Non-coding RNA

John S. Mattick* and Igor V. Makunin

Australian Research Council Centre for Functional and Applied Genomics, Institute for Molecular Bioscience, University of Queensland, St Lucia, QLD 4072, Australia

The term non-coding RNA (ncRNA) is commonly employed for RNA that does not encode a protein, but this does not mean that such RNAs do not contain information nor have function. Although it has been generally assumed that most genetic information is transacted by proteins, recent evidence suggests that the majority of the genomes of mammals and other complex organisms is in fact transcribed into ncRNAs, many of which are alternatively spliced and/or processed into smaller products. These ncRNAs include microRNAs and snoRNAs (many if not most of which remain to be identified), as well as likely other classes of yet-to-be-discovered small regulatory RNAs, and tens of thousands of longer transcripts (including complex patterns of interlacing and overlapping sense and antisense transcripts), most of whose functions are unknown. These RNAs (including those derived from introns) appear to comprise a hidden layer of internal signals that control various levels of gene expression in physiology and development, including chromatin architecture/epigenetic memory, transcription, RNA splicing, editing, translation and turnover. RNA regulatory networks may determine most of our complex characteristics, play a significant role in disease and constitute an unexplored world of genetic variation both within and between species.

INTRODUCTION

Until recently most of the known non-coding RNAs (ncRNAs) fulfilled relatively generic functions in cells, such as the rRNAs and tRNAs involved in mRNA translation, small nuclear RNAs (snRNAs) involved in splicing and small nucleolar RNAs (snoRNAs) involved in the modification of rRNAs. The central tenet of molecular biology, developed from the study of simple organisms like *Escherichia coli*, has been that RNA functions mainly as an informational intermediate between a DNA sequence ('gene') and its encoded protein. The presumption has been that most genetic information that specifies biological form and phenotype is expressed as proteins, which not only fulfill diverse catalytic and structural functions, but also regulate the activity of the system in various ways. This is largely true in prokaryotes and presumed also to be true in eukaryotes. Reciprocally, the extensive sequences in the higher eukaryotes that do not encode proteins or *cis*-acting regulatory elements (i.e. the majority of the vast tracts of intronic and intergenic sequences) have been regarded as simply accumulated evolutionary debris arising from the early assembly of genes and/or the insertion of mobile genetic elements.

However, most of these supposedly inert sequences are transcribed. It is also increasingly evident that RNA itself can and does have a very wide repertoire of biological functions (1) and, in particular—as first predicted by Jacob and Monod 45 years ago (2)—that it is widely employed as a means of gene regulation, both in *cis* and in *trans*, especially in the higher eukaryotes. These RNAs are the subject of this review.

EXPANSION OF ncRNAs AND RNA METABOLISM IN EUKARYOTES

A limited number of *trans*-acting small ncRNAs have been described in prokaryotes that appear mainly to regulate mRNA translation or stability. Over 60 such RNAs have been identified during the past few years in *E. coli*, with another 200 or so predicted bioinformatically (3–7). Some of these RNAs are co-expressed with mRNAs and released by cleavage after transcription (4,7), examples of a parallel output of regulatory RNAs that appears to be widespread in the higher eukaryotes (8). Small ncRNAs have also been identified in other bacteria (see e.g. 9,10) and archaea (11), which interestingly have homologs of Argonaute, a family of RNA-binding endonucleases central to the action of micro-RNAs (miRNAs) and small interfering RNAs (siRNAs) in eukaryotes (12). However, ncRNAs do not dominate

*To whom correspondence should be addressed. Tel: +61 733462079; Fax: +61 733462111; Email: j.mattick@imb.uq.edu.au

genomic output in prokaryotes, representing, as far as one can tell, only a small fraction of their genomes, which are generally dominated (80–95%) by protein-coding sequences (13), whose repertoire can vary widely even between closely related strains (14).

In contrast, the higher organisms have a relatively stable proteome, and a relatively static number of protein-coding genes, which is not only much lower than expected but also varies by less than 30% between the simple nematode worm *Caenorhabditis elegans* (which has only 10^3 cells) and humans ($\sim10^{14}$ cells), which have far greater developmental and physiological complexity (15). Moreover, only a minority of the genomes of multicellular organisms is occupied by protein-coding sequences, the proportion of which declines with increasing complexity, with a concomitant increase in the amount of non-coding intergenic and intronic sequences, most of which are in fact transcribed [(15,16); discussed in more detail subsequently]. Thus, there seems to be a progressive shift in transcriptional output between microorganisms and multicellular organisms from mainly protein-coding mRNAs to mainly non-coding RNAs, including intronic RNAs.

The eukaryotes, particularly the higher eukaryotes, also have a far more developed RNA processing and signaling system than prokaryotes, which appears to be linked to the more sophisticated pathways of gene regulation and complex genetic phenomena in eukaryotes, transcriptional and post-transcriptional gene silencing, including RNA interference (RNAi), DNA methylation and chromatin modification, imprinting, and other phenomena such as transvection, trans-induction, dosage compensation and position effect variegation (8,17,18). The higher eukaryotes also have a large repertoire of RNA-binding proteins as well as many nucleic acid- and chromatin-binding proteins whose exact specificity is unknown or uncertain, but which may recognize different types of RNA:RNA and RNA:DNA complexes (8,18).

Both theoretic considerations and empirical evidence indicate that the amount of regulatory overhead scales non-linearly with complexity in all integrated systems, and that regulatory architecture will progressively dominate the information content of more complex systems, leading to complexity limits, until and unless there is a change in the physical basis of the regulatory architecture itself (19). The generic solution to this accelerating regulatory problem is the superimposition of digital communication and control systems, which have only been broadly established in the human intellectual lexicon during the past 20–30 years, well after the central tenets of molecular biology were developed and after introns were discovered. Interestingly, although it is widely appreciated that DNA itself is a digital storage medium, it has not been considered that some of its outputs may themselves be digital signals, communicated via ncRNA, in addition to the mRNAs encoding analog components (i.e. the proteins), albeit with many design variations elaborated by alternative splicing (which itself requires regulation).

Regulatory proteins scale almost quadratically with genome size in prokaryotes (20,21), and extrapolation of this relationship suggests that prokaryotes have been limited in their complexity by their reliance on a protein-based regulatory architecture, probably for most of their evolutionary history (13,19,20,22). Conversely, it appears that the eukaryotes

breached this limit by the co-option of RNA as a digital regulatory solution, in concert with the evolution of the necessary protein infrastructure to recognize and act on these signals (13). Indeed, both logic and evidence suggest that both developmental programming and the phenotypic difference between species and individuals is heavily influenced, if not fundamentally controlled, by the repertoire of regulatory ncRNAs (13,16–18,23), which are only now being recognized and beginning to be studied in any systematic way.

INFRASTRUCTURAL ncRNAs

Some infrastructural ncRNAs have been known for a long time and have well-established functions. These include tRNAs, rRNAs, spliceosomal uRNAs or 'snRNAs' and the common 'snoRNAs'. Both translation and splicing require core infrastructural RNAs not only for sequence-specific recognition of RNA substrates, but also for the catalytic process itself (1,24–27). Recent findings indicate that some of these RNAs, not surprisingly, may also be involved in regulatory processes. For example, besides its role in splicing, the *U1* snRNA is involved in the regulation of transcription initiation by RNA polymerase II through interaction with the transcription initiation factor TFIIH (28). *U1* RNA also interacts with cyclin H (29), raising the possibility that ncRNA might be involved in cell cycle regulation. In addition, the small conserved nuclear RNA *7SK* inhibits the kinase activity of the CDK9/cyclin T complex, leading to reduced phosphorylation of RNA polymerase II and a reduction in transcription (30). The *7SK* RNA acts in concert with the HEXIM1 and HEXIM2 proteins, both of which show distinct expression patterns in various human tissues (31–33), and depletion of *7SK* RNA by siRNA causes apoptosis in HeLa cells (34).

ncRNAs also play a role in chromosome maintenance and segregation (35). A small RNA with similarity to box H/ACA snoRNAs is a component of telomerase (for review see 36) and is mutated in autosomal dominant dyskeratosis congenita (37). In human–chicken hybrid cells, mutation of *Dicer*, a key component of the siRNA/miRNA processing machinery, leads to the accumulation of transcripts derived from centromeric-satellite repetitive sequences, premature separation of sister chromatids and cell death (38). ncRNA has also been implicated in the control of chromatin architecture and epigenetic memory (35,39; discussed further below).

There are also other types of infrastructural ncRNAs that are involved in central cell biological processes. The ncRNA *7SL* RNA is a core component of the signal recognition particle (SRP), a ribonucleoprotein complex that interacts with the ribosome and is essential for targeting/transportation of nascent proteins containing signal peptides to the endoplasmic reticulum membrane for secretion or membrane insertion (40–43).

The 13 MDa vault complex (discovered in 1986) is the largest ribonucleoprotein complex described to date, three times bigger (albeit far less complex) than the ribosome. It is present in 10^4 to 10^5 copies per cell, forms a barrel-like structure predominantly localized in the cytoplasm and is presumably involved in transport (for review see 44).

Different species have between one and three vault RNAs, ranging in length from 86 to 141 nucleotides. In multi-drug resistant cells, the vault complex is upregulated and has a different ratio of vault RNAs in comparison with normal (44). Moreover, two human vault RNAs, *hvg-1* and *hvg-2*, specifically bind to mitoxantrone (45), a chemotherapeutic agent commonly used for treatment of breast cancer, myeloid leukemia and non-Hodgkin's lymphoma.

cis-ACTING REGULATORY SEQUENCES IN NON-CODING REGIONS OF mRNAs AND PRE-mRNAs

Regulatory RNAs function in most cases by base-pairing with complementary sequences in other RNAs and DNA, to form RNA:RNA (and probably RNA:DNA) complexes that are recognized, and acted upon, by a relatively generic infrastructure [such as RNA-induced silencing complex (RISC) complexes or RNA editing enzymes]. There are many well-characterized examples of regulatory RNA sequences in the untranslated regions (UTRs) of mRNAs that act in *cis* as receivers of other *trans*-acting signals, by forming secondary structures that bind regulatory proteins or small molecular weight ligands. Examples of the former include sequences in UTRs that can bind regulatory proteins or be the targets of RNA editing to control the stability, translatability or localization of mRNAs (46–49). Examples of the latter are the so-called 'riboswitches' that regulate metabolic pathways by binding metabolites such as vitamins, amino acids and purines, to effect allosteric changes in the mRNA to control its translation or stability. These have been well documented in bacteria (50–52), but also occur in eukaryotes (53,54).

UTRs in mRNAs (as well as the coding sequences themselves) can also be the sensors of *trans*-acting regulatory RNAs, specifically miRNAs (at least some of which are encoded in introns of other genes), by base sequence recognition (8,55,56), which appear to have significant influence on their evolution (57). That is, ncRNAs can either be receivers or transmitters, or both, of regulatory signals. Interestingly, the average length of the UTRs in mRNAs increase with developmental complexity in animals, and is almost equivalent to the length of the protein-coding sequences in human (total 34 Mb of coding sequences and 32 Mb of UTR at last count) (15), indicative of the much greater sophistication of mRNA regulation in the higher organisms.

There are also *cis*-acting regulatory sequences in and around splice junctions, some of which (the so-called 'exon-splicing enhancers' or ESEs) occur within protein-coding sequences (58). Nucleotide sequence conservation is higher around alternative splice sites than constitutive splice sites, albeit in complex patterns (59–61). These sequences are thought to bind regulatory proteins that influence splice selection, but two recent papers have suggested that such selection may, at least in some cases, involve complex RNA:RNA interactions, which are themselves presumably regulated by other *trans*-acting signals, including other RNAs (62–64). Consistent with this, small artificial antisense RNAs and introduced riboswitches have been shown to easily regulate splicing *in vitro* and *in vivo* (65–68), with obvious
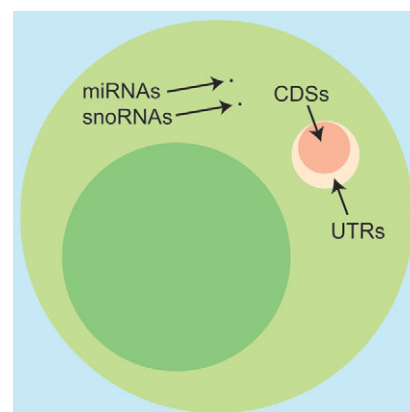

Figure 1. Graphical representation of the transcription in mammals. The area of the box represents the genome. The area of large green circle is equivalent to the documented extent of transcription, with the darker green area corresponding to that on both strands. It should be noted that these estimates may and probably will increase as more information comes to hand. The function of most of these transcripts is unknown. CDSs are protein-coding sequences, and UTRs are 5′- and 3′-untranslated sequences in mRNAs. The dots indicate (and in fact overstate) the proportion of the genome occupied by known snoRNAs and miRNAs.

implications for the natural mechanisms of splicing control (8). A snoRNA has also been shown to control splicing of *serotonin receptor 5-HT(2C)R* mRNA (64). In addition, a significant number of ultra-conserved sequences in mammals and insects are located at splice sites (63,69). It should be borne in mind that some protein-coding sequences may have dual function, and be themselves the targets of regulatory molecules, such as miRNAs and siRNAs, as has been well documented in plants (70) and has been recently shown to occur in mammals (71–73). It should also be borne in mind that many RNAs may combine both digital (i.e. sequence-specific) and analog (structure-based ligand/protein binding or catalytic) functions, and that we have barely yet scratched the surface of these functions and networks.

LARGE NUMBERS OF ncRNAs EXPRESSED FROM THE MAMMALIAN GENOME

The Ensembl 34b version of Human Genome annotation lists 22 287 known or predicted protein-coding gene loci. The coding regions occupy ∼34 Mb (∼1.2%) of the euchromatic genome, and the total fraction of bases occupied by known protein-coding transcripts is only about 2% (15,74). However, summation of the sequences covered by known genes, 'mRNAs' and spliced ESTs indicates that (at least) 60–70% of the mammalian genome is transcribed on one or both strands (15,75), noting that introns are also actually transcribed (as distinct from generating stable transcripts) (Fig. 1). These estimates are conservative, as it is clear from both cDNA and genome tiling array studies that we have not yet come close to plumbing the full depth or breadth of the expressed transcripts in different types of cells under different developmental and physiological conditions (75–79).

Large-scale cDNA cloning studies have recently shown that there are many tens of thousands of transcripts expressed from
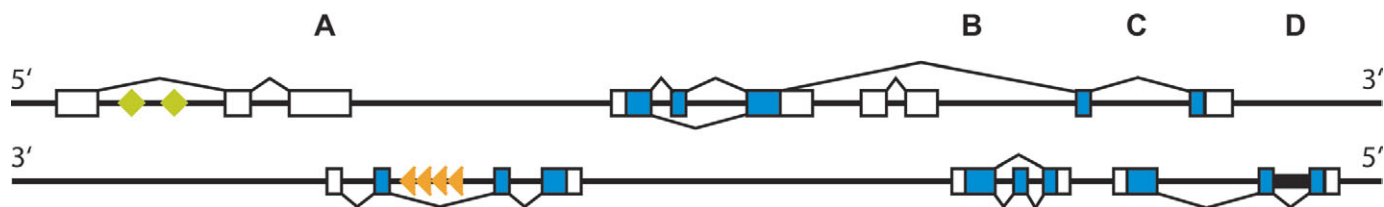
Figure 2. Graphical representation of the complexity of the transcriptional landscape in mammals. White boxes represent non-coding exonic sequences and dark blue boxes protein-coding exonic sequences. Green diamonds represent snoRNAs and orange triangles represent miRNAs. Indicated are (**A**) antisense transcripts with overlapping exons, (**B**) nested transcripts on both strands, (**C**) antisense transcripts with interlacing exons and (**D**) retained introns.

the mouse genome, a large fraction of which (over 34 000) do not appear to encode proteins (75). These studies involved aggressive normalization to enrich for rare transcripts, which introduces the possibility of contamination from pre-mRNA sequences (i.e. introns), but the findings were generally supported by the results of large-scale promoter/transcription start site mapping, suggesting that the observed transcriptional complexity of the genome is real and extends far beyond what had been previously imagined (75). It should be noted that most putative ncRNAs are expressed at lower levels than mRNAs, and many are rare, consistent with the suggestion that these RNAs mainly fulfil regulatory functions. It should also be noted that these studies, as is traditional, were orientated towards cytoplasmic polyA + RNA (75,76), for technical reasons (to exclude infrastructural RNAs and primary transcripts), on the assumption that nearly all transcripts are processed to polyadenylated RNAs that are exported to the cytoplasm for translation, which may not be correct.

It is also apparent that much of the mammalian genome is transcribed from both strands. It is estimated that 5880 human transcription clusters (22% of those analyzed) form sense–antisense pairs with most antisense transcripts being ncRNA (80), an arrangement that exhibits considerable evolutionary conservation between the human and pufferfish genomes (81). A detailed analysis of the mouse transcriptome indicated that 43 553 (72%) transcriptional units overlap with transcripts coming from opposite strand (82). In fact, there is evidence from spliced ESTs, annotated 'mRNAs' and protein-coding genes listed on the UCSC Genome Database (83) that at least 2.4 Gb of the human genome is transcribed, at least 25% from both strands (Fig. 1; M. Pheasant and J.S. Mattick, unpublished analysis). It would not be surprising if the true extent of transcription was greater than the size of the genome itself, noting that the upper limit is twice the genome size.

Genome tiling array (76,77) and massively parallel signature sequencing (MPSS) (78) studies of various tissues and cell lines have independently revealed many thousands of non-coding transcripts from intergenic and intronic sequences in the human genome. Over 37% of the MPSS signatures matched known loci, but outside of annotated exons, with another 20% matching the complementary strand of known transcripts, indicating the presence of as many as 50 000 additional non-annotated RNAs in analyzed human tissues (78). These findings are reinforced by the analysis of conserved RNA secondary structures which predict thousands of functional ncRNAs in the human genome (84,85).

High-density genome tiling array studies of 10 human chromosomes (approximately one-third of the human genome) showed that 9% of the non-repetitive sequences were expressed as detectable transcripts ('transfrags') in individual cell lines, and that 16.5% of non-repetitive bases were transcribed in at least one out of eight cell lines analyzed, indicating that many of the observed RNAs are cell-type specific (77), consistent with MPSS studies (78). It should be noted that this figure is much higher than the total length of all mRNAs expected from these chromosomes. Over 56% of the detected transfrags do not overlap with any well-characterized exon, mRNA or EST annotation; 30% map with 'intergenic' regions and 26% with introns of known genes. The latter do not appear to represent pre-mRNA contamination, as the signals were not generally spread across the introns, but rather showed discrete foci, indicative of previously unknown exons or of other RNAs (perhaps regulatory ncRNAs or their precursors) derived from these regions (77). Moreover, for technical reasons these analyses are likely to overlook many important small regulatory RNAs such as miRNAs which may be present in only trace amounts and are difficult to label by reverse transcription.

Rapid amplification of cDNA ends (RACE) analysis of selected genomic regions (79) confirmed the existence of these RNAs, and revealed an amazingly complex landscape of interlacing and overlapping transcripts, not only on opposite strands, but also on the same strand, so that there is often no clear distinction between splice variants and overlapping and neighboring genes, which had also been indicated by cDNA cloning studies (75,82). This study also showed that there are many hitherto unrecognized exons and splice variants even in very well-studied genes, such as that encoding Sonic Hedgehog, and that it is not unusual for a single base pair to be part of an intricate network of multiple isoforms of overlapping sense and antisense transcripts (Fig. 2). These observations all have important and challenging implications for genotype–phenotype correlations, the complexity of the transcriptional regulation, and the definition of a gene (79), which may now be best viewed as fuzzy transcription clusters with multiple products (18).

Just as disturbingly, it appears that almost a large proportion of the transcripts in human and mouse are unique to the largely unstudied polyA− and the nuclear polyA+ fractions of the transcriptome (77,86), which have escaped detection in most transcriptomic studies. It seems that we have barely begun to uncover the extraordinary complexity of the mammalian transcriptome.

TRANSCRIPTIONAL NOISE OR MEANINGFUL OUTPUT?

The observation that there are literally tens of thousands of ncRNAs expressed in mammals, and that most of the genome is transcribed, confronts and very largely contradicts the traditional protein-centric view of genetic information and genome organization. There are two opposing alternatives—either the bulk of the transcription which does not yield mRNAs is 'transcriptional noise' and/or (in the case of introns) the residue of evolutionary baggage retained or accumulated within genes, or this transcription comprises another level of expression and transaction of RNA information that is important to the evolution and developmental ontogeny of the higher organisms (13,16,18,23,87–90).

Most of the ncRNAs identified in genomic transcriptome studies have not been studied and have yet to be ascribed any function. However, there are many lines of evidence that suggest that these RNAs are biologically meaningful.

First, most intensively studied gene loci, including both those that are imprinted and conventional loci such as beta-globin, have been shown to express non-coding transcripts (91–96). This includes some enhancers and conserved intergenic sequences (92,97).

Second, it is clear that many of these transcripts are cell-type specific, with specific subcellular locations, and are developmentally regulated (77,98,99). A large number of ncRNAs are specifically expressed from either the paternal or maternal allele at imprinted loci, and some are associated with human diseases, such as the Prader–Willi and Angelman syndromes (39). Hence, the genetic cause for some, and perhaps many, diseases may be associated with mutations within ncRNAs. An imprinted ncRNA, *LANCAT*, spanning more than megabase in the murine region orthologous to the human Prader–Willi/Angelman syndrome locus, exhibits a distinct expression pattern in brain, as well as a cytoplasmic location (100). It has also been shown that some snoRNAs and miRNAs may be encoded within the introns of imprinted ncRNA genes (95,101). The snoRNA *HBII-52* which regulates the splicing of the serotonin receptor *5-HT(2C)R* gene is not expressed in Prader–Willi syndrome patients which have different *5-HT(2C)R* mRNA isoforms from normal, suggesting that this defect contributes to the Prader–Willi syndrome (64,102). Antisense transcripts associated with eight transcription factor genes involved in eye development also display specific expression patterns in brain, and in the retina in particular (103). Another non-coding antisense transcript, which has several alternatively spliced isoforms, shows an expression pattern similar to the sense-strand *Foxl2* gene, which encodes a forkhead transcription factor involved in development of eyelid and ovary (104).

Third, the upstream regions of ncRNA transcripts show many of the features normally associated with promoters (75,105,106) and, somewhat surprisingly, may be more highly conserved than the promoters of protein-coding genes (75). A recent large-scale study of the binding sites for the transcription factors, Sp1, cMyc and p53, found that a large proportion (36%) correlate with ncRNA transcripts, a significant number of which are regulated in response to retinoic acid, leading to the general conclusion that the human genome contains comparable numbers of protein-coding and non-coding genes that are bound by common transcription factors and regulated by common environmental signals (106).

Finally, an increasing number of ncRNAs have been shown to be functional, including the well-characterized ncRNAs *Xist* and *Tsix* that control X-chromosome inactivation in mammals (107,108). They also include a number of well-characterized antisense transcripts which appear to play regulatory roles in relation to their sense gene, including those opposite FGF-2 (fibroblast growth factor-2), HIF-1 (hypoxia inducible factor-1) and myosin heavy chain [for review see (109)]. Increasing numbers of functional studies of ncRNAs are being conducted using ectopic expression and RNAi-mediated knockdowns. For example, ectopic expression of the murine brain-specific ncRNA *SCA8*, which has been implicated in Spinocerebellar Ataxia Type 8 (110), under the control of a promoter specific to photoreceptors, results in late-onset, progressive neurodegeneration in the *Drosophila* eye (111). Moreover, using this neurodegenerative phenotype as a sensitized background for a genetic modifier screen, mutations were identified in four genes, all of which encode neuronally expressed RNA binding proteins conserved in *Drosophila* and humans (111). The knockdown by RNAi of a 6.7 kb spliced and polyadenylated murine ncRNA (*TUG1*) that is expressed in the retina and brain and upregulated by taurine in developing retinal cells RNA resulted in malformed or non-existent outer segments of transfected photoreceptors in mice (112).

This approach has recently been extended into large-scale screening strategies of ncRNAs. Pairs of siRNAs directed against 512 ncRNA sequences from the RIKEN Fantom2 mouse cDNA collection (113) were used to interrogate a battery of 12 cell-based reporter assays representing key cellular processes and signaling pathways (114). Eight functional ncRNAs were identified (114; J.B. Hogenesch and P.G. Schultz, personal communication), a good rate of return given the limited functional scope of the assays: six essential for cell viability, one repressor of Hedgehog signaling, and one (termed *NRON*) which acts as a repressor of the transcription factor NFAT, which itself is required for T-cell receptor-mediated immune response, and the development of the heart, vasculature, musculature and nervous tissue. *NRON* occurs as a variety of alternatively spliced transcripts ranging from 0.8 to 3.7 kb, and interacts with 11 different proteins, possibly as scaffolding for a complex including a translation initiation factor, RNA helicase and proteins involved in nucleocytoplasmic transport, proteolysis and signal transduction (114).

The number of known functional ncRNA genes has risen dramatically in recent years and over 800 ncRNAs (excluding tRNAs, rRNAs and snRNAs) have been catalogued in mammals, at least some of which are alternatively spliced (115,116). ncRNAs have also been implicated in many diseases, including various cancers and neurological diseases (18,115).

There is a rapidly looming nomenclature problem for the large number of ncRNAs (117), especially as the function and mode of action of the vast majority are unknown, and their complex structures and interlacing/overlapping nature make discrete classification difficult. As a considerable

fraction of eukaryotic transcripts are spliced, most approaches used, including cDNA cloning, detect only portions of transcripts, which often correspond to exons. Depending upon the method used these detected sites of transcription have been called an assortment of terms, such as ditags, CAGE tags, transfrags and ESTs, to mention a few. In some cases, experiments are used to connect these fragments into full-length or near full-length transcript structures [see e.g. (79)]. When transcripts are found to contain reduced protein-coding potential these have also been given various names including npcRNA (non-protein-coding RNA), utRNA (untranslated RNA) (117) or TUF (transcript of unknown function) (77). A structured system that may be used to catalog and refer to ncRNAs until they can be grouped and re-classified into recognized structural and/or functional classes is currently being considered by the HUGO Gene Nomenclature Committee (see http://www.gene.ucl.ac.uk/nomenclature/).

SMALL REGULATORY ncRNAs

The past few years have seen an explosion in the discovery of small regulatory RNAs in animals and plants (8,118–120) that, at present, largely fall into two classes: snoRNAs and miRNAs/siRNAs.

Small nucleolar RNAs

snoRNAs generally range from 60 to 300 nucleotides in length and guide the site-specific modification of nucleotides in target RNAs via short regions of base-pairing. There are two major classes, the box C/D snoRNAs which guide 2′-*O*-ribose-methylation, and the box H/ACA snoRNAs which guide pseudouridylation of target RNAs (36,121–123). Initially, it was thought that the role of snoRNAs was restricted to rRNA modification in ribosome biogenesis, but it is now evident that they can target other RNAs, including snRNAs and mRNAs (36,64,121–123). Most mammalian snoRNAs come from the introns of either protein-coding or non-coding genes (124) but apparently some human C/D snoRNAs are independently transcribed as indicated by the presence of methylated guanosine caps at their 5′ ends (125). Although the snoRNAs involved in ribosome biogenesis are located in the nucleolus where this type of ncRNA was first characterized (hence their name), a subset of H/ACA snoRNAs is located in Cajal bodies (a class of small nuclear organelle) and are sometimes called scaRNAs (small Cajal body RNAs) (36). Telomerase RNA is also found in Cajal bodies in a cell-cycle dependent manner (126,127).

At least some snoRNAs exhibit tissue-specific and developmental regulation, and/or imprinting (101,102,128,129), indicative of a regulatory function. There are also a number of so-called orphan snoRNAs without known targets (101,102,123,128,130,131). As noted earlier, one of these snoRNAs is linked to the aberrant splicing of the serotonin receptor *5-HT(2C)R* gene in Prader–Willi syndrome patients (64,102). It is also evident that there are many other snoRNAs, as well as likely, other as yet functionally uncharacterized classes of small regulatory RNAs, that have yet to be discovered (36,132).

MicroRNAs and small interfering RNAs

miRNAs and siRNAs are short, approximately 22 nucleotides long RNA molecules derived either from hairpin or double-stranded RNA precursors. Details of miRNA and siRNA biology and biochemistry can be found in a number of recent reviews (8,133–135). miRNAs suppress translation via non-perfect pairing with target mRNAs—usually involving a seed pairing of just six to eight nucleotides in length (56)—or (as with siRNAs) cause degradation of target RNAs by the RISC complex in the case of perfect complementarity with the target site—the phenomenon known as RNAi. It is estimated that approximately one-third of human protein-coding genes are controlled by miRNAs [reviewed in (119)]. In addition, siRNAs derived from repeats participate in the establishment of silenced (heterochromatic) chromatin, as well as in other aspects of chromosome dynamics, phenomena best studied in yeast [for reviews see (8,136)].

miRNAs are derived from the introns and exons of both protein-coding and non-coding transcripts that are synthesized by RNA polymerase II (8,137,138). It has also recently been shown that a number of mammalian miRNAs are derived from repeats, mainly various transposons (139), which may lead to a re-examination of the functional role of transposons, especially since it also appears that transposon sequences can play a significant role in the developmental processes and epigenetic variation (140,141). Some miRNAs also appear to be derived from processed pseudogenes (142).

The expression of many miRNAs is regulated and miRNAs have been shown to be central to a wide range of developmental processes, including developmental timing, cell proliferation, left–right patterning, neuronal cell fate, apoptosis and fat metabolism [for reviews see (8,133–135,143)], as well as neuronal gene expression (144), brain morphogenesis (145), muscle differentiation (146) and stem cell division (147). Not surprisingly, therefore, alterations in the expression, sequence or target sites for miRNAs may be a significant but hitherto unrecognized source of human genetic disease, including cancer. Sequence variants in the binding site for the miRNA *miR-189* in the *SLITRK1* mRNA have recently been shown to be associated with Tourette's syndrome (148). miRNA expression is dysregulated in cancer cells (143,149,150) and miRNA profiling can be used as a very accurate diagnostic tool for cancer classification (151,152). The proto-oncogene *c-Myc* has been shown to activate expression of an miRNA cluster on human chromosome 13, and two miRNAs (*miR-17-5p* and *miR-20a*) from this cluster downregulate expression of the transcription factor *E2F1* that activates cell cycle progression (153). Enforced expression of the same *miR-17-92* miRNA cluster has also been shown to promote tumor development (154), as has misexpression of the *Drosophila* miRNA *mirvana/mir-278* (155), indicating that some miRNAs may also function as proto-oncogenes.

Until recently, it was believed that the post-transcriptional suppression of gene expression by miRNA in vertebrates occurs through translation suppression directed by a non-perfect duplex formed between miRNA and mRNA in the 3′-UTR. However, in 2004, two groups described suppression of *HOX* gene expression by mRNA degradation because of a

perfect match between miRNA and mRNA in 3′-UTR (71,72). Another example of mRNA degradation because of a perfect match with a *trans*-acting miRNA has been reported for the imprinted *Rtl1/Peg11* locus (73). The maternally transcribed *anti-Peg11* transcript is processed into several miRNAs, which cause RISC-mediated cleavage of paternally expressed *Rtl1/Peg11* mRNA. Interestingly, the miRNAs are complementary to the coding region, not to the 3′-UTR (73), indicating that miRNA target sites may be located anywhere in the transcript, and indeed in any functional transcript, not just mRNAs. In addition, it has recently been shown that certain miRNA precursors are edited by ADAR1 and ADAR2, resulting in both suppression of processing by Drosha, and degradation by Tudor-SN, which is a component of RISC (156).

The miRBase database (http://microrna.sanger.ac.uk/) lists over 300 experimentally verified miRNAs in human as well as predicted miRNA target genes (157). However, many more miRNAs have been identified computationally, with a proportion validated *post hoc* (158). Most miRNA prediction methods rely on identification of a stable stem–loop precursor and phylogenetic conservation [see e.g. (158)]. However, these criteria may be far too narrow. Although many of the known miRNAs are highly conserved (and have been mainly identified on this basis), there is no reason why they all should be, as (as far as one can tell) these short RNAs have no intrinsic catalytic activity and function simply by target recognition, and thus should be able to evolve relatively quickly by co-variation with their targets, and by positive selection for new connections in regulatory networks underpinning adaptive radiation. Consistent with this, the known miRNAs appear to have many targets, thereby making co-variation difficult, and explaining their strong conservation, which in many cases surpasses that of protein-coding sequences (108). A recent study that did not require substantial evolutionary conservation identified many new human miRNAs, a significant number of which appear to be primate-specific (159).

The number of predicted human miRNAs is rising rapidly (8,135,159). Sensitive genetic screens in *C. elegans* have also identified rare miRNAs with limited evolutionary conservation such as *lys-6* which is required for left–right neuronal patterning (160), suggesting that many miRNAs may be cell-type specific and that many more remain to be found.

BIOLOGICAL ROLES OF ncRNAs

As outlined earlier, ncRNAs are already known to fulfill a wide range of functions, including the control of chromosome dynamics, splicing, RNA editing, translational inhibition and mRNA destruction. It is obvious that we have only begun to explore the true extent of RNA regulation of these processes. It also appears that RNA may play a role in virtually all levels of gene regulation in eukaryotes.

A range of evidence suggests that RNA signaling underpins chromatin remodeling and epigenetic memory, although the mechanisms are unknown, and the matter is not without controversy [for reviews and discussion see (8,18,35,161–163)].

There is evidence that transcription from upstream regions can affect the expression of the adjacent gene, either by promoter interference (164) or by altering chromatin structure (165–167), leading to the hypothesis that it is the act of transcription which is responsible for the regulatory effects, and that the transcript itself (an ncRNA) is just a by-product (168). However, it is hard to imagine how transcription *per se* could convey sufficient information to account for the precise and quite complex changes in histone modification and chromatin remodeling that are observed at most loci. Indeed, there are only a limited number of chromatin-modifying enzymes in animals, suggesting that these enzymes must be targeted to their sites of action, which vary at thousands of loci around the genome during differentiation and development, by another level of sequence-specific signals, most logically RNA. In agreement with this prediction small RNAs have been shown to induce transcriptional silencing and alterations to DNA methylation in human cells (169,170).

There are also good reasons to expect that splicing is regulated, at least in part, by *trans*-acting RNAs that guide splice site selection (8,18,64,171) or modify sequences around splice sites to render them accessible or otherwise to the splicing machinery (64).

Evidence is also emerging that transcription itself may be regulated by ncRNAs (18,163). As noted earlier, RNA polymerase II itself appears to be regulated in part by ncRNA signaling (30–34). A ncRNA has been reported to be required for the repression of RNA polymerase II-dependent transcription in primordial germ cells in *Drosophila* (172). At least some transcription factors (and chromatin-modifying proteins) appear to have affinity for structures involving RNA (173–179). A small double-stranded RNA termed *NRSE* activates transcription of neuron-specific genes (180) and short artificial RNAs have been shown to inhibit transcription of targeted genes in the absence of concomitant DNA methylation, with considerable potential for therapeutic use (181,182). An interesting case is the *steroid receptor RNA activator* (*SRA*) which was originally described as functional non-coding RNA involved in the regulation of gene expression by steroid hormones (183). The gene produces several transcripts of which one encodes a protein (184) and both the ncRNA and its encoded protein affects the activity of estrogen receptor in breast cancer cells (185). Recently, it was shown that pseudo-uridine synthase mPus1p (an enzyme that converts uridine to pseudouridine in RNA) is a coactivator for the retinoic acid receptor, which acts by pseudouridinilation of *SRA* RNA (186). In addition, the thyroid hormone receptor has an RNA-binding domain which binds *SRA*, and the binding enhances expression of reporter genes (187).

ncRNAs also play a role in stress responses. The small non-coding transcript *B2* is produced by RNA polymerase III from murine short interspersed elements (SINE) under heat shock. The *B2* RNA binds to RNA polymerase II and represses transcription after heat shock (188,189). In primates, RNA polymerase III also produces the brain-specific *Alu*-derived transcript *BC200* (190). Non-coding repetitive RNAs are also transcribed in stressed human cells and are localized in 'nuclear stress bodies' that are assembled on specific pericentromeric heterochromatic domains that change their epigenetic

status from heterochromatin to euchromatin in response to stress (191). The non-coding RNA *omega* is among few heat-shock-inducible genes in *Drosophila* (192), and although its exact role is unknown, it binds to a number of RNA-binding proteins involved in processing of nuclear RNA (hnRNPs complexes) (193).

ncRNAs may also act as scaffolding for the assembly of macromolecular complexes. Examples include rRNA in ribosomes, the *7SL* RNA in the SRP (40), and possibly RNAs involved in the assembly of chromatin complexes (35), as well as *NRON*, recently shown to interact with a number of proteins involved in nuclear transcription factor trafficking (114).

INTRONS AS A SOURCE OF FUNCTIONAL ncRNAs

Introns account for at least 30% of the human genome and may be a significant, perhaps major, source of regulatory ncRNAs (17,87), produced in parallel with protein-coding sequences (and others) as efference signals to convey regulatory information to other genes and transcripts (16,18). Almost all snoRNAs and a large proportion of miRNAs in animals are encoded in introns (138,194–196), located in both protein-coding and non-protein-coding genes [for review see (8)]. Although introns are thought to be simply degraded after being excised from primary transcripts, there is good evidence that intronic RNAs may actually be processed to smaller RNAs (which were not anticipated or detected when introns were first studied) with significant half-lives and specific subcellular locations (197,198). Recently, it was shown that ectopic expression of intronic sequences derived from the *CFTR* gene causes specific changes in transcription of various genes in HeLa cells (199). Interestingly, each of the three intron sequences tested resulted in a distinctive pattern of effects on specific subsets of genes (199). The idea that introns may be a rich source of regulatory information is consistent with the fact that the density of introns scales with developmental complexity (87), and many highly conserved sequences, including ultraconserved sequences, are found in introns (69,200–203). However, at present, it is simply not known what proportion of transcribed introns are subsequently processed into smaller functional RNAs, although many intronic sequences are detected in whole genome tiling array analyses of human transcription (77).

CONCLUSION

We may have fundamentally misunderstood the nature of genetic programming in the higher organisms. It appears that the human genome and those of other complex organisms express an enormous repertoire of ncRNAs, and that their cells are awash with these RNAs, which constitute a hidden layer of molecular genetic signals. Although the functions of these RNAs are likely to be many and varied, both logic and evidence strongly suggest that their main role is to regulate and direct the complex pathways of developmental ontogeny, which must require enormous amounts of information in an organism as precisely sculptured as a human (13).

The existence of a sophisticated RNA-based regulatory system would also largely explain the paradox of the tremendous diversity of characteristics observed among mammals and other complex organisms, despite the relative commonality of their proteomes. That such RNAs have remained hidden from view for so long appears to have been a consequence of their sheer numbers and population complexity which makes biochemical detection of individual sequences difficult, combined with the subtlety of their genetic signatures. Indeed, with few exceptions, until recently most known ncRNAs were those that are present in relatively large amounts, such as rRNAs, tRNAs and the common snoRNAs and snRNAs, and it has only been the combination of sensitive genetic screens (such as those that first identified miRNAs), large-scale cDNA and whole genome sequencing, new sensitive analytical methods (such as RT–PCR and genome tiling arrays) and bioinformatics, based on clues from known examples, that has begun to reveal the true complexity of what lies under the surface.

It is also evident that many ncRNAs, including those of demonstrated functionality like *Xist*, are evolving quickly (108). This rapid evolution has been considered as evidence of lack of functionality (204). This may be incorrect, and these sequences may in fact be simply able to drift easily because of different constraints and/or be subject to positive selection related to phenotypic variation. Recent analyses of the *Drosophila* genome have indicated that, contrary to long-held expectation, a large fraction of the non-coding sequence is functionally important and subject to various levels of purifying selection and adaptive evolution (205).

The extent of non-coding sequence conservation in mammals is also much higher than that of protein-coding sequences (202,206), perhaps as high as 10% by some estimates (207). This conservation includes ultraconserved sequences (69) and long transposon-free regions that have remained refractory to transposon insertions throughout mammalian evolution (208), observations which are difficult to reconcile with orthodox protein-based conceptions of gene regulation. As noted earlier, there is increasing evidence that transposon-derived sequences may also contribute to mammalian genetic activity. Indeed, it may be that much, if not most of the sequences comprising the human genome are functional, albeit having arrived at different times in our evolutionary history and be evolving at different rates.

The problem has been compounded by the fact that most mutations in regulatory sequences may be both subtle and difficult to track, particularly given the expectational and practical bias to date in genome scanning projects on exonic lamp-posts of protein-coding genes, and the fact that the relevant mutations may be quite distal to these lamp-posts, hidden in the dark of the vast tracts of intergenic and intronic sequences. The mutations underlying the callipyge ('beautiful bottom') phenotype in sheep and the enhanced muscling of domestic pigs, which are single base substitutions within non-coding sequences (a long intergenic sequence of unknown transcriptional status in the *DLK1-GTL2* imprinted region, and the third intron of the *IGF2* gene, respectively), the identification of which involved tour-de-force analyses in well structured pedigrees (209–211).

It is clear that different types of genetic information will be subject to different structure–function relationships and

therefore different constraints on their variation related to their role and the number of interacting partners. We predict that mutations/variations in many if not most ncRNA sequences, especially those that are involved in regulatory networks, will lead to a variety of milder phenotypes than the usually severe consequences of mutations in proteins, and will have a major influence on quantitative trait variation, developmental differences and abnormalities, cancer and other complex diseases such as neurological disorders.

The functional genomics of ncRNAs will be a daunting task, an equal or greater challenge than that we already face in working out the biochemical functions and biological roles of all of the known and predicted proteins and their isoforms (212). Bioinformatics will be key, as it should be possible to use sequence homology (albeit in small patches, and obeying a rather broader set of rules than simply Watson–Crick DNA base pairing) to identify transmitters and their receivers in RNA regulatory networks, as is already the case for miRNAs. This also means that it should be possible to develop generic approaches, applicable to any regulatory RNA or its target, to intersect and modulate gene activity at various levels for therapeutic purposes, which may revolutionize the pharmaceutical industry. The advent of large-scale whole genome (re-)sequencing, which is at an advanced stage of development (213,214), while creating enormous informatic challenges, will soon also provide the density of genomic data required to identify sequences directly associated with different characteristics in structured populations, without assumptions about the genomic position of these sequences or their mode of action.

ACKNOWLEDGEMENTS

Conflict of Interest statement. None declared.

REFERENCES

1. Gesteland, R.F., Cech, T.R. and Atkins, J.F. (eds) (2006) *The RNA World*, 3rd edn. Cold Spring Harbor Laboratory Press.
2. Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.*, **3**, 318–356.
3. Rivas, E., Klein, R.J., Jones, T.A. and Eddy, S.R. (2001) Computational identification of non-coding RNAs in *E. coli* by comparative genomics. *Curr. Biol.*, **11**, 1369–1373.
4. Vogel, J., Bartels, V., Tang, T.H., Churakov, G., Slagter-Jager, J.G., Huttenhofer, A. and Wagner, E.G. (2003) RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res.*, **31**, 6435–6443.
5. Gottesman, S. (2004) The small RNA regulators of *Escherichia coli*: roles and mechanisms. *Annu. Rev. Microbiol.*, **58**, 303–328.
6. Storz, G., Opdyke, J.A. and Zhang, A. (2004) Controlling mRNA stability and translation with small, non-coding RNAs. *Curr. Opin. Microbiol.*, **7**, 140–144.
7. Kawano, M., Reynolds, A.A., Miranda-Rios, J. and Storz, G. (2005) Detection of 5′- and 3′-UTR-derived small RNAs and *cis*-encoded antisense RNAs in *Escherichia coli. Nucleic Acids Res.*, **33**, 1040–1050.
8. Mattick, J.S. and Makunin, I.V. (2005) Small regulatory RNAs in mammals. *Hum. Mol. Genet.*, **14**, R121–R132.
9. Wilderman, P.J., Sowa, N.A., FitzGerald, D.J., FitzGerald, P.C., Gottesman, S., Ochsner, U.A. and Vasil, M.L. (2004) Identification of tandem duplicate regulatory small RNAs in *Pseudomonas aeruginosa* involved in iron homeostasis. *Proc. Natl Acad. Sci. USA*, **101**, 9792–9797.
10. Axmann, I.M., Kensche, P., Vogel, J., Kohl, S., Herzel, H. and Hess, W.R. (2005) Identification of cyanobacterial non-coding RNAs by comparative genome analysis. *Genome Biol.*, **6**, R73.
11. Dennis, P.P. and Omer, A. (2005) Small non-coding RNAs in Archaea. *Curr. Opin. Microbiol.*, **8**, 685–694.
12. Hall, T.M. (2005) Structure and function of argonaute proteins. *Structure*, **13**, 1403–1408.
13. Mattick, J.S. (2004) RNA regulation: a new genetics? *Nat. Rev. Genet.*, **5**, 316–323.
14. Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.G., Ohtsubo, E., Nakayama, K., Murata, T. *et al.* (2001) Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res.*, **8**, 11–22.
15. Frith, M.C., Pheasant, M. and Mattick, J.S. (2005) The amazing complexity of the human transcriptome. *Eur. J. Hum. Genet.*, **13**, 894–897.
16. Mattick, J.S. (2001) Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep.*, **2**, 986–991.
17. Mattick, J.S. and Gagen, M.J. (2001) The evolution of controlled multitasked gene networks: the role of introns and other non-coding RNAs in the development of complex organisms. *Mol. Biol. Evol.*, **18**, 1611–1630.
18. Mattick, J.S. (2003) Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays*, **25**, 930–939.
19. Mattick, J.S. and Gagen, M.J. (2005) Accelerating networks. *Science*, **307**, 856–858.
20. Croft, L.J., Lercher, M.J., Gagen, M.J. and Mattick, J.S. (2003) Is prokaryotic complexity limited by accelerated growth in regulatory overhead? *Genome Biology Preprint Depository*, http://genomebiology.com/qc/2003/5/1/p2.
21. van Nimwegen, E. (2003) Scaling laws in the functional content of genomes. *Trends Genet.*, **19**, 479–484.
22. Gagen, M.J. and Mattick, J.S. (2005) Inherent size constraints on prokaryote gene networks due to 'accelerating' growth. *Theory Biosci.*, **123**, 381–411.
23. Claverie, J.M. (2005) Fewer genes, more non-coding RNA. *Science*, **309**, 1529–1530.
24. Steitz, T.A. and Moore, P.B. (2003) RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem. Sci.*, **28**, 411–418.
25. Noller, H.F. (2005) RNA structure: reading the ribosome. *Science*, **309**, 1508–1514.
26. Nilsen, T.W. (2003) The spliceosome: the most complex macromolecular machine in the cell? *Bioessays*, **25**, 1147–1149.
27. Butcher, S.E. and Brow, D.A. (2005) Towards understanding the catalytic core structure of the spliceosome. *Biochem. Soc. Trans.*, **33**, 447–449.
28. Kwek, K.Y., Murphy, S., Furger, A., Thomas, B., O'Gorman, W., Kimura, H., Proudfoot, N.J. and Akoulitchev, A. (2002) U1 snRNA associates with TFIIH and regulates transcriptional initiation. *Nat. Struct. Biol.*, **9**, 800–805.
29. O'Gorman, W., Thomas, B., Kwek, K.Y., Furger, A. and Akoulitchev, A. (2005) Analysis of U1 snRNA interaction with cyclin H. *J. Biol. Chem.*, **280**, 36920–36925.
30. Yang, Z., Zhu, Q., Luo, K. and Zhou, Q. (2001) The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature*, **414**, 317–322.
31. Michels, A.A., Fraldi, A., Li, Q., Adamson, T.E., Bonnet, F., Nguyen, V.T., Sedore, S.C., Price, J.P., Price, D.H., Lania, L. *et al.* (2004) Binding of the 7SK snRNA turns the HEXIM1 protein into a P-TEFb (CDK9/cyclin T) inhibitor. *EMBO J.*, **23**, 2608–2619.
32. Yik, J.H., Chen, R., Pezda, A.C. and Zhou, Q. (2005) Compensatory contributions of HEXIM1 and HEXIM2 in maintaining the balance of active and inactive positive transcription elongation factor b complexes for control of transcription. *J. Biol. Chem.*, **280**, 16368–16376.

33. Li, Q., Price, J.P., Byers, S.A., Cheng, D., Peng, J. and Price, D.H. (2005) Analysis of the large inactive P-TEFb complex indicates that it contains one 7SK molecule, a dimer of HEXIM1 or HEXIM2, and two P-TEFb molecules containing Cdk9 phosphorylated at threonine 186. *J. Biol. Chem.*, **280**, 28819–28826.

34. Haaland, R.E., Herrmann, C.H. and Rice, A.P. (2005) siRNA depletion of 7SK snRNA induces apoptosis but does not affect expression of the HIV-1 LTR or P-TEFb-dependent cellular genes. *J. Cell Physiol.*, **205**, 463–470.

35. Bernstein, E. and Allis, C.D. (2005) RNA meets chromatin. *Genes Dev.*, **19**, 1635–1655.

36. Meier, U.T. (2005) The many facets of H/ACA ribonucleoproteins. *Chromosoma*, **114**, 1–14.

37. Vulliamy, T., Marrone, A., Goldman, F., Dearlove, A., Bessler, M., Mason, P.J. and Dokal, I. (2001) The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. *Nature*, **413**, 432–435.

38. Cam, H.P., Sugiyama, T., Chen, E.S., Chen, X., FitzGerald, P.C. and Grewal, S.I. (2005) Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nat. Genet.*, **37**, 809–819.

39. Morison, I.M., Ramsay, J.P. and Spencer, H.G. (2005) A census of mammalian imprinting. *Trends Genet.*, **21**, 457–465.

40. Nagai, K., Oubridge, C., Kuglstatter, A., Menichelli, E., Isel, C. and Jovine, L. (2003) Structure, function and evolution of the signal recognition particle. *EMBO J.*, **22**, 3479–3485.

41. Doudna, J.A. and Batey, R.T. (2004) Structural insights into the signal recognition particle. *Annu. Rev. Biochem.*, **73**, 539–557.

42. Wild, K., Halic, M., Sinning, I. and Beckmann, R. (2004) SRP meets the ribosome. *Nat. Struct. Mol. Biol.*, **11**, 1049–1053.

43. Halic, M. and Beckmann, R. (2005) The signal recognition particle and its interactions during protein targeting. *Curr. Opin. Struct. Biol.*, **15**, 116–125.

44. van Zon, A., Mossink, M.H., Scheper, R.J., Sonneveld, P. and Wiemer, E.A. (2003) The vault complex. *Cell Mol. Life Sci.*, **60**, 1828–1837.

45. Gopinath, S.C., Matsugami, A., Katahira, M. and Kumar, P.K. (2005) Human vault-associated non-coding RNAs bind to mitoxantrone, a chemotherapeutic compound. *Nucleic Acids Res.*, **33**, 4874–4881.

46. Kuersten, S. and Goodwin, E.B. (2003) The power of the 3′ UTR: translational control and development. *Nat. Rev. Genet.*, **4**, 626–637.

47. Gebauer, F. and Hentze, M.W. (2004) Molecular mechanisms of translational control. *Nat. Rev. Mol. Cell. Biol.*, **5**, 827–835.

48. Moore, M.J. (2005) From birth to death: the complex lives of eukaryotic mRNAs. *Science*, **309**, 1514–1518.

49. Prasanth, K.V., Prasanth, S.G., Xuan, Z., Hearn, S., Freier, S.M., Bennett, C.F., Zhang, M.Q. and Spector, D.L. (2005) Regulating gene expression through RNA nuclear retention. *Cell*, **123**, 249–263.

50. Vitreschak, A.G., Rodionov, D.A., Mironov, A.A. and Gelfand, M.S. (2004) Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends Genet.*, **20**, 44–50.

51. Tucker, B.J. and Breaker, R.R. (2005) Riboswitches as versatile gene control elements. *Curr. Opin. Struct. Biol.*, **15**, 342–348.

52. Winkler, W.C. (2005) Riboswitches and the role of non-coding RNAs in bacterial metabolic control. *Curr. Opin. Chem. Biol.*, **9**, 594–602.

53. Sudarsan, N., Barrick, J.E. and Breaker, R.R. (2003) Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA*, **9**, 644–647.

54. Kubodera, T., Watanabe, M., Yoshiuchi, K., Yamashita, N., Nishimura, A., Nakai, S., Gomi, K. and Hanamoto, H. (2003) Thiamine-regulated gene expression of *Aspergillus oryzae* thiA requires splicing of the intron containing a riboswitch-like domain in the 5′-UTR. *FEBS Lett.*, **555**, 516–520.

55. Vella, M.C., Choi, E.Y., Lin, S.Y., Reinert, K. and Slack, F.J. (2004) The *C. elegans* microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3′UTR. *Genes Dev.*, **18**, 132–137.

56. Lewis, B.P., Burge, C.B. and Bartel, D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15–20.

57. Farh, K.K., Grimson, A., Jan, C., Lewis, B.P., Johnston, W.K., Lim, L.P., Burge, C.B. and Bartel, D.P. (2005) The widespread impact of mammalian microRNAs on mRNA repression and evolution. *Science*, **310**, 1817–1821.

58. Wang, J., Smith, P.J., Krainer, A.R. and Zhang, M.Q. (2005) Distribution of SR protein exonic splicing enhancer motifs in human protein-coding genes. *Nucleic Acids Res.*, **33**, 5053–5062.

59. Sorek, R. and Ast, G. (2003) Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. *Genome Res.*, **13**, 1631–1637.

60. Sugnet, C.W., Kent, W.J., Ares, M., Jr. and Haussler, D. (2004) Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pac. Symp. Biocomput.*, 66–77.

61. Sugnet, C.W., Srinivasan, K., Clark, T.A., O'Brien, G., Cline, M.S., Wang, H., Williams, A., Kulp, D., Blume, J.E., Haussler, D. *et al.* (2006) Unusual intron conservation near tissue-regulated exons found by splicing microarrays. *PLoS Comput. Biol.*, **2**, e4.

62. Graveley, B.R. (2005) Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell*, **123**, 65–73.

63. Glazov, E.A., Pheasant, M., McGraw, E.A., Bejerano, G. and Mattick, J.S. (2005) Ultraconserved elements in insect genomes: a highly conserved intronic sequence implicated in the control of homothorax mRNA splicing. *Genome Res.*, **15**, 800–808.

64. Kishore, S. and Stamm, S. (2006) The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science*, **311**, 230–232.

65. Sazani, P. and Kole, R. (2003) Therapeutic potential of antisense oligonucleotides as modulators of alternative splicing. *J. Clin. Invest.*, **112**, 481–486.

66. Gebski, B.L., Mann, C.J., Fletcher, S. and Wilton, S.D. (2003) Morpholino antisense oligonucleotide induced dystrophin exon 23 skipping in mdx mouse muscle. *Hum. Mol. Genet.*, **12**, 1801–1811.

67. Kole, R., Vacek, M. and Williams, T. (2004) Modification of alternative splicing by antisense therapeutics. *Oligonucleotides*, **14**, 65–74.

68. Kim, D.S., Gusti, V., Pillai, S.G. and Gaur, R.K. (2005) An artificial riboswitch for controlling pre-mRNA splicing. *RNA*, **11**, 1667–1677.

69. Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S. and Haussler, D. (2004) Ultraconserved elements in the human genome. *Science*, **304**, 1321–1325.

70. Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.

71. Yekta, S., Shih, I.H. and Bartel, D.P. (2004) MicroRNA-directed cleavage of HOXB8 mRNA. *Science*, **304**, 594–596.

72. Mansfield, J.H., Harfe, B.D., Nissen, R., Obenauer, J., Srineel, J., Chaudhuri, A., Farzan-Kashani, R., Zuker, M., Pasquinelli, A.E., Ruvkun, G. *et al.* (2004) MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression. *Nat. Genet.*, **36**, 1079–1083.

73. Davis, E., Caiment, F., Tordoir, X., Cavaille, J., Ferguson-Smith, A., Cockett, N., Georges, M. and Charlier, C. (2005) RNAi-mediated allelic *trans*-interaction at the imprinted Rtl1/Peg11 locus. *Curr. Biol.*, **15**, 743–749.

74. Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature*, **431**, 931–945.

75. Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C. *et al.* (2005) The transcriptional landscape of the mammalian genome. *Science*, **309**, 1559–1563.

76. Bertone, P., Stolc, V., Royce, T.E., Rozowsky, J.S., Urban, A.E., Zhu, X., Rinn, J.L., Tongprasit, W., Samanta, M., Weissman, S. *et al.* (2004) Global identification of human transcribed sequences with genome tiling arrays. *Science*, **306**, 2242–2246.

77. Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G. *et al.* (2005) Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science*, **308**, 1149–1154.

78. Jongeneel, C.V., Delorenzi, M., Iseli, C., Zhou, D., Haudenschild, C.D., Khrebtukova, I., Kuznetsov, D., Stevenson, B.J., Strausberg, R.L., Simpson, A.J. *et al.* (2005) An atlas of human gene expression from massively parallel signature sequencing (MPSS). *Genome Res.*, **15**, 1007–1014.

79. Kapranov, P., Drenkow, J., Cheng, J., Long, J., Helt, G., Dike, S. and Gingeras, T.R. (2005) Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res.*, **15**, 987–997.

80. Chen, J., Sun, M., Kent, W.J., Huang, X., Xie, H., Wang, W., Zhou, G., Shi, R.Z. and Rowley, J.D. (2004) Over 20% of human transcripts might form sense-antisense pairs. *Nucleic Acids Res.*, **32**, 4812–4820.

81. Dahary, D., Elroy-Stein, O. and Sorek, R. (2005) Naturally occurring antisense: transcriptional leakage or real overlap? *Genome Res.*, **15**, 364–368.

82. Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J. *et al.* (2005) Antisense transcription in the mammalian transcriptome. *Science*, **309**, 1564–1566.

83. Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F. *et al.* (2006) The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.*, **34**, D590–D598.

84. Washietl, S., Hofacker, I.L. and Stadler, P.F. (2005) Fast and reliable prediction of non-coding RNAs. *Proc. Natl Acad. Sci. USA*, **102**, 2454–2459.

85. Washietl, S., Hofacker, I.L., Lukasser, M., Huttenhofer, A. and Stadler, P.F. (2005) Mapping of conserved RNA secondary structures predicts thousands of functional non-coding RNAs in the human genome. *Nat. Biotechnol.*, **23**, 1383–1390.

86. Kiyosawa, H., Mise, N., Iwase, S., Hayashizaki, Y. and Abe, K. (2005) Disclosing hidden transcripts: mouse natural sense-antisense transcripts tend to be poly(A) negative and nuclear localized. *Genome Res.*, **15**, 463–474.

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

88. Dennis, C. (2002) The brave new world of RNA. *Nature*, **418**, 122–124.

89. Huttenhofer, A., Schattner, P. and Polacek, N. (2005) Non-coding RNAs: hope or hype? *Trends Genet.*, **21**, 289–297.

90. Werner, A. and Berdal, A. (2005) Natural antisense transcripts: sound or silence? *Physiol. Genomics*, **23**, 125–131.

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

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

93. Charlier, C., Segers, K., Wagenaar, D., Karim, L., Berghmans, S., Jaillon, O., Shay, T., Weissenbach, J., Cockett, N., Gyapay, G. *et al.* (2001) Human-ovine comparative sequencing of a 250-kb imprinted domain encompassing the callipyge (clpg) locus and identification of six imprinted transcripts: DLK1, DAT, GTL2, PEG11, antiPEG11, and MEG8. *Genome Res.*, **11**, 850–862.

94. Holmes, R., Williamson, C., Peters, J., Denny, P. and Wells, C. (2003) A comprehensive transcript map of the mouse Gnas imprinted complex. *Genome Res.*, **13**, 1410–1415.

95. Seitz, H., Youngson, N., Lin, S.P., Dalbert, S., Paulsen, M., Bachellerie, J.P., Ferguson-Smith, A.C. and Cavaille, J. (2003) Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. *Nat. Genet.*, **34**, 261–262.

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

97. Jones, E.A. and Flavell, R.A. (2005) Distal enhancer elements transcribe intergenic RNA in the IL-10 family gene cluster. *J. Immunol.*, **175**, 7437–7446.

98. Blackshaw, S., Harpavat, S., Trimarchi, J., Cai, L., Huang, H., Kuo, W.P., Weber, G., Lee, K., Fraioli, R.E., Cho, S.H. *et al.* (2004) Genomic analysis of mouse retinal development. *PLoS Biol.*, **2**, E247.

99. Ravasi, T., Suzuki, H., Pang, K.C., Katayama, S., Furuno, M., Okunishi, R., Fukuda, S., Ru, K., Frith, M.C., Gongora, M.M. *et al.* (2006) Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res.*, **16**, 11–19.

100. Le Meur, E., Watrin, F., Landers, M., Sturny, R., Lalande, M. and Muscatelli, F. (2005) Dynamic developmental regulation of the large non-coding RNA associated with the mouse 7C imprinted chromosomal region. *Dev. Biol.*, **286**, 587–600.

101. Cavaille, J., Seitz, H., Paulsen, M., Ferguson-Smith, A.C. and Bachellerie, J.P. (2002) Identification of tandemly repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. *Hum. Mol. Genet.*, **11**, 1527–1538.

102. Cavaille, J., Buiting, K., Kiefmann, M., Lalande, M., Brannan, C.I., Horsthemke, B., Bachellerie, J.P., Brosius, J. and Huttenhofer, A. (2000) Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Natl Acad. Sci. USA*, **97**, 14311–14316.

103. Alfano, G., Vitiello, C., Caccioppoli, C., Caramico, T., Carola, A., Szego, M.J., McInnes, R.R., Auricchio, A. and Banfi, S. (2005) Natural antisense transcripts associated with genes involved in eye development. *Hum. Mol. Genet.*, **14**, 913–923.

104. Cocquet, J., Pannetier, M., Fellous, M. and Veitia, R.A. (2005) Sense and antisense Foxl2 transcripts in mouse. *Genomics*, **85**, 531–541.

105. Gagnon, M.L., Moy, G.K. and Klagsbrun, M. (1999) Characterization of the promoter for the human antisense fibroblast growth factor-2 gene; regulation by Ets in Jurkat T cells. *J. Cell. Biochem.*, **72**, 492–506.

106. Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J. *et al.* (2004) Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of non-coding RNAs. *Cell*, **116**, 499–509.

107. Chureau, C., Prissette, M., Bourdet, A., Barbe, V., Cattolico, L., Jones, L., Eggen, A., Avner, P. and Duret, L. (2002) Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine. *Genome Res.*, **12**, 894–908.

108. Pang, K.C., Frith, M.C. and Mattick, J.S. (2006) Rapid evolution of non-coding RNAs: lack of conservation does not mean lack of function. *Trends Genet.*, **22**, 1–5.

109. Werner, A. (2005) Natural antisense transcripts. *RNA Biol.*, **2**, 53–62.

110. Mutsuddi, M. and Rebay, I. (2005) Molecular genetics of spinocerebellar ataxia type 8 (SCA8). *RNA Biol.*, **2**, 49–52.

111. Mutsuddi, M., Marshall, C.M., Benzow, K.A., Koob, M.D. and Rebay, I. (2004) The spinocerebellar ataxia 8 non-coding RNA causes neurodegeneration and associates with staufen in *Drosophila*. *Curr. Biol.*, **14**, 302–308.

112. Young, T.L., Matsuda, T. and Cepko, C.L. (2005) The non-coding RNA taurine upregulated gene 1 is required for differentiation of the murine retina. *Curr. Biol.*, **15**, 501–512.

113. Numata, K., Kanai, A., Saito, R., Kondo, S., Adachi, J., Wilming, L.G., Hume, D.A., Hayashizaki, Y. and Tomita, M. (2003) Identification of putative non-coding RNAs among the RIKEN mouse full-length cDNA collection. *Genome Res.*, **13**, 1301–1306.

114. Willingham, A.T., Orth, A.P., Batalov, S., Peters, E.C., Wen, B.G., Aza-Blanc, P., Hogenesch, J.B. and Schultz, P.G. (2005) A strategy for probing the function of non-coding RNAs finds a repressor of NFAT. *Science*, **309**, 1570–1573.

115. Pang, K.C., Stephen, S., Engström, P.G., Tajul-Arifin, K., Chen, W., Wahlestedt, C., Lenhard, B., Hayashizaki, Y. and Mattick, J.S. (2005) RNAdb—a comprehensive mammalian non-coding RNA database. *Nucleic Acids Res.*, **33** (database issue), D125–D130.

116. Liu, C., Bai, B., Skogerbo, G., Cai, L., Deng, W., Zhang, Y., Bu, D., Zhao, Y. and Chen, R. (2005) NONCODE: an integrated knowledge database of non-coding RNAs. *Nucleic Acids Res.*, **33** (database issue), D112–D115.

117. Brosius, J. and Tiedge, H. (2004) RNomenclature. *RNA Biol.*, **1**, 81–83.

118. Storz, G., Altuvia, S. and Wassarman, K.M. (2005) An abundance of RNA regulators. *Annu. Rev. Biochem.*, **74**, 199–217.

119. Du, T. and Zamore, P.D. (2005) microPrimer: the biogenesis and function of microRNA. *Development*, **132**, 4645–4652.

120. Lu, C., Tej, S.S., Luo, S., Haudenschild, C.D., Meyers, B.C. and Green, P.J. (2005) Elucidation of the small RNA component of the transcriptome. *Science*, **309**, 1567–1569.

121. Bachellerie, J.P., Cavaille, J. and Huttenhofer, A. (2002) The expanding snoRNA world. *Biochimie*, **84**, 775–790.

122. Henras, A.K., Dez, C. and Henry, Y. (2004) RNA structure and function in C/D and H/ACA s(no)RNPs. *Curr. Opin. Struct. Biol.*, **14**, 335–343.

123. Kiss, A.M., Jady, B.E., Bertrand, E. and Kiss, T. (2004) Human box H/ACA pseudouridylation guide RNA machinery. *Mol. Cell. Biol.*, **24**, 5797–5807.

124. Kiss, T. (2002) Small nucleolar RNAs: an abundant group of non-coding RNAs with diverse cellular functions. *Cell*, **109**, 145–148.

125. Tycowski, K.T., Aab, A. and Steitz, J.A. (2004) Guide RNAs with 5′ caps and novel box C/D snoRNA-like domains for modification of snRNAs in metazoa. *Curr. Biol.*, **14**, 1985–1995.

126. Jady, B.E., Darzacq, X., Tucker, K.E., Matera, A.G., Bertrand, E. and Kiss, T. (2003) Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal body following import from the cytoplasm. *EMBO J.*, **22**, 1878–1888.

127. Jady, B.E., Bertrand, E. and Kiss, T. (2004) Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal. *J. Cell. Biol.*, **164**, 647–652.

128. Cavaille, J., Vitali, P., Basyuk, E., Huttenhofer, A. and Bachellerie, J.P. (2001) A novel brain-specific box C/D small nucleolar RNA processed from tandemly repeated introns of a non-coding RNA gene in rats. *J. Biol. Chem.*, **276**, 26374–26383.

129. Rogelj, B. and Giese, K.P. (2004) Expression and function of brain specific small RNAs. *Rev. Neurosci.*, **15**, 185–198.

130. Huttenhofer, A., Kiefmann, M., Meier-Ewert, S., O'Brien, J., Lehrach, H., Bachellerie, J.P. and Brosius, J. (2001) RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J.*, **20**, 2943–2953.

131. Vitali, P., Royo, H., Seitz, H., Bachellerie, J.P., Huttenhofer, A. and Cavaille, J. (2003) Identification of 13 novel human modification guide RNAs. *Nucleic Acids Res.*, **31**, 6543–6551.

132. Deng, W., Zhu, X., Skogerbo, G., Zhao, Y., Fu, Z., Wang, Y., He, H., Cai, L., Sun, H., Liu, C. *et al.* (2006) Organization of the *Caenorhabditis elegans* small non-coding transcriptome: genomic features, biogenesis, and expression. *Genome Res.*, **16**, 20–29.

133. Berezikov, E. and Plasterk, R.H. (2005) Camels and zebrafish, viruses and cancer: a microRNA update. *Hum. Mol. Genet.*, **14** (Suppl. 2), R183–R190.

134. Bartel, B. (2005) MicroRNAs directing siRNA biogenesis. *Nat. Struct. Mol. Biol.*, **12**, 569-571.

135. Zamore, P.D. and Haley, B. (2005) Ribo-gnome: the big world of small RNAs. *Science*, **309**, 1519-1524.

136. Verdel, A. and Moazed, D. (2005) RNAi-directed assembly of heterochromatin in fission yeast. *FEBS Lett.*, **579**, 5872-5878.

137. Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H. and Kim, V.N. (2004) MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.*, **23**, 4051–4060.

138. Rodriguez, A., Griffiths-Jones, S., Ashurst, J.L. and Bradley, A. (2004) Identification of mammalian microRNA host genes and transcription units. *Genome Res.*, **14**, 1902–1910.

139. Smalheiser, N.R. and Torvik, V.I. (2005) Mammalian microRNAs derived from genomic repeats. *Trends Genet.*, **21**, 322–326.

140. Whitelaw, E. and Martin, D.I. (2001) Retrotransposons as epigenetic mediators of phenotypic variation in mammals. *Nat. Genet.*, **27**, 361–365.

141. Peaston, A.E., Evsikov, A.V., Graber, J.H., de Vries, W.N., Holbrook, A.E., Solter, D. and Knowles, B.B. (2004) Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell*, **7**, 597–606.

142. Devor, E.J. (2006) Primate microRNAs miR-220 and miR-492 lie within processed pseudogenes. *J. Hered.*, **97**, 186–190.

143. Croce, C.M. and Calin, G.A. (2005) miRNAs, cancer, and stem cell division. *Cell*, **122**, 6–7.

144. Klein, M.E., Impey, S. and Goodman, R.H. (2005) Role reversal: the regulation of neuronal gene expression by microRNAs. *Curr. Opin. Neurobiol.*, **15**, 507–513.

145. Giraldez, A.J., Cinalli, R.M., Glasner, M.E., Enright, A.J., Thomson, J.M., Baskerville, S., Hammond, S.M., Bartel, D.P. and Schier, A.F. (2005) MicroRNAs regulate brain morphogenesis in zebrafish. *Science*, **308**, 833–838.

146. Naguibneva, I., Ameyar-Zazoua, M., Polesskaya, A., Ait-Si-Ali, S., Groisman, R., Souidi, M., Cuvellier, S. and Harel-Bellan, A. (2006) The microRNA miR-181 targets the homeobox protein Hox-A11 during mammalian myoblast differentiation. *Nat. Cell Biol.*, **8**, 278–284.

147. Hatfield, S.D., Shcherbata, H.R., Fischer, K.A., Nakahara, K., Carthew, R.W. and Ruohola-Baker, H. (2005) Stem cell division is regulated by the microRNA pathway. *Nature*, **435**, 974–978.

148. Abelson, J.F., Kwan, K.Y., O'Roak, B.J., Baek, D.Y., Stillman, A.A., Morgan, T.M., Mathews, C.A., Pauls, D.L., Rasin, M.R., Gunel, M. *et al.* (2005) Sequence variants in SLITRK1 are associated with Tourette's syndrome. *Science*, **310**, 317–320.

149. Iorio, M.V., Ferracin, M., Liu, C.G., Veronese, A., Spizzo, R., Sabbioni, S., Magri, E., Pedriali, M., Fabbri, M., Campiglio, M. *et al.* (2005) MicroRNA gene expression deregulation in human breast cancer. *Cancer Res.*, **65**, 7065–7070.

150. Jiang, J., Lee, E.J., Gusev, Y. and Schmittgen, T.D. (2005) Real-time expression profiling of microRNA precursors in human cancer cell lines. *Nucleic Acids Res.*, **33**, 5394–5403.

151. Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A. *et al.* (2005) MicroRNA expression profiles classify human cancers. *Nature*, **435**, 834–838.

152. Calin, G.A., Ferracin, M., Cimmino, A., Di Leva, G., Shimizu, M., Wojcik, S.E., Iorio, M.V., Visone, R., Sever, N.I., Fabbri, M. *et al.* (2005) A microRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N. Engl. J. Med.*, **353**, 1793–1801.

153. O'Donnell, K.A., Wentzel, E.A., Zeller, K.I., Dang, C.V. and Mendell, J.T. (2005) c-Myc-regulated microRNAs modulate E2F1 expression. *Nature*, **435**, 839–843.

154. He, L., Thomson, J.M., Hemann, M.T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S.W., Hannon, G.J. *et al.* (2005) A microRNA polycistron as a potential human oncogene. *Nature*, **435**, 828–833.

155. Nairz, K., Rottig, C., Rintelen, F., Zdobnov, E., Moser, M. and Hafen, E. (2006) Overgrowth caused by misexpression of a microRNA with dispensable wild-type function. *Dev. Biol.*, Epub ahead of print (doi:10.1016/j.ydbio.2005.11.047).

156. Yang, W., Chendrimada, T.P., Wang, Q., Higuchi, M., Seeburg, P.H., Shiekhattar, R. and Nishikura, K. (2006) Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat. Struct. Mol. Biol.*, **13**, 13–21.

157. Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. and Enright, A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34** (database issue), D140–D144.

158. Berezikov, E., Guryev, V., van de Belt, J., Wienholds, E., Plasterk, R.H. and Cuppen, E. (2005) Phylogenetic shadowing and computational identification of human microRNA genes. *Cell*, **120**, 21–24.

159. Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U., Meiri, E. *et al.* (2005) Identification of hundreds of conserved and non-conserved human microRNAs. *Nat. Genet.*, **37**, 766–770.

160. Johnston, R.J. and Hobert, O. (2003) A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature*, **426**, 845–849.

161. Morey, C. and Avner, P. (2004) Employment opportunities for non-coding RNAs. *FEBS Lett.*, **567**, 27–34.

162. Bayne, E.H. and Allshire, R.C. (2005) RNA-directed transcriptional gene silencing in mammals. *Trends Genet.*, **21**, 370–373.

163. Corey, D.R. (2005) Regulating mammalian transcription with RNA. *Trends Biochem. Sci.*, **30**, 655–658.

164. Martens, J.A., Laprade, L. and Winston, F. (2004) Intergenic transcription is required to repress the *Saccharomyces cerevisiae SER3* gene. *Nature*, **429**, 571–574.

165. Rank, G., Prestel, M. and Paro, R. (2002) Transcription through intergenic chromosomal memory elements of the *Drosophila* bithorax complex correlates with an epigenetic switch. *Mol. Cell. Biol.*, **22**, 8026–8034.

166. Schmitt, S., Prestel, M. and Paro, R. (2005) Intergenic transcription through a polycomb group response element counteracts silencing. *Genes Dev.*, **19**, 697–708.

167. Vakoc, C.R., Mandat, S.A., Olenchock, B.A. and Blobel, G.A. (2005) Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. *Mol. Cell*, **19**, 381–391.

168. Schmitt, S. and Paro, R. (2004) A reason for reading nonsense. *Nature*, **429**, 510–511.

169. Morris, K.V., Chan, S.W., Jacobsen, S.E. and Looney, D.J. (2004) Small interfering RNA-induced transcriptional gene silencing in human cells. *Science*, **305**, 1289–1292.

170. Imamura, T., Yamamoto, S., Ohgane, J., Hattori, N., Tanaka, S. and Shiota, K. (2004) Non-coding RNA directed DNA demethylation of Sphk1 CpG island. *Biochem. Biophys. Res. Commun.*, **322**, 593–600.

171. Holliday, R. and Murray, V. (1994) Specificity in splicing. *Bioessays*, **16**, 771–774.

172. Martinho, R.G., Kunwar, P.S., Casanova, J. and Lehmann, R. (2004) A non-coding RNA is required for the repression of RNApolII-dependent transcription in primordial germ cells. *Curr. Biol.*, **14**, 159–165.

173. Shi, Y. and Berg, J.M. (1995) Specific DNA–RNA hybrid binding by zinc finger proteins. *Science*, **268**, 282–284.

174. Ladomery, M. (1997) Multifunctional proteins suggest connections between transcriptional and post-transcriptional processes. *Bioessays*, **19**, 903–909.

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

176. Muchardt, C., Guillemé, M., Seeler, J., Trouche, D., Dejean, A. and Yaniv, M. (2002) Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1. *EMBO Rep.*, **3**, 975–981.

177. Jeffery, L. and Nakielny, S. (2004) Components of the DNA methylation system of chromatin control are RNA-binding proteins. *J. Biol. Chem.*, **279**, 49479–49487.

178. Krajewski, W.A., Nakamura, T., Mazo, A. and Canaani, E. (2005) A motif within SET-domain proteins binds single-stranded nucleic acids and transcribed and supercoiled DNAs and can interfere with assembly of nucleosomes. *Mol. Cell. Biol.*, **25**, 1891–1899.

179. Brown, R.S. (2005) Zinc finger proteins: getting a grip on RNA. *Curr. Opin. Struct. Biol.*, **15**, 94–98.

180. Kuwabara, T., Hsieh, J., Nakashima, K., Taira, K. and Gage, F.H. (2004) A small modulatory dsRNA specifies the fate of adult neural stem cells. *Cell*, **116**, 779–793.

181. Janowski, B.A., Huffman, K.E., Schwartz, J.C., Ram, R., Hardy, D., Shames, D.S., Minna, J.D. and Corey, D.R. (2005) Inhibiting gene expression at transcription start sites in chromosomal DNA with antigene RNAs. *Nat. Chem. Biol.*, **1**, 216–222.

182. Ting, A.H., Schuebel, K.E., Herman, J.G. and Baylin, S.B. (2005) Short double-stranded RNA induces transcriptional gene silencing in human cancer cells in the absence of DNA methylation. *Nat. Genet.*, **37**, 906–910.

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

184. Chooniedass-Kothari, S., Emberley, E., Hamedani, M.K., Troup, S., Wang, X., Czosnek, A., Hube, F., Mutawe, M., Watson, P.H. and Leygue, E. (2004) The steroid receptor RNA activator is the first functional RNA encoding a protein. *FEBS Lett.*, **566**, 43–47.

185. Chooniedass-Kothari, S., Hamedani, M.K., Troup, S., Hube, F. and Leygue, E. (2006) The steroid receptor RNA activator protein is expressed in breast tumor tissues. *Int. J. Cancer*, **118**, 1054–1059.

186. Zhao, X., Patton, J.R., Davis, S.L., Florence, B., Ames, S.J. and Spanjaard, R.A. (2004) Regulation of nuclear receptor activity by a pseudouridine synthase through post-transcriptional modification of steroid receptor RNA activator. *Mol. Cell. Biol.*, **15**, 549–558.

187. Xu, B. and Koenig, R.J. (2004) An RNA-binding domain in the thyroid hormone receptor enhances transcriptional activation. *J. Biol. Chem.*, **279**, 33051–33056.

188. Espinoza, C.A., Allen, T.A., Hieb, A.R., Kugel, J.F. and Goodrich, J.A. (2004) B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nat. Struct. Mol. Biol.*, **11**, 822–829.

189. Allen, T.A., Von Kaenel, S., Goodrich, J.A. and Kugel, J.F. (2004) The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nat. Struct. Mol. Biol.*, **11**, 816–821.

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

191. Valgardsdottir, R., Chiodi, I., Giordano, M., Cobianchi, F., Riva, S. and Biamonti, G. (2005) Structural and functional characterization of non-coding repetitive RNAs transcribed in stressed human cells. *Mol. Biol. Cell*, **16**, 2597–2604.

192. Lakhotia, S.C., Rajendra, T.K. and Prasanth, K.V. (2001) Developmental regulation and complex organization of the promoter of the non-coding *hsr(omega)* gene of *Drosophila melanogaster. J. Biosci.*, **26**, 25–38.

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

194. Cai, X., Hagedorn, C.H. and Cullen, B.R. (2004) Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA*, **10**, 1957–1966.

195. Baskerville, S. and Bartel, D.P. (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA*, **11**, 241–247.

196. Ying, S.Y. and Lin, S.L. (2005) Intronic microRNAs. *Biochem. Biophys. Res. Commun.*, **326**, 515–520.

197. Clement, J.Q., Qian, L., Kaplinsky, N. and Wilkinson, M.F. (1999) The stability and fate of a spliced intron from vertebrate cells. *RNA*, **5**, 206–220.

198. Clement, J.Q., Maiti, S. and Wilkinson, M.F. (2001) Localization and stability of introns spliced from the *Pem homeobox* gene. *J. Biol. Chem.*, **276**, 16919–16930.

199. Hill, A.E., Hong, J.S., Wen, H., Teng, L., McPherson, D.T., McPherson, S.A., Levasseur, D.N. and Sorscher, E.J. (2006) Micro-RNA-like effects of complete intronic sequences. *Front. Biosci.*, **11**, 1998–2006.

200. Dermitzakis, E.T., Reymond, A., Scamuffa, N., Ucla, C., Kirkness, E., Rossier, C. and Antonarakis, S.E. (2003) Evolutionary discrimination of mammalian conserved non-genic sequences (CNGs). *Science*, **302**, 1033–1035.

201. Thomas, J.W., Touchman, J.W., Blakesley, R.W., Bouffard, G.G., Beckstrom-Sternberg, S.M., Margulies, E.H., Blanchette, M., Siepel, A.C., Thomas, P.J., McDowell, J.C. *et al.* (2003) Comparative analyses of multi-species sequences from targeted genomic regions. *Nature*, **424**, 788–793.

202. Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S. *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.*, **15**, 1034–1050.

203. Sironi, M., Menozzi, G., Comi, G.P., Cagliani, R., Bresolin, N. and Pozzoli, U. (2005) Analysis of intronic conserved elements indicates that functional complexity might represent a major source of negative selection on non-coding sequences. *Hum. Mol. Genet.*, **14**, 2533–2546.

204. Wang, J., Zhang, J., Zheng, H., Li, J., Liu, D., Li, H., Samudrala, R., Yu, J. and Wong, G.K. (2004) Neutral evolution of 'non-coding' cDNAs from the mouse transcriptome. *Nature*, doi:10.1038/nature03016.

205. Andolfatto, P. (2005) Adaptive evolution of non-coding DNA in *Drosophila. Nature*, **437**, 1149–1152.

206. Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas, E.J., III, Zody, M.C. *et al.* (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, **438**, 803–819.

207. Smith, N.G., Brandstrom, M. and Ellegren, H. (2004) Evidence for turnover of functional non-coding DNA in mammalian genome evolution. *Genomics*, **84**, 806–813.

208. Simons, C., Pheasant, M., Makunin, I.V. and Mattick, J.S. (2006) Transposon-free regions in mammalian genomes. *Genome Res.*, **16**, 164–172.

209. Smit, M., Segers, K., Carrascosa, L.G., Shay, T., Baraldi, F., Gyapay, G., Snowder, G., Georges, M., Cockett, N. and Charlier, C. (2003) Mosaicism of Solid Gold supports the causality of a non-coding A-to-G transition in the determinism of the callipyge phenotype. *Genetics*, **163**, 453–456.

210. Georges, M., Charlier, C. and Cockett, N. (2003) The callipyge locus: evidence for the trans interaction of reciprocally imprinted genes. *Trends Genet.*, **19**, 248–252.

211. Van Laere, A.S., Nguyen, M., Braunschweig, M., Nezer, C., Collette, C., Moreau, L., Archibald, A.L., Haley, C.S., Buys, N., Tally, M. *et al.* (2003) A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature*, **425**, 832–836.

212. Mattick, J.S. (2005) The functional genomics of non-coding RNA. *Science*, **309**, 1527–1528.

213. Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.

214. Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D. and Church, G.M. (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*, **309**, 1728–1732.