CMLS, Cell. Mol. Life Sci. 58 (2001) 960–977
1420-682X/01/070960-18 $ 1.50 + 0.20/0
© Birkhäuser Verlag, Basel, 2001

CMLS Cellular and Molecular Life Sciences

Review

Regulatory RNAs

V. A. Erdmann[a], M. Z. Barciszewska[b], A. Hochberg[c], Nathan de Groot[c] and J. Barciszewski[b,*]

[a] Institute of Biochemistry, Freie Universität Berlin, Thielallee 63, 14195 Berlin (Germany)
[b] Institute of Bioorganic Chemistry of the Polish Academy of Sciences, Noskowskiego 12, 61704 Poznań (Poland), Fax +48 61 852 05 32, e-mail: jbarcisz@ibch.poznan.pl
[c] Department of Biological Chemistry, Institute of Life Sciences, The Hebrew University, 91904 Jerusalem (Israel)

Received 20 November 2000; received after revision 16 January 2001; accepted 14 February 2001

Abstract. In addition to mRNA, rRNA and tRNA, which play central roles within cells, there are a number of regulatory, non-coding RNAs (ncRNAs). Of varying lengths, ncRNAs have no long open reading frame. While not encoding proteins, they may act as riboregulators, and their main function is posttranscriptional regulation of gene expression. Many ncRNAs have been identified and characterized both in prokaryotes and eukaryotes, and are involved in the specific recognition of cellular nucleic acid targets through complementary base pairing, controlling cell growth and differentiation. Some are associated with the abnormalities in imprinted inheritance that occur in several well-known developmental and neurobehavioral disorders. Other ncRNAs accomplish regulation by modulating the activity of proteins. Several rRNAs are able to sustain enzymatic reactions implicated in the translation process including synthesis of peptide bonds within the ribosome. The different roles played by widely distributed RNAs acting in diverse ways, suggest the flexibility and versality of these molecules in regulatory processes. This review summarizes the available biochemical and structural data on known regulatory RNAs.

Key words. Non-coding RNA; antisense RNA; regulatory RNA.

Introduction

RNA plays a variety of structural, informational, catalytic, and regulatory roles in the cell. Certain RNAs show catalytic properties and act as ribozymes. 23S and 28S rRNAs have peptidyltransferase activity. As aptamers, some RNAs bind small molecular compounds, e. g., metal ions and nucleotides. Much data has accumulated in recent years showing that various non-translatable, noncoding RNA (ncRNA) transcripts are synthesized in different cells. These RNAs lack protein-coding capacity and may exert their action mainly or exclusively at the RNA level. RNAs functioning as genetic regulators (riboregulators) by acting either in cis or in trans have been demonstrated widely in bacteria, and eukaryotes including nematodes, plants, and mammals. Regulatory RNAs are synthesized in most cases by pol II, are polyadenylated and spliced. Although much effort has been invested in trying to identify functions and, in particular, mechanisms of action of ncRNAs, they still remain mysterious. This knowledge will come through better understanding of their origin, synthesis, structure, and the nature of their interactions with other components of the cell. An analysis of the origin, functions, and properties of ncRNAs suggests they may be divided five groups:

1) DNA markers. These are involved in dosage compensation and imprinting. This group consists of *Xist*, *roX*, *PAT-1*, *Tsix*, *XistAS*, *H19*, and *IPW*.

2) Gene regulators. These affect the activity of genes by different mechanisms such as silencing or RNA-RNA interaction. The group includes *NTT*, *DGCR5*, *KvLQTI-AS*, *Nesp/GNAS*, *SCA8*, *CMPD*, *lin-4*, *let-7*, *UBE3A*,

* Corresponding author.

ZNF127-AS, ScYc, DISC2, sok CopA, RNAI, pnd, and *RNA-OUT.*
3) Abiotic stress signals. These RNAs are synthesized in response to abiotic stress (e.g., oxidation) and include *gadd7/adapt15*, *adapt33*, *hsrω*, *G90*, *OxyS, DsrA,* and *PCGEM1*.
4) Biotic stress signals. RNAs that belong to this group are inducible by biologically active molecules, e.g., cytokinins. They are *His-1, ENOD40*, *lbi*, *CR20,* and *GUT15*.
5) Other RNAs in this group have various origins and functions. It includes *Bsr, BC1, BC200, SRA, meiRNA, UHG* and *Xlsirt.*
The above grouping of ncRNAs is not very precise. Lacking structural information, it is based on limited biochemical data (Table 1) [1, 2]. Their nucleotide sequences are stored in the Non-Coding RNA database at http://biobases.ibch.poznan.pl/ncRNA/ [1].

Table 1. List of non-coding RNAs and their characteristics (kb, kilobases; n/d, not determined; n/a, not available; nt, nucleotides).

Non-coding RNA	Size	EMBL/GenBank Acc. No. or Reference	Remarks
DNA markers			
1) Dosage compensation RNAs			
Homo sapiens	16.5 kb	M97168	
Mus musculus	14.7 kb	L04961	
Bos taurus	n/d	AF104906	partial sequence
Equus caballus	n/d	U50911	partial sequence
Oryctolagus cuniculus	n/d	U50910	partial sequence
Drosophila melanogaster roX1	3749 nt	U85980	
Drosophila melanogaster roX2	1293 nt	U85981	
HZ-1 virus *PAT-1*	2937 nt	U03488	
Tsix (*Homo sapiens*) Tsix	40 kb	[12]	
XistAS mouse *Mus musculus*	n/d	[13]	
2) *H19*			
Homo sapiens	2313 nt	M32053	
Mus musculus	1899 nt	X58196	
Rattus rattus	2297 nt	X59864	
Oryctolagus cuniculus	1842 nt	M97348	partial sequence
Pongo pygmaeus	1644 nt	AF190058	partial sequence
Felis catus	1747 nt	AF190057	partial sequence
Lynx lynx	879 nt	AF190056	partial sequence
Ovis aries	397 nt	AF105429	partial sequence
Thomomys monticola	875 nt	AF190055	partial sequence
Elephantidae Gen. sp	856 nt	AF190054	partial sequence
Peromyscus maniculatus	2094 nt	AF214115	
3) IPW			
Homo sapiens	2075 nt	U12897	
Mus musculus	734 nt	U69888	partial sequence
Gene regulators			
1) *NTT*			
Homo sapiens	17 kb	U54776	
2) *DGCR5*			
Homo sapiens	1284 nt	X91348	
3) K*KvLQTA-AS*			
Homo sapiens	n/d	n/a	
Mus musculus	n/d	AF119385	partial intron sequence
4) Nesp/*GNAS*			
Homo sapiens	828 nt	AJ251760	partial sequence
Mus musculus	1083 nt	AF173359	
5) *SCA8*			
Homo sapiens	32.3 kb	AF252279	partial sequence
6) C*MPD* associated RNA			
Homo sapiens	3414 nt	D43770	
7) Developmental timing			
Caenorhabditis brigsae let-7	21 nt	AF210771	
Caenorhabditis elegans lin-4	22 nt/61 nt	U01830	

Table 1 (continued)

Non-coding RNA	Size	EMBL/GenBank Acc. No. or Reference	Remarks
8) Other non-coding RNA transcripts			
Homo sapiens UBE3A antisense	n/d	n/a	
Homo sapiens DISC2	15 kb	AF222981	
Homo sapiens ZNF127-AS	1827 nt	U19107	
Styela clava ScYc RNA	1.1 kb	L42757	
9) Antisense plasmid			
sok RNA	66 nt	AP000342	R100 plasmid
finP RNA	2778 nt	AP000342	R100 plasmid
CopA	186 nt	V00326	R1 plasmid
RNAI	107 nt	J01566	ColE1 plasmid
Selenomonas ruminantium CtRNA	88 nt	Z49917	pJJM1 plasmid
Escherichia coli Incl	363 nt	M34837	Col Ib-P9 plasmid
Streptococcus pneumoniae RNAII	111 nt	S81045	pLS1 plasmid
Streptococcus agalactiae RNAII		L03355	pIP501 plasmid
Escherichia coli RNAI	73 nt	M28718	pMU 720 plasmid
Abiotic stress signals			
1) *gadd7/adapt15, adapt33, vseap1*			
Cricetulus griseus gadd7	754 nt	L40430	
Cricetulus griseus adapt15	746 nt	U26833	adapt15-P9
	753 nt	U26834	adapt15-P8
Cricetulus griseus adapt33	1290 nt	U29660	adapt33A
	1186 nt	U29661	adapt33B
Cric etulus griseus vseap1	0.9 kb	AJ003192	
	3.1 kb		
2) *hsr-ω*			
Drosophila melanogaster	1174 nt	U18307	alternative poly(A) sites
	1190 nt		
Drosophila hydei	1129 nt	M14558; J02629	
Drosophila pseudoobscura	1213 nt	X16337; X16157;	
3) *G90*			
Mus musculus	1357 nt	AJ132433	
4) *OxyS*			
Escherichia coli	110 nt	U87390	
5) *DsrA*			
Escherichia coli	86 nt	U17136	
Salmonella typhimurium	82 nt	AF090431	putative
Klebsiella pneumoniae	82 nt	AF090431	
6) *DD3/PCGEM1*			
Homo sapiens	3800 nt	AF103907	
	1600 nt	AF22389	
Biotic stress signals			
1) *His-1*			
Homo sapiens	n/d	U56440	gene sequence, exon structure
Mus musculuss	3053 nt	U09772	unknown alternatively spliced
	3003 nt	U10269	forms of the same pre-mRNA
2) *ENOD40*			
Glycine max	679 nt	X69154	ENOD40-1
	617 nt	X69155	ENOD40-2
Pisum sativum	702 nt	X81064	
Phaseolus vulgaris	600 nt	X86441	
Vicia sativa	718 nt	X83683	
Trifolium repens	631 nt	AJ000268	
Lotus japonicus	770 nt	AF013594	
Medicago sativa	626 nt	X80263	
	733 nt	L32806	
Medicago truncatula	920 nt	X80262	
Nicotiana tabacum	470 nt	X98716	
Vigna radiata	331 nt	AF061818	partial sequence
Sesbania rostrata	638 nt	Y12714	
3) lbiRNA			
Bacteriophage Acm1	97 nt	Z30964	

Table 1 (continued)

Non-coding RNA	Size	EMBL/GenBank Acc. No. or Reference	Remarks
4) *CR20*			
Cucumis sativus	1108 nt	D79216	
Arabidopsis thaliana	758 nt	D79218	
5) *GUT15*			
Arabidopsis thaliana	1377 nt	U84973	
Nicotiana tobaccum	1670 nt	U84972	
Other functions			
1) Bsr RNA			
Rattus norvegicus	4723 nt	AB014883	the isolated clones
	920 nt	AB014882	contrain various number of
	2032 nt	AB014881	~ 0.9-kb repeats units
	1198 nt	AB014880	
	1773 nt	AB014879	
	2244 nt	AB014878	
	1755 nt	AB014877	
BC1 RNA			
Rattus rattus	152 nt	M16113	
Peromyscus maniculatus	391 nt	U33851	
Peromyscus californicus	359 nt	U33850	
Meriones unguiculatus	350 nt	U33852	
Mus musculuss	152 nt	U01310	
Mesocricetus auratus	142 nt	U01309	
Cavia porcellus	165 nt	U01304	
BC200 RNA			
Homo sapiens	200 nt	AF020057, U01306	
Saguinus oedipus	195 nt	AF067788	
Saguinus imperator	194 nt	AF067787	
Aotus trivirgatus	196 nt	AF067786	
Macaca fascicularis	200 nt	AF067785	
Macaca mulatta	200 nt	AF067784	
Chlorocebus aethiops	205 nt	AF067783	
Papio hamadryas	197 nt	AF067782	
Hylobates lar	203 nt	AF067781	
Pongo pygmaeus	198 nt	AF067780	
Gorilla gorilla	204 nt	AF067779	
Pan paniscus	205 nt	AF067778	
2) *SRA*			
Homo sapiens	875 nt	AF092038	
Mus musculuss	829 nt	AF092039	
3) *meiRNA*			
Schizosaccharomyces pombe	508 nt	D31852	
4) UHG			
Homo sapiens U22HG	1114 nt	U40580	
Mus musculus U22HG	590 nt	U40654	
Homo sapiens U17HG	885 nt	AJ006834	variant A
	2139 nt	AJ006835	variant AB
Mus musculus U17HG	1682 nt	AJ006836	
	383 nt	AJ006837	
Homo sapiens U19HG	681 nt	AJ224167	
	785 nt	AJ224166	
	310 nt	AJ224170	
	375 nt	AJ224169	
	666 nt	AJ224168	
Homo sapiens Gas5	4055 nt	AF141346	
5) *Xlsirt* RNA			
Xenopus laevis	76 nt	S67412	single repeat sequences
	79 nt	S67413	
	78 nt	S67414	
	80 nt	S67415	

DNA markers

Dosage compensation RNAs

Most genes are inherited in two copies, one from each parent. The sex-linked genes in males and mitochondrial genes are exceptions. Whereas most genes have identical properties, imprinted genes usually function only when inherited from either the mother or the father. These genes are 'imprinted' by one of the parents, a phenomenon termed parental, genomic, or gametic imprinting. It is an epigenetic event in which gene activity is reversibly modified depending on the sex of the parent who transmits it. Genomic imprinting leads to allele-specific expression depending on the parent of origin of the allele. Thus, in contrast to the expectations of simple Mendelian inheritance, imprinting results in the unequal expression of maternally and paternally derived alleles in an individual: one of the alleles is usually suppressed during development. The imprinted copy of the gene is considered to be the silent one. In the simplest case, the silent copy of the gene is the one actively regulated, or imprinted, in one of the gametes, while the active copy is in the default state. However, in some cases, the active copy may be actively regulated in the gamete, with the silent copy in the default state. Over 40 imprinted genes have been identified to date in the combined human and mouse genomes (http://www.geneimprint.com/ and http://www.mgu.har.mrc.ac.uk/anomaly/anomaly.html).

Imprinted genes are not distributed as single units throughout the genome, but tend to cluster together [3–5]. Normal mammalian development requires the correct parental contribution of imprinted genes. Their lack or aberrant expression leads to a variety of abnormalities [6]. One fundamental difference between male and female cells is the number for X chromosomes. The difference in their content and the requirement for equal expression of X genes in both sexes has led to the evolution of several types of dosage-compensation mechanism. In mammals and *Drosophila*, this process involves expression of specific genes, whose products (RNA transcripts) do not contain long open reading frames [1, 2, 7–13]. They act as RNA and together with specific proteins are responsible for chromatin remodeling. In mammals, both X chromosomes are transcriptionally active at the early stages of XX embryo development. However, upon differentiation, a single X chromosome is inactivated in early embryogenesis and that state is then stably maintained through subsequent cell generations. Inactivation of the paternally or maternally inherited X chromosome is random in somatic cells. Spontaneous non-programmed reactivation is extremely rare [14].

The inactive X chromosome is distinguished from the active one by the following properties: transcriptional down-regulation of nearly the entire chromosome, heterochromatic condensation at interphase of the cell cycle, late replication during S phase, DNA methylation of cytosine residues at CpG dinucleotides in the 5′ region of X-linked genes, hypoacetylation of histone H4 and expression of *Xist* (X-inactive specific transcript) located at the X inactivation centre (*Xic*).

Xic loci on the X chromosome are initially 'counted', given that at least two copies of *Xic* are present per diploid genome for inactivation to occur. The counting process ensures that one X chromosome remains active in diploid cells [14]. The essential function of *Xic* is assured by a 450-kb region containing the *Xist* gene (in mouse) or *XIST* (in human). *XIST*/*Xist* is expressed exclusively from the inactive X chromosome in somatic cells and produces a large RNA of 19.3 kb in humans [15] and 17.4 kb in the mouse [16], for which no protein products have been identified [17]. The presence of the *Xist* gene in cis is both necessary and sufficient for X inactivation. *Xist* RNA is not exported from the nucleus and is associated with the inactive X chromosome. In differentiated cells, *Xist* RNA is produced solely from the inactive X chromosome (*Xi*), binding to and coating it. It does not interact directly with DNA but is likely to participate in a ribonucleoprotein complex. The protein content of *Xi* differs from that of the active X chromosome (*Xa*) and the autosomes [18]. The large increase in quantity of *Xist* RNA in the nucleus at the onset of X chromosome inactivation results from stabilization of the *Xist* RNA rather than an increase in the rate of transcription. Multiple copies of *Xist* transgenes inserted into mouse chromosome 12 produced an RNA transcript that coated the autosome, which became hypoacetylated, was replicated late, and transcriptionally inactive [19, 20].

In mouse embryonic stem cells, *Xist* is expressed from two active X chromosomes, but *Xist* is up-regulated prior to the inactivation of one X chromosome. This developmental activation is a consequence of the stabilization of *Xist* RNA due to a promoter switch, which is an effect of chromatin hyperacetylation upstream of *Xist*. On the other hand, hypoacetylation leads to the further progression of X inactivation involving a cascade of events. An increase in the level of *Xist* RNA associated with the X that will become inactivated is essential for the downstream events such as late replication, H4 hypoacetylation, DNA methylation, and down-regulation of gene expression which seems to proceed in a different order during differentiation. Histone deacetylases are also components of an X inactivation machinery that is recruited to the promoters of genes that are silenced on *Xi* [19–21]. Two functionally distinct promoters, P1 and P0, have been suggested to act within the *Xist* gene. P1 (and P2) is used in differentiated female cells for production of a stable transcript that coats the inactive X chromosome. A second promoter (P0) is located 6.6 kb upstream of P1 and produces unstable *Xist* transcripts in undifferentiated cells that have not undergone X inactivation. The location

of P0 coincides with a ribosomal protein S12 pseudogene, and transcription of the *Xist* P1 promoter is antisense to *Xist* and represents the 3′ end of the *Tsix* gene [22]. The transcribed *Xist* allele on *Xi* is more compact than the silenced *Xist* allele on *Xa* but its promoter region remains accessible to nucleases [23].

Imprinted genes in mammals are expressed from only one chromosome, but either the maternally or paternally inherited one is silenced. Such genes include *Igf2* and other factors involved in growth control. Imprinting might work to balance maternal and paternal demands on the size of the embryo. Differential methylation which distinguishes the two copies of an imprinted gene originates in the parental germ cells, in the egg or sperm. Gene methylation ensures that it cannot be translated. DNA sequences involved in spreading of RNA and silencing into attached autosomes are not specific to the X chromosome. However, because the spread and inactivation are less efficient in autosomes, they must be promoted by something other than X chromatin. It turned out that the X chromosome of human and mouse are rich in L1 LINE elements (26%), of which there are only 13% in autosomal DNA [24–27]. These properties of L1 distribution on the X chromosome provide strong evidence that L1 elements may serve as DNA signals to propagate X inactivation along the chromosome [27, 28]. Monitoring with antibodies against acetylated H4 histone shows accelerated shortening of telomere repeats in the inactive X chromosome. This suggests that epigenetic factors modulate not only the length but also the rate of age-associated telomere shortening in human cells in vivo [29].

Deletion of a 65-kb region downstream of *Xist* results in its constitutive expression and X inactivation, implying the presence of a *cis*-regulatory element. In this region, the *Tsix* gene codes for an RNA anti-sense to *Xist* RNA [30, 31]. *Tsix* regulates *Xist* in *cis* and determines X chromosome choice without affecting silencing. Therefore, counting, choice, and silencing are genetically separable [14, 32]. *Tsix* is a 40-kb RNA originating 15 kb downstream of *Xist* and transcribed across the *Xist* locus [33–35]. Human *Tsix* RNA, which has no conserved open reading frames (ORFs), is seen exclusively in the nucleus, is localized at *Xic* and dynamically regulated during. Similar characteristics are found for anti-sense *Xist (XistAS*) in mouse [30, 31, 36, 37].

X chromosome inactivation and genome imprinting are associated with DNA methylation but the mechanism that initiates de novo modification remains obscure. One can speculate that non-coding RNAs (*Xist* RNAs) which are involved in imprinting direct DNA methylation. Before X inactivation *Tsix* is biallelically expressed but becomes monoallelically expressed at the onset of X inactivation, marking only the future active X and therefore raising the possibility that *Tsix* blocks *Xist* accumulation. Fluorescence in situ hybridization (FISH) experiments revealed that *Tsix* RNA co-localized with *Xist* RNA in the X inactivation centre [30]. Although the methylation status of *Xist* gene promoters of the active and inactive X chromosomes have not been precisely correlated with the presence or absence of *Tsix*, one can assume that an RNA-directed DNA methylation mechanism can contribute to the X inactivation process. *Xist* RNA probably triggers X chromosome methylation, leading to its inactivation [4].

In *Drosophila*, compensation for the reduced dosage of genes located on the single male X chromosome involves doubling their expression relative to their counterparts on female X chromosomes. This mechanism utilizes noncoding, spliced and polyadenylated *roX* RNAs. The location of genes encoding these unusual RNAs mark the X chromosome, which is the target of regulation. *roX* RNAs are associated with a specific acetylated form of histone 4 and elevated gene expression [38]. Dosage compensation is an epigenetic process based on the specific acetylation of histone 4 at lysine 16 by histone acetyltransferase (HAT), which is part of a chromosome-associated dosage compensation complex (DCC) comprising male-specific lethal (MSL) proteins and *roX* RNA. The HAT chromodomain interacts with rox RNA in vivo, contributing to the integration of HAT into the DCC on the male X chromosome. Chromodomains are important for the function of a number of chromatin regulators, but their modes of action remain enigmatic [39]. The dosage compensation process involves products of two genes, *roX1* and *roX2*, which are male specific and do not encode proteins [25, 40]. *roX1* RNA becomes associated with the X chromosome at sites determined by the binding of the *msl* (male-specific lethal) gene products complex. One of the MSL proteins is HAT, the other is DNA helicase, and their action probably leads to the chromatin remodeling allowing increased transcription [12].

rox RNAs are expressed in all somatic cells in males and in diploid cells they co-localize with the MSL proteins in a subregion of the nucleus corresponding to the X chromosome [41]. A two-step process for recognition of the X chromosome by the MSL protein complex has been proposed [38]. In the first step, MSL 1 and MSL 2 together recognize 30–40 sites distributed along the length of the X chromosome, to which they recruit other MSL proteins. Two of these sites encode *rox* RNAs which are incorporated into the growing MSL complex. This can associate with chromatin entry sites on the X chromosome and then spread along the chromosome in *cis* to locate all genes utilizing MSL-mediated dosage compensation [38, 42, 43].

Both flies and mammals face the same problem of equalizing gene expression of two X chromosome in females and one in males. Female mammals inactivate most of the genes of one X chromosome in each cell forming the Barr body. Female flies actively transcribe both X chromosomes and the male is forced to hypertranscribe his single

X. Both mechanisms apparently utilize ncRNA to coat the dosage compensated chromosome: *Xist* in mammals and *roX* RNA in flies [18]. In summary, the RNA species currently knwon to be involved in gene regulation are human *Xist* and *Tsix* RNA, mouse *Xist* and *XistAS* RNA and *Drosophila roX1 and roX2* RNA. Partial *Xist* RNA sequences have been determined for *Oryctolagus cuniclulus* and *Equus caballus*. The Hz-1 virus persistence associated gene 1 (*pag1*) encoding a 2.9-kb non-coding viral *PAT1* RNA shows similarity to *Xist* [44].

RNA in imprinting

The *H19* encodes a capped, spliced and polyadenylated oncofetal 2.3-kb-long RNA [45–48]. It is a developmentally regulated gene and highly expressed in several fetal tissues, except the nervous system and thymus, and is repressed after birth in most organs. In adulthood, basal *H19* gene expression has been detected only in mammary gland, cardiac and skeletal muscles and, to a lesser extent, in kidney, adrenal gland, and lung. Cells with a high *H19* level stop proliferation after 48 h when cultivated in a low-serum-containing medium, while those lacking *H19* continue proliferation [49]. *H19* RNA has been suggested as an adjuvant tumor marker for diagnosis, staging, and follow-up of patients with ovarian serous carcinoma [50, 51].

The maternally expressed *H19* is located 90 kb downstream from the paternally expressed *Igf2* gene on human chromosome 11 and mouse chromosome 7. Due to their reciprocal imprinting and similar expression, these two genes are assumed to be functionally coupled. The intrinsic purpose of *H19* is apparently to imprint *Igf2*. Its 5′ region forms a boundary on the maternal chromosome that prevents downstream enhancers regulating the *Igf2* promoters. Knock-out of the *Igf2* upstream region releases the dominance of *H19*, identifying a mesoderm-specific silencer regulated by methylation and operating independently [52].

The imprinting results from methylation of a 7 to 9-kb domain on the paternal allele of *H19*. The 5′ flank of the *H19* gene contains an imprinting mark region characterized by paternal-allele-specific methylation. An evolutionarily conserved 42-bp upstream element might play a role in imprinting and/or transcriptional regulation of *H19* [53, 54]. The G-rich repeat 1.5 kb upstream of mouse *H19* is present in rats but not in humans and is not essential for *H19* imprinting [55]. The transcriptionally active maternal allele is unmethylated whereas the inactive paternal is methylated. The differentially methylated region (DMR) in the mouse *H19* locus extends over 2 kb at about 4 kb upstream of the gene promoter. A similar pattern of methylation is observed for human *H19*. The exact timing of the erasure of the old and establishment of the new modification pattern remains a very interesting question. *H19* and another imprinted gene, *MESTP/PEG1*, are unmethylated in fetal spermatogonia, suggesting that all pre-existing methylation imprints are already erased by this stage. Methylation of *H19* typical of the paternal allele first appears in a subset of adult spermatogonia and is then maintained in spermatocytes, spermatids and mature spermatozoa. This means that the methylation imprint inherited from the parents is first erased in the male germ line at an early fetal stage and that the paternal-specific imprint is reestablished only later, during spermatogonial differentiation in adult testes [56]. The loss or inactivation of the maternal copy of *H19* has been found to be associated with several tumors, including Wilms' tumor [57]. Human *H19* RNA contains four attachment sites for the oncofetal IGF2 mRNA-binding proteins with K_ds of 0.4–1.3 nM. They are located within a 700-nucleotide segment encoded by exons 4 and 5 [58]. The *H19* genes isolated from human and mouse consist of five exons separated by four unusually short introns. No conserved ORF was found by sequence comparisons. There is a putative ORF in human *H19* that could encode a 256-amino-acid-long protein, but no translation product has been identified. The gene product seems to act as a regulatory RNA and its function requires the ability to fold into a specific secondary structure [59]. Absolutely conserved sequences often lie next to helical parts in both hairpin loops and join regions, suggesting that the structure serves as a framework for presenting these invariant nucleosides [59].

H19 RNA is associated with polysomes in a variety of cell types both in mouse and human [60] and may be an antagonist of *IGF2* expression in trans [57, 61]. Comparison of the mouse and rat gene versions indicates that the RNA is under stabilizing selection and hence is most likely functional [57]. Recently, a chromatin boundary model of genomic imprinting has been proposed. In this scheme, chromatin boundary domains (insulators) act as cis-acting elements that insulate a gene and its regulatory elements from position effects and block transcription when placed between a gene and its enhancer. An insulator located upstream of the *H19* gene isolates *Igf2* from its enhancers. When the regulatory elements are moved between the two genes (upstream of the putative insulator), the enhancers are accessible to *Igf2* but not to *H19*. When the imprinted-control region (ICR) is unmethylated on the maternally inherited chromosome, there are two nuclease-hypersensitive regions which overlap with several short CG-rich repetitive elements conserved in human and mouse. These repeats are targets for the conserved DNA-binding zinc finger protein CTCF. The CCCTC-binding factor interacts with the core insulator sequence via a domain containing 11 zinc fingers. The resulting insulator blocks activation of the maternal copy of *Igf2* by *H19*. The methylated ICR contains no hypersensitive sites and this prevents CTCF binding, thereby inactivat-

ing the insulator and allowing the *H19* enhancer to activate *Igf2* [62–70]. *H19* gene expression was recently demonstrated to be highly expressed in steroid-hormone-sensitive organs when the hormonal stimulation is accompanied by morphological repair [71]. To date, *H19* RNA sequences are known for human, mouse, rat, rabbit and, recently, for lynx, cat, orangutan, elephant, and gopher [65].

IPW RNA

IPW (imprinted gene in the Prader Willi syndrome) has been mapped in the Prader-Willi Syndrome (PWS) region, where it is exclusively expressed from the paternal chromosome. Clinical features of PWS include postnatal failure to thrive with childhood onset of hyperphagia, severe obesity and short stature with neurosecretory growth hormone deficiency. PWS is caused by genomic alterations that inactivate imprinted, paternally expressed genes on human chromosome 15q11–q13. *IPW*, a paternally expressed gene cloned from that region, is not expressed in individuals with PWS, and is therefore a good candidate for involvement in this disorder [72]. The *IPW* (human) and *ipw* (mouse) transcripts are alternatively spliced and do not show long ORFs. They contain multiple copies of a 147-bp repeat arranged in a head-to-tail orientation [45, 46]. However, unlike *H19* and *Xist*, *IPW* is rather poorly conserved between mouse and human in both overall structure and nucleotide sequence. Similarity is restricted to a 500-bp region [72]. A mechanism for *IPW* RNA action has not yet been elucidated.

Gene regulators

NTT RNA

NTT is a human gene (non-coding transcript in T cells) expressed in activated CD4+ T cells. There is a single copy of the *NTT* gene per haploid genome and both alleles are transcriptionally active. It produces a 17-kb non-coding, polyadenylated and non-spliced nuclear RNA. *NTT* has no ORF larger than 270 bp. It may play a role in regulating of neighbouring gene expression, e.g. the interferon-gamma receptor gene [73].

DGCR5

Most cases of DiGeorge syndrome are associated with deletions within 22q1. A critical region of 500 kb is presumed to contain genes of major effect in haplo-insufficiency syndromes. It also contains sequences disrupted by a balanced translocation, which is associated with the ADU breakpoint. The gene encoding alternatively spliced transcripts of sequences including the ADU breakpoint has been named *DGCR5*. It codes non-translated *DGCR5* RNA and its function is not known [74, 75].

KvLQT1-AS

A relatively small number of genes are imprinted and expressed from either the paternally or the maternally inherited allele. Recently, transcription of the silenced allele of the imprinted gene, antisense RNA (AS-RNA), has been implicated in the imprinting mechanism of some of these marked genes. Specific synthesis of these AS-RNAs is made possible through the existence of regions in the imprinted genes whose methylation pattern differs in the two alleles (DMR).

Beckwith-Wiedeman syndrome (BWS) is a human disease characterized by tissue overgrowth, macroglossia, abnormal wall defects, and predisposition to embryonal tumors. Loss of imprinting at the *Igf2* locus is associated with the cancer predisposition condition, BWS. *IGF-2* expression may be responsible for the tissue hypertrophy, but a high level of IGF2 mRNA is connected with Wilms' tumor, rhabdomyosarcoma, and hepatoblastoma [76]. BWS is also related to hypomethylation at KvDMR1, a maternally methylated CpG island within an intron of the *KvLQT1* gene. Imprinting control elements are proposed to exist within this *KvLQT1* locus. Recently, an antisense transcript, *KvLQT1-AS* or *LIT1*, has been found. It is transcribed, exclusively from the paternal allele starting at the above-mentioned locus and in reverse orientation with respect to the maternally expressed *KVLQT1* gene located on chromosome 11p15.5. Its function is not known [6]. *KvLQT1*-AS (*LIT1*) could be an imprinter gene that competes with the target-imprinted gene *KvLQT1* for expression and is silenced by DNA methylation, probably directed by RNA [4]. Down-regulation of *KvLQT1-AS* expression during developmental relaxation of *KvLQT1* imprinting would support the notion of a functional role for antisense RNA in this epigenetic process [6, 77, 78].

Nesp/GNAS

The *Gnas* locus in the distal part of chromosome 2 contains three imprinted genes in the order *Nesp-Gnasxl-Gnas*. These form a single transcription unit because RNA transcripts of *Nesp* and *Gnasxl* are alternatively spliced into exon 2 of *Gnas*. They are expressed from opposite parental alleles and are good candidates for imprinting phenotypes. *Nesp* is paternally imprinted and maternally expressed whereas *Gnasxl* is maternally imprinted and paternally expressed. *Nesp* provides an antisense transcript from the paternally derived allele, which could act as a regulatory element [79–81].

SCA8

Spinocerebellar ataxia type 8 (SCA8) is a neurodegenerative disorder caused by the expansion of a CTG trinucleotide repeat that is transcribed as part of the untranslated RNA. It has been found that the SCA8 is tran-

scribed only through the repeat in the CTG orientation, as in the case for myotonic dystrophy (DM), but not in the CAG orientation, as in the other dominantly inherited ataxias SCA1, 2, 3, 6 and 7. In these latter cases, the CAG expansion is translated into a polyglutamine tract that adds a toxic gain of function to the respective proteins, whereas the CTG expansions in DM and SCA8 are not translated. A mechanism for DM is not known. The RNA transcripts containing the SCA8 CTG are alternatively spliced, polyadenylated, and finally expressed in various brain tissues. No extended ORF has been found [82, 83]. The SCA8 CTG repeat is preceded by a polymorphic but stable CTA tract with the configuration $(CTA)_{1-21}(CTG)_n$. Analyzed affected individuals had an uninterrupted CTG repeat tract or an allele with one or more CCG, CTA, CTC, CCA, or CTT interruptions. In addition, the SCA8 repeat tract in sperm undergoes contractions, with nearly all the resulting expanded alleles having repeat lengths of <100 CTG, a size that is not often associated with disease. These repeat contractions in sperm likely underlie the reduced penetrance associated with paternal transmission [84].

CMPD-associated RNA

Analysis of chromosomal translocations in patients with campomelic dysplasia (CMPD) suggested that gene disruption at 17q24–q25 might be responsible for this disease, which often involves sex reversal in genotypical males. Positional cloning allowed the identification and isolation of a 3.5-kb cDNA that was located close to the SOX9 gene. Northern blot analysis showed that it is specifically expressed in adult testes. No long ORFs were detected in this RNA and no protein was produced in in vitro translation experiments. These results suggest that RNA plays an important role in differentiation and sex determination [85].

Developmental timing RNA

The *Caenorhdabditis elegans* heterochronic gene pathway consists of a cascade of regulatory genes that are temporally controlled to specify the timing of developmental events. The products of the heterochronic genes include transcriptional and translational regulators and two different cases of novel small translational regulatory RNAs [86–90]. Other genes of the pathway encode evolutionarily conserved proteins [88].

The *C. elegans* heterochronic gene *lin-14* generates a temporal gradient of LIN-14 and LIN-28 proteins to control stage-specific patterns of cell lineage during development. Down-regulation of LIN-14 is mediated by the *lin-14* and *lin-28* 3′-untranslated region through interactions with *lin-4* RNAs of 22 and 61 nucleotides [86, 91–94].

Another heterochronic swich gene is *let-7.* Loss of its activity causes reiteration of larval cell fates during the adult stage, whereas increased *let-7* gene dosage causes precocious expression of adult fates during larval stages. *let-7* encodes a temporally regulated 21-nucleotide RNA complementary to the 3′-untranslated regions of *lin-14*, *lin-28, lin-41, lin-42*, and *daf-12.* The sequential stage-specific expression of the *lin-4* and *let-7* regulatory RNAs triggers transitions in the complement of heterochronic regulatory proteins to coordinate developmental timing [86–88]. The *lin-4* and *let-7* genes are the only ones known to encode small RNAs that specifically regulate others, but how they affect gene expression is not known [89, 90]. However, *lin-14* is clearly regulated by *lin-4* RNA at a step after transcription [92–96]. This observation is similar to the *nanos* gene [97]. In that case, the 90-nucleotide translational control element, which has a bipartite secondary structure, is recognized by a repressor and an activator. Two separate but overlapping recognition motifs for translational repressors and localization factors have been identified and provide a mechanism for the switch between translational deactivation and activation [94].

Other ncRNA transcripts

A non-coding antisense RNA overlapping with the ubiquitin protein ligase gene (*UBE3A*) has been found in association with Angelman syndrome [98]. It covers the 3′ half of *UBE3A* and additional sequence downstream. The 5′ end of the antisense transcript is 6.5 kb from the stop codon of *UBE3A* [99].

A novel locus in the human PWS region encodes the antisense *ZNF127* gene [100, 101].

A balanced translocation segregating with schizophrenia can disrupt gene function. Two genes are known: *DISC1*, a protein encoding gene and *DISC2,* which is an ncRNA antisense to *DISC1* [102].

The myoplasm is a cytoplasmic region involved in axis determination, gastrulation, muscle cell specification, and pattern cell division during ascidian development. One-cell zygotes of the ascidian *Styela clava* contain 1.2-kb polyadenylated RNA (yellow crescent RNA, *ScYc* RNA), present throughout embryonic development and associated with the cytoskeleton. *ScYc* RNA contains no long ORF, but a short ORF could encode a putative peptide of 49 amino acids without significant homology to known proteins [103]. It is localized in the cytoplasm and segregates with the larval muscle cells during cleavage. Probes containing the 3′ region of *Yc* RNA were used to identify maternal *Yc*-related RNAs. A cDNA clone encoding the ascidian proliferating cell nuclear antigen (PCNA) has a long 3′-untranslated region containing a 521-nucleotide segment with antisense complementary to part of the 3′ region of *Yc* RNA. This suggests that dif-

ferential cell proliferation in the embryo may be limited by localization of maternal PCNA in RNA and protein [104]. The role of *Yc* RNA may be to promote PCNA mRNA degradation in myoplasm.

Antisense plasmid RNAs

Many genes are involved in programmed cell death in bacteria. Their function has primarily been ascribed to their ability to mediate plasmid maintenance by killing plasmid-free cells. At the same time, bacterial chromosomes encode numerous genes that are homologous to plasmid-encoded killer genes. Two types of loci that mediate plasmid stabilization by postsegregational killing (PSK) have been described. The first, the toxin-antitoxin gene system, encodes a stable toxin and an unstable protein anti-toxin. The other type of PSK genes are regulated by antisense RNAs. The toxins are encoded by stable mRNAs, whose translation is inhibited by unstable antisense RNAs. Non-translated RNAs exert regulatory functions via binding to complementary sequences. Most antisense RNAs have been identified in prokaryotic cells, mainly in their plasmids, transposons, and bacteriophages. Transient structures in RNA can be functionally important, as the final structure of an RNA often depends on a specific folding pathway determined by RNA itself.

The transient structures in RNA play a very important key role in the *hok-sok* system of plasmid R1. The locus encodes a very stable mRNA, which specifies the toxic Hok (host-killing) protein that can kill cells by damaging their membranes. Translation of *hok* is regulated by *sok* (suppression of killing) RNA, an unstable antisense RNA of 63 nucleotides complementary to the *hok* mRNA leader [105–107].

The initiation frequency of plasmid R1 is controlled by antisense RNA (*CopA*) which binds to the leader region of the *repA* mRNA (*CopT*). The *CopA-CopT* association process is woven as a series of reactions leading to progressively more stable complexes. *CopA* and *CopT* are fully complementary and both RNAs contain a major stem-loop structure that is essential for high pairing rates. The initial step involves a transient loop-loop interaction (kissing complex) between complementary hairpin loops [108, 109].

In the plasmids of the ColE1 family, control of replication is mediated by an antisense RNA, *RNAI*, that interacts with the preprimer, *RNAII*, via initial and transient base pairing between complementary loops [110, 111]. Crystallography showed that the stem-loop structures had melted to a duplex. In the stem-loop conformation, the RNA oligos bind the plasmid-encoded four-helix bundle protein rop [112].

Very similar interactions are observed in other plasmids: *pnd-pndB* of R483 [107] and *RNA-IN/RNA-OUT* of IS10 [113].

Abiotic stress signals

Oxidative stress response RNAs (*gadd7/adapt15, adapt33*)

The oxidative stress caused by exposure to UV radiation or reactive oxygen species is responsible for significant damage in biological cells. The effects of stress involve the inactivation of enzymes and transport proteins, peroxidation of lipids, DNA damage including mutations, as well as cleavage of cellular macromolecules. Oxidative stress has been shown to be responsible for several human disorders including cataract, arteriosclerosis, cancer, and aging. Although much is known about the chemical and biochemical consequences of oxidant exposure, little information is available about its effect on gene expression. In bacteria, a number of genes are involved in regulation by oxidative stress. These include catalase, superoxide dismutase, and glutathione reductase, which are involved in detoxification of cells. In mammals, several groups of genes are expressed in response to UV irradiation or the exposure to hydrogen peroxide and other reactive oxygen species. However, transcripts of some of these genes apparently lack protein products and likely act as riboregulators.

One them, *gadd7*, is expressed in response to treatment with UV radiation. It belongs to the family of *gadd* genes induced by various types of growth arrest signals and by DNA damage. The *gadd7* transcript is a 0.9-kb-long polyadenylated RNA species, lacking a long ORF [114]. Sequence analysis showed that there might be three short ORFs for 38-, 37- and 43-amino-acid long peptides. However, in vitro translation of the *gadd7* transcript did not reveal any protein product. *gadd7* RNA may play its function in the regulation of other genes following DNA damage [114].

RNA species produced in response to oxidative stress induced by hydrogen peroxide are *adapt15* and *adapt33*. Their expression is correlated with the adaptive response to H_2O_2 [115]. Transcription products of both these genes lack long ORFs and are polyadenylated. The *adapt15* RNA is 0.95 kb long and is almost identical to *gadd7* RNA, whereas for *adapt33*, two homologous RNA species of 1.46 and 0.99 kb have been isolated [116]. *gadd/adapt* RNAs known to date have been isolated from Chinese hamster cells and include *gadd7*, *adapt15* and *adapt33* (two species). v-src end-associated peptide 1 RNA (*vseap1*) shows high similarity to *gadd7/adapt15* RNA. This RNA is translated and a peptide is formed. Its specific function is not known [117].

Heat shock response RNA (*hsrω*)

Protection against environmental stress is mainly conferred by the induction of heat shock genes. This mechanism is common in both prokaryotic and eukaryotic organisms.

The amino acid sequences of heat-shock-induced proteins isolated from a variety of organisms are highly conserved. In *Drosophila*, a major site of transcription in temperature stress is the *hsrω* gene. It is located in the polytene region 93D. An interesting feature of this domain is that it is induced independently by the rest of the heat shock genes encoding all major groups of heat shock proteins [118]. The product of this gene is a polyadenylated and spliced RNA with very short ORFs. It shows very little conservation among different species. The expression of *hsrω* is constitutive and elevated by heat shock. Most of the *hsrω* transcripts are located in the nucleus [119]. In both normal and stressed cells, transcription from the *hsrω* locus gives rise to three transcripts – *ω1*, *ω2*, and *ω3*. The *hsrω1* transcript is about 10 kb long and results from transcription of the entire locus. This RNA is present at very high levels in the nucleus. At the 3′-end of the *hsrω1* transcript there is a 7 to 8-kb-long region consisting of a short tandem repeat unit. The *hsrω2* (about 1.9 kb) RNA also accumulates in the nucleus and results from the use of an alternative polyadenylation site located just upstream of the tandem repeat region. The *ω3* transcript (1.2–1.3 kb) is a cytoplasmic species produced from *ω2* by removal of a single intron. This RNA does not contain any long ORF. One of the short ORFs that is conserved in three *Drosophila* species would encode a 23- to 27-mino-acid-long peptide. This RNA likely does not act as a message for protein synthesis but performs other functions as regulatory RNA. However, no accumulation of a protein product was observed but some results suggest that the short ORF is translated. The act of translation alone and not the generation of a functional protein product may be important [120, 121]. The sequences of *hsrω* genes have been determined for *Drosophila melanogaster, D. hydei,* and *D. pseudoobscura*.

G90 RNA

The murine *G90* gene has been identified by substractive hybridization based on the differential presence of its transcript in large and small intestine. The *G90* gene was mapped to a region of mouse chromosome 6. G90 is transcribed at very high levels in the small intestine and at lower levels in large intestine, testis, and kidney of the mouse [122]. The full-length cDNA produces a 1.5-kb transcript that is polyadenylated but has no ORF larger than 249 bp. Its highly specific expression pattern might suggest that *G90* is involved in cell cycle arrest in intestinal epithelium and perhaps in testis. Another possibility is that *G90* is just a marker for cellular differentiation [122].

OxyS RNA

During exponential growth, the expression of several hydrogen-peroxide-inducible proteins in *Escherichia coli* is controlled by the transcriptional regulator, OxyR [123, 124]. This induces the expression of an abundant 109-nucleotide-long untranslated regulatory *OxyS* RNA. This RNA acts as a global regulator by affecting the expression of as many as 40 genes in *E. coli*, which coordinate adaptation to hydrogen peroxide with other cellular stress responses and help to protect cells against oxidative damage. *OxyS* inhibits the translation of fhlA, the transcriptional activator for formate metabolism, by an antisense mechanism. *OxyS* contacts two separate target RNA regions more than 40 nucleotides apart and both interactions are important for *OxyS*-target binding as well as for blocking ribosome access to the *fhlA* translation initiation site [125, 126]. A main target of *OxyS* is an mRNA encoding the σ_s subunit of RNA polymerase. *OxyS* RNA is one of the elements in the cascade of factors that modulate the stability of a region of σ_s-encoding mRNA. *OxyS* RNA can pair with a short sequence overlapping the Shine-Dalgarno sequence in that mRNA [127–131].

DsrA RNA

Regulation of capsular polysaccharide synthesis in *E. coli* K-12 depends on the level of the unstable positive regulator RcsA. The amount of RcsA protein is small because of its rapid degradation and low synthesis level due to transcriptional silencing by the histone-like protein H-NS. A small 85-nucleotide-long *DsrA* RNA activates *RcsA* transcription by counteracting H-NS silencing [132–136]. *DsrA* contains regions of sequence complementary to at least five different genes: *hns, argR, ilvIH, rpoS,* and *rbsD,* but acts in trans by RNA-RNA interactions with only two mRNAs: *hns* and *rpoS.* H-NS is a major nucleoid-structuring, histone-like protein responsible for silencing a number of bacterial genes, RpoS(G^s) in the stationary phase, and the stress response sigma-factor of RNA polymerase. *DsrA* antagonizes HN-S function by decreasing the level of H-NS protein in the cell and increasing RpoS especially at low temperature. Thus *DsrA* has opposite effects on these two targets, both mediated by RNA, with global regulatory consequences for the transcriptional state of the cell. Whereas the mechanism of *DsrA* action at *Hns* is not known, *DsrA* binds the operator of *rpoS* to open a stable stem-loop of *rps* RNA [132–136] enabling access to the Shine-Dalgarno sequence and thus enhancing translation. Structure predictions using thermodynamic calculations suggest that *DsrA* consists of three stem-loop structures, the last being the transcription terminator of *DsrA* [132]. The *hns* complementary region in the center of the molecule resides within the predicted second stem-loop but the *rpoS* region occupies the predicted first stem-loop and the base of the second stem [135, 136]. Footprinting in the presence or absence of *hns* RNA in vitro shows a distinct *DsrA-rpoS* interaction involving stem-loop 1, whereas a

two-parts *DsrA-hns* interaction involves stem-loop 2. There is base pairing between *DsrA* and two regions of *hns*, near the 5′ and 3′ ends of its coding region, resulting in a contiguous coaxial stack, looping out in the middle part of *hns* mRNA exposed to nucleases. These interactions are proposed to circularize *hns* mRNA and define the structural basis of *DsrA* activity at *hns*. On the other hand, the structure of stem-loop 2 of *DsrA* remains intact while base paired to *rpoS* mRNA, stabilizing *rpoS* and reducing mRNA turnover [137].

DD3/PCGEM1 RNA

The prostate-specific genes, *DD3* and *PCGEM1*, are specifically expressed in human prostate tissue. The *PCGEM1* locus has been mapped to chromosome 2q32. cDNA sequence analysis revealed that *PCGEM1* lacks protein-coding capacity and suggests that it may belong to the ncRNAs. There are two RNAs which have highly prostate specific expression in cancer versus normal prostate [138, 140].
Another gene, *DD3* maps to chromosome 9q21–22. No extensive ORF has been found, and the transcript may function as a ncRNA [139, 140].

Biotic stress signals

His-1 RNA

His-1 is a single-copy gene that has been found in a variety of vertebrate species. Upon viral insertion in murine myeloid leukemias, the *His-1* gene is activated. It is not expressed in uninfected cells. *His-1* RNA is expressed as a 3-kb-long, spliced and polyadenylated RNA. An analysis of the RNA sequence did not reveal any ORF that would exceed 219 nucleotides (73 amino acids). The lack of a long ORF suggests that it can function as a riboregulator. *His-1* is expressed at low levels in the epithelial cells of the adult mouse stomach, prostate, seminal vesicle, and developing choroid plexus. The presence of *His-1* RNA is correlated with a viral insertion and carcinogenesis, since no transcripts were detected in normal tissues. They can be readily identified in mouse leukemias and carcinomas. This finding suggests that expression of the *His-1* gene is highly restricted and that its inappropriate activation may contribute to carcinogenesis [141–143]. *His-1* RNAs known to date are those from human and mouse.

Early nodulin 40 (*ENOD40*)

Genes that are activated in plants by Nod (nodulin) factors are called nodulins. In the symbiosis between rhizobia and legumes, nitrogen-fixing nodules are formed the an outcome of a complex process that includes new organ development, microbial invasion of plant tissues, internalization of bacteria in plant cells, and functional differentiation of the two partners. During nodule development, several plant genes are expressed in subsequent stages. Genes that are transcribed early in the interaction (*ENOD* genes) seem to play a role in organogenesis and bacterial invasion. Several *ENOD* clones have been isolated, but their individual contribution to nodule formation is often unclear. Some nodulins have been employed as early markers to study the initiation of nodulation [144–146]. One of them, *ENOD40*, an early nodulin gene, is expressed following inoculation with *Rhizobium meliloti* or by adding *R. meliloti*-produced Nod factors or the plant hormone cytokinin to non-inoculated roots [147, 148]. It is detectable in the root pericycle opposite the nodule primordium even before the appearance of infection threads, and is also found later, associated with vascular strands in mature nodules. Comparison of the *ENOD40* sequence isolated from several legumes and one non-legume did not reveal any conserved large ORF. Instead, a conserved region, which might allow the production of stable cytoplasmic RNA was found. *ENOD40* has therefore been proposed to encode RNA with a regulatory function. The different *ENOD40* RNAs have only one short ORF for peptides 10–13 amino acids long, and the 3′ end region appears to have important functions as a riboregulator [149–152].
ENOD40 RNAs known to date are those from *Glycine max*, *Pisum sativum*, *Phaseolus vulgaris*, *Vicia sativa*, *Trifolium repens*, *Lotus japonicus*, *Medicago sativa*, *M. truncatula*, *Nicotiana tabacum*, *Vigna radiata,* and *Sesbania rostrata*.

lbi RNA

The *lbi* (lipopolysaccharide biosynthesis interfering) RNA of phage Acm1, a non-translated RNA species of 97 nucleotides, was shown to affect the biosynthesis of the O-specific polysaccharide of lipopolysaccharide in various *E. coli* strains and to down-regulate the synthesis of the D-glucan component of the O-specific polysaccharide in *Klebsiella pneumoniae* serotype 01. *lbi* RNA consists of two consecutive stem-loop structures. The 5′-proximal hairpin loop function is a key structural element in the mechanism leading to the inhibition of D-glucan biosynthesis due to its antisense interactions with cellular target RNAs [153, 154].

Cytokinin response RNA (*CR20*)

CR20 is one of several genes repressed by cytokinins in excised cotyledons of cucumber. Detailed analysis showed that there are several *CR20* transcripts generated by alternative splicing of precursor RNA. However, none of the isolated and sequenced *CR20* transcripts contained a

long ORF. This suggests that these RNAs are not translated into a protein and may function as riboregulators. A comparison of the two known *CR20* sequences from cucumber and *Arabidopsis thaliana* revealed that although they show little overall homology, there is a highly conserved 180-nt region that seems to form a stable secondary structure [155].

GUT15 RNA

The majority of transcripts in eukaryotic cells are stable, with half-lives in the order of hours. In contrast to these relatively stable mRNAs, a subset of transcripts is rapidly degraded, with half-lives of the order of minutes. The tobacco genes with unstable transcripts (*GUT*) have been isolated and characterized. *GUT15* RNA has a polyA tail and one intron, but does not contain an ORF. It has a long segment similar to that of *CR20* [156, 157]. Similar RNAs may also be present in aspen, poplar, soybean, rice and maize [Ambro van Hoof, personal communication].

Other functions of ncRNAs

Brain-specific RNAs (*Bsr, BC1, BC200* RNA)

Brain-specific repetitive RNA (*Bsr* RNA) is an ncRNA that consists of tandem repeats of similar sequences, approximately 0.9 kb in length. The isolated clones may contain one or more repeats. In rat, *Bsr* RNA is preferentially expressed in the central nervous system [158].
BC1 RNA is generated by retroposition of $tRNA^{Ala}$. It is a 152-nucleotides-long transcript expressed preferentially in the brain by polymerase III. The sequence similarity between $tRNA^{Ala}$ and the 5′ half of *BC1* RNA is approximately 80%. However, this domain does not fold in a tRNA-like manner, but forms a stable stem-loop structure. The level of this RNA increases during the period of synaptogenesis [159, 160].
BC200 RNA is a small cytoplasmic ncRNA identified in a number of primate species. Its expression is almost exclusively restricted to neural tissues. In the cell, it is present as a ribonucleoprotein particle. The gene encoding *BC200* arose from a monomeric Alu element, and the RNA it encodes has been recruited into a function in the nervous system [160–162]. The 5′ domain consists of about 120 nt with identity to monomeric Alu element. *BC200* RNA has been identified in all *Anthropoidea* [163].

SRA RNA

SRA RNA, a steroid receptor RNA activator, is selective for steroid hormone receptors and mediates transactivation via its amino-terminal domain. The transcript is a 0.7- to 0.85-kb RNA. It is a *bona fide* transcriptional co-activator selective for the AF1 domain of steroid receptors. It is expressed as multiple isoforms in a cell-specific manner and present in a steady-state co-regulator complex with the AF2 co-activator SRC-1 [164].

meiRNA

Fission yeast protein Mei2 is an RNA-binding protein required for both premeiotic DNA synthesis and meiosis I. It binds to polyadenylated *meiRNA* of about 0.5 kb, the loss of which blocks meiosis I, and is required for nuclear localization of Mei2 [165–167].

UHG RNAs

Small nucleolar RNAs (snoRNAs) are essential for maturation of rRNA. They are encoded within different introns of the unusually compact mammalian *U22* host gene (*UHG*). *U22* RNA is essential for the maturation of 18S rRNA. The *U22* snoRNA gene lies within an intron of the single-copy gene *UHG* that encodes a polyadenylated ncRNA. *UHG* RNAs in human and mouse are 1114 and 590 nucleotides, respectively [168]. Recently, three other non-protein-coding snoRNA host genes, *Gas5* [169], *U17HG* [170], and *U19HG* [171], have been identified. *Gas5* spliced RNA becomes polyadenylated and associates with translating ribosomes [169]. The transcript of *U17HG* is enriched in the cytoplasm but does not associate with translating ribosomes [170]. Both the *Gas5* and *H17HG* transcripts initiate with a C nucleotide followed by a pyrimidine-rich tract and belong to the 5′-terminal oligopyrimidine gene family (5′-TOP) [172]. *U19HG* remains in the nucleus after splicing and has a 5′ sequence reminiscent of 5′-TOP [171].

Xlsirt RNA

Xlsirt RNAs have been identified as RNA species localized to the vegetal cortex of *Xenopus* oocytes during early stages of oogenesis. They have been proposed to play a structural role and to be used to localize other RNAs such as *Vg1* mRNA. *Xlsirt* RNAs form a family of heterogenous molecules originating from both strands of the genes. Those from one strand (sense) are localized in the vegetal cortex, while others are found throughout the cytoplasm. The transcripts contain 3–13 repeat units that are flanked by unique sequences [173, 174].

Conclusion

To date, more than 100 ncRNAs have been identified, some of which are partly characterized. They differ in origin, function size and structure. None have long ORFs and toghether form a collection of stable RNA molecules

involved in many cellular regulatory pathways. These ncRNAs are developmentally regulated or show highly restricted patterns of gene expression. They provide important insights into RNA-based mechanisms of gene expression, genomic imprinting, cell cycle progression, and differentiation. Some ncRNAs influence complex neurobehavioral phenotypes including psychiatric disorders [175]. In many cases, this is done by interaction with 3′-untranslated regions of mRNA [176]. Data exist that some ncRNAs are integral components of the 'compensasome' involved in dosage compensation and genomic imprinting. These RNAs might also function as chromatin remodeling factors. Some of the RNAs resemble very closely those involved in gene silencing mediated by double-stranded RNA (dsRNA), called interference RNA (RNAi) [177]. dsRNA is cleaved into short RNAs of 21–23 nucleotides which mediate RNA degradation of the complementary sequences [178, 179]. Although RNAi acts by a different mechanism, the structural similarity to some ncRNAs is striking.
The interesting question is why so many different noncoding regulatory RNAs with various functions occur within the cell. The answer may lie in the amazing variety of structures, mechanisms, and biological functions acquired by RNAs that provide the basis to use them in various regulatory processes. Many small RNAs require a protein component for their activity as in the case of *tmRNA* [180], *M1* RNA of RNase P [181] and the RNA chaperone Hfg (*HF1*) [182].
The energetic cost of synthesis of small RNAs may be much lower than for a protein, for which mRNA has to be synthesized first, followed by translation into a chain of amino acids [125]. However, this argument is not quite so convincing for large RNAs such as *Xist* or *H19*.
Finally, there is the general question concerning finding and analyzing ncRNAs in the era of genomics. They are not easy identifiable, and in most cases do not shown conserved motifs or signals [183]. New structural data are urgently needed. They will form a basis to improve our understanding of the mechanisms of ncRNA action.

Acknowledgements. This work was supported by the Polish Committee for Scientific Research (KBN) and by the DFG Trilateral Research Project 'Genomic Imprinting in Human Bladder Cancer'.

1 Erdmann V. A., Barciszewska M. Z., Szymanski M., Hochberg A., Groot N. de and Barciszewski J. (2001) The noncoding RNAs as riboregulators. Nucleic Acids Res. **29:** 189–193
2 Eddy S. R. (1999) Non-coding RNA genes. Curr. Opin. Genet. Dev. **9:** 695–699
3 Reik W. and Walter J. (2001) Genomic imprinting: parental influence on the genome. Nat. Rev. Genet. **2:** 21–32
4 Wassenegge M. (2000) RNA-directed DNA methylation. Plant Mol. Biol. **43:** 203–220
5 John R. M. and Surani M. A. (2000) Genomic imprinting, mammalian evolution and the mystery of egg-laying mammals. Cell **101:** 585–588
6 Smilinich N. J., Day C. D., Fitzpatrick G. V., Caldwell G. M., Lossie A. C., Cooper P. R. 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
7 Johnston C. M., Nesterova T. B., Formstone E. J., Newall A. E., Duthie S. M., Sheardown S. A. et al. (1998) Developmentally regulated *Xist* promoter switch mediates initiation of X-inactivation. Cell **94:** 809–817
8 Marahrens Y. (1999) X-inactivation by chromosomal pairing events. Genet. Dev. **13:** 2624–2632
9 Panning B. and Jaenisch R. (1998) RNA and the epigenetic regulation of X-chromosome inactivation. Cell **93:** 305–308
10 Askew D.S. and Xu F. (1999) New insight into the function of noncoding RNA and its potential role in disease pathogenesis. Histol. Histopathol. **14:** 235–245
11 Lee J. T. (2000) Disruption of imprinted X inactivation by parent-of-origin effects at Tsix. Cell **103:** 17–27
12 Kelley R. L. and Kuroda M. I. (2000) The role of chromosomal RNAs in marking the X for dosage compensation. Curr. Opin. Genet. Dev. **10:** 555–561
13 Kelley R. L. and Kuroda M. I. (2000) Noncoding RNA genes in dosage compensation and imprinting. Cell **103:** 9–12
14 Avner P. and Heard E. (2001) X-chromosome inactivation: counting, choice and initiation. Nat. Rev. Genet. **2:** 56–67
15 Hong Y.-K., Ontiveros S. D. and Strauss W. M. (2000) A revision of the human XIST gene organization and structural comparison with mouse *Xist*. Mamm. Genome **11:** 220–224
16 Hong Y.-K., Ontiveros S. D., Chen C. and Strauss W. M. (2000) A new structural for the murine *Xist* gene and its relationship to chromosome choice/counting during X-chromosome inactivation. Proc. Natl. Acad. Sci. USA **96:** 6829–6834
17 Clerc P. and Avner P. (1998) Role of the region 3′ to *Xist* exon 6 in the counting process of X-chromosome inactivation. Nat. Genet. **19:** 249–253
18 Meller V. H. (2000) Dosage compensation: making 1X equal 2X. Trends Cell Biol. **10:** 54–59
19 Gilbert S. L. and Sharp P. A. (2000) Promoter-specific hypoacetylation of X inactivated genes. Proc. Natl. Acad. Sci. USA **96:** 13825–13830
20 Schlissel M. S. (2000) A tail of histone acetylation and DNA recombination. Science **287:** 438-439
21 Chenong W. L. Briggs S. D. and Allis C. D. (2000) Acetylation and chromosomal functions. Curr. Opin. Cell Biol. **12:** 326–333
22 Warshawsky D., Stavropoulos N. and Lee J. T. (1999) Further examination of the *Xist* promoter-switch hypothesis in X inactivation: evidence against the existence and function of a P(0) promoter. Proc. Natl. Acad. Sci. USA **96:** 14424–14429
23 Endo Y., Watanabe T., Mishima Y., Yoshimura A., Takagi N. and Kominami R. (1999) Compact chromatin packaging of inactive X chromosome involves the actively transcribed *Xist* gene. Mamm. Genome **10:** 606–610
24 Lyon M. F. (2000) LINE-1 elements and X chromosome inactivation: a function for 'junk' DNA? Proc. Natl. Acad. Sci. USA **97:** 6248–6249
25 Lyon M. F. (1998) X-chromosome inactivation: a repeat hypothesis. Cytogenet. Cell Genet. **80:** 133–137
26 Sasaki H., Ishihara K. and Kato R. (2000) Mechanisms of *Igf2/H19* imprinting: DNA methylation, chromatin and long-distance gene regulation. J. Biochem. **127:** 711–715
27 Bailey J. A., Carrel L., Chakravarti A. and Eichler E. E. (2000) Molecular evidence for a relationship between LINE-1 elements and X-chromosome inactivation: the Lyon repeat hypothesis. Proc. Natl. Acad. Sci. USA **97:** 6634–6639
28 Weiner A. M. (2000) Do all SINEs lead to LINEs? Nat. Genet. **24:** 332–333
29 Surralles J., Hande M. P., Marcos R. and Lansdorp P. M. (1999) Accelerated telomere shortening in the human inactive X chromosome. Am. J. Hum. Genet. **65:** 1617–1622

30 Lee J. T., Davidov L. S. and Varshawsky D. (1999) *Tsix*, a gene antisense to *Xist* at the X-inactivation centre. Nat. Genet. **21:** 400–404
31 Mise N., Goto Y., Nakajima N. and Takagi N. (1999) Molecular cloning of antisense transcript of the mouse *Xist* gene. Biochem. Biophys. Res. Commun. **258:** 537–541
32 Lee J. T. and Lu N. (1999) Targeted mutagenesis of *Tsix* leads to nonrandom X inactivation. Cell **99:** 47–57
33 McCabe V., Formstone E. J., O'Neill L. P., Turner B. M. and Brockdorff N. (1999) Chromatin structure analysis of the mouse *Xist* locus. Proc. Natl. Acad. Sci USA **96:** 7155–7160
34 Duthie S. M., Nesterova T. B., Formstone E. J., Keohane A. M. and Turner B. M., Zakian S. M. et al. (1999) *Xist* RNA exhibits a banded localisation on the inactive X-chromosome and is excluded from autosomal material in cis. Hum. Mol. Genet. **8:** 195–204
35 Keohane A. M., Barlow A. L., Waters J., Bourn D. and Turner, B. M. (1999) H4 acetylation, *Xist* RNA and replication timing are coincident and define X; autosome boundaries in two abnormal X-chromosomes. Hum. Mol. Genet. **18:** 377–383
36 O'Neill L. P., Keokhane A. M., Lavender J. S., McCabe V., Heard E., Avner P. et al. (1999) A developmental switch in H4 acetylation upstream of *Xist* plays a role in X-chromosome inactivation. EMBO J. **18:** 2897–2907
37 Lee J.T. (2000) Disruption of imprinted X inactivation by parent-of-origin effects at Tsix. Cell **103:** 17–27
38 Kelley R. L., Meller V. H., Gordadze P. R., Roman, G., Davis R. L. and Kuroda M. I. (1999) Epigenetic spreading of the *Drosophila* dosage compensation complex from *roX* RNA genes into flanking chromatin. Cell **98:** 513–522
39 Akhtar A., Zink D. and Becker P. B. (2000) Chromodomains are protein-RNA interaction modules. Nat. **407:** 405–408
40 Moulton C. M., Chow J. C., Brown C. J., and Lawrence J. B. (1998) Stabilisation and localisation of *Xist* RNA are controlled by separate mechanisms and are not sufficient for X-inactivation. J. Cell Biol. **142:** 13–23
41 Franke A. and Baker B. S. (1999) The *rox1* and *rox2* RNAs are essential components of the compensasome, which mediates dosage compensation in *Drosophila*. Mol. Cell **4:** 117–122
42 Franke A. and Baker B. S. (2000) Dosage compensation rox. Curr. Opin. Cell Biol. **12:** 351–354
43 Meller V. H., Gordadze P. R., Park Y., Chu X., Stuckenholtz C., Kelly R. L. and Kuroda M. I. (2000) Ordered assembly of *roX* RNAs into MSL complexes on the dosage-compensated X chromosome in *Drosophila.* Curr. Biol. **10:** 136–143
44 Chao Y.-C., Lee S-.T., Chang M.-C., Chen H.-H., Chen S.-S., Wu, T.-Y. et al. (1998) A 2.9 kilobase non-coding nuclear RNA functions in the establishment of persistent Hz-1 viral infection. J. Virol. **72:** 2233–2245
45 Tilgman S. M. (1999) The sins of the fathers and mothers: genomic imprinting in mammalian development. Cell **96:** 185–193
46 Brannan C. I. and Bartolomei M. (1999) Mechanisms of genomic imprinting. Curr. Opin. Genet. Dev. **9:** 164–170
47 Adriaenssens E., Dumont L., Lottin S., Bolle D., Lepretre A., Delobelle A. et al. (1998) *H19* over-expression in breast adenocarcinoma stromal cells is associated with tumor values and steroid receptor status but independent of p53 and Ki-67 expression. Am. J. Pathol. **153:** 1597–1607
48 Ariel I., Ayesh S., Perlman E. J., Pizov G., Tanos V., Schneider T. et al. (1997) The product of the imprinted H19 gene is an oncofetal RNA. Molec. Pathol. **50:** 34–44
49 Ohana P., Kopf E., Bibi O., Ayesh S., Schneider T., Laster M. et al. (1999) The expression of the *H19* gene and its function in human bladder carcinoma cell lines. FEBS Lett. **454:** 81–84
50 TanosV., Prus D., Ayesh S., Weinstein D., Tykocinski M. L., DeGroot N. et al. (1999) Expression of the imprinted *H19* oncofetal RNA in epithelial ovarian cancer. Eur. J. Obstet. Gynecol. Reprod. Biol. **85:** 7–11
51 Ariel I., de Groot N. and Hochberg A. (2000) Imprinted *H19* gene expression in embryogenesis and human cancer: the oncofetal connection. Am. J. Med. Genet. **91:** 46–50
52 Kelsey G. (2000) The hows and whys of imprinting. Trends Genet. **16:** 15–16
53 Frevel M. A. E., Hornberg J. J. and Reeve A. E. (1999) A potential imprint control element identification of conserved 42 bp sequence upstream of *H19*. Trends Genet. **15:** 216–218
54 Kanduri C., Holmgren C., Pilartz M., Franklin G., Kanduri M., Liu L. et al. (2000) The 5′ flank of mouse H19 in an unusual chromatin conformation undirectionally blocks enhancer-promoter communication. Curr. Biol. **10:** 449–457
55 Stadnick M. P., Pieracci F. M., Cranston M. J., Taksel E., Thorvaldsen J. L. and Bartolomei M. S. (1999) Role of a 461-bp G-rich repetitive element in *H19* transgene imprinting. Dev. Genes Evol. **209:** 239–248
56 Kerjean A., Dupont J-M., Vasseur C., Le Tessier D., Cuisset L., Paldi A. et al. (2000) Establishment of the paternal methylation imprint of the human H19 and MEST/PEG1 genes during spermatogenesis. Hum. Mol. Genet. **9:** 2183–2187
57 Hurst L. D. and Smith N. G. C. (1999) Molecular evolutionary evidence that *H19* mRNA is functional. Trends Genet. **15:** 134–135
58 Runge S., Nielsen F.C., Nielsen J., Lykke-Andersen J., Wewer U. M. and Christiansen J. (2000) H19 RNA binds four molecules of IGF-II mRNA-binding protein. J. Biol. Chem**. 275:** 29562–29569
59 Juan V., Crain Ch. and Wilson Ch. (2000) Evidence for evolutionarily conserved secondary structure in the *H19* tumor suppressor RNA. Nucleic. Acids Res. **28:** 1221–1227
60 Li Y. M., Franklin G., Cin H. M., Svenson K., He X. B., Adam G. et al. (1998) The H19 transcript is associated with polysomes and may regulate IGF-2 expression in trans. J. Biol. Chem. **273:** 28247–28252
61 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
62 Bell A.C. and Felsenfeld G. (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. Nat. **405:** 482–485
63 Hark A. T., Schoenherr C. J., Katz D. J., Ingram R. S., Levorse J. M. and Tilgman S. M. (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. Nat. **405:** 486–489
64 Thorvaldsen J. L. and Bartolomei M. S. (2000) Mothers setting boundaries. Science **288:** 2145–2146
65 Wilkin F., Paquette J., Ledru E., Mamelin C., Pollak M. and Deal C. L. (2000) H19 sense and antisense transgenes modify insulin like growth factor-II mRNA levels. Eur. J. Biochem. **267:** 4020–4027
66 Srivastava M., Hsieh S., Grinberg A., Williams-Simons L., Huang S. P. and Pfeifer K. (2000) *H19* and *Igf2* monoallelic expression is regulated in two distinct ways by a shared *cis* acting regulatory region upstream of *H19*. Genes Dev. **14:** 1186–1195
67 Reik W. and Murrell A. (2000) Science across the border. Nat. **405:** 408–409
68 Wolffe A. P. (2000) Transcriptional control: imprinting insulation. Curr. Biol. **10:** R463–R465
69 Kanduri C., Pant V., Loukinov D., Pugacheva E., Qi C.-F., Wolffe A. et al. (2000) Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation sensitive. Curr. Biol. **10:** 853–856
70 Allshire R. and Bickmore W. (2000) Pausing for thought on the boundaries of imprinting. Cell **102:** 705–708
71 Adriaenssens E., Lottin S., Dugimont T., Fauquette W., Coll J., Dupouy J. P. et al. (1999) Steroid hormones modulate H19 gene expression in both mammary gland and uterus. Oncogene **18:** 4460–4473

72 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
73 Li A. Y., Torchia B. S., Migeon B. R. and Siliciano R. F. (1997) The human NTT gene: identification of a novel 17 kb noncoding nuclear RNA expressed in activated CD+ T cells. Genomics **39:** 171–184
74 Southerland H. F., Wadey R., McKie J. M., Taylor C., Atif U., Johnstone K. A. et al. (1996) Identification of a novel transcript disrupted by a balanced translocation associated with DiGeorge syndrome. Am. J. Hum. Genet. **59:** 23–31
75 Southerland H. F., Kim U. J. and Scambler P. J. (1998) Cloning and comparative mapping of the DiGeorge syndrome critical region in the mouse. Genomic **52:** 37–43
76 Sperandeo M. P., Ungaro P., Vernucci M., Pedone P. V., Cerrato F., Perone L. et al. (2000) Relaxation of insulin-like growth factor 2 imprinting and discordant methylation at *KvDMR1* in two first cousins affected by Beckwith-Wiedemann and Klippel-Trenaunay-Weber syndromes. Am. J. Hum. Genet. **66:** 841–847
77 Mitsuya K., Meguro M., Lee M. P., Katoh M., Schulz T. C., Kugoh H. 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
78 Lee M. P., DeBaun M. R., Mitsuya K., Galonek H. L., Brandenburg S., Oshimura M. et al. (1999) Loss of imprinting of a paternally expressed transcript, with antisense orientation to *KVLQT1*, occurs frequently in Beckwith-Wiedemann syndrome and is independent of insulin-like growth factor II imprinting. Proc. Natl. Acad. Sci.USA **96:** 5203–5208
79 Wroe S. F., Kelsey G., Skinner J. A., Bodle D., Ball S. T., Beechey C. V. 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
80 Hayward B. E. and Bonthron D. T. (2000) An imprinted antisense transcript at the human GNAS1 locus. Hum. Mol. Genet. **9:** 835–841
81 Peters J., Wroe S. F., Wells C. A., Miller H. J., Bodle D., Beechey C. V. 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
82 Nemes J. P., Benzow K. A. and Koob M. D. (2000) The SCA8 transcript is an antisense RNA to a brain-specific transcript encoding a novel actin-binding protein. Hum. Mol. Genet. **9:** 1543–1551
83 Robinson D. N and Cooley L. (1997) *Drosophila* kelch is an oligomeric ring canal actin organizer. J. Cell Biol. **138:** 799–810
84 Moseley M. L., Schut L. J., Bird T. D., Koob M. D., Day J. W. and Ranum L. P .W. (2000) SCA8 CTG repeat: en masse contractions in sperm and intergenerational sequence changes may play a role in reduced penetrance. Hum. Mol. Genet. **9:** 2125–2130
85 Ninomiya S., Isomura M., Narahara K., Seino Y. and Nakamara Y. (1996) Isolation of a testis-specific cDNA on chromosome 17q from a region adjacent to the breakpoint of t(12,17) observed in a patient with acampomelic campomelic dysplasia and sex reversal. Hum. Mol. Genet. **5:** 69–72
86 Reinhart B. J., Slack F. J., Basson M., Pasquinelli A. E., Bettinger J. C., Rougvie A. E. et al. (2000) The 21-nucleotide *let-7 RNA* regulates developmental timing in *Caenorhabditis elegans.* Nat. **403:** 901–906
87 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
88 Ambros V. (2000) Control of developmental timing in *Caenorhabditis elegans.* Curr. Opin. Genet. Dev. **10:** 428–433
89 Slack F. J., Basson M., Liu Z., Ambros V., Horvitz H. R. and Ruvkun G. (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
90 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
91 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
92 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
93 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
94 Moss E. G. (2000) Non-coding RNAs: lightning strikes twice. Curr. Biol. **10:** R436–439
95 Adoutte A. (2000) Small but mighty timekeepers. Nat. **408:** 37–38
96 Pasquinelli A. E., Reinhart B. J., Slack F., Martindale M. Q., Kuroda M. I., Maller B. et al. (2000) Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA. Nat. **408:** 86–89
97 Crucs S., Chatterjee S. and Gavis E. R. (2000) Overlapping but distinct RNA elements control repression and activation of *nanos* translation. Mol. Cell **5:** 457–467
98 Lalande M., Minnassian B. A., DeLorey T. M. and Olsen R. W. (1999) Parental imprinting and Angelman syndrome. Adv. Neurol. **79:** 421–429
99 Rougeulle C., Cardoso C., Fontez M., Colleaux L. and Lalande M. (1998) An imprinted antisense RNA overlaps UDE3A and a second maternally expressed transcript. Nat. Genet. **19:** 15–16
100 Jong M. T., Carey A. H., Caldwell K. A., Lau M. H., Handel M. A., Driscoll D. J. et al. (1999) Imprinting of a RING zincfinger encoding gene in the mouse chromosome region homologous to the Prader-Willi syndrome genetic region. Hum. Mol. Genet. **8:** 795–803
101 Jong M. T., Gray T. A., Ji Y., Glenn C. C., Saitoh S., Driscoll D. J. et al. (1999) A novel imprinted gene, encoding a RING zinc-finger protein, and overlapping antisense transcript in the Prader-Willi syndrome critical region. Hum. Mol. Genet. **8:** 783–793
102 Millar J. K., Wilson-Annan J. C., Anderson S., Christie S., Taylor M. S., Semple C. A. et al. (2000) Disruption of two novel genes by a translocation co-segregating with schizophrenia. Hum. Mol. Genet. **9:** 1415–1423
103 Swalla B. J. and Jeffery W. R. (1995) A maternal RNA localized in the yellow crescent is segregated to the larval muscle cells during ascidian development. Dev. Biol. **170:** 353–364
104 Swalla B. J. and Jeffery W. R. (1996) PCNA mRNA has a 3′ UTR antisense to yellow crescent RNA and is localized in ascidian eggs and embryos, Dev. Biol. **178:** 23–34
105 Pedersen K. and Gerdes K. (1999) Multiple *hok* genes on the chromosome of *Escherichia coli*. Mol. Microbiol. **32:** 1090–1102
106 Franch T., Thisted T. and Gerdes K. (1999) Ribonuclease III processing of coaxially stacked RNA helices. J. Biol. Chem. **274:** 26572X–26578
107 Nagel J. H., Gultayev A. P., Gerdes K. and Pleij C. W. (1999) Metastable structures and refolding kinetics in hok mRNA of plasmid R1. RNA **5:** 1408–1418

108 Ohman M. and Wagner E. G. H. (1989) Secondary structure analysis of RepA mRNA leader transcript involved in control of plasmid replication. Nucleic Acids Res. **17:** 2557–2579
109 Kolb F. A., Malmgren C., Westhof E., Ehresmann C., Ehresmann B., Wagner E. G. H. et al. (2000) An unsusual structure formed by antisense-target RNA binding involves an extended kissing complex with a four way junction and side-by-side helical alignment. RNA **6:** 311–324
110 Eguchi Y. and Tomizawa J. (1991) Complexes formed by complementary RNA stem-loops: their formations, structures and interactions with ColE1 Rom protein. J. Mol. Biol. **229:** 831–842
111 Lee A. J. and Crothers D. M. (1998) The solution structure of an RNA loop-loop complex: the ColE1 inverted loop sequence. Structure **6:** 993–1005
112 Klosterman P. S., Shah S. A. and Steitz T. A. (1999) Crystal structure of two plasmid copy control related RNA duplexes: an 18 base pair duplex at 1.2 Å resolution and a 19 base pair duplex at 1.55 Å resolution. Biochemistry **38:** 14784–14797
113 Franch T., Petersen M., Wagner E. G., Jacobsen J. P. and Gerdes K. (1999) Antisense regulation in prokaryotes: rapid RNA/RNA interaction facilitated by a general U-turn loop structure. J. Mol. Biol. **294:** 1115–1125
114 Hollander M. C., Alamo I. and Fornace A. J. Jr. (1996) A novel DNA damage-inducible transcript, *gadd7*, inhibits cell growth, but lacks a protein product. Nucleic Acids Res. **24:** 1589–1593
115 Wang Y., Crawford D. R. and Davies K. J. A (1996) *adap33*, a novel oxidant inducible RNA from hamster HA-1 cells. Arch. Biochem. Biophys. **332:** 255–260
116 Crawford D. R., Schools G. P., Salmon S. L. and Davies K. J. A. (1996) Oxidant-inducible *adap*15 RNA is associated with growth arrest- and DNA damage-inducible *gadd*153 and *gadd*45. Arch. Biochem. Biophys. **325:** 256–264
117 Mizenina O., Yanushevich Y., Musatkina E., Rodina A., Camonis J., Tavitian A. et al. (1998) C-terminal end of *v-src* protein interacts with peptide coded by *gadd*7/*adapt*15-like RNA in two-hybrid system, FEBS Lett. **422:** 79–84
118 Fini M. E., Bendena W. G. and Pardue M. L. (1989) Unusual behaviour of the cytoplasmic transcript of *hsrω:* an abundant stress-inducible RNA that is translated but yields no detectable protein product. J. Cell Biol. **108:** 2045–2057
119 Lakhotia S. C. and Sharma A. (1995) The 93 D (hsr-omega) locus of *Drosophila*: non-coding gene with house-keeping functions. Chrom. Res. **3:** 151–161
120 Lakhotia S. C. and Sharma A. (1996) RNA metabolism in situ at the 93D heat shock locus in polytene nuclei of *Drosophila melanogaster* after various treatments. Genetica **97:** 339–348
121 McKechnie S. W., Halford M. M., McColl G. and Hoffmann A. A. (1998) Both allelic variation and expression of nuclear and cytoplasmic transcripts of *Hsr*-omega are closely associated with thermal phenotype in *Drosophila*. Proc. Natl. Acad. Sci. USA **95:** 2423–2428
122 Krause R., Hemberger M., Himmelbauer H., Kalscheuer V. and Fundele R. H. (1999) Identification and characterisation of G90, a novel mouse RNA that lacks an extensive open reading frame. Gene **232:** 35–42
123 Hengge-Aromis R. (1999) Interplay of global regulators and cell physiology in the general response of *Escherichia coli*. Curr. Opin. Microbiol. **2:** 148–152
124 Nogueira T. and Springer M. (2000) Post transcriptional control by global regulators of gene expression in bacteria. Curr. Opin. Microbiol. **3:** 154–158
125 Altuvia S. and Wagner E. G. H. (2000) Switching on and off with RNA. Proc. Natl. Acad. Sci. USA **97:** 9824–9826
126 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
127 Altuvia S., Weinstein-Fischer D., Zhang A., Postow L. and Storz G. (1997) A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. Cell **90:** 43–53
128 Zhang A., Altuvia S., Tiwari A., Argaman L., Hengge-Aronis R. and Storz G. (1998) The *OxyS* regulatory RNA represses *rpoS* translation and binds the Hfg (HF-1) protein. EMBO J. **17:** 6061–6068
129 Altuvia S., Zhang A., Argaman L., Tiwari A. and Storz G. (1998) The *Escherichia coli OxyS* regulatory RNA represses fhlA translation by blocking ribosome binding. EMBO J. **17:** 6069–6075
130 Loewen P.C., Strutinsky J. and Sparling R. (1998) Regulation in *rpoS* regulon of *Escherichia coli*. Can. J. Microbiol. **44:** 707–717
131 Gonzalez-Flecha B. and Demple B. (1999) Role for the *oxyS* gene in regulation of intracellular hydrogen peroxide in *Escherichia coli*. J. Bacteriol. **181:** 3833–3836
132 Sledjeski D. and Gottesman S. (1995) A small RNA acts as an anti silencer of the H-NS-silencer rcsA gene of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **92:** 2003–2007
133 Sledjeski D., Gupta A. and Gottesman S. (1996) The small RNA, DsrA, is essential for the low temperature expression of *RpoS* during exponential growth in *Escherichia coli*. EMBO J. **15:** 3993–4000
134 Klauck E., Bohringer J. and Hengge-Aronis R. (1997) The *LysR*-like regulator *LeuO* in *Escherichia coli* is involved in the translational regulation of rpoS by affecting the expression of the small regulatory DsrA-RNA. Mol. Microbiol. **25:** 559–569
135 Majdalani N., Cunning C., Sledjeski D., Elliott T. and Gottesman S. (1998) DsrA RNA regulates translation of RpoS message by an antisense mechanism, independent of its action as an anti silencer of transcription. Proc. Natl. Acad. Sci. USA **95:** 12462–12467
136 Lease R. A., Cusick M. E. and Belfort M. (1995) Riboregulation in *Escherichia coli:* DsrA RNA acts by RNA:RNA interactions at multiple loci. Proc. Natl. Acad. Sci. USA **95:** 12456–12461
137 Lease R.A. and Belfort M. (2000) 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
138 Srikantan V., Zou Z., Petrovics G., Xu L., Augustus M., Davis L., et al. (2000) PCGEM1, a prostate-specific gene, is overexpressed in prostate cancer. Proc. Natl. Acad. Sci. USA **97:** 12216–12221
139 Bussemakers M. J., Bokhoven A. van, Verhaegh G. W., Smit F. P., Karthaus H. F., Schalken J. A. et al. (1999) DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. Cancer Res. **59:** 5975–5979
140 Verhaegh G. W., van Bokhoven A., Smit F., Schalken J. A. and Bussemakers M. J. (2000) Isolation and characterization of the promoter of the human prostate-cancer-specific DD3 gene. J. Biol. Chem. **275:** 37496–37503
141 Askew D. S., Li J. and Ihle J. N. (1994) Retroviral insertions in the murine *His-1* locus activate the expression of novel RNA that lacks an extensive open reading frame. Mol. Cell Biol. **14:** 1743–1751
142 Li J., Rhodes J. C. and Askew D. S. (1997) Evolutionary conservation of putative functional domains in the human homolog of murine *His-1* gene. Gene, **184:** 169–176
143 Li J., Witte D. P., Van Dyke T. and Askew D. S. (1997) Expression of the putative proto-oncogene *His-1* in normal and neoplastic tissues. Am. J. Pathol. **150:** 1297–1305
144 Crespti M. D., Jurkevitch E., Poiret M., d'Aubenton-Carafa Y., Petrovics G., Kondorosi E. et al. (1994) *enod40*, a gene expressed during nodule organogenesis, codes for a non-translable RNA involved in plant growth. EMBO J. **13:** 5099–5112

145 Sande K. van de, Pawlowski K., Czaja I., Wieneke U., Schell J., Schmidt J. et al. (1996) Modification of phytohormone response by a peptide encoded by ENOD40 of legumes and a non-legume. Science **273:** 370–373
146 Schell J., Bisseling T., Dülz M., Fransen H., Fritze K., John M. et al. (1999) Re-evaluation of phytochrome-independent division of tobacco protoplast-derived cells. Plant J. **17:** 461–466
147 Sande K. van de and Bisseling T. (1997) Signalling in symbiotic root nodule formation. Essays Biochem. **32:** 127–142
148 Corich V., Goormachting S., Lievens S., Van Montagu M. and Holsters M. (1998) Patterns of ENOD40 gene expression in stem-borne nodules of *Sesbania rostrata*. Plant. Mol. Biol **37:** 67–76
149 Bladergroen M. R. and Spaink H. P. (1998) Genes and signal molecules involved in the rhizobia-leguminoseae symbiosis. Curr. Opin. Plant Biol. **1:** 353–359
150 Mirabella R., Martirani L., Lamberti A., Iaccarino M. and Chiurazzi M. (1999) The soybean ENOD40(2) promoter is active in *Arabidopsis thaliana* and is temporally and spatially regulated. Plant Mol. Biol. **39:** 177–181
151 Kouchi H., Takane K., So R. B., Ladha J. K. and Reddy P. M. (1999) Rice ENOD40: isolation and expression analysis in rice and transgenic soybean root nodules. Plant J. **18:** 121–129
152 Flemetakis E., Kavroulakis N., Quaedvelleg N. E. M., Spaink H. P., Dimov M., Roussis A. et al. (2000) *Lotus japonicus* contains two distinct ENOD 40 genes that are expressed in symbiotic, nonsymbiotic and embryonic tissues, Mol. Plant Microbiol. Inter. **13:** 987–994
153 Mamat U., Rietschel E. T. and Schmidt G. (1995) Repression of lipopolysaccharide biosynthesis in *Escherichia coli* by an antisense RNA of *Acetobacter methanolicus* phage Acm1. Mol. Microbiol. **15:** 1115–1125
154 Warnecke J. M., Nitschke M., Moolenaar C. E., Ritschel E. T., Hartmann R. K. and Mamat U. (2000) The 5′-proximal hairpin loop of lbi RNA is a key structural element in repression of D-glucan II biosynthesis *Klebsiella pneumoniae* serotype O_1. Mol. Microbiol. **36:** 697–709
155 Teramoto H., Toyama T., Takeba G. and Tsuji H., (1996) Noncoding RNA for CR20, a cytokinin-repressed gene of cucumber. Plant Mol. Biol. **32:** 797–808
156 Taylor C. B. and Green P. J. (1995) Identification and characterisation of genes with unstable transcripts (GUTs) in tobacco. Plant Mol. Biol. **28:** 27–28
157 Hoff A. van, Kastenmayer J. P., Crispin B., Taylor C. B. and Green P. J. (1995) GUT15 cDNAs from tobacco and *Arabidopsis* correspond to transcripts with unusual metabolism and a short conserved OR. Plant Gene Reg. Plant Physiol. **113:** 1004
158 Komine Y., Tanaka N. K., Yano R., Takai S., Yuasa S., Shiiroishi T. et al. (1999) A novel type of non-coding RNA expressed in the rat brain. Brain Res. Mol. Brain Res. **66:** 1–13
159 Kobayashi S. and Anzai K. (1998) An E-box sequence acts as transcriptional activator for BC-1 RNA expression by RNA polymerase III in the brain. Biochem. Biophys. Res. Commun. **245:** 59–63
160 Brosius J. (1999) RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. Gene **238:** 115–134
161 Kremerskothen J., Nettermann M., Bekke A. op de, Bachmann M. and Brosius J. (1998) Identification of human auto antigen La/SS-B as BC1/BC200 RNA-binding protein. DNA Cell Biol. **17:** 751–759
162 Skryabin B. V., Kremerskothen J., Vassilacopoulou D., Disotell T. R., Kapitonov V. V., Jurka J. et al. (1998) The BC200 RNA gene and its neural expression are conserved in Anthropoidea (Primates). J. Mol. Evol. **47:** 677–685
163 Brosius J. (1999) RNAs from all categories generate retrosequences that may be expanded as novel genes or regulatory elements. Gene **238:** 115–134
164 Lanz R. B., McKenna N. J., Onate S. A., Albrecht U., Wong J., Tsai S. Y. et al. (1999) A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. Cell **97:** 17–27
165 Watanabe Y. and Yamamoto M. (1994) *S. pombe* mei2+ encodes an RNA-binding protein essential for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA species meiRNA. Cell **78:** 487–498
166 Yamashita A., Watanabe Y., Nukina N. and Yamamoto M. (1998) RNA-assisted nuclear transport of the meiotic regulator Mei2p in fission yeast, Cell **95:** 115–123
167 Ohno M. and Mattaj I. W. (1999) Meiosis: MeiRNA hits the spot. Curr. Biol. **9:** R66–R69
168 Tycowski K., Shu M-D. and Steitz J. A. (1996) A mammalian gene with introns instead of exons generating stable RNA products. Nat. **379:** 464–466
169 Smith C. M. and Steitz J. A. (1998) Classification of gas5 as a multi-small-nuclear (snoRNA) host gene and a member of the 5′-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. Mol. Cell. Biol. **18:** 6897–6909
170 Pelczar P. and Filipowicz W. (1998) The host gene for intronic U17 small nucleolar RNA in mammals has no protein-coding potential and is a member of the 5′-terminal oligopyrimidine gene family. Mol. Cell. Biol. **18:** 4509–4518
171 Bortolin M-L. and Kiss T. (1998) Human U19 intron-encoded snoRNA is processed from a long primary transcript that possesses little potential for protein coding. RNA **4:** 445–454
172 Weinstein L. B. and Steitz J. A. (1999) Guided tours: from precursor snoRNA to functional snoRNP, Curr. Opin. Cell Biol. **11:** 378–384
173 Kloc M., Spohr G. and Etkin L. D. (1993) Translocation of repetitive RNA sequences with the germ plasm in *Xenopus* oocytes. Science **262:** 1712–1714
174 Kloc M. and Etkin L. D. (1994) Delocalisation of Vg1 mRNA from the vegetal cortex in *Xenopus* oocytes after destruction of *Xlsirt* RNA. Science **265:** 1101–1103
175 Nicholls R. D. (2000) The impact of genomic imprinting for neurobehavioral and developmental disorders. J. Clin. Invest. **105:** 413–418
176 Conne B., Stutz A. and Vassalli J.-D. (2000) The 3′ untranslated region of messenger RNA: a molecular hotspot for pathology. Nat. Med. **6:** 637–641
177 Bosher J. M. and Labouesse M. (2000) RNA interference: genetic wand and genetic watchdog, Nat. Cell Biol. **2:** E31–E36
178 Sijen T. and Kooter J. M. (2000) Post-transcriptional gene-silencing: RNAs on the attack or on the defense? BioEssays **22:** 520–531
179 Kuwabara P. E. and Coulson A. (2000) RNAi – prospects for a general technique for determining gene function. Parasitol. Today **16:** 347–349
180 Karzai A. W., Susskind M. M. and Sauer T. T. (199) SmpB, a unique RNA-binding protein essential for the peptide-tagging activity of SsrA (tm RNA). EMBO J. **18:** 3793–3799
181 Hsu A. W., Kilani A. F., Liou K., Lee J. and Liu F. (2000) Differential effects of the protein cofactor on the interactions between and RNase P ribozyme an its target mRNA substrate. Nucleic Acids Res. **28:** 3105–3116
182 Zhang A., Altuvia S., Tiwari A., Argaman L., Hengge-Aronis R. and Stocz G. (1998) The *Escherichia coli* Oxy S regulatory RNA represses fh1A translation by blocking ribosome binding. EMBO J. **17:** 6061–6068
183 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