The new world of RNAs

Danyella Barbosa Dogini¹, Vinícius D’Avila Bittencourt Pascoal², Simoni Helena Avansini¹, André Schwambach Vieira¹, Tiago Campos Pereira³ and Iscia Lopes-Cendes¹

¹Departamento de Genética Médica, Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, SP, Brazil.
²Faculdade de Ciências Básicas, Universidade Federal Fluminense, Nova Friburgo, RJ, Brazil.
³Departamento de Biologia, Faculdade de Filosofia Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, Brazil.

Abstract

One of the major developments that resulted from the human genome sequencing projects was a better understanding of the role of non-coding RNAs (ncRNAs). NcRNAs are divided into several different categories according to size and function; however, one shared feature is that they are not translated into proteins. In this review, we will discuss relevant aspects of ncRNAs, focusing on two main types: i) microRNAs, which negatively regulate gene expression either by translational repression or target mRNA degradation, and ii) small interfering RNAs (siRNAs), which are involved in the biological process of RNA interference (RNAi). Our knowledge regarding these two types of ncRNAs has increased dramatically over the past decade, and they have a great potential to become therapeutic alternatives for a variety of human conditions.

Keywords: non-coding RNAs, microRNAs, RNA interference, RNA-based drugs.

Introduction

One of the main goals of the Human Genome Project was to identify all human genes (Collins et al., 1998). It was estimated that the human genome contained approximately 100,000-150,000 genes. However, only nearly 20,000 genes could be identified at the end of the project; therefore, protein-coding sequences represent only 1.5% of the human genome. The rest of the genome, approximately 98%, was considered junk DNA because it is composed of sequences that are not translated into proteins. These sequences were believed to be mainly part of regulatory regions or non-coding regions that did not play an important role in cell function (Lander et al., 2001). These conclusions were based on the central dogma of molecular biology (Crick, 1958), which considers that all relevant information contained in the DNA will be transcribed into messenger RNA (mRNA) molecules that will be subsequently translated into proteins, which are the important molecular players in cell function (including the regulation of gene expression). The concept of two main classes of RNA molecules already existed: i) RNAs which were translated into proteins, i.e., mRNAs, and ii) the group of non-coding RNAs (ncRNAs), mainly comprising transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs). However, as additional studies were performed, it became clear that ncRNAs were much more abundant than expected, and they were present in several different organisms, with a particularly high abundance in Homo sapiens (Hüttenhofer et al., 2005). These findings prompted the development of a new specific field in molecular biology, the study of ncRNAs, which has become increasingly relevant in the past decade. As knowledge advances, it becomes clear that discoveries in the field of ncRNAs are likely to make significant contributions to the biomedical sciences, including the possibility of novel therapeutic alternatives for a variety of human conditions.

Non-coding RNAs

The length of ncRNAs can vary from 21 to several thousand nucleotides (nt) and these molecules are divided into i) long or large RNAs, such as transfer RNA, ribosomal RNA and X-inactive specific transcript RNA (XIST RNA), and ii) small ncRNAs, such as microRNAs (miRNAs), small interfering RNAs (siRNAs), repeat associated small interfering RNA (rasiRNAs), small nucleolar RNAs (snoRNAs), small nuclear RNAs (snRNAs), piwi-interacting RNAs (piRNAs) and others (Gavazzo et al., 2013). Table 1 summarizes the different types of ncRNAs described, as well as some of their main characteristics.
siRNAs are approximately 21 nt in length and are produced from the processing of double-stranded RNA molecules by the enzyme Dicer. siRNAs are involved in gene regulation, as well as viral defense and transposon activity (Ghildiyal and Zamore, 2009; Malone and Hannon, 2009). rasiRNAs are approximately 24-27 nt long and play a role in heterochromatin orientation during the formation of centromeres (Theurkauf et al., 2006; Josse et al., 2007). snoRNAs consist of approximately 80 nt and are believed to be involved in the guidance of uridylation and/or site-specific methylation during ribosomal RNA maturation (Dieci et al., 2009; Taft et al., 2009); in addition, there is evidence that snoRNAs can also play a role in the regulation of gene expression (Matera et al., 2007). snRNAs have been shown to be part of the spliceosome complex (Peters and Robson, 2008; Tazi et al., 2009) and are important for removing introns from immature mRNAs (Valadkhan, 2005). piRNAs are approximately 26 to 30 nt in length, are restricted to germ line cells and function with AGO and PIWI proteins to regulate transposon activity and chromatin states (Ghildiyal and Zamore, 2009; Malone and Hannon, 2009). PiRNAs are also highly expressed in mammalian cells at the pachytene stage (Megosh et al., 2006).

Next, we will present additional information about miRNAs and siRNAs, because these are better characterized in terms of their functions and their impact in human health and disease.

**miRNAs**

MicroRNAs (miRNAs) are small endogenous ncRNAs that regulate gene expression post-transcriptionally in a sequence-specific manner (Bartel, 2004). In 1993, Victor Ambros and colleagues observed a mutant strain of Caenorhabditis elegans that exhibited developmental abnormalities, such as the inability to form the vulva (Lee et al., 1993). These authors observed that the gene responsible for these phenotypes, lin-4, had two transcripts: one, which was more abundant and 61 nt in length and a second 21 nt transcript that was not translated. lin-14 was shown to encode a nuclear protein that participates in the regulation of the transition from the first (L1) to the second (L2) larval stages in C. elegans (Ruvkun and Giusto, 1989; Lee et al., 1993). Subsequent studies revealed that the 21 nt transcript is complementary to the 3’ untranslated region (3’UTR) of lin-14 and, most interestingly, negatively regulates the expression of lin-14 (Lee et al., 1993). Initially, these findings were not adequately appreciated by the scientific community, because it was believed to be a rare process occurring only in C. elegans. However, in 2000, another such 22 nt non-coding RNA, let-7, was identified in C. elegans (Reinhart et al., 2000). Surprisingly, it was also found to be complementary to the 3’UTR of another gene, lin-41, promoting its translational knockdown. In addition, the let-7 sequence was shown to be highly conserved in most organisms, including non-nematodes (Pasquinelli et al., 2000). These findings officially started a new field of investigation in small ncRNAs, more specifically, miRNAs. Currently, the existence of miRNAs has been extensively shown in insects and mammals (Ambros, 2001; Ambros et al., 2003; Lagos-Quintana et al., 2003; Mattick and Makunin, 2005), and the database of miRNAs has registered approximately 25,000 sequences found in 32 organisms representing vertebrates, invertebrates, plants and viruses (http://www.mirbase.org/).

**miRNA biogenesis**

Most miRNAs described to date are transcribed from sequences present in intergenic regions (Lagos-Quintana et al., 2001; Lau et al., 2001), but 25% of human miRNAs identified are located in intronic regions and are transcribed in the same direction as the pre-messenger RNA, leading to the hypothesis that these miRNAs may use the promoter region of mRNAs for their transcription (Aravin et al., 2003; Lai et al., 2003).

miRNAs are transcribed by RNA polymerase II from what are called MIR genes. The first intermediate transcripts are ‘hairpin’ molecules called pri-miR. Similar to mRNAs, these transcripts undergo capping and polyadenylation at the 5’ and 3’ ends of the transcript, respectively (Lee et al., 2002; Cai et al., 2004; Kim, 2005).

### Table 1 - Different types, main characteristics and functions of non-coding RNAs.

<table>
<thead>
<tr>
<th>Type</th>
<th>Mean size</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long ncRNA</td>
<td>~1.9 kb</td>
<td>Essential for protein synthesis</td>
</tr>
<tr>
<td>XIST RNA</td>
<td>~17 kb</td>
<td>Chromosome X inactivation</td>
</tr>
<tr>
<td>Other IncRNA</td>
<td>&gt; 200 nt</td>
<td>Involved in epigenetic modification, post-transcriptional processing, modulation of chromatin structure, etc.</td>
</tr>
<tr>
<td>Small ncRNA</td>
<td>18-21 nt</td>
<td>Gene regulation</td>
</tr>
<tr>
<td>miRNAs</td>
<td>~21 nt</td>
<td>Gene regulation; defense against viruses and transposon activity</td>
</tr>
<tr>
<td>siRNA</td>
<td>24-27 nt</td>
<td>Orientation of heterochromatin in the formation of centromeres</td>
</tr>
<tr>
<td>rasiRNA</td>
<td>60-300 nt</td>
<td>Methylation and pseudo uridylation of other RNAs</td>
</tr>
<tr>
<td>snoRNA</td>
<td>100-300 nt</td>
<td>Involved in spliceosome complex</td>
</tr>
<tr>
<td>snRNA</td>
<td>26-30 nt</td>
<td>Regulation of transposon activity and chromatin state</td>
</tr>
<tr>
<td>piRNA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Unlike mRNAs, miRNA maturation begins in the nucleus and finishes in the cytoplasm. In animals, the nuclear processing is performed by DROSHA, an RNase type III enzyme with endonucleolytic activity, which, in combination with PASHA (Lee et al., 2003), recognizes the ‘hairpin’ structure of the pri-miR and cleaves it. This generates a pre-miR of approximately 60-70 nt in length. Pre-miRs are transported to the cytoplasm by exportin-5 and the cofactor Ran-GTP (Yi et al., 2003; Lund et al., 2004; Shibata et al., 2006). Once in the cytoplasm, pre-miRs are processed by the enzyme DICER, thus generating dimeric miRNA:miRNA* molecules approximately 22 nt in length (Hutvagner et al., 2001). Whereas mammals typically encode a single DICER that can generate several classes of small RNAs, Drosophila and C. elegans have two types of Dicer (Lee et al., 2004). Plants require multiple types of Dicers; four and six different classes of DICER-like (DCL) enzymes have been identified in Arabidopsis thaliana and rice, respectively. In Arabidopsis, DCL 2, 3 and 4 are involved in the formation of different species of siRNAs, whereas DCL 1 is exclusively responsible for the biogenesis of miRNAs (Bernstein et al., 2001). In addition, it has been demonstrated that each Dicer produces siRNAs with specific sizes: DCL2, DCL4 and DCL3 form siRNAs with 22, 24 and 21 nt, respectively (Xie et al., 2005; Deleris et al., 2006; Blevins et al., 2006). The cleavage performed by DICER and Drosha will ultimately define the dimer miRNA:miRNA* extensity; the strand with a lower free energy at the 5’ end is incorporated by RISC (RNA-induced silencing complex), while the other strand is degraded (Khrvorova et al., 2003). The miRNA bound to the RISC structure is guided to the complementary sequence of the target mRNA. If the pairing is complete (100% sequence complementarity between the miRNA and mRNA), the miRNA will be degraded; however, if the sequence is only partially complementary, this will result in translational inhibition of the mRNA (Cannell et al., 2008).

One unique aspect of miRNA regulation is its complexity. It has been observed that a single miRNA can regulate expression of different mRNAs. Additionally, one mRNA can be regulated by multiple miRNAs (Yanaihara et al., 2006). In C. elegans, the miRNA lin-4 regulates the expression of the lin-14 gene as well as the heterochronic gene lin-28. A homolog of lin-14 exists in animals, and in mouse and human cells, it is regulated by miR-125a and let-7b (Moss and Tang, 2003).

miRNAs can also be produced by an alternative pathway. In this new model, sequences present in introns are capable of being transcribed as miRNAs. This hypothesis is derived from experiments in which an analysis was performed in pooled sequences from Drosophila S2 cells, leading to the mapping of miR:miR* duplexes (Ruby et al., 2007). It is believed that nearly 80% of animal miRNAs are transcribed from introns, and these miRNAs are known as mirtrons (; Rodriguez et al., 2004; Kim and Kim, 2007; Okamura et al., 2007; Ruby et al., 2007). The formation of mirtrons differs from classical miRNA biogenesis because it does not require DROSHA proteins (Han et al., 2006; Kim and Kim, 2007; Wang et al., 2007); instead, mirtrons require AGO1 proteins for maturity (Okamura et al., 2007).

The pathway through which miRNAs regulate mRNA translation involves the mechanism used by cells as a defense against exogenous mRNA (such as in a viral infection), called the post-transcriptional gene silencing (PTGS) pathway (Fire et al., 1998). When used as a defense mechanism, PTGS induces the cleavage of double-stranded RNA (dsRNA) and allows for translational inhibition of exogenous mRNA (Hamilton and Baulcombe, 1999). dsRNAs are cleaved into smaller molecules, small interfering RNAs (siRNAs), which are subsequently associated with RISC, leading to gene silencing by RNA interference (RNAi). This will be discussed in more detail in a later section.

miRNA function

MiRNAs exhibit temporal- and tissue-specific expression (Lagos-Quintana et al., 2002; Lagos-Quintana et al., 2003; Liu et al., 2005; Mehler and Mattick, 2006; Schrott et al., 2006). One of the first functions associated with miRNA was the temporal regulation of development (Lee et al., 1993; Wightman et al., 1993). Negative post-transcriptional regulation of lin-14 by the lin-4 miRNA is essential for the formation of a temporal gradient of LIN-14 protein to ensure correct transition between the larval stages in C. elegans. The second miRNA discovered, let-7, also proved to be a key driver in the temporal pattern of development of C. elegans (Abrahante et al., 2003; Lin et al., 2003; Grosshans et al., 2005).

To date, it has been shown that miRNAs regulate the expression of at least 1/3 of all human genes, and these are known to play an important role in several biological processes, including cell cycle regulation, apoptosis, cell differentiation, and embryonic development, etc. (Ketting et al., 2001; Wienholds et al., 2003; Ambros, 2004; Bartel, 2004; Lee et al., 2004; Wienholds and Plasterk, 2005). De-regulation of miRNA expression is often associated with human cancers (Volinia et al., 2006; Lu et al., 2008) and can induce activation of oncogenes or inactivation of tumor suppressor genes (Esquela-Kerscher and Slack, 2006). DICER-deficient mice die because they lose their stem cell pluripotency (Bernstein et al., 2003). Experiments with mutant DICER-1 in Drosophila show that the miRNA pathway is critical for cell division and for the passage from the G1 phase to the S phase of the cell cycle (Hatfield et al., 2005). There is also evidence of miRNA involvement in metabolism as well as in the regulation of apoptosis. In flies, the miRNA bantam accelerates proliferation and prevents apoptosis by regulation of a proapoptotic gene, hid (Brennecke et al., 2003). In vertebrates, miR-375 is ex-
pressed in pancreatic islets and suppresses the secretion of insulin induced by glucose (Poy et al., 2004).

miRNAs have also been shown to play an important role in myogenesis and cardiogenesis. miR-1 is conserved from worms to mammals and is highly expressed in human muscle, fly muscle and the mouse heart (Aboobaker et al., 2005; Zhao et al., 2008). Interestingly, knockdown of this miRNA in Drosophila does not affect the formation and function of muscles during the larval stages, but instead affects the formation of muscles during growth in the adult animal (Sokol and Ambros, 2005). These results clearly show the role of miRNAs not only during development but also for further growth and tissue maintenance.

The association between miRNA deregulation and the development of pathological states was first discovered through studies in the field of oncology. One of the initial findings that showed that miRNAs indeed play a role in the regulation of oncogenes came from studies in chronic lymphocytic leukemia (Calin et al., 2002). These studies showed that several human miRNA transcription sites are located in genomic regions involved in cancers, such as regions of chromosomal breakpoints and fragile sites (Calin et al., 2004). Further studies demonstrated that differential expression of miRNAs is associated with tumor formation (Michael et al., 2003; Calin et al., 2004, 2005; Lu et al., 2005). In a few cases, a correlation between specific miRNA expression patterns and cell type could be found (Lu et al., 2005). In addition, most miRNAs are found to be down-regulated in tumor tissue when compared with normal tissue, which may lead to loss of cell differentiation (Lu et al., 2005).

Recent studies have demonstrated that some viruses encode miRNAs (Pfeffer et al., 2004) and that in the Epstein-Barr virus (EBV), a member of the herpesvirus family, these miRNAs are encoded in intergenic regions at specific clusters (Edwards et al., 2008; Feederle et al., 2011). These virus-encoded miRNAs have been shown to regulate their own genes (Barth et al., 2008; Umbach et al., 2008), as well as genes involved in virus-cell interactions (Murphy et al., 2008), leading to a modulation of the host immune system (Stern-Ginossar et al., 2007). In this way, it has been demonstrated that the BHRF1 miRNA cluster plays an important role in the transition from the latent virus state by enhancing expansion of the virus reservoir and reducing the viral antigenic load (Feederle et al., 2011). Therefore, these features have the potential to facilitate persistence of the virus in the infected host and can be used as new therapeutic targets for the treatment of some EBV-associated lymphomas (Feederle et al., 2011).

To better understand the function of miRNAs, it is also important to know their regulatory targets (i.e., the genes regulated by specific miRNAs). Because it has been estimated that each miRNA could regulate a large number of targets (Kim, 2005; Baek et al., 2008), bioinformatic algorithms have become a powerful tool for identifying miRNA-regulated genes and predicting gene function. Therefore, a large number of predictive algorithms are available, such as TargetScan (Lewis et al., 2003), miranda (John et al., 2004), Pictar (Krek et al., 2005), RNA22 (Miranda et al., 2006), PITA (Kertesz et al., 2007), DIANA-microT (Maragkakis et al., 2009) and Tarbase (Hsu et al., 2011). The main algorithm used to predict miRNA:miRNA interactions involves pairing of the 5' region of the miRNA - a 2-8 nt region known as the 'seed region' - to the 3' untranslated region (3'-UTR) of the mRNA (Thomson et al., 2011). However, evidence suggests that the miRNA seed region pairing is not always a reliable predictor of miRNA:miRNA interactions (Didiano and Hobert, 2006). Indeed, precision of these algorithms is estimated to be approximately 50% when tested against experimental-proven miRNA targets (Alexiou et al., 2009). Therefore, it is highly suggested that additional experiments should be performed to validate potential miRNA targets. These may include, but are not limited to, reporter gene assays, evaluation of miRNA and target mRNA co-expression (e.g., northern blotting, qPCR or in situ hybridization) and assessment of miRNA effects on target protein expression (e.g., ELISA, western blotting, immunohistochemistry) (Kertesz et al., 2007; Kuhn et al., 2008; Nuovo, 2010; Thomson et al., 2011 Hébert and Nelson, 2012; Vergoulis et al., 2012).

Small Interfering RNAs and RNA Interference

RNAi is frequently used as a technique to promote effective and specific post-transcriptional gene silencing through the administration of double-stranded RNAs (dsRNAs). Although the phenomenon was first observed in plants and fungi, the clear triggering mechanism was originally reported in the nematode C. elegans (Fire et al., 1998). In this work, the authors observed that after injecting long dsRNA molecules into the worm’s gonad, the matching target mRNA was destroyed and a corresponding phenotypical change could be subsequently observed (Fire et al., 1998).

After the initial description, RNAi was shown to be functional in nearly all eukaryotic species tested so far. Indeed, RNAi has been shown to be widely present, from viruses, unicellular organisms, fungi, plants and other animals. The technology promoted a revolution in molecular biology and medical sciences because it can be used to identify gene function or to silence essential genes present in a pathogen. The impact of RNAi is such that less than a decade after the seminal report, the discoverers were awarded the Nobel Prize. Investors also noted the immense potential behind this technique. As a result, several biotechnology start-ups emerged, devoted to the development of RNAi-based therapies (Check, 2004; Bonetta, 2007; Osborne, 2007).

RNAi can be triggered by two main types of double stranded RNA molecules. The first class encompasses long
molecules, approximately 300-800 base pairs (bp) in length, which may be produced by several processes, such as: i) RT-PCR followed by in vitro transcription (Goto et al., 2003), ii) expression from a cDNA cloned in special vectors (Kamath et al., 2001) or iii) a transgenic cassette (Chuang and Meyerowitz, 2000). These long dsRNAs are the molecules of choice when using the technology in non-mammalian models. Because dsRNAs longer than 30 bp promote lethal effects in mammalian cells, 21nt RNA duplexes, known as siRNAs, are the molecule of choice for use in mammals (Elbashir et al., 2001). siRNAs can also be used in non-mammalian cells, but this molecule must be designed for the target gene and its functionality must first be tested in vitro.

RNAi can be used to create genetically modified animals in an attempt to recapitulate the null phenotype. Moreover, hypomorphous animals, i.e., displaying intermediate levels of mRNA knockdown (from 0.1 to 99%), may also be generated via RNAi, because the efficiency of silencing can be controlled (Khvorova et al., 2003, Schwarz et al., 2003). Such genetic constructions, presenting intermediary phenotypes, may be of great biological value when the null animal is not viable (Baker et al., 2004).

RNAi is frequently used as a strategy to identify gene function, but there are many other possibilities: i) to combat several classes of pathogens (viruses, Palliser et al., 2006; bacterial diseases, Escobar et al., 2001; parasites, Pereira et al., 2008), ii) to generate plants and animals of interest (Minton, 2004; Peng et al., 2006) and iii) to control genetic diseases and tumors (Ptasznik et al., 2006; Raoul et al., 2005). All these new developments have recently led to the first published human clinical trials, with very promising results (Koldehoff et al., 2007; Koldehoff and Elmagabi, 2009; Davis et al., 2010; DeVincenzo et al., 2010; Leachman et al., 2010).

Future Directions

Progress in the area of ncRNAs seems to have occurred faster than in any major biological discipline in recent memory. The field has moved from virtual ignorance about an abundant class of regulatory molecules to a reasonably advanced understanding of the mechanisms of miRNA biogenesis and an emerging consensus about the numbers of miRNAs and their targets in several species, including humans. Recently, miRNAs have been used as a biomarker for several diseases. This is one of the most promising approaches in the use of these molecules (Zho and Wang, 2010; Cheng et al., 2011; Ohyashiki et al., 2011).

RNAi has promoted an enormous advancement in the field of molecular biology in the past decade. This technique allows a fast, cost-effective and simple alternative to promote down-regulation of virtually any gene from many species. RNAi-based drugs constitute the next big gamble of pharmaceutical companies for two main reasons: i) the promise of being highly specific, because RNAi relies on total sequence complementarity, and ii) the fact that the associated pharmacodynamics may not be problematic, because RNA is a biological molecule. Therefore, it is very likely that RNAi may lead to innovative medical treatments in the near future.

Acknowledgments

Dr. Iscia Lopes-Cendes is supported by grants from FAPESP, FAPESP-CEPID (BRAINN), CNPq and EpimirNA International Project.

References


License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.