REVIEW

Institute of Biochemistry, Department of Chemistry and Biology, Justus Liebig University of Giessen, Germany

# Eukaryotic non-coding RNAs: new targets for diagnostics and therapeutics?

O. Rossbach, A. Bindereif

*Received June 30, 2015, accepted August 19, 2015*

*Prof. Dr. Albrecht Bindereif, Institute of Biochemistry, Department of Chemistry and Biology, Justus Liebig University of Giessen, Heinrich-Buff-Ring 58, 35392 Giessen, Germany*
*albrecht.bindereif@chemie.bio.uni-giessen.de*

*Pharmazie 71: 3–7 (2016)* *doi: 10.1691/ph.2016.5736*

Non-coding RNAs (ncRNAs) in eukaryotes have recently developed to a very active research area in RNA biology, opening up new strategies for diagnosis and therapies of human disease. Here we introduce and describe the most important classes of eukaryotic ncRNAs: microRNAs (miRNAs), long non-coding RNAs (lncRNAs), and circular RNAs (circRNAs). We further discuss new RNA-based diagnostic and therapeutic concepts.

## 1. Introduction: Coding and non-coding RNAs in eukaryotes

The deciphering of the human genome sequence in 2000 marked a unique breakthrough in modern molecular biology. An important insight was the unexpectedly low number of protein-coding genes (about 20,000), which lagged far behind previous estimates. This number revealed that only approximately two percent of the human genome is transcribed into messenger RNA (mRNA).

However, the human genome sequence itself did not provide an explanation for the enormous complexity of the human organism compared with, for example, the more primitive worm *Caenorhabditis elegans* (*C. elegans)* that carries a similar number of protein-coding genes. Only in the following years, we gradually learned more about the human transcriptome, which is defined as the total ensemble of RNA transcripts. This resulted in the surprising finding that much more of the human genome is actively transcribed than initially assumed, a phenomenon called "pervasive transcription" (Jensen et al. 2013).

The rapid development of high-throughput sequencing ("deep sequencing") techniques was essential for this progress, in particular the so-called RNA-Seq method, by which the total RNA of a cell is transcribed into complementary DNA (cDNA), followed by DNA amplification *via* the polymerase chain reaction (PCR) and massive parallel sequencing. Thereby, a quantitative overview on all transcribed sequences at single-nucleotide resolution can be obtained within a single experiment, thus creating a snapshot of all RNAs in a certain cell, tissue or entire organism. As a consequence, transcriptomes of cells in different developmental stages, or from pathological *versus* healthy control samples, can be directly compared with each other in a quantitative and qualitative manner.

This opened up a new chapter in RNA biology. In addition to the classical RNA species known for decades, which are mRNAs, ribosomal RNAs (rRNAs), and transfer RNAs (tRNAs), multiple groups of RNAs with new functions were discovered, most prominently microRNAs (miRNAs), long non-coding RNAs (lncRNAs), and circular RNAs (circRNAs), which will be introduced in the following. Single representatives had been known already before 2000, but only after the introduction of high-throughput sequencing these few examples were recognized as members of entire classes of ncRNAs, each consisting of hundreds to thousands of different RNA species (see Fig. 1).

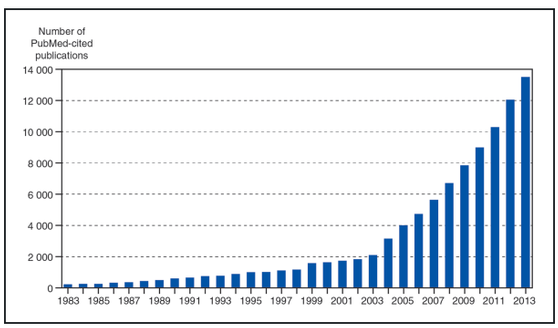


Fig. 1: Non-coding RNAs as an emerging field. The rapid development of the non-coding RNA field can be inferred from the number of PubMed-cited publications (http://www.ncbi.nlm.nih.gov/pubmed), using *non-coding RNA* as a keyword in searches for publications between 1983 and 2013.

The rapid development of the last few years, also called "Non-coding RNA revolution" will likely continue or even intensify (Cech and Steitz 2014), concomitantly with a shift in human genome research activities, that is, from the focus on mRNAs and their protein products to functional studies on the cellular networks of all RNAs including ncRNAs. Despite all enthusiasm one should keep in mind that our knowledge on ncRNAs currently increases mainly on a descriptive level, owing to the massive data output of deep sequencing studies. More and more new classes of ncRNAs are currently discovered that await their functional investigation. This includes the very basic question of how many of the thousands of new ncRNAs are biologically relevant at all. On the other hand, even without knowing or understanding their functions, new RNA species are interesting and relevant as potential biomarkers.

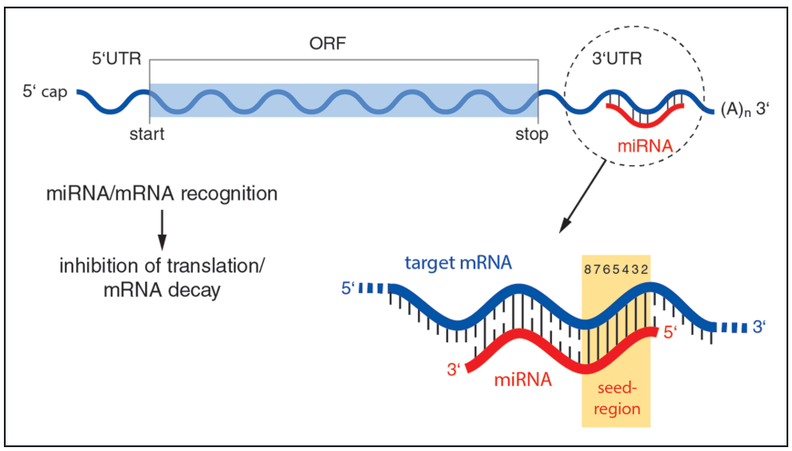


Fig. 2: Functional mechanism of microRNAs (miRNAs). The figure illustrates the characteristic binding of a miRNA within the 3'-untranslated region (3'-UTR) of a typical eukaryotic mRNA, consisting of a 5'-cap structure (5'-cap), 5' and 3'-UTRs, and the protein-coding open reading frame (ORF). MiRNA:mRNA recognition minimally proceeds through base-pairing of the seed region (nucleotides 2-8) of the miRNA, resulting in inhibition of translation or mRNA degradation.

This review will first introduce miRNAs and lncRNAs, followed by circRNAs as an example of a very recently discovered class of ncRNAs. Finally, we will discuss how these new RNA classes may be utilized for novel diagnostic and therapeutic concepts.

## 2. MiRNAs

Mature miRNAs are short single-stranded ncRNAs of 19-24 nucleotides in length that are incorporated into larger protein complexes in the cytoplasm of eukaryotic cells (Zamore and Haley 2005). The first examples of miRNAs, lin-4 and let-7, have been known as ncRNAs since the early 1990s. They were discovered in *C. elegans* through their essential role in the larval development and guide the miRNA-associated protein complexes to their specific mRNA target sites *via* Watson-Crick base pairing with the mRNA (Reinhart et al. 2000). These initially identified miRNA species already exemplified the most important characteristics of miRNAs: (1) their target sequences normally resides in the 3'-untranslated region (3'-UTR) of mRNAs; (2) miRNA-mRNA pairing does not cover the entire length of the miRNA strand; (3) miRNAs are evolutionarily conserved; (4) multiple targets are recognized, i.e. a single miRNA can specifically recognize several target sites in different mRNAs or even within the same mRNA; (5) miRNA binding to mRNAs negatively regulates protein biosynthesis (translation); and (6) mature miRNAs are processed from longer precursor RNAs (Fig. 2).

Genome-wide systematic searches since 2000 by use of miRNA-tailored sequencing approaches, as well as functional studies, quickly extended these few miRNA representatives to large classes of evolutionarily conserved miRNAs in all investigated eukaryotic systems. It became clear that miRNA-mRNA interactions in animal systems essentially involve a short, perfect base pairing between the mRNA and the so-called seed sequence at the 5' end of the miRNA (nucleotides 2-8). Additional base pairing between the 3'-proximal region of the miRNA and the mRNA may occur, but is not essential (Filipowicz et al. 2008). By today, more than 28,000 miRNAs have been identified in various eukaryotic organisms, ranging from *C. elegans*, *Drosophila*, the human system to plants, such as *Arabidopsis*. This information is documented in databases that are constantly updated, a prominent example being the miRBase (www.mirbase.org/). The relatively well-characterized human system contains around 2,500 individual miRNAs, many of which are embedded in regulatory networks. It is estimated that one third of all human protein-coding genes are regulated by miRNAs, which explains why miRNAs are sometimes called *micromanagers* of gene expression.

The biogenesis pathway leading to mature miRNAs requires several ordered processing stages, which take place in the eukaryotic nucleus and subsequently in the cytoplasm. First, starting from long, partially double-stranded precursor RNAs (pri-miRNAs), the RNase DROSHA generates intermediates of 60-80 nucleotides in length (pre-miRNAs) in the nuclear compartment. Second, after export of these pre-miRNAs into the cytoplasm, the DICER complex produces short, double-stranded RNAs (19-24 nucleotides per strand). One of the strands, the so-called guide strand, is loaded onto an effector complex, called RISC (*RNA-induced silencing complex*). This enables RISC complexes to find their specific target sites in cytoplasmic mRNAs. Numerous detailed studies have revealed that miRNAs can either mediate inhibition of translation (*translational silencing*) or induce degradation of the specific mRNA (*mRNA decay*; Huntzinger and Izaurralde 2011). During target recognition, RISC-bound miRNAs are associated with so-called Ago proteins, particularly Ago2 (also called *slicer*). If miRNA:mRNA base pairing is perfect over the entire miRNA length, Ago2 can endonucleolytically cleave the mRNA target sequence, thus mimicking the siRNA-mediated cleavage mechanism (see article by Grünweller and Hartmann in this issue).

Since miRNAs show cell type- and tissue-specific expression patterns and perform multiple regulatory roles in differentiation and development, it is not surprising that more and more examples are discovered where aberrant expression of miRNAs turns out to be a major cause of disease states (Garzon et al. 2009), including diabetes, neuro- and immunodegenerative diseases, immune defects and cancer. In the case of tumor tissues, pathologically overexpressed miRNAs can promote tumor progression and thus function as oncogenes ("oncomiRs"); on the other hand, miRNAs with tumor-suppressive activity may be downregulated in tumor tissues. This explains why miRNAs and their expression profiles are intensely studied and applied as tumor-specific biomarkers.

## 3. Long non-coding RNAs (lncRNAs)

Long non-coding RNAs (lncRNAs) constitute another large class of ncRNAs which are functionally and structurally more diverse than miRNAs. LncRNAs are defined by their size exceeding 200 nucleotides (often several kb and up to 100 kb

in extreme cases) and their lack of protein-coding potential (Nagano and Fraser 2011; Rinn and Chang 2012). The latter criterion, however, has recently been questioned as short open reading frames within lncRNAs may be utilized for the expression of peptides (Andrews et al. 2014; Anderson et al. 2015). As described for miRNAs (see above), a few representatives of the lncRNA family had been known already since 1991, such as lncRNAs Xist, H19 and AIR.

Some lncRNAs have been implicated in chromatin association and modification, such as Xist, which is responsible for transcriptional silencing of one X-chromosome in female mammals by associating over the entire length of the X-chromosome. Since several lncRNAs are chromatin-associated, the model of a global RNA-chromatin network has been put forward in which lncRNAs play a central regulatory role.

Based on the multiple functions of various lncRNAs in chromatin structure and activation it is not too surprising that lncRNA expression is deregulated in many tumor types (Cheetham et al. 2011). A relatively well characterized lncRNA, *MALAT1*, is involved in alternative splicing through regulating the phosphorylation of certain alternative splicing factors to control their localization and activity (Tripathi et al. 2010). The *MALAT1* lncRNA provides a nice example how various levels of RNA-mediated regulation are linked with each other.

The functions of lncRNAs may be summarized and grouped as follows: (1) a so-called *decoy function*, whereby the lncRNA serves as a bait to prevent specific proteins from DNA binding, thus blocking their normal activity; (2) a *scaffold function* for the assembly of multiple protein components on the lncRNA, including proteins with or without RNA-binding activity; (3) a *guide function*: many lncRNAs with bound protein factors (for example, with chromatin organization and modification roles) are guided *via* certain proteins of the lncRNP complex to specific genome regions (Fig. 3).

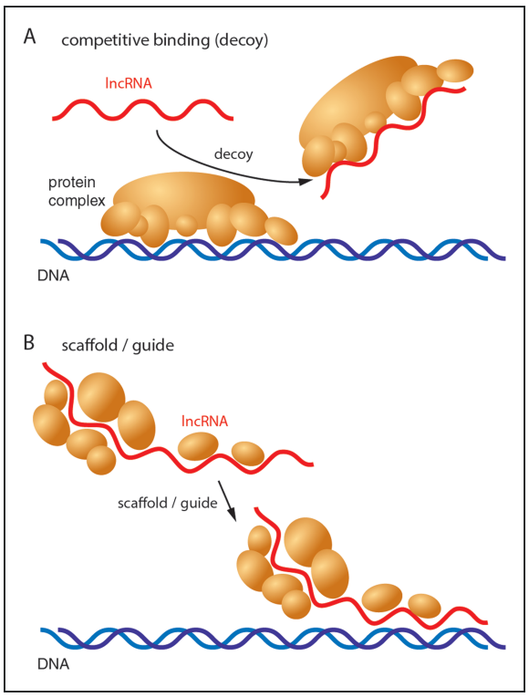


Fig. 3: Functions of long non-coding RNAs (lncRNAs). LncRNAs can compete with genomic DNA regions for binding to protein complexes. Some lncRNAs sequester proteins or protein complexes, thereby impeding their normal function (**A**; "*decoy*"). In addition, there are examples for lncRNAs acting as a scaffold for proteins and/or for guiding bound proteins to specific cell areas (**B**; "*scaffold*"/"*guide*"). As shown here, both functions may also be combined within the same lncRNA molecule, whereby lncRNA-associated proteins are delivered to certain DNA or chromatin regions.

## 4. Circular RNAs (circ-RNAs)

Circular RNAs (circ-RNAs) represent the most recently discovered ncRNA class, and they perfectly illustrate how a few single representatives, some already known for decades, suddenly (in 2013) opened up a large new ncRNA class. Already in 1976, the plant-pathogenic so-called viroid RNAs were discovered, representing the first examples of naturally occurring circular RNAs (Sänger et al. 1976). In the following decades a number of circ-RNAs were found in diverse biological systems, for example the *SRY* RNA in mouse, which is important for testis development (Capel et al. 1993), various circRNAs in *Archaea* (Kjems and Garrett 1988; Danan et al. 2012), and the human *cANRIL* circRNA (Burd et al. 2010).

With the advent of tailored RNA-Seq technologies applied to genome-wide investigations, circRNAs emerged as a novel and wide-spread ncRNA class (Hansen et al. 2013; Memczak et al. 2013; Jeck et al. 2013; Glazar et al. 2014; summarized by Hentze and Preiss 2013; Lasda and Parker 2014). CircRNAs are released from primary transcripts of protein-coding mRNAs, when a single exon or two adjacent exons circularize in an alternative splicing process (Starke et al. 2015; see Fig. 4). As numerous pre-mRNAs give rise to circRNAs, hundreds to thousands of such different circRNAs coexist in cells and accumulate primarily in the cytoplasm. Several studies have shown that circRNAs are expressed in a cell type- and tissue-specific manner, immediately suggesting their potential use as novel RNA-based biomarkers in human diseases. Another feature beneficial to their use as biomarkers is the high metabolic stability of circRNAs, because they are protected from exoribonucleolytic degradation due to the absence of 5' and 3' ends.

At present, little is known about circRNA functions. At least for one human circRNA, ciRS-7 (= *circular RNA sponge for miR-7*; Hansen et al. 2011), a role as miRNA sponge was convincingly demonstrated: ciRS-7 contains a cluster of around 70 miR-7 binding sites, which can efficiently sequester and inactivate this miRNA (Hanssen et al. 2013). A similar miRNA sponge function is suggested for circRNA SRY and miR-138 (Hansen et al. 2013; Memczak et al. 2013). A recent study revealed that circRNAs may also be integrated in autoregulatory loops, as demonstrated for the *muscleblind* circRNA (circMbl) from *Drosophila*: the muscleblind protein (Mbl) regulates the circular processing pathway of its own *muscleblind* pre-mRNA as a splicing factor by binding to its own pre-mRNA in the region of the circularizing exon (Ashwal-Fluss et al. 2014). The cytoplasmic localization of circRNAs, observed in numerous studies, raises questions as to the mechanism of circRNA export from the nucleus.

Apart from miRNA sponge and autoregulatory functions we can currently only speculate on the functional relevance and diversity of circRNAs. Some might act as protein sponges (Fig. 4) to inactivate specific proteins, possibly combined with a controlled release option, for example through the regulated degradation of the circRNA component. Other (hypothetical) features of circRNAs include regulatory antisense activities, circRNA-mediated assembly of protein complexes, a role as transport vehicles or allosteric regulation of proteins or other RNAs *via* aptamer domains as part of circRNAs (aptamer vector, see Fig. 4 and summary by Hentze and Preiss 2013). Whether the numerous circRNAs identified by deep sequencing in fact exist in biologically significant copy numbers, which circRNAs are functional in which cell types and tissues, and how broad the functional

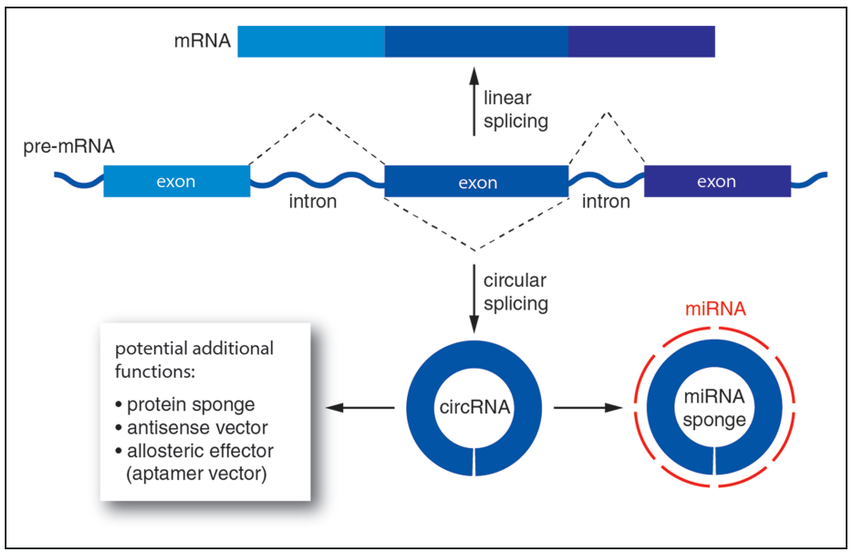


Fig. 4: Biogenesis and functions of circular RNAs (circRNAs). CircRNAs are generated by an alternative splicing process, in which single exons, or adjacent multiple exons, of a pre-mRNA circularize, instead of being spliced in a linear manner to produce a protein-coding mRNA. Such exon-derived circRNAs normally accumulate in the cytoplasm, where they could act as miRNA sponges, if they carry a cluster of miRNA binding sites. By binding to the circRNA, the miRNA would be inactivated. Other potential functions, still completely hypothetical, include roles as vectors for antisense sequences or as vectors for aptamer RNAs, which may allosterically regulate the function of an RNA-binding protein or another RNA (aptamer vector).

spectrum of circRNAs will turn out to be are currently open questions.

## 5. New targets for diagnosis and therapy?

How can we utilize ncRNAs as biomarkers for the diagnosis of human diseases and for stratifying patient groups in personalized medicine? An important criterion for the suitability as biomarker is a cell type- or tissue-specific expression pattern, which indeed applies to many ncRNAs. Numerous well-characterized miRNAs and also some lncRNAs have already been implemented as disease-specific biomarkers. Due to their high metabolic stability, circRNAs are particularly attractive as biomarkers in body fluids with considerable ribonuclease activities, such as blood, urine or liquor. Note that RNA-based biomarkers do not require any prior knowledge on the RNA's detailed function or biological role. The high suitability of RNA-based biomarkers is based on the following features:

(a) Simple and rapid development of assays: inexpensive high-throughput screening approaches can easily be set up, since DNA oligonucleotides for RT-PCR assays can be rapidly and cost-effectively synthesized. This is, for example, an advantage over antibody-based protein detection assays, which require a more cost-intensive development of specific antibodies. Furthermore, RNA-based RT-PCR assays require only minimal amounts of biological material.

(b) High specificity and sensitivity of assays as well as simple quantification of specific RNAs by real-time PCR assays, even allowing determination of absolute copy numbers.

(c) Localization: Specific ncRNAs are detectable by *in situ* hybridization, either in tissue samples (e.g. derived from tumor biopsies) or even in single cells and at the single-molecule level. NcRNAs are embedded in regulatory networks and strongly participate in developmental and differentiation processes, such that their altered expression often potentiates disease progression. Therefore, understanding the functions of ncRNAs opens up new perspectives for therapeutic interventions. For miRNAs, a relatively well characterized class of ncRNAs, specific therapeutic strategies have already been developed and are about to enter the clinic. An example is miR-122, a miRNA specifically expressed in the liver, which stimulates (instead of suppressing) translation of hepatitis C virus (HCV) RNA. In this case, modified antisense oligonucleotides were designed to specifically block miR-122 and thereby HCV replication. *Miravirsen*, an LNA *(locked nucleic acid)*-modified antisense oligonucleotide that functions as a miR-122 antagonist (generally termed "antimiR"), is presently in advanced clinical testing phases (Janssen et al. 2013; Conrad and Niepmann 2014). Other variants of miRNA-specific antisense oligonucleotides developed for therapeutic applications are the so-called *antagomiRs* (see article by Grünweller and Hartmann in this issue).

CircRNAs might as well be utilized for therapeutic applications. One may envision "designer" circRNAs, either expressed *in vivo* or chemically synthesized, which serve as sponges for specific miRNAs or RNA-binding proteins. Conversely, by using antisense nucleic acids, one may even be able to induce the linearization of circRNAs to trigger their degradation and the release of bound miRNAs or RNA-binding proteins. Finally, circRNAs also appear suitable as new antisense vectors, designed to block or modulate translation of specific mRNAs or alternative splicing processes.

### References

Anderson DM, Anderson KM, Chang CL, Makarewich CA, Nelson BR, McAnally JR, Kasaragod P, Shelton JM, Liou J, Bassel-Duby R, Olson EN (2015) A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. Cell 160: 595–606.

Andrews SJ, Rothnagel JA (2014) Emerging evidence for functional peptides encoded by short open reading frames. Nat Rev Genet 15: 193–204.

Ashwal-Fluss R, Meyer M, Pamudurti NR, Ivanov A, Bartok O, Hanan M, Evantal N, Memczak S, Rajewsky N, Kadener S (2014) CircRNA biogenesis competes with pre-mRNA splicing. Mol Cell 56: 55–66.

Burd CE, Jeck WR, Liu Y, Sanoff HK, Wang Z, Sharpless NE (2010) Expression of linear novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. PLoS Genet 6: e1001233.

Capel B, Swain A, Nicolis S, Hacker A, Walter M, Koopman P, Goodfellow P, Lovell-Badge R (1993) Circular transcripts of the testis-determining gene Sry in adult mouse testis. Cell 73: 1019–1030.

Cech TR, Steitz JA (2014) The noncoding RNA revolution-trashing old rules to forge new ones. Cell 157: 77–94.

Cheetham SW, Gruhl F, Mattick JS, Dinger ME (2013) Long noncoding RNAs and the genetics of cancer. Br J Cancer 108: 2419–2425.
Conrad KD, Niepmann M (2014) The role of microRNAs in hepatitis C virus RNA replication. Arch Virol 159: 849–862.
Danan M, Schwartz S, Edelheit S, Sorek R (2012) Transcriptome-wide discovery of circular RNAs in Archaea. Nucleic Acids Res 40: 3131–3142.
Filipowicz W, Bhattacharyya SN, Sonenberg N. (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat Rev Genet 9: 102–114.
Garzon R, Calin GA, Croce CM (2009) MicroRNAs in Cancer. Annu Rev Med 60: 167–179.
Glazar P, Papavasileiou P, Rajewsky N (2014) circBase: a database for circular RNAs. RNA 20: 1666–1670.
Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J (2013) Natural RNA circles function as efficient microRNA sponges. Nature 495: 384–388.
Hansen TB, Wiklund ED, Bramsen JB, Villadsen SB, Statham AL, Clark SJ, Kjems J (2011) miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. EMBO J 30: 4414–4422.
Hentze MW, Preiss T (2013) Circular RNAs: splicing's enigma variations. EMBO J 32: 923–925.
Huntzinger E, Izaurralde E (2011) Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat Rev Genet 12: 99–110.
Janssen HL, Reesink HW, Lawitz EJ, Zeuzem S, Rodriguez-Torres M, Patel K, van der Meer AJ, Patick AK, Chen A, Zhou Y, Persson R, King BD, Kauppinen S, Levin AA, Hodges MR (2013) Treatment of HCV infection by targeting microRNA. New Engl J Med 18: 1685–1694.
Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, Marzluff WF, Sharpless NE (2013) Circular RNAs are abundant, conserved, and associated with ALU repeats. RNA 19: 141–157.
Jensen TH, Jacquier A, Libri D (2013) Dealing with pervasive transcription. Mol Cell 52: 473–484.
Kjems J, Garrett RA (1988) Novel splicing mechanism for the ribosomal RNA intron in the archaebacterium Desulfurococcus mobilis. Cell 54: 693–703.
Lasda E, Parker R (2014) Circular RNAs: diversity of form and function. RNA 20: 1829–1842.
Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M, Loewer A, Ziebold U, Landthaler M, Kocks C, le Noble F, Rajewsky N (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. Nature 495: 333–338.
Nagano T, Fraser P (2011) No-nonsense functions for long noncoding RNAs. Cell 145: 178–81.
Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature 403: 901–906.
Rinn JL, Chang HY (2012) Genome regulation by long noncoding RNAs. Annu Rev Biochem 81: 145–166.
Sänger HL, Klotz G, Riesner D, Gross HJ, Kleinschmidt AK (1976) Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. Proc Natl Acad Sci USA 73: 3852–3856.
Starke S, Jost I, Rossbach O, Schneider T, Schreiner S, Hung L-H, Bindereif A (2015) Exon circularization requires canonical splice signals. Cell Rep 10: 103–111.
Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA, Blencowe BJ, Prasanth SG, Prasanth KV (2010) The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol Cell 39: 925–938.
Zamore PD, Haley B (2005) Ribo-gnome: the big world of small RNAs. Science 309: 1519–1524.