TIMELINE

The rise of regulatory RNA

Kevin V. Morris and John S. Mattick

Abstract | Discoveries over the past decade portend a paradigm shift in molecular biology. Evidence suggests that RNA is not only functional as a messenger between DNA and protein but also involved in the regulation of genome organization and gene expression, which is increasingly elaborate in complex organisms. Regulatory RNA seems to operate at many levels; in particular, it plays an important part in the epigenetic processes that control differentiation and development. These discoveries suggest a central role for RNA in human evolution and ontogeny. Here, we review the emergence of the previously unsuspected world of regulatory RNA from a historical perspective.

RNA has long been at the centre of molecular biology and was likely the primordial molecule of life, encompassing both informational and catalytic functions. Its informational functions are thought to have subsequently devolved to the more stable and easily replicable DNA, and its catalytic functions to the more chemically versatile polypeptides[1]. The idea that the contemporary role of RNA is to function as the intermediary between the two had its roots in the early 1940s with the entry of chemists into the study of biology, notably Beadle and Tatum[2], whose work underpinned the one gene–one enzyme hypothesis (FIG. 1 (TIMELINE)). This idea later matured into the more familiar one gene–one protein concept and became widely accepted despite the prescient misgivings of experienced geneticists, notably McClintock[3]. The concept that genes encode only the functional components of cells (that is, the 'enzymes') itself had deeper roots in the mechanical zeitgeist of the era, which was decades before the widespread understanding of the use of digital information for systems control.

Although the one gene–one protein hypothesis has long been abandoned owing to the discovery of alternative splicing in the 1970s, the protein-centric view of molecular biology has persisted. Such persistence was aided by phenotypic and ascertainment bias towards protein-coding mutations in genetic studies and by the assumption that these mutations affected *cis*-acting regulatory protein-binding sites[4]. However, this view was challenged by the discovery of nuclear introns and RNA interference (RNAi), as well as by the advent of high-throughput sequencing, which led to the identification of large numbers and different types of large and small RNAs, the functions of which are still under investigation.

> "emerging evidence suggests … that the amount and type of gene regulation in complex organisms have been substantially misunderstood"

In this Timeline article, we examine the history of, and report the shift in thinking that is still underway about, the role of RNA in cell and developmental biology, especially in animals. The emerging evidence suggests that there are more genes encoding regulatory RNAs than those encoding proteins in the human genome, and that the amount and type of gene regulation in complex organisms have been substantially misunderstood for most of the past 50 years.

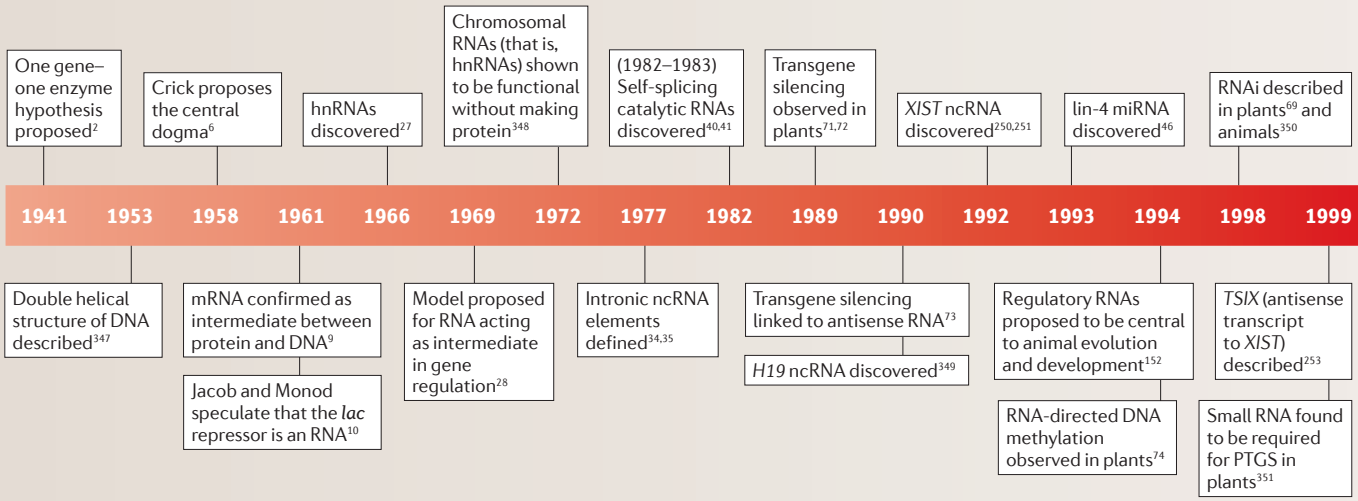
Early ideas for the role of RNA

RNA — the central dogma and gene regulation. After the elucidation of the double helical structure of DNA in 1953 (REF. 5), the following years were preoccupied with deciphering the 'genetic code' and establishing the mechanistic pathway between genes and proteins: the identification of a transitory template (mRNA), an adaptor (tRNA) and the ribosome 'factory' comprised of ribosomal RNAs and proteins for translating the code into a polypeptide. In 1958, Crick published the celebrated central dogma to describe the flow of genetic information from DNA to RNA to protein, which has proved remarkably accurate and durable, including the prediction of reverse transcription[6]. Nonetheless, in conceptual terms, RNA was tacitly consigned to be the template and an infrastructural platform (with regard to rRNAs and tRNAs) for protein synthesis or has at least been interpreted in this way by most people since that time.

In the mid-1950s, the link was established between rRNA (which is highly expressed in essentially all cells) and the structures termed ribosomes as the platform for protein synthesis[7]. The roles of tRNA and mRNA were experimentally confirmed in 1958 (REF. 8) and 1961 (REF. 9), respectively. The latter occurred in the same year that Jacob and Monod published their classic paper on the *lac* operon of *Escherichia coli*[10], which was the first locus to be characterized at the molecular genetic level. These studies confirmed that at least some, but presumably most, genes encoded proteins and supported the emerging idea that gene expression is controlled by regulating the transcription of the gene, as indicated by the locus encoding the *lac* repressor in the repressor–operator model. At the time, Jacob and Monod did not know the chemical identity of the repressor and speculated in passing that it "may be a polyribonucleotide" (that is, RNA)[10]. However, Gilbert later showed that the repressor is a polypeptide that allosterically binds to the lactose substrate, and the brief idea faded[11].

These studies reinforced and extended the concept that proteins are not only enzymes but also the primary analogue components and control factors that constitute the cellular machinery. This, in turn, has led to the prevailing transcription factor

Timeline | **The rise of regulatory RNA**



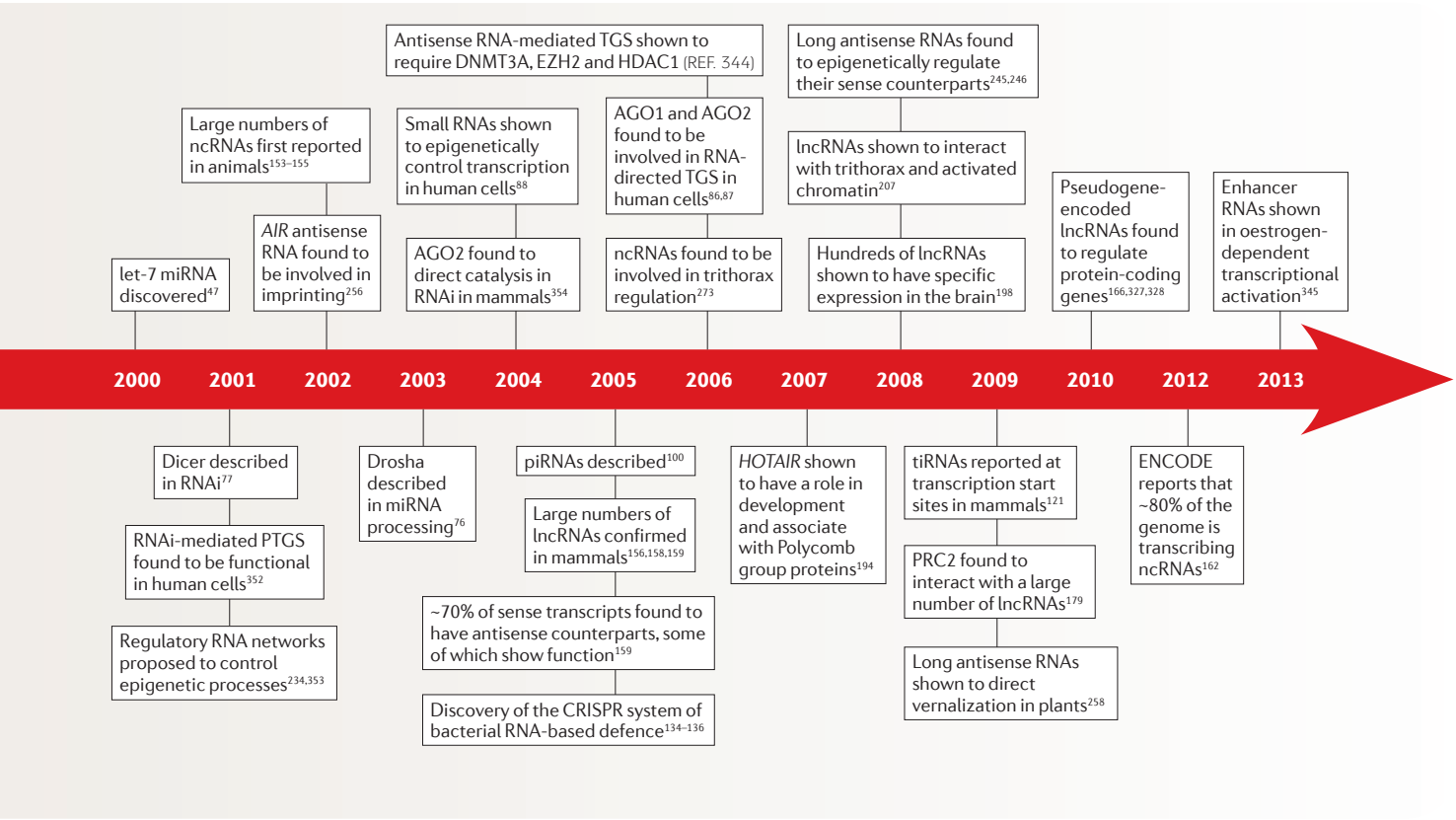
AGO, Argonaute; *AIR*, also known as *AIRN* (antisense of *IGF2R* non-protein coding RNA); CRISPR, clustered regularly interspaced short palindromic repeat; DNMT3A, DNA (cytosine-5)-methyltransferase 3A; ENCODE, Encyclopedia of DNA Elements; EZH2, enhancer of Zeste 2; *H19*, H19 imprinted maternally expressed transcript; HDAC1, histone deacetylase 1; hnRNA, heterogeneous nuclear RNA; *HOTAIR*, HOX transcript antisense RNA; lncRNA, long non-coding RNA; miRNA, microRNA; ncRNA, non-coding RNA; piRNA, PIWI-interacting RNA; PRC2, Polycomb repressive complex 2; PTGS, post-transcriptional gene silencing; RNAi, RNA interference; TGS, transcriptional gene silencing; tiRNA, transcription initiation RNA; *XIST*, X inactive specific transcript.

paradigm of gene regulation, including the derived assumption that combinatorial interactions would provide an enormous range of regulatory possibilities[12] that are more than enough to control human ontogeny. However, this assumption has not been substantiated theoretically or mechanistically, and both the observed scaling of regulatory genes and the extent of the regulatory challenge in programming human developmental architecture seem to be different from these expectations[13]. In this context, it is noteworthy that genome-wide association studies have shown that most haplotype blocks influencing complex diseases are outside the known boundaries of protein-coding genes[14].

Small nuclear RNAs and small nucleolar RNAs. Following the discovery and functional description of tRNAs and rRNAs, new classes of common small RNAs in the nucleus were identified by biochemical fractionation[15]. Many of these small RNAs were found to be part of ribonucleoprotein (RNP) complexes (reviewed in REF. 16). One class — the small nuclear RNAs (snRNAs) (FIG. 2) — was later found to be a central cofactor in RNA splicing[17] (see below) and was therefore given the newer designation as spliceosomal RNAs. The snRNAs U1, U2, U4, U5 and U6 participate in various RNA–RNA and RNA–protein interactions in the assembly and function of canonical spliceosomes: U1 and U2 recognize the 5′ splice site and the branch point, respectively, followed by the recruitment of U4, U5 and U6, which displace U1 and interact with U2 (through U6) as well as the 5′ and 3′ splice sites (through U5)[18]. A set of less abundant snRNAs (U11, U12, U4atac and U6atac) and U5 are found in a variant 'minor' spliceosome termed U12-type[19].

Other small RNAs were found to be localized to the nucleolus and to guide the methylation (the box C/D subclass) and pseudouridylation (the box H/ACA subclass) of rRNAs, tRNAs and snRNAs[20–22] (FIG. 2). The chemical modifications of rRNAs, tRNAs and snRNAs proved to be essential in ribosomal and cellular function, particularly in tRNA and mRNA maturation, and in pre-mRNA splicing (which requires modification of the U2 snRNA). Notably, the disruption of small nucleolar RNAs (snoRNAs) was found to cause a loss of processing of the 5.8S, 18S and 28S (or 25S in plants) rRNAs[20]. Early studies found that some snoRNAs are subject to parental imprinting and/or differentially expressed (for example, in the brain[23,24]), and that they seem to target a wide range of RNAs (including mRNAs[25]), which suggests a regulatory role. Related small RNAs have also been identified in subnuclear structures called Cajal bodies (which process telomerase RNA), and these were termed small Cajal body-specific RNAs (scaRNAs)[26]. However, none of these studies suggested anything other than that the role of RNA was limited to protein synthesis.

The emergence of heterogeneous nuclear RNAs. The first hint that RNA may have additional roles in complex organisms was the discovery of heterogeneous nuclear RNA (hnRNA)[27] and the observation that the complexity of this population, as determined by denaturation–renaturation hybridization kinetics, was much greater in the nucleus than in the cytoplasm. The existence of hnRNA and the concomitant discovery of the large amount of repetitive sequences (that is, different classes of retrotransposon sequences with similar composition that occupy large portions of plant and animal genomes) led Britten and Davidson to speculate in 1969 that animal cells contain extensive RNA-based regulatory networks[28–30]. Although this hypothesis attracted a great deal of interest at the time, it also quickly



lapsed. Its proponents did not revisit the hypothesis even after the subsequent discovery of introns (see below) and instead focused on regulatory networks controlled by transcription factors[31,32] or on the importance of transposons in protein evolution[33].

The discovery of introns. The discovery of introns in 1977 (REFS 34,35) was perhaps the biggest surprise in the history of molecular biology[36] (FIG. 1 (TIMELINE)), as no one expected that the genes of higher organisms would be mosaics of coding and non-coding sequences, all of which are transcribed. However, the prevailing concept of the flow of genetic information was not overly disturbed, as the removal of the intervening sequences (that is, introns) and the reconstruction of a mature mRNA by splicing preserved the conceptual status quo; that is, genes still made proteins. In parallel, it was assumed that the excised intronic RNAs were simply degraded, although the technology of the time was too primitive to confirm this. In any case, introns were immediately and universally dismissed as genomic debris, and their presence was rationalized as evolutionary remnants involved in the prebiotic modular assembly of protein-coding RNAs that have remained (and been expanded by transposition) in complex organisms[37]. This notion was consistent, at least superficially, with the implication of the C-value enigma that eukaryotes contained varying amounts of DNA 'baggage'. It is also in agreement with the accompanying conclusion that retrotransposon sequences are mainly 'selfish', parasitic DNA[38,39].

RNA as a catalyst. A few years later, Cech, Altman and colleagues demonstrated that RNA itself was capable of enzymatic catalysis (that is, they are ribozymes)[40,41], which provided evidence in support of the RNA early hypothesis. They also showed that RNA catalysis exists and has persisted in particular contexts, notably at the core of RNA splicing[42] and mRNA translation[43]. This finding reinforced both the mechanical concept of molecular biology and the role of RNA as the platform for protein synthesis, but did not give any hint of RNA as a widespread regulatory factor, although that possibility is perfectly feasible. Indeed, there is increasing evidence that catalytic RNA exists in animal and plant cells, in introns, untranslated regions (UTRs) and elsewhere, and that these RNAs may have various roles, for example, in the regulation of post-transcriptional cleavage reactions[44,45].

The small RNA revolution

The discovery of microRNAs. In 1993, Ambros and colleagues showed the first evidence for small (~22-nucleotide) regulatory RNAs with the discovery of the genetic loci *lin-4* and *let-7*, which regulate the timing of *Caenorhabditis elegans* development[46,47] (FIG. 1 (TIMELINE)). Although *let-7* is highly conserved from nematodes to humans[48], very few microRNAs (miRNAs) were discovered genetically[49,50], and these RNAs remained interesting idiosyncrasies until the discovery of RNAi (see below). This discovery led to the targeted cloning after size selection of many more miRNAs[51–53] and the demonstration that these miRNAs act, at least partly, by imperfect base-pairing — typically with the 3′UTRs of target mRNAs — to inhibit their translation and to accelerate their degradation[54].

Current databases list large numbers of evolutionarily widespread miRNAs[55], almost all of which had evaded prior detection by genetic screens but many were subsequently validated by reverse genetics. Although many miRNAs can be identified by conservation, it is also evident that many are tissue and lineage specific[56,57], and that there may be many more to be discovered.

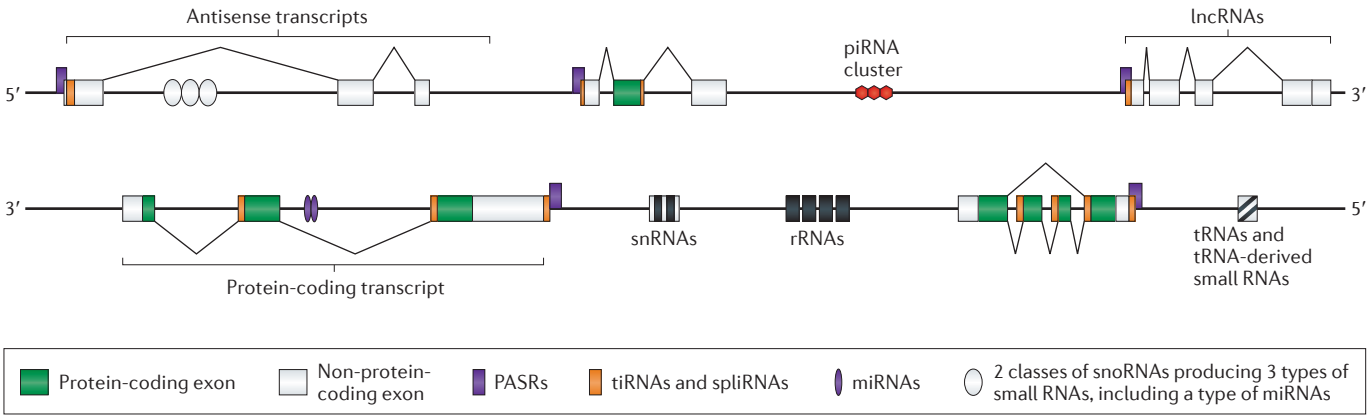


Figure 2 | **Complex expression of the genome and examples of non-coding RNA expression.** The mammalian transcriptional landscape is represented graphically with genes expressing ribosomal RNAs, tRNAs, small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), various protein-coding and non-coding genes (which encode mRNAs and long non-coding RNAs (lncRNAs), respectively), as well as genes expressing small regulatory RNAs such as microRNAs (miRNAs), PIWI-interacting RNAs (piRNAs), promoter-associated short RNAs (PASRs), transcription initiation RNAs (tiRNAs) and splice site RNAs (spliRNAs), snoRNA-derived small RNAs and tRNA-derived small RNAs. The transcriptional units are not depicted to scale.

There is also evidence that many, if not most, protein-coding transcripts are targets for miRNA regulation[58,59]. In some cases, miRNAs can regulate large numbers of target mRNAs[60] and, reciprocally, many mRNAs contain target sites for a large number of miRNAs[61], although the implied regulatory logic of this complex multiplex arrangement has not been explained. The targets of miRNAs are usually thought to be mRNAs but may also include other types of RNAs[62]. Biologically, miRNAs have been shown to regulate many physiological, developmental and disease processes, including pluripotency[63], epithelial–mesenchymal transition and metastasis[64], testis differentiation[65], diabetes[66], and neural plasticity and memory[67], among others[68].

The RNA interference pathway. miRNAs are only one aspect of the phenomenon of RNAi, which silences gene expression after the introduction of sense–antisense RNA pairs. This process was discovered in 1998 in plants[69] and *C. elegans*[70] (FIG. 1 (TIMELINE)). These discoveries were presaged by the curious phenomenon of transgene silencing, which is mainly found in plants[71,72] and linked to both antisense RNA and small RNA-directed DNA methylation, thus indicating transcriptional and post-transcriptional silencing[73,74]. Mechanistic analyses of these silencing mechanisms showed that exogenous double-stranded RNA (dsRNA) is processed into short fragments (known as small interfering RNAs (siRNAs)) with similar sizes to miRNAs, which implies that miRNAs may represent a similar endogenous system.

This hypothesis was confirmed and led to the elucidation of natural dsRNA precursors in stem–loop structures[75], as well as the identification of key genes and enzymes involved in their biogenesis and function, notably Drosha[76], Dicer[77] and several Argonaute (AGO) proteins[78]. AGO proteins were already known to have central roles in differentiation and development[79] but are now known to also be involved in defence against RNA viruses in many organisms[80]. Drosha and exportin 5 are involved in the cleavage and export of dsRNA precursors from the nucleus to the cytoplasm[76], where they are further processed by Dicer to small (21–24-nucleotide) dsRNA moieties. One strand of the dsRNA is loaded into the AGO component of the RNA-induced silencing complex (RISC), which also comprises other proteins[77]. The RISC is guided by the small RNA strand to complementary RNA targets, which are subsequently silenced by translational repression and/or RNA destabilization[81,82] (FIG. 3).

Although still under discussion, the current view is that siRNAs (and short hairpin RNAs (shRNAs)) — which seem to naturally occur more commonly in plants — act primarily by perfect base-pairing and by AGO-mediated cleavage of complementary target RNAs; hence, they are used widely as experimental tools and potential therapeutic agents[83]. By contrast, miRNAs have incomplete homology with their target sequences and act primarily at the translational level[81,82] (FIG. 3).

Both miRNAs and siRNAs are thought to act post-transcriptionally in the cytoplasm, but the existence of AGO in the nucleus[84–87] and the role of the RNAi pathway in epigenetic modulation[88] suggest that the system is more complex and multifaceted than expected. For example, it has been shown that miRNA isoforms are developmentally regulated[89], that the target 'seed' sequence is only one factor in target recognition[90,91] and that miRNAs can also impose transcriptional gene silencing[92] (FIG. 3). There is also growing evidence of intersecting pathways, such as RNA editing and modification, in these networks[93–96].

PIWI-associated small RNAs. Although most AGO proteins are expressed ubiquitously and associate with both miRNAs and siRNAs, there is a subclade of AGO proteins termed PIWI that are required for germ cell development[97–100]. PIWI and PIWI-like proteins associate with a distinctive class of small (26–30-nucleotide) RNAs termed PIWI-interacting RNAs (piRNAs), which epigenetically and post-transcriptionally silence transposons in germ cells[101–110]. PIWI is found predominantly in the nucleus[111], colocalizes in an RNA-dependent manner with Polycomb group proteins[112] and seems to be expressed in other tissues (including the brain[113]), which suggests a role beyond genome protection in epigenetic processes[114,115].

Other classes of small RNAs in eukaryotes. The molecular genetics, biochemistry and structural biology of the RNAi system are still being unravelled but indicate an ancient, widespread and multilaterally adapted system that controls many cellular processes, the dimensions of which are still being explored. These include potentially lineage-specific

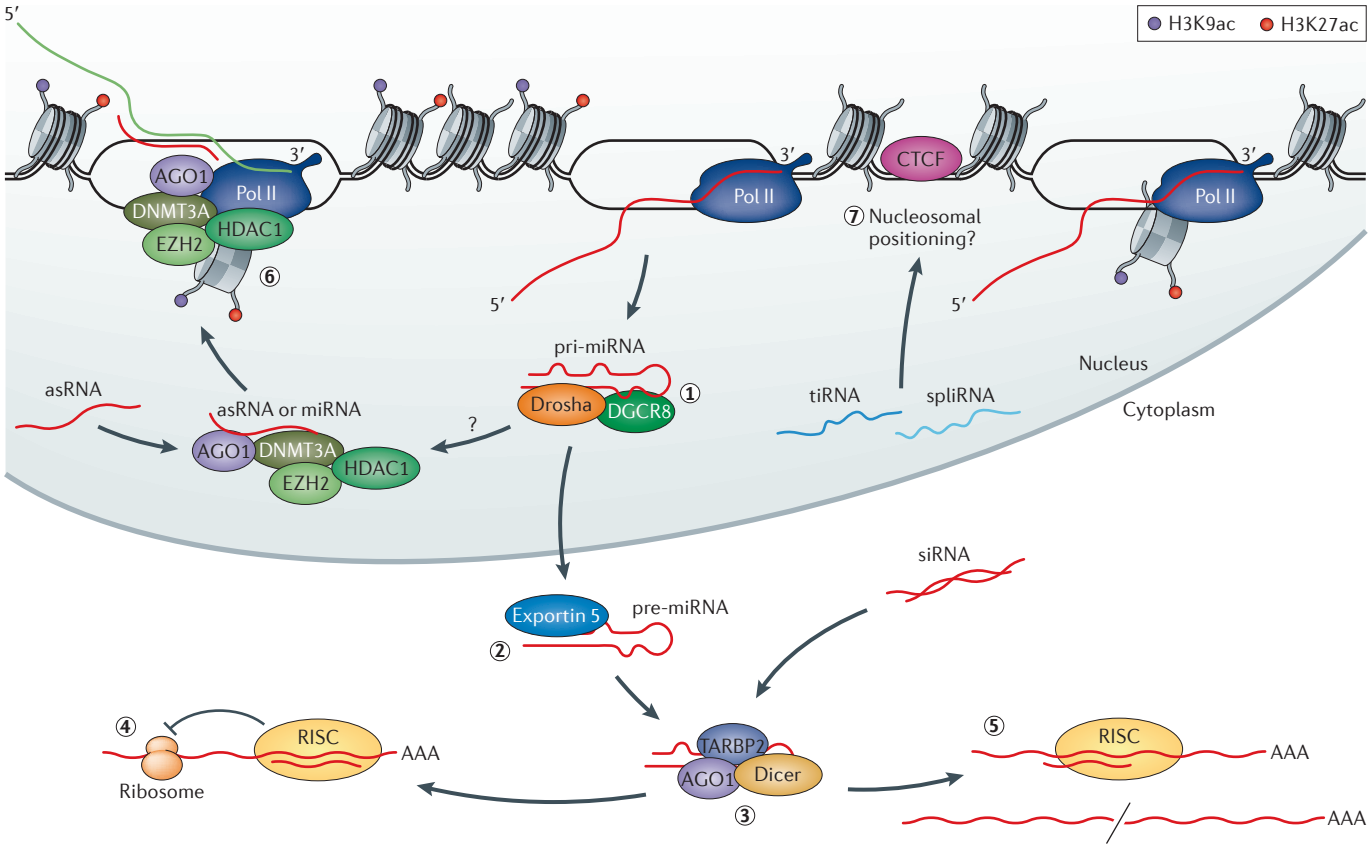


Figure 3 | **Functional pathways of small regulatory RNAs.** MicroRNA (miRNA) precursors (that is, pri-miRNAs) are expressed as stem–loop structures[75], which interact with Drosha[76] and DGCR8 (also known as Pasha) (step 1). They are then processed into pre-miRNAs and exported from the nucleus by exportin 5 (step 2). These transcripts are further processed by Dicer to small (21–23-nucleotide) double-stranded RNAs, one strand of which is loaded into the Argonaute (AGO) component of the RNA-induced silencing complex (RISC) (step 3). Exogenously introduced small interfering RNAs (siRNAs) can also be processed by RISC. The endogenous miRNA or siRNA, or exogenously added siRNA, can then target the repression of translation (step 4) and/or cleavage of homology-containing transcripts[81,82] (step 5). Some small RNAs are functional in the nucleus. Exogenously introduced small antisense RNAs (asRNAs) can induce epigenetic silencing of targeted loci[88,342,343] — a pathway that miRNAs may also use in the nucleus[92] (step 6). Transcription initiation RNAs (tiRNAs) and splice site RNAs (spliRNAs)[121,122] are expressed through an unknown pathway that may involve RNA polymerase II (Pol II) backtracking and TFIIS cleavage[123] (not shown); tiRNAs and spliRNAs are shown to modulate CCCTC-binding factor (CTCF) chromatin localization and to be associated with nucleosome positioning[124] (step 7). DNMT3A, DNA (cytosine-5)-methyltransferase 3A; EZH2, enhancer of Zeste 2; H3K9ac, histone H3 lysine 9 acetylation; HDAC1, histone deacetylase 1; TARBP2, RISC-loading complex subunit TARBP2 (also known as TRBP).

variations such as the 21U RNAs in *C. elegans*[116]. Surprisingly, it seems that all snoRNAs from fission yeast to humans produce at least three different subclasses of small RNAs[117], one of which has the same size and functions as miRNAs[118], and another that is similar in size to piRNAs[117]. There are also intriguing and recurring reports of tRNA fragments that are produced in tissue-specific patterns[119] and that are associated with AGO proteins[120].

More recently, deep sequencing of small RNA populations has revealed the existence of two other classes of small RNAs in animals but not in plants, which are 17–18 nucleotides in length and associated with transcription initiation[121] and splice sites[122] (termed transcription initiation RNAs (tiRNAs) and splice site RNAs (spliRNAs), respectively) (FIG. 3). The origin and function of these RNAs are uncertain, but preliminary evidence suggests that they play a part in nucleosome positioning[123] and/or in other levels of chromatin organization[124]. There are also other reports of less distinct classes of promoter-associated RNAs called promoter-associated short RNAs (PASRs)[125], transcription start site-associated RNAs (TSSa-RNAs)[126] and promoter upstream transcripts (PROMPTS)[127], some of which may have a role in RNA-directed transcriptional gene silencing[128].

Regulatory RNAs in bacteria and archaea. Many small regulatory RNAs have been identified in bacteria, in which they regulate a wide variety of adaptive responses. Bacterial small regulatory RNAs generally function by simple antisense mechanisms to regulate translation or stability of target mRNAs through altering their secondary structure to expose or sequester *cis*-acting sites[129,130]. Studies in bacteria have also identified *cis*-acting regulatory RNA sequences known as riboswitches, which act allosterically by binding metabolites to regulate gene expression[131,132] and almost certainly exist as part of the RNA regulatory landscape in all kingdoms of life.

Very recently, the bacterial and archaeal kingdoms have once again surprised us with the sophistication of their molecular machinery. Many bacterial and most

archaeal genomes have loci comprised of regularly spaced repeats that are interspersed by other virus-derived DNA sequences[133–136] (termed clustered regularly interspaced short palindromic repeats (CRISPRs)). These loci act as an innate immune system by incorporating fragments of viral DNA between the repeats, which are then transcribed and processed to produce small guide RNAs that are linked to their effector complexes through the repeat sequence and that target and destroy viral DNA[137–140] or RNA[141]. This system has recently been adapted for RNA programmable sequence-specific genome manipulation in eukaryotes (including mammals[142–145]) with extraordinary versatility, including targeted gene excision and fusion, as well as engineered CRISPRs that can recruit silencing and activating proteins to target loci[146–150]. Moreover, the biological 'arms race' continues, as bacteriophages encode their own CRISPR system to evade host innate immunity[151].

Long non-coding RNAs

The eukaryotic transcriptome. Noting that the density and size of introns (and, as it turned out later, intergenic sequences) increased with developmental complexity, Mattick posited in 1994 that introns had evolved to express an expanding range of *trans*-acting regulatory RNAs (FIG. 1 (TIMELINE)). He postulated that some genes subsequently evolved to express only intronic or exonic regulatory RNAs, and that this RNA-based regulatory system was the essential prerequisite for the emergence of developmentally complex organisms[152]. Subsequently, the application of genome tiling array technology and deep sequencing to the characterization of the transcriptome showed that tens of thousands of loci in mammals express long transcripts that do not encode proteins, which are located intergenic, intronic and antisense to protein-coding genes. The initial findings[153–155] were confirmed in 2005 (REFS 156–159) and extended by the Encyclopedia of DNA Elements (ENCODE) project[160–162], all of which showed that the vast majority (at least 80%) of the human and mouse genomes are differentially transcribed in one context or another; other studies also reported similar findings in all organisms examined. Indeed, it seems that most intergenic and, by definition, intronic sequences are differentially transcribed, and that the extent of the transcriptome therefore expands with developmental complexity[163].

Using more focused deep sequencing methodologies, it has become evident that the full range of the protein-coding and non-protein-coding transcriptome is still vastly under-sampled[164]. In addition, many transcripts are not polyadenylated and represent a largely different sequence class[156,165], some of which seem to be relevant to development (for example, the *POU5F1* (also known as *OCT4*) transcript[166,167]). Moreover, 95% of human transcription initiation sites are not associated with mRNA transcription but rather mainly with transcription of non-polyadenylated non-coding RNAs (ncRNAs)[168]. These non-polyadenylated transcripts are so far mostly uncharacterized because of the historical use of poly(A) tails to remove the overwhelming rRNA contamination in RNA preparations. This issue is being alleviated by more sophisticated approaches such as cap trapping[169], oligonucleotide subtraction[170] and array capture[164,171].

Defining long non-coding RNAs. Long ncRNAs (lncRNAs) are operationally defined as non-protein-coding RNAs that are >200 nucleotides in length, which corresponds to a convenient cutoff in biochemical fractionation and excludes all known classes of small RNAs[172]. Transcripts are classified as non-coding if they lack long open reading frames (traditionally >100 codons) and/or do not show codon conservation, although there was considerable uncertainty, as genomic and transcriptomic data were initially limited for comparison. However, recent analyses provide strong evidence that most annotated lncRNAs do not encode proteins; nonetheless, some specify small proteins that had not been identified previously using bioinformatic approaches[173–175].

These ncRNAs can be parsed into intronic, antisense or intergenic (that is, large intergenic non-coding RNA (lincRNA)) subsets in experimental studies and databases[159,176,177], partly because of mechanistic expectations[178] and because of a desire to reduce ambiguity and overlap with protein-coding loci in functional analyses[179–181]. However, there is no evidence of any intrinsic difference between RNAs that are intronic, intergenic or antisense, or that overlap with protein-coding transcripts (FIG. 2), for example, in their interaction with chromatin-activating or chromatin-repressive complexes (see below). Nonetheless, ncRNA subclasses will inevitably exist and be defined, some of which may be biased in relation to genomic origin.

Long non-coding RNAs: transcriptional noise or functional? The unexpected discovery of large numbers of non-coding transcripts in eukaryotes, some of which span tens or hundreds of kilobases[182], led to debates about their functionality[183,184]. In particular, as many lncRNAs were shown to have fairly low evolutionary conservation and low levels of expression, some have posited that they represent transcriptional noise and/or redundant transcripts with no biological importance. This hypothesis remains, at least partly, a possibility. Nevertheless, lncRNAs show a wide range of evolutionary conservation, from ultraconserved ones[185] to primate-specific ones[186–188], which can be explained as the result of different structure–function constraints and lineage-specific adaptive radiation[189]. Indeed, there is now considerable evidence that lack of primary sequence conservation in lncRNAs does not indicate lack of function[190,191], and many lncRNAs show evidence of structural conservation[192,193].

Loci that express lncRNAs show all of the hallmarks of bona fide genes[4], including conservation of promoters[169], indicative chromatin structure[194], and regulation by conventional morphogens and transcription factors[195]. Moreover, lncRNAs were found to have a similar range of cellular half-lives as mRNAs[196] and to be differentially expressed in a tissue-specific manner[158,197], especially in the brain[198]. The study in the brain showed that, although the expression levels of many lncRNAs seem to be lower than those of mRNAs in whole tissues, lncRNAs are highly expressed and easily detectable in particular cell types[198]. In addition, lncRNAs were found to have, on average, higher cell specificity than proteins[165,199]; this is consistent with their proposed role in architectural (as opposed to 'cell-type') regulation, in which each cell has a unique positional identity in precisely sculpted organs, bones and muscles[200].

Many lncRNAs are alternatively spliced[201], which is further evidence of the precision of their expression and is hard to reconcile with the suggestion that they are simply transcriptional noise. It should also be noted that some functionally validated lncRNAs can have isoforms that encode proteins[202] and that, reciprocally, some (perhaps many) mRNAs have intrinsic functions as *trans*-acting regulatory RNAs[203–205]. In some contexts, 3′UTRs can be separately expressed and convey genetic functions in *trans*[204], and both lncRNAs and mRNAs may be further processed to produce subsidiary species[206].

lncRNAs have been shown to be dynamically expressed in a range of differentiating systems, including embryonic stem cells[207], muscles[208], T cells[209], breast tissues[210,211], the erythroid system[211] and neurons[212–214], as well as in cancer and other diseases[210,215–222]. Such dynamic expression of lncRNAs is at least partly controlled by conventional transcription factors[195,213].

Emerging roles of non-coding RNAs

The validation of ncRNA functions has so far mainly relied on knockdown of candidate ncRNAs. Knockdown of ncRNA expression has proved to be surprisingly easy using chemically engineered antisense oligonucleotides, or using siRNA- or shRNA-mediated approaches, frequently resulting in phenotypic changes in cultured cells, in which most studies have been carried out.

Development and differentiation. Many small ncRNAs[63–65] and most functionally analysed lncRNAs[223] seem to have a role in the regulation of differentiation and development. On the basis of studies in cell culture, these include the regulation of apoptosis and metastatic processes[211,218,220,221,224], retinal and erythroid development[211,225], breast development[210,226] and epidermal differentiation[227], among many others. Antisense knockdown of some lncRNAs in zebrafish and deletion of sequences that specify lncRNAs in mice have resulted in visible developmental defects[181,191,228,229]. However, knockouts of the widely expressed nuclear paraspeckle assembly transcript 1 (*Neat1*)[230] or of some of the most highly conserved sequences in the mammalian genome[231] have not shown any detrimental effect on development. These results suggest that more sophisticated phenotypic screens are required to delineate functions, especially cognitive ones, because most mammalian lncRNAs are expressed in the brain[198] and many are specific to mammals or primates[188,232]. A good example is brain cytoplasmic RNA 1 (*BC1*) — a retrotransposon-derived lncRNA that is widely expressed in the brain — the knockout of which causes no visible anatomical abnormality but leads to behavioural changes that would be lethal in the wild[233].

Epigenetic roles. Consistent with their roles in differentiation and development, a range of genetic and biochemical evidence suggests that a major function of many small RNAs and lncRNAs is the regulation of epigenetic processes[234,235], probably by guiding chromatin-modifying enzymes to their sites of action and/or by acting as scaffolds for chromosomal organization[179,235–238] (FIG. 4).

RNAs were shown to induce transcriptional gene silencing first in plants[74,239], then in fungi[240] and human cells[88], and both small RNAs and the RNAi machinery were implicated in the underlying epigenetic processes[240–242]. These studies were consistent with the observations that small RNAs interact with Polycomb group proteins[243] and that AGO proteins are found in the nucleus[86,87] (FIG. 3). In parallel, dating back to 1990, antisense RNAs were shown to affect gene expression, again initially in plants[73] and later in animals[159,166,244–246]. Similar to small ncRNAs[247], some lncRNAs have been shown to control alternative splicing[248,249]. Other naturally occurring lncRNAs were shown to control epigenetic processes *in vivo*, notably in X chromosome dosage compensation[250–254] and parental imprinting in mammals[255–257], and vernalization in plants[258]. Subsequent studies showed that intergenic and antisense RNAs bind to Polycomb repressive complexes (PRCs)[194,259–261], to trithorax chromatin-activating complexes and activated forms of histones[207], and to DNA methyltransferases[201,262,263]. These observations were writ large in 2009 when it was shown that ~20% of ~3,300 lncRNAs examined were bound by PRC2 and that others were bound by different chromatin-modifying complexes. siRNA-mediated knockdown of PRC2-associated lncRNAs was found to result in gene expression changes, and the upregulated genes were enriched for those normally silenced by PRC2 (REF. 179). Polycomb group proteins were also discovered to bind to RNA with high affinity but low specificity[264], which is consistent with the idea that many RNAs interact with these proteins.

One of the notable lncRNAs to emerge — HOX transcript antisense RNA (*HOTAIR*) — is derived from the *HOXC* locus and regulates *HOXD* in *trans*[194]. It is involved in cancer metastasis[220] and, when inactivated, results in homeotic transformation *in vivo*[229]. lncRNAs have also been shown to act as scaffolds for the assembly of histone modification complexes[265], and the widespread alternative splicing of these RNAs suggests that the cargo and/or target specificity can be varied in a context-dependent and differentiation-specific manner.

Glossary

Antisense RNA
A single-stranded RNA that is complementary to an mRNA or a gene.

Encyclopedia of DNA Elements
(ENCODE). An international consortium involved in building a comprehensive list of functional elements in the human genome.

Heterogeneous nuclear RNA
(hnRNA). A type of RNA that is similar to mRNA or pre-mRNA but that is retained predominantly in the nucleus.

Introns
A term first coined by Gilbert to describe the RNA regions that are removed, by being spliced out, to produce mRNAs.

PIWI-interacting RNAs
(piRNAs). Small RNAs that are associated with the PIWI protein complex and that emanated from transposon-like elements

RNA CaptureSeq
A method that combines the ability to capture RNA (that is, to isolate and enrich for certain types of RNA) with deep sequencing technology to mine the human transcriptome.

RNA-directed DNA methylation
An epigenetic process whereby processed double-stranded small (21–24-nucleotide) RNAs guide the methylation of homologous DNA loci.

Small interfering RNAs
(siRNAs). Small interfering, double-stranded RNAs that can be used to suppress homology-containing transcripts in a transcriptional and post-transcriptional manner.

Splice site RNAs
(spliRNAs). Small RNAs that are derived from the 3′ ends of exons adjacent to splice sites and that are similar to transcription initiation RNAs (tiRNAs).

Transcriptional gene silencing
The regulation of a gene at the transcriptional level, in contrast to post-transcriptional gene silencing, in which silencing of gene expression occurs at the mRNA or translational level, after transcription has occurred.

Transcription initiation RNAs
(tiRNAs). Small RNAs associated with promoters with peak density at ~15–35 nucleotides downstream of transcription start sites.

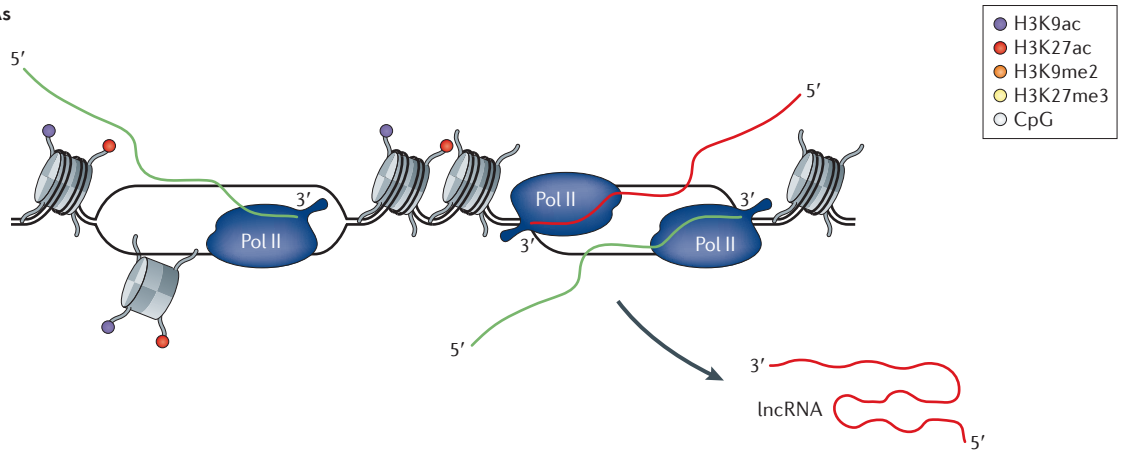
Transinduction
A genetic phenomenon whereby mRNA transcription induces transcription of nearby enhancers and intergenic non-coding RNAs.

Transposons
Mobile genetic elements with evolutionary links to retroviruses.

Transvection
A genetic phenomenon whereby non-coding regions can induce transcription of coding regions on other chromosomes.

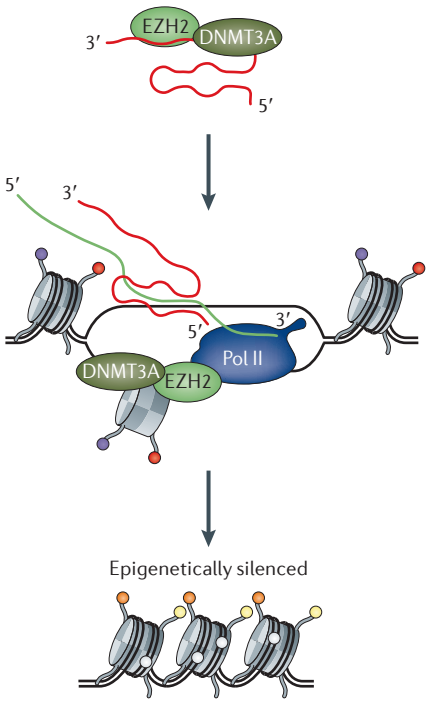
Untranslated regions
(UTRs). Sequences either side of a coding sequence on a strand of mRNA; these can be 5′ leader sequences or 3′ trailer sequences.

A Transcription of lncRNAs

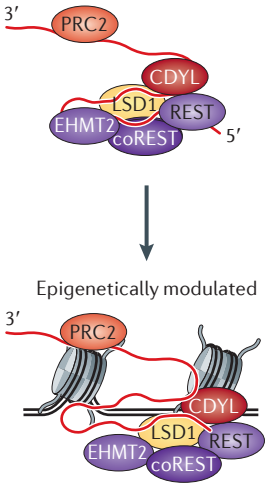


B Nuclear functional lncRNAs

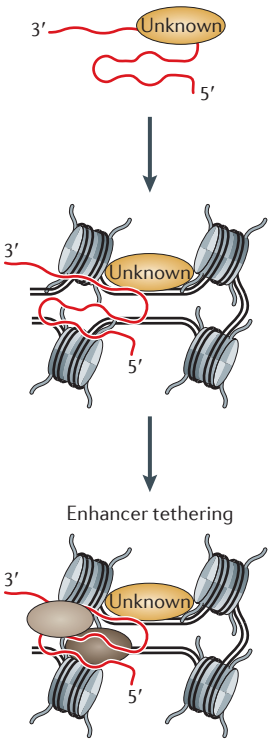
Ba



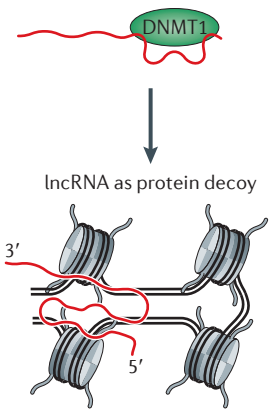
Bb



Bc

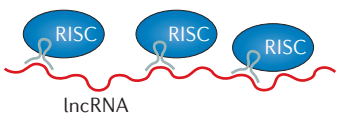


Bd

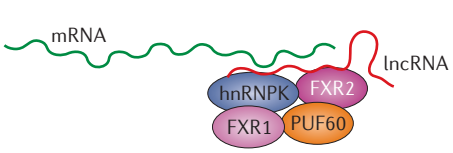


C Nuclear and cytoplasmic functional lncRNAs

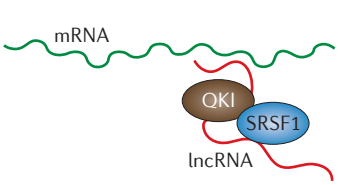
Ca lncRNA as miRNA decoy



Cb lncRNA as translational regulator



Cc lncRNA as splicing regulator



◀ Figure 4 | **Various roles for long non-coding RNAs in cellular regulation. A** | Long non-coding RNAs (lncRNAs) are expressed from many loci in the genome — sense and antisense, intronic, overlapping and intergenic with respect to nearby protein-coding loci — and function in both *cis* and *trans*. **B** | Nuclear functional lncRNAs can modulate gene expression both transcriptionally and epigenetically. Some lncRNAs interact with proteins to control the access of chromatin to cellular components and/or guide epigenetic regulatory complexes to target loci, which results in both transcriptional suppression[201] (part **Ba**) and activation or suppression (that is, bimodal control)[194] (part **Bb**). Proteins involved in chromatin modification — such as DNA (cytosine-5)-methyltransferase 3A (DNMT3A), enhancer of Zeste 2 (EZH2), euchromatic histone-lysine *N*-methyltransferase 2 (EHMT2; also known as G9a), chromodomain Y-like protein (CDYL), repressor element 1-silencing transcription factor (REST), co-repressor of REST (coREST), trithorax-activating complex MLL1 (REF. 207) (not shown) and Polycomb repressive complex 2 (PRC2) — have been associated with lncRNA-mediated epigenetic silencing[194,201,265]; the histone demethylase LSD1 (also known as KDM1A) has been associated with activation of silent loci. Enhancer functional lncRNAs tether distal enhancer elements with their promoters[344,345], presumably in concert with a protein component that has yet to be determined (shown as 'unknown') (part **Bc**). Decoy functional lncRNAs affect transcription by binding to proteins such as DNMT1 to sequester them from their sites of action, which leads to a loss of maintenance of DNA methylation and gene activation[263] (part **Bd**). **C** | Some lncRNAs can function in both nuclear and cytoplasmic compartments of the cell to affect gene expression and translation of mRNAs. Decoy functional lncRNA complexes affect microRNA (miRNA) targeting of mRNAs (part **Ca**). Some lncRNAs can interact with each other or with mRNAs to sequester small regulatory RNAs, such as miRNAs and therefore RNA-induced silencing complex (RISC), from protein-coding mRNAs[201,337,338]. Translational regulatory lncRNAs have been observed to recruit protein complexes that consist of heterogeneous nuclear ribonucloprotein K (hnRNPK), fragile X mental retardation syndrome-related protein 1 (FXR1), FXR2 and Poly(U)-binding splicing factor (PUF60) to homology-containing protein-coding mRNAs, where they bind to and sequester the mRNAs from the translational machinery[346] and regulate translation (part **Cb**). lncRNAs can also bind to homology-containing mRNAs and recruit proteins such as QKI and serine/arginine-rich splicing factor 1 (SRSF1), both of which modulate the splicing of the targeted mRNA[341] (part **Cc**). H3K9ac, histone H3 lysine 9 acetylation; me, methylation; Pol II, RNA polymerase II.

lncRNAs may also be involved in orchestrating the highly dynamic spatial structure of chromatin during differentiation and development[164,266], which would explain their often highly cell-specific expression patterns[200]. Developmental enhancers, as well as Polycomb- and trithorax-response elements, are transcribed in the cells in which they are active[203,267–272]. These elements may not only be scaffolds for the recruitment of epigenetic regulators[273] but also be the physical mediators of the complex phenomena of transvection and transinduction[234].

Moreover, many lncRNAs show the properties of enhancers[180]. These RNAs might guide the physical looping that occurs between enhancers, target promoters and exons with precise positioning of nucleosomes[274–278] to control transcription and alternative splicing[237,279,280]. Indeed, the emerging picture is of a chromatin and transcriptional landscape that is exquisitely and precisely controlled in four dimensions by a range of regulatory RNAs that assemble fairly generic (albeit often cell- or differentiation state-specific) enzyme complexes and isoforms to their sites of action in a context-dependent manner[238].

A substantial proportion of lncRNAs reside within, or are dynamically shuttled to, the cytoplasm, which indicates roles in other cellular processes (BOX 1), including the regulation of protein localization[281], mRNA translation[282] and mRNA stability[283].

RNA modification, evolution and inheritance. Regulatory RNAs may also be influenced by environmental signals and transmitted between cells and generations, which has important implications for understanding gene–environment interactions and evolution. There is evidence that plasticity has been superimposed on RNA-directed epigenetic networks by the expansion of RNA editing, especially during cognitive evolution[284,285], and by the use and mobility of retrotransposons[114,286–289], which is consistent with the insights of McClintock and of Britten and Davidson. The 'raw material' for evolution is gene duplication and transposition; the latter has the advantage of being able to mobilize functional cassettes in regulatory networks[290], which seems to be the main 'driver' of adaptive radiation[234,291]. Indeed, many lncRNAs may have originated from retrotransposons, and the evolution of mRNAs and lncRNAs may have been accelerated by retrotransposition of functional modules[292–296].

Moreover, apart from snoRNA-directed modifications, there are more than 100 other documented modifications of RNA[297,298], including cytosine and adenosine methylation that have known physiological and cognitive effects[299–302]. This indicates an additional layer of RNA informational code and epitranscriptomics — an exciting field that is just beginning to emerge[303,304].

There is evidence for systemic transmission of RNA[305,306] and RNA-mediated epigenetic inheritance in plants and animals[307–311]. There is also the intriguing possibility of RNA-directed DNA recoding, which may place RNA at the centre not only of gene regulation in the developmental ontogeny of higher organisms but also of both 'hard-wired' and 'soft-wired' somatic and germline evolution[312–314].

Conclusions and outlook

Our understanding of the previously hidden and unanticipated world of ncRNAs has greatly expanded in the past two decades. Indeed, in retrospect, it seems that we may have fundamentally misunderstood the nature of the genetic programming in complex organisms because of the assumption that most genetic information is transacted by proteins. This may be true to a large extent in simpler organisms but is turning out not to be the case in more complex organisms, the genomes of which seem to be progressively dominated by regulatory RNAs that orchestrate the epigenetic trajectories of differentiation and development.

The emerging picture is one of an extraordinarily complex transcriptional landscape in mammals and other multicellular organisms. Such a landscape is comprised of overlapping, intergenic and intronic, sense and antisense, small and large RNAs with interlaced exons[315,316], which have varying promoters, splicing patterns, polyadenylation sites and localization in different cells and developmental contexts (see below). As there seem to be few distinct boundaries to genes in humans, it might be better to change the focus of analysis to the transcript and to redefine genetic loci as 'fuzzy' transcription clusters[165,316,317] that are nonetheless semantically anchored or related to an enclosed or nearby protein-coding locus. However, this can only be stretched to a certain extent, and non-protein-coding loci raise problems for existing schema of human genome nomenclature.

Indeed, even the notion of a simple protein-coding sequence needs to be reassessed. It is becoming evident not only that mRNAs can have multiple functions[205] but also that

protein-coding sequences themselves can have other embedded functions, as suggested by constraints on synonymous codon usage[318,319], including regulatory functions as epigenetic modulators[203], tissue-specific enhancers[319,320] and transcription factor binding sites[321]. The possibility, if not likelihood, is that there is a very complex functional and evolutionary interplay between the protein-coding and regulatory functions of RNAs[200], and that some lncRNAs may have evolved, at least partly, from protein-coding genes — as in the case of X inactive specific transcript (*XIST*) — by duplication or pseudogenization and the subsequent emergence of paralogous regulatory and/or coding functions[201,322]. Conversely, new protein-coding capacity may also appear in lncRNAs[174].

The sheer number and diversity of RNAs juxtaposed with their extraordinarily complex molecular functions (FIG. 4) — for example, in regulating epigenetic processes, subcellular organelles, protein-coding and non-coding gene transcription, translation, RNA turnover, chromosomal organization and integrity, and genome defence — suggests that we have a long way to go to understand the structure and functions of what is surely a highly interconnected system. Tens of thousands (if not more) of individual non-coding RNAs exist, and their roles in cell and developmental biology, as well as in brain function, remain to be determined. Moreover, many (if not most) regulatory RNAs have yet to be identified, especially in complex organisms. These include new classes such as the circular RNAs and others that may function as miRNA 'sponges' (REFS 62,323–328), the identification of which will require targeted deep sequencing of small and large RNAs that are derived from different genomic locations in various cell types, using targeted techniques such as RNA CaptureSeq[164,171].

RNA is not a linear molecule but can fold into complex and allosterically responsive three-dimensional structures that can both recruit generic effector proteins and guide the resulting complexes in a sequence-specific manner to other RNAs and DNA through duplex or triplex formation. Important issues that remain include the identification of functional domains in RNA and their interacting partners, so that we can predict and explain RNA functional interactions in the same way that has already been done by recognition of well-characterized motifs and domains in proteins. One way to do this, which is already underway in many laboratories, is to combine immunoprecipitation of different types of RNA-binding proteins (for example, chromatin-modifying proteins, transcription factors and RNA transport proteins) with deep sequencing of the associated RNAs, followed by analysis of primary and predicted secondary structures, and ultimately by biochemical validation and characterization.

Determination of the structure of RNA species, RNA–RNA, RNA–DNA and RNP complexes will be a rapidly growing field that requires the development of new technologies, such as RNA footprinting using high-throughput sequencing[329] and *in vivo* studies using RNA-based genetic techniques, for example, CRISPR-mediated mutation[143]. Other objectives include determination of whether small RNA pathways are used in viral defence in humans[80]; the functions of tiRNAs, spliRNAs and snoRNA-derived small RNAs; the roles of piRNAs in retrotransposon dynamics and genome remodelling by retrotransposons in the brain[114]; the mechanisms and extent of RNA-mediated transgenerational epigenetic inheritance[330]; the locations of RNA-binding sites (that is, RNA–DNA duplexes and RNA–DNA:DNA triplexes) in the genome; the crosstalk between different types of regulatory RNAs; the logic and hierarchy of RNA- and protein-mediated regulation of gene expression; and finally, the extent, mechanisms and information content of RNA-mediated communication between cells both within[306] and between organisms (that is, 'social RNA')[331].

Indeed, it seems that RNA is the computational engine of cell biology, developmental biology, brain function and perhaps even evolution itself[313]. The complexity and interconnectedness of these systems should not be cause for concern but rather the motivation for exploring the vast unknown universe of RNA regulation, without which we will not understand biology.

Kevin V. Morris is at the School of Biotechnology and Biomedical Sciences, University of New South Wales, Sydney, NSW 2052, Australia; and Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, California 92037, USA.

John S. Mattick is at the Garvan Institute of Medical Research, 384 Victoria Street, Darlinghurst, NSW 2010, Australia; the School of Biotechnology and Biomedical Sciences, and St. Vincent's Clinical School, University of New South Wales, Sydney, NSW 2052, Australia.

Correspondence to K.V.M.
e-mail: kmorris@scripps.edu

doi:10.1038/nrg3722
Published online 29 April 2014

Box 1 | **Examples of specific long non-coding RNAs and their functions**

Long non-coding RNAs have a role in a wide range of biological processes in the cell, for example:

- Template RNAs guide chromosomal rearrangements in ciliates[332]
- Telomeric repeat-containing RNA (TERRA) is involved in telomere biology[333]
- 7S RNA is an essential component of the signal recognition particle, which is involved in protein export[334]
- 7SK is a highly expressed structured RNA that acts as a scaffold to assemble a multimeric protein complex containing SR splicing proteins and positive transcription elongation factor b (P-TEFb, which is a cyclin-dependent kinase required for transcriptional elongation by RNA polymerase II and other factors)[335]
- Nuclear paraspeckle assembly transcript 1 (*NEAT1*) is an essential component of paraspeckles, which are enigmatic subnuclear organelles that appear in mammalian differentiated cells but not stem cells[336,337]
- Metastasis-associated lung adenocarcinoma transcript 1 (*MALAT1*) is localized to the nucleus and regulates alternative splicing[338] and cell cycle progression[339]
- Myocardial infarction-associated transcript (*MIAT*; also known as gomafu) is expressed in an unknown subnuclear structure, possibly a specialized spliceosome, in a subset of neurons[340] and has recently been implicated in schizophrenia[341]

1. Gilbert, W. Origin of life: the RNA world. *Nature* **319**, 618 (1986).
2. Beadle, G. W. & Tatum, E. L. Genetic control of biochemical reactions in *Neurospora*. *Proc. Natl Acad. Sci. USA* **27**, 499–506 (1941).
3. Comfort, N. C. *The Tangled Field: Barbara McClintock's Search for the Patterns of Genetic Control* (Harvard Univ. Press, 2003).
4. Mattick, J. S. The genetic signatures of noncoding RNAs. *PLoS Genet.* **5**, e1000459 (2009).
5. Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).
6. Crick, F. H. On protein synthesis. *Symp. Soc. Exp. Biol.* **12**, 138–163 (1958).
7. Palade, G. E. A small particulate component of the cytoplasm. *J. Biophys. Biochem. Cytol.* **1**, 59–68 (1955).
8. Hoagland, M. B., Stephenson, M. L., Scott, J. F., Hecht, L. I. & Zamecnik, P. C. A soluble ribonucleic acid intermediate in protein synthesis. *J. Biol. Chem.* **231**, 241–257 (1958).
9. Brenner, S., Jacob, F. & Meselson, M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* **190**, 576–581 (1961).
10. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356 (1961).

11. Gilbert, W. & Muller-Hill, B. Isolation of the *lac* repressor. *Proc. Natl Acad. Sci. USA* **56**, 1891–1898 (1966).
12. Levine, M. & Tjian, R. Transcription regulation and animal diversity. *Nature* **424**, 147–151 (2003).
13. Mattick, J. S. & Gagen, M. J. Accelerating networks. *Science* **307**, 856–858 (2005).
14. Freedman, M. L. *et al.* Principles for the post-GWAS functional characterization of cancer risk loci. *Nature Genet.* **43**, 513–518 (2011).
15. Weinberg, R. A. & Penman, S. Small molecular weight monodisperse nuclear RNA. *J. Mol. Biol.* **38**, 289–304 (1968).
16. Dreyfuss, G., Philipson, L. & Mattaj, I. W. Ribonucleoprotein particles in cellular processes. *J. Cell Biol.* **106**, 1419–1425 (1988).
17. Butcher, S. E. & Brow, D. A. Towards understanding the catalytic core structure of the spliceosome. *Biochem. Soc. Trans.* **33**, 447–449 (2005).
18. Wang, Z. & Burge, C. B. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* **14**, 802–813 (2008).
19. Pessa, H. K. *et al.* Minor spliceosome components are predominantly localized in the nucleus. *Proc. Natl Acad. Sci. USA* **105**, 8655–8660 (2008).
20. Maxwell, E. S. & Fournier, M. J. The small nucleolar RNAs. *Annu. Rev. Biochem.* **64**, 897–934 (1995).
21. Henras, A. K., Dez, C. & Henry, Y. RNA structure and function in C/D and H/ACA s(no)RNPs. *Curr. Opin. Struct. Biol.* **14**, 335–343 (2004).
22. Meier, U. T. The many facets of H/ACA ribonucleoproteins. *Chromosoma* **114**, 1–14 (2005).
23. Cavaille, J., Seitz, H., Paulsen, M., Ferguson-Smith, A. C. & Bachellerie, J. P. 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 (2002).
24. Rogelj, B., Hartmann, C. E., Yeo, C. H., Hunt, S. P. & Giese, K. P. Contextual fear conditioning regulates the expression of brain-specific small nucleolar RNAs in hippocampus. *Eur. J. Neurosci.* **18**, 3089–3096 (2003).
25. Bachellerie, J. P., Cavaille, J. & Huttenhofer, A. The expanding snoRNA world. *Biochimie* **84**, 775–790 (2002).
26. Jady, B. E., Bertrand, E. & Kiss, T. Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal. *J. Cell Biol.* **164**, 647–652 (2004).
27. Warner, J. R., Soeiro, R., Birnboim, H. C., Girard, M. & Darnell, J. E. Rapidly labeled HeLa cell nuclear RNA. I. Identification by zone sedimentation of a heterogeneous fraction separate from ribosomal precursor RNA. *J. Mol. Biol.* **19**, 349–361 (1966).
28. Britten, R. J. & Davidson, E. H. Gene regulation for higher cells: a theory. *Science* **165**, 349–357 (1969).
29. Britten, R. J. & Davidson, E. H. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q. Rev. Biol.* **66**, 111–138 (1971).
30. Davidson, E. H., Klein, W. H. & Britten, R. J. Sequence organization in animal DNA and a speculation on hnRNA as a coordinate regulatory transcript. *Dev. Biol.* **55**, 69–84 (1977).
31. Howard, M. L. & Davidson, E. H. *Cis*-regulatory control circuits in development. *Dev. Biol.* **271**, 109–118 (2004).
32. Davidson, E. H. *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution* (Academic Press, 2006).
33. Britten, R. Transposable elements have contributed to thousands of human proteins. *Proc. Natl Acad. Sci. USA* **103**, 1798–1803 (2006).
34. Berget, S. M., Moore, C. & Sharp, P. A. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proc. Natl Acad. Sci. USA* **74**, 3171–3175 (1977).
35. Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* **12**, 1–8 (1977).
36. Williamson, B. DNA insertions and gene structure. *Nature* **270**, 295–297 (1977).
37. Gilbert, W., Marchionni, M. & McKnight, G. On the antiquity of introns. *Cell* **46**, 151–154 (1986).
38. Doolittle, W. F. & Sapienza, C. Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**, 601–603 (1980).
39. Orgel, L. E. & Crick, F. H. Selfish DNA: the ultimate parasite. *Nature* **284**, 604–607 (1980).
40. Kruger, K. *et al.* Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* **31**, 147–157 (1982).
41. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857 (1983).
42. Fica, S. M. *et al.* RNA catalyses nuclear pre-mRNA splicing. *Nature* **503**, 229–234 (2013).
43. Steitz, T. A. & Moore, P. B. RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem. Sci.* **28**, 411–418 (2003).
44. Webb, C. H., Riccitelli, N. J., Ruminski, D. J. & Luptak, A. Widespread occurrence of self-cleaving ribozymes. *Science* **326**, 953 (2009).
45. de la Pena, M. & Garcia-Robles, I. Intronic hammerhead ribozymes are ultraconserved in the human genome. *EMBO Rep.* **11**, 711–716 (2010).
46. Lee, R. C., Feinbaum, R. L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843–854 (1993).
47. Reinhart, B. J. *et al.* The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**, 901–906 (2000).
48. Pasquinelli, A. E. *et al.* Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408**, 86–89 (2000).
49. Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B. & Cohen, S. M. *bantam* encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in *Drosophila*. *Cell* **113**, 25–36 (2003).
50. Johnston, R. J. & Hobert, O. A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature* **426**, 845–849 (2003).
51. Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**, 858–862 (2001).
52. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. & Tuschl, T. Identification of novel genes coding for small expressed RNAs. *Science* **294**, 853–858 (2001).
53. Lee, R. C. & Ambros, V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294**, 862–864 (2001).
54. Williams, T. M. *et al.* The regulation and evolution of a genetic switch controlling sexually dimorphic traits in *Drosophila*. *Cell* **134**, 610–623 (2008).
55. Kaya, K. D., Karakulah, G., Yakicier, C. M. & Acar, A. C. & Konu, O. mESAdb: microRNA expression and sequence analysis database. *Nucleic Acids Res.* **39**, D170–D180 (2011).
56. Berezikov, E. *et al.* Diversity of microRNAs in human and chimpanzee brain. *Nature Genet.* **38**, 1375–1377 (2006).
57. Heimberg, A. M., Sempere, L. F., Moy, V. N., Donoghue, P. C. & Peterson, K. J. MicroRNAs and the advent of vertebrate morphological complexity. *Proc. Natl Acad. Sci. USA* **105**, 2946–2950 (2008).
58. John, B. *et al.* Human microRNA targets. *PLoS Biol.* **2**, e363 (2004).
59. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
60. Fabian, M. R., Sonenberg, N. & Filipowicz, W. Regulation of mRNA translation and stability by microRNAs. *Annu. Rev. Biochem.* **79**, 351–379 (2010).
61. Schnall-Levin, M. *et al.* Unusually effective microRNA targeting within repeat-rich coding regions of mammalian mRNAs. *Genome Res.* **21**, 1395–1403 (2011).
62. Hansen, T. B. *et al.* miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO J.* **30**, 4414–4422 (2011).
63. Leonardo, T. R., Schultheisz, H. L., Loring, J. F. & Laurent, L. C. The functions of microRNAs in pluripotency and reprogramming. *Nature Cell Biol.* **14**, 1114–1121 (2012).
64. Bracken, C. P., Gregory, P. A., Khew-Goodall, Y. & Goodall, G. J. The role of microRNAs in metastasis and epithelial–mesenchymal transition. *Cell. Mol. Life Sci.* **66**, 1682–1699 (2009).
65. Rakoczy, J. *et al. MicroRNAs-140-5p/140-3p* modulate Leydig cell numbers in the developing mouse testis. *Biol. Reprod.* **88**, 143 (2013).
66. Fernandez-Valverde, S. L., Taft, R. J. & Mattick, J. S. MicroRNAs in β-cell biology, insulin resistance, diabetes and its complications. *Diabetes* **60**, 1825–1831 (2011).
67. Bredy, T. W., Lin, Q., Wei, W., Baker-Andresen, D. & Mattick, J. S. MicroRNA regulation of neural plasticity and memory. *Neurobiol. Learn. Mem.* **96**, 89–94 (2011).
68. Park, C. Y., Choi, Y. S. & McManus, M. T. Analysis of microRNA knockouts in mice. *Hum. Mol. Genet.* **19**, R169–R175 (2010).
69. Waterhouse, P. M., Graham, M. W. & Wang, M. B. Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc. Natl Acad. Sci. USA* **95**, 13959–13964 (1998).
70. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
71. Napoli, C., Lemieux, C. & Jorgensen, R. Introduction of a chimeric chalcone synthase gene into Petunia results in reversible co-suppression of homologous genes in *trans*. *Plant Cell* **2**, 279–289 (1990).
72. Matzke, M. A., Primig, M., Trnovsky, J. & Matzke, A. J. M. Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. *EMBO J.* **8**, 643–649 (1989).
73. van der Krol, A. R., Mur, L. A., de Lange, P., Mol, J. N. & Stuitje, A. R. Inhibition of flower pigmentation by antisense *CHS* genes: promoter and minimal sequence requirements for the antisense effect. *Plant Mol. Biol.* **14**, 457–466 (1990).
74. Wassenegger, M., Heimes, S., Riedel, L. & Sanger, H. L. RNA-directed *de novo* methylation of genomic sequences in plants. *Cell* **76**, 567–576 (1994).
75. Basyuk, E., Suavet, F., Doglio, A., Bordonne, R. & Bertrand, E. Human *let-7* stem–loop precursors harbor features of RNase III cleavage products. *Nucleic Acids Res.* **31**, 6593–6597 (2003).
76. Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).
77. Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).
78. Doi, N. *et al.* Short-interfering-RNA-mediated gene silencing in mammalian cells requires Dicer and eIF2C translation initiation factors. *Curr. Biol.* **13**, 41–46 (2003).
79. Peters, L. & Meister, G. Argonaute proteins: mediators of RNA silencing. *Mol. Cell* **26**, 611–623 (2007).
80. Maillard, P. V. *et al.* Antiviral RNA interference in mammalian cells. *Science* **342**, 235–238 (2013).
81. Zeng, Y., Yi, R. & Cullen, B. R. MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc. Natl Acad. Sci. USA* **100**, 9779–9784 (2003).
82. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835–840 (2010).
83. Ramachandran, P. V. & Ignacimuthu, S. RNA interference — a silent but an efficient therapeutic tool. *Appl. Biochem. Biotechnol.* **169**, 1774–1789 (2013).
84. Ahlenstiel, C. L. *et al.* Direct evidence of nuclear Argonaute distribution during transcriptional silencing links the actin cytoskeleton to nuclear RNAi machinery in human cells. *Nucleic Acids Res.* **40**, 1579–1595 (2012).
85. Ameyar-Zazoua, M. *et al.* Argonaute proteins couple chromatin silencing to alternative splicing. *Nature Struct. Mol. Biol.* **19**, 998–1004 (2012).
86. Rudel, S., Flatley, A., Weinmann, L., Kremmer, E. & Meister, G. A multifunctional human Argonaute2-specific monoclonal antibody. *RNA* **14**, 1244–1253 (2008).
87. Kim, D. H., Villeneuve, L. M., Morris, K. V. & Rossi, J. J. Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nature Struct. Mol. Biol.* **13**, 793–797 (2006).
88. Morris, K. V., Chan, S. W., Jacobsen, S. E. & Looney, D. J. Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* **305**, 1289–1292 (2004).
89. Fernandez-Valverde, S. L., Taft, R. J. & Mattick, J. S. Dynamic isomiR regulation in *Drosophila* development. *RNA* **16**, 1881–1888 (2010).
90. Didiano, D. & Hobert, O. Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nature Struct. Mol. Biol.* **13**, 849–851 (2006).
91. Shin, C. *et al.* Expanding the microRNA targeting code: functional sites with centered pairing. *Mol. Cell* **38**, 789–802 (2010).

92. Kim, D. H., Saetrom, P., Snove, O. Jr & Rossi, J. J. MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc. Natl Acad. Sci. USA* **105**, 16230–16235 (2008).
93. Blow, M. J. *et al.* RNA editing of human microRNAs. *Genome Biol.* **7**, R27 (2006).
94. Hundley, H. A. & Bass, B. L. ADAR editing in double-stranded UTRs and other noncoding RNA sequences. *Trends Biochem. Sci.* **35**, 377–383 (2010).
95. Kawahara, Y., Zinshteyn, B., Chendrimada, T. P., Shiekhattar, R. & Nishikura, K. RNA editing of the *microRNA-151* precursor blocks cleavage by the Dicer–TRBP complex. *EMBO Rep.* **8**, 763–769 (2007).
96. Ota, T. *et al.* Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nature Genet.* **36**, 40–45 (2004).
97. Lin, H. & Spradling, A. C. A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the *Drosophila* ovary. *Development* **124**, 2463–2476 (1997).
98. Cox, D. N. *et al.* A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes Dev.* **12**, 3715–3727 (1998).
99. Kuramochi-Miyagawa, S. *et al. Mili*, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* **131**, 839–849 (2004).
100. Kim, J. K. *et al.* Functional genomic analysis of RNA interference in *C. elegans*. *Science* **308**, 1164–1167 (2005).
101. Pal-Bhadra, M., Bhadra, U. & Birchler, J. A. RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in *Drosophila*. *Mol. Cell* **9**, 315–327 (2002).
102. Vagin, V. V. *et al.* A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**, 320–324 (2006).
103. Aravin, A. *et al.* A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**, 203–207 (2006).
104. Girard, A., Sachidanandam, R., Hannon, G. J. & Carmell, M. A. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199–202 (2006).
105. Grivna, S. T., Beyret, E., Wang, Z. & Lin, H. A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev.* **20**, 1709–1714 (2006).
106. Watanabe, T. *et al.* Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev.* **20**, 1732–1743 (2006).
107. Aravin, A. A., Hannon, G. J. & Brennecke, J. The Piwi–piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **318**, 761–764 (2007).
108. Brennecke, J. *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128**, 1089–1103 (2007).
109. Brennecke, J. *et al.* An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* **322**, 1387–1392 (2008).
110. Kuramochi-Miyagawa, S. *et al.* DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev.* **22**, 908–917 (2008).
111. Cox, D. N., Chao, A. & Lin, H. *piwi* encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development* **127**, 503–514 (2000).
112. Grimaud, C. *et al.* RNAi components are required for nuclear clustering of Polycomb group response elements. *Cell* **124**, 957–971 (2006).
113. Rajasethupathy, P. *et al.* A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. *Cell* **149**, 693–707 (2012).
114. Baillie, J. K. *et al.* Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* **479**, 534–537 (2011).
115. Ross, R. J., Weiner, M. M. & Lin, H. PIWI proteins and PIWI-interacting RNAs in the soma. *Nature* **505**, 353–359 (2014).
116. Ruby, J. G. *et al.* Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* **127**, 1193–1207 (2006).
117. Taft, R. J. *et al.* Small RNAs derived from snoRNAs. *RNA* **15**, 1233–1240 (2009).
118. Ender, C. *et al.* A human snoRNA with microRNA-like functions. *Mol. Cell* **32**, 519–528 (2008).
119. Kawaji, H. *et al.* Hidden layers of human small RNAs. *BMC Genomics* **9**, 157 (2008).
120. Haussecker, D. *et al.* Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA* **16**, 673–695 (2010).
121. Taft, R. J. *et al.* Tiny RNAs associated with transcription start sites in animals. *Nature Genet.* **41**, 572–578 (2009).
122. Taft, R. J. *et al.* Nuclear-localized tiny RNAs are associated with transcription initiation and splice sites in metazoans. *Nature Struct. Mol. Biol.* **17**, 1030–1034 (2010).
123. Taft, R. J., Kaplan, C. D., Simons, C. & Mattick, J. S. Evolution, biogenesis and function of promoter-associated RNAs. *Cell Cycle* **8**, 2332–2338 (2009).
124. Taft, R. J., Hawkins, P. G., Mattick, J. S. & Morris, K. V. The relationship between transcription initiation RNAs and CCCTC-binding factor (CTCF) localization. *Epigenetics Chromatin* **4**, 13 (2011).
125. Kapranov, P. *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484–1488 (2007).
126. Seila, A. C. *et al.* Divergent transcription from active promoters. *Science* **322**, 1849–1851 (2008).
127. Preker, P. *et al.* RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**, 1851–1854 (2008).
128. Han, J., Kim, D. & Morris, K. V. Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells. *Proc. Natl Acad. Sci. USA* **104**, 12422–12427 (2007).
129. Wassarman, K. M., Zhang, A. & Storz, G. Small RNAs in *Escherichia coli*. *Trends Microbiol.* **7**, 37–45 (1999).
130. Gottesman, S. Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends Genet.* **21**, 399–404 (2005).
131. Tucker, B. J. & Breaker, R. R. Riboswitches as versatile gene control elements. *Curr. Opin. Struct. Biol.* **15**, 342–348 (2005).
132. Winkler, W. C. Riboswitches and the role of noncoding RNAs in bacterial metabolic control. *Curr. Opin. Chem. Biol.* **9**, 594–602 (2005).
133. Mojica, F. J., Diez-Villasenor, C., Soria, E. & Juez, G. Biological significance of a family of regularly spaced repeats in the genomes of archaea, bacteria and mitochondria. *Mol. Microbiol.* **36**, 244–246 (2000).
134. Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**, 174–182 (2005).
135. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551–2561 (2005).
136. Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**, 653–663 (2005).
137. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
138. Brouns, S. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
139. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843–1845 (2008).
140. Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).
141. Hale, C. R. *et al.* RNA-guided RNA cleavage by a CRISPR RNA–Cas protein complex. *Cell* **139**, 945–956 (2009).
142. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
143. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
144. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
145. Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910–918 (2013).
146. Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
147. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
148. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR–Cas9-based transcription factors. *Nature Methods* **10**, 973–976 (2013).
149. Cheng, A. W. *et al.* Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.* **23**, 1163–1171 (2013).
150. Hu, J. *et al.* Direct activation of human and mouse *Oct4* genes using engineered TALE and Cas9 transcription factors. *Nucleic Acids Res.* http://dx.doi.org/10.1093/nar/gku109 (2014).
151. Seed, K. D., Lazinski, D. W., Calderwood, S. B. & Camilli, A. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494**, 489–491 (2013).
152. Mattick, J. S. Introns: evolution and function. *Curr. Opin. Genet. Dev.* **4**, 823–831 (1994).
153. Kapranov, P. *et al.* Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296**, 916–919 (2002).
154. Okazaki, Y. *et al.* Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**, 563–573 (2002).
155. Rinn, J. L. *et al.* The transcriptional activity of human chromosome 22. *Genes Dev.* **17**, 529–540 (2003).
156. Cheng, J. *et al.* Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308**, 1149–1154 (2005).
157. Kapranov, P. *et al.* Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res.* **15**, 987–997 (2005).
158. Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–1563 (2005).
159. Katayama, S. *et al.* Antisense transcription in the mammalian transcriptome. *Science* **309**, 1564–1566 (2005).
160. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
161. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
162. Rosenbloom, K. R. *et al.* ENCODE whole-genome data in the UCSC Genome Browser: update 2012. *Nucleic Acids Res.* **40**, D912–D917 (2012).
163. Taft, R. J., Pheasant, M. & Mattick, J. S. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* **29**, 288–299 (2007).
164. Mercer, T. R. *et al.* Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nature Biotech.* **30**, 99–104 (2012).
165. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).
166. Hawkins, P. G. & Morris, K. V. Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. *Transcription* **1**, 165–175 (2010).
167. Sheik Mohamed, J., Gaughwin, P. M., Lim, B., Robson, P. & Lipovich, L. Conserved long noncoding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells. *RNA* **16**, 324–337 (2010).
168. Venters, B. J. & Pugh, B. F. Genomic organization of human transcription initiation complexes. *Nature* **502**, 53–58 (2013).
169. Carninci, P. *et al.* Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genet.* **38**, 626–635 (2006).
170. Huang, R. *et al.* An RNA-seq strategy to detect the complete coding and non-coding transcriptome including full-length imprinted macro ncRNAs. *PLoS ONE* **6**, e27288 (2011).
171. Roberts, A. & Pachter, L. RNA-seq and find: entering the RNA deep field. *Genome Med.* **3**, 74 (2011).
172. Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nature Rev. Genet.* **10**, 155–159 (2009).
173. Frith, M. C. *et al.* The abundance of short proteins in the mammalian proteome. *PLoS Genet.* **2**, e52 (2006).
174. Gascoigne, D. K. *et al.* Pinstripe: a suite of programs for integrating transcriptomic and proteomic datasets identifies novel proteins and improves differentiation of protein-coding and non-coding genes. *Bioinformatics* **28**, 3042–3050 (2012).
175. Banfai, B. *et al.* Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res.* **22**, 1646–1657 (2012).

176. Dinger, M. E. *et al.* NRED: a database of long noncoding RNA expression. *Nucleic Acids Res.* **37**, D122–D126 (2009).
177. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).
178. Wahlestedt, C. Natural antisense and noncoding RNA transcripts as potential drug targets. *Drug Discov. Today* **11**, 503–508 (2006).
179. Khalil, A. M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl Acad. Sci. USA* **106**, 11667–11672 (2009).
180. Ørom, U. A. *et al.* Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**, 46–58 (2010).
181. Sauvageau, M. *et al.* Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *Elife* **2**, e01749 (2013).
182. Furuno, M. *et al.* Clusters of internally primed transcripts reveal novel long noncoding RNAs. *PLoS Genet.* **2**, e37 (2006).
183. van Bakel, H., Nislow, C., Blencowe, B. J. & Hughes, T. R. Most "dark matter" transcripts are associated with known genes. *PLoS Biol.* **8**, e1000371 (2010).
184. Clark, M. B. *et al.* The reality of pervasive transcription. *PLoS Biol.* **9**, e1000625 (2011).
185. Calin, G. A. *et al.* Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell* **12**, 215–229 (2007).
186. Tay, S. K., Blythe, J. & Lipovich, L. Global discovery of primate-specific genes in the human genome. *Proc. Natl Acad. Sci. USA* **106**, 12019–12024 (2009).
187. Lipovich, L. *et al.* Developmental changes in the transcriptome of human cerebral cortex tissue: long noncoding RNA transcripts. *Cereb. Cortex* http://dx.doi.org/10.1093/cercor/bhs414 (2013).
188. Necsulea, A. *et al.* The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**, 635–640 (2014).
189. Pheasant, M. & Mattick, J. S. Raising the estimate of functional human sequences. *Genome Res.* **17**, 1245–1253 (2007).
190. Pang, K. C., Frith, M. C. & Mattick, J. S. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet.* **22**, 1–5 (2006).
191. Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H. & Bartel, D. P. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* **147**, 1537–1550 (2011).
192. Smith, M. A., Gesell, T., Stadler, P. F. & Mattick, J. S. Widespread purifying selection on RNA structure in mammals. *Nucleic Acids Res.* **41**, 8220–8236 (2013).
193. Johnsson, P., Lipovich, L., Grander, D. & Morris, K. V. Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim. Biophys. Acta* **1840**, 1063–1071 (2014).
194. Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human *HOX* loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
195. Cawley, S. *et al.* Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116**, 499–509 (2004).
196. Clark, M. B. *et al.* Genome-wide analysis of long noncoding RNA stability. *Genome Res.* **22**, 885–898 (2012).
197. Ravasi, T. *et al.* Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res.* **16**, 11–19 (2006).
198. Mercer, T. R., Dinger, M. E., Sunkin, S. M., Mehler, M. F. & Mattick, J. S. Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl Acad. Sci. USA* **105**, 716–721 (2008).
199. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927 (2011).
200. Mattick, J. S., Taft, R. J. & Faulkner, G. J. A global view of genomic information — moving beyond the gene and the master regulator. *Trends Genet.* **26**, 21–28 (2010).
201. Johnsson, P. *et al.* A pseudogene long-noncoding-RNA network regulates *PTEN* transcription and translation in human cells. *Nature Struct. Mol. Biol.* **20**, 440–446 (2013).
202. Chooniedass-Kothari, S. *et al.* The steroid receptor RNA activator is the first functional RNA encoding a protein. *FEBS Lett.* **566**, 43–47 (2004).
203. Ashe, H. L., Monks, J., Wijgerde, M., Fraser, P. & Proudfoot, N. J. Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev.* **11**, 2494–2509 (1997).
204. Mercer, T. R. *et al.* Expression of distinct RNAs from 3′ untranslated regions. *Nucleic Acids Res.* **39**, 2393–2403 (2011).
205. Dinger, M. E., Gascoigne, D. K. & Mattick, J. S. The evolution of RNAs with multiple functions. *Biochimie* **93**, 2013–2018 (2011).
206. Mercer, T. R. *et al.* Regulated post-transcriptional RNA cleavage diversifies the eukaryotic transcriptome. *Genome Res.* **20**, 1639–1650 (2010).
207. Dinger, M. E. *et al.* Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* **18**, 1433–1445 (2008).
208. Sunwoo, H. *et al.* MEN ε/β nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res.* **19**, 347–359 (2009).
209. Pang, K. C. *et al.* Genome-wide identification of long noncoding RNAs in $CD8^+$ T cells. *J. Immunol.* **182**, 7738–7748 (2009).
210. Askarian-Amiri, M. E. *et al.* SNORD-host RNA *Zfas1* is a regulator of mammary development and a potential marker for breast cancer. *RNA* **17**, 878–891 (2011).
211. Hu, W., Yuan, B., Flygare, J. & Lodish, H. F. Long noncoding RNA-mediated anti-apoptotic activity in murine erythroid terminal differentiation. *Genes Dev.* **25**, 2573–2578 (2011).
212. Mercer, T. R. *et al.* Long noncoding RNAs in neuronal–glial fate specification and oligodendrocyte lineage maturation. *BMC Neurosci.* **11**, 14 (2010).
213. Johnson, R. *et al.* Regulation of neural macroRNAs by the transcriptional repressor REST. *RNA* **15**, 85–96 (2009).
214. Ng, S.-Y., Johnson, R. & Stanton, L. W. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J.* **31**, 522–533 (2012).
215. Takeda, K. *et al.* Identification of a novel bone morphogenetic protein-responsive gene that may function as a noncoding RNA. *J. Biol. Chem.* **273**, 17079–17085 (1998).
216. Bussemakers, M. J. *et al. DD3*: a new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Res.* **59**, 5975–5979 (1999).
217. Pasmant, E. *et al.* Characterization of a germ-line deletion, including the entire *INK4/ARF* locus, in a melanoma-neural system tumor family: identification of *ANRIL*, an antisense noncoding RNA whose expression coclusters with *ARF*. *Cancer Res.* **67**, 3963–3969 (2007).
218. Wang, F., Li, X., Xie, X., Zhao, L. & Chen, W. *UCA1*, a non-protein-coding RNA up-regulated in bladder carcinoma and embryo, influencing cell growth and promoting invasion. *FEBS Lett.* **582**, 1919–1927 (2008).
219. Mourtada-Maarabouni, M., Pickard, M. R., Hedge, V. L., Farzaneh, F. & Williams, G. T. *GAS5*, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. *Oncogene* **28**, 195–208 (2009).
220. Gupta, R. A. *et al.* Long non-coding RNA *HOTAIR* reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076 (2010).
221. Khaitan, D. *et al.* The melanoma-upregulated long noncoding RNA *SPRY4-IT1* modulates apoptosis and invasion. *Cancer Res.* **71**, 3852–3862 (2011).
222. Kerin, T. *et al.* A noncoding RNA antisense to moesin at 5p14.1 in autism. *Sci. Transl Med.* **4**, 128ra40 (2012).
223. Amaral, P. P. & Mattick, J. S. Noncoding RNA in development. *Mamm. Genome* **19**, 454–492 (2008).
224. Mourtada-Maarabouni, M., Hedge, V. L., Kirkham, L., Farzaneh, F. & Williams, G. T. Growth arrest in human T-cells is controlled by the non-coding RNA growth-arrest-specific transcript 5 (*GAS5*). *J. Cell Sci.* **121**, 939–946 (2008).
225. Young, T. L., Matsuda, T. & Cepko, C. L. The noncoding RNA taurine upregulated gene 1 is required for differentiation of the murine retina. *Curr. Biol.* **15**, 501–512 (2005).
226. Ginger, M. R. *et al.* A noncoding RNA is a potential marker of cell fate during mammary gland development. *Proc. Natl Acad. Sci. USA* **103**, 5781–5786 (2006).
227. Kretz, M. *et al.* Control of somatic tissue differentiation by the long non-coding RNA *TINCR*. *Nature* **493**, 231–235 (2013).
228. Gutschner, T. *et al.* The noncoding RNA *MALAT1* is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res.* **73**, 1180–1189 (2013).
229. Li, L. *et al.* Targeted disruption of *Hotair* leads to homeotic transformation and gene derepression. *Cell Rep.* **5**, 3–12 (2013).
230. Nakagawa, S., Naganuma, T., Shioi, G. & Hirose, T. Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice. *J. Cell Biol.* **193**, 31–39 (2011).
231. Ahituv, N. *et al.* Deletion of ultraconserved elements yields viable mice. *PLoS Biol.* **5**, e234 (2007).
232. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).
233. Lewejohann, L. *et al.* Role of a neuronal small non-messenger RNA: behavioural alterations in *BC1* RNA-deleted mice. *Behav. Brain Res.* **154**, 273–289 (2004).
234. Mattick, J. S. & Gagen, M. J. 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 (2001).
235. Mattick, J. S., Amaral, P. P., Dinger, M. E., Mercer, T. R. & Mehler, M. F. RNA regulation of epigenetic processes. *Bioessays* **31**, 51–59 (2009).
236. Koziol, M. J. & Rinn, J. L. RNA traffic control of chromatin complexes. *Curr. Opin. Genet. Dev.* **20**, 142–148 (2010).
237. Mercer, T. R. & Mattick, J. S. Structure and function of long noncoding RNAs in epigenetic regulation. *Nature Struct. Mol. Biol.* **20**, 300–307 (2013).
238. Mercer, T. R. & Mattick, J. S. Understanding the regulatory and transcriptional complexity of the genome through structure. *Genome Res.* **23**, 1081–1088 (2013).
239. Wassenegger, M. RNA-directed DNA methylation. *Plant Mol. Biol.* **43**, 203–220 (2000).
240. Hall, I. M. *et al.* Establishment and maintenance of a heterochromatin domain. *Science* **297**, 2232–2237 (2002).
241. Volpe, T. A. *et al.* Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833–1837 (2002).
242. Verdel, A. *et al.* RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**, 672–676 (2004).
243. Kanhere, A. *et al.* Short RNAs are transcribed from repressed Polycomb target genes and interact with Polycomb repressive complex-2. *Mol. Cell* **38**, 675–688 (2010).
244. Imamura, T. *et al.* Non-coding RNA directed DNA demethylation of *Sphk1* CpG island. *Biochem. Biophys. Res. Commun.* **322**, 593–600 (2004).
245. Morris, K. V., Santoso, S., Turner, A. M., Pastori, C. & Hawkins, P. G. Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. *PLoS Genet.* **4**, e1000258 (2008).
246. Yu, W. *et al.* Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* **451**, 202–206 (2008).
247. Allo, M. *et al.* Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nature Struct. Mol. Biol.* **16**, 717–724 (2009).
248. Beltran, M. *et al.* A natural antisense transcript regulates *Zeb2/Sip1* gene expression during *Snail1*-induced epithelial–mesenchymal transition. *Genes Dev.* **22**, 756–769 (2008).
249. Morrissy, A. S., Griffith, M. & Marra, M. A. Extensive relationship between antisense transcription and alternative splicing in the human genome. *Genome Res.* **21**, 1203–1212 (2011).
250. Brockdorff, N. *et al.* The product of the mouse *Xist* gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* **71**, 515–526 (1992).
251. Brown, C. J. *et al.* The human *XIST* gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71**, 527–542 (1992).
252. Meller, V. H., Wu, K. H., Roman, G., Kuroda, M. I. & Davis, R. L. *roX1* RNA paints the X chromosome of male *Drosophila* and is regulated by the dosage compensation system. *Cell* **88**, 445–457 (1997).

253. Lee, J. T., Davidow, L. S. & Warshawsky, D. *Tsix*, a gene antisense to *Xist* at the X-inactivation centre. *Nature Genet.* **21**, 400–404 (1999).
254. Sado, T., Wang, Z., Sasaki, H. & Li, E. Regulation of imprinted X-chromosome inactivation in mice by *Tsix*. *Development* **128**, 1275–1286 (2001).
255. Ripoche, M. A., Kress, C., Poirier, F. & Dandolo, L. Deletion of the *H19* transcription unit reveals the existence of a putative imprinting control element. *Genes Dev.* **11**, 1596–1604 (1997).
256. Sleutels, F., Zwart, R. & Barlow, D. P. The non-coding *Air* RNA is required for silencing autosomal imprinted genes. *Nature* **415**, 810–813 (2002).
257. Thakur, N. *et al.* An antisense RNA regulates the bidirectional silencing property of the *Kcnq1* imprinting control region. *Mol. Cell. Biol.* **24**, 7855–7862 (2004).
258. Swiezewski, S., Liu, F., Magusin, A. & Dean, C. Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. *Nature* **462**, 799–802 (2009).
259. Nagano, T. *et al.* The *Air* noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**, 1717–1720 (2008).
260. Mohammad, F. *et al. Kcnq1ot1*/*Lit1* noncoding RNA mediates transcriptional silencing by targeting to the perinucleolar region. *Mol. Cell. Biol.* **28**, 3713–3728 (2008).
261. Kotake, Y. *et al.* Long non-coding RNA *ANRIL* is required for the PRC2 recruitment to and silencing of *p15*(*INK4B*) tumor suppressor gene. *Oncogene* **30**, 1956–1962 (2011).
262. Mohammad, F., Mondal, T., Guseva, N., Pandey, G. K. & Kanduri, C. *Kcnq1ot1* noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. *Development* **137**, 2493–2499 (2010).
263. Di Ruscio, A. *et al.* DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* **503**, 371–376 (2013).
264. Davidovich, C., Zheng, L., Goodrich, K. J. & Cech, T. R. Promiscuous RNA binding by Polycomb repressive complex 2. *Nature Struct. Mol. Biol.* **20**, 1250–1257 (2013).
265. Tsai, M. C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693 (2010).
266. Zhang, H. *et al.* Long noncoding RNA-mediated intrachromosomal interactions promote imprinting at the *Kcnq1* locus. *J. Cell Biol.* **204**, 61–75 (2014).
267. Sanchez-Herrero, E. & Akam, M. Spatially ordered transcription of regulatory DNA in the bithorax complex of *Drosophila*. *Development* **107**, 321–329 (1989).
268. Bae, E., Calhoun, V. C., Levine, M., Lewis, E. B. & Drewell, R. A. 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 (2002).
269. Jones, E. A. & Flavell, R. A. Distal enhancer elements transcribe intergenic RNA in the IL-10 family gene cluster. *J. Immunol.* **175**, 7437–7446 (2005).
270. Petruk, S. *et al.* Transcription of *bxd* noncoding RNAs promoted by trithorax represses *Ubx* in *cis* by transcriptional interference. *Cell* **127**, 1209–1221 (2006).
271. Feng, J. *et al.* The *Evf-2* noncoding RNA is transcribed from the *Dlx-5/6* ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes Dev.* **20**, 1470–1484 (2006).
272. Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
273. Sanchez-Elsner, T., Gou, D., Kremmer, E. & Sauer, F. Noncoding RNAs of trithorax response elements recruit *Drosophila* Ash1 to Ultrabithorax. *Science* **311**, 1118–1123 (2006).
274. Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C. & Komorowski, J. Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res.* **19**, 1732–1741 (2009).
275. Nahkuri, S., Taft, R. J. & Mattick, J. S. Nucleosomes are preferentially positioned at exons in somatic and sperm cells. *Cell Cycle* **8**, 3420–3424 (2009).
276. Schwartz, S., Meshorer, E. & Ast, G. Chromatin organization marks exon–intron structure. *Nature Struct. Mol. Biol.* **16**, 990–995 (2009).
277. Spies, N., Nielsen, C. B., Padgett, R. A. & Burge, C. B. Biased chromatin signatures around polyadenylation sites and exons. *Mol. Cell* **36**, 245–254 (2009).
278. Tilgner, H. *et al.* Nucleosome positioning as a determinant of exon recognition. *Nature Struct. Mol. Biol.* **16**, 996–1001 (2009).
279. Luco, R. F. *et al.* Regulation of alternative splicing by histone modifications. *Science* **327**, 996–1000 (2010).
280. Mercer, T. R. *et al.* DNase I-hypersensitive exons colocalize with promoters and distal regulatory elements. *Nature Genet.* **45**, 852–859 (2013).
281. Willingham, A. T. *et al.* A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science* **309**, 1570–1573 (2005).
282. Carrieri, C. *et al.* Long non-coding antisense RNA controls *Uchl1* translation through an embedded SINEB2 repeat. *Nature* **491**, 454–457 (2012).
283. Gong, C. & Maquat, L. E. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. *Nature* **470**, 284–288 (2011).
284. Mattick, J. S. RNA as the substrate for epigenome-environment interactions: RNA guidance of epigenetic processes and the expansion of RNA editing in animals underpins development, phenotypic plasticity, learning, and cognition. *Bioessays* **32**, 548–552 (2010).
285. Mattick, J. S. The central role of RNA in human development and cognition. *FEBS Lett.* **585**, 1600–1616 (2011).
286. Muotri, A. R. *et al.* Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435**, 903–910 (2005).
287. Coufal, N. G. *et al.* L1 retrotransposition in human neural progenitor cells. *Nature* **460**, 1127–1131 (2009).
288. Muotri, A. R. *et al.* L1 retrotransposition in neurons is modulated by MeCP2. *Nature* **468**, 443–446 (2010).
289. Faulkner, G. J. *et al.* The regulated retrotransposon transcriptome of mammalian cells. *Nature Genet.* **41**, 563–571 (2009).
290. Feschotte, C. Transposable elements and the evolution of regulatory networks. *Nature Rev. Genet.* **9**, 397–405 (2008).
291. Carroll, S. B. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008).
292. Brosius, J. The contribution of RNAs and retroposition to evolutionary novelties. *Genetica* **118**, 99–116 (2003).
293. Krull, M., Brosius, J. & Schmitz, J. Alu–SINE exonization: en route to protein-coding function. *Mol. Biol. Evol.* **22**, 1702–1711 (2005).
294. Cordaux, R., Udit, S., Batzer, M. A. & Feschotte, C. Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. *Proc. Natl Acad. Sci. USA* **103**, 8101–8106 (2006).
295. Kelley, D. & Rinn, J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol.* **13**, R107 (2012).
296. Kapusta, A. *et al.* Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet.* **9**, e1003470 (2013).
297. Czerwoniec, A. *et al.* MODOMICS: a database of RNA modification pathways. 2008 update. *Nucleic Acids Res.* **37**, D118–D121 (2009).
298. Cantara, W. A. *et al.* The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res.* **39**, D195–D201 (2011).
299. Motorin, Y., Lyko, F. & Helm, M. 5-methylcytosine in RNA: detection, enzymatic formation and biological functions. *Nucleic Acids Res.* **38**, 1415–1430 (2010).
300. Abbasi-Moheb, L. *et al.* Mutations in *NSUN2* cause autosomal-recessive intellectual disability. *Am. J. Hum. Genet.* **90**, 847–855 (2012).
301. Meyer, K. D. *et al.* Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell* **149**, 1635–1646 (2012).
302. Jia, G., Fu, Y. & He, C. Reversible RNA adenosine methylation in biological regulation. *Trends Genet.* **29**, 108–115 (2013).
303. Saletore, Y. *et al.* The birth of the epitranscriptome: deciphering the function of RNA modifications. *Genome Biol.* **13**, 175 (2012).
304. Saletore, Y., Chen-Kiang, S. & Mason, C. E. Novel RNA regulatory mechanisms revealed in the epitranscriptome. *RNA Biol.* **10**, 342–346 (2013).
305. Brosnan, C. A. *et al.* Nuclear gene silencing directs reception of long-distance mRNA silencing in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **104**, 14741–14746 (2007).
306. Dinger, M. E., Mercer, T. R. & Mattick, J. S. RNAs as extracellular signaling molecules. *J. Mol. Endocrinol.* **40**, 151–159 (2008).
307. Alleman, M. *et al.* An RNA-dependent RNA polymerase is required for paramutation in maize. *Nature* **442**, 295–298 (2006).
308. Rassoulzadegan, M. *et al.* RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* **441**, 469–474 (2006).
309. Chandler, V. L. Paramutation: from maize to mice. *Cell* **128**, 641–645 (2007).
310. Nadeau, J. H. Transgenerational genetic effects on phenotypic variation and disease risk. *Hum. Mol. Genet.* **18**, R202–R210 (2009).
311. Buckley, B. A. *et al.* A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality. *Nature* **489**, 447–451 (2012).
312. Herbert, A. & Rich, A. RNA processing in evolution: the logic of soft-wired genomes. *Ann. NY Acad. Sci.* **870**, 119–132 (1999).
313. Herbert, A. & Rich, A. RNA processing and the evolution of eukaryotes. *Nature Genet.* **21**, 265–269 (1999).
314. Mattick, J. S. Has evolution learnt how to learn? *EMBO Rep.* **10**, 665 (2009).
315. Mattick, J. S. & Makunin, I. V. Non-coding RNA. *Hum. Mol. Genet.* **15**, R17–29 (2006).
316. Gingeras, T. R. Origin of phenotypes: genes and transcripts. *Genome Res.* **17**, 682–690 (2007).
317. Mattick, J. S. Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays* **25**, 930–939 (2003).
318. Chamary, J. V., Parmley, J. L. & Hurst, L. D. Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nature Rev. Genet.* **7**, 98–108 (2006).
319. Lin, M. F. *et al.* Locating protein-coding sequences under selection for additional, overlapping functions in 29 mammalian genomes. *Genome Res.* **21**, 1916–1928 (2011).
320. Birnbaum, R. Y. *et al.* Coding exons function as tissue-specific enhancers of nearby genes. *Genome Res.* **22**, 1059–1068 (2012).
321. Stergachis, A. B. *et al.* Exonic transcription factor binding directs codon choice and affects protein evolution. *Science* **342**, 1367–1372 (2013).
322. Duret, L., Chureau, C., Samain, S., Weissenbach, J. & Avner, P. The *Xist* RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science* **312**, 1653–1655 (2006).
323. Capel, B. *et al.* Circular transcripts of the testis-determining gene *Sry* in adult mouse testis. *Cell* **73**, 1019–1030 (1993).
324. Hansen, T. B. *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384–388 (2013).
325. Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338 (2013).
326. Zhang, Y. *et al.* Circular intronic long noncoding RNAs. *Mol. Cell* **51**, 792–806 (2013).
327. Wang, J. *et al.* CREB up-regulates long non-coding RNA, *HULC* expression through interaction with *microRNA-372* in liver cancer. *Nucleic Acids Res.* **38**, 5366–5383 (2010).
328. Poliseno, L. *et al.* A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**, 1033–1038 (2010).
329. Wan, Y. *et al.* Genome-wide measurement of RNA folding energies. *Mol. Cell* **48**, 169–181 (2012).
330. Arteaga-Vazquez, M. A. & Chandler, V. L. Paramutation in maize: RNA mediated *trans*-generational gene silencing. *Curr. Opin. Genet. Dev.* **20**, 156–163 (2010).
331. Sarkies, P. & Miska, E. A. Is there social RNA? *Science* **341**, 467–468 (2013).
332. Nowacki, M. *et al.* RNA-mediated epigenetic programming of a genome-rearrangement pathway. *Nature* **451**, 153–158 (2008).
333. Lopez de Silanes, I., Stagno d'Alcontres, M. & Blasco, M. A. *TERRA* transcripts are bound by a complex array of RNA-binding proteins. *Nature Commun.* **1**, 33 (2010).
334. Walter, P. & Blobel, G. Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. *Nature* **299**, 691–698 (1982).
335. Ji, X. *et al.* SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. *Cell* **153**, 855–868 (2013).
336. Fox, A. H., Bond, C. S. & Lamond, A. I. P54nrb forms a heterodimer with PSP1 that localizes to paraspeckles in an RNA-dependent manner. *Mol. Biol. Cell* **16**, 5304–5315 (2005).

337. Mao, Y. S., Sunwoo, H., Zhang, B. & Spector, D. L. Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nature Cell Biol.* **13**, 95–101 (2011).
338. Tripathi, V. *et al.* The nuclear-retained noncoding RNA *MALAT1* regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* **39**, 925–938 (2010).
339. Tripathi, V. *et al.* Long noncoding RNA *MALAT1* controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. *PLoS Genet.* **9**, e1003368 (2013).
340. Sone, M. *et al.* The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons. *J. Cell Sci.* **120**, 2498–2506 (2007).
341. Barry, G. *et al.* The long non-coding RNA Gomafu is acutely regulated in response to neuronal activation and involved in schizophrenia-associated alternative splicing. *Mol. Psychiatry* **19**, 486–494 (2014).
342. Weinberg, M. S. *et al.* The antisense strand of small interfering RNAs directs histone methylation and transcriptional gene silencing in human cells. *RNA* **12**, 256–262 (2006).
343. Janowski, B. A. *et al.* Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. *Nature Struct. Mol. Biol.* **13**, 787–792 (2006).
344. Ling, J., Baibakov, B., Pi, W., Emerson, B. M. & Tuan, D. The HS2 enhancer of the beta-globin locus control region initiates synthesis of non-coding, polyadenylated RNAs independent of a *cis*-linked globin promoter. *J. Mol. Biol.* **350**, 883–896 (2005).
345. Li, W. *et al.* Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**, 516–520 (2013).
346. Gumireddy, K. *et al.* Identification of a long non-coding RNA-associated RNP complex regulating metastasis at the translational step. *EMBO J.* **32**, 2672–2684 (2013).
347. Watson, J. D. & Crick, F. H. Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171**, 964–967 (1953).
348. Holmes, D. S., Mayfield, J. E., Sander, G. & Bonner, J. Chromosomal RNA: its properties. *Science* **177**, 72–74 (1972).
349. Brannan, C. I., Dees, E. C., Ingram, R. S. & Tilghman, S. M. The product of the *H19* gene may function as an RNA. *Mol. Cell. Biol.* **10**, 28–36 (1990).
350. Fire, A. *et al.* Potent and specific genetic interference by double stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
351. Hamilton, A. J. & Baulcombe, D. C. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**, 950–952 (1999).
352. Elbashir, S. M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).
353. Mattick, J. S. Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep.* **2**, 986–991 (2001).
354. Liu, J. *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437–1441 (2004).

Acknowledgements
This work was supported by the US National Institutes of Health grant PO1 AI099783-01 and the Australian Research Council Future Fellowship FT130100572 to K.V.M., and by the National Health and Medical Research Council Australia Fellowship 631688 to J.S.M.

Competing interests statement
The authors declare no competing interests.