Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms

John S. Mattick

Summary

The central dogma of biology holds that genetic information normally flows from DNA to RNA to protein. As a consequence it has been generally assumed that genes generally code for proteins, and that proteins fulfil not only most structural and catalytic but also most regulatory functions, in all cells, from microbes to mammals. However, the latter may not be the case in complex organisms. A number of startling observations about the extent of non-protein-coding RNA (ncRNA) transcription in the higher eukaryotes and the range of genetic and epigenetic phenomena that are RNA-directed suggests that the traditional view of the structure of genetic regulatory systems in animals and plants may be incorrect. ncRNA dominates the genomic output of the higher organisms and has been shown to control chromosome architecture, mRNA turnover and the developmental timing of protein expression, and may also regulate transcription and alternative splicing. This paper re-examines the available evidence and suggests a new framework for considering and understanding the genomic programming of biological complexity, autopoietic development and phenotypic variation. ***BioEssays*** **25:930–939, 2003.**

Introduction

Since the foundation studies on the genetic code and the regulation of the lac operon in the middle of the last century, it has been generally assumed that ''genes'' are synonymous with proteins, i.e., that genes are repositories of protein-coding sequences, except for those that specify infrastructural RNAs (rRNAS, tRNAs, snoRNAs, spliceosomal RNAs etc.) that are directly or indirectly required for messenger RNA (mRNA) processing and translation.

ARC Special Research Centre for Functional and Applied Genomics, Institute for Molecular Bioscience, University of Queensland, St. Lucia QLD 4072, Australia. E-mail: j.mattick@imb.uq.edu.au
DOI 10.1002/bies.10332
Published online in Wiley InterScience (www.interscience.wiley.com).

Abbreviations: ncRNA, non-protein-coding RNA; miRNA, microRNA; snoRNA, small nucleolar RNA; dsRNA, double-stranded RNA; RNAi, RNA interference; LCR, locus control region.

The notion that genes encode proteins via an mRNA intermediate is derived from the central dogma (DNA>RNA> protein). This essentially holds true in prokaryotes, whose genomes are almost entirely composed of closely packed protein-coding sequences with associated 5′ and 3′ *cis*-regulatory sequences, although recently it has been found that prokaryotes do in fact contain a number of non-protein-coding RNA genes, apart from those encoding rRNAs and tRNAs. These may number 200 or more in *Escherichia coli*, but account for no more than about 0.5% of the total number of genes and about 0.2% of the transcriptional output.[1] In addition, phenotypic variation in bacteria is achieved by varying the proteome, sometimes massively.[2] However, this is not the case in the higher organisms, whose proteomes appear to be relatively stable,[3,4] and whose coding sequences occupy only a tiny fraction of the genome.

This article reviews the recent evidence that non-protein-coding RNAs (ncRNAs) derived from introns of protein-coding genes and the introns and exons of non-protein-coding genes constitute the majority of the genomic programming in the higher organisms. It extends earlier articles on this topic,[4–6] firstly, by exploring the idea that complex organisms require two levels of programming, the specification of the functional components of the system (mainly proteins) and the orchestration of the expression and assembly of these components during differentiation and development and, secondly, by making a series of explicit predictions about the mechanistic basis of the latter, specifically the different levels of genetic regulation controlled by RNA, including chromatin architecture, transcription and alternative splicing. These matters go to the heart of understanding how the trajectories of differentiation and development may be controlled, and are linked to three assertions: (1) that the majority of the genomic sequence in the higher organisms (the non-protein-coding DNA) is devoted to the control of developmental programming, (2) that the majority of the regulatory transactions in the higher organisms are conveyed by RNAs, not proteins, although the two classes of regulatory controls work in concert, and (3) that the combinatorics of protein regulators intersecting with environmental signals in itself provides insufficient state information for the programming of differentiation and development. Rather I suggest that the importance of protein signaling is to provide contextual cues to guide and to tune the (RNA-directed)

endogenously programmed pathways by providing positional information and correcting stochastic errors. Indeed, the problem is not generating complexity per se, but navigating among the explosive numbers of possible trajectories to generate ordered outcomes, such as a human.

Noncoding RNAs constitute the vast majority of the genomic output of complex organisms

Around 97–98% of the transcriptional output of the human genome is non-protein-coding RNA (ncRNA).[6] This estimate is based upon the fact that intronic RNA constitutes 95% of primary protein-coding transcripts (pre-mRNAs),[7,8] and on a range of observations that suggest that there are large numbers of ncRNA transcripts that do not contain substantial open reading frames and which may represent at least half of all transcripts. Detailed analysis of particular loci, such as the β-globin cluster and various imprinted loci in mammals show that almost all of these regions are transcribed, mainly into ncRNAs.[9–12] Most of the discovered microRNAs (miRNAs) are derived from ''intergenic'' regions that were not previously recognized as being transcribed,[13,14] and the known miRNAs are considered to be the tip of an iceberg.[15,16] Whole chromosome analysis using oligonucleotide arrays has revealed that the level of transcription from human chromosomes 21 and 22 is an order of magnitude higher than can be accounted for by known or predicted exons.[17] Finally it appears that almost half of all transcripts identified from well-constructed mouse cDNA libraries are ncRNAs[18] (see below), many of which are spliced, some alternatively, and many are evolutionarily conserved.

It is possible that eukaryotic transcription is inherently sloppy and that there is a large background of nonspecific transcription that could account for some of these observations,[19] but, in all cases that have been examined, these ncRNAs have been found to be developmentally regulated, i.e. expressed in a gender-, tissue- or cell-specific manner, which suggests that these transcripts are not spurious. Thus, it may be that the true number of ''genes'', as defined as segments of the genome that are transcribed to produce functional information, may be much higher than anticipated, with a significant proportion, and probably the majority in complex organisms, producing ncRNAs. This would resolve at least part of the discrepancy between the estimates of mammalian gene numbers based on genome sequence analysis (30–40,000)[7,8] and cDNA cluster analysis (65–70,000),[20] and indicates that the number of genes/transcription units in mammals may in fact be of the order of 60–100,000, except that many (if not most) do not code for protein. Indeed, conclusions about genetic regulatory networks derived from microarray analyses of transcriptome activity in complex organisms may be substantially incomplete if ncRNA sequences are not included in the interrogating set.

This also means that, although protein-coding sequences constitute only about 1.5% of the human genome, a very large proportion of the genome is in fact expressed.[21] Given that the average human protein-coding transcript is 95% intronic,[7,8] at least 30% of the genome must be transcribed, a rather astounding figure that is generally unappreciated. If one also allows a similar number of non-protein-coding transcripts, as appears to be the case, then at least half of the genome is transcribed, an estimate supported by a recent analysis of human chromosomes 7, 21 and 22.[17,22] What fraction of the expressed RNAs serves some biological purpose, or is simply degraded, is unknown. However, it is hard to escape the general conclusion that either the human genome is replete with useless transcription, or that these RNAs are fulfilling some unexpected function(s).

ncRNAs also appear to constitute the majority of the transcriptional output in other animals, such as *Drosophila*, where the intron:exon ratio is approximately 1:1,[23] and where molecular genetic analyses of well-studied loci have revealed a significant proportion of ncRNAs. For example, only three of the seven major transcripts identified in the *bithorax-abdominalAB* region encode proteins, but all seven are developmentally regulated and the interruption or deletion of the DNA that encodes them has known phenotypic consequences.[9,10] In the pufferfish, *Fugu rubripes*, which has a minimal genome with little repetitive DNA, analysis of the genome sequence shows that the average intron:exon ratio in protein-coding genes is around 2:1, i.e. there is still much more unique sequence information in introns than in protein-coding sequences in this organism.[24] In addition, conventional protein-coding genes (including their introns) occupy only one third of the *Fugu* genome[24] and it seems highly likely that much of the remainder, which has been referred to as ''gene bare'',[24] or in the case of the human genome as ''gene deserts'',[8] may in fact be expressed as ncRNAs and/or be *cis*-acting regulatory sequences receiving signals from RNAs and proteins.

Identification of noncoding RNAs

It is only recently that noncoding RNAs have begun to be studied in any systematic way and that it has been recognized that noncoding RNAs (ncRNAs) may have a significant role in cell and developmental biology. Relatively few ncRNAs have been identified or described in any detail since, in most cases, these RNAs either have not been anticipated or, if detected, have been set aside in favour of focusing on the proteins in the pathway or genome concerned.[25] In some cases, the presence of unusual intronic, antisense or intergenic transcripts have been noted in published reports, but in most cases not. In essence, most researchers have not known what do with these RNAs, which have been considered of uncertain significance, rather than potentially important components of cellular systems.

Despite this, a number of functional ncRNAs have been documented by different laboratories and provide some insight into this largely unexplored facet of the genomic output of the higher organisms. In animals these include: (1) *Pgc*, required for germ cell formation in *Drosophila*,[26] (2) *7H4, BC1, Bsr* and *Ntab*, which are expressed in particular cell types within the rat nervous system,[27–30] (3) *adapt33*, induced by oxidants in hamster HA-1 cells,[31] (4) *BC200*, which is limited to the primates and is expressed specifically in neuronal cells,[32] (5) *Bic*, which is strongly upregulated in certain B-cell lymphomas,[33] (6) *NTT*, which is expressed in activated T cells,[34] (7) *DISC2*, implicated in the molecular etiology of schizophrenia,[35] (8) the Y-chromosome-specific *TTY*2 family, which is expressed in human testis and kidney,[36] (9) *roX1/2* and *Xist/Tsix*, which are involved in X chromosome dosage compensation in insects and mammals, respectively,[37,38] (10) *ncR-uPAR*, which regulates human protease-activated receptor-1 gene during embryogenesis,[39] (11) *FGF-AS*, which is differentially spliced to produce coding and noncoding RNAs in different tissues and which regulates the expression of fibroblast growth factor-2,[40] (12) *BORG*, induced following exposure to bone morphogenetic proteins,[41] (13) *H19*, an abundant hepatic fetal-specific ncRNA implicated as a tumor suppressor,[42] (14) *SCA8*, within which an expansion of a CTG trinucleotide repeat underlies the neurodegenerative disorder spinocerebellar ataxia type 8,[43] and (15) *RMRP*, mutations of which underlie the recessively inherited developmental disorder, cartilage-hair hypoplasia (CHH), with manifestations including short stature, defective cellular immunity, and predisposition to several cancers.[44] In addition, it has recently been shown that one of the loci associated with autism encodes a large ncRNA.[22] There are also many ncRNAs documented at imprinted loci in animals (see e.g. Ref. 45), and it has recently been shown that genes can be regulated by homologous noncoding pseudogene transcripts,[46] of which over 20,000 may exist in the human genome.[47] Such observations also resonate with the increased awareness that antisense transcription is widespread in mammalian genomes,[48] not just at imprinted loci, suggesting RNA-mediated local regulatory loops.

The actual number of ncRNA genes in mammals is probably in the order of tens of thousands (see below). Why have more of these RNAs not been discovered? There appear to be many reasons. Apart from the general lack of alertness mentioned above, most biochemical analyses of cell fractions are not designed to detect ncRNAs. RNA is also labile, and there are no well-established protocols for analysing the functions of ncRNAs, apart from gene knockouts, which are technically demanding, often ambiguous and not undertaken lightly. Their genetic effects may also be subtle and/or may have been interpreted in terms of the prevailing wisdom, as for example in the *bxd* region in Drosophila, which expresses an ncRNA, but is normally considered simply to be the promoter region for the adjacent *Ubx* gene.[4] In addition, many ncRNAs are expressed at low levels,[17,18] and thus poorly represented in cDNA libraries, unless normalized carefully.

Most cDNA collections are mainly composed of incomplete reverse transcripts (''expressed sequence tags'' or ESTs) in which it has been impossible to conclude that an open reading frame is absent. However, over the past few years, RIKEN in Tokyo have taken a systematic approach to determining the mouse transcriptome, by cloning and sequencing full-length cDNAs in normalized libraries from different tissues. In the latest round of annotation,[18] 60,770 full-length cDNAs were clustered into 33,409 transcription units, of which 15,815 appear to be ncRNAs in that they do not code for a substantial open reading frame (ORF) (greater than 100 codons). Over 4,200 of these are strong candidates for ncRNAs by stringent criteria, which is clearly an underestimate of the total, as some known ncRNAs fail these criteria. A large number of these transcripts are represented by more than one independent clone and many have been shown to be differentially expressed in different tissues, and thus are real transcripts, not genomic contamination of the cDNA library. Almost 30% are spliced. In addition, there are over 2,400 pairs of overlapping sense-antisense transcripts, of which almost 1,600 had not previously been identified,[18] an apparently common phenomenon.[48] Some of these transcripts, such as novel antisense transcripts from the imprinted Gnas locus, have since been experimentally verified.[49]

It has also been shown that both plants and animals contain large numbers of microRNAs (miRNAs), short RNAs (21–22 nucleotides) that are produced by post-transcriptional processing of ncRNAs by enzymes, including Dicer and RDE homologs, which also participate in the phenomenon of RNA interference (RNAi), and which involves double-stranded RNA binding and cleavage.[13,14,50–53] These miRNAs are derived from precursors that in some cases are known ncRNAs, such as *Bic*.[14] Many others are sourced from ''intergenic regions'' of the genome, from sequences that are antisense to known genes, and from sequences that have been annotated as introns.[13,52] At least some transcripts give rise to multiple miRNAs,[13] and analysis of their sequence strongly suggests that the specificity of miRNA processing is guided by the secondary (stem-loop) structure of the precursor RNA.[13,51,54] Taken together, this suggests firstly that many genomic regions that have been annotated as ''intergenic'' (i.e. sequences between protein-coding genes) contain ncRNA genes, and secondly that many RNAs may be processed by pathways involving double-stranded RNases, helicases and exonucleases. This appears to be the tip of a much bigger iceberg of large numbers of small *trans*-acting RNAs that are produced from ncRNA and intronic precursors,[4,5] and whose incidence has until recently gone unnoticed, because of the complexity of this population and the fact that small RNAs often escape detection in conventional biochemical analyses.

miRNAs may have a role in tissue specification or cell-lineage decisions.[52,53] In plants, most miRNAs appear to target protein-coding sequences, frequently of transcription factors, leading to the suggestion that in plants miRNAs have a major function in targeting mRNAs for destruction, perhaps as part of a differentiation program,[52,54,55] although other small dsRNAs clearly have a role in directing DNA methylation.[52,56,57] In animals, many miRNAs are tissue-specific and many are conserved among different species, in some cases over large evolutionary distances (e.g. nematodes, insects, fish and mammals).[13,14] Some of these RNAs, such as the archetypal *lin-4* and *let-7* in *C. elegans*, have known developmental functions, and act by binding to the 3′ end of target mRNAs.[53] Others (termed "small interfering RNAs") participate in targeted destruction of other RNAs via the RNAi pathway, with the difference apparently being in the degree of sequence complementarity with the target.[54,58] On this basis, it has been believed that miRNAs function primarily as modulators of mRNA translation and stability, but recent evidence indicates that miRNAs also play a key role in epigenetic modification of chromatin (see below). There are probably tens or even hundreds of thousands of small RNAs produced by processing of expressed noncoding RNA sequences, including introns.[4] Since most remain to be identified, it is highly likely that such RNAs transmit a variety of signals to different targets and fulfil a wide variety of functions in the regulation of gene expression during eukaryotic development[4,5] (see below).

Introns

Since they do not code for protein, introns have been generally regarded as non-functional, and thus were originally rationalized as evolutionary hangovers from the early assembly of genes, maintained in part because of their utility in enabling protein domain shuffling. However, while the mosaic structure of genes has been exploited in this way, it now appears that nuclear pre-mRNA introns are the descendants of self-splicing group II introns which expanded in eukaryotes late in evolution, and which evolved in situ in parallel with the host genome,[5] as also appears to have happened with transposons and other types of molecular parasites.[7]

It is likely that intronic RNAs are processed post-splicing and also participate in RNA-mediated transactions within the cell. The importance of this is that such a system would in theory and in practice comprise a parallel network of efference signals which allow transcriptional activity at one locus to directly influence others.[4,54] Introns have interesting patterns of conservation, often in blocks,[4,59] which have become globally evident from sequence comparisons of syntenic regions between the human and mouse genomes[25] and can be readily viewed in genome browsers such as Ensembl (http://www.ensembl.org/). In some cases, intron-derived RNAs have been shown to be stable and trafficked to different subcellular compartments.[60,61] In addition, at least some introns are known to produce functional RNAs. Most small nucleolar RNAs in the higher eukaryotes are derived from introns, some of which are excised from primary transcripts encoding ribosomal proteins and other proteins involved in nucleolar biology and the cell cycle, but many of which have no protein-coding capacity,[4,62] including the known ncRNA, *Bsr*.[63] At least some of these snoRNAs are developmentally regulated and appear to have a role in the processing of mRNAs to produce even greater functional diversity of their encoded products,[64] as well as being implicated in the control of epigenetic imprinting.[65] These and other RNAs are processed from longer precursors in large complexes called exosomes, and involve dsRNAses, helicases and exonucleases,[66,67] reminiscent of the pathways involved in the production of microRNAs and small interfering RNAs.[50,68]

Genes as fuzzy transcription clusters with multiple products

The classification of ncRNAs is a problem. Many ncRNAs are simply unprocessed primary transcripts but, in other cases, they are formed from the exons of spliced transcripts, which may also be alternatively spliced or polyadenylated in a developmentally regulated fashion. Some ncRNAs are derived from the further processing of exons, such as the miRNAs that are produced from the *Bic* transcript,[14] and of introns of both protein-coding and ncRNA genes, as exemplified by the snoRNAs.

Such considerations also require a reassessment of what is meant by a "gene". Traditionally, the genetic definition has been based on an inheritable phenotype irrespective of how the relevant information and variation is encoded, whereas the biochemical definition has tended to emphasize the definition of a gene as a protein-coding region with associated regulatory signals. It is now clear that the assumption that most genetic information is expressed as proteins is incorrect, at least in the higher organisms. Recognition of this fact has led to the re-definition of a gene as a "transcription unit"[18] or "a complete chromosomal segment responsible for making a functional product".[69] Moreover given the fact that many transcripts are initiated from alternate promoters with alternate splicing and polyadenylation signals, the definition of a gene perhaps needs to be expanded to encompass a transcription cluster, defined as a separate locus producing one or more related transcripts from the same DNA strand (together with associated *cis*-acting regulatory elements), which may encode one or more related protein products (alternatively spliced isoforms) and/or one or more different ncRNAs (such as transcripts that specify seven different miRNAs,[13] or multiple snoRNAs[64,65] in parallel with exonic-encoded proteins) (Fig. 1). Even this definition is inadequate in that overlapping transcripts may encode different products or a different repertoire of products with different functional

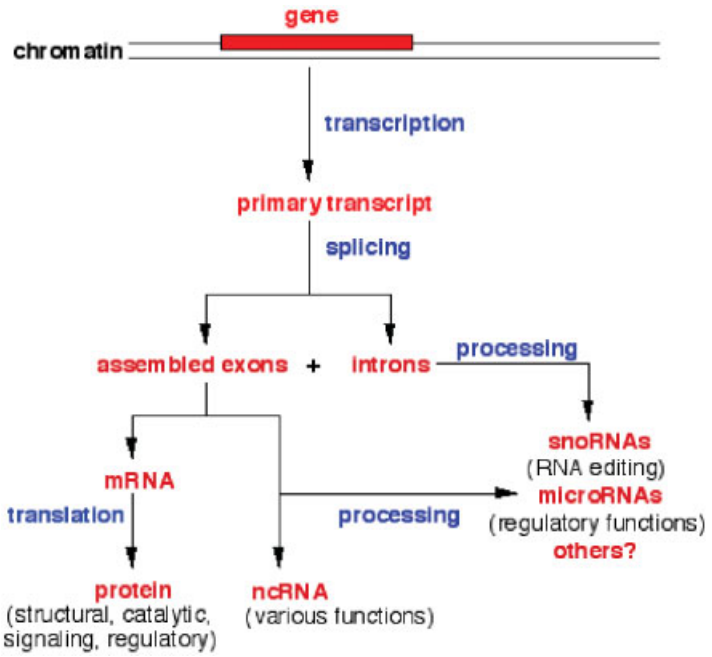

Figure 1. A revised view of the flow of genetic information in the higher eukaryotes. Primary transcripts may be (alternatively) spliced and further processed to produce a range of protein isoforms and/or ncRNAs of various types, which are involved in complex networks of structural, functional and regulatory interactions.

consequences,[69] and different mRNAs may be produced by *trans*-splicing between different primary transcripts.[70,71] It is also problematical whether an overlapping antisense transcript that may be involved in the *cis*-regulation of its sense pair should be considered part of the same "gene" or not, and we are left with the notion that a "gene" is at best a fuzzy concept to describe a sequence of genetic information, expressed either as RNA (per se) and/or protein (via mRNA), that has functional consequences. It is also clear that, in the higher organisms, genetic information is encoded in very complex ways, and that the readouts are not discrete. In this light, it is also simplistic to assume that any transcriptional output that is not (yet) understood is noise.

Complex genetic phenomena involving RNA

The central importance of RNA signaling to eukaryotic cell and developmental biology has become apparent from the range of complex genetic phenomena in the higher eukaryotes that have been shown to involve RNA.[4,68,72] Co-suppression, transcriptional and post-transcriptional gene silencing, and RNAi are all related and involve dsRNA-targeted destruction of mRNAs and, at least in plants, transcriptional silencing via DNA methylation.[4,68] These processes are often considered to be a defense against transposons and viruses, but it is also clear that they play an important role in normal development.[73]

Methylation is RNA-directed in plants and probably also in animals.[56,72] Imprinting in animals involves antisense RNAs and DNA methylation,[37] and the loss of expression of ncRNAs has been implicated in the imprinting disease Prader-Willi syndrome.[65,74] The link between DNA methylation, antisense RNAs, imprinting and transcriptional gene silencing suggests that RNA-directed DNA methylation is involved in epigenetic gene regulation throughout the eukaryotes.[56] Co-suppression has also been reported in animals and, at least in Drosophila and C. elegans, is dependent on Polycomb-group proteins,[4,68,75] as is transgene silencing,[76] which implicates not only RNA but also the structure of chromatin complexes in these processes.[4,68,77] Polycomb-group proteins are also involved in transvection, a poorly understood genetic phenomenon involving cross-talk between regulatory and protein-coding regions on sister chromosomes that occurs throughout the eukaryotes, and which has also been implicated in genomic imprinting and X-chromosome inactivation, and which is probably linked to RNA signaling.[4] In addition, it has been shown that the locus control region (LCR) and the intergenic regions of the human β-globin locus are transcribed in erythroid but not in non-erythroid cells, but can be induced to do so in the latter following transient transfection with a β-globin gene via a process called "transinduction" that depends on the transcription but not translation of the plasmid-encoded transgene.[11]

RNA binding and signaling proteins

RNA binding, processing and signaling proteins are common in the higher eukaryotes. In addition, at least some proteins that are considered to be "transcription factors" have high affinity for RNA or for higher order nucleic acid structures that

contain RNA. These include (1) zinc finger proteins, such as the classical housekeeping transcription factor Sp1, which have comparable or greater affinity for RNA–DNA hybrids than for double-stranded DNA, which is strand-specific,[78] (2) WT1, whose splice variants can apparently bind selectively to chromatin and spliceosomes, and (3) the Y-box (cold-shock) proteins, which bind RNA.[79] The adenosine deaminases that modify or edit dsRNAs (ADARS) and play a role in RNAi[80] have been shown to contain domains related to winged helix-turn-helix domains and the globular domain of histone H5, which binds Z-DNA and/or catalyzes its formation.[81]

It has also been shown that the chromodomain intersects with RNA signals.[82] The chromodomain determines the specificity of binding to particular chromatin locations[83] via modified histones[84] and is found in a wide variety of chromatin-associated regulatory proteins with a wide variety of other functional domains, including Polycomb-group proteins, the chromodomain-helicase-DNA binding family, the retinoblastoma-binding protein family, the histone acetytransferase and histone methyltransferase families, the heterochromatin protein 1 family, and the SWI3 family.[77] This implicates RNA signaling in the targeting of chromatin complexes and in the positive and negative regulation of gene expression during differentiation and development, as well as in chromosome remodeling, histone methylation and acetylation. Recent evidence has shown that the programmed DNA elimination that accompanies macronuclear development in *Tetrahymena* and the initiation of heterochromatin formation in *Schizosaccharomyces pombe* are mediated by components of the RNAi pathway.[85–88] These processes are accompanied by histone H3 methylation at lysine 9 following the recruitment of Pdd1p, Pdd3p and Swi6/Hp1, all chromodomain-containing proteins. In addition, the RNAi pathway is also required for the proper regulation of mitosis and meiosis in *S. pombe*,[89] the latter of which has previously been reported to involve a number of ncRNAs.[90]

Functions of ncRNAs

ncRNAs will undoubtedly fulfil a wide range of functions within cells, as it would be expected that nature will exploit any interactions and chemistries that have functional value. Many different ncRNAs with different functions in eukaryotic cell and developmental biology have already been described.[91,92] However, the dominance of ncRNAs in the genomic output of the higher organisms suggests that they are not simply occasional transcripts with idiosyncratic functions, but rather that they may constitute an extensive but hitherto unrecognized regulatory network within higher organisms. If so, there are specific predictions about the types of functions that these RNAs might fulfil at various levels in the control of molecular genetic activity, which one would expect to occur via a variety of RNA–DNA, RNA–RNA and RNA–protein interactions.

Chromatin modification and epigenetic memory

Firstly, as noted above, there is a strong case that most if not all epigenetic phenomenon, i.e. the cellular memory of developmental history and its effect on current gene activation status, is directed by *trans*-acting RNAs, as RNA signaling appears to be involved in DNA methylation, imprinting, transvection, position-effect variegation, chromatin remodeling and the activation and repression of chromosomal domains. Epigenetic memory is undoubtedly central to the process of differentiation and development and appears to be a major target of the emerging RNA regulatory network.

Transcription

One can also make the prediction that *trans*-acting RNAs are central to the initiation of transcription, i.e. that sequence-specific RNA signals are involved in the targeting and recruitment of transcription factors and transcription complexes at appropriate places in the genome, whose general availability has already been determined by longer-term epigenetic mechanisms also controlled by RNA signaling. The β-globin LCR, which is considered to be the archetypal long-distance transcriptional "enhancer", is itself specifically transcribed in erythroid cells.[11] It has also recently been shown that steroid receptor transactivation requires an ncRNA called SRA, and it appears that this may be just one example of a large family of such RNAs.[93,94] The ncRNA *7SK* is involved in the transcriptional activation of the proto-oncogene *c-myc*.[95] The idea that a *trans*-acting ncRNA may act to recruit transcription factors to sequence-specific locations in the chromosome has some merit, as it would potentially solve the problem that most transcription factors have relatively loose consensus binding sites that occur in very many places in the genome, as well as provide an explanation as to why many transcription factors appear have affinity for nucleic acid structures involving RNA. This would also provide an internal mechanism for programmed transcription during differentiation and development, in conjunction with epigenetic changes to chromatin, and suggests that the cellular receptor and signaling cascades that intersect with nuclear transcription factors may serve mainly to integrate the endogenous programming with exogenous information about cell position, morphogen gradients and physiological status.

Alternative splicing

Alternative splicing may also be regulated by *trans*-acting RNAs. Alternative splicing is another programmed response that is of fundamental importance to development, but the basis of its specificity is presently not understood. Hundreds, if not thousands, of mRNAs are alternatively spliced in differentiated cells. However, few protein factors that control the alternate splicing of specific genes have been described, one exception being *sex-lethal* in *Drosophila*, and the process is usually considered to be combinatorially determined by

changes in the concentration or activity of one or more splicing factors that either activate or repress particular splice site choices in different contexts.[96,97] The *cis*-acting sequences involved in splice site selection within pre-mRNAs are usually located at or near the intron–exon boundary.[98] One way to target these sequences, therefore, is by RNA itself, i.e. *trans*-acting RNAs (produced from other loci or from antisense transcripts), which specifically address the site in question, may activate or (more likely) inhibit exon selection. Indeed, it has been shown that antisense oligoribonucleotides can alter the splicing patterns of specific transcripts in both cultured cells and transgenic animals,[99–101] in which case it is perfectly plausible that this also occurs naturally in vivo. It has also been suggested that snoRNAs may be able to alter the splicing patterns of specific transcripts by RNA editing.[64]

Developmental timing of mRNA turnover, translation and cell signaling

As noted above, it is well established that small *trans*-acting RNAs (such as *lin-4* and *let-7*) can affect mRNA stability and translation, and play a central role in developmental timing.[53–55] It would also be predicted that many ncRNAs will interact with proteins and participate in cytoplasmic cell signaling pathways. Many polyadenylated ncRNAs are found in the cytoplasm (T. Gingeras, personal communication), and there are known proteins that bind RNA that include signaling domains, such as SH3-domain-binding domains.[102]

Genetic programming of complex organisms

Complex organisms require two levels of genetic programming for their autopoietic development (i.e. self-referential assembly)—firstly the components of the system, and secondly the deployment of these components during differentiation and development. The structural and functional components of the system are mainly proteins, and their synthesized products (lipids, carbohydrates and nucleic acids), which have been the primary focus of biochemical research to date. Damage to proteins by mutations that render them non-functional or misfunctional is very obvious and usually gives a severe phenotype, for example in monogenic diseases such as cystic fibrosis and thalassemia, or in cancer, just as damaging or compromising the functional components of any structure is usually obvious.

Complex organisms, or indeed any complex objects, have a wide range of components, many of which are in fact common (or very similar), in related organisms or objects. This is reflected in the similarity of the proteome between, for example, human and mouse where only about 1% of the proteins do not have recognizable homologs in the other species.[25] Many of these components are also re-used in different contexts (multitasked),[3,4] which also has the impact of freezing their design, as one cannot re-design a core component without re-designing all of the contexts in which it is used. The solution to this problem in the higher organisms is alternate splicing, which allows design variations in new contexts, while still preserving the original design in pre-existing contexts.

The genomes of complex organisms must also contain all of the information required to specify the timing, patterns, variations and amounts of expression of these components during development, and therefore must also program the overall design of the organism and individual variations. This is no trivial matter. Every cell in *C. elegans* has a defined ontogeny and fate, and this is likely to be true for most cells in animals, except those that clonally expand under (e.g.) immune pressure or nutritional conditions. Traditionally it has simply been assumed that the programming of animal and plant development is embedded in *cis*-acting control sequences (promoters and enhancers), which regulate gene expression in conjunction with various combinations of *trans*-acting proteins that relay environmental cues. This assumption is not necessarily correct. On the contrary, the massive amount of ncRNA that is expressed from the genomes of higher organisms, and the complex genetic phenomena that involve RNA, suggests that ncRNAs may constitute an endogenous control system that regulates the programmed patterns of gene expression during their development.

The regulatory regions of higher organisms, including promoters and therefore potentially also introns and ncRNAs, are generally much less conserved than protein-coding regions, and are presumably far more plastic, but this does not mean that they are nonfunctional. The vast majority of human genomic polymorphisms occur outside of protein-coding regions[8] but known mutations or polymorphisms in promoters or other noncoding sequences that have strong effects on phenotype are rare, for the simple reason that, while a single base change can have disastrous effects on the structure and function of a vital protein component, it may have only subtle effects on *cis*- and *trans*-interactions in regulatory networks that are also likely to be intrinsically robust.[103] Sequence differences in this control architecture would be expected primarily to affect quantitative trait variation, and also probably includes a large proportion of the genetic contributions to susceptibility to complex diseases.

Conclusions

The presumed universality of the central dogma and the flow of genetic information was encapsulated by Jacques Monod's famous statement (reported in reference 104) that "What was true for *E. coli* would also be true for the elephant" and ever since has dominated our conception of the nature of genetic information and the structure of genetic systems. Although Monod did suggest that RNA itself may have (other) functions, the prevailing orthodoxy has been that proteins not only constitute the primary structural and functional components of living cells, but also constitute most of the regulatory control

system, in both simple and complex organisms. The central dogma has therefore not only been taken to mean that most genes encode proteins, but also that proteins are sufficient in themselves to specify and organize the autopoietic programming of complex biological entities, an assumption that has pervaded molecular biology for decades. This assumption must now be reassessed.

The fact that introns and noncoding RNAs carry out the majority of the transcription of the genomes of humans and other complex organisms suggests that a second tier of genetic output and a network of parallel RNA-mediated interactions has evolved in these organisms, which may enable the integration and coordination of sophisticated suites of gene expression required for differentiation and development.[4–6] The expansion of the complement of ncRNAs in the higher organisms also suggests that the evolution of complexity may not have been simply dependent on an expanded repertoire of proteins and protein isoforms, but on a (much) larger set of genomic design instructions embedded in *trans*-acting RNAs (and *cis*-acting receiver sequences), which form the basis of a cascade of programmed response networks capable of implementing stored sequences of dynamical activities in response to internal and external stimuli.[4–6] It is also likely that alteration in this control architecture is responsible for much of the phenotypic variation that is observed between individuals and species (such as vertebrates) that use a relatively common set of functional components, with the remainder of the variation (and the majority of catastrophic problems) due to variation in the components (the proteins) themselves.

Such considerations also bring an entirely new perspective on how self-programming information may be stored and communicated in complex organisms. The principles that emerge have much in common with other advanced information-processing systems that underlie neural and computer function, including component multitasking and endogenous control signals (referred to as the "hidden layer" in neural systems),[4] and may also underpin the regulatory and epigenetic interactions controlling the ontogeny of complex organisms. An elephant may not be like *E. coli*.

References

1. Argaman L, Hershberg R, Vogel J, Bejerano G, Wagner EG, Margalit H, Altuvia S. Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. Curr Biol 2001;11:941–950.
2. Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K, Han CG, Ohtsubo E, Nakayama K, Murata T, et al. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res 2001;8:11–22.
3. Duboule D, Wilkins AS. The evolution of ''bricolage''. Trends Genet 1998;14:54–59.
4. Mattick JS, Gagen MJ. The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. Mol Biol Evol 2001;18:1611–1630.
5. Mattick JS. Introns: evolution and function. Curr Opin Genet Dev 1994; 4:823–831.
6. Mattick JS. Non-coding RNAs: the architects of eukaryotic complexity. EMBO Reports 2001;2:986–991.
7. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. Initial sequencing and analysis of the human genome. Nature 2001;409:860–921.
8. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, et al. The sequence of the human genome. Science 2001;291:1304–1351.
9. Lipshitz HD, Peattie DA, Hogness DS. Novel transcripts from the Ultrabithorax domain of the bithorax complex. Genes Dev 1987;1:307–322.
10. Sanchez-Herrero E, Akam M. Spatially ordered transcription of regulatory DNA in the bithorax complex of *Drosophila*. Development 1989; 107:321–329.
11. Ashe HL, Monks J, Wijgerde M, Fraser P, Proudfoot NJ. Intergenic transcription and transinduction of the human beta-globin locus. Genes Dev 1997;11:2494–2509.
12. Wroe SF, Kelsey G, Skinner JA, Bodle D, Ball ST, Beechey CV, Peters J, Williamson CM. An imprinted transcript, antisense to Nesp, adds complexity to the cluster of imprinted genes at the mouse Gnas locus. Proc Natl Acad Sci USA 2000;97:3342–3346.
13. Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science 2001;294:858–862.
14. Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T. Identification of tissue-specific microRNAs from mouse. Curr Biol 2002;12:735–739.
15. Ruvkun G. Glimpses of a tiny RNA world. Science 2001;294:797–799.
16. Moss EG. MicroRNAs: hidden in the genome. Curr Biol 2002;12:R138–R140.
17. Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, Fodor SP, Gingeras TR. Large-scale transcriptional activity in chromosomes 21 and 22. Science 2002;296:916–919.
18. Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, Nikaido I, Osato N, Saito R, Suzuki H, et al. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature 2002;420:563–573.
19. Dennis C. The brave new world of RNA. Nature 2002;418:122–124.
20. Zhuo D, Zhao WD, Wright FA, Yang HY, Wang JP, Sears R, Baer T, Kwon DH, Gordon D, Gibbs S, et al. Assembly, annotation, and integration of UNIGENE clusters into the human genome draft. Genome Res 2001;11:904–918.
21. Wong GK, Passey DA, Yu J. Most of the human genome is transcribed. Genome Res 2001;11:1975–1977.
22. Scherer SW, Cheung J, MacDonald JR, Osborne LR, Nakabayashi K, Herbrick JA, Carson AR, Parker-Katiraee L, et al. Human chromosome 7: DNA sequence and biology. Science 2003;300:767–772.
23. Deutsch M, Long M. Intron-exon structures of eukaryotic model organisms. Nucleic Acids Res 1999;27:3219–3228.
24. Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, et al. Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. Science 2002;297:1301–1310.
25. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, et al. Initial sequencing and comparative analysis of the mouse genome. Nature 2002;420:520–562.
26. Nakamura A, Amikura R, Mukai M, Kobayashi S, Lasko PF. Requirement for a noncoding RNA in *Drosophila* polar granules for germ cell establishment. Science 1996;274:2075–2079.
27. Velleca MA, Wallace MC, Merlie JP. A novel synapse-associated noncoding RNA. Mol Cell Biol 1994;14:7095–7104.
28. Lin Y, Brosius J, Tiedge H. Neuronal BC1 RNA: co-expression with growth-associated protein-43 messenger RNA. Neuroscience 2001; 103:465–479.
29. Komine Y, Tanaka NK, Yano R, Takai S, Yuasa S, Shiroishi T, Tsuchiya K, Yamamori T. A novel type of non-coding RNA expressed in the rat brain. Brain Res Mol Brain Res 1999;66:1–13.
30. French PJ, Bliss TV, O'Connor V. Ntab, a novel non-coding RNA abundantly expressed in rat brain. Neuroscience 2001;108:207–215.

31. Wang Y, Crawford DR, Davies KJ. adapt33, a novel oxidant-inducible RNA from hamster HA-1 cells. Arch Biochem Biophys 1996;332:255–260.
32. Kuryshev VY, Skryabin BV, Kremerskothen J, Jurka J, Brosius J. Birth of a gene: locus of neuronal BC200 snmRNA in three prosimians and human BC200 pseudogenes as archives of change in the Anthropoidea lineage. J Mol Biol 2001;309:1049–1066.
33. Tam W. Identification and characterization of human BIC, a gene on chromosome 21 that encodes a noncoding RNA. Gene 2001;274:157–167.
34. Liu AY, Torchia BS, Migeon BR, Siliciano RF. The human NTT gene: identification of a novel 17-kb noncoding nuclear RNA expressed in activated CD4+ T cells. Genomics 1997;39:171–184.
35. Millar JK, Wilson-Annan JC, Anderson S, Christie S, Taylor MS, Semple CA, Devon RS, Clair DM, Muir WJ, Blackwood DH, Porteous DJ. Disruption of two novel genes by a translocation co-segregating with schizophrenia. Hum Mol Genet 2000;9:1415–1423.
36. Makrinou E, Fox M, Lovett M, Haworth K, Cameron JM, Taylor K, Edwards YH. Tty2: a multicopy y-linked gene family. Genome Res 2001;11:935–945.
37. Kelley RL, Kuroda MI. Noncoding RNA genes in dosage compensation and imprinting. Cell 2000;103:9–12.
38. Lee JT, Davidow LS, Warshawsky D. Tsix, a gene antisense to Xist at the X-inactivation centre. Nature Genet 1999;21:400–404.
39. Madamanchi NR, Hu ZY, Li F, Horaist C, Moon SK, Patterson C, Runge MS. A noncoding RNA regulates human protease-activated receptor-1 gene during embryogenesis. Biochim Biophys Acta 2002;1576:237–245.
40. Li AW, Murphy PR. Expression of alternatively spliced FGF-2 antisense RNA transcripts in the central nervous system: regulation of FGF-2 mRNA translation. Mol Cell Endocrinol 2000;162:69–78.
41. Takeda K, Ichijo H, Fujii M, Mochida Y, Saitoh M, Nishitoh H, Sampath TK, Miyazono K. Identification of a novel bone morphogenetic protein-responsive gene that may function as a noncoding RNA. J Biol Chem 1998;273:17079–17085.
42. Hurst LD, Smith NG. Molecular evolutionary evidence that H19 mRNA is functional. Trends Genet 1999;15:134–135.
43. Nemes JP, Benzow KA, Koob MD. The SCA8 transcript is an antisense RNA to a brain-specific transcript encoding a novel actin-binding protein (KLHL1). Hum Mol Genet 2000;9:1543–1551.
44. Ridanpaa M, van Eenennaam H, Pelin K, Chadwick R, Johnson C, Yuan B, vanVenrooij W, Pruijn G, Salmela R, Rockas S, et al. Mutations in the RNA component of RNase MRP cause a pleiotropic human disease, cartilage-hair hypoplasia. Cell 2001;104:195–203.
45. Charlier C, Segers K, Wagenaar D, Karim L, Berghmans S, Jaillon O, Shay T, Weissenbach J, Cockett N, Gyapay G, Georges M. Human-ovine comparative sequencing of a 250-kb imprinted domain encompassing the callipyge (clpg) locus and identification of six imprinted transcripts: DLK1, DAT, GTL2, PEG11, antiPEG11, and MEG8. Genome Res 2001;11:850–862.
46. Hirotsune S, Yoshida N, Chen A, Garrett L, Sugiyama F, Takahashi S, Yagami K, Wynshaw-Boris A, Yoshiki A. An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. Nature 2003;423:91–96.
47. Harrison PM, Hegyi H, Balasubramanian S, Luscombe NM, Bertone P, Echols N, Johnson T, Gerstein M. Molecular fossils in the human genome: identification and analysis of the pseudogenes in chromosomes 21 and 22. Genome Res 2002;12:272–280.
48. Yelin R, Dahary D, Sorek R, Levanon EY, Goldstein O, Shoshan A, Diber A, Biton S, Tamir V, Khosravi R, et al. Widespread occurrence of antisense transcription in the human genome. Nature Biotechnol 2003;21:379–386.
49. Holmes R, Williamson C, Peters J, Members RGG, Wells C. A comprehensive transcript map of the mouse Gnas imprinted complex. Genome Res 2003;13: in press.
50. Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, Ha I, Baillie DL, Fire A, Ruvkun G, Mello CC. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell 2001;106:23–34.
51. Mourelatos Z, Dostie J, Paushkin S, Sharma A, Charroux B, Abel L, Rappsilber J, Mann M, Dreyfuss G. miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev 2002;16:720–728.
52. Llave C, Kasschau KD, Rector MA, Carrington JC. Endogenous and silencing-associated small RNAs in plants. Plant Cell 2002;14:1605–1619.
53. Pasquinelli AE, Ruvkun G. Control of developmental timing by microRNAs and their targets. Annu Rev Cell Dev Biol 2002;18:495–513.
54. Llave C, Xie Z, Kasschau KD, Carrington JC. Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science 2002;297:2053–2056.
55. Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP. MicroRNAs in plants. Genes Dev 2002;16:1616–1626.
56. Wassenegger M. RNA-directed DNA methylation. Plant Mol Biol 2000;43:203–220.
57. Matzke MA, Matzke AJ, Pruss GJ, Vance VB. RNA-based silencing strategies in plants. Curr Opin Genet Dev 2001;11:221–227.
58. Hutvagner G, Zamore PD. A microRNA in a multiple-turnover RNAi enzyme complex. Science 2002;297:2056–2060.
59. Hare MP, Palumbi SR. High intron sequence conservation across three Mammalian orders suggests functional constraints. Mol Biol Evol 2003;20:969–978.
60. Clement JQ, Qian L, Kaplinsky N, Wilkinson MF. The stability and fate of a spliced intron from vertebrate cells. RNA 1999;5:206–220.
61. Clement JQ, Maiti S, Wilkinson MF. Localization and stability of introns spliced from the Pem homeobox gene. J Biol Chem 2001;276:16919–16930.
62. Tycowski KT, Shu MD, Steitz JA. A mammalian gene with introns instead of exons generating stable RNA products. Nature 1996;379:464–466.
63. Cavaille J, Vitali P, Basyuk E, Huttenhofer A, Bachellerie JP. 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 2001;276:26374–26383.
64. Cavaille J, Buiting K, Kiefmann M, Lalande M, Brannan CI, Horsthemke B, Bachellerie JP, Brosius J, Huttenhofer A. Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. Proc Natl Acad Sci USA 2000;97:14311–14316.
65. Cavaille J, Seitz H, Paulsen M, Ferguson-Smith AC, Bachellerie JP. 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 2002;11:1527–1538.
66. Chanfreau G, Rotondo G, Legrain P, Jacquier A. Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. EMBO J 1998;17:3726–3737.
67. Allmang C, Petfalski E, Podtelejnikov A, Mann M, Tollervey D, Mitchell P. The yeast exosome and human PM-Scl are related complexes of 3′→5′ exonucleases. Genes Dev 1999;13:2148–2158.
68. Sharp PA. RNA interference-2001. Genes Dev 2001;15:485–490.
69. Snyder M, Gerstein M. Defining genes in the genomics era. Science 2003;300:258–260.
70. Finta C, Zaphiropoulos PG. Intergenic mRNA molecules resulting from trans-splicing. J Biol Chem 2002;277:5882–5890.
71. Pirrotta V. Trans-splicing in Drosophila. Bioessays 2002;24:988–991.
72. Fire A. RNA-triggered gene silencing. Trends Genet 1999;15:358–363.
73. Fagard M, Boutet S, Morel JB, Bellini C, Vaucheret H. AGO1, QDE-2, and RDE-1 are related proteins required for post- transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. Proc Natl Acad Sci USA 2000;97:11650–11654.
74. Gallagher RC, Pils B, Albalwi M, Francke U. Evidence for the role of PWCR1/HBII-85 C/D box small nucleolar RNAs in Prader-Willi syndrome. Am J Hum Genet 2002;71:669–678.
75. Pal-Bhadra M, Bhadra U, Birchler JA. Cosuppression in Drosophila: gene silencing of Alcohol dehydrogenase by white-Adh transgenes is Polycomb dependent. Cell 1997;90:479–490.
76. Birchler JA, Bhadra MP, Bhadra U. Making noise about silence: repression of repeated genes in animals. Curr Opin Genet Dev 2000;10:211–216.
77. Jones DO, Cowell IG, Singh PB. Mammalian chromodomain proteins: their role in genome organisation and expression. Bioessays 2000;22:124–137.

78. Shi Y, Berg JM. Specific DNA–RNA hybrid binding by zinc finger proteins. Science 1995;268:282–284.
79. Ladomery M. Multifunctional proteins suggest connections between transcriptional and post-transcriptional processes. Bioessays 1997;19: 903–909.
80. Bass BL. Double-stranded RNA as a template for gene silencing. Cell 2000;101:235–238.
81. Kim YG, Lowenhaupt K, Maas S, Herbert A, Schwartz T, Rich A. The Zab domain of the human RNA editing enzyme ADAR1 recognizes Z-DNA when surrounded by B-DNA. J Biol Chem 2000;275:26828–26833.
82. Akhtar A, Zink D, Becker PB. Chromodomains are protein–RNA interaction modules. Nature 2000;407:405–409.
83. Platero JS, Hartnett T, Eissenberg JC. Functional analysis of the chromo domain of HP1. EMBO J 1995;14:3977–3986.
84. Boggs BA, Cheung P, Heard E, Spector DL, Chinault AC, Allis CD. Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes. Nature Genet 2002;30:73–76.
85. Mochizuki K, Fine NA, Fujisawa T, Gorovsky MA. Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. Cell 2002;110:689–699.
86. Taverna SD, Coyne RS, Allis CD. Methylation of histone h3 at lysine 9 targets programmed DNA elimination in tetrahymena. Cell 2002;110: 701–711.
87. Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science 2002;297:1833–1837.
88. Hall IM, Shankaranarayana GD, Noma K, Ayoub N, Cohen A, Grewal SI. Establishment and maintenance of a heterochromatin domain. Science 2002;297:2232–2237.
89. Hall IM, Noma K, Grewal SI. RNA interference machinery regulates chromosome dynamics during mitosis and meiosis in fission yeast. Proc Natl Acad Sci USA 2003;100:193–198.
90. Watanabe T, Miyashita K, Saito TT, Yoneki T, Kakihara Y, Nabeshima K, Kishi YA, Shimoda C, Nojima H. Comprehensive isolation of meiosis-specific genes identifies novel proteins and unusual non-coding transcripts in Schizosaccharomyces pombe. Nucleic Acids Res 2001;29: 2327–2337.
91. Eddy SR. Non-coding RNA genes and the modern RNA world. Nature Rev Genet 2001;2:919–929.
92. Erdmann VA, Barciszewska MZ, Szymanski M, Hochberg A, de Groot N, Barciszewski J. The non-coding RNAs as riboregulators. Nucleic Acids Res 2001;29:189–193.
93. Lanz RB, McKenna NJ, Onate SA, Albrecht U, Wong J, Tsai SY, Tsai MJ, O'Malley BW. A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. Cell 1999;97:17–27.
94. Lanz RB, Razani B, Goldberg AD, O'Malley BW. Distinct RNA motifs are important for coactivation of steroid hormone receptors by steroid receptor RNA activator (SRA). Proc Natl Acad Sci USA 2002;99: 16081–16086.
95. Krause MO. Chromatin structure and function: the heretical path to an RNA transcription factor. Biochem Cell Biol 1996;74:623–632.
96. Smith CW, Valcarcel J. Alternative pre-mRNA splicing: the logic of combinatorial control. Trends Biochem Sci 2000;25:381–388.
97. Caceres JF, Kornblihtt AR. Alternative splicing: multiple control mechanisms and involvement in human disease. Trends Genet 2002;18: 186–193.
98. McKeown M. Alternative mRNA splicing. Annu Rev Cell Biol 1992;8: 133–155.
99. Kole R, Sazani P. Antisense effects in the cell nucleus: modification of splicing. Curr Opin Mol Ther 2001;3:229–234.
100. De Angelis FG, Sthandier O, Berarducci B, Toso S, Galluzzi G, Ricci E, Cossu G, Bozzoni I. Chimeric snRNA molecules carrying antisense sequences against the splice junctions of exon 51 of the dystrophin pre-mRNA induce exon skipping and restoration of a dystrophin synthesis in Delta 48-50 DMD cells. Proc Natl Acad Sci USA 2002;99: 9456–9461.
101. Sazani P, Gemignani F, Kang SH, Maier MA, Manoharan M, Persmark M, Bortner D, Kole R. Systemically delivered antisense oligomers upregulate gene expression in mouse tissues. Nature Biotechnol 2002;20: 1228–1233.
102. Tourriere H, Gallouzi IE, Chebli K, Capony JP, Mouaikel J, van der Geer P, Tazi J. RasGAP-associated endoribonuclease G3Bp: selective RNA degradation and phosphorylation-dependent localization. Mol Cell Biol 2001;21:7747–7760.
103. Albert R, Jeong H, Barabasi AL. Error and attack tolerance of complex networks. Nature 2000;406:378–382.
104. Judson HF. The Eighth Day of Creation: The Makers of the Revolution in Biology. New York: Simon and Schuster. 1979.