

Small regulatory RNAs in mammals

John S. Mattick* and Igor V. Makunin

ARC Special Research Centre for Functional and Applied Genomics, Institute for Molecular Bioscience, University of Queensland, Brisbane QLD 4072, Queensland, Australia

Received January 3, 2005; Revised and Accepted February 23, 2005

Mammalian cells harbor numerous small non-protein-coding RNAs, including small nucleolar RNAs (snoRNAs), microRNAs (miRNAs), short interfering RNAs (siRNAs) and small double-stranded RNAs, which regulate gene expression at many levels including chromatin architecture, RNA editing, RNA stability, translation, and quite possibly transcription and splicing. These RNAs are processed by multistep pathways from the introns and exons of longer primary transcripts, including protein-coding transcripts. Most show distinctive temporal- and tissue-specific expression patterns in different tissues, including embryonal stem cells and the brain, and some are imprinted. Small RNAs control a wide range of developmental and physiological pathways in animals, including hematopoietic differentiation, adipocyte differentiation and insulin secretion in mammals, and have been shown to be perturbed in cancer and other diseases. The extent of transcription of non-coding sequences and the abundance of small RNAs suggests the existence of an extensive regulatory network on the basis of RNA signaling which may underpin the development and much of the phenotypic variation in mammals and other complex organisms and which may have different genetic signatures from sequences encoding proteins.

INTRODUCTION

Although only 1.2% of the human genome encodes protein, a large fraction of it is transcribed. Indeed, ~98% of the transcriptional output in humans and other mammals consists of non-protein-coding RNAs (ncRNA) from the introns of protein-coding genes and the exons and introns of non-protein-coding genes (1,2), including many that are antisense to or overlapping protein-coding genes (3–5).

Until recently, the non-coding RNA fraction was considered mainly useless with the exception of the common infrastructural RNAs involved in protein synthesis, transport and splicing. Introns have long been regarded as evolutionary debris with intronic RNA assumed to be simply degraded after splicing excision, and the increasing number of non-protein-coding transcripts being detected in mammalian cells has been suggested, at least by some, to be largely 'transcriptional noise' (6). However, a significant proportion of ncRNAs appears to be stable in eukaryotic cells. For example, some excised introns have half-lives comparable with mRNA and are even exported from the nucleus to the cytoplasm (7,8). Whole chromosome tiling chip arrays have shown that the range of detectable ncRNAs in human cells is much greater than can be accounted for by mRNAs (9) and that there

appear to be roughly equal numbers of protein-coding and non-coding transcripts regulated by common transcription factors in the human genome (4,10). Similar data have been reported in *Drosophila* (11).

All intensively studied gene loci, including those that are imprinted and conventional loci such as beta-globin have been shown to contain a majority of non-coding transcripts (12–15). The number of known functional ncRNA genes has risen dramatically in recent years and over 800 ncRNAs [excluding, transfer RNAs (tRNAs), ribosomal RNAs (rRNAs) and small nuclear RNAs (snRNAs)] have been catalogued in mammals, at least some of which are alternatively spliced (16,17), along with almost 20 000 putative ncRNA transcripts identified in cDNA libraries (16,18). ncRNAs have been implicated in diseases including various cancers and neurological diseases (2,16), and at least some are processed into smaller functional molecules (19,20).

Apart from tRNAs and spliceosomal snRNAs, which are housekeeping RNAs involved in mRNA splicing and translation, there are several functionally and structurally distinct classes of short RNAs in eukaryotic cells. In most, if not all, cases, their function is based on recognition of RNA or DNA target sequences by specific base pairing, analogous to digital signaling (21). Because of this feature, even short

*To whom correspondence should be addressed. Tel: +61 733462110; Fax: +61 73346 2111; Email: j.mattick@imb.uq.edu.au

RNAs contain sufficient information to specify individual targets in the genome and the transcriptome, in a much more compact and energy-efficient manner than proteins, which may have been a necessary adaptation to address the accelerating regulatory requirements of more complex organisms (21,22) and have been crucial to their evolution and development (23,24). These small RNAs and their role in mammalian cell and developmental biology are the subjects of the current review.

SMALL NUCLEOLAR RNAs

Small nucleolar RNAs (snoRNAs) guide the site-specific modification of nucleotides in target RNAs. Two types of modification occur, 2'-*O*-ribose methylation and pseudouridylation, directed by two large families of snoRNAs termed box C/D and box H/ACA snoRNAs, respectively (reviewed in 25,26). SnoRNAs recognize target sequences by formation of a canonical guide RNA duplex and recruit associated proteins to perform the corresponding modification at the target site. Generally, snoRNAs range between 60 and 300 nucleotides in length, but only short sequences participate in target recognition via antisense interactions. 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 possibly mRNAs (25,27). Interestingly, a pseudouridine synthase and pseudouridylation of the steroid receptor RNA activator SRA have recently been shown to be involved in retinoic acid and other nuclear receptor-dependent transactivation (28).

Over 300 different snoRNAs are known in humans and almost 200 in mouse (see databases at <http://noncode.bioinfo.org.cn> and <http://www.sanger.ac.uk/Software/Rfam>) (17,29), many of which occur in polycistronic clusters and at least some of which show tissue-specific expression, suggesting that they are specifically regulated and in turn have specific regulatory roles in the differential modification of selected target RNAs in different tissues including brain. Interestingly, a number of brain-specific snoRNAs come from imprinted regions and some of them represent orphan snoRNAs with unknown targets (30,31). One of these snoRNAs has an 18 nucleotide phylogenetically conserved sequence complementary to a critical alternative splice site and adenosine-to-inosine (A-I) RNA editing site in the serotonin 2C receptor mRNA whose gene also encodes another snoRNA that itself has an unknown target (30). Moreover, at least some brain-specific snoRNAs appear to have evolved recently and to be restricted in their phylogenetic distribution (32), suggesting their importance in the epigenetic control of behavior.

Mammalian snoRNAs are derived from introns of pre-mRNA transcripts, and are produced by processing of the excised intron (debranched lariat) as well as by endonucleolytic cleavage of unspliced primary transcripts, by a complex pathway involving endonucleases, exonucleases and helicases (25,26). In many cases, the snoRNA-containing introns occur within protein-coding transcripts, such as those encoding ribosomal proteins and others involved in ribosomal biogenesis/nucleolar function, a clear example of a parallel genetic output. However, in many other cases, snoRNAs are derived

from the introns of transcripts that do not have any protein-coding capacity (25–27,33), the function of whose exons (if any) is unknown.

No mutations with phenotypic consequences have been recorded in snoRNA sequences, which suggest that they are lethal, functionally redundant or (most likely) cause subtle effects, for example on growth or brain function. It has been reported that certain snoRNAs, present as multicopy genes (designated as *HBII-52* and *HBII-85* snoRNAs, respectively), are absent from the cortex of a patient with Prader–Willi Syndrome (PWS) and from a PWS mouse model, demonstrating their paternal imprinting and pointing to their potential role in the etiology of PWS (30). However, a recent study of a genomic deletion of the *HBII-52* snoRNA gene cluster in humans indicates that these snoRNAs do not play a major role in PWS on their own, or do so only in connection with the *HBII-85* snoRNA cluster (34).

In relation to RNA regulation, it should be noted that another form of RNA editing, A–I conversion, catalyzed by adenosine deaminases that act on RNA (ADARs), is also common in the brain, and aberrant editing has been associated with certain cancers and a range of abnormal behaviors including epilepsy and depression (35). In humans, A–I editing has recently been shown to be much more widespread than was previously thought and to occur primarily in *Alu* elements (which are primate-specific) in non-coding RNA sequences in both protein-coding and non-coding transcripts (36–38). A–I editing also appears to modulate RNA interference (RNAi) (described subsequently) by altering the sequence of introduced (39,40) and naturally occurring (41) double-stranded RNAs.

miRNAs AND siRNAs

The other broad class of small regulatory RNAs in mammals (and indeed in animals and plants generally) are the tiny, generally 21–25 nt long, molecules named microRNAs (miRNAs) and short interfering RNAs (siRNA), which have been the subject of intense recent interest (42–48). miRNAs come from endogenous short hairpin precursor structures (described subsequently) and usually target other loci with similar but not identical sequences for translational repression. siRNAs are produced from longer double-stranded (bimolecular) RNAs or long hairpins, often of exogenous origin, and usually target homologous sequences at the same locus or elsewhere in the genome for destruction (gene silencing) (43–45), the phenomenon termed RNAi (49). However, the distinction between miRNAs and siRNAs is becoming blurred as both are produced by similar pathways and have similar mechanisms of action (43–45). Recent observations have shown that both miRNAs and siRNAs can suppress translation of mRNAs (in the case of an imperfect match) and can cleave target RNAs (in the case of a perfect match) (44,50,51). miRNAs have been shown to be involved in a variety of developmental processes in animals and plants (43,46) (described subsequently). In contrast, siRNAs were originally proposed to act mainly as an antiviral defense and transposon repression system via the phenomenon of RNAi (52), but recent findings

indicate that such RNAs may play a much broader role in gene and genome regulation (described subsequently).

A database of known and predicted endogenous miRNAs in various animal and plant species is available at <http://www.sanger.ac.uk/Software/Rfam/mirna> (29,53). The database (release 6.1) currently lists 222 human and 224 mouse miRNAs (as well as 186 rat miRNAs), many of which are orthologous. Roughly half are conserved in fishes and a quarter are conserved in invertebrates (54), which presumably have evolutionarily conserved functions in vertebrate and metazoan development. However, new human and rodent miRNAs are constantly being identified (55). Very recently, 976 candidate miRNAs were identified by scanning whole-genome human/mouse and human/rat alignments, most of which are also conserved in other vertebrates, and around 20% of which have experimental support (56).

Two databases of siRNAs directed against human genes have also recently been published (see <http://www.human-siRNA-database.net>, and <http://siRNA.cgb.ki.se>) (57,58). These databases contain several hundred siRNAs that have been experimentally verified to be active against human genes (57,58) and thousands of siRNA sequences designed computationally to be active against the *RefSeq* curated human gene set (58).

Biogenesis of miRNAs and siRNAs

miRNAs are processed by the RNAi machinery in a two-step cleavage process (59) from longer primary transcripts that have been termed 'pri-miRNAs' (60) but that in reality appear to be conventional pre-mRNAs and ncRNAs (15,20,55, 61–66), including antisense transcripts (47,55), at least some of which are polycistronic (15,60,61,66–68). miRNA-containing primary transcripts have been shown to be synthesized by RNA polymerase II and to be polyadenylated and capped (62,65). Many human and mouse miRNAs are derived from the introns of protein-coding genes and the remainder from the introns and the exons of mRNA-like ncRNA genes (63,66).

These transcripts, presumably after splicing (64), are processed by the RNase III endonuclease Droscha (69). Droscha cleaves RNA hairpins that contain a large (≥ 10 nt) terminal loop approximately two helical turns into the stem, to excise 65–75 nt precursors called 'pre-miRNAs' (69,70). Droscha appears to occur in two complexes in the nucleus. The larger of these complexes includes a variety of RNA-associated proteins including RNA helicases, proteins that bind double-stranded RNA, novel heterogeneous nuclear ribonucleoproteins and the Ewing's sarcoma family of proteins (71), whereas the smaller complex is composed of Droscha and the double-stranded-RNA-binding protein, DGCR8 (also called Pasha), the product of a gene deleted in DiGeorge syndrome (71–74). The pre-miRNAs are then exported from the nucleus by Exportin 5 (75,76) and processed by the cytoplasmic RNase III endonuclease Dicer (77) into ~ 22 bp (imperfect) duplexes with a 2 nt overhang at their 3' ends (43,46,59).

siRNAs are also processed by Dicer from double-stranded RNA precursors but do not require Droscha (reviewed in 43,46). These precursors may be produced endogenously, for example, from sense–antisense transcripts. They may

also be supplied exogenously, as occurred in the initial discovery of RNAi, whereby such RNAs can act catalytically to destroy endogenous RNAs with matching sequence (78), now a widespread tool for probing gene function by siRNA-induced target knockdown (57,58).

The Dicer-processed short duplex RNAs are incorporated into the RISC ribonucleoprotein complex, which contains a member of the Argonaute family. There are many Argonaute homologs in animals, plants and fungi, implying that there may be many forms of such complexes that may recognize different RNA substrates (reviewed in 79). Recent evidence suggests that different RISC complexes containing different Argonaute proteins may be involved in miRNA- and siRNA-mediated RNAi (80), although there are conflicting reports (81). Argonaute proteins intersect with the Wingless/Wnt and Hedgehog pathways that control cell fate and developmental patterning (82,83). Mutations in Argonaute family members affect a variety of developmental processes including germ cell development and stem cell fate, as well as being implicated in various human cancers and developmental abnormalities (79). The fragile X mental retardation protein (FMRP), an RNA-binding protein which associates with hundreds of mRNAs in neurons via a G-quartet structure and/or U-rich sequences (84–88), is also associated with the RISC complex, as well as with Dicer itself (89–92), and is involved with the control of behavior via a process involving Argonaute2 (93). There is a strong enrichment of predicted miRNA targets in mRNAs associated with FMRP in mammals (94), and it has also recently been reported that FMRP is phosphorylated by casein kinase II (95), hinting at the enormous complexity in these RNA processing and signaling pathways and their regulation.

In general, only one strand from the processed duplex is retained in the RISC complex, the selection of which appears to be determined by the relative stability of the two ends of the duplex, favoring the one whose 5'-end is less tightly paired (96,97). The RISC complex then forms a complex with a target RNA and either leads to its translational repression by an, as yet, unknown mechanism, but which may involve interaction with polyribosomes (98), or in the case of (near) perfect identity cleaves the target RNA approximately in the middle of the paired region. Recent evidence suggests that the endonuclease activity within the RISC complex, dubbed 'slicer', is in fact mediated by an RNase H-like domain (piwi) in Argonaute (99–102).

The roles of miRNAs and siRNAs in development and disease

Some mammalian miRNAs appear to be ubiquitously expressed, but most have been found to exhibit developmentally regulated expression patterns in a variety of cells and tissues, including brain, lung, liver, spleen, heart and skeletal muscle, using northern blots, PCR, microarray chips and sensor transgenes (15,46,68,103–112). Many miRNAs are specifically expressed during embryonal stem cell differentiation (68,105) and embryogenesis (112), as well as during brain development (104,108,113,114), neuronal differentiation (103,115) and hematopoietic lineage differentiation (106,111).

Studies in model organisms have shown that miRNAs are involved in the control of developmental timing, cell proliferation, left–right patterning, neuronal cell fate, apoptosis and fat metabolism in invertebrates (as well as in a variety of developmental processes in plants) (reviewed in 43,45–47,116), and there is every reason to expect a similar range of functions in vertebrates. Indeed the archetypal miRNAs, *lin-4* and *let-7*, which were first discovered by genetic screens to control developmental timing in *Caenorhabditis elegans*, have been shown to have close homologs in other species, including mammals (117–119), as do many other miRNAs (54,67,118–120). The target of *lin-4* and *let-7*, *lin-28*, is also conserved in mammals (121). siRNAs targeted against *let-7* cause developmental abnormalities in fish and frogs (122), and human *let-7* paralogs are able to suppress a sensor transgene containing the human homolog of *lin-28* (123). It has also been shown that *let-7b* miRNA associates together with *lin-28* mRNA in polyribosomes in human cells, indicating a possible physical interaction between this miRNA and its target mRNA (98). Reduced expression of *let-7* is observed in certain human lung cancers in association with shortened postoperative survival, and over-expression of *let-7* in a lung adenocarcinoma cell line inhibited lung cancer cell growth *in vitro* (124).

Knockout of the miRNA-producing enzyme Dicer1 in mice leads to lethality early in development, with Dicer1-null embryos depleted of stem cells (125). These observations and the apparent inability to generate viable Dicer1-null embryonic stem cells *in vitro* suggest a role for Dicer, and, by implication, miRNAs, in maintaining stem cell populations during early mouse development (125). Dicer-defective ES cells also exhibit severe defects in differentiation *in vitro* as well as in centromeric silencing (126). Inactivation of Dicer also causes developmental arrest in zebrafish embryos (127).

In mammals, miRNAs have been shown to regulate B-cell differentiation (106), adipocyte differentiation (128) and insulin secretion (129). Chen *et al.* have shown that three miRNAs are differentially expressed during mouse hematopoiesis and that ectopic over-expression of one of them (*miR-181*) in hematopoietic stem/progenitor cells increases the fraction of B-lineage cells both *in vitro* and *in vivo* (106). Reduction in the level of *miR-143*, one of whose predicted targets is the MAP kinase *BMK1/ERK5* mRNA (130), resulted in an increase in the level of *BMK1/ERK5* and inhibited adipocyte differentiation in culture (128). In pancreatic endocrine cells, inhibition of *miR-375* enhanced glucose-induced insulin secretion and conversely, over-expression of *miR-375* suppressed insulin secretion, an effect that could be mimicked by siRNAs directed against *Myotrophin*, the putative target of *miR-375* (129). The authors suggest that many of 67 miRNA sequences cloned from pancreatic cells (11 of which had not been previously identified) may regulate endocrine pancreas development (129).

Some miRNAs are embedded in *Hox* clusters and exhibit expression patterns that are reminiscent of *Hox* genes (51,112). At least one of these miRNAs (*miR-196*) has extensive, evolutionarily conserved complementarity to *HoxB8*, *HoxC8* and *HoxD8* sequences and has been shown to negatively regulate *HoxB8* and other *Hox* genes, suggesting a miRNA-mediated mechanism for the posttranscriptional

restriction of *Hox* gene expression during vertebrate development (51,112).

It has also been shown that some mammalian miRNAs are imprinted (15,61). The imprinting process clearly involves transactions with non-coding RNAs (131), although the mechanisms remain unknown. Mouse *miR-127* and *miR-136* are transcribed antisense to a reciprocally imprinted retrotransposon-like gene, which are expressed from maternal and paternal chromosomes, respectively. In addition, the neighboring region contains maternally expressed clusters of snoRNA and miRNAs, and it has been suggested that these miRNAs may play a role in the imprinting process, either by directing allele-specific chromatin modification or by targeting particular transcripts (15,61).

These observations all indicate that miRNAs are part of the molecular circuitry and complex regulatory networks that control cell fate during mammalian development, a conclusion supported by a variety of studies in other animals. For example, the miRNAs, *miR-7* and *miR-2a/b*, have been shown to regulate the Notch pathway and proapoptotic genes *reaper*, *grim* and *sickle* in *Drosophila* (132).

Consistent with their role in developmental processes, perturbations of miRNA expression are observed in aberrant developmental states, i.e. oncogenesis (133), in human B-cell chronic lymphocytic leukemia (109,134), Burkitt lymphoma (20), colorectal cancer (135), lung cancer (124) and in a number of cancer cell lines (110). A high proportion of known miRNAs are located at fragile sites or in cancer-associated genomic regions (minimal regions of loss of heterozygosity, minimal regions of amplification or common breakpoint regions) (136). Interestingly, some miRNAs from the same genomic cluster show different expression patterns in cancer cells, indicating that the regulation of some miRNAs might occur post transcriptionally (109), perhaps, themselves regulated by other miRNAs as part of more complex regulatory networks (described subsequently).

It has been reported that some siRNAs can induce sequence-dependent off-target effects on proteins such as p53 and p21 that are sensitive markers of cell state (137), which suggest care in the interpretation of such experiments. These effects may occur either because of partial complementary sequence matches to other genes (137) or because perturbation of miRNA-regulated pathways can have pleiotropic effects on wider networks of gene expression.

Target prediction in mammals

Thus far, only a few miRNA targets have been identified in mammals and the rules of interaction are largely unknown. The majority of identified mammalian miRNAs has non-perfect matches to target mRNAs but the rules are complex (138). Because of the very short target sequences and the presence of mismatches, the bioinformatic prediction of miRNA targets, especially in the complex genomes of mammals, is a very challenging task. Most studies predict miRNA targets on the basis of an evolutionarily conserved sequence complementarity and low free energy of interaction, usually focused on 3'-UTRs of known genes (94,123,130,132), on the presumption that most miRNAs are involved in translational

repression via UTRs, by extension from the well-studied examples of *lin-4* and *let-7*.

Such approaches potentially minimize the numbers of false positives but may well seriously underestimate the actual numbers of such RNAs (described subsequently). On the basis of minimal binding energies, Kiriakidou *et al.* (123) predicted 5031 target sequences for 94 miRNAs. More than 400 targets were predicted by Lewis *et al.* (130) and 11 out of 15 were confirmed experimentally. Predictions from a bigger miRNA dataset (218 known mammalian miRNAs) identified 2273 target genes with one or more target sites showing 90% sequence conservation between human, mouse and rat in aligned UTRs (94). The predicted target genes had diverse functions, but were enriched for genes encoding mRNAs coding for transcription factors, components of the miRNA machinery, other proteins involved in translational regulation and components of the ubiquitin protein-degradation machinery (94,130), many of which are known to play important roles in developmental regulation and some of which are involved in the molecular etiology of cancer. Very recently, sophisticated bioinformatics analyses based on the overabundance of conserved adenosines flanking the complementary sites in mRNAs have implicated more than 5300 human genes (>20% of all known or predicted human protein-coding genes) as potential miRNA targets, most of which occur in 3'-UTRs, but some of which occur in coding sequences (139). The fact that many miRNAs are predicted to have multiple cognate mRNAs, and vice versa, suggests that the regulatory networks in which they participate are very complex indeed.

Small RNAs regulate chromosome dynamics and chromatin structure

As described previously, miRNAs and siRNAs target mRNAs for either translational inhibition or destruction by RISC-mediated cleavage (43,45,46). However, there is also considerable evidence that small RNAs also regulate chromosome dynamics, chromatin modification and epigenetic memory, including imprinting, DNA methylation and transcriptional gene silencing (2).

The RNAi pathway and non-coding RNAs have been shown to be central to the formation of silenced chromatin and chromosomal dynamics in animals, plants, fungi and protozoa (reviewed in 140–142). In the fission yeast, *Schizosaccharomyces pombe*, the RNAi pathway has been shown to be involved in heterochromatin formation, as well as in centromere function in meiosis and mitosis, via the methylation of histone H3 on lysine-9 and the RITS (RNA-induced initiation of transcriptional gene silencing) complex which contains Argonaute, the Chp1 chromodomain protein (among others) and Dicer-produced small RNAs homologous to the target DNA in heterochromatic regions (143–148). Meiosis in *S. pombe* has also been reported to require a number of specific non-coding RNAs (149). Similar observations have been made in *Drosophila* where mutations in components of the RNAi machinery affect silencing and heterochromatin formation, accompanied by reduction in histone H3 lysine-9 methylation and delocalization of the heterochromatin proteins HP1 and HP2 (150). Short 25–27 nt RNAs, derived

from dsRNA of *Drosophila Su(Ste)* repeats from the Y chromosome, suppress the *Stellate* gene on the X chromosome, and complementarity between the *Stellate* transcript and the *Su(Ste)* repeats is essential for silencing (151). Knockout of Dicer has been recently reported to affect centromeric heterochromatin formation in mouse (126).

These observations suggest that small RNAs may be central to chromatin regulation in all eukaryotes including mammals. Indeed, it has been shown that the localization of mammalian HP1 to heterochromatin involves its co-ordinate binding to methylated histone H3 and RNA, involving interactions in the hinge region between its chromodomains (152,153). This is consistent with previous reports that chromodomains (which are present in many different types of chromatin-binding and chromatin remodeling proteins, including the polycomb family, the histone methyltransferase and histone acetyltransferase families, the retinoblastoma binding protein 1 family, the CHD family and the SWI3 family) bind RNA as well as modified histones (154–156). RNA-interacting proteins are also components of the mammalian DNA methylation system (157).

Moreover, it has recently been shown that synthetic siRNAs targeted to CpG islands in the *E-cadherin* promoter reduced the expression of the gene and induced significant DNA methylation and histone H3 lysine-9 methylation in human cultured cells (158). Similar results were obtained with siRNAs directed against the *erbB2/HER2* promoter (158) and the *elongation factor 1alpha* promoter (159), providing strong support for the notion that endogenous small RNAs may perform similar functions *in vivo*. However, these results remain to be confirmed, and there is a recent report that over-expression of fragments of a non-coding antisense RNA *Khps1* results in demethylation of CG sites and methylation of CC(A/T)GG sites in a region of the *Sphk1* gene promoter that is subjected to tissue-specific differential methylation (160), suggesting that RNA signaling may control epigenetic modifications in more than one way. RNA signaling may also play a key role in the control of transcription and splicing (see below), although such RNAs have yet to be identified.

OTHER PLAYERS

Many other small RNAs appear to exist in animal and plant cells. Ambros and colleagues (116) recently described 33 small non-coding RNAs in *C. elegans*, which are similar in size to miRNAs and are developmentally regulated but which are not derived from hairpin precursors and are not evolutionarily conserved. They also described over 700 small RNAs that are antisense to known protein-coding sequences (compared with only 49 from sense strands), some of which are detectable as ~22 nt species in northern blots, which are potential endogenous siRNAs (116). Endogenous siRNAs derived from the sense and antisense strands of non-coding RNA have been described in *Arabidopsis* (161). A number of small RNAs, including new snoRNAs and 22 others ranging from 70 to 450 bp derived from intergenic and genic regions, including splice junctions, were detected and confirmed by northern blot analyses in *Drosophila* (162).

It seems that other small RNAs also exist in mammals. Only 30% of 179 small RNA sequences cloned from mouse ES cells appear to come from hairpin precursors (68). Approximately 20% show similarity to tRNA or rRNA, which leaves close to half that may act as small regulatory RNAs in some other capacity. Of 733 non-redundant sequences isolated from human ES cells, only 36 could have been derived from hairpin precursors (105). More than 50 unknown short non-coding RNAs were cloned from neural stem cells (163). One of these sequences is present in more than 60 copies in the mouse genome and has similarity to the NRSE/RE1 sequence, which is preferentially localized in promoter regions of neuron-specific genes. This RNA, which occurs in the nucleus as a small ~20 nt dsRNA, controls the differentiation of adult neural stem cells and activates the transcription of genes containing NRSE/RE1 sequence, apparently mediated through dsRNA-protein interactions, rather than through siRNA or miRNA (163).

PARALLEL OUTPUTS AND REGULATORY NETWORKS

Approximately two-thirds of annotated mammalian miRNAs are encoded within known genes and (like snoRNAs) mainly occur within introns of protein-coding and non-coding genes, with some residing within exons of non-coding genes (63). Some of these transcripts may be very long and contain multiple miRNAs (61,67). The situation appears to be even more complicated as some miRNAs from intronic regions are derived from anti-sense transcripts (54). The most common molecular function for mammalian protein-coding host genes are those annotated as 'purine nucleotide binding', 'DNA binding' (63) and those containing homeobox and RNA-binding domains (94), all of which point to the parallel output of proteins and regulatory RNA sequences as part of complex networks which underpin mammalian biology (1,24) (Fig. 1). Moreover, it has also been suggested, on the basis of sequence homologies, that some miRNAs may regulate other miRNAs rather than mRNAs via a network of regulatory interactions at the RNA level (164).

THE TIP OF THE ICEBERG?

There are only limited numbers of known miRNAs in mammals. The fact that some of these miRNAs show up repeatedly in cloning experiments had led some to suggest that the set of miRNAs in mammals and other organisms is small. However, there are good reasons to think that this is not the case and that there are in fact tens or even hundreds of thousands of RNA signals which constitute a hitherto hidden control network that regulates chromatin architecture and gene expression during mammalian ontogeny (24). Some miRNAs are present in large amounts (54), but it is clear that at least some miRNAs and other small regulatory RNAs are present at very low levels (45,165), and it is possible, indeed likely, that most will exhibit very restricted expression patterns in specific cell types such as observed in hematopoietic and pancreatic cells (106,129). Some miRNAs, such as that encoded by the *lys-6* locus, which

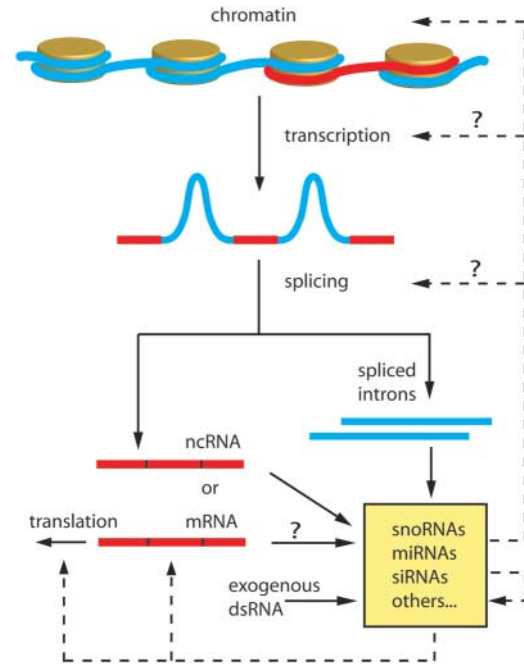


Figure 1. Regulatory networks involving small non-coding RNAs. Small non-coding RNAs regulate genome structure and gene expression at many levels. miRNAs, siRNAs, snoRNAs and other small RNAs are involved in the regulation of translation, mRNA stability and chromatin structure, as well as self-regulation (dashed lines) and possibly also the control of transcription and splicing (question marks).

controls the asymmetry of chemosensory neurons in *C. elegans* (166), and that encoded by the *bantam* locus in *Drosophila* (167), were only discovered by sensitive genetic screens, which are difficult if not impossible to carry out in mammals. The miRNA encoded by *lys-6* is very scarce and cannot be detected by normal biochemical procedures, and loss-of-function mutations in it result in a very subtle phenotype (166).

The lack of known mutations in miRNAs (and no doubt other types of regulatory RNAs) in mammals is likely to be due to a combination of ascertainment bias focused on exons of protein-coding genes and the difficulty of mutation screening across large tracts of non-coding sequences in regions identified by genome scanning for quantitative trait or disease associations. In this context, it is worth noting that the mutations underlying the callipyge ('beautiful bottom') phenotype in sheep or the enhanced muscling of domestic pigs are single base substitutions within non-coding sequences (a long intergenic sequence of unknown transcriptional status in the *DLKI-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 (168–170).

The problems of cloning small RNAs (171) and the contamination of cDNA libraries with rRNAs and other common RNA sequences have led to the conclusion that not many more miRNAs will be identified by this approach (54). Bioinformatic predictions (at least to date) have been limited by the tight constraints on the search parameters,

including their focus on hairpin precursors, mRNA/UTR targets and strong evolutionary conservation, although improved filters based on the different patterns of miRNA and flanking sequence conservation in different species have recently identified almost 1000 candidates (56). Moreover, new algorithms based on secondary structural parameters are being developed, which appear to have the potential to identify other types of non-coding RNAs (172) that presumably also have regulatory functions.

There is strong evidence that chromatin dynamics and heterochromatin formation are controlled by small regulatory RNAs and that local chromatin architecture (in promoter regions) can also be directed by small RNAs. Indeed, this would make a lot of sense. It is well established that chromatin modification occurs at many different loci in different cells and that this is central to developmental ontogeny. There must either be an army of sequence-specific DNA binding proteins that carry out these modifications, which is not the case—there are only a limited number of DNA and histone modifying enzymes (methylases, acetylases and deacetylases etc.) (173)—or these enzymes must be directed to their sites of action by some other signal, most logically sequence-specific RNAs. Such signals would also potentially solve the conundrum of how to select from the huge number of transcription factor binding sites that exist in the genome. In this context, it is interesting to note that triplexes, which may contain RNA, are very common in human chromosomes (174) and many transcription factors have high affinity for RNA (175,176).

Trans-acting guide RNAs may also regulate alternative splicing (177), which is currently mainly thought to be controlled by the combinatorial effects of protein ‘splicing factors’ but is not at all well understood in these terms (2,178,179). Consistent with the possibility that site-specific *trans*-acting RNAs are involved, the nucleotide sequences around alternative splice sites are often highly conserved between species (180,181), and it has been shown by many studies that splicing patterns may be easily altered in cultured cells and in whole animals by introducing small antisense RNAs, an approach which is showing considerable promise for gene therapy of splice site mutations in muscular dystrophy and other human genetic diseases (182–186). It is not a big leap of faith to conclude that RNA control of splice site selection is also likely to happen naturally and that the reason that it has not yet been demonstrated to be the case is because of the sheer complexity of the numbers and variety of such signals in regulatory networks in different cells. If cells are awash with small RNA signals processed from longer precursors, which (as such) have short half-lives, identification of these signals will be difficult, although bioinformatics using appropriate search algorithms may provide a means to do so.

The known miRNAs tend to be highly conserved (54,56), presumably because their sequence is constrained by functional interactions with multiple targets (94,187), which is possibly also the case for the ultra-conserved elements that are far more conserved than protein-coding sequences in the mammalian genome (188). In contrast, endogenous siRNAs are less conserved presumably because these RNAs and their homologous targets can easily co-vary and still maintain specificity (43), which would make them difficult to identify

bioinformatically, at least on the basis of evolutionary conservation. The level of selection pressure on such sequences (as signaling molecules largely dependent on primary sequence recognition and secondary structure) will be a function of the number of interactions that must be maintained, rather than the precise sequence itself. Those with one or few interacting partners will be able to evolve relatively freely and also explore new connections in regulatory networks, which themselves can evolve to explore new developmental space and which (given a relatively stable proteome) may be the major route to higher complexity and phenotypic variation.

Thus, many small regulatory RNAs, including possibly the majority of miRNAs, may not show strong evidence of sequence conservation over significant evolutionary distances. In this context, it is worth pointing out that known non-coding RNAs with conserved functions, such as *Xist*, are not highly conserved at the primary sequence level among mammals (189,190). This will, of course, also contribute to the perception that genomic sequences encoding such ncRNAs are (in general) drifting neutrally and that the majority of the genome which is transcribed is non-functional, which may not be the case at all (21). Indeed, a recent analysis suggests that the proportion of the human genome which is under purifying selection for functions held in common with other mammals is (at least) an order of magnitude higher in non-coding than in protein-coding sequences (191), an observation which is hard to reconcile with protein-based models of regulation of gene expression.

We have argued elsewhere that the majority of the genome of humans and other complex organisms is in fact devoted to extensive, but hitherto largely hidden, regulatory networks that are *trans*-acted by non-coding RNAs and that were essential to the evolution of complex organisms (1,2,21). We predicted that regulatory ncRNAs would be derived from the vast tracts of transcribed introns (23,24), which has now been confirmed in principle, as well as from non-protein-coding transcripts, which also appear to be the case. Indeed, the evidence is accumulating that the major advance in the evolution of complex organisms was the co-option of RNA as a digital signaling network, which was required to overcome the limitations of an analog (protein) based regulatory system (21,22,24). These RNA networks are likely to be intrinsically robust and their perturbation, particularly in the case of single nucleotide polymorphisms, may lead to a range of subtle phenotypes (in contrast to the generally severe effects of non-synonymous mutations in protein-coding sequences) (2). These may underpin much of the variation observed between species and individuals, including differences in quantitative and behavioral traits, and variation in susceptibility to complex diseases. If this is correct, most of our conceptions of gene regulation and approaches to molecular genetic analysis will have to be revised.

CONCLUSIONS

We have barely begun to investigate RNA regulatory networks in mammals. The majority of the mammalian genome is transcribed into non-protein-coding RNA. Numerous short RNAs are processed from longer transcripts and possess various

expression patterns, but the biological functions and targets are known for very few of them. The mechanism of RNAi is relatively well studied but our knowledge of chromatin modification by siRNAs is incomplete to say the least. It also seems likely that other short non-coding RNAs exist in the cell that may regulate many other processes and utilize other mechanisms, which remain to be discovered. Our understanding of the mammalian genome is undergoing major change. We used to consider most non-coding regions to be junk, but the extent of non-coding RNA transcription, the rapidly emerging evidence of regulatory networks controlled by RNA and a new logic about the genetic structure of complex organisms suggest that most of the mammalian genome may in fact be functional or at least that this possibility should be more seriously considered.

ACKNOWLEDGEMENTS

The authors thank the Australian Research Council and the Queensland State Government for financial support and their colleagues for many stimulating discussions. They also thank Alex Hüttenhofer and two anonymous reviewers for alerting us to some important omissions and for some very helpful suggestions.

REFERENCES

- Mattick, J.S. (2001) Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep.*, **2**, 986–991.
- Mattick, J.S. (2003) Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays*, **25**, 930–939.
- Yelin, R., Dahary, D., Sorek, R., Levanon, E.Y., Goldstein, O., Shoshan, A., Diber, A., Biton, S., Tamir, Y., Khosravi, R. *et al.* (2003) Widespread occurrence of antisense transcription in the human genome. *Nat. Biotechnol.*, **21**, 379–386.
- 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 noncoding RNAs. *Cell*, **116**, 499–509.
- Lavorgna, G., Dahary, D., Lehner, B., Sorek, R., Sanderson, C.M. and Casari, G. (2004) In search of antisense. *Trends Biochem. Sci.*, **29**, 88–94.
- Dennis, C. (2002) The brave new world of RNA. *Nature*, **418**, 122–124.
- 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.
- 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.
- Kapranov, P., Cawley, S.E., Drenkow, J., Bekiranov, S., Strausberg, R.L., Fodor, S.P. and Gingeras, T.R. (2002) Large-scale transcriptional activity in chromosomes 21 and 22. *Science*, **296**, 916–919.
- Kampa, D., Cheng, J., Kapranov, P., Yamanaka, M., Brubaker, S., Cawley, S., Drenkow, J., Piccolboni, A., Bekiranov, S., Helt, G. *et al.* (2004) Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res.*, **14**, 331–342.
- Stolc, V., Gauhar, Z., Mason, C., Halasz, G., van Batenburg, M.F., Rifkin, S.A., Hua, S., Herreman, T., Tongprasit, W., Barbano, P.E. *et al.* (2004) A gene expression map for the euchromatic genome of *Drosophila melanogaster*. *Science*, **306**, 655–660.
- Ashe, H.L., Monks, J., Wijgerde, M., Fraser, P. and Proudfoot, N.J. (1997) Intergenic transcription and transduction of the human beta-globin locus. *Genes Dev.*, **11**, 2494–2509.
- 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.
- 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.
- 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.
- 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 noncoding RNA database. *Nucleic Acids Res.*, **33**(Database issue), D125–D130.
- 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.
- Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H. *et al.* (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature*, **420**, 563–573.
- van den Berg, A., Kroesen, B.J., Kooistra, K., de Jong, D., Briggs, J., Blokzijl, T., Jacobs, S., Kluiver, J., Diepstra, A., Maggio, E. *et al.* (2003) High expression of B-cell receptor inducible gene *BIC* in all subtypes of Hodgkin lymphoma. *Genes Chromosomes Cancer*, **37**, 20–28.
- Metzler, M., Wilda, M., Busch, K., Viehmann, S. and Borkhardt, A. (2004) High expression of precursor microRNA-155/*BIC* RNA in children with Burkitt lymphoma. *Genes Chromosomes Cancer*, **39**, 167–169.
- Mattick, J.S. (2004) RNA regulation: a new genetics? *Nat. Rev. Genet.*, **5**, 316–323.
- Mattick, J.S. and Gagen, M.J. (2005) Accelerating networks. *Science*, **307**, 856–858.
- Mattick, J.S. (1994) Introns: evolution and function. *Curr. Opin. Genet. Dev.*, **4**, 823–831.
- Mattick, J.S. and Gagen, M.J. (2001) The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. *Mol. Biol. Evol.*, **18**, 1611–1630.
- Bachellerie, J.P., Cavaille, J. and Huttenhofer, A. (2002) The expanding snoRNA world. *Biochimie*, **84**, 775–790.
- Kiss, T. (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell*, **109**, 145–148.
- 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.
- 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 posttranscriptional modification of steroid receptor RNA activator. *Mol. Cell*, **15**, 549–558.
- Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S.R. and Bateman, A. (2005) Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.*, **33**(Database issue), D121–D124.
- Cavaille, J., Buiting, K., Kieffmann, M., Lalonde, 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.
- 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.
- 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 noncoding RNA gene in rats. *J. Biol. Chem.*, **276**, 26374–26383.

33. 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.
34. Runte, M., Varon, R., Horn, D., Horsthemke, B. and Buiting, K. (2004) Exclusion of the C/D box snoRNA gene cluster *HBII-52* from a major role in Prader–Willi syndrome. *Hum. Genet.*, **116**, 228–230.
35. Bass, B.L. (2002) RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.*, **71**, 817–846.
36. Levanon, E.Y., Eisenberg, E., Yelin, R., Nemzer, S., Hallegger, M., Shemesh, R., Fligelman, Z.Y., Shoshan, A., Pollock, S.R., Sztybel, D. *et al.* (2004) Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat. Biotechnol.*, **22**, 1001–1005.
37. Athanasiadis, A., Rich, A. and Maas, S. (2004) Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol.*, **2**, e391.
38. Kim, D.D., Kim, T.T., Walsh, T., Kobayashi, Y., Matise, T.C., Buyske, S. and Gabriel, A. (2004) Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome Res.*, **14**, 1719–1725.
39. Scadden, A.D. and Smith, C.W. (2001) RNAi is antagonized by A → I hyper-editing. *EMBO Rep.*, **2**, 1107–1111.
40. Yang, W., Wang, Q., Howell, K.L., Lee, J.T., Cho, D.S., Murray, J.M. and Nishikura, K. (2005) ADAR1 RNA deaminase limits siRNA efficacy in mammalian cells. *J. Biol. Chem.*, **280**, 3946–3953.
41. Tonkin, L.A. and Bass, B.L. (2003) Mutations in RNAi rescue aberrant chemotaxis of ADAR mutants. *Science*, **302**, 1725.
42. Shi, Y. (2003) Mammalian RNAi for the masses. *Trends Genet.*, **19**, 9–12.
43. Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
44. Cullen, B.R. (2004) Derivation and function of small interfering RNAs and microRNAs. *Virus Res.*, **102**, 3–9.
45. Ambros, V. (2004) The functions of animal microRNAs. *Nature*, **431**, 350–355.
46. He, L. and Hannon, G.J. (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.*, **5**, 522–531.
47. He, Z. and Sontheimer, E.J. (2004) “siRNAs and miRNAs”: a meeting report on RNA silencing. *RNA*, **10**, 1165–1173.
48. Kidner, C.A. and Martienssen, R.A. (2005) The developmental role of microRNA in plants. *Curr. Opin. Plant Biol.*, **8**, 38–44.
49. Fire, A. (1999) RNA-triggered gene silencing. *Trends Genet.*, **15**, 358–363.
50. Zeng, Y., Yi, R. and Cullen, B.R. (2003) MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc. Natl Acad. Sci. USA*, **100**, 9779–9784.
51. Yekta, S., Shih, I.H. and Bartel, D.P. (2004) MicroRNA-directed cleavage of HOXB8 mRNA. *Science*, **304**, 594–596.
52. Hunter, C.P. (2000) Gene silencing: shrinking the black box of RNAi. *Curr. Biol.*, **10**, R137–R140.
53. Griffiths-Jones, S. (2004) The microRNA Registry. *Nucleic Acids Res.*, **32**, D109–D111.
54. Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A. and Tuschl, T. (2003) New microRNAs from mouse and human. *RNA*, **9**, 175–179.
55. Weber, M.J. (2005) New human and mouse microRNA genes found by homology search. *FEBS J.*, **272**, 59–73.
56. 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.
57. Truss, M., Swat, M., Kielbasa, S.M., Schafer, R., Herzel, H. and Hagemeyer, C. (2005) HuSiDa—the human siRNA database: an open-access database for published functional siRNA sequences and technical details of efficient transfer into recipient cells. *Nucleic Acids Res.*, **33**(Database issue), D108–D111.
58. Chalk, A.M., Warfinge, R.E., Georgii-Hemming, P. and Sonnhammer, E.L. (2005) siRNAdb: a database of siRNA sequences. *Nucleic Acids Res.*, **33**(Database issue), D131–D134.
59. Murchison, E.P. and Hannon, G.J. (2004) miRNAs on the move: miRNA biogenesis and the RNAi machinery. *Curr. Opin. Cell Biol.*, **16**, 223–229.
60. Lee, Y., Jeon, K., Lee, J.T., Kim, S. and Kim, V.N. (2002) MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J.*, **21**, 4663–4670.
61. Seitz, H., Royo, H., Bortolin, M.L., Lin, S.P., Ferguson-Smith, A.C. and Cavaille, J. (2004) A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. *Genome Res.*, **14**, 1741–1748.
62. 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.
63. 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.
64. Ying, S.Y. and Lin, S.L. (2005) Intronic microRNAs. *Biochem. Biophys. Res. Commun.*, **326**, 515–520.
65. 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.
66. Baskerville, S. and Bartel, D.P. (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA*, **11**, 241–247.
67. Lau, N.C., Lim, L.P., Weinstein, E.G. and Bartel, D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, **294**, 858–862.
68. Houbaviv, H.B., Murray, M.F. and Sharp, P.A. (2003) Embryonic stem cell-specific microRNAs. *Dev. Cell*, **5**, 351–358.
69. Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S. *et al.* (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature*, **425**, 415–419.
70. Zeng, Y., Yi, R. and Cullen, B.R. (2005) Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *EMBO J.*, **24**, 138–148.
71. Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N. and Shiekhattar, R. (2004) The microprocessor complex mediates the genesis of microRNAs. *Nature*, **432**, 235–240.
72. Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F. and Hannon, G.J. (2004) Processing of primary microRNAs by the microprocessor complex. *Nature*, **432**, 231–235.
73. Han, J., Lee, Y., Yeom, K.H., Kim, Y.K., Jin, H. and Kim, V.N. (2004) The Drosha–DGCR8 complex in primary microRNA processing. *Genes Dev.*, **18**, 3016–3027.
74. Landthaler, M., Yalcin, A. and Tuschl, T. (2004) The human DiGeorge Syndrome critical region gene 8 and its *D. melanogaster* homolog are required for miRNA biogenesis. *Curr. Biol.*, **14**, 2162–2167.
75. Bohnsack, M.T., Czaplinski, K. and Gorlich, D. (2004) Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA*, **10**, 185–191.
76. Yi, R., Doehle, B.P., Qin, Y., Macara, I.G. and Cullen, B.R. (2005) Overexpression of Exportin 5 enhances RNA interference mediated by short hairpin RNAs and microRNAs. *RNA*, **11**, 220–226.
77. Kolb, F.A., Zhang, H., Jaronczyk, K., Tahbaz, N., Hobman, T.C. and Filipowicz, W. (2005) Human dicer: purification, properties, and interaction with PAZ PIWI domain proteins. *Methods Enzymol.*, **392**, 316–336.
78. Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E. and Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, **391**, 806–811.
79. Carmell, M.A., Xuan, Z., Zhang, M.Q. and Hannon, G.J. (2002) The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev.*, **16**, 2733–2742.
80. Okamura, K., Ishizuka, A., Siomi, H. and Siomi, M.C. (2004) Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.*, **18**, 1655–1666.
81. Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G. and Tuschl, T. (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell*, **15**, 185–197.
82. Kataoka, Y., Takeichi, M. and Uemura, T. (2001) Developmental roles and molecular characterization of a *Drosophila* homologue of *Arabidopsis* Argonaute1, the founder of a novel gene superfamily. *Genes Cells*, **6**, 313–325.
83. King, F.J., Szakmary, A., Cox, D.N. and Lin, H. (2001) Yb modulates the divisions of both germline and somatic stem cells through piwi- and hh-mediated mechanisms in the *Drosophila* ovary. *Mol. Cell*, **7**, 497–508.

84. Brown, V., Jin, P., Ceman, S., Darnell, J.C., O'Donnell, W.T., Tenenbaum, S.A., Jin, X., Feng, Y., Wilkinson, K.D., Keene, J.D. *et al.* (2001) Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell*, **107**, 477–487.
85. Chen, L.S., Tassone, F., Sahota, P. and Hagerman, P.J. (2003) The (CGG)_n repeat element within the 5' untranslated region of the FMR1 message provides both positive and negative cis effects on *in vivo* translation of a downstream reporter. *Hum. Mol. Genet.*, **12**, 3067–3074.
86. Chen, L., Yun, S.W., Seto, J., Liu, W. and Toth, M. (2003) The fragile X mental retardation protein binds and regulates a novel class of mRNAs containing U rich target sequences. *Neuroscience*, **120**, 1005–1017.
87. Denman, R.B. (2003) Deja vu all over again: FMRP binds U-rich target mRNAs. *Biochem. Biophys. Res. Commun.*, **310**, 1–7.
88. Miyashiro, K.Y., Beckel-Mitchener, A., Purk, T.P., Becker, K.G., Barret, T., Liu, L., Carbonetto, S., Weiler, I.J., Greenough, W.T. and Eberwine, J. (2003) RNA cargoes associating with FMRP reveal deficits in cellular functioning in Fmr1 null mice. *Neuron*, **37**, 417–431.
89. Ishizuka, A., Siomi, M.C. and Siomi, H. (2002) A *Drosophila* fragile X protein interacts with components of RNAi and ribosomal proteins. *Genes Dev.*, **16**, 2497–2508.
90. Caudy, A.A., Myers, M., Hannon, G.J. and Hammond, S.M. (2002) Fragile X-related protein and VIG associate with the RNA interference machinery. *Genes Dev.*, **16**, 2491–2496.
91. Jin, P., Zarnescu, D.C., Ceman, S., Nakamoto, M., Mowrey, J., Jongens, T.A., Nelson, D.L., Moses, K. and Warren, S.T. (2004) Biochemical and genetic interaction between the fragile X mental retardation protein and the microRNA pathway. *Nat. Neurosci.*, **7**, 113–117.
92. Schaeffer, C., Beaulande, M., Ehresmann, C., Ehresmann, B. and Moine, H. (2003) The RNA binding protein FMRP: new connections and missing links. *Biol. Cell*, **95**, 221–228.
93. Xu, K., Bogert, B.A., Li, W., Su, K., Lee, A. and Gao, F.B. (2004) The fragile X-related gene affects the crawling behavior of *Drosophila* larvae by regulating the mRNA level of the DEG/ENaC protein pickpocket1. *Curr. Biol.*, **14**, 1025–1034.
94. John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C. and Marks, D.S. (2004) Human microRNA targets. *PLoS Biol.*, **2**, e363.
95. Siomi, M.C., Higashijima, K., Ishizuka, A. and Siomi, H. (2002) Casein kinase II phosphorylates the fragile X mental retardation protein and modulates its biological properties. *Mol. Cell. Biol.*, **22**, 8438–8447.
96. Khvorova, A., Reynolds, A. and Jayasena, S.D. (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell*, **115**, 209–216.
97. Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N. and Zamore, P.D. (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, **115**, 199–208.
98. Nelson, P.T., Hatzigeorgiou, A.G. and Mourelatos, Z. (2004) miRNP:mRNA association in polyribosomes in a human neuronal cell line. *RNA*, **10**, 387–394.
99. Song, J.J., Smith, S.K., Hannon, G.J. and Joshua-Tor, L. (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. *Science*, **305**, 1434–1437.
100. Lingel, A. and Izaurralde, E. (2004) RNAi: finding the elusive endonuclease. *RNA*, **10**, 1675–1679.
101. Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L. and Hannon, G.J. (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science*, **305**, 1437–1441.
102. Parker, J.S., Roe, S.M. and Barford, D. (2004) Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. *EMBO J.*, **23**, 4727–4737.
103. Dostie, J., Mourelatos, Z., Yang, M., Sharma, A. and Dreyfuss, G. (2003) Numerous microRNPs in neuronal cells containing novel microRNAs. *RNA*, **9**, 180–186.
104. Krichevsky, A.M., King, K.S., Donahue, C.P., Khrapko, K. and Kosik, K.S. (2003) A microRNA array reveals extensive regulation of microRNAs during brain development. *RNA*, **9**, 1274–1281.
105. Suh, M.R., Lee, Y., Kim, J.Y., Kim, S.K., Moon, S.H., Lee, J.Y., Cha, K.Y., Chung, H.M., Yoon, H.S., Moon, S.Y. *et al.* (2004) Human embryonic stem cells express a unique set of microRNAs. *Dev. Biol.*, **270**, 488–498.
106. Chen, C.Z., Li, L., Lodish, H.F. and Bartel, D.P. (2004) MicroRNAs modulate hematopoietic lineage differentiation. *Science*, **303**, 83–86.
107. Liu, C.G., Calin, G.A., Meloon, B., Gamliel, N., Sevignani, C., Ferracin, M., Dumitru, C.D., Shimizu, M., Zupo, S., Dono, M. *et al.* (2004) An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. *Proc. Natl Acad. Sci. USA*, **101**, 9740–9744.
108. Sempere, L.F., Freemantle, S., Pitha-Rowe, I., Moss, E., Dmitrovsky, E. and Ambros, V. (2004) Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. *Genome Biol.*, **5**, R13.
109. Calin, G.A., Liu, C.G., Sevignani, C., Ferracin, M., Felli, N., Dumitru, C.D., Shimizu, M., Cimmino, A., Zupo, S., Dono, M. *et al.* (2004) MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. *Proc. Natl Acad. Sci. USA*, **101**, 11755–11760.
110. Schmittgen, T.D., Jiang, J., Liu, Q. and Yang, L. (2004) A high-throughput method to monitor the expression of microRNA precursors. *Nucleic Acids Res.*, **32**, e43.
111. Kasashima, K., Nakamura, Y. and Kozu, T. (2004) Altered expression profiles of microRNAs during TPA-induced differentiation of HL-60 cells. *Biochem. Biophys. Res. Commun.*, **322**, 403–410.
112. Mansfield, J.H., Harfe, B.D., Nissen, R., Obenaus, 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.
113. Miska, E.A., Alvarez-Saavedra, E., Townsend, M., Yoshii, A., Sestan, N., Rakic, P., Constantine-Paton, M. and Horvitz, H.R. (2004) Microarray analysis of microRNA expression in the developing mammalian brain. *Genome Biol.*, **5**, R68.
114. Rogelj, B. and Giese, K.P. (2004) Expression and function of brain specific small RNAs. *Rev. Neurosci.*, **15**, 185–198.
115. Kim, J., Krichevsky, A., Grad, Y., Hayes, G.D., Kosik, K.S., Church, G.M. and Ruvkun, G. (2004) Identification of many microRNAs that copurify with polyribosomes in mammalian neurons. *Proc. Natl Acad. Sci. USA*, **101**, 360–365.
116. Ambros, V. (2003) MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. *Cell*, **114**, 269.
117. Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Muller, P. *et al.* (2000) Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA. *Nature*, **408**, 86–89.
118. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. and Tuschl, T. (2001) Identification of novel genes coding for small expressed RNAs. *Science*, **294**, 853–858.
119. Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W. and Tuschl, T. (2002) Identification of tissue-specific microRNAs from mouse. *Curr. Biol.*, **12**, 735–739.
120. Lee, R.C. and Ambros, V. (2001) An extensive class of small RNAs in *Caenorhabditis elegans*. *Science*, **294**, 862–864.
121. Moss, E.G. and Tang, L. (2003) Conservation of the heterochronic regulator Lin-28, its developmental expression and microRNA complementary sites. *Dev. Biol.*, **258**, 432–442.
122. Kloosterman, W.P., Wienholds, E., Ketting, R.F. and Plasterk, R.H. (2004) Substrate requirements for *let-7* function in the developing zebrafish embryo. *Nucleic Acids Res.*, **32**, 6284–6291.
123. Kiriakidou, M., Nelson, P.T., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z. and Hatzigeorgiou, A. (2004) A combined computational-experimental approach predicts human microRNA targets. *Genes Dev.*, **18**, 1165–1178.
124. Takamizawa, J., Konishi, H., Yanagisawa, K., Tomida, S., Osada, H., Endoh, H., Harano, T., Yatabe, Y., Nagino, M., Nimura, Y. *et al.* (2004) Reduced expression of the *let-7* microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res.*, **64**, 3753–3756.
125. Bernstein, E., Kim, S.Y., Carmell, M.A., Murchison, E.P., Alcorn, H., Li, M.Z., Mills, A.A., Elledge, S.J., Anderson, K.V. and Hannon, G.J. (2003) Dicer is essential for mouse development. *Nat. Genet.*, **35**, 215–217.
126. Kanelloupolou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M. and Rajewsky, K. (2005) Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev.*, **19**, 489–501.

127. Wienholds, E., Koudijs, M.J., van Eeden, F.J., Cuppen, E. and Plasterk, R.H. (2003) The microRNA-producing enzyme Dicer1 is essential for zebrafish development. *Nat. Genet.*, **35**, 217–218.
128. Esau, C., Kang, X., Peralta, E., Hanson, E., Marcusson, E.G., Ravichandran, L.V., Sun, Y., Koo, S., Perera, R.J., Jain, R. *et al.* (2004) MicroRNA-143 regulates adipocyte differentiation. *J. Biol. Chem.*, **279**, 52361–52365.
129. Poy, M.N., Eliasson, L., Krutzfeldt, J., Kuwajima, S., Ma, X., Macdonald, P.E., Pfeffer, S., Tuschl, T., Rajewsky, N., Rorsman, P. *et al.* (2004) A pancreatic islet-specific microRNA regulates insulin secretion. *Nature*, **432**, 226–230.
130. Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P. and Burge, C.B. (2003) Prediction of mammalian microRNA targets. *Cell*, **115**, 787–798.
131. Sleutels, F., Zwart, R. and Barlow, D.P. (2002) The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature*, **415**, 810–813.
132. Stark, A., Brennecke, J., Russell, R.B. and Cohen, S.M. (2003) Identification of *Drosophila* microRNA targets. *PLoS Biol.*, **1**, E60.
133. McManus, M.T. (2003) MicroRNAs and cancer. *Semin. Cancer Biol.*, **13**, 253–258.
134. Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K. *et al.* (2002) Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc. Natl Acad. Sci. USA*, **99**, 15524–15529.
135. Michael, M.Z., SM, O'Connor, S.M., van Holst Pellekaan, N.G., Young, G.P. and James, R.J. (2003) Reduced accumulation of specific microRNAs in colorectal neoplasia. *Mol. Cancer Res.*, **1**, 882–891.
136. Calin, G.A., Sevignani, C., Dumitru, C.D., Hyslop, T., Noch, E., Yendamuri, S., Shimizu, M., Rattan, S., Bullrich, F., Negrini, M. *et al.* (2004) Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc. Natl Acad. Sci. USA*, **101**, 2999–3004.
137. Scacheri, P.C., Rozenblatt-Rosen, O., Caplen, N.J., Wolfsberg, T.G., Umayam, L., Lee, J.C., Hughes, C.M., Shanmugam, K.S., Bhattacharjee, A., Meyerson, M. *et al.* (2004) Short interfering RNAs can induce unexpected and divergent changes in the levels of untargeted proteins in mammalian cells. *Proc. Natl Acad. Sci. USA*, **101**, 1892–1897.
138. Doench, J.G. and Sharp, P.A. (2004) Specificity of microRNA target selection in translational repression. *Genes Dev.*, **18**, 504–511.
139. 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.
140. Schramke, V. and Allshire, R. (2004) Those interfering little RNAs! Silencing and eliminating chromatin. *Curr. Opin. Genet. Dev.*, **14**, 174–180.
141. Lippman, Z. and Martienssen, R. (2004) The role of RNA interference in heterochromatic silencing. *Nature*, **431**, 364–370.
142. Mochizuki, K. and Gorovsky, M.A. (2004) Small RNAs in genome rearrangement in Tetrahymena. *Curr. Opin. Genet. Dev.*, **14**, 181–187.
143. Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I. and Martienssen, R.A. (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science*, **297**, 1833–1837.
144. Hall, I.M., Shankaranarayana, G.D., Noma, K., Ayoub, N., Cohen, A. and Grewal, S.I. (2002) Establishment and maintenance of a heterochromatin domain. *Science*, **297**, 2232–2237.
145. Volpe, T., Schramke, V., Hamilton, G.L., White, S.A., Teng, G., Martienssen, R.A. and Allshire, R.C. (2003) RNA interference is required for normal centromere function in fission yeast. *Chromosome Res.*, **11**, 137–146.
146. Hall, I.M., Noma, K. and Grewal, S.I. (2003) RNA interference machinery regulates chromosome dynamics during mitosis and meiosis in fission yeast. *Proc. Natl Acad. Sci. USA*, **100**, 193–198.
147. Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I. and Moazed, D. (2004) RNAi-mediated targeting of heterochromatin by the RITS complex. *Science*, **303**, 672–676.
148. Noma, K., Sugiyama, T., Cam, H., Verdel, A., Zofall, M., Jia, S., Moazed, D. and Grewal, S.I. (2004) RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing. *Nat. Genet.*, **36**, 1174–1180.
149. Watanabe, T., Miyashita, K., Saito, T.T., Yoneki, T., Kakhara, Y., Nabeshima, K., Kishi, Y.A., Shimoda, C. and Nojima, H. (2001) Comprehensive isolation of meiosis-specific genes identifies novel proteins and unusual non-coding transcripts in *Schizosaccharomyces pombe*. *Nucleic Acids Res.*, **29**, 2327–2337.
150. Pal-Bhadra, M., Leibovitch, B.A., Gandhi, S.G., Rao, M., Bhadra, U., Birchler, J.A. and Elgin, S.C. (2004) Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery. *Science*, **303**, 669–672.
151. Aravin, A.A., Klenov, M.S., Vagin, V.V., Bantignies, F., Cavalli, G. and Gvozdev, V.A. (2004) Dissection of a natural RNA silencing process in the *Drosophila melanogaster* germ line. *Mol. Cell Biol.*, **24**, 6742–6750.
152. 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.
153. Meehan, R.R., Kao, C.F. and Pennings, S. (2003) HP1 binding to native chromatin *in vitro* is determined by the hinge region and not by the chromodomain. *EMBO J.*, **22**, 3164–3174.
154. Akhtar, A., Zink, D. and Becker, P.B. (2000) Chromodomains are protein–RNA interaction modules. *Nature*, **407**, 405–409.
155. Tajul-Arifin, K., Teasdale, R., Ravasi, T., Hume, D.A. and Mattick, J.S. (2003) Identification and analysis of chromodomain-containing proteins encoded in the mouse transcriptome. *Genome Res.*, **13**, 1416–1429.
156. Brehm, A., Tufteland, K.R., Aasland, R. and Becker, P.B. (2004) The many colours of chromodomains. *Bioessays*, **26**, 133–140.
157. 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.
158. Kawasaki, H. and Taira, K. (2004) Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. *Nature*, **431**, 211–217.
159. 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.
160. 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.
161. Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gascioli, V., Mallory, A.C., Hilbert, J.L., Bartel, D.P. and Crete, P. (2004) Endogenous trans-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. *Mol. Cell*, **16**, 69–79.
162. Yuan, G., Klambt, C., Bachellerie, J.P., Brosius, J. and Huttenhofer, A. (2003) RNomics in *Drosophila melanogaster*: identification of 66 candidates for novel non-messenger RNAs. *Nucleic Acids Res.*, **31**, 2495–2507.
163. 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.
164. Lai, E.C., Wiel, C. and Rubin, G.M. (2004) Complementary miRNA pairs suggest a regulatory role for miRNA:miRNA duplexes. *RNA*, **10**, 171–175.
165. Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T. and Jewell, D. (2003) MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Curr. Biol.*, **13**, 807–818.
166. Johnston, R.J. and Hobert, O. (2003) A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature*, **426**, 845–849.
167. Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B. and Cohen, S.M. (2003) Bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell*, **113**, 25–36.
168. Smit, M., Segers, K., Carrasco, L.G., Shay, T., Baraldi, F., Gyapay, G., Snower, G., Georges, M., Cockett, N. and Charlier, C. (2003) Mosaicism of solid gold supports the causality of a noncoding A-to-G transition in the determinism of the callipyge phenotype. *Genetics*, **163**, 453–456.
169. 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.

170. 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.
171. Ambros, V. and Lee, R.C. (2004) Identification of microRNAs and other tiny noncoding RNAs by cDNA cloning. *Methods Mol. Biol.*, **265**, 131–158.
172. Washietl, S., Hofacker, I.L. and Stadler, P.F. (2005) From The Cover: Fast and reliable prediction of noncoding RNAs. *Proc. Natl Acad. Sci. USA*, **102**, 2454–2459.
173. Khorasanizadeh, S. (2004) The nucleosome: from genomic organization to genomic regulation. *Cell*, **116**, 259–272.
174. Ohno, M., Fukagawa, T., Lee, J.S. and Ikemura, T. (2002) Triplex-forming DNAs in the human interphase nucleus visualized *in situ* by polypurine/polypyrimidine DNA probes and antitriplex antibodies. *Chromosoma*, **111**, 201–213.
175. Shi, Y. and Berg, J.M. (1995) Specific DNA–RNA hybrid binding by zinc finger proteins. *Science*, **268**, 282–284.
176. Lodomery, M. (1997) Multifunctional proteins suggest connections between transcriptional and post-transcriptional processes. *Bioessays*, **19**, 903–909.
177. Holliday, R. and Murray, V. (1994) Specificity in splicing. *Bioessays*, **16**, 771–774.
178. Lopez, A.J. (1998) Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu. Rev. Genet.*, **32**, 279–305.
179. Singh, R. (2002) RNA–protein interactions that regulate pre-mRNA splicing. *Gene Expr.*, **10**, 79–92.
180. Sorek, R. and Ast, G. (2003) Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. *Genome Res.*, **13**, 1631–1637.
181. 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.
182. Duncley, M.G., Manoharan, M., Villiet, P., Eperon, I.C. and Dickson, G. (1998) Modification of splicing in the dystrophin gene in cultured Mdx muscle cells by antisense oligoribonucleotides. *Hum. Mol. Genet.*, **7**, 1083–1090.
183. Mann, C.J., Honeyman, K., Cheng, A.J., Ly, T., Lloyd, F., Fletcher, S., Morgan, J.E., Partridge, T.A. and Wilton, S.D. (2001) Antisense-induced exon skipping and synthesis of dystrophin in the mdx mouse. *Proc. Natl Acad. Sci. USA*, **98**, 42–47.
184. Kole, R. and Sazani, P. (2001) Antisense effects in the cell nucleus: modification of splicing. *Curr. Opin. Mol. Ther.*, **3**, 229–234.
185. Suwanmanee, T., Sierakowska, H., Lacerra, G., Svasti, S., Kirby, S., Walsh, C.E., Fucharoen, S. and Kole, R. (2002) Restoration of human beta-globin gene expression in murine and human IVS2–654 thalassemic erythroid cells by free uptake of antisense oligonucleotides. *Mol. Pharmacol.*, **62**, 545–553.
186. Sazani, P., Gemignani, F., Kang, S.H., Maier, M.A., Manoharan, M., Persmark, M., Bortner, D. and Kole, R. (2002) Systemically delivered antisense oligomers upregulate gene expression in mouse tissues. *Nat. Biotechnol.*, **20**, 1228–1233.
187. Lim, L.P., Lau, N.C., Garrett-Engle, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S. and Johnson, J.M. (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, **433**, 769–73.
188. 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.
189. Nesterova, T.B., Slobodyanyuk, S.Y., Elisaphenko, E.A., Shevchenko, A.I., Johnston, C., Pavlova, M.E., Rogozin, I.B., Kolesnikov, N.N., Brockdorff, N. and Zakian, S.M. (2001) Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. *Genome Res.*, **11**, 833–849.
190. 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.
191. Smith, N.G., Brandstrom, M. and Ellegren, H. (2004) Evidence for turnover of functional noncoding DNA in mammalian genome evolution. *Genomics*, **84**, 806–813.