosophila* bithorax complex. *Development* **129,** 4923–4930.

Bennani-Baiti, I. M., Asa, S. L., Song, D., *et al.* (1998). DNase I-hypersensitive sites I and II of the human growth hormone locus control region are a major developmental activator of somatotrope gene expression. *Proc. Natl. Acad. Sci. USA* **95,** 10655–10660.

Bernstein, E., Caudy, A. A., Hammond, S. M., *et al.* (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409,** 363–366.

Blin-Wakkach, C., Lezot, F., Ghoul-Mazgar, S., *et al.* (2001). Endogenous *Msx1* antisense transcript: *In vivo* and *in vitro* evidences, structure, and potential involvement in skeleton development in mammals. *Proc. Natl. Acad. Sci. USA* **98,** 7336–7341.

Borsani, G., Tonlorenzi, R., Simmler, M. C., *et al.* (1991). Characterization of a murine gene expressed from the inactive X chromosome. *Nature* **351,** 325–329.

Brantl, S. (2002). Antisense-RNA regulation and RNA interference. *Biochim. Biophys. Acta* **1575,** 15–25.

Brescia, C. C., and Sledjeski, D. D. (2003). We are legion noncoding regulatory RNAs and Hfq. *In* "Noncoding RNAs: Molecular Biology and Molecular Medicine" (J. Barciszewski and V. A. Erdmann, Eds.), pp. 260–268. Kluwer Academic/Plenum Publishers, Dordrecht, The Netherlands.

Brockdorff, N. (2002). X-chromosome inactivation: Closing in on proteins that bind Xist RNA. *Trends Genet.* **18,** 352–358.

Brockdorff, N., Ashworth, A., Kay, G. F., *et al.* (1991). Conservation of position and exclusive expression of mouse *Xist* from the inactive X chromosome. *Nature* **351,** 329–331.

Brockdorff, N., Ashworth, A., Kay, G. F., *et al.* (1992). 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,** 515–526.

Brown, C. J., and Baldry, S. E. (1996). Evidence that heteronuclear proteins interact with *XIST* RNA in vitro. *Somat. Cell Mol. Genet.* **22,** 403–417.

Brown, C. J., Ballabio, A., Rupert, J. L., *et al.* (1991a). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349,** 38–44.

Brown, C. J., Lafreniere, R. G., Powers, V. E., *et al.* (1991b). Localization of the X inactivation centre on the human X chromosome in Xq13. *Nature* **349,** 82–84.

Burns, C. M., Chu, H., Rueter, S. M., *et al.* (1997). Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature* **387,** 303–308.

Bussemakers, M. J., van Bokhoven, A., Verhaegh, G. W., *et al.* (1999). *DD3:* A new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Res.* **59,** 5975–5979.

Calin, G. A., Dumitru, C. D., Shimizu, M., *et al.* (2002). Frequent deletions and down-regulation of microRNA genes *mir 15* and *mir 16* at 13q14 in chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. USA* **99,** 15524–15529.

Carmell, M. A., Xuan, Z., Zhang, M. Q., *et al.* (2002). The Argonaute family: Tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev.* **16,** 2733–2742.

Carter, R. J., Dubchak, I., and Holbrook, S. R. (2001). A computational approach to identify genes for functional RNAs in genomic sequences. *Nucleic Acids Res.* **29,** 3928–3938.

Caspary, T., Cleary, M. A., Baker, C. C., *et al.* (1998). Multiple mechanisms regulate imprinting of the mouse distal chromosome 7 gene cluster. *Mol. Cell. Biol.* **18,** 3466–3474.

Cavaillé, J., Buiting, K., Kiefmann, M., *et al.* (2000). Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Natl. Acad. Sci. USA.* **97,** 14311–14316.

Chamberlain, S. J., and Brannan, C. I. (2001). The Prader-Willi syndrome imprinting center activates the paternally expressed murine *Ube3a* antisense transcript but represses paternal *Ube3a*. *Genomics* **73,** 316–322.

Chan, A. S., Thorner, P. S., Squire, J. A., *et al.* (2002). Identification of a novel gene *NCRMS* on chromosome 12q21 with differential expression between rhabdomyosarcoma types. *Oncogene* **21,** 3029–3037.

Chao, W., Huynh, K. D., Spencer, R. J., *et al.* (2002). CTCF, a candidate trans-acting factor for X-inactivation choice. *Science* **295,** 345–347.

Cheng, J. G., Tiedge, H., and Brosius, J. (1996). Identification and characterization of BC1 RNP particles. *DNA Cell Biol.* **15,** 549–559.

Chouabe, C., Neyroud, N., Guicheney, P., *et al.* (1997). Properties of *KvLQT1* K+ channel mutations in Romano-Ward and Jervell and Lange-Nielsen inherited cardiac arrhythmias. *EMBO J.* **16,** 5472–5479.

Chureau, C., Prissette, M., Bourdet, A., *et al.* (2002). Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine. *Genome Res.* **12,** 894–908.

Crawford, D. R., and Davies, K. J. A. (2003). Adapt gene RNA transcripts as riboregulators. *In* ''Noncoding RNAs: Molecular Biology and Molecular Medicine'' (J. Barciszewski and V. A. Erdmann, Eds.), pp. 221–230. Kluwer Academic/Plenum Publishers, Dordrecht, The Netherlands.

Crawford, D. R., Schools, G. P., and Salmon, S. L. (1996). Hydrogen peroxide induces the expression of adapt15, a novel RNA associated with polysomes in hamster HA-1 cells. *Arch. Biochem. Biophys.* **325,** 256–264.

Cumberledge, S., Zaratzian, A., and Sakonju, S. (1990). Characterization of two RNAs transcribed from the cis-regulatory region of the *abd-A* domain within the *Drosophila* bithorax complex. *Proc. Natl. Acad. Sci. USA* **87,** 3259–3263.

DeBaun, M. R., Niemitz, E. L., McNeil, D. E., *et al.* (2002). Epigenetic alterations of *H19* and *LIT1* distinguish patients with Beckwith-Wiedemann syndrome with cancer and birth defects. *Am. J. Hum. Genet.* **70,** 604–611.

Delihas, N., and Forst, S. (2001). MicF: An antisense RNA gene involved in response of *Escherichia coli* to global stress factors. *J. Mol. Biol.* **313,** 1–12.

Dong, C., and Flavell, R. A. (2000). Cell fate decision: T-helper 1 and 2 subsets in immune responses. *Arthritis Res.* **2,** 179–188.

Dostie, J., Mourelatos, Z., Yang, M., *et al.* (2003). Numerous microRNPs in neuronal cells containing novel microRNAs. *RNA* **9,** 180–186.

Doudna, J. A., and Cech, T. R. (2002). The chemical repertoire of natural ribozymes. *Nature* **418,** 222–228.

Drewell, R. A., Bae, E., Burr, J., *et al.* (2002). Transcription defines the embryonic domains of cis-regulatory activity at the *Drosophila* bithorax complex. *Proc. Natl. Acad. Sci. USA* **99,** 16853–16858.

Duthie, S. M., Nesterova, T. B., Formstone, E. J., *et al.* (1999). *Xist* RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in cis. *Hum. Mol. Genet.* **8,** 195–204.

Eddy, S. R., and Durbin, R. (1994). RNA sequence analysis using covariance models. *Nucleic Acids Res.* **22,** 2079–2088.

Elmendorf, H. G., Singer, S. M., and Nash, T. E. (2001). The abundance of sterile transcripts in *Giardia lamblia*. *Nucleic Acids Res.* **29,** 4674–4683.

Escobar, M. A., and Dandekar, A. M. (2003). Post-transcriptional gene silencing in plants. *In* "Noncoding RNAs: Molecular Biology and Molecular Medicine" (J. Barciszewski and V. A. Erdmann, Eds.), pp. 129–141. Kluwer Academic/Plenum Publishers, Dordrecht, The Netherlands.

Feinbaum, R., and Ambros, V. (1999). The timing of lin-4 RNA accumulation controls the timing of postembryonic developmental events in *Caenorhabditis elegans*. *Dev. Biol.* **210,** 87–95.

Fichant, G. A., and Burks, C. (1991). Identifying potential tRNA genes in genomic DNA sequences. *J. Mol. Biol.* **220,** 659–671.

Fini, M. E., Bendena, W. J., and Pardue, M. L. (1989). Unusual behavior of the cytoplasmic transcript of hsrω. An abundant stress-inducible RNA that is translated, but which yields no detectable protein product. *J. Cell Biol.* **108,** 2045–2057.

Franke, A., and Baker, B. S. (1999). The roxl and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in *Drosophila*. *Mol. Cell* **4,** 117–122.

Frazer, K. A., Ueda, Y., Zhu, Y., *et al.* (1997). Computational and biological analysis of 680 kb of DNA sequence from the human 5q31 cytokine gene cluster region. *Genome Res.* **7,** 495–512.

Fung, P. A., Gaertig, J., Gorovsky, M. A., *et al.* (1995). Requirement of a small cytoplasmic RNA for the establishment of thermotolerance. *Science* **268,** 1036–1038.

Garbe, J. C., Bendena, W. G., and Pardue, M. L. (1986). A *Drosophila* heat shock gene with a rapidly diverging sequence but a conserved structure. *J. Biol. Chem.* **261,** 16889–16894.

Gautheret, D., Major, F., and Cedergren, R. (1990). Pattern searching/alignment with RNA primary and secondary structures: An effective descriptor for tRNA. *Comput. Appl. Biosci.* **64,** 325–331.

Gilbert, S. L., and Sharp, P. A. (1999). Promoter-specific hypoacetylation of X-inactivated genes. *Proc. Natl. Acad. Sci. USA* **96,** 13825–13830.

Ginger, M. R., Gonzalez-Rimbau, M. F., Gay, J. P., *et al.* (2001). Persistent changes in gene expression induced by estrogen and progesterone in the rat mammary gland. *Mol. Endocrinol.* **15,** 1993–2009.

Glickman, J. N., Howe, J. G., and Steitz, J. A. (1988). Structural analyses of EBER1 and EBER2 ribonucleoprotein particles present in Epstein-Barr virus-infected cells. *J. Virol.* **62,** 902–911.

Goto, Y., and Takagi, N. (2000). Maternally inherited X chromosome is not inactivated in mouse blastocysts due to parental imprinting. *Chromosome Res.* **8,** 101–109.

Graves, J. A. (1996). Mammals that break the rules: Genetics of marsupials and monotremes. *Annu. Rev. Genet.* **30,** 233–260.

Guerrier-Takada, C., Gardiner, K., and Marsh, T. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35,** 849–857.

Ha, I., Wightman, B., and Ruvkun, G. (1996). A bulged *lin-4*/*lin-14* RNA duplex is sufficient for *Caenorhabditis elegans lin-14* temporal gradient formation. *Genes Dev.* **10,** 3041–3050.

Hall, I. M., Shankaranarayana, G. D., Noma, K., *et al.* (2002). Establishment and maintenance of a heterochromatin domain. *Science* **297,** 2232–2237.

Hanel, M. L., and Wevrick, R. (2001). The role of genomic imprinting in human developmental disorders: Lessons from Prader-Willi syndrome. *Clin. Genet.* **59,** 156–164.

Hannon, G. J. (2002). RNA interference. *Nature* **418,** 244–251.

Hayward, B. E., and Bonthron, D. T. (2000). An imprinted antisense transcript at the human *GNAS1* locus. *Hum. Mol. Genet.* **9,** 835–841.

Hayward, B. E., Barlier, A., Korbonits, M., *et al.* (2001). Imprinting of the Gsα gene *GNAS1* in the pathogenesis of acromegaly. *J. Clin. Invest.* **107,** R31–R36.

Heard, E., Rougeulle, C., Arnaud, D., *et al.* (2001). Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. *Cell* **107,** 727–738.

Hodgson, J. W., Argiropoulos, B., and Brock, H. W. (2001). Site-specific recognition of a 70-base-pair element containing d(GA)(n) repeats mediates bithoraxoid polycomb group response element-dependent silencing. *Mol. Cell. Biol.* **21,** 4528–4543.

Hoe, N. P., and Goguen, J. D. (1993). Temperature sensing in *Yersinia pestis:* Translation of the LcrF activator protein is thermally regulated. *J. Bacteriol.* **175,** 7901–7909.

Hogga, I., and Karch, F. (2002). Transcription through the iab-7 cis-regulatory domain of the bithorax complex interferes with maintenance of polycomb-mediated silencing. *Development* **129,** 4915–4922.

Horike, S., Mitsuya, K., Meguro, M., *et al.* (2000). Targeted disruption of the human *LIT1* locus defines a putative imprinting control element playing an essential role in Beckwith-Wiedemann syndrome. *Hum. Mol. Genet.* **9,** 2075–2083.

Hurst, L. D., and Smith, N. G. (1999). Molecular evolutionary evidence that H19 mRNA is functional. *Trends Genet.* **15,** 134–135.

Hutvagner, G., McLachlan, J., Pasquinelli, A. E., *et al.* (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293,** 834–838.

Jacobsen, S. E., Running, M. P., and Meyerowitz, E. M. (1999). Disruption of an RNA helicase/RNAse III gene in *Arabidopsis* causes unregulated cell division in floral meristems. *Development* **126,** 5231–5243.

Jin, Y., Wang, Y., Walker, D. L., *et al.* (1999). JIL-1: A novel chromosomal tandem kinase implicated in transcriptional regulation in *Drosophila*. *Mol. Cell* **4,** 129–135.

Jin, Y., Wang, Y., Johansen, J., *et al.* (2000). JIL-1, a chromosomal kinase implicated in regulation of chromatin structure, associates with the male specific lethal (MSL) dosage compensation complex. *J. Cell Biol.* **149,** 1005–1010.

Johansson, J., Mandin, P., Renzoni, A., *et al.* (2002). An RNA thermosensor controls expression of virulence genes in *Listeria monocytogenes*. *Cell* **110,** 551–561.

Johnston, C. M., Nesterova, T. B., Formstone, E. J., *et al.* (1998). Developmentally regulated *Xist* promoter switch mediates initiation of X inactivation. *Cell* **94,** 809–817.

Johnston, C. M., Newall, A., Brockdorff, N., *et al.* (2002). *Enox*, a novel gene that maps 10 kb upstream of *Xist* and partially escapes X inactivation. *Genomics* **80,** 236–244.

Jones, K. A. (1997). Taking a new TAK on tat transactivation. *Genes Dev.* **11,** 2593–2599.

Joubel, A., Curgy, J. J., Pelczar, H., *et al.* (1996). The 5′ part of the human H19 RNA contains cis-acting elements hampering its translatability. *Cell Mol. Biol.* **42,** 1159–1172.

Juan, V., Crain, C., and Wilson, C. (2000). Evidence for evolutionarily conserved secondary structure in the H19 tumor suppressor RNA. *Nucleic Acids Res.* **28,** 1221–1227.

Kelley, R. L., and Kuroda, M. I. (2000). Noncoding RNA genes in dosage compensation and imprinting. *Cell* **103,** 9–12.

Ketting, R. F., Fischer, S. E., Bernstein, E., *et al.* (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* **15,** 2654–2659.

Khochbin, S., Brocard, M. P., Grunwald, D., *et al.* (1992). Antisense RNA and *p53* regulation in induced murine cell differentiation. *Ann. N. Y. Acad. Sci.* **660,** 77–87.

Kidner, C. A., and Martienssen, R. A. (2003). Macro effects of micro RNAs in plants. *Trends Genet.* **19,** 13–16.

Kim, J., Martignetti, J. A., Shen, M. R., *et al.* (1994). Rodent BC1 RNA gene as a master gene for ID element amplification. *Proc. Natl. Acad. Sci. USA* **91,** 3607–3611.

Kindel, D., and Amrein, H. (2003). Dosage compensation in *Drosophila* a ribonucleoprotein complex mediates transcriptional up-regulation. *In* "Noncoding RNAs: Molecular Biology and Molecular Medicine" (J. Barciszewski and V. A. Erdmann, Eds.), pp. 67–82. Kluwer Academic/Plenum Publishers, Dordrecht, the Netherland.

Klein, R. J., Misulovin, Z., and Eddy, S. R. (2002). Noncoding RNA genes identified in AT-rich hyperthermophiles. *Proc. Natl. Acad. Sci. USA* **99,** 7542–7547.

Kobayashi, S., Goto, S., and Anzai, K. (1991). Brain-specific small RNA transcript of the identifier sequences is present as a 10 S ribonucleoprotein particle. *J. Biol. Chem.* **266,** 4726–4730.

Kosaki, K., Kosaki, R., Craigen, W. J., *et al.* (2000). Isoform-specific imprinting of the human *PEG1/MEST* gene. *Am. J. Hum. Genet.* **66,** 309–312.

Kramer, C., Loros, J. L., Dunlap, J. C., *et al.* (2003). Role for antisense RNA in regulating circadian clock function in *Neurospora crassa*. *Nature* **421,** 948–952.

Krause, R., Hemberger, M., Himmelbauer, H., *et al.* (1999). Identification and characterization of G90, a novel mouse RNA that lacks an extensive open reading frame. *Gene* **232,** 35–42.

Kruger, K., Grabowski, P. J., Zaug, A. J., *et al.* (1982). Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* **31,** 147–157.

Krystal, G. W., Armstrong, B. C., and Battey, J. F. (1990). *N-myc* mRNA forms an RNA-RNA duplex with endogenous antisense transcripts:. *Mol. Cell. Biol.* **10,** 4180–4191.

Kuryshev, V. Y., Skryabin, B. V., Kremerskothen, J., *et al.* (2001). Birth of a gene: Locus of neuronal BC200 snmRNA in three prosimians and human BC200 pseudogenes as archives of change in the *Anthropoidea* lineage. *J. Mol. Biol.* **309,** 1049–1066.

Kwek, K. Y., Murphy, S., and Furger, A. (2002). U1 snRNA associates with TFIIH and regulates transcription initiation. *Nat. Struct. Biol.* **9,** 800–805.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., *et al.* (2001). Identification of novel genes coding for small expressed RNAs. *Science* **294,** 853–858.

Lagos-Quintana, M., Rauhut, R., Yalcin, A., *et al.* (2002). Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* **12,** 735–739.

Lagos-Quintana, M., Rauhut, R., Meyer, J., *et al.* (2003). New microRNAs from mouse and human. *RNA* **9,** 175–179.

Lai, E. C. (2002). Micro RNAs are complementary to 3′ UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat. Genet.* **30,** 363–364.

Lakhotia, S. C. (1989). The 93D heat shock locus of *Drosophila melanogaster* modulation by genetic and developmental factors. *Genome* **31,** 677–683.

Lakhotia, S. C. (2003). The noncoding developmentally active and stress inducible hsrω gene of *Drosophila melanogaster* integrates post-transcriptional processing of other nuclear transcripts. *In* "Noncoding RNAs: Molecular Biology and Molecular Medicine" (J. Barciszewski and V. A. Erdmann, Eds.), pp. 203–220. Kluwer Academic/Plenum Publishers, Dordrecht, The Netherlands.

Lander, E. S., Linton, L. M., Birren, B., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921.

Lanz, R. B., and Rusconi, S. (1994). A conserved carboxy-terminal subdomain is important for ligand interpretation and transactivation by nuclear receptors. *Endocrinology* **135,** 2183–2195.

Lanz, R. B., McKenna, N. J., Onate, S. A., *et al.* (1999). A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell* **97,** 7–27.

Lanz, R., Razani, B., Goldberg, A. D., *et al.* (2002). Distinct RNA motifs are important for coactivation of steroid hormone receptors by steroid receptor RNA activator (SRA). *Proc. Natl. Acad. Sci. USA* **99,** 16081–16086.

Lau, N. C., Lim, L. P., Weinstein, E. G., *et al.* (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294,** 858–862.

Le, S. Y., Chen, J. H., Currey, K. M., *et al.* (1988). A program for predicting significant RNA secondary structures. *Comput. Appl. Biosci.* **4,** 153–159.

Lease, R. A., and Belfort, M. (2000a). A trans-acting RNA as a control switch in *Escherichia coli:* DsrA modulates function by forming alternative structures. *Proc. Natl. Acad. Sci. USA.* **97,** 9919–9924.

Lease, R. A., and Belfort, M. (2000b). Riboregulation by DsrA RNA: Trans-actions for global economy. *Mol. Microbiol.* **38,** 667–672.

Lease, R. A., Cusick, M. E., and Belfort, M. (1998). Riboregulation in *Escherichia coli:* DsrA acts by RNA:RNA interactions at multiple loci. *Proc. Natl. Acad. Sci. USA* **95,** 12456–12461.

Lee, J. T., and Jaenisch, R. (1997). Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. *Nature* **386,** 275–279.

Lee, J. T., Davidow, L. S., and Warshawsky, D. (1999). *Tsix*, a gene antisense to *Xist* at the X-inactivation centre. *Nat. Genet.* **21,** 400–404.

Lee, R. C., and Ambros, V. (2001). An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294,** 862–864.

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75,** 843–854.

Lee, Y. J., Park, C. W., Hahn, Y., *et al.* (2000). *Mit1/Lb9* and *Copg2*, new members of mouse imprinted genes closely linked to *PEG1/MEST*. *FEBS Lett.* **472,** 230–234.

Leighton, P. A., Ingram, R. S., Eggenschwiler, J., *et al.* (1995). Disruption of imprinting caused by deletion of the *H19* gene region in mice. *Nature* **375,** 34–39.

Lewis, E. B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* **276,** 565–570.

Li, T., Vu, T. H., Lee, K. O., *et al.* (2002). An imprinted *PEG1/MEST* antisense expressed predominantly in human testis and in mature spermatozoa. *J. Biol. Chem.* **277,** 13518–13527.

Li, Y. M., Franklin, G., Cui, H. M., *et al.* (1998). The *H19* transcript is associated with polysomes and may regulate IGF2 expression in trans. *J. Biol. Chem.* **273,** 28247–28252.

Lim, L. P., Glasner, M. E., Yekta, S., *et al.* (2003). Vertebrate microRNA genes. *Science* **299,** 1540.

Lipshitz, H. D., Peattie, D. A., and Hogness, D. S. (1987). Novel transcripts from the ultrabithorax domain of the bithorax complex. *Genes Dev.* **1,** 307–322.

Liu, A. Y., Torchia, B. S., Migeon, B. R., *et al.* (1997). The human *NTT* gene: Identification of a novel 17-kb noncoding nuclear RNA expressed in activated CD4+ T cells. *Genomics.* **39,** 171–184.

Llave, C., Kasschau, K. D., Rector, M. A., *et al.* (2002a). Endogenous and silencing-associated small RNAs in plants. *Plant Cell.* **14,** 1–15.

Llave, C., Xie, Z., Kasschau, K. D., *et al.* (2002b). Cleavage of *Scarecrow*-like mRNA targets directed by a class of *Arabidopsis* miRNA. *Science* **297,** 2053–2056.

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25,** 955–964.

Lowe, T. M., and Eddy, S. R. (1999). A computational screen for methylation guide snoRNAs in yeast. *Science* **283,** 1168–1171.

Luikenhuis, S., Wutz, A., and Jaenisch, R. (2001). Antisense transcription through the *Xist* locus mediates *Tsix* function in embryonic stem cells. *Mol. Cell Biol.* **21,** 8512–8520.

Lyle, R., Watanabe, D., te Vruchte, D., *et al.* (2000). The imprinted antisense RNA at the *Igf2r* locus overlaps but does not imprint *Mas1*. *Nat. Genet.* **25,** 19–21.

Lyon, M. F. (1961). Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* **190,** 372–373.

MacIntosh, G. C., Wilkerson, C., and Green, P. J. (2001). Identification and analysis of *Arabidopsis* expressed sequence tags characteristic of non-coding RNAs. *Plant Phys.* **127,** 765–776.

Macke, T. J., Ecker, D. J., Gutell, R. R., *et al.* (2001). RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Res.* **29,** 4724–4735.

Madamanchi, N. R., Hu, Z. Y., Li, F., *et al.* (2002). A noncoding RNA regulates human protease-activated receptor-1 gene during embryogenesis. *Biochim. Biophys. Acta* **1576,** 237–245.

Majdalani, N., Cunning, C., Sledjeski, D., *et al.* (1998). DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. *Proc. Natl. Acad. Sci. USA* **95,** 12462–12467.

Majdalani, N., Hernandez, D., and Gottesman, S. (2002). Regulation and mode of action of the second small RNA activator of RpoS translation, RprA. *Mol. Microbiol.* **46,** 813–826.

Martignetti, J. A., and Brosius, J. (1993). BC200 RNA: A neural RNA polymerase III product encoded by a monomeric Alu element. *Proc. Natl. Acad. Sci. USA* **90,** 11563–11567.

Martignetti, J. A., and Brosius, J. (1995). BC1 RNA: Transcriptional analysis of a neural cell-specific RNA polymerase III transcript. *Mol. Cell Biol.* **15,** 1642–1650.

Masse, E., and Gottesman, S. (2002). A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **99,** 4620–4625.

Mattick, J. S. (1994). Introns: Evolution and function. *Curr. Opin. Genet. Dev.* **4,** 823–831.

Mattick, J. S. (2001). Non-coding RNAs: The architects of eukaryotic complexity. *EMBO Rep.* **2,** 986–991.

Mattick, J. S. (2003). Introns and noncoding RNAs, the hidden layer of eukaryotic complexity. *In* "Noncoding RNAs: Molecular Biology and Molecular Medicine" (J. Barciszewski and V. A. Erdmann, Eds.), pp. 12–33. Kluwer Academic/Plenum Publishers, Dordrecht, The Netherlands.

Mattick, J. S., and Gagen, M. J. (2001). 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,** 1611–1630.

Matzke, M. A., Mette, M. F., Kanno, T., *et al.* (2003). RNA-directed DNA methylation and chromatin modifications. *In* "Noncoding RNAs: Molecular Biology and Molecular Medicine" (J. Barciszewski and V. A. Erdmann, Eds.), pp. 142–160. Kluwer Academic/Plenum Publishers, Dordrecht, The Netherlands.

McKenna, N. J., and O'Malley, B. W. (2002). Combinatorial control of gene expression by nuclear receptors and coregulators. *Cell* **108,** 465–474.

Meller, V. H., and Rattner, B. P. (2002). The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *EMBO J.* **21,** 1084–1091.

Meller, V. H., Gordadze, P. R., Park, Y., *et al.* (2000). Ordered assembly of roX RNAs into MSL complexes on the dosage-compensated X chromosome in *Drosophila*. *Curr. Biol.* **10,** 136–143.

Mermoud, J. E., Costanzi, C., Pehrson, J. R., *et al.* (1999). Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation. *J. Cell Biol.* **147,** 1399–1408.

Michel, U., Kallmann, B., Rieckmann, P., *et al.* (2002). UM 9(5)h and UM 9(5)p, human and porcine noncoding transcripts with preferential expression in the cerebellum. *RNA* **8,** 1538–1547.

Migeon, B. R., Chowdhury, A. K., Dunston, J. A., *et al.* (2001). Identification of *TSIX*, encoding an RNA antisense to human *XIST*, reveals differences from its murine counterpart: Implications for X inactivation. *Am. J. Hum. Genet.* **69,** 951–960.

Migeon, B. R., Lee, C. H., Chowdhury, A. K., *et al.* (2002). Species differences in *TSIX/Tsix* reveal the roles of these genes in X-chromosome inactivation. *Am. J. Hum. Genet.* **71,** 286–293.

Millar, J. K., Wilson-Annan, J. C., Anderson, S., *et al.* (2000). Disruption of two novel genes by a translocation co-segregating with schizophrenia. *Hum. Mol. Genet.* **9,** 1415–1423.

Mironov, A. S., Gusarov, I., Rafikov, R., *et al.* (2002). Sensing small molecules by nascent RNA: A mechanism to control transcription in bacteria. *Cell* **111,** 747–756.

Mitsuya, K., Meguro, M., Lee, M. P., *et al.* (1999). LIT1, an imprinted antisense RNA in the human *KvLQT1* locus identified by screening for differentially expressed transcripts using monochromosomal hybrids. *Hum. Mol. Genet.* **8,** 1209–1217.

Møller, T., Franch, T., Udesen, C., *et al.* (2002). Spot 42 RNA mediates discoordinate expression of the *E. coli* galactose operon. *Genes Dev.* **16,** 1696–1706.

Moore, P. B., and Steitz, T. A. (2002). The involvement of RNA in ribosome function. *Nature* **418,** 229–235.

Morfeldt, E., Taylor, D., von Gabain, A., *et al.* (1995). Activation of alpha-toxin translation in *Staphylococcus aureus* by the trans-encoded antisense RNA, RNAIII. *EMBO J.* **14,** 4569–4577.

Moss, E. G. (2003). MicroRNAs. *In* "Noncoding RNAs: Molecular Biology and Molecular Medicine" (J. Barciszewski and V. A. Erdmann, Eds.), pp. 98–115. Kluwer Academic/Plenum Publishers, Dordrecht, The Netherlands.

Moss, E. G., Lee, R. C., and Ambros, V. (1997). The cold shock domain protein LIN-28 controls developmental timing in *C. elegans* and is regulated by the lin-4 RNA. *Cell* **88,** 637–646.

Mourelatos, Z., Dostie, J., Paushkin, S., *et al.* (2002). miRNPs: A novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev.* **16,** 720–728.

Muddashetty, R., Khanam, T., Kondrashov, A., *et al.* (2002). Poly(A)-binding protein is associated with neuronal BC1 and BC200 ribonucleoprotein particles. *J. Mol. Biol.* **321,** 433–445.

Murphy, S. K., Wylie, A. A., and Jirtle, R. L. (2001). Imprinting of PEG3, the human homologue of a mouse gene involved in nurturing behavior. *Genomics* **71,** 110–117.

Nakabayashi, K., Bentley, L., Hitchins, M. P., *et al.* (2002). Identification and characterization of an imprinted antisense RNA (MESTIT1) in the human MEST locus on chromosome 7q32. *Hum. Mol. Genet.* **11,** 1743–1756.

Nanbo, A., and Takada, K. (2002). The role of Epstein-Barr virus-encoded small RNAs (EBERs) in oncogenesis. *Rev. Med. Virol.* **12,** 321–326.

Nesterova, T. B., Slobodyanyuk, S. Y., and Elisaphenko, E. A. (2001). Characterization of the genomic *Xist* locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. *Genome Res.* **11,** 833–849.

Nesterova, T. B., Mermoud, J. E., Hilton, K., *et al.* (2002). Xist expression and macroH2A1.2 localisation in mouse primordial and pluripotent embryonic germ cells. *Differentiation* **69,** 216–225.

Newall, A. E., Duthie, S., Formstone, E., *et al.* (2001). Primary non-random X inactivation associated with disruption of *Xist* promoter regulation. *Hum. Mol. Genet.* **10,** 581–589.

Nguyen, V. T., Kiss, T., Michels, A. A., *et al.* (2001). 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature* **414,** 322–325.

Nicholls, R. D. (2000). The impact of genomic imprinting for neurobehavioral and developmental disorders. *J. Clin. Invest.* **105,** 413–418.

Noonan, F. C., Goodfellow, P. J., Staloch, L. J., *et al.* (2003). Antisense transcripts at the *EMX2* locus in human and mouse. *Genomics* **81,** 58–66.

Ohlsson, R., Nystrom, A., Pfeifer-Ohlsson, S., *et al.* (1993). *IGF2* is parentally imprinted during human embryogenesis and in the Beckwith-Wiedemann syndrome. *Nat. Genet.* **4,** 94–97.

Okazaki, Y., Furuno, M., Kasukawa, T., *et al.* (2002). Analysis of the mouse transcriptome based on functional annotation of 60,770 full length cDNAs. *Nature* **420,** 563–573.

Okutsu, T., Kuroiwa, Y., Kagitani, F., *et al.* (2000). Expression and imprinting status of human *PEG8/IGF2AS*, a paternally expressed antisense transcript from the *IGF2* locus, in Wilms' tumors. *J. Biochem.* **127,** 475–483.

Olivas, W. M., Muhlard, D., and Parker, R. (1997). Analysis of the yeast genome: Identification of new non-coding and small ORF-containing RNAs. *Nucleic Acids Res.* **25,** 4619–4625.

Olsen, P. H., and Ambros, V. (1999). The lin-4 regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev. Biol.* **216,** 671–680.

Omer, A. D., Lowe, T. M., Russell, A. G., *et al.* (2000). Homologs of small nucleolar RNAs in Archaea. *Science* **288,** 517–522.

Park, W., Li, J., Song, R., *et al.* (2002). CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr. Biol.* **12,** 1484–1495.

Park, Y., Kelley, R. L., Oh, H., *et al.* (2002). Extent of chromatin spreading determined by *roX* RNA recruitment of MSL proteins. *Science* **298,** 1620–1623.

Pasquinelli, A. E., Reinhart, B. J., Slack, F., *et al.* (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408,** 86–89.

Paulsen, M., Takada, S., Youngson, N. A., *et al.* (2001). Comparative sequence analysis of the imprinted *Dlk1-Gtl2* locus in three mammalian species reveals highly conserved genomic elements and refines comparison with the *Igf2-H19* region. *Genome Res.* **11,** 2085–2094.

Pavesi, A., Conterio, F., Bolchi, A., *et al.* (1994). Identification of new eukaryotic tRNA genes in genomic DNA databases by a multistep weight matrix analysis of transcriptional control regions. *Nucleic Acids Res.* **22,** 1247–1256.

Peters, J., Wroe, S. F., Wells, C. A., *et al.* (1999). A cluster of oppositely imprinted transcripts at the *Gnas* locus in the distal imprinting region of mouse chromosome 2. *Proc. Natl. Acad. Sci. USA* **96,** 3830–3835.

Pibouin, L., Villaudy, J., Ferbus, D., *et al.* (2002). Cloning of the mRNA of overexpression in colon carcinoma-1: A sequence overexpressed in a subset of colon carcinomas. *Cancer Genet. Cytogenet.* **133,** 55–60.

Plant, K. E., Routledge, S. J., and Proudfoot, N. J. (2001). Intergenic transcription in the human beta-globin gene cluster. *Mol. Cell. Biol.* **21,** 6507–6514.

Prasanth, K. V., Rajendra, T. K., Lal, A. K., *et al.* (2000). Omega speckles—a novel class of nuclear speckles containing hnRNPs associated with non-coding hsr-omega RNA in *Drosophila*. *J. Cell Sci.* **113,** 3485–3497.

Ramakrishnan, V. (2002). Ribosome structure and the mechanism of translation. *Cell* **108,** 557–572.

Rasmussen, T. P., Mastrangelo, M. A., Eden, A., *et al.* (2000). Dynamic relocalization of histone MacroH2A1 from centrosomes to inactive X chromosomes during X inactivation. *J. Cell Biol.* **150,** 1189–1198.

Raymond, C. S., Kettlewell, J. R., Hirsch, B., *et al.* (1999). Expression of *Dmrt1* in the genital ridge of mouse and chicken embryos suggests a role in vertebrate sexual development. *Dev. Biol.* **215,** 208–220.

Reinhart, B. J., and Bartel, D. P. (2002). Small RNAs correspond to centromere heterochromatic repeats. *Science* **297,** 1831.

Reinhart, B. J., Slack, F. J., Basson, M., *et al.* (2000). The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403,** 901–906.

Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., *et al.* (2002). MicroRNAs in plants. *Genes Dev.* **16,** 1616–1626.

Ripoche, M. A., Kress, C., Poirier, F., *et al.* (1997). Deletion of the *H19* transcription unit reveals the existence of a putative imprinting control element. *Genes Dev.* **11,** 1596–1604.

Rivas, E., and Eddy, S. R. (2000). Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs. *Bioinformatics* **16,** 583–605.

Rivas, E., and Eddy, S. R. (2001). Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics* **21,** 8.

Rivas, E., Klein, R. J., Jones, T. A., *et al.* (2001). Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Curr. Biol.* **11,** 1369–1373.

Rogan, D. F., Cousins, D. J., and Staynov, D. Z. (1999). Intergenic transcription occurs throughout the human IL-4/IL-13 gene cluster. *Biochem. Biophys. Res. Commun.* **255,** 556–561.

Routledge, S. J., and Proudfoot, N. J. (2002). Definition of transcriptional promoters in the human beta globin locus control region. *J. Mol. Biol.* **323,** 601–611.

Roy, S. W., Fedorov, A., and Gilbert, W. (2002). The signal of ancient introns is obscured by intron density and homolog number. *Proc. Natl. Acad. Sci. USA* **99,** 15513–15517.

Rozhdestvensky, T. S., Kopylov, A. M., Brosius, J., *et al.* (2001). Neuronal BC1 RNA structure: Evolutionary conversion of a tRNA-Ala domain into an extended stem-loop structure. *RNA.* **7,** 722–730.

Runte, M., Huttenhofer, A., Gross, S., *et al.* (2001). The *IC-SNURF-SNRPN* transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A. *Hum. Mol. Genet.* **10,** 2687–2700.

Ruvkun, G., Ambros, V., Coulson, A., *et al.* (1989). Molecular genetics of the *Caenorhabditis elegans* heterochronic gene *lin-14*. *Genetics* **121,** 501–516.

Sado, T., Wang, Z., Sasaki, H., *et al.* (2001). Regulation of imprinted X-chromosome inactivation in mice by *Tsix*. *Development* **128,** 1275–1286.

Saha, S., Ansari, A. Z., Jarell, K. A., *et al.* (2003). RNA sequences that work as transcriptional activating regions. *Nucleic Acids Res.* **31,** 1565–1570.

Sanchez-Herrero, E., and Akam, M. (1989). Spatially ordered transcription of regulatory DNA in the bithorax complex of *Drosophila*. *Development* **107,** 321–329.

Sawata, M., Yoshino, D., Takeuchi, H., *et al.* (2002). Identification and punctate nuclear localization of a novel noncoding RNA, Ks-1, from the honeybee brain. *RNA* **8,** 772–785.

Schattner, P. (2003). Computational gene-finding for noncoding RNAs. *In* "Noncoding RNAs: Molecular Biology and Molecular Medicine" (J. Barciszewski and V. A. Erdmann, Eds.), pp. 34–49. Kluwer Academic/Plenum Publishers, Dordrecht, The Netherlands.

Schmidt, J. V., Matteson, P. G., Jones, B. K., *et al.* (2000). The *Dlk1* and *Gtl2* genes are linked and reciprocally imprinted. *Genes Dev.* **14,** 1997–2002.

Schoenherr, C. J., Levorse, J. M., and Tilghman, S. M. (2003). CTCF maintains differential methylation at the *Igf2/H19* locus. *Nat. Genet.* **33,** 66–69.

Seggerson, K., Tang, L., and Moss, E. G. (2002). Two genetic circuits repress the *C. elegans* heterochronic gene *lin-28* after translation initiation. *Dev. Biol.* **243,** 215–225.

Sharp, T. A., Schwemmle, M., Jeffrey, I., *et al.* (1993). Comparative analysis of the regulation of the interferon-inducible protein kinase PKR by Epstein-Barr virus RNAs EBER-1 and EBER-2 and adenovirus VAI RNA. *Nucleic Acids Res.* **21,** 4483–4490.

Shi, Y., Downes, M., Xie, W., *et al.* (2001). Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. *Genes Dev.* **15,** 1140–1151.

Shibata, S., and Lee, J. T. (2003). Characterization and quantitation of differential *Tsix* transcripts: Implications for *Tsix* function. *Hum. Mol. Genet.* **12,** 125–136.

Slack, F. J., Basson, M., Liu, Z., *et al.* (2000). The *lin-41* RBCC gene acts in the *C. elegans* heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. *Mol. Cell* **5,** 659–669.

Sledjeski, D. D., and Gottesman, S. (1995). A small RNA acts as an antisilencer of the H-NS-silenced *rcsA* gene of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **92,** 2003–2007.

Sledjeski, D. D., Whitman, C., and Zhang, A. (2001). Hfq is necessary for regulation by the untranslated RNA DsrA. *J. Bacteriol.* **183,** 1997–2005.

Sleutels, F., Zwart, R., and Barlow, D. P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* **415,** 810–813.

Smilinich, N. J., Day, C. D., Fitzpatrick, G. V., *et al.* (1999). A maternally methylated CpG island in *KvLQT1* is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome. *Proc. Natl. Acad. Sci. USA* **96,** 8064–8069.

Smith, C. A., Katz, M., and Sinclair, A. H. (2003). *DMRT1* is upregulated in the gonads during female-to-male sex reversal in ZW chicken embyos. *Biol. Reprod.* **68,** 560–570.

Srikantan, V., Zou, Z., Petrovics, G., *et al.* (2000). *PCGEM1*, a prostate-specific gene, is overexpressed in prostate cancer. *Proc. Natl. Acad. Sci. USA* **97,** 12216–12221.

Stavropoulos, N., Lu, N., and Lee, J. T. (2001). A functional role for *Tsix* transcription in blocking *Xist* RNA accumulation but not in X-chromosome choice. *Proc. Natl. Acad. Sci. USA* **98,** 10232–10237.

Stevenson, D. S., and Jarvis, P. (2003). Chromatin silencing: RNA in the driving seat. *Curr. Biol.* **13,** R13–R15.

Storz, G. (2002). An expanding universe of noncoding RNAs. *Science* **296,** 1260–1263.

Stougaard, P., Molin, S., and Nordström, K. (1981). RNAs involved in copy-number control and incompatibility of plasmid R1. *Proc. Natl. Acad. Sci. USA* **78,** 6008–6012.

Szymanski, M., and Barciszewski, J. (2002). Beyond the proteome: Non-coding regulatory RNAs. *Genome Biol.* **3,** reviews 0005.

Szymanski, M., Erdmann, V. A., and Barciszewski, J. (2003). Noncoding regulatory RNAs database. *Nucleic Acids Res.* **31,** 429–431.

Takada, S., Tevendale, M., Baker, J., *et al.* (2000). Delta-like and gtl2 are reciprocally expressed, differentially methylated linked imprinted genes on mouse chromosome 12. *Curr. Biol.* **10,** 1135–1138.

Takeda, K., Ichijo, H., Fujii, M., *et al.* (1998). Identification of a novel bone morphogenetic protein-responsive gene that may function as a noncoding RNA. *J. Biol. Chem.* **273,** 17079–17085.

Takemoto, N., Kamogawa, Y., Lee, H. J., *et al.* (2000). Cutting edge: Chromatin remodeling at the IL-4/IL-13 intergenic regulatory region for Th2-specific cytokine gene cluster. *J. Immunol.* **165,** 6687–6691.

Teranishi, M., Shimada, Y., Hori, T., *et al.* (2001). Transcripts of the MHM region on the chicken Z chromosome accumulate as non-coding RNA in the nucleus of female cells adjacent to the *DMRT1* locus. *Chromosome Res.* **9,** 147–165.

The *C. elegans* Sequencing Consortium. (1998). Genome sequence of the nematode *C. elegans:* A platform for investigating biology. *Science* **282,** 2012–2018.

Tiedge, H., Fremeau, R. T., Weinstock, P. H., *et al.* (1991). Dendritic location of neural BC1 RNA. *Proc. Natl. Acad. Sci. USA.* **88,** 2093–2097.

Tiedge, H., Chen, W., and Brosius, J. (1993). Primary structure, neural-specific expression, and dendritic location of human BC200 RNA. *J. Neurosci.* **13,** 2382–2390.

Toczyski, D. P., Matera, A. G., Ward, D. C., *et al.* (1994). The Epstein-Barr virus (EBV) small RNA EBER1 binds and relocalizes ribosomal protein L22 in EBV-infected human B lymphocytes. *Proc. Natl. Acad. Sci. USA* **91,** 3463–3467.

Tomizawa, J., Itoh, T., Selzer, G., *et al.* (1981). Inhibition of ColE1 RNA primer formation by a plasmid-specified small RNA. *Proc. Natl. Acad. Sci. USA* **78,** 1421–1425.

Tommasi, S., and Pfeifer, G. P. (1999). *In vivo* structure of two divergent promoters at the human *PCNA* locus. *J. Biol. Chem.* **274,** 27829–27838.

Tora, L., Gronemeyer, H., Turcotte, B., *et al.* (1988). The N-terminal region of of the chicken progesterone receptor specifies target gene activation. *Nature* **333,** 185–188.

Tuan, D., Kong, S., and Hu, K. (1992). Transcription of the hypersensitive site HS2 enhancer in erythroid cells. *Proc. Natl. Acad. Sci. USA* **89,** 11219–11223.

Turner, B. M., Birley, A. J., and Lavender, J. (1992). Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in *Drosophila* polytene nuclei. *Cell* **69,** 375–384.

Venter, C. J., Adams, M. D., Myers, E. W., *et al.* (2001). The sequence of the human genome. *Science* **291,** 1304–1351.

Verona, R. I., and Bartolomei, M. S. (2003). The structure regulation and function of the imprinted H19 RNA. *In* "Noncoding RNAs: Molecular Biology and Molecular Medicine" (J. Barciszewski and V. A. Erdmann, Eds.), pp. 83–93. Kluwer Academic/Plenum Publishers, Dordrecht, The Netherlands.

Vincent, J. B., Herbrick, J. A., Gurling, H. M., *et al.* (2000). Identification of a novel gene on chromosome 7q31 that is interrupted by a translocation breakpoint in an autistic individual. *Am. J. Hum. Genet.* **67,** 510–514.

Vincent, J. B., Petek, E., Thevarkunnel, S., *et al.* (2002). The *RAY1/ST7* tumor-suppressor locus on a chromosome 7q31 represents a complex multi-transcript system. *Genomics* **80,** 283–294.

Volpe, T. A., Kidner, C., Hall, I. M., *et al.* (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297,** 1833–1837.

Vuyisich, M., Spanggord, R. J., and Beal, P. A. (2002). The binding site of the RNA-dependent protein kinase (PKR) on EBER1 RNA from Epstein-Barr virus. *EMBO Rep.* **3,** 622–627.

Wagner, E. G. H., and Vogel, J. (2003). Noncoding RNAs encoded by bacterial chromosomes. *In* "Noncoding RNAs: Molecular Biology and Molecular Medicine" (J. Barciszewski and V. A. Erdmann, Eds.), pp. 243–259. Kluwer Academic/Plenum Publishers, Dordrecht, The Netherlands.

Wang, H., Iacoangeli, A., Popp, S., *et al.* (2002). Dendritic BC1 RNA: Functional role in regulation of translation initiation. *J. Neurosci.* **22,** 10232–10241.

Wang, Y., Crawford, D. R., and Davies, K. J. (1996). adapt33, a novel oxidant-inducible RNA from hamster HA-1 cells. *Arch. Biochem. Biophys.* **332,** 255–260.

Wassarman, K. M., and Storz, G. (2000). 6S RNA regulates E. coli RNA polymerase activity. *Cell* **101,** 613–623.

Wassarman, K. M., Repoila, F., Rosenow, C., *et al.* (2001). Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev.* **15,** 1637–1651.

Watanabe, M., Yanagisawa, J., Kitagawa, H., *et al.* (2001). A subfamily of RNA-binding DEAD-box proteins acts as an estrogen receptor α coactivator through the N-terminal activation domain (AF-1) with an RNA coactivator, SRA. *EMBO J.* **20,** 1341–1352.

Wevrick, R., and Francke, U. (1997). An imprinted mouse transcript homologous to the human imprinted in Prader-Willi syndrome *(IPW)* gene. *Hum. Mol. Genet.* **6,** 325–332.

Wevrick, R., Kerns, J. A., and Francke, U. (1994). Identification of a novel paternally expressed gene in the Prader-Willi syndrome region. *Hum. Mol. Genet.* **3,** 1877–1882.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* **75,** 855–862.

Winker, S., Overbeek, R., Woese, C. R., *et al.* (1990). Structure detection through automated covariance search. *Comput. Appl. Biosci.* **64,** 365–371.

Winkler, W., Nahvi, A., and Breaker, R. R. (2002). Thiamine derivatives bind messenger RNA directly to regulate bacterial gene expression. *Nature* **419,** 952–956.

Wolf, S., Mertens, D., Schaffner, C., *et al.* (2001). B-cell neoplasia associated gene with multiple splicing *(BCMS)*: The candidate B-CLL gene on 13q14 comprises more than 560 kb covering all critical regions. *Hum. Mol. Genet.* **10,** 1275–1285.

Wroe, S. F., Kelsey, G., Skinner, J. A., *et al.* (2000). An imprinted transcript, antisense to *Nesp*, adds complexity to the cluster of imprinted genes at the mouse *Gnas* locus. *Proc. Natl. Acad. Sci. USA* **97,** 3342–3346.

Wu, X., Weigel, D., and Wigge, P. A. (2002). Signaling in plants by intracellular RNA and protein movements. *Genes Dev.* **16,** 151–158.

Wutz, A. (2003). *Xist* RNA associates with chromatin and causes gene silencing. *In* "Noncoding RNAs: Molecular Biology and Molecular Medicine" (J. Barciszewski and V. A. Erdmann, Eds.), pp. 50–66. Kluwer Academic/Plenum Publishers, Dordrecht, The Netherlands.

Wutz, A., Smrzka, O. W., Schweifer, N., *et al.* (1997). Imprinted expression of the *Igf2r* gene depends on an intronic CpG island. *Nature* **389,** 745–749.

Wutz, A., Rasmussen, T. P., and Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat. Genet.* **30,** 167–174.

Wylie, A. A., Murphy, S. K., Orton, T. C., *et al.* (2000). Novel imprinted *DLK1/GTL2* domain on human chromosome 14 contains motifs that mimic those implicated in *IGF2/H19* regulation. *Genome Res.* **10,** 1711–1718.

Yamasaki, K., Hayashida, S., Miura, K., *et al.* (2000). The novel gene, *gamma2-COP (COPG2)*, in the 7q32 imprinted domain escapes genomic imprinting. *Genomics* **68,** 330–335.

Yamashita, M., Ukai-Tadenuma, M., Kimura, M., *et al.* (2002). Identification of a conserved GATA3 response element upstream proximal from the interleukin-13 gene locus. *J. Biol. Chem.* **277,** 42399–42408.

Yang, Z., Zhu, Q., Luo, K., *et al.* (2001). The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature* **414,** 317–322.

Yi-Brunozzi, H. Y., Easterwood, L. M., Kamilar, G. M., *et al.* (1999). Synthetic substrate analogs for the RNA-editing adenosine deaminase ADAR-2. *Nucleic Acids Res.* **27,** 2912–2917.

Yu, S., Yu, D., Lee, E., *et al.* (1998). Variable and tissue-specific hormone resistance in heterotrimeric Gs protein alpha-subunit (Gsα) knockout mice is due to tissue-specific imprinting of the gsalpha gene. *Proc. Natl. Acad. Sci. USA* **95,** 8715–8720.

Zalfa, F., Giorgi, M., Primerano, B., *et al.* (2003). The fragile X syndrome protein FMRP associates with BC1 RNA and regulates the translation of specific mRNAs at synapses. *Cell* **112,** 317–327.

Zdobnov, E. M., von Mering, C., Letunic, I., *et al.* (2002). Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* **298,** 149–159.

Zenklusen, J. C., Conti, C. J., and Green, E. D. (2001). Mutational and functional analyses reveal that *ST7* is a highly conserved tumor suppressor gene on human chromosome 7q31. *Nat. Genet.* **27,** 392–398.

Zhang, A., Altuvia, S., Tiwari, A., *et al.* (1998). The OxyS regulatory RNA represses rpoS translation and binds the Hfq (HF-I) protein. *EMBO J.* **17,** 6061–6068.

Zhang, A., Wassarman, K. M., Ortega, J., *et al.* (2002). The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Mol. Cell.* **9,** 11–22.

Zheng, W., and Flavell, R. A. (1997). The transcription factor GATA-3 is necessary and sufficient for Th2 cytokine gene expression in CD4 T cells. *Cell* **89,** 587–596.

Zhuo, D., Zhao, W. D., Wright, F. A., *et al.* (2001). Assembly, annotation, and integration of UNIGENE clusters into the human genome draft. *Genome Res.* **11,** 904–918.

Zilberman, D., Cao, X., and Jacobsen, S. E. (2003). *ARGONAUTE4* control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299,** 716–719.

Zwart, R., Sleutels, F., Wutz, A., *et al.* (2001). Bidirectional action of the *Igf2r* imprint control element on upstream and downstream imprinted genes. *Genes Dev.* **15,** 2361–2366.