Published in final edited form as:

Mamm Genome. 2008 August ; 19(7-8): 449–453. doi:10.1007/s00335-008-9151-8.

The Noncoding RNAs – A Genomic Symphony of Transcripts

Linda D. Siracusa and Arthur M. Buchberg

Department of Microbiology and Immunology, Thomas Jefferson University, Kimmel Cancer Center, 233 South 10th, Street Philadelphia, PA 191067-5541

Linda D. Siracusa: siracusa@mail.jci.tju.edu; Arthur M. Buchberg: buchberg@mail.jci.tju.edu

Introduction

Imagine you are playing a piano. Your fingers sequentially move from one end of the keyboard to the other. Now imagine there is another person, playing the same keys in the opposite direction. Add yet another set of hands, which start and stop playing different keys within the center of the keyboard. Now imagine the room has many pianos, each with thousands of keys, and imagine each piano with thousands of players. The sound delicately balanced in sequence, with some melodies played very loud and others played very softly. So too, are the many transcripts of the genome, especially those of the class of noncoding RNAs (ncRNAs).

The regulation of protein production has many more checks and balances than previously imagined. At the heart of this control lies the ncRNAs, which include the familiar tRNAs and rRNAs as well as the more recently discovered microRNAs (miRNAs) and piRNAs (Table 1). The transcriptome has proven to be much more complex than the initial sequencing estimates of approximately 20,000 protein-coding genes in the human genome. Studies have estimated that almost all of the transcriptional output of the human genome is RNAs that do not encode proteins (Mattick 2005). The articles in this issue focus on many aspects of the regulation and functions of miRNAs and other ncRNAs in the mammalian genome.

The miRNA genes

Since the first identification of miRNAs by Victor Ambros and colleagues (Lee et al. 1993), a novel mechanism for controlling translation was discovered. MiRNAs are a group of small ncRNA molecules that are distinct from, but related to, small interfering RNAs (siRNAs) and that have been identified in a variety of organisms (Ambros 2003; Bartel 2004; He and Hannon 2004). These small 19–24 nucleotides (nt) RNAs are transcribed as parts of longer molecules, several kilobases (kb) in length, which are then capped and polyadenylated. These pri-miRNA transcripts are processed in the nucleus into hairpin RNAs of 70–100 nt by the ribonuclease Drosha and the RNA binding protein Pasha (Cullen 2004; Gregory et al. 2004). The hairpin RNAs are transported to the cytoplasm via an exportin-5 dependent mechanism; there they are digested by the double-strand-specific ribonuclease, Dicer.

In animals, single-stranded miRNA binds specific messenger RNA (mRNA) through sequences that are complementary to the target mRNA, mainly to the 3' untranslated region (3' UTR). By a mechanism that is not fully characterized, the bound mRNA remains untranslated, resulting in reduced levels of the corresponding protein; alternatively, if the sequence match between the microRNA and its target is exact, the bound mRNA can be degraded, resulting in reduced levels of both the corresponding transcript and its encoded protein.

Mechanism of miRNA interaction with protein-coding transcripts

It is now estimated that 1–3% of the human genome encodes miRNAs (<http://microrna.sanger.ac.uk/>) and that miRNAs regulate 20–30% of the human genome (Lewis et al. 2005), although others have predicted that miRNAs regulate 70–90% of all gene transcripts in an organism (Miranda et al. 2006). The major mechanism by which miRNAs control the translation of protein-coding genes and impact mRNA stability is at the post-transcriptional level, although exceptions to this rule exist (Carthew 2006; Borel and Antonarakis 2008 – this issue).

This exquisite regulation of gene expression involves multiple levels of complexity. A single mature miRNA can have one or many mRNA targets (Lim et al 2005). Thus, a single miRNA may simultaneously coordinate the regulation of a network of genes or may selectively inhibit genes in specific pathways. Conversely, a target mRNA may have several miRNAs that can bind to it. Thus, the ultimate translational output of a single mRNA may be determined by the balance between the expression of its miRNA binding partners. This delicate regulation of translation by miRNAs takes on yet another level of complexity, since regulatory feedback loops between miRNAs and their target genes are beginning to be discovered (Carthew 2006). The next level of complexity will involve intersections between pathways, where translation can be turned on and off in rapid fashion. The biggest challenge, however, will be to evaluate the phenotypic effects of subtle variations in miRNA expression in different cell types over time.

More ncRNAs

MiRNA genes are just a subset within the class of non-coding genes. The classical ncRNAs, such as tRNA and rRNA, fall into this category as well. However, a number of genes that do not encode proteins, starting with H19 (Brannan et al. 1990), have now and are continuing to be identified (Prasanth and Spector 2008; Yasada and Hayashaziki 2008). Some of these non-coding transcripts are very long, such as the antisense Igfar (*Air*) transcript which is 108 kb in length (Sleutels et al. 2002). The ncRNAs can be transcribed from within protein-coding genes (primarily from within their introns), from the opposite DNA strand as protein-coding genes (anti-sense transcripts), or independently from regions that were formerly thought to be “spacer” or “junk” DNA. Indeed, some of these transcripts overlap or are intertwined with each other. However, the mechanisms that regulate the expression of these ncRNAs remain a mystery.

Much analysis initially has relied on bioinformatics or computational methods to analyze existing sequences, mostly comparing expressed sequence tags (ESTs) or computational predictions to the sequence of the genome (genome.ucsc.edu; www.ensembl.org; www.ncbi.nlm.nih.gov). Transcripts falling into the class of long, large or macro ncRNAs were identified by *in silico* analyses, and verified by Northern blot and RT-PCR assays (Furuno et al. 2006) or microarray and stem-loop RT-PCR (He et al. 2008). The application of genome-wide tiling arrays coupled with mapping of cDNA sequences to assess transcript production and deep sequencing of transcriptomes are revealing that most, if not all, of the human genome is transcribed (Amaral and Mattick 2008 – this issue). Add to the existing collection of transcripts several new experimental large-scale approaches that attempt to capture all RNAs for cloning and sequencing (Yasuda and Hayashizaki 2008), and the total number of transcripts (not just within a cell or an organ but within the organisms as a whole) is staggering. It is clear from these studies that what was once thought to be long stretches of “silent DNA” or “gene deserts” are areas that are alive and brimming with encoded transcripts. Distinguishing the origins of each ncRNA transcript, their start and stop points along the length of each chromosome, and dissecting their regulation, timing, and sites of expression are only the beginning steps necessary for discovering their biological functions.

Evolution and microRNAs

Over the past decade, the number of manuscripts describing miRNAs and other ncRNAs has gone from a handful to several thousand in 2008 to date (www.pubmed.com). These RNA transcripts exist in animals, plants and viruses. Some of these transcripts are highly conserved, being derived from ultraconserved regions of the genome (Rossi et al. 2008 – this issue), whereas others appear to have undergone recent evolution and are species-specific (Berezikov et al. 2006; Lee et al. 2007). Indeed, there is evidence that some miRNAs have undergone rapid evolution, suggesting that their emergence is an ongoing process. Differences in miRNA copy number also impact the diversity between species. However, it is the ability of these miRNAs to coordinately regulate sets of genes or networks that makes them likely to not just be passive players in evolution, but to actually drive the evolutionary process.

Roles for ncRNAs in development, imprinting, and organogenesis

NcRNAs have essential roles in developmental programs in animals (Amaral and Mattick 2008– this issue). Indeed, the study of development in *Caenorhabditis elegans* was what led to the finding of the *lin-4* gene, which inhibits the function of the *lin-14* gene, a gene normally involved in a switch governing the control of larval development (Lee et al. 1993). MiRNAs have since been shown to be key players in cellular processes including, but not limited to, proliferation, growth, morphogenesis, differentiation, and apoptosis. NcRNAs have been shown to be involved in organ development as well, and this issue highlights reviews on miRNAs found in the eye (Huang et al. 2008 – this issue) as well as in the brain (Barbato et al. 2008 – this issue).

A subset of ncRNAs have roles in the modulation of chromatin architecture and epigenetic processes (Amaral and Mattick 2008– this issue; Rouhi et al. 2008 this issue). For example, the inactivation of the X-chromosome is controlled in part by the non-coding genes *Xist* and *Tsix* (Avner and Heard 2001). Furthermore, both long and short ncRNAs have essential roles in the imprinting process, as described by Peters and Robson (2008 this issue). Control of the epigenome during development and differentiation is thus another function of ncRNAs. The challenge is to elucidate the precise mechanisms governing changes in the epigenome and to understand how these coordinated regulatory mechanisms unfold during developmental programs and disease processes.

Memory and learning

Another exciting area of investigation is the development of memory and learning. The impact of ncRNAs, especially miRNAs, extends to higher-order executive functions of the human brain. For example, Barbato and colleagues (2008– this issue) describe miRNAs regulating gene expression within the nervous system. Studies of miRNA levels have demonstrated both temporal and spatial restricted expression within the nervous system of several organisms. The diverse functions of miRNAs extend to the process of neurogenesis, the control of neuronal cell fate specification and differentiation, as well as synaptic development and plasticity. However, the key mRNA targets of the differentially expressed miRNAs have yet to be identified. Nevertheless, these intriguing findings have implications for the study and treatment of neurologic and psychiatric disorders.

MiRNAs have been implicated in Tourette's and Fragile X Syndromes, as well as in neurodegenerative diseases such as Alzheimer's and Parkinson's disease. Aside from measuring expression levels, results of genetic studies have pointed to the association of sequence variants in miRNA genes with diseases such as schizophrenia. Many challenges remain for the task of fully deciphering the complexity of the human brain and understanding the role of ncRNAs, including the ability to dissect specific sets of neurons and to capture as

images what might be instantaneous changes *in situ*. One approach that has been described is to use a model organism, such as the mouse, to characterize miRNA function in the brain (Parsons et al. 2008 – this issue). No doubt, the strategy of conditional gene targeting in the brain is crucial to this undertaking as is the ability to precisely quantitate phenotypic traits.

Based on the ability of miRNAs to regulate synaptic activity and to establish or maintain neuronal diversity, it has been proposed that miRNAs are key players in learning and memory formation (Mattick 2005; Ashraf and Kunes 2006; Presutti et al. 2006). What better way to learn than to instantaneously “turn on” tiny genes that control a network of transcripts waiting for directions to make memories? The mystery of how we learn and what triggers recall may thus be intricately tied to the function of ncRNAs.

The impact of ncRNAs in cancer

Perhaps the area where ncRNAs, predominantly the miRNAs, have been most studied is cancer. The first report of a role for miRNAs in cancer was about chronic lymphocytic leukemia (Calin et al. 2002). Further studies showed that some miRNAs can act as oncogenes (oncomirs) in cancer by promoting the transformation of cells, whereas others act as tumor suppressor genes, and some may even act as both (Fabbri et al. 2008).

Many studies have since been published describing the results of gene expression profiling of miRNAs. They revealed that miRNA expression signatures are more accurate in distinguishing among cancer types than expression signatures of protein-coding genes (Lu et al. 2005; Volinia et al. 2006). Such signatures have been associated with the diagnosis, prognosis, and/or response to treatment of several human cancers. In addition, miRNA genes should be considered candidates for tumor susceptibility loci and may thus represent a new family of genes involved in cancer penetrance (Sevignani et al. 2007). Although the regulation of transcription of most miRNA genes remains unknown, epigenetic changes can influence miRNA expression; conversely, expression of miRNAs can influence epigenetic changes within a cell leading to cancer (Rouhi et al. 2008–this issue). The articles in this issue highlight the application of bioinformatic tools to study the characteristics and functions of miRNAs and ultraconserved regions (UCRs) in the process of tumorigenesis and describe the use of miRNA and UCR transcripts as tumor biomarkers and therapeutic targets (Rossi et al. 2008– this issue).

The future of ncRNAs

The term “RNA continent” was invented to describe the complex nature of the entire mouse transcriptome (Yasuda and Hayashizaki 2008), although it is clear that this concept applies to all genomes. The number of ncRNAs keeps growing and comprehensive databases are being developed to compile and annotate the findings (Table 2). Bioinformatic and experimental approaches are revealing many new long ncRNAs, implying that we have discovered only the tip of the iceberg. Computational approaches have suggested that the number of miRNA transcripts alone is in the tens of thousands for mammalian genomes (Miranda et al. 2006). Furthermore, Rigoutsos and colleagues (2006) used an unsupervised pattern-discovery method to process sequences in the human genome and found “unique functional connections between the coding and non-coding parts of the human genome.” Thus, understanding the full complement of fundamental cellular processes regulated by ncRNAs remains a considerable challenge.

Indeed, the complexity within the genome is telling us that there are many ways to “play” a strand of DNA and even more ways to compose a symphony. Imagine you are back in the room of pianos. In addition to the grade of softness or loudness, and although each song is played with a different tempo, the harmony is evident. You walk down a hallway with multiple doors, each room filled with pianos. The music coming from one room influences the songs played

in the next. The building you are in is a multicellular organism, with each floor a different organ and each room a different cell. The symphonies change as you walk down the hallway, mimicking the passage of time. This issue of *Mammalian Genome* is just the beginning of a musical notebook, with many melodies waiting to be discovered.

References

- Amaral PP, Mattick JS. Noncoding RNA in development. *Mamm Genome* 2008;19:454–492. [PubMed: 18839252]
- Ambros V. MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. *Cell* 2003;113:673–676. [PubMed: 12809598]
- Ashraf SI, Kunes S. A trace of silence: memory and microRNA at the synapse. *Curr Opin Neurobiology* 2006;16:535–539.
- Avner P, Heard E. X-chromosome inactivation, counting, choice and initiation. *Nature Rev Genet* 2001;2:59–67. [PubMed: 11253071]
- Barbato C, Giorgi C, Catalanotto C, Cogoni C. Thinking about RNA? MicroRNAs in the brain. *Mamm Genome* 2008;19:541–551. [PubMed: 18670804]
- Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004;116:281–297. [PubMed: 14744438]
- Berezikov E, Thummmler F, van Laake LW, Kondova I, Bontrop R, et al. Diversity of microRNAs in human and chimpanzee brain. *Nat Gen* 2006;38:1375–1377.
- Borel C, Antonarakis SE. Functional genetic variation of human miRNAs and phenotypic consequences. *Mamm Genome* 2008;19:503–509. [PubMed: 18787897]
- Brannan CI, Dees EC, Ingram RS, Tilghman SM. The product of the H19 gene may function as an RNA. *Mol Cell Biol* 1990;10:28–36. [PubMed: 1688465]
- Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, et al. Frequent deletions and down-regulation of microRNA genes *miR15* and *miR16* at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci USA* 2002;99:15524–15529. [PubMed: 12434020]
- Carthew RW. Gene regulation by microRNAs. *Curr Opin Genet Dev* 2006;16:203–208. [PubMed: 16503132]
- Cullen BR. Transcription and processing of human microRNA precursors. *Mol Cell* 2004;16:861–865. [PubMed: 15610730]
- Fabbri M, Croce CM, Calin GA. MicroRNAs. *Cancer Journal* 2008;14:1–6.
- Furuno M, Pang KC, Ninomiya N, Fukuda S, Frith MC, et al. Clusters of internally primed transcripts reveal novel long noncoding RNAs. *PLoS Genet* 2006;2:e37. [PubMed: 16683026]
- Gregory RI, Yan KP, Amuthan G, Chendrimada TP, Doratotaj B, et al. The microprocessor complex mediates the genesis of microRNAs. *Nature* 2004;432:235–240. [PubMed: 15531877]
- He L, Hannon GJ. MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* 2004;5:522–531. [PubMed: 15211354]
- He S, Su H, Liu C, Skogerbo G, He H, et al. MicroRNA-encoding long non-coding RNAs. *BMC Genomics* 2008;9:236. [PubMed: 18492288]
- Huang KM, Dentchev T, Stambolian D. MiRNA expression in the eye. *Mamm Genome* 2008;19:510–516. [PubMed: 18648874]
- Lee CT, Risom T, Strauss WM. Evolutionary conservation of microRNA regulatory circuits: an examination of microRNA gene complexity and conserved microRNA-target interactions through metazoan phylogeny. *DNA Cell Biology* 2007;26:209–218.
- Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 1993;75:843–854. [PubMed: 8252621]
- Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates thousands of human genes are microRNA targets. *Cell* 2005;120:15–20. [PubMed: 15652477]
- Lim LP, Lau NC, Garrett-Engle P, Grimson A, Schelter JM, et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 2005;433:769–773. [PubMed: 15685193]

- Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, et al. MicroRNA expression profiles classify human cancers. *Nature* 2005;435:834–838. [PubMed: 15944708]
- Mattick JS. The functional genomics of noncoding RNA. *Science* 2005;309:1527–1528. [PubMed: 16141063]
- Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, et al. A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. *Cell* 2006;126:1203–1217. [PubMed: 16990141]
- Parsons MJ, Grimm CH, Paya-Cano JL, Sugden K, Nietfeld W, et al. Using hippocampal microRNA expression differences between mouse inbred strains to characterise miRNA function. *Mamm Genome* 2008;19:552–560. [PubMed: 18594913]
- Peters J, Robson JE. Imprinted noncoding RNAs. *Mamm Genome* 2008;19:493–502. [PubMed: 18815833]
- Prasanth KV, Spector DL. Eukaryotic regulatory RNAs: an answer to the ‘genome complexity’ conundrum. *Genes Dev* 2008;21:11–42. [PubMed: 17210785]
- Presutti C, Rosati J, Vincenti S, Nasi S. Non coding RNA and brain. *BMC Neuroscience* 2006;7(Suppl 1):S5. [PubMed: 17118159]
- Rigoutsos I, Huynh T, Miranda K, Tsirigos A, McHardy A, et al. Short blocks from the noncoding parts of the human genome have instances within nearly all known genes and relate to biological processes. *Proc Natl Acad USA* 2006;103:6605–6610.
- Rossi S, Seignani C, Siracusa LD, Calin G. Cancer-associated genomic regions (CAGRs) and noncoding RNAs: bioinformatics and therapeutic implications. *Mamm Genome* 2008;19:526–540. [PubMed: 18636290]
- Rouhi A, Mager DL, Humphries RK, Kuchenbauer F. MiRNAs, epigenetics, and cancer. *Mamm Genome* 2008;19:517–525. [PubMed: 18688563]
- Seignani C, Calin GA, Nnadi SC, Shimizu M, Davuluri RV, et al. MicroRNA genes are frequently located near mouse cancer susceptibility loci. *Proc Natl Acad USA* 2007;104:8017–8022.
- Sleutels F, Zwart R, Barlow DP. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* 2002;415:810–813. [PubMed: 11845212]
- Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, et al. microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci USA* 2006;103:2257–2261. [PubMed: 16461460]
- Yasuda J, Hayashizaki Y. The RNA Continent. *Adv Cancer Res* 2008;99:77–112. [PubMed: 18037407]

Table 1

General classes of short noncoding RNAs in mammals.

Gene symbol	Gene Name	Function	Average length of functional segment ^a
miRNA	microRNA	Translational control of mRNAs	19–24 nt
piRNA	Piwi-interacting RNA	Associate with Piwi proteins of the Argonaute family	26–31 nt
rRNA	Ribosomal RNA	Translation of mRNA into protein	121 nt, 2 kb, and 4.5 kb
scaRNA	Small Cajal body-specific RNAs	Chemical modifications of snRNAs	120–500 nt
snRNA	Small nuclear RNA	RNA splicing	100–300 nt
snoRNA	Small nucleolar RNA	Chemical modifications of RNAs	50 nt
tRNA	Transfer RNA	Deliver amino acids to the site of translation	75–95 nt
TERC	Telomerase RNA component	Regulation of telomere length	450–550 nt

^a kb = kilobases, nt = nucleotides.

Table 2

Publically-available noncoding RNA databases.

Database name and url	Database Description
NONCODE http://www.noncode.org/	Comprehensive database of all noncoding RNAs, except rRNA and tRNA; manually annotated
frnAdb http://www.ncrna.org/frnadb/	Comprehensive database of ncRNAs
Noncoding RNA database http://biobases.ibch.poznan.pl/ncRNA/	Sequence and functional information of regulatory noncoding RNA; does not have miRNA and snoRNA
RNAdb http://jsm-research.imb.uq.edu.au/rnadb/	Comprehensive mammalian noncoding RNA database
ncRNA expression database http://jsm-research.imb.uq.edu.au/abancrna	Brain-specific expression database of ncRNAs in mice and humans
Rfam http://rfam.sanger.ac.uk/	Database of RNA families
Software ncRNA http://software.ncrna.org/	Portal site for software tools and servers for sequence and structural analyses of ncRNAs