

PERSPECTIVE

The Eukaryotic Genome as an RNA Machine

Paulo P. Amaral, Marcel E. Dinger, Tim R. Mercer, John S. Mattick*

The past few years have revealed that the genomes of all studied eukaryotes are almost entirely transcribed, generating an enormous number of non-protein-coding RNAs (ncRNAs). In parallel, it is increasingly evident that many of these RNAs have regulatory functions. Here, we highlight recent advances that illustrate the diversity of ncRNA control of genome dynamics, cell biology, and developmental programming.

RNAs are an integral component of chromosomes and contribute to their structural organization (1, 2). It is now becoming apparent that chromatin architecture and epigenetic memory are regulated by RNA-directed processes that, although the exact mechanisms are yet to be understood, involve the recruitment of histone-modifying complexes and DNA methyltransferases to specific loci (3). Whereas long non-protein-coding RNAs (ncRNAs) have been classically implicated in the regulation of dosage compensation and genomic imprinting in animals (4), they seem to play a much broader role in the epigenetic control of developmental trajectories (3). For example, it was recently shown that 231 long ncRNAs associated with human *HOX* gene clusters are colinearly expressed along developmental axes (5), one of which, termed *HOTAIR*, transcribed from the *HOXC* locus, was studied in detail and found to recruit Polycomb complexes to repress gene expression in trans at the *HOXD* cluster (5) (Fig. 1). Other ncRNAs will likely perform similar functions, such as the intergenic transcripts from globin and antigen receptor loci, which have been associated with complex epigenetic phenomena (6, 7).

Small ncRNAs have been consistently linked with heterochromatin formation via the RNA interference (RNAi) pathway (8), including Piwi-interacting RNAs (piRNAs) (9), which guide PIWI family proteins to control transposon activity from flies to vertebrates (10). However, piRNAs might also regulate euchromatin formation, given that PIWI is required for establishing euchromatin in certain subtelomeric regions in *Drosophila* (11).

Higher-level nuclear organization and chromosome dynamics are also regulated by ncRNAs in a variety of systems. For example, the formation of the kinetochore and centromeric heterochromatin in fission yeast is dependent on cell cycle-regulated centromeric repeat-derived RNAs and the RNAi pathway, whereas kinetochore assembly and chromosome segregation require

the ribonuclease activity of a component of the exosome (12–15). These findings reveal an RNA-based mechanistic link between these processes in mitosis. In *Tetrahymena*, RNAs direct heterochromatin formation and DNA elimination via RNAi-dependent recruitment of Polycomb complexes and histone methylation (16). The RNAi pathway along with directed histone modifications also regulates the organization of the nucleolus in *Drosophila* (17).

Likewise, long ncRNAs direct programmed whole-genome rearrangements during ciliate dif-

ferentiation (18). In mammals, transcription of long ncRNAs contributes to various processes including T cell receptor recombination (7), maintenance of telomeres (19, 20), X-chromosome pairing required for dosage compensation (21) and inactive X-chromosome perinucleolar localization (22).

The functional organization of chromatin can also be regulated by ncRNAs derived from repetitive elements. In mice, bidirectional transcription of a retrotransposed SINE B2 sequence by RNA polymerase (RNAP)II and RNAPIII relocates the associated growth hormone locus into nuclear compartments and locally defines the heterochromatin-euchromatin boundary, regulating the expression of the gene during organogenesis (23) (Fig. 1). Given the abundance of transcribed repetitive sequences, this may represent a genome-wide strategy for the control of chromatin domains that may be conserved throughout eukaryotes (23–25). Moreover, such observations and others suggest that a large portion of the genome may, in fact, be functionally active and that transposon-derived sequences may not be reliable indices of the rate of neutral evolution (26).

Transcription

Noncoding RNAs can regulate transcription by interacting with transcription factors, RNAP, or

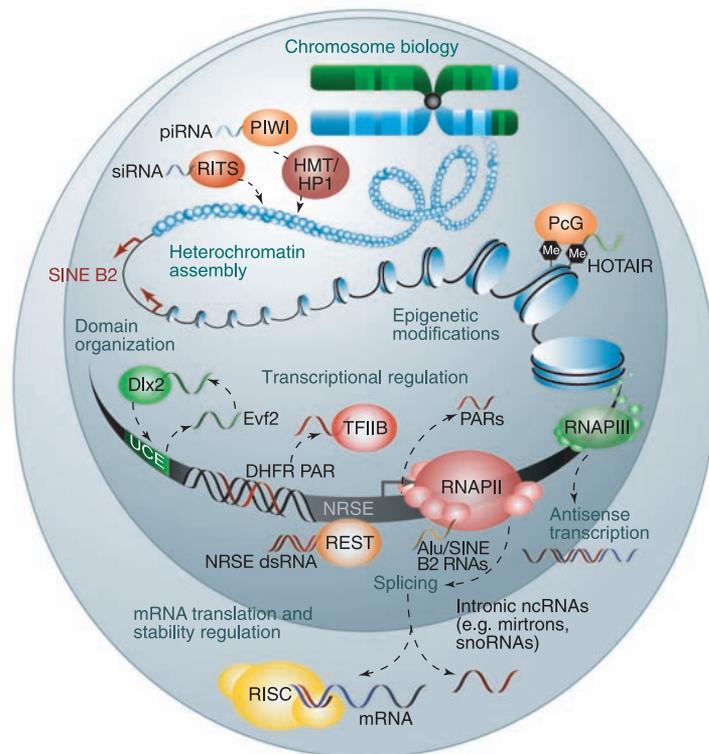


Fig. 1. Recent examples of the various levels of regulation of eukaryotic gene expression and cell biology by ncRNAs. dsRNA, double-stranded RNA; HMT, histone methyltransferases; HP1, heterochromatin protein 1; PARs, promoter-associated RNAs; PcG, Polycomb group proteins; RISC, RNA-induced silencing complex; RITS, RNA-induced initiation of transcriptional gene silencing; siRNA, small interfering RNA; TFIIIB, transcription factor IIB; and UCE, ultraconserved element. See text for details, other acronyms, and references. For additional examples, see (3, 58).

Institute for Molecular Bioscience, University of Queensland, St. Lucia QLD 4072, Australia.

*To whom correspondence should be addressed. E-mail: j.mattick@imb.uq.edu.au

Gene Regulation

DNA itself (Fig. 1). The small double-stranded RNA *NRSE* directs the activation of neuronal genes containing the upstream conserved *NRSE* sequence by triggering the REST transcriptional machinery (27), components of which are also targets of microRNAs (miRNAs) (28). Analogously, the ncRNA *Eyf-2* is transcribed from an ultraconserved enhancer associated with the *Dlx-5/6* locus and interacts with *Dlx-2* transcription factor to induce enhancer activity (29). Indeed, it appears that most ultraconserved elements in the human genome (sequences that, in many cases, act as enhancers and have remained essentially unchanged throughout amniote evolution) (30) are transcribed in a regulated manner whose aberrant expression may be involved in pathological processes such as cancer (31).

In addition, promoter-directed sequence-specific RNAs, including miRNAs (32), have been shown to induce (32, 33) or repress transcription (34, 35), the latter involving epigenetic modification and targeting of promoter-associated low-copy RNA (34, 35) [in one case by triplex formation (35), which appears common in human chromatin (36)]. The abundance of promoter-associated transcripts in the human (37, 38), *Arabidopsis* (39), and yeast (40) genomes suggests that such examples may represent widespread phenomena.

RNAPIII can also produce various types of regulatory RNAs. In mice, heat shock induces RNAPIII transcription of a B2 SINE element producing a structured modular RNA that represses RNAPII in trans at specific loci (41). These same properties and function in human cells under heat shock were identified for ncRNAs derived from Alu elements, primate-specific SINEs that comprise 10.5% of our genome, and show a striking convergent evolution of SINE function as trans-repressors of gene expression (42).

RNA Processing and Translation

MiRNAs regulate a wide range of processes in animals and plants by directing translational repression or degradation of mRNAs (Fig. 1), with consequent effects on global regulatory circuits (28). However, miRNAs can also promote activation of gene expression under conditions such as stress, depending on the association with regulatory factors (43), and even activate translation during cell cycle arrest (44).

The formation and target specificity of miRNAs can be modulated by tissue-specific A-I editing (45) and RNA binding proteins (46, 47). Furthermore, miRNAs may be differentially processed from both sense and antisense strands of the same hairpin or from sense and antisense transcripts from the same locus (48, 49), which expands the potential of a single genomic locus to generate multiple miRNAs with different targets.

In addition, many miRNAs derive from introns of protein-coding genes, in some cases by splicing rather than the canonical Drosha pathway (50–52), thus, along with small nucleolar RNAs

(snoRNAs), making up intronic parallel outputs of gene expression (3), which may be much more extensive than expected.

Other classes of ncRNAs can also act post-transcriptionally. In a recent report, a number of RNAPIII-transcribed short ncRNAs were found to have sequence complementarity to protein-coding genes and provided evidence for the existence of a sense-antisense-based regulatory network wherein RNAPIII transcripts control their RNAPII counterparts (Fig. 1) (53).

A World of Noncoding RNAs

The examples above provide proof-of-principle that RNA can regulate gene expression at many levels and by using a wide array of mechanisms (Fig. 1). The ENCODE project showed that at least 93% of analyzed human genome nucleotides are transcribed in different cells (54), with similar findings in mouse (55) and other eukaryotes, which indicate that there may be a vast reservoir of biologically meaningful RNAs that could greatly exceed the ~1.2% encoding proteins. A fraction of RNAs with short open reading frames (ORFs) potentially encodes peptides (56) but on the other side of the ledger many currently annotated ORFs are not conserved and may be false, which could reduce the number of protein-coding genes in the human genome (57).

There has been debate about whether these ncRNAs are (in the main) functional or simply noise. In some cases, it may be the transcript or merely the act of transcription, or both (35), that are relevant. Nevertheless, many observations indicate that substantial numbers of ncRNAs are intrinsically functional. These include the fact that many loci produce spliced (and alternatively spliced) transcripts that are developmentally regulated (5, 55, 58) and that at least some antisense and intergenic ncRNAs can function in trans (5, 41, 53, 59). A large fraction of ncRNAs are expressed in specific regions of the brain, exhibiting precise cellular locations (60). Some mark new domains within the cell (61, 62), which means that ncRNAs are also set to have a major impact in cell biology.

Comparative analyses indicate that ncRNA promoters are, on average, more conserved than those of protein-coding genes (55) and that ncRNA sequences, secondary structures, and splice site motifs have been subject to purifying selection (63, 64). Moreover, many ncRNAs are evolving quickly (65), and some have undergone recent positive selection, as exemplified by *HARI* RNA expressed in the human brain, which contains the sequence conserved in mammals that most rapidly diverged after the human-chimpanzee separation (66).

Given the functional versatility of RNAs, it is plausible that ncRNAs have represented a rich substrate for evolutionary innovations in eukaryotes. In support of this idea, regulatory RNAs are centrally involved in the ontogeny of many orga-

nisms, from unique developmental pathways in protozoa (18) to the control of conserved or clade-specific developmental regulators in multicellular animals (5, 52). In addition, there is mounting evidence that RNA can transmit information intergenerationally, so as to mediate non-Mendelian inheritance of epigenetic changes in mice (67) and plants (68).

Although the need for large-scale approaches to explore the function of ncRNAs is evident, a glance at the genome browser will show non-coding expressed sequence tags associated with most genes of interest that may have regulatory functions. As an example, a recent study of ncRNAs associated with tumor suppressor genes focused on an RNA antisense to the *p15* gene and showed that it acts to alter histone methylation to silence expression of this gene, with important implications for cell differentiation and tumorigenesis (59).

Indeed, ncRNAs are already being identified as markers for cancer (69, 70) and associated with other complex diseases such as coronary disease and diabetes (71, 72). The elucidation of their function may significantly contribute to the understanding and treatment of such conditions. It may also transform our understanding of the genetic programming of multicellular organisms, particularly as it appears that regulation dominates the information content of complex systems (3).

References and Notes

1. J. A. Nickerson, G. Krochmalnik, K. M. Wan, S. Penman, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 177 (1989).
2. A. Rodríguez-Campos, F. Azorin, *PLoS ONE* **2**, e1182 (2007).
3. J. S. Mattick, *J. Exp. Biol.* **210**, 1526 (2007).
4. P. K. Yang, M. I. Kuroda, *Cell* **128**, 777 (2007).
5. J. L. Rinn *et al.*, *Cell* **129**, 1311 (2007).
6. H. L. Ashe, J. Monks, M. Wijgerde, P. Fraser, N. J. Proudfoot, *Genes Dev.* **11**, 2494 (1997).
7. I. Abarrategui, M. S. Krangel, *EMBO J.* **26**, 4380 (2007).
8. M. Buhler, D. Moazed, *Nat. Struct. Mol. Biol.* **14**, 1041 (2007).
9. B. Brower-Toland *et al.*, *Genes Dev.* **21**, 2300 (2007).
10. A. A. Aravin, G. J. Hannon, J. Brennecke, *Science* **318**, 761 (2007).
11. H. Yin, H. Lin, *Nature* **450**, 304 (2007).
12. M. Buhler, W. Haas, S. P. Gygi, D. Moazed, *Cell* **129**, 707 (2007).
13. H. Murakami *et al.*, *PLoS ONE* **2**, e317 (2007).
14. H. D. Folco, A. L. Pidoux, T. Urano, R. C. Allshire, *Science* **319**, 94 (2008).
15. E. S. Chen *et al.*, *Nature* **451**, 734 (2008).
16. Y. Liu *et al.*, *Genes Dev.* **21**, 1530 (2007).
17. J. C. Peng, G. H. Karpen, *Nat. Cell Biol.* **9**, 25 (2007).
18. M. Nowacki *et al.*, *Nature* **451**, 153 (2008).
19. S. Schoeffer, M. A. Blasco, *Nat. Cell Biol.* **10**, 228 (2008).
20. C. M. Azzalin, P. Reichenbach, L. Khorai, E. Giulotto, J. Lingner, *Science* **318**, 798 (2007).
21. N. Xu, M. E. Donohoe, S. S. Silva, J. T. Lee, *Nat. Genet.* **39**, 1390 (2007).
22. L. F. Zhang, K. D. Huynh, J. T. Lee, *Cell* **129**, 693 (2007).
23. V. V. Lunyak *et al.*, *Science* **317**, 248 (2007).
24. K. Noma, H. P. Cam, R. J. Maraia, S. I. Grewal, *Cell* **125**, 859 (2006).
25. D. A. Willoughby, A. Vilalta, R. G. Oshima, *J. Biol. Chem.* **275**, 759 (2000).
26. M. Pheasant, J. S. Mattick, *Genome Res.* **17**, 1245 (2007).
27. T. Kuwabara, J. Hsieh, K. Nakashima, K. Taira, F. H. Gage, *Cell* **116**, 779 (2004).

28. E. V. Makeyev, T. Maniatis, *Science* **319**, 1789 (2008).
29. J. Feng *et al.*, *Genes Dev.* **20**, 1470 (2006).
30. S. Stephen, M. Pheasant, I. V. Makunin, J. S. Mattick, *Mol. Biol. Evol.* **25**, 402 (2008).
31. G. A. Calin *et al.*, *Cancer Cell* **12**, 215 (2007).
32. R. F. Place, L. C. Li, D. Pookot, E. J. Noonan, R. Dahiya, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1608 (2008).
33. L. C. Li *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17337 (2006).
34. J. Han, D. Kim, K. V. Morris, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12422 (2007).
35. I. Martianov, A. Ramadass, A. Serra Barros, N. Chow, A. Akoulitchev, *Nature* **445**, 666 (2007).
36. M. Ohno, T. Fukagawa, J. S. Lee, T. Ikemura, *Chromosoma* **111**, 201 (2002).
37. P. Kapranov *et al.*, *Science* **316**, 1484 (2007).
38. M. G. Guenther, S. S. Levine, L. A. Boyer, R. Jaenisch, R. A. Young, *Cell* **130**, 77 (2007).
39. J. A. Chekanova *et al.*, *Cell* **131**, 1340 (2007).
40. C. A. Davis, M. Ares Jr., *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3262 (2006).
41. C. A. Espinoza, J. A. Goodrich, J. F. Kugel, *RNA* **13**, 583 (2007).
42. P. D. Mariner *et al.*, *Mol. Cell* **29**, 499 (2008).
43. A. K. Leung, P. A. Sharp, *Cell* **130**, 581 (2007).
44. S. Vasudevan, Y. Tong, J. A. Steitz, *Science* **318**, 1931 (2007).
45. Y. Kawahara *et al.*, *Science* **315**, 1137 (2007).
46. M. Kedde *et al.*, *Cell* **131**, 1273 (2007).
47. S. R. Viswanathan, G. Q. Daley, R. I. Gregory, *Science*, 21 February 2008 (10.1126/science.1154040).
48. D. M. Tyler *et al.*, *Genes Dev.* **22**, 26 (2008).
49. A. Stark *et al.*, *Genome Res.* **17**, 1865 (2007).
50. J. G. Ruby, C. H. Jan, D. P. Bartel, *Nature* **448**, 83 (2007).
51. K. Okamura, J. W. Hagen, H. Duan, D. M. Tyler, E. C. Lai, *Cell* **130**, 89 (2007).
52. E. Berezikov, W. J. Chung, J. Willis, E. Cuppen, E. C. Lai, *Mol. Cell* **28**, 328 (2007).
53. A. Pagano *et al.*, *PLoS Genet.* **3**, e1 (2007).
54. E. Birney *et al.*, *Nature* **447**, 799 (2007).
55. P. Carninci *et al.*, *Science* **309**, 1559 (2005).
56. M. C. Frith *et al.*, *PLoS Genet.* **2**, e52 (2006).
57. M. Clamp *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 19428 (2007).
58. K. V. Prasanth, D. L. Spector, *Genes Dev.* **21**, 11 (2007).
59. W. Yu *et al.*, *Nature* **451**, 202 (2008).
60. T. R. Mercer, M. E. Dinger, S. M. Sunkin, M. F. Mehler, J. S. Mattick, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 716 (2008).
61. M. Sone *et al.*, *J. Cell Sci.* **120**, 2498 (2007).
62. H. Royo *et al.*, *Mol. Biol. Cell* **18**, 2817 (2007).
63. E. Torarinsson *et al.*, *Genome Res.* **18**, 242 (2008).
64. J. Ponjavic, C. P. Ponting, G. Lunter, *Genome Res.* **17**, 556 (2007).
65. K. C. Pang, M. C. Frith, J. S. Mattick, *Trends Genet.* **22**, 1 (2006).
66. K. S. Pollard *et al.*, *Nature* **443**, 167 (2006).
67. M. Rassoulzadegan *et al.*, *Nature* **441**, 469 (2006).
68. C. J. Hale, J. L. Stonaker, S. M. Gross, J. B. Hollick, *PLoS Biol.* **5**, 2156 (2007).
69. R. Lin, S. Maeda, C. Liu, M. Karin, T. S. Edgington, *Oncogene* **26**, 851 (2007).
70. D. S. Perez *et al.*, *Hum. Mol. Genet.* **17**, 642 (2007).
71. N. Ishii *et al.*, *J. Hum. Genet.* **51**, 1087 (2006).
72. H. M. Broadbent *et al.*, *Hum. Mol. Genet.* **17**, 806 (2007).
73. The authors acknowledge the support of the Australian Research Council and the New Zealand Foundation for Research Science and Technology.

10.1126/science.1155472

PERSPECTIVE

Multilevel Regulation of Gene Expression by MicroRNAs

Eugene V. Makeyev and Tom Maniatis

MicroRNAs (miRNAs) are ~22-nucleotide-long noncoding RNAs that normally function by suppressing translation and destabilizing messenger RNAs bearing complementary target sequences. Some miRNAs are expressed in a cell- or tissue-specific manner and may contribute to the establishment and/or maintenance of cellular identity. Recent studies indicate that tissue-specific miRNAs may function at multiple hierarchical levels of gene regulatory networks, from targeting hundreds of effector genes incompatible with the differentiated state to controlling the levels of global regulators of transcription and alternative pre-mRNA splicing. This multilevel regulation may allow individual miRNAs to profoundly affect the gene expression program of differentiated cells.

There is a surprisingly small difference in the number of protein-encoding genes between organisms of vastly different morphological and behavioral complexity. A plausible explanation for this paradox may be in the increased elaboration of gene regulatory networks at the levels of transcription (1) and alternative pre-mRNA splicing (2). More recently, miRNAs have been proposed to play a role in the expansion of organismal complexity (3). Indeed, some miRNAs are expressed in a cell- or tissue-specific manner during embryonic development, suggestive of a role in cellular differentiation (4) [also see essay by O. Hobert in this issue (5)]. Certain of these tissue-specific miRNAs, such as miR-1 and miR-124, which are expressed in muscle cells and neurons, respectively, have been shown to

stimulate differentiation of the corresponding cell types (6–8).

Targeting Gene Batteries

Several distinct molecular mechanisms may underlie the biological functions of miRNAs. Individual miRNAs can repress large sets of mRNAs that are not required at a particular developmental stage (9, 10). A number of these miRNA targets fall into the category of “gene batteries,” sets of functionally related effector genes that represent outputs of gene regulatory networks (11). Indeed, when miR-124 is introduced into nonneuronal mammalian cells a preferential reduction in the amounts of multiple nonneuronal mRNAs, for example, those encoding proteins required for cell proliferation or neural stem cell function, is observed (7, 9, 12). Conversely, depletion of miR-124 from primary neurons leads to the accumulation of a number of nonneuronal mRNA targets (13). A similar mode of action has been reported

for miR-1 (9). Thus, in cells undergoing differentiation, miRNAs can effectively deplete unwanted mRNAs left over from progenitor cells. Furthermore, miRNAs may regulate a fraction of mRNA targets at the level of translation without affecting their stability (14, 15).

Targeting Regulators of Transcription

MiRNAs can also control the expression of critical transcriptional regulators, as first shown by the repression of transcription factor lin-14 by miRNA lin-4 (16, 17). Another example is miR-124, which targets mRNA of the small C-terminal domain phosphatase 1 (SCP1/CTDSP1) protein, a component of the RE1-silencing transcription repressor (REST/NRSF) (6) (Fig. 1). REST represses the transcription of a large number of neuron-specific genes in nonneuronal cells (13). Thus, by reducing SCP1 expression in differentiating neurons, miR-124 may facilitate the derepression of the neuronal transcription program (6). The REST complex may also repress miR-124 gene expression and thus establish a double-negative feedback loop between miR-124 and the REST complex (6, 13) (Fig. 1). Similarly, miR-1 and another muscle-specific miRNA, miR-133, regulate several transcription factors essential for muscle development (4, 8).

Targeting Regulators of Alternative Splicing

Recent studies have identified yet another activity of miRNAs: the induction of large-scale changes in gene expression by targeting global regulators of alternative pre-mRNA splicing (7, 18). In early muscle cell precursors, a repressor of alternative splicing called polypyrimidine tract-binding protein 1 (PTBP1/PTB/hnRNP-1) and its homolog, PTBP2 (nPTB/brPTB/PTBLP), repress the inclusion of a number of muscle-specific exons into mature mRNAs. However, upon myotube differentiation, PTBP1 and PTBP2

Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA. E-mail: makeyev@mcb.harvard.edu (E.V.M.); maniatis@mcb.harvard.edu (T.M.)