Review

The central role of RNA in human development and cognition

John S. Mattick

Institute for Molecular Bioscience, University of Queensland, Brisbane 4072, Australia

A B S T R A C T

It appears that the genetic programming of humans and other complex organisms has been misunderstood for the past 50 years, due to the assumption that most genetic information is transacted by proteins. However, the human genome contains only about 20,000 protein-coding genes, similar in number and with largely orthologous functions as those in nematodes that have only 1000 somatic cells. By contrast, the extent of non-protein-coding DNA increases with increasing complexity, reaching 98.8% in humans. The majority of these sequences are dynamically transcribed, mainly into non-protein-coding RNAs, with tens if not hundreds of thousands that show specific expression patterns and subcellular locations, as well as many classes of small regulatory RNAs. The emerging evidence indicates that these RNAs control the epigenetic states that underpin development, and that many are dysregulated in cancer and other complex diseases. Moreover it appears that animals, particularly primates, have evolved plasticity in these RNA regulatory systems, especially in the brain. Thus, it appears that what was dismissed as 'junk' because it was not understood holds the key to understanding human evolution, development, and cognition.

© 2011 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

1. Introduction

The purpose of this article is to make the case that the genomic programming of humans and other complex organisms has been misunderstood, because of the assumption that most genetic information is transacted by proteins. This assumption derived from the early studies of the *lac* operon in *Escherichia coli*, and from the ensuing common interpretation of the central dogma, i.e., that information mostly flows from DNA through the temporary intermediate of RNA, which is then translated into proteins that effect all of the major structural, catalytic and (notably, for the purposes of this discussion) regulatory functions of the cell. While it has always been clear that some RNAs are end-point gene products in themselves, and Crick recognized this in the central dogma [1], this has traditionally been thought to be limited to infrastructural RNAs such as ribosomal, transfer, spliceosomal and small nucleolar RNAs, involved directly or indirectly in protein expression and other core cellular functions.

The protein-centric view of molecular genetics and cell biology, extended later into developmental biology, has deep roots, dating back to the early biochemical studies, even prior to the elucidation of the double helical structure of DNA in 1953. It was fashioned in a mechanical age that had little appreciation that genetic information may be transacted (as opposed to inherited) in other ways, although the use of codes and the code-breaking successes of World War II, from Morse code through to the Enigma machine,

led to the ready acceptance of the concept of the 'genetic code', at least insofar as it applied to protein-coding sequences, and later to *cis*-acting sites in DNA and RNA recognized by regulatory proteins.

This protein-centric view became entrenched within the first few years of molecular biology, but was not without its contemporary challengers. The Nobel Laureate Barbara McClintock, who was celebrated for her insight that transposons are 'controlling elements' in corn, and was possibly the most original thinker in the early years of molecular biology, wrote in 1950 [2]: "Are we letting a philosophy of the [protein-coding] gene control [our] reasoning? What, then, is the philosophy of the gene? Is it a valid philosophy? ... When one starts to question the reasoning behind the origin of the present notion of the gene (held by most geneticists), the opportunity for questioning its validity becomes apparent." It seems few heeded her admonition, especially as her insights were apparently discredited in the eyes of others by her promotion of the idea that 'controlling elements' are the key to understanding development.

A few others also kept an open mind. François Jacob and Jacques Monod, who received the Nobel Prize for their work on the *lac* operon, mooted the possibility that the *lac* repressor, which they had identified genetically but not biochemically, might be an RNA [3]. However, the idea was lost when it was subsequently shown that the *lac* repressor was a protein [4], which undoubtedly was an important factor in the entrenchment of the protein-centric orthodoxy, especially as it pertained to gene regulation and the subsequent rise of the concept of 'transcription factors'.

E-mail address: j.mattick@imb.uq.edu.au

Occasionally related ideas emerged. Most notably, in the late 1960s, Roy Britten and Eric Davidson, using renaturation hybridization kinetics to study the sequence complexity of DNA and RNA, noted that the extent of genomic DNA broadly increased with developmental complexity, and made two unexpected observations, specifically (i) that the population of 'heterogenous nuclear RNA' (hnRNA) is far more complex than messenger RNA (mRNA), and (ii) that a substantial proportion of the genome is comprised of low complexity/high copy number 'repetitive' sequences, some of which at least are differentially expressed at different developmental stages. This led them to propose that there may be considerable regulatory RNA in the nucleus of eukaryotic cells, and that repetitive sequences (later found to be transposon-derived) may comprise parts of regulatory networks [5,6]. They also predicted that many of these putative nuclear regulatory RNAs would be chromatin-associated, which has now been shown to be the case (see below). Unfortunately, these ideas could not be tested at the time and, although their papers have been highly cited, they appear to have been not well received, or ignored, by the mainstream, similar to McClintock's experience. In any case, and somewhat surprisingly, these ideas were not re-visited later, when the discovery of introns explained, at least in part, the nature of hnRNA, and the genome projects later revealed the full repertoire of transposon-derived sequences.

2. The great surprises

This protein-centric conceptual framework imbues almost every aspect of the 'philosophy' (as McClintock put it) of molecular biology, and has persisted until the present, despite a number of subsequent surprises that, like Britten and Davidson's observations, should have given serious pause for thought, and that collectively paint an entirely different picture.

The first of these was the discovery in late 1977 that most protein-coding genes in mammals and other complex organisms are not co-linear with their encoded products, but are mosaics of small segments of protein-coding sequences ('exons') interspersed with often vast tracts of non-protein-coding sequences ('intervening sequences' or 'introns') [7,8]. Without question, this was then and still remains the biggest surprise in the history of molecular biology [9]. However, within a very short period it was universally concluded that, because they did not code for protein, and despite the fact that they are transcribed into RNA, these sequences are mostly non-functional evolutionary debris [10,11], which was then rationalized as the likely hangover of the early evolution/assembly of genes [12,13] and/or retrotransposition of 'selfish DNA' (see below). The fact that these sequences were excised (and apparently discarded) and a contiguous mRNA assembled by splicing allowed the community to breathe easy, as the core flow of information from DNA to protein was left operationally undisturbed. There is, of course, another far more interesting interpretation, with potentially profound consequences – that the excised RNAs are also transmitting information, which would mean that the system, at least in the higher eukaryotes, is far more complex and sophisticated than conventionally thought, and by logical extension that there is a hidden layer of RNA regulatory networks that might be important, even central, to developmental ontogeny [14].

The second surprise, extending from the earlier work of Britten and Davidson, was that the genomes of humans and other complex eukaryotes are full of transposon-derived sequences of various classes, pejoratively referred to as 'repetitive elements' or more simply 'repeats'. Again, on the same logic (that they do not encode proteins, except occasionally for their own mobilization) and despite McClintock's insights, these sequences were quickly assumed to be mostly non-functional and their presence rationalized as

'selfish DNA' [15,16], part of the emerging view of most of the human genome as an evolutionary graveyard. The vast tracts of intergenic and intronic sequences in the human genome, riddled with transposon-derived sequences, have been described, among other things, as 'junk … in the attic', on the proposition that some have been exapted for function but many or most only have potential future value [17], although the proportion that may have been exapted for function was and is still unknown. Again, of course, there is a more interesting and potentially more profound explanation.

The raw material for evolution is duplication and transposition, with the latter providing far more flexibility in terms of dissemination of functional cassettes and re-structuring of regulatory networks for phenotypic divergence. This appears to be the major place that evolution has played in relation to the radiation of animals, which generally have a similar proteome (see next). Nonetheless, and again despite McClintock's insights, on the dubious and entirely circular assumption of the non-functionality of the transposon-derived 'ancient repeats' (ARs) that are common to the genomes of human and other mammals (which have persisted with us for tens if not hundreds of millions of years), and the independent but equally dubious assumption that the recognizable extant population of ARs is representative of the original distribution, not simply its more conserved end [18], it has been estimated that only ∼5% of the human genome is under purifying selection [19,20] (although other analyses, which do not rely on this assumption, have indicated that evolutionary selection is widespread across the genome [21]). That is, the assumed non-functionality of a small proportion of the genome has been used to impute the non-functionality of most of the remainder, because they are evolving at similar rates. If either this or the other assumption mentioned above is wrong, then so is the conclusion [18].

The third great surprise, which was entirely contrary to expectations, is that the number and repertoire of protein-coding genes remains relatively static across the metazoan lineage, despite enormous increases in developmental and cognitive complexity [22]. The simple nematode *Caenorhabiditis elegans*, which only has ∼1000 somatic cells, has almost 20,000 protein-coding genes, similar to that in the human [23–25]. However, humans have approximately 10^{14} cells, sculpted into a myriad of different muscles, bones and organs that have complex architectures, as well as a brain with approximately 10^{11} neurons [26]. Moreover, despite some interesting expansions and innovations, such as RNA editing in vertebrates (see below), the majority of these genes are orthologous (i.e., have similar functions), even in sponge, the most primitive metazoan [27], including most of those involved in the cell signaling and homeotic pathways that underpin multicellular development. That is, all animals have a similar protein toolkit [28], and therefore the relevant information that programs progressively more complex organisms must lie elsewhere in the genome, presumably in an expanded regulatory architecture.

3. Scaling of regulatory architecture

The response to this lack of gene scaling and limited proteomic diversification has been to assert the power of combinatoric control – specifically that the power of combinatorial control by transcription factors (and other regulatory proteins, such as those involved in alternative splicing), will lead to "a dramatic expansion in regulatory complexity" [29]. As a logical extension, it is then simply proposed that the developmental programming of more complex organisms has been enabled by an expansion in the numbers and complexity of the *cis*-acting sequences recognized by these regulatory proteins, alteration of which, along with alterations in the expression of the regulatory proteins and the networks

in which they participate, lies at the heart of phenotypic diversity [29,30].

While there is little doubt that regulatory alterations and expansions underpin the emergence and divergence of more developmentally complex organisms, the essential argument here (although not spelled out explicitly in mathematical terms) is that the range of regulatory options scales factorially with the numbers of regulatory proteins, and that in animals this number is so great ($\geqslant 1000$ regulatory proteins in human or *C. elegans* [29] = potentially $\geqslant 1000!$ combinations $\equiv$ a number far bigger than the estimated number of atoms in the universe) as to be superficially more than capable, even if heavily discounted, of providing sufficient regulatory complexity to program human development, so there is no need for further concern or further justification. The necessary power is implicit in the assumption. However, while it is clear that many factors can influence a decision to transcribe a gene, so there is some sort of 'combinatorial' control, it is by no means clear that this scales factorially. Indeed, the implied assertion has not only never been clearly articulated, but it has also (consequently) never been justified, mathematically, by reference to decision theory, or mechanistically; nor has it been subjected to critical scrutiny, as the lack of articulation obscures the issue, and the assertion fits comfortably with mainstream preconceptions.

How do regulatory factors really scale with gene number? This is a difficult question to answer, but the available evidence suggests that it is quite the opposite to that which is assumed. Prokaryotic genomes are predominantly composed of protein-coding sequences, and therefore it is possible to do a first approximation analysis of the relationship between the numbers of genes encoding regulatory factors and the total numbers of genes in cells of different genetic complexity (despite the existence of a limited set of regulatory RNAs in these organisms). This is not possible with eukaryotes, which contain indeterminate but apparently large numbers of regulatory RNAs, both large and small (see below). In prokaryotes, however, it is clear that the number of regulatory genes R (those encoding proteins with characteristic motifs such as DNA-binding domains) scales quadratically with gene number N ($R \propto N^2$, not as some sort of inverse factorial $R \propto 1/N!$), over the entire range of bacterial genome sizes [31–33]. This empirical fact has many implications. Since regulatory genes are scaling twice as fast as total genes, and there is no hint of any deviation at the top end of the range, there must be a limit (as the number of regulatory genes cannot exceed the total), which is ostensibly the observed upper limit of bacterial genome sizes (about 10 Mb, ~9000 genes), where over 20% of the genes are regulatory. This limit was likely reached early in evolution.

The quadratic relationship and inferred limit also implies that higher complexity eukaryotes must have solved this problem, in all likelihood by moving to a more genomically efficient (and evolutionarily flexible) RNA-based regulatory system (that separates regulatory signal from consequent action, as exemplified by RNA interference pathways [34,35]), together with the introduction of other levels of control and compartmentalization, all hallmarks of eukaryotes, to mitigate the global scaling problem. Finally, this relationship implies that combinatorial action of regulatory factors, as envisaged to allow dramatic expansions in regulatory capacity, does not hold, at least in prokaryotes (as the scaling of regulatory factors would be entirely different) and therefore also probably not in eukaryotes, unless one can conceive of chemical space (specifically protein–protein interaction space) that is permitted in eukaryotes but not prokaryotes. If so, the clear implication is that in all cells, and indeed in all functionally integrated complex systems, the proportion of regulatory information increases with increasing complexity, i.e., occupies a progressively greater amount of the total information required to program the ontogeny and operation of the system [33].

Few would disagree that increased developmental complexity requires an expansion of regulatory information and that this information resides mainly in the non-protein-coding portion of the genome. As noted already, most molecular biologists have assumed that this information is (largely) restricted to *cis*-acting regulatory sequences that interact with 'regulatory' proteins to control gene expression, and, since it is inconceivable that such sequences could cover a huge fraction of the human genome, have been generally comfortable with the proposition that the majority of the genome is non-functional. However, while the number of protein-coding genes and the extent of protein-coding sequences remains surprisingly static, it is clear that the extent of non-protein-coding intronic and intergenic sequences does scale with increased developmental complexity [22,36] (Fig. 1), and indeed is the only variable yet demonstrated to do so (along with the complement of regulatory RNAs, see below). This does not prove anything, but is at least consistent with the proposition that high developmental complexity requires a vastly expanded regulatory architecture, which can only be falsified by a downward exception, i.e., a complex organism that has much less non-coding sequence that demonstrably simpler ones (as opposed to an upward outlier that has more non-coding sequence than organisms of equivalent complexity), which so far has not been observed. Here it is important to note that *Fugu rubripes* and *Arabidopsis thaliana*, which possess the smallest known vertebrate and angiosperm genomes, respectively, do not deviate substantially from the broad relationship described above [22].

4. Pervasive transcription of the genome

The fourth great surprise of genomic analyses has been that, irrespective of the extent of non-protein-coding sequences in different genomes, the vast majority of these sequences are transcribed, apparently in a developmentally regulated manner, mainly into non-protein-coding RNAs (ncRNAs). A variety of transcriptomic studies, primarily using high throughput sequence analyses of normalized cDNA libraries [37–42] and cDNA interrogation of genome tiling arrays [43–48], have shown that the mammalian transcriptome is amazingly complex and consists of a myriad of overlapping and interlacing transcripts from both strands, including intronic, antisense and intergenic transcripts that exhibit different start sites, termination points, splicing and expression patterns in different cells [49–52].

A frequent response to these observations, especially as many of these unexpected ncRNAs appear to be expressed at relatively low levels, has been to suspect or assert that they are 'transcriptional noise', an alternative hypothesis with the appeal of not disturbing the prevailing orthodoxy of gene regulation. Despite the substantial evidence that genomic sequences specifying these ncRNAs have all of the hallmarks of conventional genes and the rapidly accumulating numbers of functionally validated examples ([53], see below), the debate about the relevance of these transcripts continues. A recent article, for example, that compared the signals derived from transcriptomic interrogation of genome tiling arrays with next generation short read sequencing data, concluded that the former had significant problems with false positives, that many of the short sequence 'singleton' tags in the latter were so scattered as to appear "of random character" (i.e., consistent with noise) and that much of the genome is "not pervasively transcribed" [54].

However, apart from problems with the methodology employed, the data are equally consistent with low level sampling of rare transcripts (that may, e.g., be cell type-specific), considering that such deep sequencing suffers from diminishing sensitivity for uncommon transcripts by the dominance of common highly expressed (usually protein-coding) RNAs [55], especially in the brain
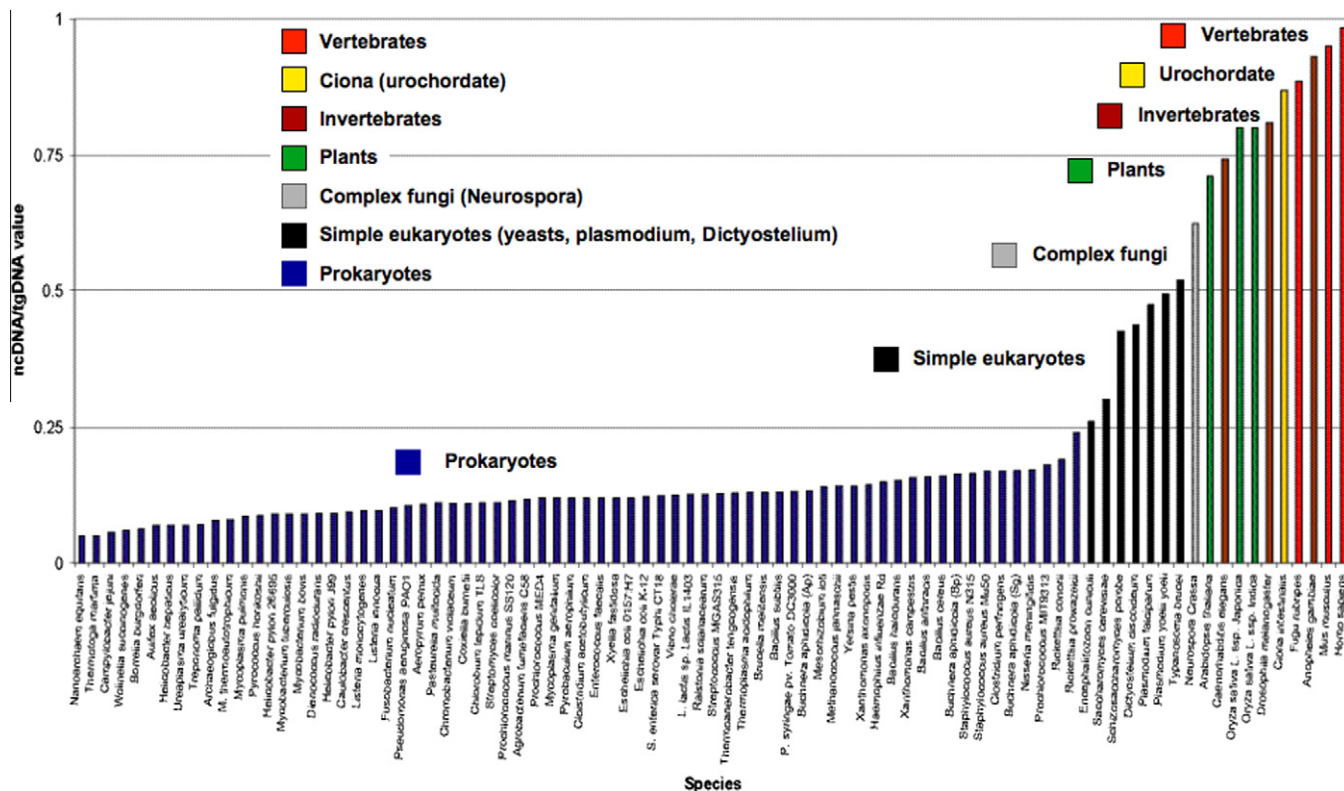
Fig. 1. The proportion of non-coding DNA increases with developmental complexity. Adapted from [22].

where much of the analysis was done [54]. Indeed some ncRNAs are easily detectable by the relatively insensitive technique of in situ hybridization in particular subregions of the brain, such as the dentate gyrus [56] (Fig. 2), which occupies a tiny fraction of the brain and would be difficult to detect in whole brain transcriptomic analysis, even by 'deep' sequencing. Moreover, a new technique called RNA Capture-Seq, which combines array capture of cDNAs with deep sequencing to increase the sensitivity of transcriptomic interrogation of targeted loci, has revealed that so-called 'gene deserts' actually express a wide range of spliced transcripts, many of which were not detected by a single tag in pre-capture libraries, as well as previously unknown isoforms of intensively studied protein-coding loci such as *p53* [57].

5. A world of long non-coding RNAs

There are tens if not hundreds of thousands of long antisense, intergenic and intronic ncRNAs (lncRNAs) expressed from mammalian genomes [41,57], with abundant evidence of their involvement in eukaryotic cell and developmental biology (for recent review see [58]). A large fraction of lncRNAs is expressed in the brain [56,59–61]. Many lncRNAs are also derived from enhancers [62–65], enigmatic non-coding regulatory elements that act at a distance to control gene expression during development, and which are thought to act by recruiting transcription factors and inducing chromosomal looping to bring these factors into contact with target promoters [66]. Reciprocally some lncRNAs emulate the properties of enhancers [67], which suggests that enhancer action may integrally involve a derived RNA [68], possibly to mediate higher order chromatin interactions and/or epigenetic changes [69,70]. These effects may explain the equally enigmatic genetic phenomena of transvection [68,69] and transinduction [63], the latter (along with clear evidence of selection on synonymous codon sites [71]) suggesting that even mRNAs

may have embedded regulatory functions in addition to their protein-coding capacity.

It has also recently become evident that conventional protein-coding loci may also produce both small and large non-coding RNAs, by regulated post-transcriptional cleavage of mRNAs [72,73], including within 3′ untranslated regions (3′UTRs) [74,75]. The latter can be expressed in a highly cell-specific manner (e.g., in the cortex and hippocampus in the brain, or Sertoli cells in the testis) [75] (Fig. 3), with genetic evidence dating back many years showing that these sequences can transmit information in trans separately from their normally associated protein-coding sequences and independently of their normal *cis*-acting functions in the regulation of mRNA translation and stability. This includes the restoration of the oogenesis defect in Drosophila *oskar* mutants [76] and the inhibition of cell division, suppression of malignancy and induction of differentiation by ectopically expressed 3′UTRs from a variety of genes in mammalian cells [77–81].

In addition, genome tiling array-based transcriptomic studies, which do not require ribosomal RNA depletion by oligo(dT) purification, showed that almost half of the transcripts found in human cells are not polyadenylated, and are largely of a different sequence composition from the RNA polymerase II-derived polyA+ fraction [46]. These transcripts are possibly synthesized by RNA polymerase III [82], indicating that for a generation a large portion of the transcriptome has been hidden from view for technical reasons. The same applies to repetitive sequences, which are frequently masked out of such analyses, but for which there is increasing evidence of differential expression [83–86], dating back many years [87,88].

6. Functions of lncRNAs

Although the vast majority of lncRNAs await characterization, there are compelling genome-wide indices of their functionality [53], as indicated by:
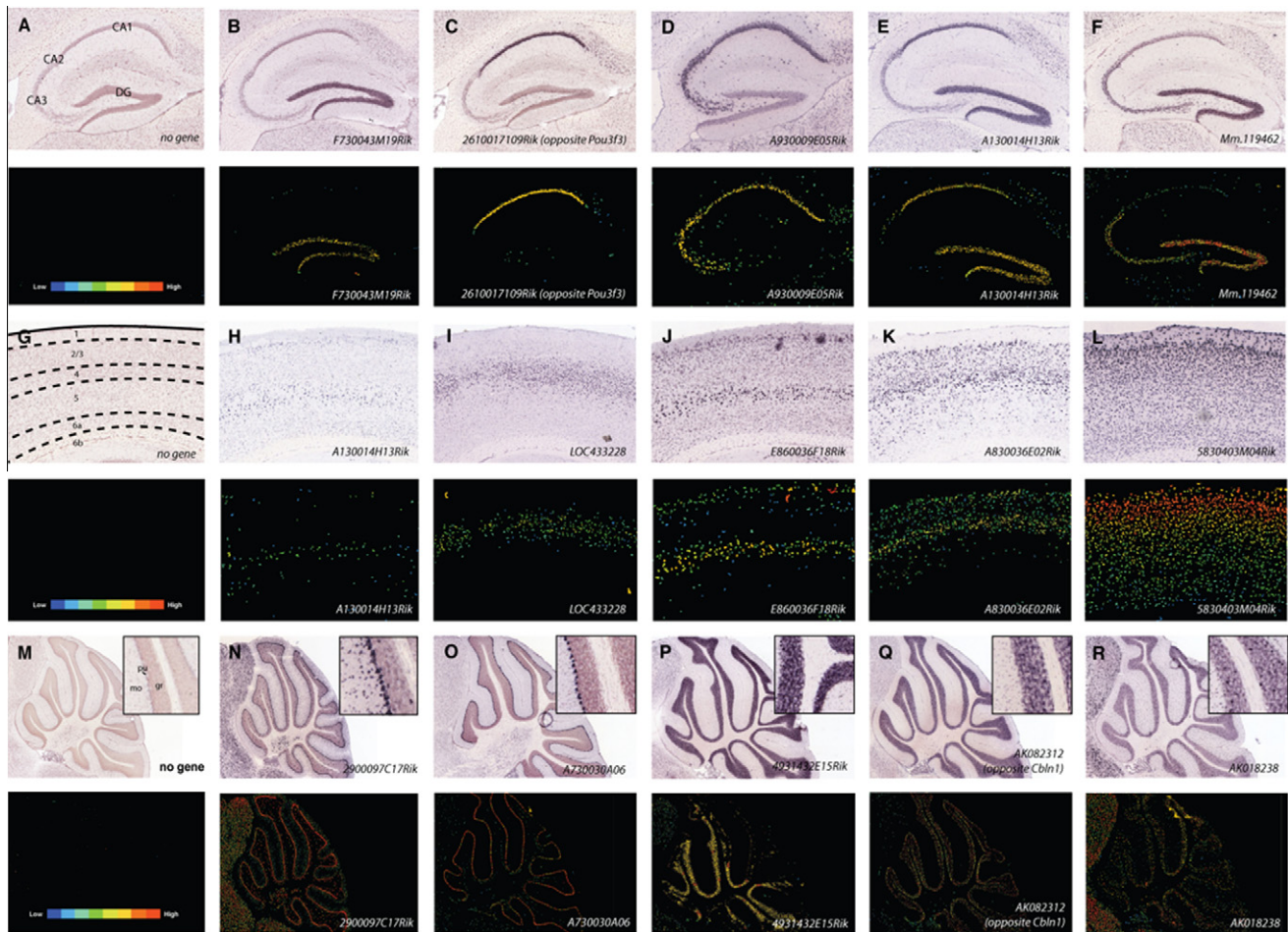
Fig. 2. Regionally enriched expression of ncRNAs in mouse hippocampus, cerebral cortex, and cerebellum. ISH images of ncRNA expression (accession numbers indicated) in sagittal plane accompanied by false-color heat map below. (A) No probe control in hippocampus, with the functionally distinct CA1, CA2, CA3, and dentate gyrus (DG) subfields indicated. (B–F) Enriched ncRNA expression in DG (B), CA1 (C), CA1–CA3 (D) and DG and CA1/CA2-3 in combination (E and F). (G) No probe control with labeled cortical layer boundaries. (H–L) Enriched ncRNA expression that correlates with specific cortical laminae. (M) No probe control in cerebellum, with the molecular (MO), Purkinje (PU), and granular (GR) layers indicated on detailed Inset. (N–R) Enriched ncRNA expression associated with cerebellar subregions. Adapted from [56].
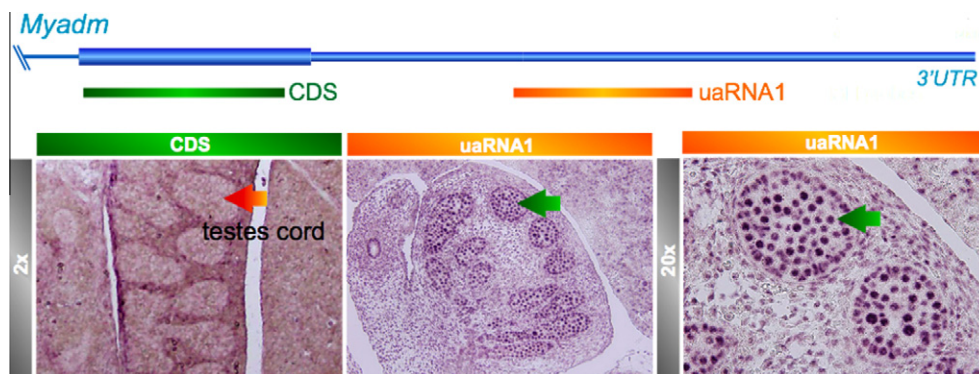

Fig. 3. Separate expression of Myadm coding sequences and 3′UTR in mouse testis [75]. Image courtesy of Dagmar Wilhelm.

(i) Conservation of their promoters, splice junctions, exons, predicted structures, genomic position and expression patterns [41,89–100].

(ii) Dynamic expression and alternative splicing during differentiation [94,95,101–103].

(iii) Altered expression or splicing patterns in cancer and other diseases [103–117].

(iv) Association with particular chromatin signatures that are indicative of actively transcribed genes [94,95].

(v) Regulation of their expression by key morphogens, transcription factors and hormones [94,95,116,118,119] and

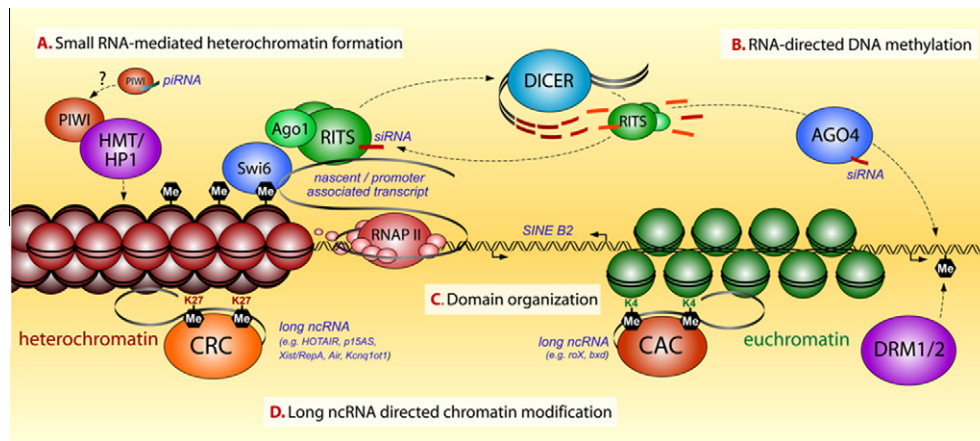(vi) Tissue- and cell-specific expression patterns and subcellular localization [59,102,116,120–130].

Fig. 4. RNA regulation of epigenetic processes. Reproduced from [184] with permission.

Indeed, different but substantial subsets of all of polled lncRNAs are differentially and dynamically expressed in all differentiation or disease systems that have been examined. These include the differentiation of embryonic stem cells [94], neuronal cells [131], muscle cells [102], T cells [132], breast epithelial cells [103] and cancer [117]. A survey of the expression of over 1300 lncRNAs in mouse brain showed that over 600 were expressed in highly specific locations, such as different regions of the hippocampus, different layers of the cortex, or different parts of the cerebellum (Fig. 2), with most (where the resolution was sufficient to determine) showing specific subcellular locations [56]. Since lack of conservation is uninformative [18,89], the only reliable index of function, although by no means proof of function, is differential expression [18]. By this criterion, most of the genome is imputed to be functional.

Many lncRNAs are associated with particular subcellular structures [133], including novel subnuclear domains in a subset of neurons [125] or Purkinje cells [56], and paraspeckles, which are as yet not well understood mammal-specific subnuclear domains involved in the retention of edited transcripts (containing Alu elements, see below) and induced in differentiated cells [102, 126–129,134–136]. What feature of mammalian differentiation and development that involves these structures (and is not required in other vertebrates) is not known, but is an intriguing question, especially in view of the recent report that mice lacking paraspeckles appear superficially normal [137]. Preliminary evidence suggests that paraspeckles may have a role in regulating the nuclear-cytoplasmic shuttling of RNAs that are subject to RNA editing [134,138], which appears to be associated with the rise of cognition (see below), and may be a feature not only of mammalian brain function, but also other aspects of mammalian reproduction, development and physiology associated with the exchange of high reproductive rates for extended nurturing, presumably with a net advantage for survival and reproductive fitness.

Although very few have yet been studied, there are rapidly increasing numbers of reports showing that lncRNAs have biological function [139], usually using siRNA-mediated knockdown and/or ectopic expression in cell-based assays (for recent compilations see [53,140]). These include functions, for example, in the modulation of the reprogramming of induced pluripotent stem cells [141], regulation of homeotic gene expression [101], cancer metastasis [117,142], breast development [103,143], retinal development [144,145], parental imprinting [146–151], X-chromosome dosage compensation [152–157], paraspeckle assembly [102,126–129] and p53 regulation [158–160].

In some cases, mechanistic insights into the mode of action of lncRNAs are emerging. For example, the nuclear lncRNA Malat1 ('metastasis-associated lung adenocarcinoma transcript 1') is expressed in numerous tissues and is highly abundant in neurons, where it has been shown to play a role in synaptogenesis [161] and to regulate alternative splicing by interacting with SR family pre-mRNA-splicing factors [161,162] and other nuclear RNA binding proteins [163]. It also yields a tRNA-like cytoplasmic RNA by 3′ end processing, although the function of this small RNA is unknown [164].

Most functionally characterized lncRNAs appear to play a role in differentiation and development, which is the broad prediction from the presence and progressive expansion of these transcripts in developmentally complex organisms. Consistent with this, a major, if not the major, function of these lncRNAs appears to be the regulation of epigenetic processes.

7. RNA regulation of epigenetic processes

Epigenetic processes are central to differentiation and development [165–170], long-term responses to environmental variables [171,172], and brain function [173–179]. Epigenetic information is encoded by the methylation [180] and hydroxymethylation [181,182] of cytosines in DNA, and in a wide range of modifications of the histones that package DNA into nucleosomes [183]. These are catalyzed by a suite of ∼60 generic enzymes/chromatin modifying complexes that impose a myriad of different chemical marks at hundreds of thousands of genomic locations in different cells at different stages of differentiation [183].

What determines the site-selectivity of these chromatin remodeling complexes, how is the position of nucleosomes regulated, and what is the molecular basis of environment–epigenome interactions? The likely answer to all of these questions is RNA [184–186]. It had been thought that these processes were directed by transcription factors, which clearly can exert powerful effects on cell state, capable of (re-)programming stem cells [187] or forcing differentiation into e.g., myoblasts [188]. However, the enormous and underappreciated challenge for genetic programming is not simply to define the phenotypic state of a cell, or to wind it back or forward, but rather to organize the 4-dimensional growth and differentiation of cells into a myriad of precisely 3-and 4-dimensionally sculpted organs and tissues [189] – different vertebrae or muscles, for example, have a unique architecture tuned to their functional position. While many proteins, including transcription factors and homeotic proteins, which are often differentially expressed and

state-specific (see e.g., [190,191]), have a role in this process, these clearly also require additional information for their site-specificity (as only a subset of potential sites are actually addressed in any given context). This additional information increasingly appears to be embedded in regulatory RNAs that control chromatin architecture by directing effector 'regulatory' proteins to their sites of action [184] (Fig. 4).

The evidence implicating RNAs in the epigenetic control of chromatin architecture and transcription dates back many years [69,184,192–194] and includes the physical association of RNA with chromatin [195–198] and chromatin regulatory proteins such as DNA methyltransferases and REST [199,200], the presence of RNA binding domains in proteins in chromatin remodeling complexes [201–206] and the observations that major classes of transcription factors, such as zinc finger and Y-box proteins, have RNA or RNA:DNA binding activity [207–209]. The direct evidence of lncRNA involvement in epigenetic processes includes the association of subsets of lncRNAs with Trithorax-group proteins and forms of modified histones in active chromatin [94,210], or with Polycomb-group proteins and histone modifications associated with repressed chromatin [101,150,151,198,200,211–217], which also appears to involve elements of the RNA interference pathway [218–220].

Apart from lncRNAs that regulate allele-specific expression at imprinted loci and sex chromosomes, a number of interesting examples have emerged over the past few years. These include the epigenetic regulation of the tumour suppressor gene *p15* by its antisense RNA [217,221], called ANRIL, independently shown to be associated with susceptibility to range of complex diseases, including coronary disease, intracranial aneurysm, type 2 diabetes, gliomas and basal cell carcinomas, among others, with preliminary evidence suggesting that the common factor is its role in the regulation of cell proliferation and senescence [222]. Indeed, most genetic variants associated with complex diseases are non-coding [223], presumably regulatory, many if not most of which may affect intergenic, intronic and antisense lncRNAs [53].

Another that has been well studied (by lncRNA standards) is HOTAIR, a relatively rapidly evolving mammal-specific transcript [224], which is derived from the *HOXC* locus and was first identified among 231 *HOX*-associated lncRNAs that are differentially expressed along spatiotemporal axes during development [101]. It represses transcription at the *HOXD* locus in trans [101], is up-regulated in breast cancer and increases cancer invasiveness and metastasis [142], which has also been observed with another lncRNA SPRY4-IN1, which is up-regulated in melanoma [117].

These examples are undoubtedly the tip of a very big iceberg, especially when one considers that, in all likelihood, most of the 10^{14} cells in a human, like the 10^3 in *C. elegans*, will have a unique ontogeny and positional identity (a proposition that is supported by the phenotypic similarity of monozygotic twins) that is epigenetically controlled by such regulatory RNAs.

8. A world of small RNAs

There are also many classes of small RNAs that have been discovered in recent years. The best characterized is microRNAs (miRNAs), ∼23 nt small RNAs derived from short hairpin precursors that generally appear to control mRNA translation and stability via their (usually imperfect) recognition of target sites usually in the 3′UTR of mRNAs) and interaction with Argonaute-containing RNA-induced silencing complexes (RISCs) [225]. Approximately 1000 miRNAs have been identified in human, but there are likely to be many more, especially if significant numbers are lineage- or species-restricted [226] and/or, like *lsy6* in *C. elegans* [227], are cell-specific.

The repertoire of miRNAs, like lncRNAs, has expanded during animal evolution [228–231]. They regulate most mRNAs [232,233] (and possibly lncRNAs), appear to influence almost every facet of animal and plant development [234–242], as well as many aspects of brain function (see e.g., [243–246]), and are often dysregulated in cancer and other complex diseases [247,248]. There is much more to be learned about their biology and functional mechanisms, including the parameters that determine target recognition [249–252], especially in view of the observations that miRNA isoforms are developmentally regulated [253] and that at least some miRNAs are localized in the nucleus [254,255].

Similar small RNAs, termed siRNAs (small interfering RNAs), derived from longer hairpin precursors or duplex RNAs, also act through the RNA interference pathway, usually by perfect matches with the target sequence, resulting in mRNA degradation [35] as well as transcriptional silencing via epigenetic effects on target sequences [184,219,256–258], with as yet not well understood interplay between different facets of the RNA interference system [35]. There appear to be a number of pathways as well as intercellular transport of small RNA signals involved, at least in plants [259], and possibly also in animals [260].

Related animal-specific small (∼24–30 nt) RNAs that also involve the RNA interference system, piRNAs (piwi-interacting RNAs) [261–263], are involved in the silencing of transposons, primarily in the germline [264–267]. PiRNAs are 2′-O-methylated at their 3′ ends, which permits their specific recognition by the PAZ domain of specific Argonautes that are expressed in the germline [268,269]. They appear to be involved in genome defense [267,270,271], although another more interesting possibility is that they have evolved to regulate retrotransposon expression in more subtle ways during early development [256,272–276]. These RNAs are also linked to epigenetic changes (DNA methylation and histone modifications) [277–279], assembly of the telomere protection complex [273], and complex genetic phenomena such as hybrid dysgenesis [280], position effect variegation [256] and trans-silencing [281].

Another class of small RNAs, termed small nucleolar RNAs (snoRNAs), which may be spread by retrotransposition [282–284], are derived from the introns of protein-coding and non-coding host transcripts [285–288], with at least two cases of the latter having independent functions as lncRNAs [103,289]. SnoRNAs guide specialized protein complexes to impart sequence-specific 2′-O-methylation (box C/D snoRNAs) or the isomerization of specific uridines to pseudouridines (box H/ACA snoRNAs) in target RNAs, and are usually localized in the nucleolus, with a related class of box H/ACA snoRNAs found in (nuclear) Cajal-bodies [290–293] using a specific localization signal that is also found in telomerase RNA [294]. SnoRNAs were initially thought simply to modify ribosomal RNAs to tune their chemical properties in translation [295], but some show imprinted, tissue-specific and/or context-dependent expression, especially in the brain [296–300], and target other cellular RNAs, including small nuclear (spliceosomal) RNAs (snRNAs), transfer RNAs (tRNAs) and possibly even mRNAs, in ways and for purposes that are not yet well-defined or understood [291]. RNA can also be modified by cytosine methylation. Dnmt2, named because of its homology to DNA methyltransferases, is in fact an RNA methyltransferase [301,302], which, although its target spectrum is not known, plays a role in the development of the brain and other organs [303] and is required for retrotransposon silencing in somatic cells of Drosophila [304].

Recently, it has been shown that snoRNAs, snRNAs and tRNAs are cleaved at specific positions to produce smaller RNAs [305–307]. Most, if not all, small nucleolar RNAs in eukaryotes, from fission yeast to human, are processed to three different subspecies of small RNAs of three different canonical sizes (∼17–19 nt and ∼30 nt from boxC/D snoRNAs, and ∼20–24 nt from box

H/ACA snoRNAs), which show altered expression patterns in mutants affecting the RNA interference pathway [306]. The size of the box H/ACA-derived small RNAs is similar to miRNAs, and some have been shown to function as miRNAs [308,309]. A number of annotated miRNA precursors have box H/ACA snoRNA features [310] and some miRNAs have a nucleolar location [254]. Small RNAs derived from box C/D snoRNAs have been implicated in the regulation of splicing [311], whereas those derived from tRNAs have been found to be preferentially associated with different Argonautes, and to affect the global regulation of the RNA silencing system [307]. These findings all point to a complex evolutionary and functional interplay between different classes of small RNAs in the regulation of gene expression, whose dimensions have barely begun to be explored.

Analysis of deep sequencing small RNAs datasets identified a new class of animal-specific nuclear-localized tiny (~18 nt) RNAs derived just downstream from transcription initiation sites (tiRNAs) [255,312]. Their size and position suggest that tiRNAs may be produced by RNA polymerase II interaction with the first downstream nucleosome and backtracking [312,313], a well established phenomenon that involves TFIIS-mediated cleavage of a small RNA from the 3′ end of the nascent transcript prior to resumption of synthesis [314]. This possibility was given strong support by the finding that the position of tiRNAs is different in human and Drosophila, reflecting the different position of the first nucleosome downstream of the transcription start site [313] (Fig. 5), with the implication that tiRNAs may be marking, and perhaps regulating, the position of the nucleosome [313].

While it has been known for some time that there is a nucleosome-free region around transcription start sites, with periodicity elsewhere in nucleosome spacing [315–319], it came as a significant surprise to discover recently that nucleosomes are in fact preferentially positioned at exons in somatic cells [320–324] and germ cells [323], the latter implying trans-generational epigenetic inheritance, for which there is increasing evidence [325–327] and which may provide selective advantage to cope with environmental challenges [328]. Presumably there is also some logic to the positioning of nucleosomes in intronic and intergenic regions, which also produce many functional (non-coding) RNAs, although this has yet to be established. In any case, this finding shows that chromatin is far more organized than previously suspected, and that the position of nucleosomes must be regulated by some active or passive mechanism(s). It also provides an explanation for the observed coupling of chromatin structure, transcription and splicing [329], and a potential basis for exon selection through various histone modifications within these nucleosomes that report the status of particular exons during differentiation and development [323] – i.e., that alternative splicing may be controlled by histone modifications [320–324], a prediction that has since been confirmed, at least in part [330,331].

Subsequent deep sequencing of nuclear small RNAs identified another class of small RNAs positioned at splice sites (spliRNAs) [255] (Fig. 6). These spliRNAs are similar in size (17–18 nt) to tiRNAs and, like tiRNAs, are only found in animals. This suggests that they may have a similar ontogeny and function, but spliRNAs are derived from the 3′ end of exons [255], whereas tiRNAs are positioned on the 5′ side (upstream) of the first nucleosome [313]. However, there is (at least) one form of modified histone (H3K36me3) that has a different position, which is associated with actively expressed genes [332] and which appears to be positioned at the exon–intron boundary [333], or just downstream of it [323]. This would (in principle) place spliRNAs on the 5′ side of these nucleosomes, like tiRNAs, a possibility that we are currently testing by deep sequencing on small RNAs associated with different forms of modified nucleosomes. It is worth noting that individual tiRNAs and spliRNAs are only present at very low levels in deep sequencing datasets [255,312]: neither would have been identified without the orthogonal intersection of small RNA datasets with specific genomic features (transcription start sites and exon–intron junctions, respectively), which also suggests that there may be many more locus-specific regulatory RNAs to be discovered.

9. Transcriptomic, epigenomic and genomic plasticity in gene-environment interactions and brain function

RNA also appears to be the substrate for environmental–epigenome interactions. There is emerging evidence that RNA is subject to a great deal of context-dependent editing, especially in the brain. RNA editing involves base deamination (as distinct from snoRNA-mediated modification) and is catalyzed by two classes of enzymes in animals: ADARs (adenosine deaminases that act on RNA) change adenosine to inosine (A > I) [334,335], which behaves similarly to guanosine (e.g., in sequencing protocols), but has different base pairing qualities; and APOBECs (named after 'ApoB editing complex', see below), which are vertebrate-specific and change cytosine to uracil (C > U) and may act on RNA or DNA [336,337].

There are three ADAR orthologs in animals. ADAR1 and ADAR2 occur in invertebrates and vertebrates [338], and are expressed in most tissues, but particularly highly expressed in the nervous system [334,335,339]. Loss of these genes in mice is embryonic or postnatally lethal [340,341]. ADAR3 is vertebrate-and brain-specific [338,342], but little is known about its function. Little is known about how RNA editing is regulated [343], but ADARs can be localized in both the nucleus and cytoplasm [344,345], and there is evidence that RNA editing activity is connected to canonical cell signaling pathways [346], implying response to external cues.

A > I editing was first discovered a generation ago by cDNA-genomic comparisons of sequences encoding important neuroreceptors, such as glutamate, GABA and serotonin receptors, where it alters the amino acid sequence, ostensibly to tune the electrophysiological properties of the synapse [334]. It has since been regarded as an interesting but somewhat idiosyncratic subfield of both molecular biology and neuroscience. However four papers published in 2004, which have attracted surprisingly little attention, showed by comparison of large scale cDNA libraries with genomic DNA that A > I editing is far more widespread than previously suspected, and occurs in thousands of transcripts [347–350]. Most of the edited sites occur in non-coding regions, implying that editing is not only modifying the structure-function properties of proteins, but also RNA-based regulatory circuits, and therefore potentially epigenetic processes, which are central to learning and brain function [173–177,179,351].

Moreover, these studies showed that there is a massive increase (~35×) in the intensity of RNA editing in humans compared to mouse. Most (>90%) of this editing occurs in primate-specific Alu sequences [347–350], which evolved from a functional RNA ancestor (the 7SL RNA of the signal recognition particle) [352,353]. Alu sequences invaded the primate lineage in three successive waves, and now comprise ~1.2 million mostly sequence unique copies that collectively occupy ~10.5% of our genome [354,355]. A subsequent study showed that the intensity of A > I editing also increased during primate evolution, and that new editable Alu insertions after the human-chimpanzee split are significantly enriched in genes related to neurological functions and neurological diseases [356]. These SINEs (short interspersed nuclear elements), consistent with the general view of such elements as junk, have long been regarded as the most recent transpositional storm to hit our lineage. However, these observations suggest a radically different and much more interesting interpretation – i.e., that such sequences, while having being recruited for many functions [353],
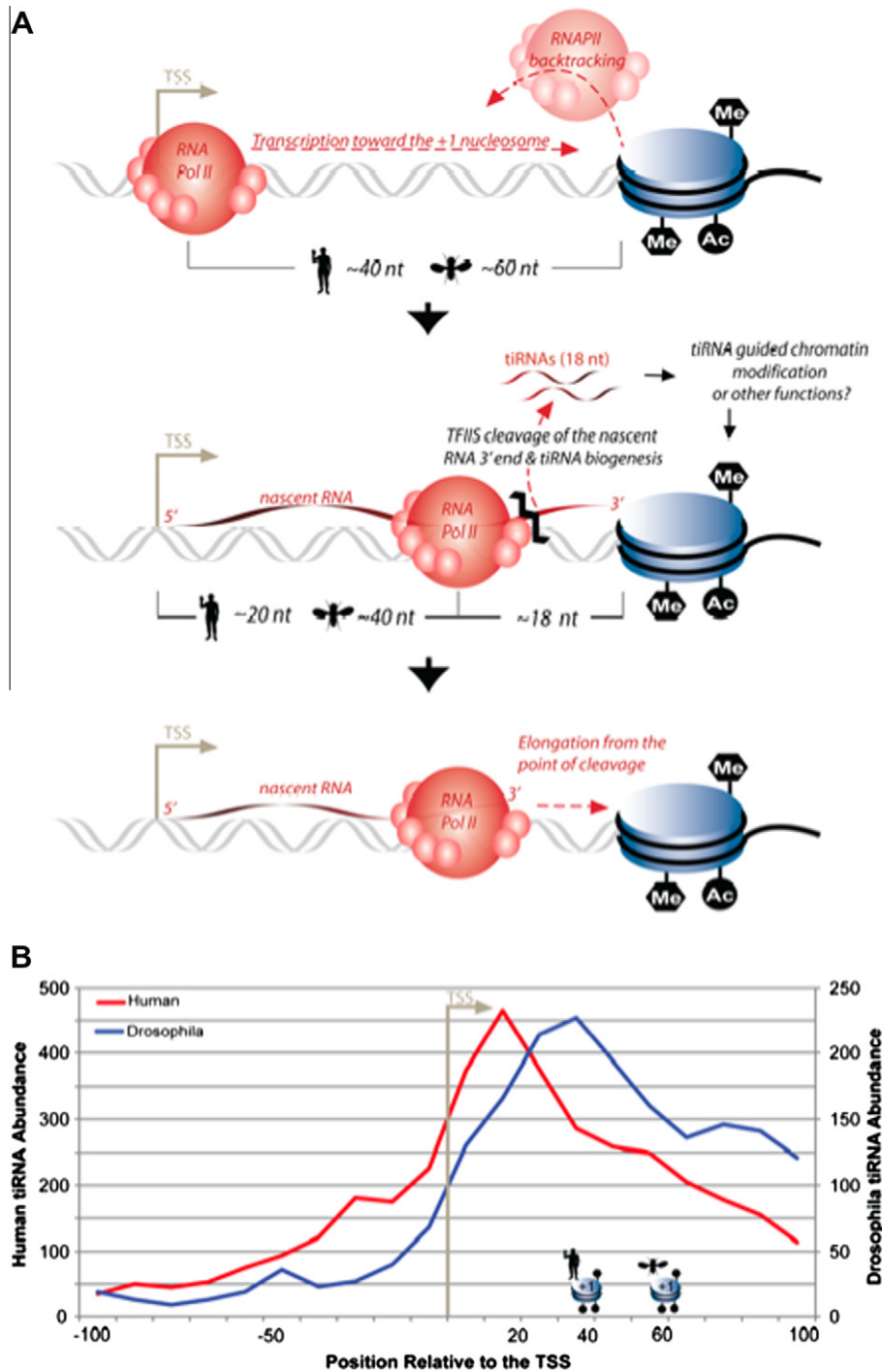
Fig. 5. A model of tiRNA biogenesis, and evidence that tiRNAs are linked to the position of the first nucleosome downstream of the transcription start site (TSS). (A) RNAPII generates a short nascent RNA, stalls at the first (+1) nucleosome and backtracks, and the 3′ end of the nascent RNA is cleaved by TFIIS to generate an ~18 nt tiRNA. Values next to human or fly silhouettes show the average distances from the TSS to the nucleosome or the putative position of tiRNA biogenesis. (B) Peak densities of the 5′ ends of tiRNAs in human and Drosophila are offset, consistent with positions of phased +1 nucleosomes, which are different between the two species, and the proposed model of biogenesis. Reproduced from [313] with permission.

have flourished as modular substrates for RNA editing, permitting the introduction and spread of the transcriptomic and epigenomic plasticity necessary for epigenome–environment interactions, driven by positive selection for cognitive function [185,328,356,357].

The APOBECs are even more intriguing. They were discovered initially by their action on ApolipoproteinB mRNA where a C > U change introduces a stop codon to generate a truncated isoform of the protein in intestine versus the longer form produced in liver

[336]. There are five families of APOBECs, two of which (APOBEC 1 and 3) are mammal-specific [336,337]. The best characterized is AID, which is involved in somatic rearrangements and hypermutation of immunoglobulins in the immune system [337]. AID appears to act on DNA but may be targeted by RNA [358]. Moreover, AID deaminates 5′ methylcytosine (to form thymine) [359] and is required for the reprogramming of cells to pluripotency [360], and APOBEC2 is required for muscle differentiation [361], suggesting a wider role
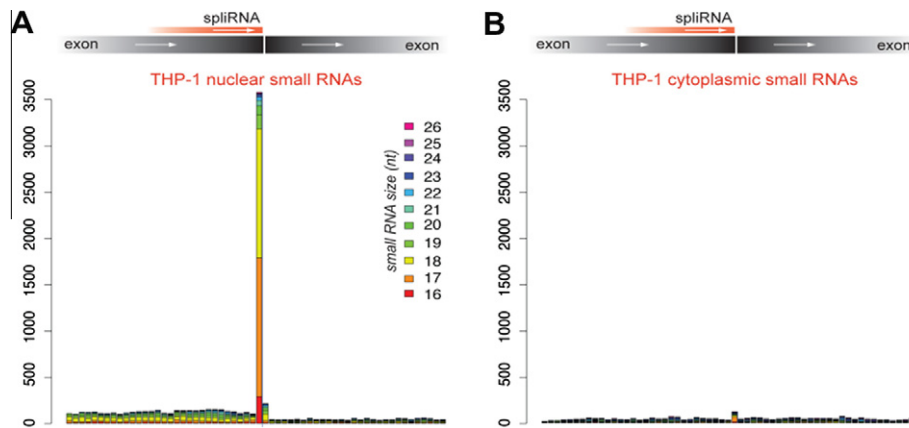
Fig. 6. Small RNAs are associated with splice sites (spliRNAs). The position of small RNA 3 ends is plotted with respect to the splice donor site (the 3 end of the exon). The schematics at top depict the position of spliRNAs detected in human THP-1 cells and their strand orientation with respect to exon–exon junctions. Small RNAs, dominantly ~17 or 18 nt, are >35-fold enriched at the 5 splice site in nuclei (A) compared to the background or (B) cytoplasmic small RNAs. Adapted from [255] with permission.
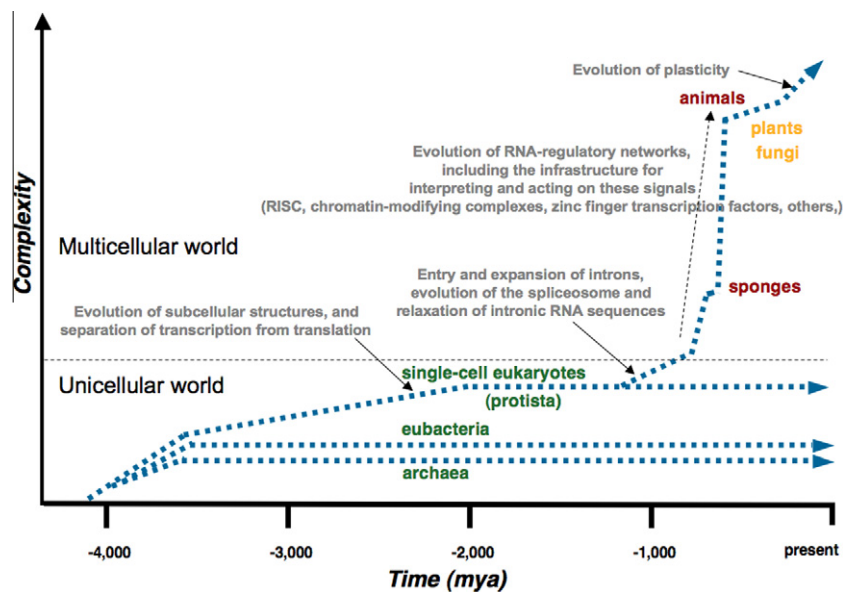

Fig. 7. A simplified biological history of the earth indicating the proposed major events and transitions on the evolutionary path from simple eukaryotes to developmentally complex and cognitively advanced organisms.

for such enzymes in developmental processes. Interestingly, there are many parallels between the nervous and adaptive immune systems, including the presence of immunoglobulin domains in many neuronal cell surface receptors [362,363], indicating that the adaptive immune system evolved (in vertebrates) from the nervous system, and that both may use similar mechanisms to tune receptor interactions. Moreover, the existence of many unusual DNA repair enzymes, many of which appear to be linked to reverse transcriptase activity, suggests that RNA-directed DNA recoding may play a role in long-term memory formation [357].

The APOBEC3 family originated after the divergence of the marsupial and placental lineages and has greatly expanded in the primate lineage, with very strong signatures of positive selection [337,364,365]. At least some (as well as APOBEC1 [366]) appear to be involved in the control of exogenous and endogenous retrotransposition, possibly by inhibition of reverse transcription, and are therefore thought to be involved in host/genome defense [337,367–370], but why this should be particularly important in mammals and especially primates is problematic. An alternative possibility is that these enzymes (one of which, APOBEC3G) is ex-

pressed in neurons [371]) have evolved to domesticate transposition. This possibility has been given strong support by the recent observations that de novo L1 retrotransposition events occur in neural progenitor cells and may therefore contribute to individual somatic mosaicism in the brain [372,373]. Moreover, this process that appears to be regulated by Wnt signaling pathways and transcription factors known to be important in neural differentiation [374]. The process is also regulated by MeCP2 (methyl-CpG-binding protein 2), which is involved in global DNA methylation and neurodevelopmental diseases [375], and is modulated in the brain itself by environmental influences [376]. That is, transposon mobilization may not simply have played a role in genome evolution but also in real time genome dynamics that enable the extraordinary in situ evolution [377] and functional complexity of the neuronal networks in the human brain [378].

10. Concluding remarks

The emerging evidence suggests that evolution has shaped the human genome in far more sophisticated ways than ever imagined,

and that most of the information it holds is involved in complex regulatory processes that underpin development and brain function. This includes the vast numbers of non-coding RNAs and transposons, which rather than being junk, appear to provide the regulatory power and plasticity required to program our ontogeny and cognition [14,69,184,185,194,379–381]. In this scenario, DNA might be viewed as a zip file/hard disc, proteins (including most 'regulatory' proteins) as the analog effectors, and RNA as the computational engine of the system [189,382]. Moreover, it seems that the major challenge that evolution had to overcome to evolve developmentally complex organisms was regulatory, and that the barriers imposed by the rising cost of regulation were overcome by moving to a hierarchical RNA-based regulatory system.

It seems plausible that the deep roots of this system lay in the compartmentalization of eukaryotic cells and the consequent separation of transcription from translation. This enabled group II self-splicing introns (the precursors of modern nuclear introns) to invade genes and, following the evolution of the spliceosome, to evolve to express regulatory RNAs along with a receptive protein infrastructure, which then allowed the progressive attainment of higher developmental complexity and the exploration of design variations [14,69,381], which exploded in the Cambrian, where recognizable ancestors of most modern phyla are observable in the fossil record [383] (Fig. 7). The subsequent competition within and between these dynasties to refine their body plans and colonize new niches, including the land and the air, resulted in the wonderful diversity of multicellular life on this planet, and is usually thought of as the exemplar of Darwinian evolution. However, perhaps the real story and the last great mountain that evolution climbed was the superimposition of environmentally responsive plasticity on this system, including the cooption and domestication of retrotransposons, to allow the development of higher order information processing, learning, language and thought, the most powerful selective advantage of all.

Acknowledgements

References

[1] Crick, F. (1970) Central dogma of molecular biology. Nature 227, 561–563.
[2] Quoted from: Comfort, N.C. (2003) The Tangled Field: Barbara McClintock's Search for the Patterns of Genetic Control, Harvard University Press, Cambridge, Massachussetts.
[3] Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. 3, 318–356.
[4] Gilbert, W. and Muller-Hill, B. (1966) Isolation of the lac repressor. Proc. Natl. Acad. Sci. USA 56, 1891–1898.
[5] Britten, R.J. and Davidson, E.H. (1969) Gene regulation for higher cells: a theory. Science 165, 349–357.
[6] Britten, R.J. and Davidson, E.H. (1971) Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. Q. Rev. Biol. 66, 111–138.
[7] Chow, L.T., Gelinas, R.E., Broker, T.R. and Roberts, R.J. (1977) An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. Cell 12, 1–8.
[8] Berget, S.M., Moore, C. and Sharp, P.A. (1977) Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. Proc. Natl. Acad. Sci. USA 74, 3171–3175.
[9] Williamson, B. (1977) DNA insertions and gene structure. Nature 270, 295–297.
[10] Gilbert, W. (1978) Why genes in pieces? Nature 271, 501.
[11] Crick, F. (1979) Split genes and RNA splicing. Science 204, 264–271.
[12] Doolittle, W.F. (1978) Genes in pieces: were they ever together? Nature 272, 581–582.
[13] Gilbert, W., Marchionni, M. and McKnight, G. (1986) On the antiquity of introns. Cell 46, 151–154.
[14] Mattick, J.S. (1994) Introns: evolution and function. Curr. Opin. Genet. Dev. 4, 823–831.
[15] Doolittle, W.F. and Sapienza, C. (1980) Selfish genes, the phenotype paradigm and genome evolution. Nature 284, 601–603.
[16] Orgel, L.E. and Crick, F.H. (1980) Selfish DNA: the ultimate parasite. Nature 284, 604–607.
[17] Brosius, J. (2003) The contribution of RNAs and retroposition to evolutionary novelties. Genetica 118, 99–116.
[18] Pheasant, M. and Mattick, J.S. (2007) Raising the estimate of functional human sequences. Genome Res. 17, 1245–1253.
[19] Waterston, R.H. et al. (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420, 520–562.
[20] Lindblad-Toh, K. et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438, 803–819.
[21] Asthana, S., Noble, W.S., Kryukov, G., Grant, C.E., Sunyaev, S. and Stamatoyannopoulos, J.A. (2007) Widely distributed noncoding purifying selection in the human genome. Proc. Natl. Acad. Sci. USA 104, 12410–12415.
[22] Taft, R.J., Pheasant, M. and Mattick, J.S. (2007) The relationship between non-protein-coding DNA and eukaryotic complexity. Bioessays 29, 288–299.
[23] Stein, L.D. et al. (2003) The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics. PLoS Biol. 1, E45.
[24] Goodstadt, L. and Ponting, C.P. (2006) Phylogenetic reconstruction of orthology, paralogy, and conserved synteny for dog and human. PLoS Comput. Biol. 2, e133.
[25] Clamp, M. et al. (2007) Distinguishing protein-coding and noncoding genes in the human genome. Proc. Natl. Acad. Sci. USA 104, 19428–19433.
[26] Azevedo, F.A. et al. (2009) Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. J. Comp. Neurol. 513, 532–541.
[27] Srivastava, M. et al. (2010) The Amphimedon queenslandica genome and the evolution of animal complexity. Nature 466, 720–726.
[28] Rokas, A. (2008) The origins of multicellularity and the early history of the genetic toolkit for animal development. Annu. Rev. Genet. 42, 235–251.
[29] Levine, M. and Tjian, R. (2003) Transcription regulation and animal diversity. Nature 424, 147–151.
[30] Carroll, S.B. (2008) Evo–devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell 134, 25–36.
[31] Croft, L.J., Lercher, M.J., Gagen, M.J. and Mattick, J.S. (2003) Is prokaryotic complexity limited by accelerated growth in regulatory overhead? Genome Biology Preprint Repository <http://genomebiology.com/qc/2003/5/1/p2>.
[32] Molina, N. and van Nimwegen, E. (2008) Universal patterns of purifying selection at noncoding positions in bacteria. Genome Res. 18, 148–160.
[33] Mattick, J.S. and Gagen, M.J. (2005) Accelerating networks. Science 307, 856–858.
[34] Tang, G. (2005) SiRNA and miRNA: an insight into RISCs. Trends Biochem. Sci. 30, 106–114.
[35] Ghildiyal, M. and Zamore, P.D. (2009) Small silencing RNAs: an expanding universe. Nat. Rev. Genet. 10, 94–108.
[36] Taft, R.J. and Mattick, J.S. (2003) Increasing biological complexity is positively correlated with the relative genome-wide expansion of non-protein-coding DNA sequences. Genome Biol. Preprint Repository <http://genomebiology.com/2003/5/1/P1>.
[37] Okazaki, Y. et al. (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature 420, 563–573.
[38] Numata, K. et al. (2003) Identification of putative noncoding RNAs among the RIKEN mouse full-length cDNA collection. Genome Res. 13, 1301–1306.
[39] Imanishi, T. et al. (2004) Integrative annotation of 21,037 human genes validated by full-length cDNA clones. PLoS Biol. 2, 856–875.
[40] Ota, T. et al. (2004) Complete sequencing and characterization of 21,243 full-length human cDNAs. Nat. Genet. 36, 40–45.
[41] Carninci, P. et al. (2005) The transcriptional landscape of the mammalian genome. Science 309, 1559–1563.
[42] Katayama, S. et al. (2005) Antisense transcription in the mammalian transcriptome. Science 309, 1564–1566.
[43] Kapranov, P., Cawley, S.E., Drenkow, J., Bekiranov, S., Strausberg, R.L., Fodor, S.P. and Gingeras, T.R. (2002) Large-scale transcriptional activity in chromosomes 21 and 22. Science 296, 916–919.
[44] Kampa, D. et al. (2004) Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. Genome Res. 14, 331–342.
[45] Bertone, P. et al. (2004) Global identification of human transcribed sequences with genome tiling arrays. Science 306, 2242–2246.
[46] Cheng, J. et al. (2005) Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science 308, 1149–1154.
[47] Kapranov, P., Drenkow, J., Cheng, J., Long, J., Helt, G., Dike, S. and Gingeras, T.R. (2005) Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. Genome Res. 15, 987–997.
[48] Birney, E. et al. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447, 799–816.
[49] Frith, M.C., Pheasant, M. and Mattick, J.S. (2005) The amazing complexity of the human transcriptome. Eur. J. Hum. Genet. 13, 894–897.

[50] Mattick, J.S. and Makunin, I.V. (2006) Non-coding RNA. Hum. Mol. Genet. 15, R17–R129.

[51] Kapranov, P., Willingham, A.T. and Gingeras, T.R. (2007) Genome-wide transcription and the implications for genomic organization. Nat. Rev. Genet. 8, 413–423.

[52] Ponting, C.P., Oliver, P.L. and Reik, W. (2009) Evolution and functions of long noncoding RNAs. Cell 136, 629–641.

[53] Mattick, J.S. (2009) The genetic signatures of noncoding RNAs. PLoS Genet. 5, e1000459.

[54] van Bakel, H., Nislow, C., Blencowe, B.J. and Hughes, T.R. (2010) Most "dark matter" transcripts are associated with known genes. PLoS Biol. 8, e1000371.

[55] Clark, M.B., Amaral, P.P., Schlesinger, F.J., Dinger, M.E., Taft, R.J., Rinn, J.L., Ponting, C.P., Stadler, P.F., Morris, K.V., Morillon, A., Rozowsky, J.S., Gerstein, M.B., Wahlestedt, C., Hayashizaki, Y., Carninci, P., Gingeras, T.R. and Mattick, J.S. (in press) The reality of pervasive transcription. PLoS Biol.

[56] Mercer, T.R., Dinger, M.E., Sunkin, S.M., Mehler, M.F. and Mattick, J.S. (2008) Specific expression of long noncoding RNAs in the mouse brain. Proc. Natl. Acad. Sci. USA 105, 716–721.

[57] Mercer, T.R., Gerhardt, D.J., Dinger, M.E., Crawford, J., Trapnell, C., Jeddeloh, J.A., Mattick, J.S. and Rinn, J.L. (submitted for publication) RNA Capture-Seq resolves the deep complexity of the human transcriptome.

[58] Amaral, P.P. and Mattick, J.S. (2008) Noncoding RNA in development. Mamm. Genome 19, 454–492.

[59] Ravasi, T. et al. (2006) Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. Genome Res. 16, 11–19.

[60] Mehler, M.F. and Mattick, J.S. (2006) Non-coding RNAs in the nervous system. J. Physiol. 575, 333–341.

[61] Mehler, M.F. and Mattick, J.S. (2007) Noncoding RNAs and RNA editing in brain development, functional diversification, and neurological disease. Physiol. Rev. 87, 799–823.

[62] Sanchez-Herrero, E. and Akam, M. (1989) Spatially ordered transcription of regulatory DNA in the bithorax complex of Drosophila. Development 107, 321–329.

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

[64] Drewell, R.A., Bae, E., Burr, J. and Lewis, E.B. (2002) Transcription defines the embryonic domains of cis-regulatory activity at the Drosophila bithorax complex. Proc. Natl. Acad. Sci. USA 99, 16853–16858.

[65] Kim, T.K. et al. (2010) Widespread transcription at neuronal activity-regulated enhancers. Nature 465, 182–187.

[66] Levine, M. (2010) Transcriptional enhancers in animal development and evolution. Curr. Biol. 20, R754–R763.

[67] Ørom, U.A. et al. (2010) Long noncoding RNAs with enhancer-like function in human cells. Cell 143, 46–58.

[68] Mattick, J.S. (2010) Linc-ing long noncoding RNAs and enhancer function. Dev. Cell 19, 485–486.

[69] Mattick, J.S. and Gagen, M.J. (2001) The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. Mol. Biol. Evol. 18, 1611–1630.

[70] Koziol, M.J. and Rinn, J.L. (2010) RNA traffic control of chromatin complexes. Curr. Opin. Genet. Dev. 20, 142–148.

[71] Chamary, J.V., Parmley, J.L. and Hurst, L.D. (2006) Hearing silence: non-neutral evolution at synonymous sites in mammals. Nat. Rev. Genet. 7, 98–108.

[72] Fejes-Toth, K. et al. (2009) Post-transcriptional processing generates a diversity of 5′-modified long and short RNAs. Nature 457, 1028–1032.

[73] Mercer, T.R. et al. (2010) Regulated post-transcriptional RNA cleavage diversifies the eukaryotic transcriptome. Genome Res. 20, 1639–1650.

[74] Carninci, P. et al. (2006) Genome-wide analysis of mammalian promoter architecture and evolution. Nat. Genet. 38, 626–635.

[75] Mercer, T.R. et al. (2011) Expression of distinct RNAs from 3′ untranslated regions. Nucleic Acids Res. 39, 2393–2403.

[76] Jenny, A., Hachet, O., Zavorszky, P., Cyrklaff, A., Weston, M.D., Johnston, D.S., Erdelyi, M. and Ephrussi, A. (2006) A translation-independent role of oskar RNA in early Drosophila oogenesis. Development 133, 2827–2833.

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

[78] Rastinejad, F., Conboy, M.J., Rando, T.A. and Blau, H.M. (1993) Tumor suppression by RNA from the 3′ untranslated region of alpha-tropomyosin. Cell 75, 1107–1117.

[79] Fan, H., Villegas, C., Huang, A. and Wright, J.A. (1996) Suppression of malignancy by the 3′ untranslated regions of ribonucleotide reductase R1 and R2 messenger RNAs. Cancer Res. 56, 4366–4369.

[80] Jupe, E.R., Liu, X.T., Kiehlbauch, J.L., McClung, J.K. and Dell'Orco, R.T. (1996) Prohibitin in breast cancer cell lines: loss of antiproliferative activity is linked to 3′ untranslated region mutations. Cell Growth Differ. 7, 871–878.

[81] Amack, J.D., Paguio, A.P. and Mahadevan, M.S. (1999) Cis and trans effects of the myotonic dystrophy (DM) mutation in a cell culture model. Hum. Mol. Genet. 8, 1975–1984.

[82] Dieci, G., Fiorino, G., Castelnuovo, M., Teichmann, M. and Pagano, A. (2007) The expanding RNA polymerase III transcriptome. Trends Genet. 23, 614–622.

[83] Lunyak, V.V. et al. (2007) Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. Science 317, 248–251.

[84] Mariner, P.D., Walters, R.D., Espinoza, C.A., Drullinger, L.F., Wagner, S.D., Kugel, J.F. and Goodrich, J.A. (2008) Human Alu RNA Is a modular transacting repressor of mRNA transcription during heat shock. Mol. Cell 29, 499–509.

[85] Chueh, A.C., Northrop, E.L., Brettingham-Moore, K.H., Choo, K.H. and Wong, L.H. (2009) LINE retrotransposon RNA is an essential structural and functional epigenetic component of a core neocentromeric chromatin. PLoS Genet. 5, e1000354.

[86] Faulkner, G.J. et al. (2009) The regulated retrotransposon transcriptome of mammalian cells. Nat. Genet. 41, 563–571.

[87] Denis, H. (1966) Gene expression in amphibian development II. Release of the genetic information in growing embryos. J. Mol. Biol. 22, 285–304.

[88] Davidson, E.H., Crippa, M. and Mirsky, A.E. (1968) Evidence for the appearance of novel gene products during amphibian blastulation. Proc. Natl. Acad. Sci. USA 60, 152–159.

[89] Pang, K.C., Frith, M.C. and Mattick, J.S. (2006) Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. Trends Genet. 22, 1–5.

[90] Ponjavic, J., Ponting, C.P. and Lunter, G. (2007) Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. Genome Res. 17, 556–565.

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

[92] Torarinsson, E., Sawera, M., Havgaard, J.H., Fredholm, M. and Gorodkin, J. (2006) Thousands of corresponding human and mouse genomic regions unalignable in primary sequence contain common RNA structure. Genome Res. 16, 885–889.

[93] Torarinsson, E. et al. (2008) Comparative genomics beyond sequence-based alignments: RNA structures in the ENCODE regions. Genome Res. 18, 242–251.

[94] Dinger, M.E. et al. (2008) Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Res. 18, 1433–1445.

[95] Guttman, M. et al. (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 458, 223–227.

[96] Trinklein, N.D., Aldred, S.F., Hartman, S.J., Schroeder, D.I., Otillar, R.P. and Myers, R.M. (2004) An abundance of bidirectional promoters in the human genome. Genome Res. 14, 62–66.

[97] Engstrom, P.G. et al. (2006) Complex loci in human and mouse genomes. PLoS Genet. 2, e47.

[98] Tupy, J.L., Bailey, A.M., Dailey, G., Evans-Holm, M., Siebel, C.W., Misra, S., Celniker, S.E. and Rubin, G.M. (2005) Identification of putative noncoding polyadenylated transcripts in Drosophila melanogaster. Proc. Natl. Acad. Sci. USA 102, 5495–5500.

[99] Louro, R., El-Jundi, T., Nakaya, H.I., Reis, E.M. and Verjovski-Almeida, S. (2008) Conserved tissue expression signatures of intronic noncoding RNAs transcribed from human and mouse loci. Genomics 92, 18–25.

[100] Amaral, P.P., Neyt, C., Wilkins, S.J., Askarian-Amiri, M.E., Sunkin, S.M., Perkins, A.C. and Mattick, J.S. (2009) Complex architecture and regulated expression of the Sox2ot locus during vertebrate development. RNA 15, 2013–2027.

[101] Rinn, J.L. et al. (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129, 1311–1323.

[102] Sunwoo, H., Dinger, M.E., Wilusz, J.E., Amaral, P.P., Mattick, J.S. and Spector, D.L. (2009) 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.

[103] Askarian-Amiri, M.E. et al. (2011) SNORD-host RNA Zfas1 is a regulator of mammary development and a potential marker for breast cancer. RNA 17, 878–891.

[104] Thrash-Bingham, C.A. and Tartof, K.D. (1999) AHIF: a natural antisense transcript overexpressed in human renal cancer and during hypoxia. J. Natl. Cancer Inst. 91, 143–151.

[105] Ji, P. et al. (2003) MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. Oncogene 22, 6087–6097.

[106] Mutsuddi, M., Marshall, C.M., Benzow, K.A., Koob, M.D. and Rebay, I. (2004) The spinocerebellar ataxia 8 noncoding RNA causes neurodegeneration and associates with staufen in Drosophila. Curr. Biol. 14, 302–308.

[107] Reis, E.M. et al. (2004) Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer. Oncogene 23, 6684–6692.

[108] Reis, E.M. et al. (2005) Large-scale transcriptome analyses reveal new genetic marker candidates of head, neck, and thyroid cancer. Cancer Res. 65, 1693–1699.

[109] Sonkoly, E. et al. (2005) Identification and characterization of a novel, psoriasis susceptibility-related noncoding RNA gene, PRINS. J. Biol. Chem. 280, 24159–24167.

[110] Szymanski, M., Barciszewska, M.Z., Erdmann, V.A. and Barciszewski, J. (2005) A new frontier for molecular medicine: noncoding RNAs. Biochim. Biophys. Acta 1756, 65–75.

[111] Angeloni, D. et al. (2006) Analysis of a new homozygous deletion in the tumor suppressor region at 3p12.3 reveals two novel intronic noncoding RNA genes. Genes Chromosomes Cancer 45, 676–691.

[112] Calin, G.A. et al. (2007) Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. Cancer Cell 12, 215–229.

[113] Pang, K.C., Stephen, S., Dinger, M.E., Engstrom, P.G., Lenhard, B. and Mattick, J.S. (2007) RNAdb 2.0 – an expanded database of mammalian non-coding RNAs. Nucleic Acids Res. 35, D178–D182.

[114] Christov, C.P., Trivier, E. and Krude, T. (2008) Noncoding human Y RNAs are overexpressed in tumours and required for cell proliferation. Br. J. Cancer 98, 981–988.

[115] Perez, D.S., Hoage, T.R., Pritchett, J.R., Ducharme-Smith, A.L., Halling, M.L., Ganapathiraju, S.C., Streng, P.S. and Smith, D.I. (2008) Long, abundantly expressed non-coding transcripts are altered in cancer. Hum. Mol. Genet. 17, 642–655.

[116] Zhang, X. et al. (2009) A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. Blood 113, 2526–2534.

[117] Khaitan, D., Dinger, M.E., Mazar, J., Crawford, J., Smith, M.A., Mattick, J.S. and Perera, R.J. (in press) The melanoma-upregulated long noncoding RNA SPRY4-IN1 modulates apoptosis and invasion. Cancer Res.

[118] Cawley, S. et al. (2004) Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell 116, 499–509.

[119] Koshimizu, T.A., Fujiwara, Y., Sakai, N., Shibata, K. and Tsuchiya, H. (2010) Oxytocin stimulates expression of a noncoding RNA tumor marker in a human neuroblastoma cell line. Life Sci. 86, 455–460.

[120] Blin-Wakkach, C. et al. (2001) Endogenous Msx1 antisense transcript: in vivo and in vitro evidences, structure, and potential involvement in skeleton development in mammals. Proc. Natl. Acad. Sci. USA 98, 7336–7341.

[121] Kohtz, J.D. and Fishell, G. (2004) Developmental regulation of EVF-1, a novel non-coding RNA transcribed upstream of the mouse Dlx6 gene. Gene Expr. Patterns 4, 407–412.

[122] Inagaki, S., Numata, K., Kondo, T., Tomita, M., Yasuda, K., Kanai, A. and Kageyama, Y. (2005) Identification and expression analysis of putative mRNA-like non-coding RNA in Drosophila. Genes Cells 10, 1163–1173.

[123] Young, T.L., Matsuda, T. and Cepko, C.L. (2005) The noncoding RNA taurine upregulated gene 1 is required for differentiation of the murine retina. Curr. Biol. 15, 501–512.

[124] Brena, C., Chipman, A.D., Minelli, A. and Akam, M. (2006) Expression of trunk Hox genes in the centipede Strigamia maritima: sense and anti-sense transcripts. Evol. Dev. 8, 252–265.

[125] Sone, M., Hayashi, T., Tarui, H., Agata, K., Takeichi, M. and Nakagawa, S. (2007) The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons. J. Cell Sci. 120, 2498–2506.

[126] Sasaki, Y.T., Ideue, T., Sano, M., Mituyama, T. and Hirose, T. (2009) MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles. Proc. Natl. Acad. Sci. USA 106, 2525–2530.

[127] Clemson, C.M., Hutchinson, J.N., Sara, S.A., Ensminger, A.W., Fox, A.H., Chess, A. and Lawrence, J.B. (2009) An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Mol. Cell 33, 717–726.

[128] Chen, L.L. and Carmichael, G.G. (2009) Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. Mol. Cell 35, 467–478.

[129] Bond, C.S. and Fox, A.H. (2009) Paraspeckles: nuclear bodies built on long noncoding RNA. J. Cell Biol. 186, 637–644.

[130] Redrup, L. et al. (2009) The long noncoding RNA Kcnq1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing. Development 136, 525–530.

[131] Mercer, T.R., Qureshi, I.A., Gokhan, S., Dinger, M.E., Li, G., Mattick, J.S. and Mehler, M.F. (2010) Long noncoding RNAs in neuronal-glial fate specification and oligodendrocyte lineage maturation. BMC Neurosci. 11, 14.

[132] Pang, K.C., Dinger, M.E., Mercer, T.R., Malquori, L., Grimmond, S.M., Chen, W. and Mattick, J.S. (2009) Genome-wide identification of long noncoding RNAs in CD8+ T cells. J. Immunol. 182, 7738–7748.

[133] Clark, M.B. and Mattick, J.S. (2011) Long noncoding RNAs in cell biology. Semin. Cell Dev. Biol. epub ahead of print, doi:10.1016/j.semcdb.2011.01.001.

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

[135] Souquere, S., Beauclair, G., Harper, F., Fox, A. and Pierron, G. (2010) Highly ordered spatial organization of the structural long noncoding NEAT1 RNAs within paraspeckle nuclear bodies. Mol. Biol. Cell 21, 4020–4027.

[136] Mao, Y.S., Sunwoo, H., Zhang, B. and Spector, D.L. (2011) Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. Nat. Cell Biol. 13, 95–101.

[137] Nakagawa, S., Naganuma, T., Shioi, G. and Hirose, T. (2011) Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice. J. Cell Biol. 193, 31–39.

[138] Rivera, M.N., Kim, W.J., Wells, J., Stone, A., Burger, A., Coffman, E.J., Zhang, J. and Haber, D.A. (2009) The tumor suppressor WTX shuttles to the nucleus and modulates WT1 activity. Proc. Natl. Acad. Sci. USA 106, 8338–8343.

[139] Mercer, T.R., Dinger, M.E. and Mattick, J.S. (2009) Long non-coding RNAs: insights into functions. Nat. Rev. Genet. 10, 155–159.

[140] Amaral, P.P., Clark, M.B., Gascoigne, D.K., Dinger, M.E. and Mattick, J.S. (2011) LncRNAdb: a reference database for long noncoding RNAs. Nucleic Acids Res. 39, D146–D151.

[141] Loewer, S. et al. (2010) Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nat. Genet. 42, 1113–1117.

[142] Gupta, R.A. et al. (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 464, 1071–1076.

[143] Ginger, M.R., Shore, A.N., Contreras, A., Rijnkels, M., Miller, J., Gonzalez-Rimbau, M.F. and Rosen, J.M. (2006) A noncoding RNA is a potential marker of cell fate during mammary gland development. Proc. Natl. Acad. Sci. USA 103, 5781–5786.

[144] Rapicavoli, N.A. and Blackshaw, S. (2009) New meaning in the message: noncoding RNAs and their role in retinal development. Dev. Dyn. 238, 2103–2114.

[145] Rapicavoli, N.A., Poth, E.M. and Blackshaw, S. (2010) The long noncoding RNA RNCR2 directs mouse retinal cell specification. BMC Dev. Biol. 10, 49.

[146] Sleutels, F., Zwart, R. and Barlow, D.P. (2002) The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature 415, 810–813.

[147] Thakur, N. et al. (2004) An antisense RNA regulates the bidirectional silencing property of the Kcnq1 imprinting control region. Mol. Cell. Biol. 24, 7855–7862.

[148] Schoenfelder, S., Smits, G., Fraser, P., Reik, W. and Paro, R. (2007) Non-coding transcripts in the H19 imprinting control region mediate gene silencing in transgenic Drosophila. EMBO Rep. 8, 1068–1073.

[149] Gabory, A., Jammes, H. and Dandolo, L. (2010) The H19 locus: role of an imprinted non-coding RNA in growth and development. Bioessays 32, 473–480.

[150] Santoro, F. and Barlow, D.P. (2011) Developmental control of imprinted expression by macro non-coding RNAs. Semin. Cell Dev. Biol. epub ahead of print, doi:10.1016/j.semcdb.2011.02.018.

[151] Kanduri, C. (2011) Kcnq1ot1: a chromatin regulatory RNA. Semin. Cell Dev. Biol. epub ahead of print, doi:10.1016/j.semcdb.2011.02.020.

[152] Kelley, R.L. and Kuroda, M.I. (2000) Noncoding RNA genes in dosage compensation and imprinting. Cell 103, 9–12.

[153] Senner, C.E. and Brockdorff, N. (2009) Xist gene regulation at the onset of X inactivation. Curr. Opin. Genet. Dev. 19, 122–126.

[154] Tian, D., Sun, S. and Lee, J.T. (2010) The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. Cell 143, 390–403.

[155] Navarro, P. et al. (2010) Molecular coupling of Tsix regulation and pluripotency. Nature 468, 457–460.

[156] Chureau, C., Chantalat, S., Romito, A., Galvani, A., Duret, L., Avner, P. and Rougeulle, C. (2011) Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region. Hum. Mol. Genet. 20, 705–718.

[157] Kim, D.H., Jeon, Y., Anguera, M.C. and Lee, J.T. (2011) X-chromosome epigenetic reprogramming in pluripotent stem cells via noncoding genes. Semin. Cell Dev. Biol. epub ahead of print, doi:10.1016/j.semcdb.2011.02.025.

[158] Zhou, Y. et al. (2007) Activation of p53 by MEG3 non-coding RNA. J. Biol. Chem. 282, 24731–24742.

[159] Zhang, X., Rice, K., Wang, Y., Chen, W., Zhong, Y., Nakayama, Y., Zhou, Y. and Klibanski, A. (2010) Maternally expressed gene 3 (MEG3) noncoding ribonucleic acid: isoform structure, expression, and functions. Endocrinology 151, 939–947.

[160] Huarte, M. et al. (2010) A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell 142, 409–419.

[161] Bernard, D. et al. (2010) A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. EMBO J. 29, 3082–3093.

[162] Tripathi, V. et al. (2010) The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol. Cell 39, 925–938.

[163] Tollervey, J.R. et al. (2011) Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. Nat. Neurosci. 14, 452–458.

[164] Wilusz, J.E., Freier, S.M. and Spector, D.L. (2008) 3′ End processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. Cell 135, 919–932.

[165] Cernilogar, F.M. and Orlando, V. (2005) Epigenome programming by Polycomb and Trithorax proteins. Biochem. Cell Biol. 83, 322–331.

[166] Bantignies, F. and Cavalli, G. (2006) Cellular memory and dynamic regulation of polycomb group proteins. Curr. Opin. Cell Biol. 18, 275–283.

[167] Kiefer, J.C. (2007) Epigenetics in development. Dev. Dyn. 236, 1144–1156.

[168] Kwon, C.S. and Wagner, D. (2007) Unwinding chromatin for development and growth: a few genes at a time. Trends Genet. 23, 403–412.

[169] Schwartz, Y.B. and Pirrotta, V. (2007) Polycomb silencing mechanisms and the management of genomic programmes. Nat. Rev. Genet. 8, 9–22.

[170] Gibney, E.R. and Nolan, C.M. (2010) Epigenetics and gene expression. Heredity 105, 4–13.

[171] Feinberg, A.P. (2007) Phenotypic plasticity and the epigenetics of human disease. Nature 447, 433–440.

[172] Petronis, A. (2010) Epigenetics as a unifying principle in the aetiology of complex traits and diseases. Nature 465, 721–727.

[173] Levenson, J.M. and Sweatt, J.D. (2005) Epigenetic mechanisms in memory formation. Nat. Rev. Neurosci. 6, 108–118.

[174] Roth, T.L., Roth, E.D. and Sweatt, J.D. (2010) Epigenetic regulation of genes in learning and memory. Essays Biochem. 48, 263–274.

[175] Dulac, C. (2010) Brain function and chromatin plasticity. Nature 465, 728–735.

[176] Qureshi, I.A., Mattick, J.S. and Mehler, M.F. (2010) Long non-coding RNAs in nervous system function and disease. Brain Res. 1338, 20–35.

[177] Covic, M., Karaca, E. and Lie, D.C. (2010) Epigenetic regulation of neurogenesis in the adult hippocampus. Heredity 105, 122–134.

[178] Ma, D.K., Marchetto, M.C., Guo, J.U., Ming, G.L., Gage, F.H. and Song, H. (2010) Epigenetic choreographers of neurogenesis in the adult mammalian brain. Nat. Neurosci. 13, 1338–1344.

[179] Auger, A.P. and Auger, C.J. (2011) Epigenetic turn ons and turn offs: chromatin reorganization and brain differentiation. Endocrinology 152, 349–353.

[180] Lister, R. et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462, 315–322.

[181] Kriaucionis, S. and Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 324, 929–930.

[182] Tahiliani, M. et al. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324, 930–935.

[183] Kouzarides, T. (2007) Chromatin modifications and their function. Cell 128, 693–705.

[184] Mattick, J.S., Amaral, P.P., Dinger, M.E., Mercer, T.R. and Mehler, M.F. (2009) RNA regulation of epigenetic processes. Bioessays 31, 51–59.

[185] Mattick, J.S. (2010) 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.

[186] Luco, R.F. and Misteli, T. (2011). More than a splicing code: integrating the role of RNA, chromatin and non-coding RNA in alternative splicing regulation. Curr. Opin. Genet. Dev. epub ahead of print, doi: 10.1016/j.gde.2011.03.004.

[187] Pei, D. (2009) Regulation of pluripotency and reprogramming by transcription factors. J. Biol. Chem. 284, 3365–3369.

[188] Tapscott, S.J. (2005) The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. Development 132, 2685–2695.

[189] Mattick, J.S., Taft, R.J. and Faulkner, G.J. (2010) A global view of genomic information – moving beyond the gene and the master regulator. Trends Genet. 26, 21–28.

[190] Giudicelli, F., Taillebourg, E., Charnay, P. and Gilardi-Hebenstreit, P. (2001) Krox-20 patterns the hindbrain through both cell-autonomous and non cell-autonomous mechanisms. Genes Dev. 15, 567–580.

[191] Narita, Y. and Rijli, F.M. (2009) Hox genes in neural patterning and circuit formation in the mouse hindbrain. Curr. Top. Dev. Biol. 88, 139–167.

[192] Mattick, J.S. (2003) Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. Bioessays 25, 930–939.

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

[194] Mattick, J.S. (2007) A new paradigm for developmental biology. J. Exp. Biol. 210, 1526–1547.

[195] Nickerson, J.A., Krochmalnic, G., Wan, K.M. and Penman, S. (1989) Chromatin architecture and nuclear RNA. Proc. Natl. Acad. Sci. USA 86, 177–181.

[196] Rodriguez-Campos, A. and Azorin, F. (2007) RNA is an integral component of chromatin that contributes to its structural organization. PLoS One 2, e1182.

[197] Mondal, T., Rasmussen, M., Pandey, G.K., Isaksson, A. and Kanduri, C. (2010) Characterization of the RNA content of chromatin. Genome Res. 20, 899–907.

[198] Zhao, J. et al. (2010) Genome-wide identification of Polycomb-associated RNAs by RIP-seq. Mol. Cell 40, 939–953.

[199] Mohammad, F., Mondal, T., Guseva, N., Pandey, G.K. and Kanduri, C. (2010) Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. Development 137, 2493–2499.

[200] Tsai, M.C. et al. (2010) Long noncoding RNA as modular scaffold of histone modification complexes. Science 329, 689–693.

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

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

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

[204] Bernstein, E., Duncan, E.M., Masui, O., Gil, J., Heard, E. and Allis, C.D. (2006) Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. Mol. Cell. Biol. 26, 2560–2569.

[205] Tresaugues, L. et al. (2006) Structural characterization of Set1 RNA recognition motifs and their role in histone H3 lysine 4 methylation. J. Mol. Biol. 359, 1170–1181.

[206] Shimojo, H., Sano, N., Moriwaki, Y., Okuda, M., Horikoshi, M. and Nishimura, Y. (2008) Novel structural and functional mode of a knot essential for RNA binding activity of the Esa1 presumed chromodomain. J. Mol. Biol. 378, 987–1001.

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

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

[209] Cassiday, L.A. and Maher 3rd., L.J. (2002) Having it both ways: transcription factors that bind DNA and RNA. Nucleic Acids Res. 30, 4118–4126.

[210] Wang, K.C. et al. (2011) A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature epub ahead of print, doi: 10.1038/nature09831.

[211] Mohammad, F., Pandey, R.R., Nagano, T., Chakalova, L., Mondal, T., Fraser, P. and Kanduri, C. (2008) Kcnq1ot1/Lit1 noncoding RNA mediates transcriptional silencing by targeting to the perinucleolar region. Mol. Cell. Biol. 28, 3713–3728.

[212] Pandey, R.R. et al. (2008) Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol. Cell 32, 232–246.

[213] Terranova, R., Yokobayashi, S., Stadler, M.B., Otte, A.P., van Lohuizen, M., Orkin, S.H. and Peters, A.H. (2008) Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. Dev. Cell 15, 1–12.

[214] Khalil, A.M. et al. (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc. Natl. Acad. Sci. USA 106, 11667–11672.

[215] Swiezewski, S., Liu, F., Magusin, A. and Dean, C. (2009) Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. Nature 462, 799–802.

[216] Mohammad, F., Mondal, T. and Kanduri, C. (2009) Epigenetics of imprinted long noncoding RNAs. Epigenetics 4, 277–286.

[217] Kotake, Y., Nakagawa, T., Kitagawa, K., Suzuki, S., Liu, N., Kitagawa, M. and Xiong, Y. (2011) Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. Oncogene 30, 1956–1962.

[218] Pal-Bhadra, M., Bhadra, U. and Birchler, J.A. (2002) RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in Drosophila. Mol. Cell 9, 315–327.

[219] Kim, D.H., Villeneuve, L.M., Morris, K.V. and Rossi, J.J. (2006) Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. Nat. Struct. Mol. Biol. 13, 793–797.

[220] Grimaud, C., Bantignies, F., Pal-Bhadra, M., Ghana, P., Bhadra, U. and Cavalli, G. (2006) RNAi components are required for nuclear clustering of Polycomb group response elements. Cell 124, 957–971.

[221] Yu, W., Gius, D., Onyango, P., Muldoon-Jacobs, K., Karp, J., Feinberg, A.P. and Cui, H. (2008) Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. Nature 451, 202–206.

[222] Pasmant, E., Sabbagh, A., Vidaud, M. and Bieche, I. (2011) ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. FASEB J. 25, 444–448.

[223] Manolio, T.A., Brooks, L.D. and Collins, F.S. (2008) A HapMap harvest of insights into the genetics of common disease. J. Clin. Invest. 118, 1590–1605.

[224] He, S., Liu, S. and Zhu, H. (2011) The sequence, structure and evolutionary features of HOTAIR in mammals. BMC Evol. Biol. 11, 102.

[225] Fabian, M.R., Sonenberg, N. and Filipowicz, W. (2010) Regulation of mRNA translation and stability by microRNAs. Annu. Rev. Biochem. 79, 351–379.

[226] Berezikov, E., Thuemmler, F., van Laake, L.W., Kondova, I., Bontrop, R., Cuppen, E. and Plasterk, R.H. (2006) Diversity of microRNAs in human and chimpanzee brain. Nat. Genet. 38, 1375–1377.

[227] Chang, S., Johnston Jr., R.J., Frokjaer-Jensen, C., Lockery, S. and Hobert, O. (2004) MicroRNAs act sequentially and asymmetrically to control chemosensory laterality in the nematode. Nature 430, 785–789.

[228] Hertel, J., Lindemeyer, M., Missal, K., Fried, C., Tanzer, A., Flamm, C., Hofacker, I.L. and Stadler, P.F. (2006) The expansion of the metazoan microRNA repertoire. BMC Genomics 7, 25.

[229] Sempere, L.F., Cole, C.N., McPeek, M.A. and Peterson, K.J. (2006) The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. J. Exp. Zool. B Mol. Dev. Evol. 306, 575–588.

[230] Prochnik, S.E., Rokhsar, D.S. and Aboobaker, A.A. (2007) Evidence for a microRNA expansion in the bilaterian ancestor. Dev. Genes Evol. 217, 73–77.

[231] Heimberg, A.M., Sempere, L.F., Moy, V.N., Donoghue, P.C. and Peterson, K.J. (2008) MicroRNAs and the advent of vertebrate morphological complexity. Proc. Natl. Acad. Sci. USA 105, 2946–2950.

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

[233] Friedman, R.C., Farh, K.K., Burge, C.B. and Bartel, D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 19, 92–105.

[234] Ambros, V. (2004) The functions of animal microRNAs. Nature 431, 350–355.

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

[236] Giraldez, A.J. et al. (2005) MicroRNAs regulate brain morphogenesis in zebrafish. Science 308, 833–838.

[237] Mattick, J.S. and Makunin, I.V. (2005) Small regulatory RNAs in mammals. Hum. Mol. Genet. 14, R121–R132.

[238] Wienholds, E. and Plasterk, R.H. (2005) MicroRNA function in animal development. FEBS Lett. 579, 5911–5922.

[239] Neilson, J.R., Zheng, G.X., Burge, C.B. and Sharp, P.A. (2007) Dynamic regulation of miRNA expression in ordered stages of cellular development. Genes Dev. 21, 578–589.

[240] Stefani, G. and Slack, F.J. (2008) Small non-coding RNAs in animal development. Nat. Rev. Mol. Cell Biol. 9, 219–230.

[241] Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A. and Burge, C.B. (2008) Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. Science 320, 1643–1647.

[242] Friedman, L.M. and Avraham, K.B. (2009) MicroRNAs and epigenetic regulation in the mammalian inner ear: implications for deafness. Mamm. Genome 20, 581–603.

[243] Chandrasekar, V. and Dreyer, J.L. (2009) MicroRNAs miR-124, let-7d and miR-181a regulate cocaine-induced plasticity. Mol. Cell. Neurosci. 42, 350–362.

[244] Edbauer, D. et al. (2010) Regulation of synaptic structure and function by FMRP-associated microRNAs miR-125b and miR-132. Neuron 65, 373–384.

[245] Konopka, W. et al. (2010) MicroRNA loss enhances learning and memory in mice. J. Neurosci. 30, 14835–14842.

[246] Hansen, K.F., Sakamoto, K., Wayman, G.A., Impey, S. and Obrietan, K. (2010) Transgenic miR132 alters neuronal spine density and impairs novel object recognition memory. PLoS One 5, e15497.

[247] Medina, P.P. and Slack, F.J. (2008) MicroRNAs and cancer: an overview. Cell Cycle 7, 2485–2492.

[248] Kocerha, J. et al. (2009) MicroRNA-219 modulates NMDA receptor-mediated neurobehavioral dysfunction. Proc. Natl. Acad. Sci. USA 106, 3507–3512.

[249] Didiano, D. and Hobert, O. (2006) Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. Nat. Struct. Mol. Biol. 13, 849–851.

[250] Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P. and Bartel, D.P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol. Cell 27, 91–105.

[251] Didiano, D. and Hobert, O. (2008) Molecular architecture of a miRNA-regulated 3′ UTR. RNA 14, 1297–1317.

[252] Shin, C., Nam, J.W., Farh, K.K., Chiang, H.R., Shkumatava, A. and Bartel, D.P. (2010) Expanding the microRNA targeting code: functional sites with centered pairing. Mol. Cell 38, 789–802.

[253] Fernandez-Valverde, S.L., Taft, R.J. and Mattick, J.S. (2010) Dynamic isomiR regulation in Drosophila development. RNA 16, 1881–1888.

[254] Politz, J.C., Hogan, E.M. and Pederson, T. (2009) MicroRNAs with a nucleolar location. RNA 15, 1705–1715.

[255] Taft, R.J. et al. (2010) Nuclear-localized tiny RNAs are associated with transcription initiation and splice sites in metazoans. Nat. Struct. Mol. Biol. 17, 1030–1034.

[256] Lippman, Z. et al. (2004) Role of transposable elements in heterochromatin and epigenetic control. Nature 430, 471–476.

[257] Zilberman, D., Cao, X., Johansen, L.K., Xie, Z., Carrington, J.C. and Jacobsen, S.E. (2004) Role of Arabidopsis ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. Curr. Biol. 14, 1214–1220.

[258] Morris, K.V., Santoso, S., Turner, A.M., Pastori, C. and Hawkins, P.G. (2008) Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. PLoS Genet. 4, e1000258.

[259] Brosnan, C.A., Mitter, N., Christie, M., Smith, N.A., Waterhouse, P.M. and Carroll, B.J. (2007) Nuclear gene silencing directs reception of long-distance mRNA silencing in Arabidopsis. Proc. Natl. Acad. Sci. USA 104, 14741–14746.

[260] Dinger, M.E., Mercer, T.R. and Mattick, J.S. (2008) RNAs as extracellular signaling molecules. J. Mol. Endocrinol. 40, 151–159.

[261] Aravin, A. et al. (2006) A novel class of small RNAs bind to MILI protein in mouse testes. Nature 442, 203–207.

[262] Girard, A., Sachidanandam, R., Hannon, G.J. and Carmell, M.A. (2006) A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature 442, 199–202.

[263] Grivna, S.T., Beyret, E., Wang, Z. and Lin, H. (2006) A novel class of small RNAs in mouse spermatogenic cells. Genes Dev. 20, 1709–1714.

[264] Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R. and Hannon, G.J. (2007) Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell 128, 1089–1103.

[265] Houwing, S. et al. (2007) A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell 129, 69–82.

[266] Malone, C.D., Brennecke, J., Dus, M., Stark, A., McCombie, W.R., Sachidanandam, R. and Hannon, G.J. (2009) Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell 137, 522–535.

[267] Senti, K.A. and Brennecke, J. (2010) The piRNA pathway: a fly's perspective on the guardian of the genome. Trends Genet. 26, 499–509.

[268] Simon, B. et al. (2011) Recognition of 2′-O-methylated 3′-end of piRNA by the PAZ domain of a Piwi protein. Structure 19, 172–180.

[269] Tian, Y., Simanshu, D.K., Ma, J.B. and Patel, D.J. (2011) Structural basis for piRNA 2′-O-methylated 3′-end recognition by Piwi PAZ (Piwi/Argonaute/Zwille) domains. Proc. Natl. Acad. Sci. USA 108, 903–910.

[270] Aravin, A.A., Hannon, G.J. and Brennecke, J. (2007) The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science 318, 761–764.

[271] Malone, C.D. and Hannon, G.J. (2009) Small RNAs as guardians of the genome. Cell 136, 656–668.

[272] Grivna, S.T., Pyhtila, B. and Lin, H. (2006) MIWI associates with translational machinery and PIWI-interacting RNAs (piRNAs) in regulating spermatogenesis. Proc. Natl. Acad. Sci. USA 103, 13415–13420.

[273] Khurana, J.S., Xu, J., Weng, Z. and Theurkauf, W.E. (2010) Distinct functions for the Drosophila piRNA pathway in genome maintenance and telomere protection. PLoS Genet. 6, e1001246.

[274] Ohnishi, Y. et al. (2010) Small RNA class transition from siRNA/piRNA to miRNA during pre-implantation mouse development. Nucleic Acids Res. 38, 5141–5151.

[275] Rouget, C. et al. (2010) Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. Nature 467, 1128–1132.

[276] Zhang, D., Duarte-Guterman, P., Langlois, V.S. and Trudeau, V.L. (2010) Temporal expression and steroidal regulation of piRNA pathway genes (mael, piwi, vasa) during Silurana (Xenopus) tropicalis embryogenesis and early larval development. Comp. Biochem. Physiol. C Toxicol. Pharmacol. 152, 202–206.

[277] Yin, H. and Lin, H. (2007) An epigenetic activation role of Piwi and a Piwi-associated piRNA in Drosophila melanogaster. Nature 450, 304–308.

[278] Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T. and Hannon, G.J. (2008) A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol. Cell 31, 785–799.

[279] Kuramochi-Miyagawa, S. et al. (2008) DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev. 22, 908–917.

[280] Brennecke, J., Malone, C.D., Aravin, A.A., Sachidanandam, R., Stark, A. and Hannon, G.J. (2008) An epigenetic role for maternally inherited piRNAs in transposon silencing. Science 322, 1387–1392.

[281] Todeschini, A.L., Teysset, L., Delmarre, V. and Ronsseray, S. (2010) The epigenetic trans-silencing effect in Drosophila involves maternally-transmitted small RNAs whose production depends on the piRNA pathway and HP1. PLoS One 5, e11032.

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

[283] Weber, M.J. (2006) Mammalian small nucleolar RNAs are mobile genetic elements. PLoS Genet. 2, e205.

[284] Schmitz, J., Zemann, A., Churakov, G., Kuhl, H., Grutzner, F., Reinhardt, R. and Brosius, J. (2008) Retroposed SNOfall – a mammalian-wide comparison of platypus snoRNAs. Genome Res. 18, 1005–1010.

[285] Nicoloso, M., Qu, L.H., Michot, B. and Bachellerie, J.P. (1996) Intron-encoded, antisense small nucleolar RNAs: the characterization of nine novel species points to their direct role as guides for the 2′-O-ribose methylation of rRNAs. J. Mol. Biol. 260, 178–195.

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

[287] Pelczar, P. and Filipowicz, W. (1998) The host gene for intronic U17 small nucleolar RNAs in mammals has no protein-coding potential and is a member of the 5′-terminal oligopyrimidine gene family. Mol. Cell. Biol. 18, 4509–4518.

[288] Runte, M., Huttenhofer, A., Gross, S., Kiefmann, M., Horsthemke, B. and Buiting, K. (2001) The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A. Hum. Mol. Genet. 10, 2687–2700.

[289] Mourtada-Maarabouni, M., Pickard, M.R., Hedge, V.L., Farzaneh, F. and Williams, G.T. (2009) GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. Oncogene 28, 195–208.

[290] Kiss, A.M., Jady, B.E., Darzacq, X., Verheggen, C., Bertrand, E. and Kiss, T. (2002) A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains. Nucleic Acids Res. 30, 4643–4649.

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

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

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

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

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

[296] Cavaille, J. et al. (2000) Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. Proc. Natl. Acad. Sci. USA 97, 14311–14316.

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

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

[299] Rogelj, B., Hartmann, C.E., Yeo, C.H., Hunt, S.P. and Giese, K.P. (2003) Contextual fear conditioning regulates the expression of brain-specific small nucleolar RNAs in hippocampus. Eur. J. Neurosci. 18, 3089–3096.

[300] Smalheiser, N.R., Lugli, G., Thimmapuram, J., Cook, E.H. and Larson, J. (2011) Endogenous siRNAs and noncoding RNA-derived small RNAs are expressed in adult mouse hippocampus and are up-regulated in olfactory discrimination training. RNA 17, 166–181.

[301] Jurkowski, T.P., Meusburger, M., Phalke, S., Helm, M., Nellen, W., Reuter, G. and Jeltsch, A. (2008) Human DNMT2 methylates tRNA(Asp) molecules using a DNA methyltransferase-like catalytic mechanism. RNA 14, 1663–1670.

[302] Schaefer, M., Hagemann, S., Hanna, K. and Lyko, F. (2009) Azacytidine inhibits RNA methylation at DNMT2 target sites in human cancer cell lines. Cancer Res. 69, 8127–8132.

[303] Rai, K., Chidester, S., Zavala, C.V., Manos, E.J., James, S.R., Karpf, A.R., Jones, D.A. and Cairns, B.R. (2007) Dnmt2 functions in the cytoplasm to promote liver, brain, and retina development in zebrafish. Genes Dev. 21, 261–266.

[304] Phalke, S., Nickel, O., Walluscheck, D., Hortig, F., Onorati, M.C. and Reuter, G. (2009) Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine-5 methyltransferase DNMT2. Nat. Genet. 41, 696–702.

[305] Kawaji, H. et al. (2008) Hidden layers of human small RNAs. BMC Genomics 9, 157.

[306] Taft, R.J., Glazov, E.A., Lassmann, T., Hayashizaki, Y., Carninci, P. and Mattick, J.S. (2009) Small RNAs derived from snoRNAs. RNA 15, 1233–1240.

[307] Haussecker, D., Huang, Y., Lau, A., Parameswaran, P., Fire, A.Z. and Kay, M.A. (2010) Human tRNA-derived small RNAs in the global regulation of RNA silencing. RNA 16, 673–695.

[308] Ender, C. et al. (2008) A human snoRNA with microRNA-like functions. Mol. Cell 32, 519–528.

[309] Saraiya, A.A. and Wang, C.C. (2008) SnoRNA, a novel precursor of microRNA in Giardia lamblia. PLoS Pathog. 4, e1000224.

[310] Scott, M.S., Avolio, F., Ono, M., Lamond, A.I. and Barton, G.J. (2009) Human miRNA precursors with box H/ACA snoRNA features. PLoS Comput. Biol. 5, e1000507.

[311] Kishore, S. et al. (2010) The snoRNA MBII-52 (SNORD 115) is processed into smaller RNAs and regulates alternative splicing. Hum. Mol. Genet. 19, 1153–1164.

[312] Taft, R.J. et al. (2009) Tiny RNAs associated with transcription start sites in animals. Nat. Genet. 41, 572–578.

[313] Taft, R.J., Kaplan, C.D., Simons, C. and Mattick, J.S. (2009) Evolution, biogenesis and function of promoter-associated RNAs. Cell Cycle 8, 2332–2338.

[314] Cheung, A.C. and Cramer, P. (2011) Structural basis of RNA polymerase II backtracking, arrest and reactivation. Nature 471, 249–253.

[315] Baldi, P., Brunak, S., Chauvin, Y. and Krogh, A. (1996) Naturally occurring nucleosome positioning signals in human exons and introns. J. Mol. Biol. 263, 503–510.

[316] Mavrich, T.N. et al. (2008) Nucleosome organization in the Drosophila genome. Nature 453, 358–362.

[317] Cairns, B.R. (2009) The logic of chromatin architecture and remodelling at promoters. Nature 461, 193–198.

[318] Jiang, C. and Pugh, B.F. (2009) Nucleosome positioning and gene regulation: advances through genomics. Nat. Rev. Genet. 10, 161–172.

[319] Sasaki, S. et al. (2009) Chromatin-associated periodicity in genetic variation downstream of transcriptional start sites. Science 323, 401–404.

[320] Schwartz, S., Meshorer, E. and Ast, G. (2009) Chromatin organization marks exon–intron structure. Nat. Struct. Mol. Biol. 16, 990–995.

[321] Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcarcel, J. and Guigo, R. (2009) Nucleosome positioning as a determinant of exon recognition. Nat. Struct. Mol. Biol. 16, 996–1001.

[322] Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C. and Komorowski, J. (2009) Nucleosomes are well positioned in exons and carry characteristic histone modifications. Genome Res. 19, 1732–1741.

[323] Nahkuri, S., Taft, R.J. and Mattick, J.S. (2009) Nucleosomes are preferentially positioned at exons in somatic and sperm cells. Cell Cycle 8, 3420–3424.

[324] Spies, N., Nielsen, C.B., Padgett, R.A. and Burge, C.B. (2009) Biased chromatin signatures around polyadenylation sites and exons. Mol. Cell 36, 245–254.

[325] Chandler, V.L. (2007) Paramutation: from maize to mice. Cell 128, 641–645.

[326] Cuzin, F., Grandjean, V. and Rassoulzadegan, M. (2008) Inherited variation at the epigenetic level: paramutation from the plant to the mouse. Curr. Opin. Genet. Dev. 18, 193–196.

[327] Nadeau, J.H. (2009) Transgenerational genetic effects on phenotypic variation and disease risk. Hum. Mol. Genet. 18, R202–R210.

[328] Mattick, J.S. (2009) Has evolution learnt how to learn? EMBO Rep. 10, 665.

[329] Allemand, E., Batsche, E. and Muchardt, C. (2008) Splicing, transcription, and chromatin: a menage a trois. Curr. Opin. Genet. Dev. 18, 145–151.

[330] Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M. and Misteli, T. (2010) Regulation of alternative splicing by histone modifications. Science 327, 996–1000.

[331] Luco, R.F., Allo, M., Schor, I.E., Kornblihtt, A.R. and Misteli, T. (2011) Epigenetics in alternative pre-mRNA splicing. Cell 144, 16–26.

[332] Kim, A., Kiefer, C.M. and Dean, A. (2007) Distinctive signatures of histone methylation in transcribed coding and noncoding human beta-globin sequences. Mol. Cell. Biol. 27, 1271–1279.

[333] Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X.S. and Ahringer, J. (2009) Differential chromatin marking of introns and expressed exons by H3K36me3. Nat. Genet. 41, 376–381.

[334] Bass, B.L. (2002) RNA editing by adenosine deaminases that act on RNA. Annu. Rev. Biochem. 71, 817–846.

[335] Nishikura, K. (2010) Functions and regulation of RNA editing by ADAR deaminases. Annu. Rev. Biochem. 79, 321–349.

[336] Navaratnam, N. and Sarwar, R. (2006) An overview of cytidine deaminases. Int. J. Hematol. 83, 195–200.

[337] Conticello, S.G. (2008) The AID/APOBEC family of nucleic acid mutators. Genome Biol. 9, 229.

[338] Jin, Y., Zhang, W. and Li, Q. (2009) Origins and evolution of ADAR-mediated RNA editing. IUBMB Life 61, 572–578.

[339] Paul, M.S. and Bass, B.L. (1998) Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. EMBO J. 17, 1120–1127.

[340] Higuchi, M. et al. (2000) Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. Nature 406, 78–81.

[341] Hartner, J.C., Schmittwolf, C., Kispert, A., Muller, A.M., Higuchi, M. and Seeburg, P.H. (2004) Liver disintegration in the mouse embryo caused by deficiency in the RNA-editing enzyme ADAR1. J. Biol. Chem. 279, 4894–4902.

[342] Chen, C.X., Cho, D.S., Wang, Q., Lai, F., Carter, K.C. and Nishikura, K. (2000) A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains. RNA 6, 755–767.

[343] Jacobs, M.M., Fogg, R.L., Emeson, R.B. and Stanwood, G.D. (2009) ADAR1 and ADAR2 expression and editing activity during forebrain development. Dev. Neurosci. 31, 223–237.

[344] Desterro, J.M., Keegan, L.P., Lafarga, M., Berciano, M.T., O'Connell, M. and Carmo-Fonseca, M. (2003) Dynamic association of RNA-editing enzymes with the nucleolus. J. Cell Sci. 116, 1805–1818.

[345] Maas, S. and Gommans, W.M. (2009) Identification of a selective nuclear import signal in adenosine deaminases acting on RNA. Nucleic Acids Res. 37, 5822–5829.

[346] Macbeth, M.R., Schubert, H.L., Vandemark, A.P., Lingam, A.T., Hill, C.P. and Bass, B.L. (2005) Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. Science 309, 1534–1539.

[347] Athanasiadis, A., Rich, A. and Maas, S. (2004) Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol. 2, e391.

[348] Blow, M., Futreal, P.A., Wooster, R. and Stratton, M.R. (2004) A survey of RNA editing in human brain. Genome Res. 14, 2379–2387.

[349] Kim, D.D., Kim, T.T., Walsh, T., Kobayashi, Y., Matise, T.C., Buyske, S. and Gabriel, A. (2004) Widespread RNA editing of embedded alu elements in the human transcriptome. Genome Res. 14, 1719–1725.

[350] Levanon, E.Y. et al. (2004) Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat. Biotechnol. 22, 1001–1005.

[351] Mercer, T.R., Dinger, M.E., Mariani, J., Kosik, K.S., Mehler, M.F. and Mattick, J.S. (2008) Noncoding RNAs in long-term memory formation. Neuroscientist 14, 434–445.

[352] Labuda, D. and Zietkiewicz, E. (1994) Evolution of secondary structure in the family of 7SL-like RNAs. J. Mol. Evol. 39, 506–518.

[353] Hasler, J. and Strub, K. (2006) Alu elements as regulators of gene expression. Nucleic Acids Res. 34, 5491–5497.

[354] Lander, E.S. et al. (2001) Initial sequencing and analysis of the human genome. Nature 409, 860–921.

[355] Umylny, B., Presting, G., Efird, J.T., Klimovitsky, B.I. and Ward, W.S. (2007) Most human Alu and murine B1 repeats are unique. J. Cell. Biochem. 102, 110–121.

[356] Paz-Yaacov, N. et al. (2010) Adenosine-to-inosine RNA editing shapes transcriptome diversity in primates. Proc. Natl. Acad. Sci. USA 107, 12174–12179.

[357] Mattick, J.S. and Mehler, M.F. (2008) RNA editing, DNA recoding and the evolution of human cognition. Trends Neurosci. 31, 227–233.

[358] Bransteitter, R., Pham, P., Scharff, M.D. and Goodman, M.F. (2003) Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. Proc. Natl. Acad. Sci. USA 100, 4102–4107.

[359] Morgan, H.D., Dean, W., Coker, H.A., Reik, W. and Petersen-Mahrt, S.K. (2004) Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. J. Biol. Chem. 279, 52353–52360.

[360] Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y. and Blau, H.M. (2010) Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature 463, 1042–1047.

[361] Sato, Y., Probst, H.C., Tatsumi, R., Ikeuchi, Y., Neuberger, M.S. and Rada, C. (2010) Deficiency in APOBEC2 leads to a shift in muscle fiber type, diminished body mass, and myopathy. J. Biol. Chem. 285, 7111–7118.

[362] Hilschmann, N., Barnikol, H.U., Barnikol-Watanabe, S., Gotz, H., Kratzin, H. and Thinnes, F.P. (2001) The immunoglobulin-like genetic predetermination of the brain: the protocadherins, blueprint of the neuronal network. Naturwissenschaften 88, 2–12.

[363] Maness, P.F. and Schachner, M. (2007) Neural recognition molecules of the immunoglobulin superfamily: signaling transducers of axon guidance and neuronal migration. Nat. Neurosci. 10, 19–26.

[364] Sawyer, S.L., Emerman, M. and Malik, H.S. (2004) Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. PLoS Biol. 2, E275.

[365] Zhang, J. and Webb, D.M. (2004) Rapid evolution of primate antiviral enzyme APOBEC3G. Hum. Mol. Genet. 13, 1785–1791.

[366] Petit, V., Guetard, D., Renard, M., Keriel, A., Sitbon, M., Wain-Hobson, S. and Vartanian, J.P. (2009) Murine APOBEC1 is a powerful mutator of retroviral and cellular RNA in vitro and in vivo. J. Mol. Biol. 385, 65–78.

[367] Muckenfuss, H. et al. (2006) APOBEC3 proteins inhibit human LINE-1 retrotransposition. J. Biol. Chem. 281, 22161–22172.

[368] Chiu, Y.L., Witkowska, H.E., Hall, S.C., Santiago, M., Soros, V.B., Esnault, C., Heidmann, T. and Greene, W.C. (2006) High-molecular-mass APOBEC3G

complexes restrict Alu retrotransposition. Proc. Natl. Acad. Sci. USA 103, 15588–15593.

[369] Schumann, G.G. (2007) APOBEC3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition. Biochem. Soc. Trans. 35, 637–642.

[370] Aguiar, R.S. and Peterlin, B.M. (2008) APOBEC3 proteins and reverse transcription. Virus Res. 134, 74–85.

[371] Hill, M.S., Mulcahy, E.R., Gomez, M.L., Pacyniak, E., Berman, N.E. and Stephens, E.B. (2006) APOBEC3G expression is restricted to neurons in the brains of pigtailed macaques. AIDS Res. Hum. Retroviruses 22, 541–550.

[372] Muotri, A.R., Chu, V.T., Marchetto, M.C., Deng, W., Moran, J.V. and Gage, F.H. (2005) Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature 435, 903–910.

[373] Coufal, N.G. et al. (2009) L1 retrotransposition in human neural progenitor cells. Nature 460, 1127–1131.

[374] Kuwabara, T. et al. (2009) Wnt-mediated activation of NeuroD1 and retro-elements during adult neurogenesis. Nat. Neurosci. 12, 1097–1105.

[375] Muotri, A.R., Marchetto, M.C., Coufal, N.G., Oefner, R., Yeo, G., Nakashima, K. and Gage, F.H. (2010) L1 retrotransposition in neurons is modulated by MeCP2. Nature 468, 443–446.

[376] Muotri, A.R., Zhao, C., Marchetto, M.C. and Gage, F.H. (2009) Environmental influence on L1 retrotransposons in the adult hippocampus. Hippocampus 19, 1002–1007.

[377] Edelman, G.M. (1993) Neural Darwinism: selection and reentrant signaling in higher brain function. Neuron 10, 115–125.

[378] Singer, T., McConnell, M.J., Marchetto, M.C., Coufal, N.G. and Gage, F.H. (2010) LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? Trends Neurosci. 33, 345–354.

[379] Mattick, J.S. (2001) Non-coding RNAs: the architects of eukaryotic complexity. EMBO Rep. 2, 986–991.

[380] Mattick, J.S. (2004) RNA regulation: a new genetics? Nat. Rev. Genet. 5, 316–323.

[381] Mattick, J.S. (2009) Deconstructing the dogma: a new view of the evolution and genetic programming of complex organisms. Ann. NY Acad. Sci. 1178, 29–46.

[382] Herbert, A. and Rich, A. (1999) RNA processing in evolution: the logic of soft-wired genomes. Ann. NY Acad. Sci. 870, 119–132.

[383] Conway Morris, S. (2000) The Cambrian "explosion": slow-fuse or megatonnage? Proc. Natl. Acad. Sci. USA 97, 4426–4429.