Discovery of novel small RNAs in the quest to unravel genome complexity

Adam E. Hall and Tamas Dalmay1
School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, U.K.

Abstract

The different RNA populations within the cell, or the ‘RNAome’, form a complex and integral layer of gene regulation. In particular, small non-coding RNAs such as microRNAs have been shown to regulate as much as half of all genes expressed in mammals. The recent discovery of novel small RNAs derived from tRNAs, snoRNAs and Y RNAs, as well as the development of techniques that can reduce sequencing bias of these molecules, is slowly helping us to unveil the secrets of the genome.

A genome governed by RNA

One of the most striking discoveries in recent times was the realization of just how few genes there are in the human genome [1]. Compared with the predicted 100000 protein-coding genes originally estimated, only 20687 were finally identified recently, following the completion of the ENCODE (Encyclopaedia of DNA Elements) project [2]. Paradoxically, the human proteome is likely to hold an estimated 500000 unique protein variants on the basis of investigations into yeast and fruitfly genomes [3,4]. The complexity of the proteome relative to the modest repertoire of bona fide genes is now becoming clear.

There is an overlapping arrangement of genetic information that is achieved by several mechanisms [5]. First, variable transcription start sites for the same locus can give rise to transcripts that share a common sequence, but are functionally independent. Secondly, genes can be differentially ‘spliced’ to give rise to varied isoforms. Thirdly, a single transcript can contain several genetic units, which are differentially processed and serve separate biological roles.

Over the last 10 years, small RNAs have emerged as molecules that have a big regulatory effect, further deepening the level of complexity by which genes are expressed. In the present review we discuss some of these small regulatory molecules and how NGS (next generation sequencing) has identified other novel RNAs whose function has yet to be determined. We also discuss how the detection of these molecules can be improved using a modified library generation protocol for small RNA cDNAs.

The advent of ‘deep sequencing’ and the discovery of the ‘small RNA world’

The birth of sRNA (small non-coding RNA) biology (<200 nt) came with the discovery that small double-stranded RNA molecules can bind to complementary mRNA targets and inhibit their expression into proteins [RNAi (RNA interference)] [6–8]. Since then, the advent of small RNA cloning and sequencing techniques on the basis of 3′ and 5′ adapter ligation and subsequent RT (reverse transcription)-PCR [9] have allowed ‘deep sequencing’ of RNA samples. Various NGS platforms such as Illumina (Solexa) and 454 pyrosequencing have been developed [10], and at present, the most popular HiSeq2000 platform can yield 100–200 million reads per sample. The improvement of this technology, the reduction of running costs and the increase in availability have resulted in the discovery of millions of novel small RNAs [11]. This in turn has completely revolutionized our view of how the cell controls the flow of genetic information and the hierarchical structure by which this is established.

miRNAs (microRNAs): the most ‘famous’ small RNAs

miRNAs are some of the most abundant well-studied small RNAs, and have long been the focus of thousands of deep sequencing studies. miRNAs are endogenously produced post-transcriptional regulators that bind to the 3′ UTR (untranslated region) of their targets. They are ~22 nt in length and form distinct stem–loop structures [12]. miRNAs are thought to regulate at least 30%, and maybe as much as half, of mammalian protein-coding genes with each miRNA
being able to regulate hundreds of targets [13]. miRNAs act either by inhibiting translation of semi-complementary transcripts in animals or by catalysing cleavage of highly complementary transcripts in plants.

Most miRNAs are transcribed by RNA polymerase II [14], producing pri-miRNAs (primary miRNAs). The pri-miRNA is enzymatically cleaved by the Microprocessor complex [15], which is composed of the RNase III-like enzyme Drosha and the RNA-binding protein DGCR8 (or pasha in Drosophila). The resultant 60–80 nt pre-miRNA (precursor miRNA) is characterized by a 3′ end 2 nt overhang. The role of the Microprocessor complex is redundant for some miRNAs, where the pre-miRNA forms a debranched intron and is generated during splicing. These ‘mirtrons’ can be transported directly to the cytoplasm [16].

The pre-miRNA is exported into the cytoplasm by the nuclear-envelope-bound exportin-5 protein [17]. In the cytoplasm, another RNase III enzyme, Dicer, along with the HIV-1 TRBP (transactivating response RNA-binding protein), processes the pre-miRNA into a 20–22 nt miRNA–miRNA* duplex [18]. Although Dicer is involved in the canonical miRNA maturation pathway, at least one miRNA circumvents the use of Dicer altogether [19].

The miRNA duplex contains a guide miRNA and a passenger miRNA*. Animal miRNAs are semi-complementary to their mRNA targets, and contains a 2–8 nt ‘seed’ region at the 5′ end. The seed region of the miRNA has a near perfect degree of complementarity to the MRE (miRNA recognition element) at the 3′ UTR of the mRNA [20]. The passenger miRNA* strand is usually discarded; however, several miRNA* strands can be actively selected for RISC (RNA-induced silencing complex) incorporation [21], TRBP also functions as a recruiting factor by bringing the Dicer–miRNA complex and the AGO (Argonaute protein)-based silencing complex together, thereby forming the RLC (RISC-loading complex) [22].

Once loaded into RISC, the process of gene silencing can take place. There is still an uncertainty about how exactly miRNA gene silencing is executed in animals, and the prevalence of translational repression compared with mRNA transcript degradation [23,24].

The revelation of extensive RNA fragmentation in cells

Since 2008, there has been an abundance of deep sequencing studies, which have identified small RNA fragments derived from longer RNAs, and this has been the subject of several excellent reviews [5,11,25]. Previous deep sequencing studies overlooked these fragments, as they were filtered out from datasets as they were assumed to be degradation products. For example, snoRNAs (small nucleolar RNAs), which are involved in ribosome biogenesis, have been shown to give rise to RNA fragments that can act as miRNAs [26,27]. The first sDNA (snoRNA-derived RNA) was discovered in HEK (human embryonic kidney)-293 cells by deep sequencing of immunoprecipitated AGO1- and AGO2-associated RNAs [26]. This sDNA was found to be processed by Dicer (but independently of Drosha) and its RNAi function was demonstrated by luciferase assay. However, other sDNAs show potential Dicer-independent maturation [27]. Further sDNA discovery confirmed that these 18–35 nt RNAs primarily mapped to the 5′ and 3′ stem of the full-length snoRNA parent [27].

Indeed, sRNA fragments from tRNAs, snoRNAs, snRNAs (small nuclear RNAs) and tRNAs seem to give fragments predominantly from the 5′ and 3′ ends of mature parental RNA transcripts [28]. This seems to occur asymmetrically in human and mouse cells, with one end appearing more prominently in sequencing studies (similar to miRNA–miRNA* level bias) [28].

Additional AGO immunoprecipitation and deep sequencing studies uncovered small RNA fragments derived from rRNAs, snRNAs, vRNAs (vault RNAs), tRNAs and even mRNAs, all of which were associated with the silencing protein family [29].

Vault particles are cellular organelles thought to be involved in nuclear export and play a role in multidrug resistance. They contain the 88–98 nt non-coding vRNAs. It was shown that these vRNAs can be cleaved to give rise to svRNAs (small vault-derived sRNAs) (23 nt), independently of Dicer processing [30]. Intriguingly, one of these svRNAs associates with an AGO protein and can down-regulate the mRNA encoding enzyme CYP3A4 (cytochrome P450 3A4), which has a role in drug metabolism [30].

An analysis of a large collection of publicly available independent deep sequencing datasets has revealed that IncRNAs (long non-coding RNAs) can be processed into sRNA fragments [31]. The sRNA clusters predominantly mapped to the 3′ end of IncRNAs. Interestingly, IncRNA-derived sRNA fragments were shown to be differentially processed in a variety of cell and tissue types, implying functionality rather than being simple products of degradation [31].

tRNA fragments

To date, the most well-studied sRNA fragments are those derived from the tRNAs. The first reports of tRNA cleavage came in a study which showed that tRNA cleavage in Escherichia coli occurs following infection by bacteriophage T4 [32]. A more recent deep sequencing study in an animal system has found that sRNA fragments derived from tRNAs are more abundant than miRNAs and are not simply the product of degradation pathways [33,34].

There are two distinct classes of tRNA fragments; the tRFs (tRNA regulatory fragments) and the tiRNAs or ‘tRNA halves’ (stress-induced tRNA-derived RNAs). The tRFs are composed of three types of fragments (17–26 nt), which were first discovered during deep sequencing of prostate cancer cell lines [34]. The tRF-1 species is released by RNase Z (ELAC2) from the 3′ trailer of pre-tRNAs as part of tRNA maturation [34], and tRF-3 and TRF-5 fragments are generated from mature tRNAs by Dicer from the 3′ and 5′ end of the mature tRNA.
respectively [33,34]. RNAi-mediated knockdown of one of the tRFs, TRF-1001, resulted in reduced cell proliferation [34]. Although tRFs have been shown to be Dicer-dependent and some tRFs can associate with AGOs [33,35], the gene-silencing capability of tRFs still requires further validation. Although modest gene silencing by tRFs has been observed despite widespread AGO association, the mechanism of tRF function is still under debate. It has been argued that rather than acting as bona fide miRNAs, tRFs globally modulate gene silencing by competing for AGO binding, thereby regulating AGO’s availability to bind with miRNAs [35]. However, the latest research has shown that AGO2-associated 3' tRFs are complementary to retroviral sequences in the human genome and can target these transcripts via RNAi [28].

The second class of tRNA fragments is the tiRNAs. These tiRNAs are longer than tRFs (31–40 nt) and are generated via cleavage in the anticodon loop of mature tRNAs following various stress stimuli [36–38]. Both 5' and 3' end-derived tRNA halves can be detected. A variety of stress conditions, including nutrient depletion, hypoxia, hypothermia, heat shock and UV irradiation give rise to tiRNAs extremely rapidly (within 20 min) [37,38]. A low level of tiRNAs was also detected in unstressed cells [36]. Furthermore, this cleavage is not a mechanism to simply reduce the number of available tRNAs as only a small proportion of the full-length tRNA population are targeted [37,38].

The RNase A family member angiogenin was found to be the nuclease responsible for stress-induced tRNA cleavage [37]. Angiogenin is bound to the ribonuclease inhibitor RNH1 under normal conditions. When angiogenin was overexpressed in cells there was an increase in tiRNA production [37], and when RNH1 was targeted via siRNA, a similar result was observed [38]. It was demonstrated that the function of tiRNAs was to inhibit translation. Specifically, transfection of 5' tiRNAs (but not 3' tiRNAs) inhibited protein synthesis and RNH1 depletion promoted translational arrest [38].

Angiogenin-induced tiRNA fragments were shown to specifically target the initiation stage of translation. Protein synthesis is inhibited by tiRNAs’ association with the translational silencer YB-1, and subsequent displacement of eIF (eukaryotic initiation factor) 4G/eIF4A [39]. A subset of these tiRNAs (5'-tiRNA<sup>Δ5Δ6</sup> and 5'-tiRNA<sup>Δ35</sup>) was found to be the most potent translation inhibitors, and uncapped mRNAs were found to be preferentially targeted. This specific targeting of certain mRNAs suggests that tiRNAs reprogramme gene expression in stressed cells [38].

To complicate the picture of tRNA cleavage fragment biogenesis and function further, a recent study has shown that several RNases, including angiogenin, can also give rise to ~20 nt fragments derived from the 3' end of mature tRNAs [28]. These 3' end tRFs can associate with AGO2 and direct gene silencing. It is truly surprising that multiple sRNAs can be derived from tRNAs with such diverse modes of function.

**Y RNA-derived small RNAs**

Small RNAs derived from non-coding RNA molecules called Y RNAs have also come under the spotlight. Y RNAs were first discovered during investigations into SLE (systemic lupus erythematous) as they were found to associate with the autoantigen Ro60 [40]. Y RNAs have since been implicated in biological processes such as chromosomal DNA replication [41] and RNA quality control [42]. YsRNAs (Y RNA-derived small RNA) are produced at high levels in apoptotic cells as well as in a range of cancerous tissues [43,44].

Recently, it was shown that YsRNAs are also produced in non-apoptotic cells at physiological levels similar to those of miRNAs, and that they are also produced in non-cancerous cells [45]. It has been proposed that fragments derived from the hairpin-containing Y RNAs could potentially enter the miRNA pathway [44,46]. However, we found that YsRNAs are generated independently of Dicer and that they are not immunoprecipitated by AGO2 [45]. Anion-exchange chromatography also demonstrated that YsRNAs and miRNAs are in different complexes [45]. Our work, along with failure of YsRNA to display gene silencing using the luciferase reporter assay by others, confirmed that YsRNAs do not enter the miRNA pathway [44,45].

**Looking deeper into deep sequencing data**

Solexa (Illumina) sequencing is the most widely used deep sequencing technique for small RNA detection. It relies on small RNA adapter molecules ligating to the 3' and 5' ends of total RNA extracted from a tissue or cell sample. RT followed by PCR then results in generation of cDNA, which can be separated on a polyacrylamide gel. The desired size class of small RNA can then be selected for sequencing. Previous studies have demonstrated that a considerable bias exists whereby the adapter molecules can ligate to small RNAs [47,48]. The cloning bias in small RNA libraries is ligase-dependent and results in failure to detect thousands of non-adapter or reduced-adapter ligated small RNAs in library preparations. In order to reduce this ligation bias, new adapters were developed, which contain degenerate nucleotides at the ends of the adapters that are ligated to the small RNAs [49,50]. These HD (high-definition) adapters were shown to double the read coverage, and data obtained correlated better with absolute quantification of sRNAs via Northern blot, compared with standard Illumina adapters [49]. We envisage that the use of HD adapters will be a highly useful tool for future investigations of novel classes of small RNAs.

**Outlook**

Following the RNA ‘renaissance’, which was marked by the discovery of gene silencing-related molecules, sequencing technologies have unveiled a plethora of other small RNAs that were once considered degradation products. Studies have already managed to determine the function of some of these
small RNAs, such as tRNA halves. However others, such as YsRNAs, remain elusive in their function. The small RNA journey of discovery continues.

Acknowledgements

A.E.H. is funded by the Biotechnology and Biological Sciences Research Council.

References


Received 5 December 2012
doi:10.1042/BST20120347