Histol Histopathol (1999) 14: 235-241

http://www.ehu.es/histol-histopathol

Histology and Histopathology

Invited Review

New insights into the function of noncoding RNA and its potential role in disease pathogenesis

D.S. Askew and F. Xu

Department of Pathology and Laboratory Medicine, University of Cincinnati, Cincinnati, OH, USA

Summary. All polyadenylated RNAs expressed in mammalian tissues are assumed to be transported to the cytoplasm where they direct the synthesis of a protein product. This mainstream view of the function of polyadenylated transcripts is currently being challenged by the identification of a novel class of genes which, although they encode polyadenylated RNA, do not make a translated protein. Many of these noncoding RNAs are developmentally regulated or show highly restricted patterns of gene expression, and their functions are providing important insight into RNA-based mechanisms of gene expression, genomic imprinting, cell cycle progression, and differentiation. The purpose of this review is to discuss the current understanding of mammalian noncoding RNAs, and to highlight their potential for identifying new pathways of human disease.

Key words: Noncoding, RNA, Disease, Imprinting, Retrovirus

Introduction

In contrast to lower eukaryotes, where most of the genome is responsible for directing the synthesis of mRNAs that are translated into proteins, the overwhelming majority of genomic DNA in higher species does not code for protein. In humans, the protein-encoding portions of genes account for only 3% of the genome, and although some of the remaining 97% may be transcribed into RNA, it does not encode protein (Nowak, 1994; Gardiner, 1995). Since most known catalytic functions in the cell are accomplished by proteins, the prevailing notion has been that noncoding DNA represents either transcriptional regulatory sequences or "junk DNA". This concept is changing however, with the discovery that "junk" DNA has sequence information with features of a complex language (Ossadnik et al., 1994), and that its presence in the genome is associated with the maintenance of chromosomal stability and human disease (Lothe, 1997; Zakian, 1997). RNA polymerase II is one of three RNA polymerases and is primarily responsible for transcribing the protein coding genes of the eukaryotic genome. RNAs that are transcribed by this enzyme generally possess a clearly identifiable open reading frame (ORF) that directs the translation of a single protein on cytoplasmic ribosomes. Although the concept that all polymerase II-transcribed genes must encode a protein is based upon a large body of evidence, a small group of genes that produce noncoding RNAs have recently been shown to provide exceptions to this generalized rule. This review will provide an overview of the known mammalian polyadenylated RNAs that are believed to function without translation into a protein product. This is an exciting area of investigation that is providing some intriguing information about new mechanisms of RNA-based gene function that are relevant to the understanding of human disease.

Offprint requests to: David S. Askew, Ph.D., Department of Pathology and Laboratory Medicine, University of Cincinnati, 231 Bethesda Ave., Cincinnati, OH 45267-0529, USA. Fax: 513-558-2141. e-mail: david.askew@uc.edu

1. His-1

Initial interest in this locus began with the demonstration that the *His-1* gene is transcriptionally activated by retroviral insertional mutagenesis in mouse retrovirus-induced leukemias (Askew et al., 1991). Since *His-1* transcripts are not normally present in hematopoietic cells, this observation indicated that *His-1* gene activation is associated with the neoplastic phenotype and may therefore contribute to leukemogenesis. The *His-1* gene is comprised of 3 exons spanning 6 kb on mouse chromosome 2, and is expressed as a 3 kb spliced and polyadenylated cytoplasmic RNA (Askew et al., 1994). Although there is currently no single unifying characteristic that can unequivocally demonstrate that an RNA does not encode a protein, several characteristics are generally shared by noncoding sequences. Instead of the single long ORF that is found in most protein-coding RNAs, noncoding RNAs contain multiple small ORFs, many of which contain direct repeats that are atypical of

protein coding exons (Brockdorff et al., 1992; Brown et al., 1992; Askew et al., 1994). It is therefore a considerable challenge to identify which, if any, of the multiple small ORFs could encode a small functional peptide. The longest *His-1* ORF could potentially encode a predicted peptide of 73 amino acids, but this ORF contains a direct repeat and evidence from in vitro translation systems have shown that it does not make a protein (Askew et al., 1994). Another common approach to assess coding capacity is to use computer sequence analysis. Recent advances in large-scale genome sequencing has shown that protein-coding exons share several sequence features that can be recognized by neural networks and other complex computer programs (Uberbacher and Mural, 1991). The entire *His-1* genomic sequence is not predicted to be coding by such programs. This finding strengthens the conclusion that the RNA does not encode a protein, and further suggests that the *His-1* sequence contains unique sequence information that is not typical of the vast majority of mammalian polyadenylated RNA.

The strongest evidence for lack of coding potential in the *His-1* RNA is provided by sequence comparisons between the mouse and human homologs. The rationale behind this approach is based upon the fact that conserved genes from different species encode related proteins from a single conserved ORF. The human homolog of the mouse *His-1* gene resides on chromosome 2q14-q21 (Askew et al., 1991), and cross hybridization studies using *His-1* DNA probes have showed that the *His-1* gene is highly conserved among many other vertebrates (Askew et al., 1994; Li et al., 1997a). Despite the fact that evolutionary constraints have restricted divergence in nucleic acid sequence, there is no conservation of the predicted amino acid sequence of the multiple small ORFs, making it highly unlikely that these ORFs are functional in vivo.

The *His-1* gene has an exceedingly narrow range of gene expression, a feature that it also shares with other noncoding RNAs (see below). The highest levels of *His-1* RNA are present in the epithelium of the embryonic choroid plexus (Li et al., 1997b). The RNA is restricted to the epithelial cells which line the developing choroid plexus, but is notably absent from the adjacent neural tissue and from the ependymal epithelium that is contiguous with the choroid plexus epithelium. The choroid plexus is a highly vascular structure arising from the wall of each of the four ventricles in the brain of all vertebrates, and it is the main source of production of cerebrospinal fluid (CSF) (Segal, 1993). The primary function of this tissue is to provide CSF for mechanical support of the brain, but there is also evidence that the choroid plexus can modify CSF composition and act as both a target and a source of neuroendocrine signalling within the brain (Nilsson et al., 1992). Lower levels of *His-1* expression have also been reported in the epithelial cells of the adult mouse prostate, seminal vesicle, and nonglandular portion of the stomach (Li et al., 1997b). The exceedingly narrow range of *His-1* gene expression suggests that powerful constraints limit the expression of this gene in most tissues, and that release of these constraints in the hematopoietic cell lineage may contribute to leukemogenesis. It will be interesting to see if targeted deletion of the *His-1* gene will interfere with the functional development of the mammalian choroid plexus.

2. Bic

Bic provides a second example of a noncoding RNA that is implicated in the pathogenesis of hematologic malignancies induced by oncogenic retroviruses. The *bic* gene was initially identified by virtue of its location within a common site of retroviral integration in B-cell lymphomas induced by Avian Leukosis Virus (ALV) (Tam et al., 1997). The transcription of the *bic* gene was specifically activated by the inserted proviruses in these tumors, resulting in high levels of a polyadenylated *bic* RNA. The activation of *bic* by these retroviruses was frequently accompanied by the transcriptional activation of the *c-myc* protooncogene, suggesting that *bic* cooperates with *c-myc* during lymphomagenesis. Preferential activation of *bic* was also identified in tumors with high metastatic potential (Tam et al., 1997), raising the possibility that the oncogenicity of *bic* is relevant for the later stages of tumor progression. The *bic* gene consists of two exons and is expressed as two spliced and alternatively polyadenylated transcripts. The expression pattern of *bic* in avian tissues is normally restricted to B and T lymphocytes of the bursa, thymus, and spleen (Tam et al., 1997). Characterization of the full length RNA revealed no signficant ORF, suggesting that *bic* is another example of a noncoding RNA. The normal function of *bic* is currently unknown, but the fact that exon 2 is predicted to form extensive double-stranded stem-loops that are energetically favorable, suggests that the function involves a requirement for RNA secondary structure. Although there is no homology between *bic* and *His-1*, it will be interesting to determine if these noncoding RNAs use similar mechanisms to promote oncogenesis.

3. H19

The *H19* gene provides one of the best characterized examples of a spliced and polyadenylated RNA that does not encode a protein. *H19* was originally identified as a mammalian gene that is coordinately regulated with the α-fetoprotein gene by the trans-acting locus *raf* (Pachnis et al., 1984). The gene is composed of 5 exons spanning 3 kb and is expressed as a single 2.5 kb cytoplasmic RNA that is spliced and polyadenylated (Brannan et al., 1990). The numerous stop codons in the *H19* RNA produce multiple small ORFs, none of which span more than 2 of the 5 exons. The *H19* RNA does not associate with ribosomes, and since none of the ORFs are conserved in homologs from other species, it has been concluded that the *H19* RNA does not make a protein

(Brannan et al., 1990).

The *H19* gene has a developmentally restricted pattern of expression in fetal and neonatal liver, visceral endoderm and fetal gut, and is repressed in all adult tissues except skeletal muscle (Pachnis et al., 1988). Expression of the *H19* gene is regulated by genomic imprinting (Bartolomei et al., 1991), an epigenetic process which silences one parental allele during transmission through the germ line (Hall, 1997). This is a rare mode of gene regulation that results in exclusive expression from a single parentally-inherited allele. Genomic imprinting is required for normal mammalian development, and abnormalites in the imprinting of certain genes are implicated in a variety of human genetic disorders and cancer (Reik, 1989; Rainier et al., 1993). The *H19* gene lies at the end of a cluster of genomically imprinted genes on mouse chromosome 7 and is itself imprinted such that expression is exclusively from the maternal allele (Bartolomei et al., 1991).

Mice which harbor a homozygous deletion of the entire *H19* structural gene revealed abnormalities in the imprinting of adjacent genes, suggesting that *H19* controls the imprinting of flanking genes (Leighton et al., 1995). The maternal inheritance of the *H19* deletion resulted in the disruption of normal imprinting of the insulin-2 (Ins-2) and insulin-like growth factor 2 (IGF2) genes which lie 100 kb upstream of *H19* (Leighton et al., 1995). Paternal inheritance of the *H19* mutation had no effect since *H19* is normally expressed from the maternal allele and the paternally inherited allele is normally silent (Bartolomei et al., 1991; Leighton et al., 1995). The identification of a role for a spliced and polyadenylated noncoding RNA as the signal that regulates the genomic imprinting of multiple flanking genes raises the possibility that other imprinted genes may be regulated by similar mechanisms that involve noncoding RNAs.

The ability of *H19* to influence the imprinting pattern of adjacent genes has been attributed to promoter competition for a common set of enhancers, but elucidating the function of the RNA itself has proven to be more difficult. A possible role for the *H19* RNA as a tumor suppressor was demonstrated by overexpressing the RNA in embryonal carcinoma cells. Expression of *H19* in these cells inhibited growth, reduced clonogenicity in soft agar, and reduced tumorigenicity in nude mice, but a mechanistic explanation for this antitumorigenic effect was not established (Hao et al., 1993). Consistent with a potential role in tumor suppression, loss of *H19* expression has been reported to be an early event in carcinogen-induced transformation of Syrian hamster cells (Cui et al., 1997) and in primary human Wilm's tumors (Hao et al., 1993; Isfort et al., 1997). Regardless of the mechanism involved, the participation of the *H19* gene in the regulation of gene expression at the chromosomal level has important implications for human diseases where genomic imprinting conveys a unique pattern of inheritance (Reik, 1989; Rainier et al., 1993).

4. XIST

The suppression of gene expression on one X-chromosome in mammals is under the control of a single cis-acting locus called the X-inactivation center, *Xic* (Lyon, 1994). The *XIST* gene maps to this region and is expressed exclusively from the inactive X-chromosome (Brown et al., 1991). The product of the 8 exons of the mouse *XIST* gene is an extremely large, spliced and polyadenylated RNA (17 kb in humans) that shares the characteristic high density of stop codons with other noncoding sequences (Brockdorff et al., 1992; Brown et al., 1992). The mouse and human *XIST* homologs contain several tandem repeats that are strikingly conserved, but the lack of a conserved ORF between the two RNAs has led to the conclusion that *XIST* is another example of a noncoding RNA (Brown et al., 1991, 1992; Brockdorff et al., 1992; Hendrich et al., 1993). This concept is further strengthened by the inability of the coding recognition program GRAIL to recognize the human *XIST* sequence as coding (Brockdorff et al., 1992; Brown et al., 1992).

A unique feature of the *XIST* RNA is its nuclear localization, despite the fact that it is polyadenylated and spliced like a typical protein-coding transcript (Brockdorff et al., 1992; Brown et al., 1992; Hendrich et al., 1993). Furthermore, the *XIST* RNA becomes physically associated with the inactive X-chromosome, suggesting that it controls X-inactivation by remodeling chromatin through interactions between the RNA and the X chromosome (Willard and Salz, 1997). Functional studies to support a direct role for the *XIST* RNA in the repression of X-linked gene expression has come from gene-targeting and transgenic approaches in embryonic stem (ES) cells. By analogy to the gene disruption approach utilized for *H19*, a large deletion mutation in the *XIST* gene was created to maximize the probability of disrupting the function of the *XIST* RNA (Penny et al., 1996). The gene targeting was performed in XX ES cells so that analysis of the initiation and progression of X-inactivation could be studied by in vitro differentiation of the mutant ES cells. The results of this deletion demonstrated that the *Xist* gene is required for X-inactivation to occur in cis, thereby providing direct evidence that a noncoding RNA could regulate gene expression on a chromosome-wide basis (Penny et al., 1996). This provides an intriguing parallel to the ability of the *H19* RNA to influence gene expression .

Recent studies have examined the possibility that the *XIST* RNA can suppress the expression of genes that are not located on the X-chromosome. This was performed by introducing copies of *XIST* into autosomes in murine ES cells. The ectopically transcribed *XIST* RNA remained nuclear, became closely associated with the autosome into which the gene was integrated, and induced long-range gene inactivation along the affected autosome (Herzing et al., 1997; Lee and Jaenisch, 1997). This indicates that the mechanism(s) involved in the suppression of genes on the X chromosome by the *XIST*

RNA can also operate on autosomes, raising the possibility that additional RNA-based methods of complex gene regulation may play a role in autosomal gene expression in higher eukaryotes. A role for *XIST* in disease is suggested by the association of mental retardation and congenital anomalies with human genetic disorders that lack *XIST* (Apacik et al., 1996; Guillen et al., 1997). It is therefore likely that further analysis of this RNA will provide insights into the causes and consequences of congenital defects that involve aberrant female dosage compensation mechanisms, and this is consistent with the observation that female mice which carry an *XIST* deletion mutation on the paternal X-chromosome fail to develop normally (Marahrens et al., 1997).

5. The 3' UTR

It is worth noting that a high density of stop codons is characteristic of the 3' untranslated regions (3' UTRs) of all eukaryotic mRNAs that encode a protein product. Although it was presumed for many years that this feature reflected a lack of biological significance, it is now known that the 3' UTR plays a crucial role in the regulation of mRNA function. At least one of these roles is the regulation of translation, a function that is mediated by RNA-binding proteins that interact with specific sequences in the UTR (Goodwin et al., 1993). A second and previously unanticipated function for the 3' UTR is the ability of noncoding sequences in the 3' UTRs of certain muscle structural genes to activate muscle specific promoters, regulate the proliferation and differentiation of some cell types, and function as a tumor suppressor (Rastinejad and Blau, 1993; Rastinejad et al., 1993). The mechanism by which the noncoding RNA of the 3' UTR can regulate gene expression and influence cell growth has not been established, but is presumed to involve protein-RNA interactions. Regardless of the mechanism involved, the discovery of a functional role for noncoding sequences in mRNA has important implications for the pathogenesis of neoplastic and congenital diseases that are associated with disruptions in the untranslated regions of specific cellular genes (Begley et al., 1989; Brook et al., 1992; Liu et al., 1993; Zelenetz et al., 1993).

6. RNA derived from introns

Evolutionary comparisons between species homologs of eukaryotic genes have shown that protein coding exons are generally the most conserved, whereas intronic sequences display the greatest variability. Exceptions to this have been reported however (Koop and Hood, 1994), and recent evidence has established precedence for conserved introns that produce noncoding RNAs. One of the best studied examples of an intron-encoded noncoding RNA is self-splicing RNA. These catalytic RNAs fold into tertiary structures and catalyze the excision and religation of the host exons without the aid of the spliceosome (Cech, 1990).

A second example is provided by the snoRNAs. These small stable noncoding RNAs are found within the introns of protein-encoding genes and are believed to play a role in rRNA processing in the nucleolus (Maxwell and Fournier, 1995). They are transcribed as part of the primary transcript, and the mature snoRNA is released by excision and subsequent trimming of the intronic sequences (Kiss and Filipowicz, 1995). One unusual member of the snoRNA family, U22, is found within the penultimate intron of the U22 host gene (UHG). In contrast to other reported snoRNA host genes, the spliced and polyadenylated UHG RNA possesses the same high density of stop codons that is characteristic of noncoding RNA, suggesting that it does not encode a protein (Tycowski et al., 1996). Strikingly, it is the introns of the UHG gene that show the greatest degree of conservation and these introns have been shown to contain several snoRNAs in addition to U22 (Tycowski et al., 1996). This is the first example of a gene for which the functional product is a noncoding RNA that is specified by the introns rather than the exons. It is conceivable that further interspecies comparisons of mammalian genomes will identify conserved intronic sequences that are flanked by noncoding exons, and that these introns may provide a previously unsuspected source of new examples of noncoding RNAs.

7. IPW

The Prader-Willi syndrome (PWS) is a neurobehavioral disorder which is characterized by growth and mental retardation, failure to thrive in infancy, and hyperphagia (Butler, 1990). Most patients with PWS have a microscopically detectable cytogenetic deletion of their paternal 15q11-q13 chromosome. The parent-of-origin dependence of this disease suggests that one or more genes in this region of chromosome 15 is expressed from a single allele in a uniparental manner. The *IPW* gene is expressed as a 2.2 kb spliced and polyadenylated RNA from 3 exons in this region of chromosome 15, and is considered to play an important role in the pathogenesis of PWS (Wevrick et al., 1994). The *IPW* RNA is believed to function as a noncoding RNA because it shares the lack of an extensive ORF with each of the noncoding RNAs reported to date (Wevrick et al., 1994). The RNA has a broad tissue distribution, but it is preferentially expressed from the paternal allele in fetal tissues suggesting that it is subject to allele-specific gene regulation during development (Rachmilewitz et al., 1996). Although genomic imprinting is not a common form of gene regulation, it is striking that three noncoding mammalian RNAs, *H19*, *XIST*, and *IPW* are imprinted. Since the *XIST* and *H19* RNAs participate in the regulation of gene expression at the chromosomal level, it will be of interest to determine if the *IPW* RNA has a related role that contributes to the complex Prader-Willi phenotype.

Table 1. Examples of noncoding RNAs that are providing insight into novel mechanisms of RNA-based gene function.

NONCODING RNA	SPECIES	POSSIBLE FUNCTION
H19	Mammalian	Gene regulation (genomic imprinting, tumorigenesis[1]
Xist	Mammalian	Gene regulation (X-inactivation)[1]
3'UTRs	Mammalian	Gene regulation, tumor suppression[1]
7H4	Mammalian	Synaptic function in muscle? [1]
IPW	Mammalian	Unknown (imprinting?) [1]
NTT	Mammalian	Gene regulation? [1]
Introns	Multiple	Gene regulation (RNA processing) [1]
Lin-4	*C. elegans*	Gene regulation (RNA-RNA antisense inhibition)[2]
RoX1 & RoX2	*Drosophila*	Gene regulation (X-chromosome)[3]
93D	*Drosophila*	Heat shock/transcript processing and translation[4]
Pgc & Xlsirts	*Drosophila/Xenopus*	Germ cell formation[5]
Sme2+	Yeast	Cell cycle regulation[6]
OxyS	*E. coli*	Protection from oxidative stress[7]

[1]: Discussed in the text; [2]: Lee et al., 1993; [3]: Meller et al., 1997; [4]: Lakhotia and Sharma, 1996; [5]: Nakamura et al., 1996; Kloc and Etkin, 1994; [6]: Watanabe and Yamamoto, 1994; [7]: Altuvia et al., 1997.

8. NTT

The *NTT* gene was identified by a differential display approach, based upon its increased expression in activated T-cell clones (Liu et al., 1997). It is located on chromosome 6q23-q24 and is specifically expressed in activated, but not resting T cells. The *NTT* gene produces a 17 kb polyadenylated transcript that has a very high density of stop codons and a predominantly nuclear localization. Extensive sequence analysis of the multiple small ORFs in the *NTT* sequence failed to identify an ORF in good context for translation, and the sequence was not predicted to be coding by coding region prediction algorithms. Athough the function of the *NTT* RNA is not yet known, the presence of tandem repeats, a high density of stop codons, a large size, and a nuclear localization are striking features that the *NTT* RNA shares with the *XIST* RNA. The ability of *XIST* to repress autosomal gene expression when ectopically expressed has raised the possibility that additional gene-regulatory RNAs have yet to be identified (Herzing et al., 1997; Lee and Jaenisch, 1997), and the identification of the *NTT* RNA has provided a likely candidate for such an RNA.

9. 7H4

Very little is known about the *7H4* RNA. The RNA was cloned during a subtractive hybridization approach that was used to enrich for RNAs that function at the motor endplate (Velleca et al., 1994). The gene is highly conserved among vertebrates, and has a very narrow range of expression. The *7H4* RNA is muscle specific and is upregulated during early postnatal development and after denervation of muscle (Velleca et al., 1994). The gene contains no introns, but the RNA is polyadenylated like a typical mRNA. The lack of a significant ORF in the sequence of the *7H4* RNA has indicated that this RNA may be another example of a mammalian noncoding RNA. Since the RNA is localized to the synapse at the muscle endplate, it is presumed that it has a synaptic function, but this remains to be established.

Conclusions

Genes that encode noncoding RNAs are providing important insight into previously unanticipated roles for polyadenylated RNA in both normal cell function and in disease. Although only a limited number of such genes have been reported in mammals, other examples have been identified in *Drosophila*, *Xenopus*, plants and yeast (Table 1). Some of these noncoding RNAs have no established function, whereas others have been implicated in the control of germ cell formation, meiosis, and adaptation to heat shock or oxidative stress. The existence of noncoding RNAs therefore appears to be a phylogenetically conserved phenomenon, but there is as yet no single characteristic that can be used to either identify them or to understand how they act. It is hoped that the large-scale sequencing of complex mammalian genomes and expressed sequence tags will provide clues about the functional significance of the sequence information that is contained in atypical RNAs, so that we can better understand the mechanism by which they interact with cellular machinery to coordinate cell functions.

Acknowledgements. Work presented from the author's laboratory was supported by National Institutes of Health grant R29CA61909 to D.S.A. and by an American Cancer Society Junior Faculty Research Award to D.S.A.

References

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.

Apacik C., Cohen M., Jakobeit M., Schmucker B., Schuffenhauer S.,

Thurn und Taxis E., Genzel-Boroviczeny O. and Stengel-Rutkowski S. (1996). Two brothers with multiple congenital anomalies and mental retardation due to disomy (X) (q12—>q13.3) inherited from the mother. Clin. Genet. 50, 63-73.

Askew D.S., Bartholomew C., Buchberg A.M., Valentine M.B., Jenkins N.A., Copeland N.G. and Ihle J.N. (1991). *His-1* and *His-2*: identification and chromosomal mapping of two commonly rearranged sites of viral integration in a myeloid leukemia. Oncogene 6, 2041-2047.

Askew D.S., Li J. and Ihle J.N. (1994). Retroviral insertions in the murine *His-1* locus activate the expression of a novel RNA that lacks an extensive open reading frame. Mol. Cell. Biol. 14, 1743-1751.

Bartolomei M.S., Zemel S. and Tilghman S.M. (1991). Parental imprinting of the mouse *H19* gene. Nature 351, 153-155.

Begley C.G., Aplan P.D., Denning S.M., Haynes B.F., Waldmann T.A. and Kirsch I.R. (1989). The gene scl is expressed during early hematopoiesis and encodes a differentiation-related DNA-binding motif. Proc. Natl. Acad. Sci. USA 86, 10128-10132.

Brannan C.I., Dees E.C., Ingram R.S. and Tilghman S.M. (1990). The product of the *H19* gene may function as an RNA. Mol. Cell. Biol. 10, 28-36.

Brockdorff N., Ashworth A., Kay G.F., McCabe V.M., Norris D.P., Cooper P.J., Swift S. and Rastan S. (1992). The product of the mouse *Xist* gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell 71, 515-526.

Brook J.D., McCurrach M.E., Harley H.G., Buckler A.J., Church D., Aburatani H., Hunter K., Stanton V.P., Thirion J.P., Hudson T., Sohn R., Zemelman B., Snell R.G., Rundle S.A., Crow S., Davies J., Shelbourne P., Buxton J., Jones C., Juvonen v., Johnson K., Harper P.S., Shaw D.J. and Housman D.E. (1992). Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3'end of a transcript encoding a protein kinase family member. Cell 68, 799-808.

Brown C.J., Ballabio A., Rupert J.L., Lafreniere R.G., Grompe M., Tonlorenzi R. and Willard H.F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature 349, 38-44.

Brown C.J., Hendrich B.D., Rupert J.L., Lafreniere R.G., Xing Y., Lawrence J. and Willard H.F. (1992). The human *XIST* gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell 71, 527-542.

Butler M.G. (1990). Prader-Willi syndrome: current understanding of cause and diagnosis. Am. J. Med. Genet. 35, 319-332.

Cech T.R. (1990). Self-splicing of group I introns. Annu. Rev. Biochem. 59, 543-568.

Cui H., Hedborg F., He L., Nordenskjold A., Sandstedt B., Pfeifer-Ohlsson S. and. Ohlsson R. (1997). Inactivation of *H19*, an imprinted and putative tumor repressor gene, is a preneoplastic event during Wilm's tumorigenesis. Cancer Res. 57, 4469-4473.

Gardiner K. (1995). Human genome organization. Curr. Opin. Genet. Dev. 5, 315-322.

Goodwin E.B., Okkema P.G., Evans T.C. and Kimble J. (1993). Translational regulation of tra-2 by its 3'untranslated region controls sexual identity in *C. elegans*. Cell 75, 329-339.

Guillen D.R., Lowichik A., Shneider N.R., Cohen D.S., Garcia S. and Zinn A.R. (1997). Prune-belly syndrome and other anomalies in a stillborn fetus with a ring X chromosome lacking *XIST*. Am. J. Med. Genet. 70, 32-36.

Hall J.G. (1997). Genomic imprinting: nature and clinical relevance. Ann. Rev. Med. 48, 35-44.

Hao Y., Crenshaw T., Moulton T., Newcomb E. and Tycko B. (1993). Tumour-suppressor activity of *H19* RNA. Nature 365, 764-767.

Hendrich B.D., Brown C.J. and Willard H.F. (1993). Evolutionary conservation of possible functional domains of the human and murine genes. Hum. Mol. Genet. 2, 663-672.

Herzing L.B.K., Romer J.T., Horn J.M. and Ashworth A. (1997). *Xist* has properties of the x-chromosome inactivation centre. Nature 386, 272-275.

Isfort R.J., Cody D.B., Kerckaert G.A., Tycko B. and LeBoeuf R.A. (1997). Role of the *H19* gene in Syrian hamster embryo cell tumorigenicity. Molec. Carcinogen 20, 189-193.

Kiss T. and Filipowicz W. (1995). Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns. Genes Dev. 9, 1411-1424.

Kloc M. and Etkin L.D. (1994). Delocalization of *Vgl* mRNA from the vegetal cortex in *Xenopus* oocytes after destruction of Xlsirt RNA. Science 265, 1101-1103.

Koop B.F. and Hood L. (1994). Striking sequence similarity over almost 100 kilobases of human and mouse T-cell receptor DNA. Nature Genet. 7, 48-53.

Lakhotia S.C. and Sharma A. (1996). The 93D (hsr-omega) locus of Drosophila: non-coding gene with house-keeping functions. Genetica 97, 339-348.

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

Lee 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.

Leighton P.A., Ingram R.S., Eggenschwiler J., Efstratiadis A. and Tilghman S.M. (1995). Disruption of imprinting caused by deletion of the *H19* gene region in mice. Nature 375, 34-39.

Li J., Rhodes J.C. and Askew D.S. (1997a). Evolutionary conservation of putative functional domains in the human homolog of the murine *His-1* gene. Gene 184, 169-176.

Li J., Witte D.P., Van Dyke T. and Askew D.S. (1997b). Expression of the putative proto-oncogene *His-1* in normal and neoplastic tissues. Am. J. Pathol. 150, 297-1305.

Liu 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 CD4+ T cells. Genomics 39, 171-184.

Liu P., Tarlé S.A., Hajra A., Claxton D.F., Marlton P., Freedman M., Siciliano M.J. and Collins F.S. (1993). Fusion between transcription factor CBFß/PEBP2ß and a myosin heavy chain in acute myeloid leukemia. Science 261, 1041-1044.

Lothe R.A. (1997). Microsatellite instability in human solid tumors. Molec. Med. Today 3, 61-68.

Lyon M.F. (1994). The x-inactivation centre and x-chromosome imprinting. Eur. J. Human Genet. 2, 255-261.

Marahrens Y., Panning, B., Dausman J., Strauss W. and Jaenisch R. (1997). *Xist*-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev. 11, 156-166.

Maxwell E.S. and Fournier M.J. (1995). The small nucleolar RNAs. Annu. Rev. Biochem. 64, 897-934.

Meller V.H., Wu K.H., Roman G., Kuroda M.I. and Davis R.L. (1997). RoX1 RNA paints the X chromosome of male Drosophila and is regulated by the dosage compensation system. Cell 88, 445-457.

Nakamura A., Amikura R., Mukai M., Kobayashi S. and Lasko P.F. (1996). Requirement for a noncoding RNA in Drosophila polar granules for germ cell establishment. Science 274, 2075-2079.

Nilsson C., Lindvall-Axelsson M. and Owman C. (1992). Neuroendocrine regulatory mechanisms in the choroid plexus-cerebrospinal fluid system. Brain Res. Rev. 17, 109-138.

Nowak R. (1994). Mining treasures from 'junk DNA'. Science 263, 608-610.

Ossadnik S.M., Buldyrev S.V., Goldberger A.L., Havlin S., Mantegna R.N., Peng C.K., Simons M. and Stanley H.E. (1994). Correlation approach to identify coding regions in DNA sequences. Biophys. J. 67, 64-70.

Pachnis V., Belayew A. and Tilghman S.M. (1984). Locus unlinked to alpha-fetoprotein under the control of the murine *raf* and Rif genes. Proc. Natl. Acad. Sci. USA 81, 5523-5527.

Pachnis V., Brannan C.I. and Tilghman S.M. (1988). The structure and expression of a novel gene activated in early mouse embryogenesis. EMBO J. 7, 673-681.

Penny G.D., Kay G.F., Sheardown S.A., Rastan S. and Brockdorff N. (1996). Requirement for *Xist* in X chromosome inactivation. Nature 379, 131-137.

Rachmilewitz J., Elkin M., Looijenga L.H.J., Verkerk A.J.M.H., Gonik B., Lustig O., Werner D., de Groot N. and Hochberg A. (1996). Characteriztion of the imprinted *IPW* gene. allelic expression in normal and tumorigenic human tissues. Oncogene 13, 1687-1692.

Rainier S., Johnson L.A., Dobry C.J., Ping A.JL, Grundy P.E. and Feinberg A.P. (1993). Relaxation of imprinted genes in human cancer. Nature 362, 747-749.

Rastinejad F. and Blau H.M. (1993). Genetic complementation reveals a novel regulatory role for 3' untranslated regions in growth and differentiation. Cell 72, 903-917.

Rastinejad F., Conboy M.J., Rando T.A. and Blau H.M. (1993). Tumor suppression by RNA from the 3' untranslated region of α-tropomyosin. Cell 75,1107-1117.

Reik W. (1989). Genomic imprining and genetic disorders in man. Trends Genet. 5, 331-336.

Segal M.B. (1993). Extracellular and cerebrospinal fluids. J. Inher. Metab. Dis. 16, 617-638.

Tam W., Ben-Yehuda D. and Hayward W.S. (1997). *bic*, a novel gene activated by proviral insertions in avian leukosis virus-induced lymphomas, is likely to function through its noncoding RNA. Mol. Cell. Biol. 17, 1490-1502.

Tycowski K.T., Shu M.D. and Steitz J.A. (1996). A mammalian gene with introns instead of exons generating stable RNA products. Nature 379, 464-466.

Uberbacher E.C. and Mural R.J. (1991). Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. Proc. Natl. Acad. Sci. USA 88, 11261-11265.

Velleca M.A., Wallace M.C. and Merlie J.P. (1994). A novel synapse-associated noncoding RNA. Mol. Cell. Biol. 14, 7095-7104.

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.

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

Willard H.F. and Salz H.K. (1997). Remodelling chromatin with RNA. Nature 386, 228-229.

Zakian V.A. (1997). Life and cancer without telomerase. Cell 91, 25-34.

Zelenetz A.D., Cleary M.L. and Levy R. (1993). A submicroscopic interstitial deletion of chromosome 14 frequently occurs adjacent to the t(14;18) translocation breakpoint in human follicular lymphoma. Gene. Chromosome. Canc. 6, 140-150.

Accepted July 3, 1998