Genome **Biology**

REVIEW

Seq and CLIP through the miRNA world

Nitish Mittal and Mihaela Zavolan[*]

Abstract

High-throughput sequencing of RNAs crosslinked to Argonaute proteins reveals not only a multitude of atypical miRNA binding sites but also of miRNA targets with atypical functions, and can be used to infer quantitative models of miRNA-target interaction strength.

Introduction

In the vast landscape of cellular RNAs of widely different sizes, microRNAs (miRNAs) are small (21 to 22 nucleotides long) RNAs that guide Argonaute proteins to target RNAs to post-transcriptionally regulate their expression [1,2]. lin-4 was the first miRNA to be reported and found to inhibit the translation of the *lin-14* mRNA at a critical stage in the development of the worm *Caenorhabditis elegans* [3,4]. It was the discovery of the evolutionarily conserved let-7 miRNA [5,6], however, that sparked a tremendous interest in RNAs with regulatory functions. Through many studies, a large catalog of miRNAs has since been compiled, from species as evolutionarily distant as viruses and mammals [7]. In the canonical biogenesis pathway, miRNAs are transcribed by the RNA polymerase II (Pol II) as long pri-miRNA. These are processed through two endonucleolytic steps involving RNase III enzymes [8], the first carried out by the Drosha-DiGeorge syndrome critical region 8 (DGCR8) complex in the nucleus to produce pre-miRNAs, and the second by the Dicer-TAR (HIV-1) RNA binding protein 2 (TRBP) complex in the cytoplasm to yield 21 to 22 nucleotide-long double-stranded RNAs. Typically one of the two strands of the duplex is picked up by an Argonaute protein to form a miRNA-guided RNA silencing complex (miRISC). The biogenesis of miRNAs has been reviewed extensively elsewhere [9]. Several alternative miRNA biogenesis pathways have also been described. Mirtrons, for example, bypass Drosha processing, being instead produced from spliced introns by the activity of the lariat debranching enzyme [10]. Another miRNA, pre-miR-451, is not processed

by Dicer but rather by the Argonaute 2 (Ago2) protein itself to yield the mature miRNA [11].

Many experimental and computational studies converged on the 5′ end (about nucleotides 1 to 8) of the miRNA (also known as the 'seed' region) being generally involved in target recognition through perfect nucleotide complementarity (see [1] for a recent review). Exceptions have also been reported: for example, the let-7 binding site in the *lin-41* 3′ UTR, in which the nucleotide located between those that base-pair with the fourth and fifth miRNA nucleotide is looped out of the miRNA-target hybrid [12,13]. Relatively rare sites that pair with the central region of the miRNA have also been found [14] and the interest in non-canonical miRNA target sites, which do not pair perfectly with the miRNA seed region, persists [15,16]. Putative sites that are computationally predicted to imperfectly pair with the miRNA seed region due to a bulged nucleotide in either the miRNA or the target site are known to exhibit some degree of evolutionary conservation relative to random 3′ UTR fragments of the same length [17,18]. However, the conservation signal as well as the apparent effect of such sites on the stability of the target mRNAs is smaller than those of canonical sites [19]. This likely indicates that only a subset of these sites is functional. Identifying this subset has so far been challenging.

Evolutionary studies of the Piwi-Argonaute-Zwille (PAZ) domain-containing proteins revealed largely two clusters, one corresponding to Argonaute and the other to the Piwi proteins [20]. Members of these clusters appear to have quite exquisite specificities for the length of the small RNAs that they bind [21]. Sequencing of the populations of small RNAs that associate with individual members of this protein family has been recently used to identify not only small guiding RNAs but also their targets. Here we review the insights into the processing of small RNAs and into their biological functions that were derived through high-throughput studies, particularly those that investigated individual protein components of small RNA-containing regulatory pathways.

* Correspondence: mihaela.zavolan@unibas.ch
Biozentrum, University of Basel, Klingelbergstrasse 50-70, Basel, Switzerland

High-throughput approaches for identifying small non-coding RNA genes and targets
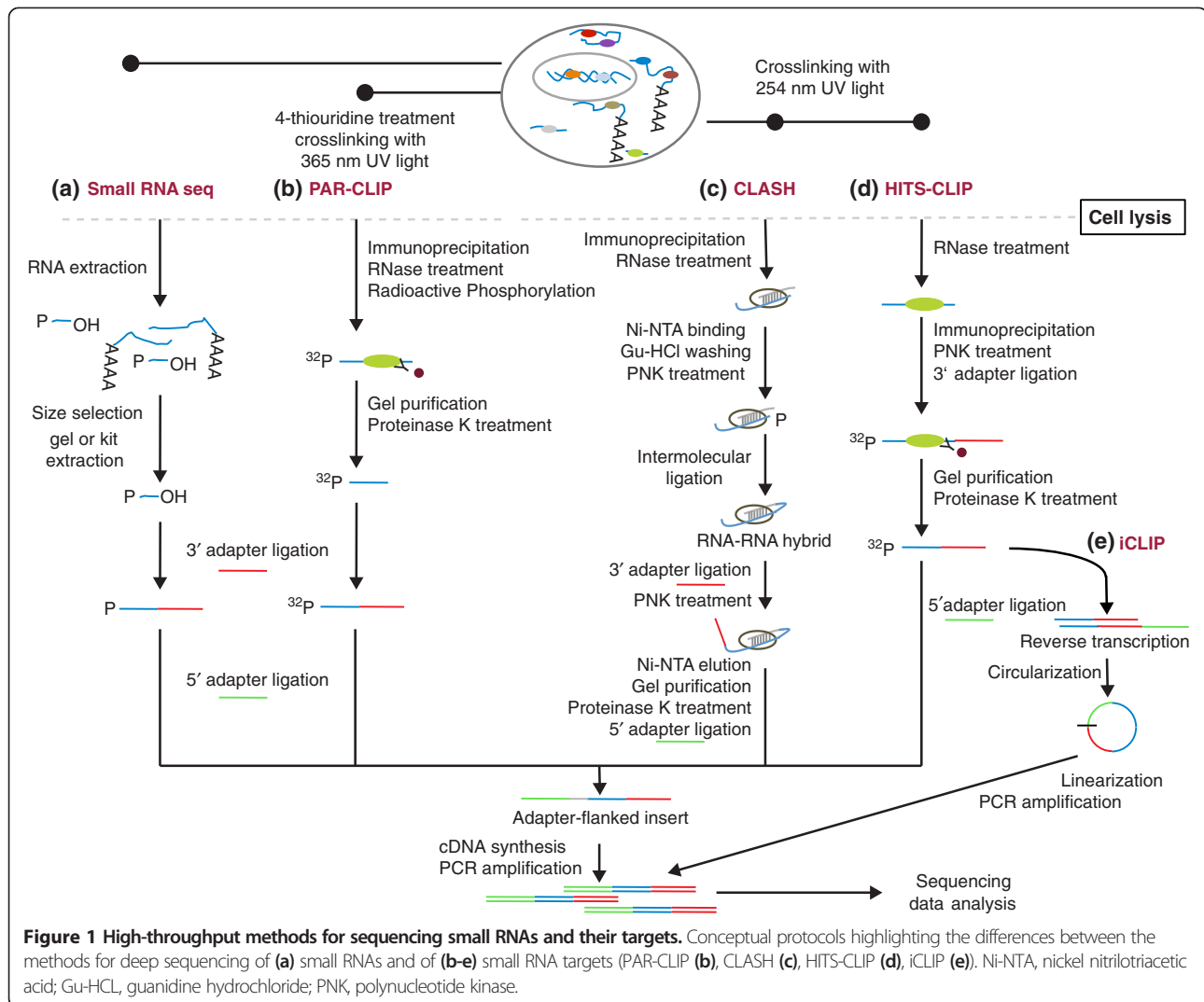
High-throughput sequencing has revolutionized molecular biology, including the study of RNA. Taking advantage of the biochemical properties of miRNAs (presence of a 5′-phosphate and 3′-hydroxyl), protocols have been developed to isolate and sequence these molecules with very little background [22-24]. The approach consisted of isolation of total RNA, followed by separation on urea-containing 15% polyacrylamide gel along with a ^{32}P-labelled ladder to allow identification of RNAs of the appropriate size. After cutting the corresponding band out of the gel and elution of the RNA overnight, 3′ and 5′ adaptors were ligated, the fragments concatamerized, and cDNA synthesized, PCR-amplified, cloned into plasmid vectors and sequenced with the Sanger method to yield 100 to 1,000 small RNAs per sample. Next generation sequencing (NGS) greatly increased the yield to 10^4 to 10^5 small RNA sequences per sample in the initial studies employing this technology [25-27]. NGS-based approaches have since been used to identify many other types of small RNAs. The basic protocol remains largely the same, except that cDNAs are sequenced without cloning and concatamerization [28].

To further remove the background of processing products of abundant cellular RNAs as well as to gain more direct insight into the functions of small RNAs, protocols that employ the pulldown of the protein of interest with a specific antibody have also been proposed (Figure 1). They have been used in the discovery of miRNAs and various other non-coding RNAs that associate with Argonaute proteins [29,30]. Building on this approach, the Darnell group [31,32] further applied a step of *in vivo* crosslinking using ultraviolet (UV) C light (254 nm) of the RNA-binding protein (RBP) to the RNAs with which it interacts in intact cells or tissues. After cell lysis, the RNA is partially digested to yield fragments in the range of 30 to 50 nucleotides, the RNA-protein complex is immunoprecipitated with an antibody specific to the protein of interest, the RNA in the complex is radioactively labeled at the 5′ end with ^{32}P, and an adapter is ligated at the 3′ end, after which the RNA-protein complex is separated on an SDS gel and transferred to a nitrocellulose membrane. This step results in the removal of unbound RNAs and retention of the covalently crosslinked RNA-protein complex. After the protein is digested from the complex with proteinase K, a 5′ adapter is ligated, cDNA is synthesized and PCR amplification is carried out with primers complementary to 3′ and 5′ adapters. The PCR adapters also carry sequences needed for attachment to the flowcell surface and for attachment of the sequencing primers, when sequencing on Illumina platforms. The resulting library is subjected to NGS. To further improve the efficiency of capture of miRNA targets, the Tuschl group proposed a modified protocol, photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP), in which photoactivatable ribonucleoside analogues such as 4-thiouridine (4-SU) or 6-thioguanosine (6-SG) are incorporated into RNAs before crosslinking [33]. These modified nucleotides can be efficiently crosslinked to proteins using UV A (365 nm). In addition, crosslinking-diagnostic mutations (T-to-C or G-to-A, respectively) are introduced during reverse transcription to allow determination of the binding sites at close-to-nucleotide resolution. This protocol has been successfully used to identify not only miRNA targets [33,34] but also the RNA targets of many RNA-binding proteins [35]. To achieve the desired single-nucleotide resolution in the identification of RBP targets, a method that takes advantage of the propensity of reverse transcriptase to stop at the position of crosslinking has been proposed [36]. This individual nucleotide resolution CLIP method (iCLIP) has only very recently been applied to the characterization of small RNA-guided interactions [37].

Although high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CLIP), PAR-CLIP and iCLIP have a similar basis, their differences make them more or less applicable in specific contexts. For instance, an important advantage of HITS-CLIP is that it can be performed with relative ease in both cultured cells and living tissues. However, the efficiency of crosslinking of Argonaute to the mRNA targets (as opposed to the guide RNAs) appears lower than with PAR-CLIP. Although PAR-CLIP is more difficult to carry out in tissues, its successful application to the identification of *in vivo germline development defective 1* (GLD-1) protein binding sites in the worm *C. elegans* has been reported [38]. Important concerns about the use of photoreactive nucleosides are that they are toxic for cells [39] and they bias the set of binding sites that can be identified. However, the concentration of 4-thiouridine that has been used in PAR-CLIP experiments has not been found to obviously affect the cells [33]. On the other hand, the bias in binding site identification remains largely unquantified. Yet this is not only an issue for PAR-CLIP because crosslinking with 254 nm UV, as in HITS-CLIP, also targets uridines preferentially [40].

Generally, it has become clear that crosslinking-induced mutations are useful in separating the signal from noise and identification of high-affinity binding sites [34,40,41], but how different CLIP methods compare in this regard needs to be further investigated. Several factors make this comparison difficult. First, the protocols are lengthy and difficult to master, which makes it difficult to obtain equally good data with all the different CLIP protocols. Second, the possible interplay between the biases of individual approaches and the sequence specificity of individual proteins makes it necessary to perform the comparison on multiple proteins. Third, it is non-trivial to obtain independent quantifications of occupancies of individual binding sites by a given protein, which is necessary for evaluating the results

Figure 1 High-throughput methods for sequencing small RNAs and their targets. Conceptual protocols highlighting the differences between the methods for deep sequencing of **(a)** small RNAs and of **(b-e)** small RNA targets (PAR-CLIP **(b)**, CLASH **(c)**, HITS-CLIP **(d)**, iCLIP **(e)**). Ni-NTA, nickel nitrilotriacetic acid; Gu-HCL, guanidine hydrochloride; PNK, polynucleotide kinase.

of different CLIP protocols. One possibility is to use an *in vitro*-derived model of the sequence specificity of the protein to predict its affinity for individual CLIPed sites [34]. The success of this approach depends on how accurately the affinity of RBP-RNA interactions can be predicted. Another approach would be to take advantage of proteins that establish crosslinks to RNA in a UV-independent manner. For example, the NOP2/Sun domain family, member 2 protein (NSUN2) normally catalyzes methylation of cytosine to 5-methylcytosine, generating a protein-RNA crosslink as an intermediate in the process. Employing a variant that can no longer resolve the covalent bond that the protein forms with the RNA, the binding sites of this protein could be determined without UV crosslinking and compared with the binding sites obtained by crosslinking the protein to its sites with UV light. Finally, in the absence of independent measures of site occupancy, comparisons of sequence biases around putative binding sites inferred for different proteins have been performed [40]. They indicate

that UVC light preferentially induces crosslinking of uridines. Furthermore, it appears that reverse transcriptase stoppage sites that are captured through iCLIP are a more accurate indicator of protein binding sites than nucleotide deletions that are introduced during HITS-CLIP.

Although the above-mentioned methods are able to identify the endogenous targets of miRNAs or other small noncoding RNAs, they do not directly reveal which small RNA guided the interaction of the RBP with individual targets. To address this issue, another experimental approach has been very recently proposed. It is known as crosslinking, ligation and sequencing of hybrids (CLASH) and it relies on the ligation of the guide RNA to the target RNA within the ternary guide RNA-target RNA-RBP complex, after the immunoprecipitation of the protein with the bound RNAs [42]. In contrast to CLIP, this protocol includes, after immunoprecipitation and partial digestion of the RNA in the RNA-protein complex, a purification step based on a 6x-histidine epitope tag that allows denaturing purification

of the RNA-protein complex on nickel beads at 6 M guanidine-HCl. This ensures that only the RNA that is covalently linked to protein is purified. In addition, an inter-molecular RNA-RNA ligation step is introduced to capture the target site and the miRNA from the RNA-protein ternary complex. After elution of the RNA-protein complex from nickel beads, the sample preparation proceeds similarly to CLIP. This method has been successfully used to identify various types of RNA-RNA hybrids [43], and its recent application to the Ago1 protein led to the suggestion that different miRNAs may have different modes of binding to their target mRNAs [42]. In its current form, CLASH has very low efficiency, with only about 2% of the reads obtained in an experiment corresponding to miRNA-target hybrids. Furthermore, the use of a 6x-histidine tag for the purification of RNA-protein complexes makes the protocol applicable only to cells that express the tagged protein.

The expanding set of miRNA targets

Following the model of worm miRNAs, initial large-scale studies of miRNA targets focused on mRNAs, first attempting to predict them computationally [44-46] and then to determine them experimentally, by virtue of the change in their expression upon miRNA transfection measured with microarrays [47]. More recently, crosslinking-based approaches are starting to bring a new understanding of miRNA-target interactions and to uncover unusual targets (Figure 2).

Identification of non-canonical miRNA target sites from CLIP data

miRNA target sites that do not perfectly pair with the miRNA seed region (so-called non-canonical sites) have been both described experimentally [5,12,15,48] and predicted based on evolutionary conservation [49]. However, recent analyses of Ago2-CLIP data underscored the relative abundance of a specific type of site, in which the nucleotide
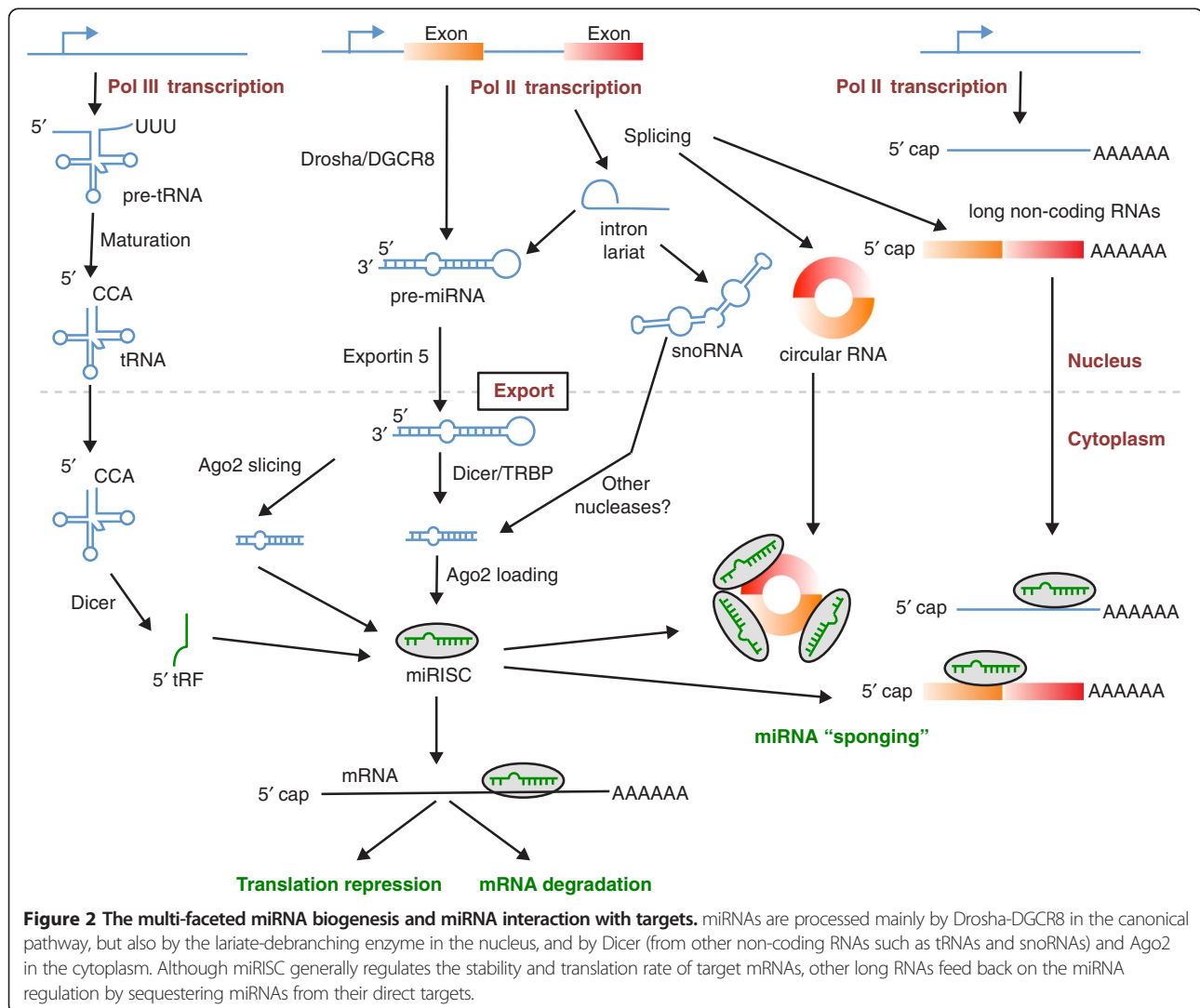

Figure 2 The multi-faceted miRNA biogenesis and miRNA interaction with targets. miRNAs are processed mainly by Drosha-DGCR8 in the canonical pathway, but also by the lariate-debranching enzyme in the nucleus, and by Dicer (from other non-coding RNAs such as tRNAs and snoRNAs) and Ago2 in the cytoplasm. Although miRISC generally regulates the stability and translation rate of target mRNAs, other long RNAs feed back on the miRNA regulation by sequestering miRNAs from their direct targets.

located between those that pair with positions 5 and 6 of the miRNA is looped out in the target [16,50]. More importantly, CLIP provided sufficient data to infer a bio-physical model of miRNA-target site interaction [19] that allows, for the first time, a quantitative evaluation of the strength of canonical and non-canonical interactions. As a result, functional non-canonical target sites could be identified with high accuracy. They amounted to approximately a quarter of the high-confidence, reproducibly CLIPed sites. Perhaps as expected, abundant miRNAs were found to have a higher proportion of non-canonical sites compared with the less expressed miRNAs. A recent study that captured and sequenced miRNA-target site pairs [42] suggested that miRNAs differ widely in their propensity to engage in non-canonical modes of interaction with their targets. miR-92a, for example, a member of the abundantly expressed miR-17/92 cluster of miRNAs, appeared to predominantly pair with targets through its 3′ end region. The response of these targets to the miR-92a depletion was, however, smaller than that of seed-type miR-92a targets, and thus the significance of these non-canonical interactions remains to be determined. Nonetheless, as more CLASH datasets emerge, it will be interesting to apply the MIRZA inference procedure described in Khorshid *et al.* [19] to CLASH data to infer miRNA-specific modes of interaction with the targets. The MIRZA approach can be further adapted to infer miRNA-target interaction parameters from measurements of interaction affinity [51]. A comparative analysis of models inferred from *in vivo* and *in vitro* data should ultimately reveal the properties of functionally relevant miRNA target sites.

Long non-coding RNA targets and miRNA sponges

Although the vast majority of Ago2 targets are mRNAs, a variety of non-coding RNA targets have also been identified. For example, about 5% of the Ago2 targets obtained in HITS-CLIP samples from mouse brain were long non-coding RNAs (lncRNAs) [32], and many lncRNA-miRNA interactions were also inferred from PAR-CLIP data of different Argonaute proteins [52]. lncRNA-Argonaute interactions (for example, between *XIST* lncRNA and hsa-miR-370-3p) are documented in the starBase database [53]. Rapidly emerging evidence points to a function of lncRNA-miRNA interactions in regulating the availability of the miRNA itself, with the lncRNA functioning as a miRNA sponge.

miRNA sponges were introduced a few years ago [54] as competitive miRNA inhibitors consisting of transgenic RNAs that contain multiple putative binding sites for a given miRNA or miRNA family. Perhaps not surprisingly, natural miRNA sponges have emerged as well, initially among viral transcripts. For example, a U-rich RNA of the *Herpesvirus saimiri* acts as a sponge for the host miR-27 [55], as does the m169 transcript of the murine cytomegalic virus [56]. In

mammals, pseudogenes such as *PTENP1* and *KRASP1* [57] have been proposed to sponge miRNAs that would otherwise act on the corresponding genes. It remains unclear, however, whether under normal or disease conditions these pseudogenes are expressed at sufficient levels to be effective as sponges [58]. Other lncRNAs do appear to accumulate at very high levels, consistent with a sponging function. For example, a very recent study showed that the lncRNA *H19* associates with the RISC complex, sequestering the let-7 miRNA and thereby modulating the expression of let-7 targets [59]. A similar interaction has been proposed to occur between lincRNA-RoR and miR-145 [60].

Circular RNA

miRNA sponges have also been found among circular RNAs (circRNAs). Although a few circRNAs, such as those derived from the *DCC* tumor suppressor gene [61], the testis-determining *SRY* gene [62], *ETS-1* [63] and the cytochrome P450 gene *2C24* [64], were described two decades ago, it was thought that such RNAs are rare, aberrant products of the splicing reaction [61,63]. Deep sequencing of RNAs from a variety of normal and malignant cells revealed, however, an abundance of such transcripts [65,66] that can be expressed at 10-fold higher levels than the mRNAs derived from the corresponding genes [67]. The biogenesis of circRNA is not yet clear. Models such as lariat-driven or intron-pairing-driven circularization have been proposed [67]. Furthermore, failure in debranching can also yield intron-derived circRNAs [68]. Interestingly, Ago2-PAR-CLIP revealed that a circRNA that is antisense to the cerebellar degeneration-related protein 1 transcript (*CDR1as*) is densely bound by Argonaute proteins, guided by a large number of conserved miR-7 binding sites [69]. The circRNA is completely resistant to miRNA-mediated target destabilization and it strongly suppresses miR-7 activity in the mouse and zebrafish brain [69,70]. Other functions of circRNAs, such as in Pol II-dependent transcription, have also been reported [68].

The adoption of high-throughput approaches is not without complications. Every method has limited accuracy and even in deep sequencing samples one expects a certain amount of contaminating RNAs, particularly originating from abundant cellular RNAs. Although *a priori* knowledge of abundant RNA species generally helps in sifting away this background, novel variants of well-studied molecules, such as tRNA-derived fragments (tRFs) and small nucleolar RNAs (snoRNAs), have also been identified recently, complicating the analysis of deep-sequencing datasets. We will describe here some non-canonically processed RNAs with biological significance, whose number appears to be more limited than initial analyses suggested [71-74].

Remodeling of the miRNA targetome upon stress

Application of Ago2-CLIP revealed a stress-dependent remodeling of miRNA-target interactions, canonical interactions becoming more prominent upon arsenite stress [75]. Increased Ago2 binding to these canonical sites was also associated with increased repression. The mechanism behind the redistribution of Ago2 binding to higher affinity, canonical sites under stress remains to be identified. The abundance of both miRNAs and Ago2 protein appears to remain unchanged between conditions and it was rather proposed that signal-induced post-translational modifications of Ago2 may alter the interaction strength at specific sites. It is conceivable that a reduction in RISC affinity for target sites leads to reduced binding to weak, non-canonical sites. However, changes in the overall abundance of miRNA target sites may also lead to changes in the stringency of competition for a limited number of RISC complexes, and to a redistribution of Ago2 between low- and high-affinity sites.

More roads leading to RISC

IsomiRs

Although mature miRNAs are typically processed very precisely from their precursor molecules, evidence is accumulating that some miRNA variants - isomiRs - that differ in a few nucleotides from the canonical, most frequently observed sequence are generated and have biological significance. Some isomiRs are templated, being the result of imprecise cropping of miRNA precursors by Drosha or Dicer [76] or of the trimming of the miRNA 3′ end by 3′-to-5′ exoribonucleases such as Nibbler in *Drosophila* [77] and QIP in *Neurospora* [78]. The Dicer partner TRBP can also modulate isomiR generation [79,80]. When the miRNA is encoded in the 3′ arm of the pre-miRNA, the Dicer-modulated change in isomiR abundance will likely lead to a change in the spectrum of mRNAs that are targeted by the miRNA. For example, the 5′ isomirs of mir-307a do seem to have distinct targets because the glycerol kinase and taranis mRNAs are repressed by mir-307a$^{23\text{-mer}}$ but not by mir-307a$^{21\text{-mer}}$ [80]. Furthermore, isomiRs and their canonical counterparts appear to associate equally with polysomal, translated RNA [81], indicating that they may indeed function as miRNAs. A variety of terminal nucleotidyl transferases, such as mitochondrial poly(A) polymerase (MTPAP), PAP associated domain containing (PAPD)4, PAPD5, zinc finger, CCHC domain containing (ZCCHC)6, ZCCHC11 and terminal uridylyl transferase 1, U6 snRNA-specific (TUT1) [82], have been implicated in the generation of non-templated 3′ isomiRs. TUT1-dependent addition of terminal U nucleotides has been implicated in the regulation of miRNA stability [83].

snoRNA-derived small RNAs and tRFs

Sequencing of small RNA populations, including those that specifically associate with RISC proteins, revealed fragments derived from abundantly expressed structural RNAs, such as snoRNAs and tRNAs, that also seem to associate with Argonaute proteins [29,84]. Among the snoRNAs, the H/ACA box-type, which forms a typical two-hairpin structure, gives rise to miRNA-like molecules that amount to a few percent of the Argonaute-associated small RNA population [84]. The H/ACA box snoRNA small Cajal body-specific RNA 15 (SCARNA15) generates the most abundant Ago2-associated snoRNA-derived small RNA, which targets the transcript encoding the Mediator coactivator complex subunit cyclin-dependent kinase 19 (CDK19) [29]. Although less abundant among the approximately 20 to 40 nucleotide-long RNAs in the cell, tRFs appear to associate more efficiently with the Ago2 protein compared with snoRNA-derived fragments [84]. Various nucleases have been implicated in the generation of tRFs, starting with Dicer, which processes the CU1276 tRF - which functions as a miRNA in B cells, repressing the replication protein A1 [85] - and the tRF-5-GlnCTG [86]. Angiogenin acts at the TΨC loop to generate 3′-end tRFs, and on the anticodon loop to produce 5′-end tRFs [87]. The latter have been implicated in the eukaryotic translation initiation factor 2 alpha (eIF2α)-independent inhibition of translation in U2OS cells upon stress [88]. Finally, the elaC ribonuclease Z 2 (ELAC2) endonuclease cleaves the 3′ trailer sequence from Ser-TGA pre-tRNAs, generating the pro-proliferative trf-1001 tRF [89].

Cleaving without a guide

Although we have extensively discussed small RNA-guided mRNA destabilization, the Drosha-DGCR8 complex that processes pri-miRNAs also cleaves hairpin structures that form within other molecules, including mRNAs, thereby inducing their destabilization. The abundance of the metastasis associated lung adenocarcinoma transcript 1 (non-protein coding) (*MALAT1*) non-coding RNA appears to be controlled through this mechanism [90], as is the expression of several genes that induce neuronal differentiation, such as neurogenin 2 [91].

Conclusions

The list of long and short functional RNAs is expanding rapidly. Here we have summarized some of the insights into the targets of the miRNA-dependent pathway that were obtained particularly though NGS-based approaches such as small RNA sequencing and different variants of RBP-CLIP methods. An increasing number of entry points into miRNA-dependent gene regulation are being discovered. Furthermore, miRNA-target interactions are plastic, and cell type- and condition-dependent. Nonetheless, quantitative analyses in the context of computational models should ultimately allow the behavior of this very complex gene regulatory system to be understood and predicted.

Abbreviations
Ago: Argonaute; circMRNA: Circular miRNA; CLASH: Crosslinking ligation and sequencing of hybrids; DGCR8: DiGeorge critical region 8; HITS-CLIP: High-throughput sequencing of RNA isolated by crosslinking immunoprecipitation; iCLIP: Individual nucleotide resolution CLIP method; lncRNA: Long noncoding RNA; miRISC: miRNA-guided RNA silencing complex; miRNA: microRNA; NGS: Next generation sequencing; PAR-CLIP: Photoactivatable ribonucleotide-enhanced crosslinking and immunoprecipitation; PCR: Polymerase chain reaction; Pol II: RNA polymerase II; RBP: RNA-binding protein; RISC: RNA silencing complex; snoRNA: Small nucleolar RNA; TRBP: TAR (HIV-1) RNA binding protein 2; tRF: tRNA-derived RNA fragments; tRNA: Transfer RNA; UTR: Untranslated region.

References
1. Bartel DP: MicroRNAs: target recognition and regulatory functions. *Cell* 2009, **136**:215–233.
2. Huntzinger E, Izaurralde E: Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat Rev Genet* 2011, **12**:99–110.
3. 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.
4. Wightman B, Ha I, Ruvkun G: Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell* 1993, **75**:855–862.
5. Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G: The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 2000, **403**:901–906.
6. Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, Hayward DC, Ball EE, Degnan B, Müller P, Spring J, Srinivasan A, Fishman M, Finnerty J, Corbo J, Levine M, Leahy P, Davidson E, Ruvkun G: Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* 2000, **408**:86–89.
7. Kozomara A, Griffiths-Jones S: miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* 2011, **39**(Database issue):D152–D157.
8. Nicholson AW: Ribonuclease III mechanisms of double-stranded RNA cleavage. *Wiley Interdiscip Rev RNA* 2014, **5**:31–48.
9. Kim VN: MicroRNA biogenesis: coordinated cropping and dicing. *Nat Rev Mol Cell Biol* 2005, **6**:376–385.
10. Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC: The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. *Cell* 2007, **130**:89–100.
11. Cheloufi S, Dos Santos CO, Chong MMW, Hannon GJ: A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature* 2010, **465**:584–589.
12. Ha I, Wightman B, Ruvkun G: A bulged lin-4/lin-14 RNA duplex is sufficient for *Caenorhabditis elegans* lin-14 temporal gradient formation. *Genes Dev* 1996, **10**:3041–3050.
13. Vella MC, Reinert K, Slack FJ: Architecture of a validated microRNA: target interaction. *Chem Biol* 2004, **11**:1619–1623.
14. Shin C, Nam J-W, Farh KK-H, Chiang HR, Shkumatava A, Bartel DP: Expanding the microRNA targeting code: functional sites with centered pairing. *Mol Cell* 2010, **38**:789–802.
15. Lal A, Navarro F, Maher CA, Maliszewski LE, Yan N, O'Day E, Chowdhury D, Dykxhoorn DM, Tsai P, Hofmann O, Becker KG, Gorospe M, Hide W, Lieberman J: miR-24 inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to "Seedless" 3′UTR microRNA recognition elements. *Mol Cell* 2009, **35**:610–625.
16. Chi SW, Hannon GJ, Darnell RB: An alternative mode of microRNA target recognition. *Nat Struct Mol Biol* 2012, **19**:321–327.
17. Lewis BP, Burge CB, Bartel DP: Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005, **120**:15–20.
18. Gaidatzis D, van Nimwegen E, Hausser J, Zavolan M: Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics* 2007, **8**:69.
19. Khorshid M, Hausser J, Zavolan M, van Nimwegen E: A biophysical miRNA-mRNA interaction model infers canonical and noncanonical targets. *Nat Methods* 2013, **10**:253–255.
20. Joshua-Tor L, Hannon GJ: Ancestral roles of small RNAs: an Ago-centric perspective. *Cold Spring Harb Perspect Biol* 2011, **3**:a003772.
21. Dueck A, Ziegler C, Eichner A, Berezikov E, Meister G: MicroRNAs associated with the different human Argonaute proteins. *Nucleic Acids Res* 2012, **40**:9850–9862.
22. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T: Identification of novel genes coding for small expressed RNAs. *Science* 2001, **294**:853–858.
23. Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T: Identification of tissue-specific microRNAs from mouse. *Curr Biol* 2002, **12**:735–739.
24. Lagos-Quintana M, Rauhut R, Meyer J, Borkhardt A, Tuschl T: New microRNAs from mouse and human. *RNA* 2003, **9**:175–179.
25. Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, Morris P, Brownstein MJ, Kuramochi-Miyagawa S, Nakano T, Chien M, Russo JJ, Ju J, Sheridan R, Sander C, Zavolan M, Tuschl T: A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* 2006, **442**:203–207.
26. Girard A, Sachidanandam R, Hannon GJ, Carmell MA: A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 2006, **442**:199–202.
27. Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, Kingston RE: Characterization of the piRNA complex from rat testes. *Science* 2006, **313**:363–367.
28. Hafner M, Landgraf P, Ludwig J, Rice A, Ojo T, Lin C, Holoch D, Lim C, Tuschl T: Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods* 2008, **44**:3–12.
29. Ender C, Krek A, Friedländer MR, Beitzinger M, Weinmann L, Chen W, Pfeffer S, Rajewsky N, Meister G: A human snoRNA with microRNA-like functions. *Mol Cell* 2008, **32**:519–528.
30. Burroughs AM, Ando Y, de Hoon MJL, Tomaru Y, Suzuki H, Hayashizaki Y, Daub CO: Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin. *RNA Biol* 2011, **8**:158–177.
31. Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB: CLIP identifies Nova-regulated RNA networks in the brain. *Science* 2003, **302**:1212–1215.
32. Chi SW, Zang JB, Mele A, Darnell RB: Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* 2009, **460**:479–486.
33. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp A-C, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T: Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 2010, **141**:129–141.
34. Kishore S, Jaskiewicz L, Burger L, Hausser J, Khorshid M, Zavolan M: A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat Methods* 2011, **8**:559–564.
35. Ascano M, Hafner M, Cekan P, Gerstberger S, Tuschl T: Identification of RNA-protein interaction networks using PAR-CLIP. *Wiley Interdiscip Rev RNA* 2011, **3**:159–177.
36. Konig J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J: iCLIP - transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. *J Vis Exp* 2011, :2638.
37. Broughton JP, Pasquinelli AE: Identifying Argonaute binding sites in *Caenorhabditis elegans* using iCLIP. *Methods* 2013, **63**:119–125.
38. Jungkamp A-C, Stoeckius M, Mecenas D, Grün D, Mastrobuoni G, Kempa S, Rajewsky N: In vivo and transcriptome-wide identification of RNA binding protein target sites. *Mol Cell* 2011, **44**:828–840.
39. Burger K, Mühl B, Kellner M, Rohrmoser M, Gruber-Eber A, Windhager L, Friedel CC, Dölken L, Eick D: 4-Thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. *RNA Biol* 2013, **10**:1623–1630.
40. Sugimoto Y, König J, Hussain S, Zupan B, Curk T, Frye M, Ule J: Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biol* 2012, **13**:R67.
41. Zhang C, Darnell RB: Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat Biotechnol* 2011, **29**:607–614.
42. Helwak A, Kudla G, Dudnakova T, Tollervey D: Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* 2013, **153**:654–665.
43. Kudla G, Granneman S, Hahn D, Beggs JD, Tollervey D: Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. *Proc Natl Acad Sci U S A* 2011, **108**:10010–10015.

44. Lewis BP, Shih I-H, Jones-Rhoades MW, Bartel DP, Burge CB: **Prediction of mammalian microRNA targets.** *Cell* 2003, **115**:787–798.

45. Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS: **MicroRNA targets in Drosophila.** *Genome Biol* 2003, **5**:R1.

46. Rajewsky N, Socci ND: **Computational identification of microRNA targets.** *Dev Biol* 2004, **267**:529–535.

47. Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM: **Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs.** *Nature* 2005, **433**:769–773.

48. Brennecke J, Stark A, Russell RB, Cohen SM: **Principles of microRNA - target recognition.** *PLoS Biol* 2005, **3**:e85.

49. Friedman RC, Farh KK-H, Burge CB, Bartel DP: **Most mammalian mRNAs are conserved targets of microRNAs.** *Genome Res* 2009, **19**:92–105.

50. Loeb GB, Khan AA, Canner D, Hiatt JB, Shendure J, Darnell RB, Leslie CS, Rudensky AY: **Transcriptome-wide miR-155 binding map reveals widespread noncanonical microRNA targeting.** *Mol Cell* 2012, **48**:760–770.

51. Wee LM, Flores-Jasso CF, Salomon WE, Zamore PD: **Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties.** *Cell* 2012, **151**:1055–1067.

52. Jalali S, Bhartiya D, Lalwani MK, Sivasubbu S, Scaria V: **Systematic transcriptome wide analysis of lncRNA-miRNA interactions.** *PLoS One* 2013, **8**:e53823.

53. Li J-H, Liu S, Zhou H, Qu L-H, Yang J-H: **starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data.** *Nucleic Acids Res* 2013, : . 10.1093/nar/gkt1248.

54. Ebert MS, Neilson JR, Sharp PA: **MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells.** *Nat Methods* 2007, **4**:721–726.

55. Cazalla D, Yario T, Steitz JA, Steitz J: **Down-regulation of a host microRNA by a Herpesvirus saimiri noncoding RNA.** *Science* 2010, **328**:1563–1566.

56. Marcinowski L, Tanguy M, Krmpotic A, Rädle B, Lisnić VJ, Tuddenham L, Chane-Woon-Ming B, Ruzsics Z, Erhard F, Benkartek C, Babic M, Zimmer R, Trgovcich J, Koszinowski UH, Jonjic S, Pfeffer S, Dölken L: **Degradation of cellular mir-27 by a novel, highly abundant viral transcript is important for efficient virus replication in vivo.** *PLoS Pathog* 2012, **8**:e1002510.

57. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP: **A coding-independent function of gene and pseudogene mRNAs regulates tumour biology.** *Nature* 2010, **465**:1033–1038.

58. Ebert MS, Sharp PA: **Emerging roles for natural microRNA sponges.** *Curr Biol* 2010, **20**:R858–R861.

59. Kallen AN, Zhou X-B, Xu J, Qiao C, Ma J, Yan L, Lu L, Liu C, Yi J-S, Zhang H, Min W, Bennett AM, Gregory RI, Ding Y, Huang Y: **The imprinted H19 lncRNA antagonizes let-7 microRNAs.** *Mol Cell* 2013, **52**:101–112.

60. Wang Y, Xu Z, Jiang J, Xu C, Kang J, Xiao L, Wu M, Xiong J, Guo X, Liu H: **Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal.** *Dev Cell* 2013, **25**:69–80.

61. Nigro JM, Cho KR, Fearon ER, Kern SE, Ruppert JM, Oliner JD, Kinzler KW, Vogelstein B: **Scrambled exons.** *Cell* 1991, **64**:607–613.

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

63. Cocquerelle C, Mascrez B, Hétuin D, Bailleul B: **Mis-splicing yields circular RNA molecules.** *FASEB J* 1993, **7**:155–160.

64. Zaphiropoulos PG: **Circular RNAs from transcripts of the rat cytochrome P450 2C24 gene: correlation with exon skipping.** *Proc Natl Acad Sci U S A* 1996, **93**:6536–6541.

65. Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO: **Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types.** *PLoS One* 2012, **7**:e30733.

66. Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO: **Cell-type specific features of circular RNA expression.** *PLoS Genet* 2013, **9**:e1003777.

67. Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, Marzluff WF, Sharpless NE: **Circular RNAs are abundant, conserved, and associated with ALU repeats.** *RNA* 2013, **19**:141–157.

68. Zhang Y, Zhang X-O, Chen T, Xiang J-F, Yin Q-F, Xing Y-H, Zhu S, Yang L, Chen L-L: **Circular intronic long noncoding RNAs.** *Mol Cell* 2013, **51**:792–806.

69. 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: **Circular RNAs are a large class of animal RNAs with regulatory potency.** *Nature* 2013, **495**:333–338.

70. Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J: **Natural RNA circles function as efficient microRNA sponges.** *Nature* 2013, **495**:384–388.

71. Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu A-L, Zhao Y, McDonald H, Zeng T, Hirst M, Eaves CJ, Marra MA: **Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells.** *Genome Res* 2008, **18**:610–621.

72. Taft RJ, Glazov EA, Lassmann T, Hayashizaki Y, Carninci P, Mattick JS: **Small RNAs derived from snoRNAs.** *RNA* 2009, **15**:1233–1240.

73. Kuchenbauer F, Morin RD, Argiropoulos B, Petriv OI, Griffith M, Heuser M, Yung E, Piper J, Delaney A, Prabhu A-L, Zhao Y, McDonald H, Zeng T, Hirst M, Hansen CL, Marra MA, Humphries RK: **In-depth characterization of the microRNA transcriptome in a leukemia progression model.** *Genome Res* 2008, **18**:1787–1797.

74. Taft RJ, Simons C, Nahkuri S, Oey H, Korbie DJ, Mercer TR, Holst J, Ritchie W, Wong JJ-L, Rasko JEJ, Rokhsar DS, Degnan BM, Mattick JS: **Nuclear-localized tiny RNAs are associated with transcription initiation and splice sites in metazoans.** *Nat Struct Mol Biol* 2010, **17**:1030–1034.

75. Karginov F, Hannon G: **Remodeling of Ago2-mRNA interactions upon cellular stress reflects miRNA complementarity and correlates with altered translation rates.** *Genes Dev* 2013, **27**:1624–1632.

76. Starega-Roslan J, Krol J, Koscianska E, Kozlowski P, Szlachcic WJ, Sobczak K, Krzyzosiak WJ: **Structural basis of microRNA length variety.** *Nucleic Acids Res* 2011, **39**:257–268.

77. Liu N, Abe M, Sabin LR, Hendriks G-J, Naqvi AS, Yu Z, Cherry S, Bonini NM: **The exoribonuclease Nibbler controls 3' end processing of microRNAs in Drosophila.** *Curr Biol* 2011, **21**:1888–1893.

78. Xue Z, Yuan H, Guo J, Liu Y: **Reconstitution of an Argonaute-dependent small RNA biogenesis pathway reveals a handover mechanism involving the RNA exosome and the exonuclease QIP.** *Mol Cell* 2012, **46**:299–310.

79. Lee HY, Doudna JA: **TRBP alters human precursor microRNA processing in vitro.** *RNA* 2012, **18**:2012–2019.

80. Fukunaga R, Han BW, Hung J-H, Xu J, Weng Z, Zamore PD: **Dicer partner proteins tune the length of mature miRNAs in flies and mammals.** *Cell* 2012, **151**:533–546.

81. Cloonan N, Wani S, Xu Q, Gu J, Lea K, Heater S, Barbacioru C, Steptoe AL, Martin HC, Nourbakhsh E, Krishnan K, Gardiner B, Wang X, Nones K, Steen JA, Matigian NA, Wood DL, Kassahn KS, Waddell N, Shepherd J, Lee C, Ichikawa J, McKernan K, Bramlett K, Kuersten S, Grimmond SM: **MicroRNAs and their isomiRs function cooperatively to target common biological pathways.** *Genome Biol* 2011, **12**:R126.

82. Wyman SK, Knouf EC, Parkin RK, Fritz BR, Lin DW, Dennis LM, Krouse MA, Webster PJ, Tewari M: **Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity.** *Genome Res* 2011, **21**:1450–1461.

83. Knouf EC, Wyman SK, Tewari M: **The human TUT1 nucleotidyl transferase as a global regulator of microRNA abundance.** *PLoS One* 2013, **8**:e69630.

84. Kishore S, Gruber AR, Jedlinski DJ, Syed AP, Jorjani H, Zavolan M: **Insights into snoRNA biogenesis and processing from PAR-CLIP of snoRNA core proteins and small RNA sequencing.** *Genome Biol* 2013, **14**:R45.

85. Maute RL, Schneider C, Sumazin P, Holmes A, Califano A, Basso K, Dalla-Favera R: **tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma.** *Proc Natl Acad Sci U S A* 2013, **110**:1404–1409.

86. Cole C, Sobala A, Lu C, Thatcher SR, Bowman A, Brown JWS, Green PJ, Barton GJ, Hutvagner G: **Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs.** *RNA* 2009, **15**:2147–2160.

87. Li Z, Ender C, Meister G, Moore PS, Chang Y, John B: **Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs.** *Nucleic Acids Res* 2012, **40**:6787–6799.

88. Yamasaki S, Ivanov P, Hu G-F, Anderson P: **Angiogenin cleaves tRNA and promotes stress-induced translational repression.** *J Cell Biol* 2009, **185**:35–42.

89. Lee YS, Shibata Y, Malhotra A, Dutta A: **A novel class of small RNAs: tRNA-derived RNA fragments (tRFs).** *Genes Dev* 2009, **23**:2639–2649.

90. Macias S, Plass M, Stajuda A, Michlewski G, Eyras E, Cáceres JF: **DGCR8 HITS-CLIP reveals novel functions for the microprocessor.** *Nat Struct Mol Biol* 2012, **19**:760–766.

91. Knuckles P, Vogt MA, Lugert S, Milo M, Chong MMW, Hautbergue GM, Wilson SA, Littman DR, Taylor V: **Drosha regulates neurogenesis by controlling neurogenin 2 expression independent of microRNAs.** *Nat Neurosci* 2012, **15**:962–969.