Functional genomic, computational and proteomic analysis of C. elegans microRNAs

Nicolas J. Lehrbach and Eric A. Miska

Advance Access publication date 19 June 2008

Abstract

MicroRNAs (miRNAs) are short non-coding RNAs that regulate gene expression in many eukaryotes. miRNAs were first discovered in Caenorhabditis elegans by Victor Ambros’ laboratory in 1993. At the same time Gary Ruvkun’s laboratory identified the first miRNA target gene. Together, these two seminal discoveries identified a novel mechanism of post-transcriptional gene regulation that has been recognized as important for development, physiology and pathology of many organisms. Here we discuss how functional genomic, computational and proteomic approaches complement classical genetic analyses to unravel miRNA biology in C. elegans.

Keywords: C. elegans; miRNA; microRNA; proteomics; SILAC; Argonaute; Ago; post-transcriptional gene regulation; let-7, Isy-6

INTRODUCTION

MicroRNAs (miRNAs) are an abundant class of small RNAs, receiving increasing attention for their important regulatory roles. Regulation by miRNAs has been implicated in a range of developmental and physiological processes and human disease, such as cancer [1].

miRNAs are endogenous small, ~22nt RNA molecules. Transcription of miRNA genes produces a primary transcript containing a hairpin structure [2]. Sequential nuclear and cytoplasmic processing steps act to release the mature miRNA from the hairpin. The mature miRNA is then loaded into an effector complex (miRNA induced silencing complex; miRISC) by interaction with an Argonaute (Ago) protein. In Caenorhabditis elegans the Ago proteins encoded byalg-1 andalg-2 are redundantly required for miRNA function [3]. The miRNA then acts as the specificity determinant of the silencing complex, hybridizing to complimentary sequences in the 3’UTRs of target mRNAs. This interaction results in reduced translation of the target mRNA and target mRNA destabilization through a poorly understood mechanism [2, 4]. miRNAs have been found in all multi-cellular organisms examined so far, as well as viruses, and a unicellular alga [5–11]. miRNA genes are relatively abundant, for example the human genome contains at least 574 miRNA genes, or more than 2% of all genes [12, 13]. Because many mRNAs have been under selective pressure to preserve pairing to a 6-mer in the 5’ region of the miRNA known as the miRNA seed (nucleotides 2–7), targets of metazoan miRNAs can be predicted above the background of false positives by searching for conserved matches to the seed region [14–18]. In nematodes, at least 10% of protein-coding messages appear to be conserved targets of miRNAs [19].

The free-living nematode C. elegans has been used as a model organism for studies in genetics, development and behaviour for over 30 years. It is an excellent choice due to its simple genetics, small genome, fast generation time, invariant cell lineage and translucent body. As such, research using C. elegans has made important contributions to our understanding of a range of processes, not least amongst which are the functions of miRNA genes. Not only were miRNAs first discovered in C. elegans but this organism provides an excellent model for future exploration of miRNA function.

DISCOVERY

The first two miRNAs, defined by the lin-4 and let-7 genes, were discovered by C. elegans
they were first termed short temporal RNAs (stRNAs), as their functions appeared specialized to developmental timing. The subsequent realization that the let-7 miRNA was conserved throughout the bilateria led to a search for additional small regulatory RNAs. Coinciding with the emergence of the RNAi-field miRNAs became established as an abundant class of regulatory RNAs [8, 9, 23]. Since the discovery of the lin-4 and let-7 miRNAs, some additional miRNAs have been discovered by forward genetic approaches, but the bulk of new miRNAs have been found by combined biochemical and computational approaches. The biggest source of new miRNAs has been the sequencing of cloned small RNA libraries. So far 118 miRNA genes have been confidently identified in C. elegans. 76 of the corresponding miRNAs can be grouped into 23 families with more than one member based on sequence identity in the ‘seed’ bases 2–7 (Figure 2) [24–27]. At least one third of all C. elegans miRNA families are conserved beyond the nematoda [24].

The abundance and diversity of miRNAs in the animal kingdom indicates the potential wealth of biological functions they may hold. The utility of C. elegans as a model for study of miRNA function is

Figure 1: Five miRNAs and two miRNA families control the timing of larval development in C. elegans.

Figure 2: Two thirds of C. elegans miRNAs are part of a family of miRNAs. Families shown here are based on identical ‘seed’ matches, i.e. identical bases 2–7.
exemplified by those miRNAs already described by forward genetics in this animal. Here we give a brief overview of known functions of miRNAs in *C. elegans*, and suggest novel approaches that might expand on this knowledge in the future.

**ROLES OF MiRNAs IN C. ELEGANS**

Forward genetics approaches have found miRNAs with important roles in specifying cell fates and functions by temporally and spatially restricting expression of target genes. The developmental timing pathway serves to specify temporal cell fates during larval development in *C. elegans*, and has been particularly well studied in the context of fate decisions made in the developing hypodermis. In these cells, the sequential expression of miRNA genes is part of a cascade of gene regulatory interactions that control temporal cell fate and proliferation (Figure 1). Mutants for the *lin-4*, and *let-7* miRNAs each show defects at different stages in development that can be suppressed by mutations in their respective targets [21–23, 28]. Additionally, the *let-7* family members *mir-48, mir-84* and *mir-241* together control gene expression independently of *let-7* [29, 30]. Some additional targets of *let-7* miRNA have been identified that may reflect additional functions of *let-7* in developmental timing outside of the hypodermis [31].

Another miRNA identified by forward genetics is the *lsy-6* miRNA. The *lsy-6* miRNA is required for the asymmetrical expression of taste receptor genes in the morphologically symmetrical sensory ASE neurons (ASEL and ASER). The *lsy-6* miRNA regulates *cog-1*, which encodes a transcription factor [32]. Two miRNA families have been implicated in vulval development, although these data rely on miRNA misexpression and have not yet been supported by loss-of-function studies. The misexpression of miR-61 in vulval progenitors drives changes in gene expression in those cells [33], and over-expression of miR-84 can suppress the multi-vulva phenotype of *let-60/RAS* gain-of-function alleles [34]. A number of additional roles for miRNAs in *C. elegans* have been identified and a complete list is shown in Table 1.

**APPROACHES TO UNDERSTAND MiRNA FUNCTION**

Central to understanding the biological roles of miRNAs is the identification of miRNA targets. As miRNA target levels are expected to be up-regulated in miRNA loss-of-function mutants, a number of miRNA target genes have been identified as suppressors of miRNA mutant phenotypes. Surprisingly, computational and genetic approaches to identify additional miRNA targets has proven remarkably difficult. However, the combination of computational, functional and biochemical methods available in *C. elegans* promises to overcome this barrier.

**GENETICS**

*C. elegans* is unique amongst model organisms in that loss-of-function alleles for the majority (>80%) of *C. elegans* miRNAs are currently available [35]. This is an important resource as it not only allows direct observation of the phenotypic consequences of loss

---

**Table 1:** Function of *C. elegans* miRNAs

<table>
<thead>
<tr>
<th>miRNA</th>
<th>Targets</th>
<th>Function</th>
<th>Evidence</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>let-7</em></td>
<td><em>lin-41, hbl-1, def-12, pha-4</em></td>
<td>Developmental timing/stem cell differentiation</td>
<td>LOF</td>
<td>[20, 28, 31, 53, 54]</td>
</tr>
<tr>
<td><em>lin-4</em></td>
<td><em>lin-14, lin-28</em></td>
<td>Developmental timing/stem cell differentiation</td>
<td>LOF</td>
<td>[21, 22]</td>
</tr>
<tr>
<td><em>lin-4</em></td>
<td><em>lin-14</em></td>
<td>Longevity</td>
<td>LOF, GOF</td>
<td>[35]</td>
</tr>
<tr>
<td><em>lsy-6</em></td>
<td><em>cog-1</em></td>
<td>Left-right asymmetry</td>
<td>LOF</td>
<td>[32]</td>
</tr>
<tr>
<td><em>mir-35-4I</em></td>
<td>Unknown</td>
<td>Embryonic development</td>
<td>LOF</td>
<td>[35]</td>
</tr>
<tr>
<td><em>mir-48, mir-84, mir-241</em></td>
<td><em>hbl-1</em></td>
<td>Developmental timing/stem cell differentiation</td>
<td>LOF</td>
<td>[29, 30]</td>
</tr>
<tr>
<td><em>mir-48, mir-84</em></td>
<td><em>nhr-23, nhr-25</em></td>
<td>Molting</td>
<td>LOF</td>
<td>[30, 56]</td>
</tr>
<tr>
<td><em>mir-6I</em></td>
<td><em>vav-1</em></td>
<td>Vulval development</td>
<td>GOF</td>
<td>[33]</td>
</tr>
<tr>
<td><em>mir-84</em></td>
<td><em>let-60</em></td>
<td>Developmental timing/stem cell differentiation</td>
<td>GOF</td>
<td>[34, 56]</td>
</tr>
<tr>
<td><em>mir-240, mir-786</em></td>
<td><em>unknown</em></td>
<td>Defecation</td>
<td>LOF</td>
<td>[35]</td>
</tr>
<tr>
<td><em>mir-273</em></td>
<td><em>die-1</em></td>
<td>Left-right asymmetry</td>
<td>GOF</td>
<td>[57]</td>
</tr>
</tbody>
</table>

This table includes all miRNAs that have been analysed using loss-of-function or gain-of-function experiments. This table does not contain miRNAs for which a target mRNA has been predicted and/or validated, but which was not analysed further. *It is unclear whether miR-273 is a miRNA.
of a specific miRNA, but also provides a resource for further functional characterization. Surprisingly, the majority of C. elegans miRNAs are individually not required for normal development; suggestive of functional redundancy with other miRNA or protein-coding genes [35].

Loss-of-function alleles of miRNA genes provide an opportunity to study their functions through forward and reverse genetic approaches. In particular, genetic interactions might indicate genes regulated by, or regulating the same process as a given miRNA. In cases where a miRNA deletion results in an observable abnormal phenotype, it is anticipated that this results from inappropriate expression of gene(s) normally repressed by that miRNA. Consequently, loss-of-function mutations in target genes could be identified in suppressor screens. Indeed, this is how the let-7 miRNA target, lin-41, was identified [28]. However, there are some difficulties with this approach; suppressors may not be identified in cases where a miRNA regulates multiple targets, or where a target gene is essential. In addition to suppressor screens, synthetic or enhancer screens, e.g. using RNAi may also provide clues to miRNA function. Analogous genome wide RNAi screens for genetic interactors have been informative for protein coding genes [36, 37]. Such an approach might prove particularly useful in cases where loss of a particular miRNA causes no obvious abnormal phenotype.

MIRNA TARGET PREDICTION

The difficulty in identifying miRNA targets by genetic approaches, and the superficially simple hybridisation-based rules of miRNA target recognition, have lead to an explosion of computational algorithms designed to predict miRNA targets. Computational target predictions have relied on the properties of the small number of genetically validated miRNA targets, identified in C. elegans and Drosophila melanogaster. miRNA targets appear to preferentially base-pair with miRNAs through their 5’ end or ‘seed’ region, which is the basis of many miRNA target prediction algorithms. Additionally, these algorithms usually include a requirement for target site conservation to reduce the number of false positives [17, 19, 38]. Subsequent iterations of these algorithms have included more sophisticated models of miRNA:mRNA interactions, taking into account such factors as local secondary structure of the miRNA, or effects of target site position within the 3’UTR [39, 40].

Current computational predictions have, however, a number of limitations. Firstly, the initial training set of known miRNA targets is very small, and so may not faithfully reflect the complexity of all miRNA:mRNA interactions. As such, they are unlikely to uncover unusual miRNA target interactions; for example, it is possible that miRNAs are able to mediate translational repression through sites outside of the 3’UTR [41]. The common algorithms also do not take into account the expression pattern of miRNAs and the predicted target genes, and so will include functionally irrelevant predictions, in cases where the miRNA and mRNA are never co-expressed. An additional concern is that many prediction algorithms have used cell-based assays for target validation in which the miRNA and its putative targets are over-expressed. Although these assays may lend some support the predictions, these are unlikely to reflect conditions under which endogenous miRNAs function. If data from over-expression assays are incorporated into further refinements of the prediction model, this may further skew predictions away from endogenous targets. For example, a study that tested the genetically validated target of the lsy-6 miRNA (cog-1), alongside 13 computationally predicted targets using an invivo reporter system, which relies on endogenously expressed miRNA, found that while the cog-1 3’UTR was indeed regulated by lsy-6, none of the other predicted targets were [42].

There is abundant evidence that miRNAs, in addition to inhibiting mRNA target translation, reduce message levels [43]. Therefore, comparison of miRNA from wild-type and miRNA mutant animals using miRNA microarrays might identify direct miRNA targets. Indeed, this approach has been used in mammalian cells [44]. However, as the effects of miRNAs on target mRNA levels are modest, direct effects might be masked by secondary effects. For whole animal studies such an approach might only be feasible for ubiquitously expressed miRNAs. Alternatively, cell populations might be sorted using GFP transgenes and fluorescent cell sorting prior to miRNA analysis [45]. This approach, in conjunction with miRNA target prediction, might aid in the experimental identification of many miRNA targets. However, some miRNA targets appear to be regulated at the translational level only and would be missed using this method [21].
BIOCHEMICAL IDENTIFICATION OF MIRNA TARGETS

The difficulties in identifying miRNA targets, and hence functions, by established techniques may be overcome by new approaches exploiting new genomic and proteomic tools. One promising means of identifying miRNA targets is through the physical interaction of targeted mRNAs with the miRISC complex. This can be achieved by immunoprecipitating (IP) miRISC-associated proteins to purify intact ribonucleoprotein complexes (RNPs) containing miRNAs and their mRNA targets. Extracting mRNAs that are enriched in the immunoprecipitated material should identify direct miRNA targets. A few recent studies have used this approach, and produced promising results. A study in *C. elegans* identified potential miRNAs targets by their association with AIN-1 and AIN-2. AIN-1 and AIN-2 each interact with ALG-1/2, and are together required for efficient mRNAs function [46, 47]. Using microarrays to analyse co-immunoprecipitated mRNAs with AIN-1 and AIN-2, this approach yielded 3500 potential miRNA targets, including most of the validated miRNA targets in *C. elegans* [46]. Similar approaches have also been used in human cell lines and *Drosophila*, using Ago proteins for immunoprecipitation followed by microarray analysis [48, 49].

High-throughput sequencing technologies now provide the opportunity to extend this approach to identify not only miRNA targets but also miRNA-binding sites in target mRNAs. A recent study of an alternative-splicing regulator used immunoprecipitation of crosslinked RNP complexes followed by partial RNAse digestion to identify the sites within primary transcripts bound by an alternative-splicing factor [50]. Similar treatment of miRISCs might identify miRNA target sites within mRNAs. A proposed protocol is shown in Figure 3. Immunoprecipitated miRISC-associated RNAs would be partially digested using RNAse treatment followed by RNA extraction from miRISC followed by RNA identification using high-throughput sequencing.

**Figure 3:** Example protocol for miRNA target site identification in *C. elegans* by RNA immunoprecipitation followed by high-throughput sequencing 'RIPseq'. Comparison of wild-type and miRNA mutant animals might identify target sites for specific miRNAs. The miRISC RNPs are immunoprecipitated and treated with RNase. RNA protected by miRISC is extracted and subjected and cloned for high-throughput sequencing.
Finally, to identify the target mRNA sites associated with a specific miRNA, this approach could be combined with miRNA loss-of-function mutants (Figure 3).

**PROTEOMICS**

miRNAs are thought to regulate mRNA translation. Consequently, measurement of protein levels is the most direct method to identify miRNA targets. However, tools for high-throughput quantification of protein levels or proteomic analyses lag behind tools for mRNA quantification or transcriptome analyses. Nevertheless, a recent study using stable isotope labelling by amino acids in cell culture (SILAC) identified a number of proteins, whose levels were altered after miR-1 transfection [51]. Interestingly, the mRNAs of the proteins down-regulated after miR-1 transfection were enriched in predicted miR-1 binding sites. A similar approach might be used to identify targets of *C. elegans* miRNAs (Figure 4). A number of alternatives to SILAC exist, for example iTRAQ [52]. In addition to such ‘shotgun’ approaches, miRNA target predictions might also be validated using a targeted

---

**Figure 4:** Suggested protocol for miRNA target identification using a proteomic approach. SILAC is shown as an example. Animals are grown on normal media or media containing stable, heavy \(^{13}\)C and \(^{15}\)N isotopes. Protein extracts are mixed prior to digestion and mass spectrometry analysis. Peptides from the two samples are distinguished by their difference in mass.
proteomics approach. The available collection of miRNA mutants in *C. elegans* might be a useful tool to identify *in vivo* miRNA targets using proteomics, although the same caveats mentioned above apply.

**CONCLUSIONS**

miRNAs may be late arrivals to biological research, but miRNA researchers can now take advantage of an advanced set of tools that only recently became available. Whole genome sequences and high-throughput sequencing aids miRNA discovery, while comparative genomics is crucial for miRNA target prediction. The identification of transcription factor networks was revolutionized by *in vivo* target identification through chromatin immunoprecipitation approaches (e.g. ‘ChIPseq’), the same might be true for miRNA regulatory networks and ‘RIPseq’. In *C. elegans* these new approaches can be combined with classic genetic analysis to discover new miRNA biology. Alas, we are still lacking nematode cell lines.

**References**


40. Lytle JR, Yario TA, Steitz JA. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 3' UTR as in the 5' UTR. Proc Natl Acad Sci USA 2007;104:9667–72.


51. Hoe Downloaded from: https://academic.oup.com/bfg/article-abstract/7/3/228/260026 by guest on 05 March 2019