

MINI REVIEW

RNA world – the dark matter of evolutionary genomics

P. MICHALAK

*Department of Biology, The University of Texas at Arlington, Arlington, TX, USA**Keywords:*

DNA microarrays;
 fitness;
 microRNAs;
 noncoding RNA;
 transcriptional profiling;
 transcriptomics.

Abstract

For a long time, molecular evolutionary biologists have been focused on DNA and proteins, whereas RNA has lived in the shadow of its famous chemical cousins as a mere intermediary. Although this perspective has begun to change since genome-wide transcriptional profiling was successfully extended to evolutionary biology, it still echoes in evolutionary literature. In this mini-review, new developments of RNA biochemistry and transcriptomics are brought to the attention of evolutionary biologists. In particular, the unexpected abundance and functional significance of noncoding RNAs is briefly reviewed. Noncoding RNAs control a remarkable range of biological pathways and processes, all with obvious fitness consequences, such as initiation of translation, mRNA abundance, transposon jumping, chromosome architecture, stem cell maintenance, development of brain and muscles, insulin secretion, cancerogenesis and plant resistance to viral infections.

Introduction

Since their introduction in the mid-1990s, microarrays have rapidly become established as an essential method of gene expression analysis in relation to physiology, development and disease. Beside high-throughput methods of DNA sequencing and analysis, the emergence of microarrays and transcriptomics is an event probably most commonly associated with the transition from conventional genetics to genomics. Recently, genome-wide surveys of transcription have made their appearance in evolutionary biology and ecology as well (reviewed by Gibson, 2002 and Stearns & Magwene, 2003; Ranz & Machado, 2006). New developments of transcriptomics, such as tiling arrays for whole genome analyses (Mockler & Ecker, 2005), are capitalizing on and – at the same time – largely contributing to the spectacular findings in the area of RNA molecular biology and RNA biochemistry.

There is a widespread view among biologists that all that really matters in the functional landscape shaping fitness is proteins, and transcript abundance can be useful as a mere proxy for the activity of the corresponding proteins, as exemplified by a recent review about limitations of transcriptomics (Feder & Walser, 2005).

This conviction originates from the central dogma of molecular biology: information flows from DNA to protein through its intermediary-RNA. From there, it is easy to generalize that ‘genes’ are ‘synonymous’ with proteins. Generally, it holds true in prokaryotes whose genomes consist of tightly packed protein-coding sequences. But there is increasing evidence that complex eukaryotes, whose genomes contain relatively rare coding sequences, have dramatically different patterns of programming and functional regulation.

We are closer than ever before to doing away with the central dogma, or at least its common generalization that ‘one gene equals one protein, one function’.

In 1969, Britten & Davidson (1969) suggested that RNA specifies which genes are turned on or off in eukaryotic cells. However, their idea that the gene regulation could be based on the complementarity rules of Watson and Crick and driven by diverse arrays of RNAs was abandoned after the subsequent discovery of protein transcription factors. But in fact RNAs, specifically various noncoding RNAs, do control plant and animal gene expression. The meaning of ‘gene expression’ and ‘gene’ themselves await a major reevaluation. To truly understand eukaryotic gene expression, a great deal more is involved than just transcription factors: noncoding RNAs (microRNAs, associated small interfering RNAs, small interfering RNAs and other unclassified non-translated RNAs), epigenetic memory (e.g. CpG

Correspondence: Pawel Michalak, Department of Biology, The University of Texas-Arlington, Box 19498, Arlington, TX 76010, USA.
 Tel.: (817) 272 2979; fax: (817) 272 2855; e-mail: michalak@uta.edu

dinucleotide methylation, histone acetylation and deacetylation), and mRNA transcript variation (e.g. alternative promoters, alternative 3' polyadenylation, differentially spliced exons and introns; e.g. Moore, 2005).

This review is by no means comprehensive. Instead, it focuses on selected examples of noncoding RNAs and their genome regulation roles; more comprehensive reviews may be found elsewhere, including a special issue of *Science* (vol. 309, September 2, 2005; see also Mattick, 2001, 2003, 2004; Costa, 2005). The paucity of the evolutionary context here reflects real scarcity of studies (but see Andolfatto, 2005) rather than the author's disinterest in evolution.

New era in transcriptomics

During the early phases of the Human Genome Project, there was a race to see who could correctly count the total number of genes, and online bids predicting the number of human protein-coding genes ranged from 30 000 to 150 000 (Editorial, 2000, *Nature Genetics*). The first bioinformatic estimates indicated—to the surprise of many—no more than 35 000 human genes, an estimate that has declined to the present 25 000 (IHGSC, 2004). There has been a clear discrepancy between the numbers of genes estimated from sequencing cDNA libraries and those derived from annotations of genomic sequences (Claverie, 2001, 2005). The genomics industry, in particular, has capitalized on sequencing millions of expressed sequence tags (ESTs); for example, Incyte Genomics came up with an estimate of 140 000 human genes, a number that did not include an additional 200 000 ESTs that were less common in the cDNA libraries (Claverie, 2005).

The solution to this paradox appears quite unexpected: recent large-scale studies based on both cDNA cloning and the interrogation of genome tiling arrays suggest that it is our traditional view of genome regulation in animals and plants that is incorrect. The startling finding of the new studies has been that a large proportion, if not a majority, of the transcriptional output has been nonprotein coding RNA (ncRNA). In *Drosophila melanogaster*, molecular studies of well-known genes have long ago demonstrated the existence of intronic expression (e.g. the bithorax complex, Lipshitz *et al.*, 1987; Sanchez-Herrero & Akam, 1989). Such findings have been regarded as exceptions to the rule that introns are nonfunctional evolutionary debris until the application of genomic tiling arrays to *Arabidopsis*, *Drosophila* and humans revealed massive expression in intronic and intergenic regions (Singh-Gasson *et al.*, 1999; Wong *et al.*, 2001; Kapranov *et al.*, 2002; Yamada *et al.*, 2003; Bertone *et al.*, 2004; Cawley *et al.*, 2004; Kampa *et al.*, 2004; Stolc *et al.*, 2004; Cheng *et al.*, 2005; Kapranov *et al.*, 2005). Whole-genome tiling arrays that allow interrogation of genome sequences at fixed intervals irrespective of predicted coding regions open a new era in

transcriptomics. High density oligonucleotide tiling arrays contain relatively short (<100-mer) probes synthesized directly on the surface of the arrays by photolithography using light-sensitive synthetic chemistry and photolithographic masks, but mask-less methods are also available (Singh-Gasson *et al.*, 1999; Mockler & Ecker, 2005). The emergence of high-density DNA tiling arrays enables not only unbiased characterization of the entire transcriptome (including regulatory and other noncoding RNAs), but also novel gene discovery, analysis of alternative splicing, mapping of regulatory DNA motifs using the chromatin immunoprecipitation assay (ChIP on chip), genome-wide DNA methylation analysis, genotyping and resequencing (reviewed by Kapranov *et al.*, 2003 and Mockler & Ecker, 2005).

Although protein encoding sequences constitute only about 1.5% of the human genome, a very large proportion of the genome is in fact expressed (Wong *et al.*, 2001; Bertone *et al.*, 2004). For example, Cheng *et al.* (2005) used a tiling array interrogating 10 human chromosomes with 5-nucleotide resolution and found that an average of 10% of the genome (compared with <2% of bona fide exonic sequences) resulted in polyadenylated transcripts, of which more than a half do not overlap with known locations. The FANTOM 3 project (FANTOM 2005 and RIKEN 2005) has revealed that as much as 62% of the mouse genome is transcribed. The consortium has identified >181,000 transcripts, of which half consist of ncRNA, including intronic sequences. It seems that rather than post-splicing-degraded and recycled, as previously believed, intronic RNAs are active players in RNA-mediated transactions within the cell (Mattick, 2003, 2004).

A brief history of small RNAs

Remarkably, it has been recognized that most small nucleolar RNAs (snoRNAs) in higher eukaryotes are derived from introns. For example, 203 of 212 snoRNAs in *Drosophila melanogaster* analysed by Huang *et al.* (2005) were intron-derived. There are two main families of snoRNPs: the box C/D snoRNPs, which mediate the formation of large numbers of 2'-O-methylated nucleotides in ribosomal RNA (rRNA), and the box H/ACA snoRNPs, responsible for the synthesis of large numbers of pseudouridines (Bertrand & Fournier, 2004). Members of both families are required for cleavage of pre-rRNA, and they have also been implicated in processing mRNAs and the control of epigenetic imprinting (Cavaille *et al.*, 2000, 2002).

Another abundant family of short (20–25 nucleotides) ncRNAs, known as microRNAs, provides an even more spectacular example of regulatory transcripts. The microRNAs originate from sequences that are antisense to known genes and they are sourced from 'intergenic regions' (i.e. sequences between annotated protein-coding genes), introns, and precursors that have been

previously identified as ncRNA (Lau *et al.*, 2001; Lagos-Quintana *et al.*, 2002). Genes encoding microRNAs are typically transcribed by RNA polymerase II, then processed into hairpins, and exported to the cytoplasm, where they are cleaved by a double-stranded RNA-specific endonuclease, Dicer, and RDE homologs, critical enzymes of the RNA interference (RNAi) pathway (Ambros, 2004; Bartel, 2004). In animals, microRNAs are thought to target partially complementary binding sites at mRNAs and repress their translation. Mainly in plants, the microRNA-RISC (RNA-induced silencing complex) induces silencing through specific mRNA cleavage. When microRNAs pair only partially with their targets, they cannot direct mRNA cleavage. Instead, translation is repressed through recruitment of mRNA to so-called processing bodies (P bodies) where they are either stored or degraded (Pillai *et al.*, 2005; Zamore & Haley, 2005). There are more than 1500 known human microRNAs and the number of newly discovered ones is growing rapidly (Bentwich *et al.*, 2005; Mattick & Makunin, 2005; Zamore & Haley, 2005). Three independent studies based on cross-species comparisons of eight vertebrates predicted that microRNAs regulate at least 20–30% of all genes (Krek *et al.*, 2005; Lewis *et al.*, 2005; Xie *et al.*, 2005).

The discovery of microRNAs has been preceded by the characterization of small interfering RNAs (siRNAs), which along with the Dicer endonuclease and small RNA-binding proteins called Argonautes are the key players in the RNAi pathway shared with microRNAs (Tabara *et al.*, 1999; Hammond *et al.*, 2001). In contrast to microRNAs that originate from long intergenic transcripts, siRNAs derive from shorter (hundreds to thousands of base pairs) double-stranded RNA. In addition to mRNA breakdown, siRNAs can affect the chromatin structure of targeted genes, resulting in transcriptional silencing through the formation of heterochromatin. How do siRNAs convert DNA to heterochromatin? Thus far, there has been no convincing evidence that siRNAs directly pair with DNA. Instead, it seems that they guide DNA-modifying enzymes through a larger protein complex including the Argonaute family (Volpe *et al.*, 2002; Zamore & Haley, 2005).

A briefer history of larger RNAs

Yet, there are reasons to believe that snoRNAs, microRNAs and siRNAs are just the tip of an iceberg. Around 97–98% of the transcriptional output of the human genome is predicted to consist of ncRNAs that largely elude easy classification (Mattick, 2001). Those ncRNAs include both polyadenylated and nonpolyadenylated transcripts, often comprising intertwined and overlapping networks from both strands, whereby the same sequences may produce differently processed sense and antisense transcripts (Cheng *et al.*, 2005; Frith *et al.*, 2005; Kapranov *et al.*, 2005). For example, Kampa *et al.* (2004) have shown that 11% of the transcriptional

output from known exons of human chromosomes 21 and 22 was antisense. The distinction between mRNA and ncRNA becomes fuzzy; so, inevitably, does the concept of a 'gene'. Many ncRNAs are unprocessed primary transcripts. Some originate from the exons of spliced transcripts that may also be alternatively spliced, and polyadenylated (or not) in a developmentally regulated manner (Mattick, 2003). A substantial proportion of exonic sequences were shown to exhibit antisense transcription in parallel to the production of exonic-encoded proteins.

A number of examples of functional ncRNA other than small RNAs (snoRNAs, microRNAs, and siRNAs) is presented in Table 1 (see Costa, 2005, for a more comprehensive review). An elegant example of large-scale screening of ncRNAs has recently been provided by Willingham *et al.* (2005). They analysed 512 ncRNAs from the RIKEN Fantom2 mouse cDNA library (Okazaki *et al.*, 2002; Numata *et al.*, 2003) that showed significant conservation with human sequences, and created an arrayed library of short hairpin RNAs (shRNAs) directed against these sequences. Via cell-based assays, they identified eight functional ncRNAs: six essential for cell viability, one repressor of Hedgehog signalling and one (NRON) that acts as a repressor of a transcription factor (NFAT) that is critical in T-cell receptor-mediated immune response and the development of various tissues. A previous screening of the RIKEN collection (Okazaki *et al.*, 2002) of 60 777 full-length cDNAs provided 33 409 transcription units, of which 15 815 did not contain sizable open reading frames, and thus most likely represented ncRNAs. Among those there were 2400 pairs of overlapping sense–antisense transcripts, of which nearly 1600 had not previously been known (Okazaki *et al.*, 2002).

It is becoming increasingly obvious that this vast contribution of ncRNA to the genomic output is not just a transcriptional 'background noise' or even an accessory toolkit of idiosyncratic functions. Rather, it constitutes a highly organized regulatory network of RNA signalling that works in concert with the protein network. Many transcription factors have been shown to exert affinity for RNA or for higher order nucleic acid structures containing RNA. For example, the zinc finger house-keeping Sp1 has comparable or higher affinity for RNA–DNA hybrids than for double-stranded DNA (Shi & Berg, 1995). RNA signalling also intersects with epigenetic memory and controls chromatin configuration. There is a clear link between antisense RNA and epigenetic gene regulation in eukaryotes; DNA methylation in plants and imprinting in animals is RNA-guided (Kelley & Kuroda, 2000; Wassenegger, 2000; Gallagher *et al.*, 2002). RNA signalling is also involved in the proper regulation of mitosis and meiosis (Hall *et al.*, 2003), heterochromatin formation (Taverna *et al.*, 2002; Volpe *et al.*, 2002), initiation of transcription (Lanz *et al.*, 2002) and alternative splicing (Mattick, 2003). In the light of the new

Table 1 Examples of functional noncoding RNAs.

Gene	Organism	Function	References
<i>Pgc</i>	<i>Drosophila</i>	Germ cell formation and migration	Nakamura <i>et al.</i> (1996), Martinho <i>et al.</i> (2004)
7SK	Mammals	Negative regulation of pol II transcription	Zieve & Penman (1976), Blencowe (2002)
<i>roX1/2</i>	Insects	X chromosome dosage compensation	Franke & Baker (1999)
<i>Xist</i>	Mammals	X chromosome dosage compensation	Lee <i>et al.</i> (1999)
<i>DISC1</i>	Human	Schizophrenia related	Hodgkinson <i>et al.</i> (2004)
ncR-uPAR	Human	Regulation of protease-activated receptor-1 gene	Madamanchi <i>et al.</i> (2002)
<i>FGF-AS</i>	Mammals	Regulation of fibroblast growth factor-2	Li & Murphy (2000)
<i>H19</i>	Human	Tumour suppressor features	Scott <i>et al.</i> (2005)
<i>BORG</i>	Mouse	Bone differentiation	Takeda <i>et al.</i> (1998)
<i>RIAN</i>	Mouse	Genomic imprinting	Hatada <i>et al.</i> (2001)
<i>GADD7</i>	Hamster	Tumour suppressor features	Hollander <i>et al.</i> (1996)
<i>KHPS1a</i>	Human	DNA demethylation	Imamura <i>et al.</i> (2004)
<i>SRA</i>	Human	Transcription co-activator	Lanz <i>et al.</i> (2002)
<i>AIR</i>	Human	Genomic imprinting	Sleutels <i>et al.</i> (2002)

findings, it should perhaps not be surprising that a large fraction of the nontranslated genome in *Drosophila melanogaster* has been found to be subject to both purifying and positive selection (Andolfatto, 2005).

The emerging model of genome regulation implies that transcriptional and post-transcriptional silencing occurs through interactions between small RNAs and mRNA, where 'messenger' RNA acquires a dramatically different meaning compared with its traditional understanding. This model requires transcription across genomic regions that were thought to lie untranscribed, such as promoters, intronic and intergenic regions. Unexpectedly, there is a lot of RNA polymerization that happens just to destroy other RNA. In plants, transcriptionally silenced DNA is transcribed by a unique type of DNA-dependent RNA polymerase, RNA polymerase IV (Herr *et al.*, 2005; Kanno *et al.*, 2005; Onodera *et al.*, 2005). This polymerase appears to transcribe 'silent' heterochromatic regions, thus providing a steady-state source of primary transcripts used to manufacture siRNAs.

What implications do these new findings have for evolutionary biology? First, the prevailing view that adaptations are mostly due to amino acid changes will need reevaluation. For example, Andolfatto (2005) estimated that as much as 83% of nucleotides in *Drosophila* UTRs have functional importance and 58% of the observed sequence divergence in UTRs was positively selected. The fact that a single sequence can give rise to both translated and untranslated functional products may necessitate a reinterpretation of tests based on the nonsynonymous rate (d_N) to synonymous rate (d_S) ratios. Although modifications of statistical tests for selection detection have been applied with success to noncoding sequences (Wong & Nielsen, 2004; Andolfatto, 2005), development of a new theoretical framework might be useful. Quantitative trait loci (QTL) mapping is another field likely to be affected by the new developments. There is accumulating evidence from high-resolution mapping that QTL may correspond to noncoding rather

than coding regions (Lyman *et al.*, 1999; Wang *et al.*, 1999; Maloof, 2006). For example, Oliver *et al.* (2005) resolved a body mass-affecting QTL in mice to a 660-kb region containing only two genes of known function, *Gpc3* and *Gpc4*, and two other putative genes of unknown function. However, they found no nonsynonymous polymorphism in any of those genes; instead, they identified 3' UTRs as strong candidates of functional importance.

Conclusions

In this brief review, I have presented evidence that although never translated, a large proportion of the transcriptome (at least in higher eukaryotes) constitutes a layer of functional regulation parallel to proteins, and as such it cannot be ignored by evolutionary biologists. Clearly, there is a long road ahead of genomics, with transcriptomics deservedly staying in its central part.

References

- Ambros, V. 2004. The functions of animal microRNAs. *Nature* **431**: 350–355.
- Andolfatto, P. 2005. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* **437**: 1149–1152.
- Bartel, D.P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**: 281–297.
- Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U., Meiri, E., Sharon, E., Spector, Y. & Bentwich, Z. 2005. Identification of hundreds of conserved and nonconserved human microRNAs. *Nat. Genet.* **37**: 766–770.
- Bertone, P., Stolc, V., Royce, T.E., Rozowsky, J.S., Urban, A.E., Zhu, X., Rinn, J.L., Tongprasit, W., Samanta, M., Weissman, S., Gerstein, M. & Snyder M. 2004. Global identification of human transcribed sequences with genome tiling arrays. *Science* **306**: 2242–2246.
- Bertrand, E. & Fournier, M.J. 2004. The snoRNPs and Related Machines: Ancient Devices That Mediate Maturation of rRNA

- and Other RNAs. In: *The Nucleolus* (M. O. J. Olson, ed.), pp. 225–261. Landes Bioscience Publishing, Georgetown, TX.
- Blencowe, B.J. 2002. Transcription: surprising role for an elusive small nuclear RNA. *Curr. Biol.* **12**: R147–R149.
- Britten, R.J. & Davidson, E.H. 1969. Gene regulation for higher cells: a theory. *Science* **165**: 349–357.
- Cavaille, J., Buiting, K., Kieffmann, M., Lalonde, M., Brannan, C.I., Horsthemke, B., Bachellerie, J.P., Brosius, J. & Huttenhofer, A. 2000. Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Natl. Acad. Sci. USA* **97**: 14311–14316.
- Cavaille, J., Seitz, H., Paulsen, M., Ferguson-Smith, A.C. & 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.
- Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J., Wheeler, R., Wong, B., Drenkow, J., Yamanaka, M., Patel, S., Brubaker, S., Tammana, H., Helt, G., Struhl, K. & Gingeras, T.R. 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.
- Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G., Sementchenko, V., Piccolboni, A., Bekiranov, S., Bailey, D.K., Ganesh, M., Ghosh, S., Bell, I., Gerhard, D.S. & Gingeras, T.R. 2005. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308**: 1149–1154.
- Claverie, J.M. 2001. Gene number. What if there are only 30,000 human genes? *Science* **291**: 1255–1257.
- Claverie, J.M. 2005. Fewer genes, more noncoding RNA. *Science* **309**: 1529–1530.
- Costa, F.F. 2005. Non-coding RNAs: new players in eukaryotic biology. *Gene* **357**: 83–94.
- Editorial 2000. The nature of the number. *Nat. Genet.* **25**: 127.
- FANTOM Consortium and RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group). 2005. The transcriptional landscape of the mammalian genome. *Science* **309**: 1559.
- Feder, M.E. & Walser, J.C. 2005. The biological limitations of transcriptomics in elucidating stress and stress responses. *J. Evol. Biol.* **18**: 901–910.
- Franke, A. & Baker, B.S. 1999. The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in *Drosophila*. *Mol. Cell* **4**: 117–122.
- Frith, M.C., Pheasant, M. & Mattick, J.S. 2005. The amazing complexity of the human transcriptome. *Eur. J. Hum. Genet.* **13**: 894–897.
- Gallagher, R.C., Pils, B., Albalwi, M. & Francke, U. 2002. Evidence for the role of PWCR1/HBII-85 C/D box small nucleolar RNAs in Prader-Willi syndrome. *Am. J. Hum. Genet.* **71**: 669–678.
- Gibson, G. 2002. Microarrays in ecology and evolution: a preview. *Mol. Ecol.* **11**: 17–24.
- Hall, I.M., Noma, K. & Grewal, S.I. 2003. RNA interference machinery regulates chromosome dynamics during mitosis and meiosis in fission yeast. *Proc. Natl. Acad. Sci. USA* **100**: 193–198.
- Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R. & Hannon, G.J. 2001. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**: 1146–1150.
- Hatada, I., Morita, S., Obata, Y., Sotomaru, Y., Shimoda, M. & Kono, T. 2001. Identification of a new imprinted gene, Rian, on mouse chromosome 12 by fluorescent differential display screening. *J. Biochem.* **130**: 187–190.
- Herr, A.J., Jensen, M.B., Dalmay, T. & Baulcombe, D.C. 2005. RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**: 118–120.
- Hodgkinson, C.A., Goldman, D., Jaeger, J., Persaud, S., Kane, J.M., Lipsky, R.H. & Malhotra, A.K. 2004. Disrupted in schizophrenia 1 (DISC1): association with schizophrenia, schizoaffective disorder, and bipolar disorder. *Am. J. Hum. Genet.* **75**: 862–872.
- Hollander, M.C., Alamo, I. & Fornace Jr, A.J., 1996. A novel DNA damage-inducible transcript, gadd7, inhibits cell growth, but lacks a protein product. *Nucleic Acids Res.* **24**: 1589–1593.
- Huang, Z.P., Zhou, H., He, H.L., Chen, C.L., Liang, D. & Qu, L.H. 2005. Genome-wide analyses of two families of snoRNA genes from *Drosophila melanogaster*, demonstrating the extensive utilization of introns for coding of snoRNAs. *RNA* **11**: 1303–1316.
- IHGSC. 2004. Finishing the euchromatic sequence of the human genome. *Nature* **431**: 931.
- Imamura, T., Yamamoto, S., Ohgane, J., Hattori, N., Tanaka, S. & Shiota, K. 2004. Non-coding RNA directed DNA demethylation of Sphk1 CpG island. *Biochem. Biophys. Res. Commun.* **322**: 593–600.
- Kampa, D., Cheng, J., Kapranov, P., Yamanaka, M., Brubaker, S., Cawley, S., Drenkow, J., Piccolboni, A., Bekiranov, S., Helt, G., Tammana, H. & Gingeras, T.R. 2004. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res.* **14**: 331–342.
- Kanno, T., Huettel, B., Mette, M.F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D.P., Matzke, M. & Matzke, A.J. 2005. Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat. Genet.* **37**: 761–765.
- Kapranov, P., Cawley, S.E., Drenkow, J., Bekiranov, S., Strausberg, R.L., Fodor, S.P. & Gingeras, T.R. 2002. Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296**: 916–919.
- Kapranov, P., Sementchenko, V.I. & Gingeras, T.R. 2003. Beyond expression profiling: next generation uses of high density oligonucleotide arrays. *Brief Funct. Genomic Proteomic.* **2**: 47–56.
- Kapranov, P., Drenkow, J., Cheng, J., Long, J., Helt, G., Dike, S. & 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.
- Kelley, R.L. & Kuroda, M.I. 2000. Noncoding RNA genes in dosage compensation and imprinting. *Cell* **103**: 9–12.
- Krek, A., Grun, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M. & Rajewsky, N. 2005. Combinatorial microRNA target predictions. *Nat. Genet.* **37**: 495–500.
- Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W. & Tuschl, T. 2002. Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* **12**: 735–739.
- Lanz, R.B., Razani, B., Goldberg, A.D. & O'Malley, B.W. 2002. Distinct RNA motifs are important for coactivation of steroid hormone receptors by steroid receptor RNA activator (SRA). *Proc. Natl. Acad. Sci. USA* **99**: 16081–16086.
- Lau, N.C., Lim, L.P., Weinstein, E.G. & Bartel, D.P. 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**: 858–862.

- Lee, J.T., Davidow, L.S. & Warshawsky, D. 1999. Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat. Genet.* **21**: 400–404.
- Lewis, B.P., Burge, C.B. & Bartel, D.P. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**: 15–20.
- Li, A.W. & Murphy, P.R. 2000. Expression of alternatively spliced FGF-2 antisense RNA transcripts in the central nervous system: regulation of FGF-2 mRNA translation. *Mol. Cell Endocrinol.* **170**: 233–242.
- Lipshitz, H.D., Peattie, D.A. & Hogness, D.S. 1987. Novel transcripts from the Ultrabithorax domain of the bithorax complex. *Genes Dev.* **1**: 307–322.
- Lyman, R.F., Lai, C., & Mackay, T.F. 1999. Linkage disequilibrium mapping of molecular polymorphisms at the scabrous locus associated with naturally occurring variation in bristle number in *Drosophila melanogaster*. *Genet. Res.* **74**: 303–311.
- Madamanchi, N.R., Hu, Z.Y., Li, F., Horaist, C., Moon, S.K., Patterson, C., Runge, M.S., Ruef, J., Fritz, P.H. & Aaron, J. 2002. A noncoding RNA regulates human protease-activated receptor-1 gene during embryogenesis. *Biochim. Biophys. Acta* **1576**: 237–245.
- Maloof, J.N. 2006. Quantitative genetics: small but not forgotten. *Heredity* **96**: 1–2.
- Martinho, R.G., Kunwar, P.S., Casanova, J. & Lehmann, R. 2004. A noncoding RNA is required for the repression of RNAPolIII-dependent transcription in primordial germ cells. *Curr Biol.* **14**: 159–165.
- Mattick, J.S. 2001. Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep.* **2**: 986–991.
- Mattick, J.S. 2003. Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays* **25**: 930–939.
- Mattick, J.S. 2004. RNA regulation: a new genetics? *Nat. Rev. Genet.* **5**: 316–323.
- Mattick, J.S. & Makunin, I.V. 2005. Small regulatory RNAs in mammals. *Hum. Mol. Genet.* **14** Spec No 1: R121–R132.
- Mockler, T.C. & Ecker, J.R. 2005. Applications of DNA tiling arrays for whole-genome analysis. *Genomics* **85**: 1–15.
- Moore, M.J. 2005. From birth to death: the complex lives of eukaryotic mRNAs. *Science* **309**: 1514–1518.
- Nakamura, A., Amikura, R., Mukai, M., Kobayashi, S. & Lasko, P.F. 1996. Requirement for a noncoding RNA in *Drosophila* polar granules for germ cell establishment. *Science* **274**: 2075–2079.
- Numata, K., Kanai, A., Saito, R., Kondo, S., Adachi, J., Wilming, L.G., Hume, D.A., Hayashizaki, Y. & Tomita, M.; RIKEN GER Group; GSL Members. 2003. Identification of putative non-coding RNAs among the RIKEN mouse full-length cDNA collection. *Genome. Res.* **13**: 1301–1306.
- Okazaki, Y., Furuno, M., Kasukawa, T. et al. 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**: 563–573.
- Oliver, F., Christians, J.K., Liu, X., Rhind, S., Verma, V., Davison, C., Brown, S.D., Denny, P. & Keightley, P.D. 2005. Regulatory variation at glypican-3 underlies a major growth QTL in mice. *PLoS Biol.* **3**: 135.
- Onodera, Y., Haag, J.R., Ream, T., Nunes, P.C., Pontes, O. & Pikaard, C.S. 2005. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**: 613–622.
- Pillai, R.S., Bhattacharyya, S.N., Artus, C.G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E. & Filipowicz, W. 2005. Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science* **309**: 1573–1576.
- Ranz, J.M. & Machado, C.A. 2006. Uncovering evolutionary patterns of gene expression using microarrays. *Trends. Ecol. Evol.* **21**: 29–37.
- RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group) and FANTOM Consortium, 2005. Antisense transcription in the mammalian transcriptome. *Science* **309**: 1564.
- Sanchez-Herrero, E. & Akam, M. 1989. Spatially ordered transcription of regulatory DNA in the bithorax complex of *Drosophila*. *Development* **107**: 321–329.
- Scott, R.E., Gao, S., Kim, C.K. & Cox, R. 2005. De-differentiation-derived mesenchymal stem cells demonstrate selective repression in H19 bioregulatory RNA gene expression. *Differentiation* **73**: 294–302.
- Shi, Y. & Berg, J.M. 1995. Specific DNA-RNA hybrid binding by zinc finger proteins. *Science* **268**: 282–284.
- Singh-Gasson, S., Green, R.D., Yue, Y., Nelson, C., Blattner, F., Sussman, M.R., Cerrina, F. 1999. Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nat. Biotechnol.* **17**: 974–978.
- Sleutels, F., Zwart, R. & Barlow, D.P., 2002. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* **415**: 810–813.
- Stearns, S.C. & Magwene, P. 2003. The naturalist in a world of genomics. *Am. Nat.* **161**: 171–180.
- Stolc, V., Gauhar, Z., Mason, C., Halasz, G., van Batenburg, M.F., Rifkin, S.A., Hua, S., Herreman, T., Tongprasit, W., Barbano, P.E., Bussemaker, H.J. & White, K.P. 2004. A gene expression map for the euchromatic genome of *Drosophila melanogaster*. *Science* **306**: 655–660.
- Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A. & Mello, C.C. 1999. The rde-1 gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* **99**: 123–132.
- Takeda, K., Ichijo, H., Fujii, M., Mochida, Y., Saitoh, M., Nishitoh, H., Sampath, T.K. & Miyazono, K. 1998. Identification of a novel bone morphogenetic protein-responsive gene that may function as a noncoding RNA. *J. Biol. Chem.* **273**: 17079–17085.
- Taverna, S.D., Coyne, R.S. & Allis, C.D. 2002. Methylation of histone h3 at lysine 9 targets programmed DNA elimination in tetrahymena. *Cell* **110**: 701–711.
- Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I. & Martienssen, R.A. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**: 1833–1837.
- Wang, R.L., Stec, A., Hey, J., Lukens, L. & Doebley J. 1999. The limits of selection during maize domestication. *Nature* **398**: 236–239.
- Wassenegger, M. 2000. RNA-directed DNA methylation. *Plant Mol. Biol.* **43**: 203–220.
- Willingham, A.T., Orth, A.P., Batalov, S., Peters, E.C., Wen, B.G., Aza-Blanc, P., Hogenesch, J.B. & Schultz, P.G. 2005. A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science* **309**: 1570–1573.
- Wong, W.S. & Nielsen, R. 2004. Detecting selection in noncoding regions of nucleotide sequences. *Genetics* **167**: 949–958.

- Wong, G.K., Passey, D. A. & Yu, J. 2001. Most of the human genome is transcribed. *Genome. Res.* **11**: 1975–1977.
- Xie, X., Lu, J., Kulbokas, E.J., Golub, T.R., Mootha, V., Lindblad-Toh, K., Lander, E.S. & Kellis, M. 2005. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* **434**: 338–345.
- Yamada, K., Lim, J., Dale, J.M., Chen, H., Shinn, P., Palm, C.J., Southwick, A.M., Wu, H.C., Kim, C., Nguyen, M., Pham, P., Cheuk, R., Karlin-Newmann, G., Liu, S.X., Lam, B., Sakano, H., Wu, T., Yu, G., Miranda, M., Quach, H.L., Tripp, M., Chang, C.H., Lee, J.M., Toriumi, M., Chan, M.M., Tang, C.C., Onodera, C.S., Deng, J.M., Akiyama, K., Ansari, Y., Arakawa, T., Banh, J., Banno, F., Bowser, L., Brooks, S., Carninci, P., Chao, Q., Choy, N., Enju, A., Goldsmith, A.D., Gurjal, M., Hansen, N.F., Hayashizaki, Y., Johnson-Hopson, C., Hsuan, V.W., Iida, K., Karnes, M., Khan, S., Koesema, E., Ishida, J., Jiang, P.X., Jones, T., Kawai, J., Kamiya, A., Meyers, C., Nakajima, M., Narusaka, M., Seki, M., Sakurai, T., Satou, M., Tamse, R., Vaysberg, M., Wallender, E.K., Wong, C., Yamamura, Y., Yuan, S., Shinozaki, K., Davis, R.W., Theologis, A. & Ecker, J.R. 2003. Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science* **302**: 842–846.
- Zamore, P.D. & Haley, B. 2005. Ribo-gnome: the big world of small RNAs. *Science* **309**: 1519–1524.
- Zieve, G. & Penman, S. 1976. Small RNA species of the HeLa cell: metabolism and subcellular localization. *Cell* **8**: 19–31.

Received 24 January 2006; revised 20 March 2006; accepted 23 March 2006