Chapter 19

44 Current Challenges in miRNomics

Bünyamin Akgül, Peter F. Stadler, Liam J. Hawkins, Hanane Hadj-Moussa, Kenneth B. Storey, Kemal Ergin, Rahmi Çetinkaya, Alexandre R. Paschoal, Pedro G. Nachtigall, Yusuf Tutar, Malik Yousef, and Jens Allmer

Abstract

Mature microRNAs (miRNAs) are short RNA sequences about 18–24 nucleotide long, which provide the recognition key within RISC for the posttranscriptional regulation of target RNAs. Considering the canonical pathway, mature miRNAs are produced via a multistep process. Their transcription (pri-miRNAs) and first processing step via the microprocessor complex (pre-miRNAs) occur in the nucleus. Then they are exported into the cytosol, processed again by Dicer (dsRNA) and finally a single strand (mature miRNA) is incorporated into RISC (miRISC). The sequence of the incorporated miRNA provides the function of RNA target recognition via hybridization. Following binding of the target, the mRNA is either degraded or translation is inhibited, which ultimately leads to less protein production. Conversely, it has been shown that binding within the 5′ UTR of the mRNA can lead to an increase in protein product. Regulation of homeostasis is very important for a cell; therefore, all steps in the miRNA-based regulation pathway, from transcription to the incorporation of the mature miRNA into RISC, are under tight control. While much research effort has been exerted in this area, the knowledgebase is not sufficient for accurately modelling miRNA regulation computationally. The computational prediction of miRNAs is, however, necessary because it is not feasible to investigate all possible pairs of a miRNA and its target, let alone miRNAs and their targets. We here point out open challenges important for computational modelling or for our general understanding of miRNA-based regulation and show how their investigation is beneficial. It is our hope that this collection of challenges will lead to their resolution in the near future.

Key words miRNomics, Challenges, miRNA prediction, miRNA targeting, Mature miRNA, RISC

1 Introduction

Cell homeostasis is vital; therefore, gene expression (RNA/protein) is under tight control [1]. The dysregulation of gene expression is associated with various disorders such as Parkinson's [2] and cancers [3, 4]. Regulatory interventions are manifold. For example, transcript abundance is influenced by cis- and trans-acting transcriptional regulators and epigenetics mechanisms while protein lifetime is controlled via, for example, ubiquitination [5]. One

Jens Allmer and Malik Yousef (eds.), *miRNomics: MicroRNA Biology and Computational Analysis*, Methods in Molecular Biology, vol. 2257, https://doi.org/10.1007/978-1-0716-1170-8_19, © Springer Science+Business Media, LLC, part of Springer Nature 2022

regulatory pathway which was discovered relatively recently [6] controls protein abundance following transcription and export of the mRNA templates. This process, termed posttranscriptional regulation, is currently under heavy investigation due to the promise it holds of generating easily accessible biomarkers and personalized therapeutic interventions [4, 7].

Checks and balances need to be in place to ensure that any regulatory mechanism cannot wreak havoc on cell homeostasis. Therefore, miRNA biogenesis and targeting are under strict regulation themselves [8]. MicroRNA biogenesis occurs in a multistep process [9, 10] starting with the transcription of pri-miRNAs. This initial transcription step is controlled by genetic and epigenetic mechanisms. Transcriptional control is more relevant to miRNAs produced from their own loci while other miRNAs might be coproducts of other transcripts such as mRNAs or long noncoding RNAs (lncRNAs). Mirtrons, microRNAs encoded within introns, are sideproducts of genes and do not influence the abundance of the host gene they are encoded in directly. This is different for miRNAs which are encoded within exons. A relevant example in this context is a miRNA (not listed in miRBase yet) encoded within the sixth exon (or seventh intron; depending on gene version) of the Di George Critical Region Eight (DGCR8) [11]. This presents a direct feedback loop for the abundance of the microprocessor and produces a miRNA which has several putative targets [12]. Both events, however, are not direct and the decreased amount of mRNA for DGCR8 will take effect during its translation and the resulting pre-miRNA needs to proceed through the miRNA biogenesis and targeting pathways to take effect. Here emerges a very important question, what is the abundance of the transcripts involved and how much ncRNA/protein will they produce? Obviously, this question is too complex to be answered in a straightforward manner. Therefore, we will break down questions into smaller, manageable chunks in the following.

2 Open Challenges

2.1 Transcription and Transcriptional Regulation

For intergenic miRNAs the primary transcript frequently remains unannotated, thus making it difficult to systematically investigate miRNAs under similar transcriptional control. The situation is also far from ideal for intronic miRNAs. Here the question remains whether miRNAs are processed equivalently from alternative overlapping transcripts. Furthermore, there are indications for a coupling of alternative splicing and miRNA processing in plants [13, 14]. Disentangling such connections will require a systematic analysis of matching sRNA-seq and RNA-seq data to see the relationships between precursors and processing products. A wealth of such data is presumably available as part of the large-scale cancer

sequencing projects, but does not seem to have been mined with such generic questions in mind. From a computational perspective, prediction of RNA-Polymerase binding sites and their resulting transcription efficiency would be beneficial to pinpointing miRNAs from their own loci. Attempts have been made to analyze large scale ChIP-seq data leading to the detection of many transcription factor binding sites [15]. However, not all possible stress responses, environmental conditions, tissues, and so on, can be queried via ChIP-seq; therefore, a computational tool for predicting such binding sites is needed. This leads to the following list of queries.

1. Improve miRNA annotation [16], for example, for the primary transcript of intergenic loci.
2. Define whether there is an effective expression difference between exons and miRNAs encoded in the introns of the same genes, considering alternative transcripts.
3. What is the effect of miRNA processing on alternative splicing?
4. Mine the large amount of available sRNA- and RNA-seq data from different perspectives, for example, to answer (1–3).
5. Detect or predict RNA-polymerase binding sites to aid prediction of pri-miRNAs.
6. Predict RNA-polymerase efficiency (based on binding sites) to predict the abundance of pri-miRNA transcripts that can be expected.

In summary, points (1–6) pertain to the questions of whether and how much initial transcript can be expected from a locus. With more than a hundred million hairpin structures that can be predicted from the human genome defining whether they are likely to be expressed in large enough amounts would aid in filtering many of the candidate miRNAs.

2.2 Steps in the miRNA Biogenesis and Targeting Pathway

In order to better understand miRNA-based regulation, a number of important investigations should be performed.

7. Define the specificity of Drosha.
8. Define the specificity of Drosha cleavage.
9. Define the specificity of Exportin-5.
10. Define the specificity of Dicer.
11. Define the possible types of mature miRNAs derivable from a pre-miRNA.
12. Define incorporation into RISC.
13. Define targeting specificity and effect.
14. Define number and intracytoplasmic localization of RISC per cell.

15. Define mRNA translation efficiency.
16. Define how often loaded RISC can modulate translation.
17. Define dissociation of miRNA from RISC.

7. Some of the challenges (7–16) are still complex and can be further broken down. For example, defining Drosha, includes whether there are structural and/or sequence requirements for hairpin recognition. Structural questions can be answered relatively easily by carefully designing a set of RNAs including one hairpin each, with varying structural features. Incubating them with Drosha for a set period of time and determining the ratio of educts to products can define the enzyme's specificity. Given knowledge from the previous experiment, the hairpins can be varied in terms of sequence while the structure stays intact. Performing the same experiment as before will show whether Drosha has a preference for sequence features. Determining these rules, will not only benefit the prediction of pre-miRNAs from a genome but will further allow for quantitative modelling.
8. It is not only important to understand which hairpins (if any bias) Drosha processes but also what the outcome may be. There are accounts in literature ranging from blunt cutting of the RNA duplex, to several nucleotide overhang on either side. Knowledge about this could be a side effect of experiments performed for (7) since sequencing the products of the known educts will lead to clearly defined cleavage rules whether based on structure, sequence, or both.
9. Any hairpin excised by Drosha or created as a miRTron, needs to be exported from the nucleus into the cytosol to progress in the miRNA biogenesis pathway. Incubating purified Exportin-5 with hairpins and performing a pull-down experiment is one approach that could be taken. Whether inserting Exportin-5 into vesicles along with other essential proteins, filled with pre-miRNAs and then measuring efflux is an easy experiment, is unlikely. Perhaps similar approaches might be performable. Unless some hairpins are rejected from transport due to structural or sequence features, which seems unlikely, working knowledge about Exportin-5 will allow quantitative modelling of the transport.
10. Similar to Drosha, it needs to be tested whether Dicer has a specificity in terms of structure, sequence, or both and what the exact outcome of the processing is. A similar approach can be taken as for Drosha (Challenges 7 and 8). In addition, the position of the cleavage event needs to be determined precisely and whether it is dependent on structure and/or sequence of

the stem-loop. Determining the cleavage rules will strongly enhance the prediction of the mature sequence from the stem-loop precursor.

11. In addition to mature microRNAs derived from the 5′-arm or the 3′-arm of the hairpin, there is evidence that the loop region can also give rise to a third mature microRNA called "loop-miR" [17, 18]. These loop-miRs have been found in different species and have been shown to be functionally active. The only known determinant of loop-miR production is the size of the loop, that is, the distance between the Dicer processing sites at the 3′-end of the 5′-arm and the 5′-end of the 3′-arm. For the future, it would be important to investigate, how many and which pre-miRNAs can give rise to this third mature microRNA, to establish their functional role and target spectrum and to elucidate whether they could be used in biotechnology to create multifunctional microRNA precursors giving rise to multiple independent microRNA sequences.
12. There are different accounts of incorporation of mature sequences into RISC. One mature sequence, both with some bias, or without can be assembled into miRISC. For computational prediction and for the understanding of the process, it is important to know which mature sequence or in which ratio they are incorporated into RISC. Here current methodologies such as HITS-CLIP and CLASH [19, 20] can help to define the amount of mature miRNAs bound in miRISC.
13. Currently, knowledge about targeting is based on a small number of early studies. There a seed sequence and an out-seed have been defined. This needs to be tested on a larger scale. For example, miRISC with different mature sequences could be used to pull down random RNAs which can then be sequenced. Thereby, the binding affinity can be defined. In a follow-up experiment, overexpression experiments can be performed and the ratio of translational repression to mRNA degradation can be measured. When done for a large variety of miRNAs and targets, this can help detect any sequence and structure biases for the miRISC-mRNA system. Furthermore, such experiments will allow for a better understanding of the targeting process and perhaps more precise computational models.
14. There exist estimates for the number of RISC complexes in a cell. Here it would be interesting to analyze the RISC content among different tissues and species. Another important finding would be the ratio of miRISC versus free RISC. Adding to this complexity is the intracytoplasmic re-localization of miRISC complexes, which might add another layer to the regulation of miRNA activity [21, 22].

15. It is important to understand how effective an mRNA is translated without miRNA influence. This is something that is being investigated computationally in the DREAM challenges in proteogenomics. Then the effects of miRNA on the translation can be better understood and perhaps delineated based on binding site motifs, location of binding sites, and binding site multiplicity. Initially, more combined investigations of mRNA and protein abundance are needed.
16. After RISC binds an mRNA, can the same RISC complex dissociate from the mRNA (perhaps after processing the mRNA) and bind another mRNA for cleavage or translational repression? This is an important question which goes hand in hand with the following one.
17. Whether an assembled miRISC complex can revert to free RISC and bind another miRNA and what are the association and dissociation constants?

Many functions can derive from the knowledge queried in (7–16), for example, when designing miRNA mimics to control diseases such as cancer. This list above pertains only to the canonical processing of miRNAs. However, there are many alternatives that only partially overlap with the canonical pathway [23]. It would be important to understand (a) how to identify the sequence processing steps in each case, (b) to what extent they can vary and admit alternatives (e.g., loop miRNAs processed from precursors that also produce canonical Dicer-cleaved products, or the production of microRNA-offset RNAs, moRNAs, from overly long precursor hairpin [24–26], and (c) to what extent the usage of pathways changes throughout evolution. One should also keep in mind that details of the relevant pathways will not have been invariant over large evolutionary time scales. We suspect, therefore, that the answers to the questions above (7–17) will show subtle differences between clades. It will be important in particular for computational approaches to understand this variability. The fact that it is possible to recognize a clade by the k-mer distribution of microRNAs [27, 28] is a clear indication for such an effect. MicroRNAs seem to have emerged independently several times in the history of Eukaryotes, thereby using the RNAi pathway(s) as an additional layer of gene regulation. What are the commonalities and differences between these innovation events? To what extent are homologous proteins used in the pertinent pathways, and to what extent have different pathways been co-opted into this new function. Is the observed plurality of processing pathways in animals a reflection of the ease with which the RNAi pathway(s) can be accessed by feeding them with custom-produced RNAs? Therefore, the questions (7–17) need to be investigated in different species from different clades to gain a better understanding of functionality and miRNA evolution.

2.3 MicroRNA and Target Prediction

Experimental miRNA and miRNA target detection has improved with many novel technologies enabling the capturing of actual miRNA target interactions [29]. Nonetheless, it is not feasible to investigate miRNA, mRNA, and protein expression under all developmental stages, all tissue types, and all different stressors. Therefore, computational prediction of miRNAs and their targets is indispensable [20]. While pre-miRNA prediction is at a quite mature state [30], the prediction of mature miRNAs [3] from these pre-miRNAs and the prediction of miRNA targets needs further improvement [31]. A deeper understanding of how the various enzymes in the miRNA biogenesis and targeting pathway work would be very beneficial for computational modeling.

18. Does the Microprocessor complex have a structural or sequential bias?
19. What are the exact cleavage rules of the Microprocessor complex, and are they dependent on sequence or structure?
20. Is there a structural or sequential bias for pre-miRNA export into the cytosol?
21. Are all transported pre-miRNAs associated with Dicer or is there a structural, sequential bias or a competition among pre-miRNAs?
22. What are the exact cleavage rules of Dicer, and are they dependent on sequence or structure?
23. On which factors does the mature miRNA incorporation into RISC depend?

18. Similarly to questions (7) and (8), question (18) investigates pri-miRNA processing into pre-miRNA. It is likely that the Microprocessor complex has a structural bias and for further investigation in this direction more cocrystals have to be produced and linked to processing efficiency. Binding often changes the structures of the partners and, therefore, the RNA structure before binding is another clue needed to better understand miRNA genesis. Whether apart from structural biases sequence biases play a role can be investigated by analyzing processing efficiency of a large number of pri-miRNAs (perhaps containing many pre-miRNAs) where the predicted structures are largely similar but the sequence varies.
19. seems similar to (18) but here the question is where the pre-miRNA is cleaved to produce the pre-miRNA. Is there a structural dependence such as cleavage n bases following the opening of the stem? Maybe the distance is measured from the loop since only a limited size structure fits into the Microprocessor complex. Perhaps some kind of sequential or energetical

motif is required for the cleavage to occur. What determines the actual cleavage leading to blunt products or products with an overhang? Answering these questions would greatly reduce the amount of false positive pre-miRNA predictions.

20. Following the excision of pre-miRNAs, are they all exported or do some remain in the nucleus and are perhaps degraded there? Here markers could help determine the cellular localization of miRNAs. Perhaps the flux of pre-miRNAs from overexpressed pri-miRNAs can be monitored live using markers and confocal laser scanning microscopy.
21. Another important question is whether hairpins are free-floating in the cytosol or whether they are handed from one protein complex to the next, for example, Microprocessor→Exportin→Dicer. Whether free floating or not, is there a competition among pre-miRNAs for Dicer processing and then incorporation into RISC? If there is competition, is it structural or sequence-based? Perhaps it is based on the binding energy distribution in the stem or maybe just miRNA abundance.
22. How does Dicer cleave the stem-loop exactly? Are there overhangs or blunt ends? Is the cleavage dependent on sequential or structural features? These questions can be queried in vitro.
23. Similarly, how is the incorporation of the mature miRNA into RISC achieved? There seems to be a bias for one of the strands of the ds-RNA following Dicer processing. What are the rules driving the selection of the strands? Are they based on binding energy, sequence-based, or structural? Given a set of miRNAs with perfect complementarity, thereby excluding structural components, sequence and energetical components can be investigated.

2.4 MiRNA Tools and Databases

MicroRNAs have seen a lot of experimental investigation in the past decades which led to humongous amounts of data available for further analysis. However, it is a challenge to incorporate all this information into a comprehensive resource. Having this in mind, we focused on exploring the available literature and providing useful and practical guidance on the miRNA database and tools [32]. From a bioinformatics perspective, the prediction of pre-miRNAs, mature miRNAs, miRNA-targeting are important areas that need improvement. Also, integrating the various databases which include miRNAs, targets, expression, and function and other information separately, would be beneficial. Despite all advances in the miRNA and miRNA-target prediction field, the next-generation of predictors should incorporate the following tasks.

24. Improve precision, by increasing the number of true positives and decreasing the number of false positives in the predictions for all pre-miRNA, mature, and target prediction.
25. Integrate the prediction into the regulatory networks and pathways by considering the expressed genes in the tissue/condition being analyzed.
26. Detect when the action of a miRNA on a specific target is direct or indirect (i.e., a miRNA targets a gene and its genes regulate another gene that also has a binding site, but this binding site is not functional).
27. Take into account the transcript isoforms present of the specific target. In this sense, the predictors should consider the existence of alternative poly-A signals in the 3′ UTR sequence, and, when possible, the tools should use transcriptome data to confirm the 3′ UTR isoform present in the tissue/condition being analyzed as well as consider alternative splicing.
28. The miRNA target predictors should consider the genomic context of the UTR and the region of the binding site therein (i.e., the position of the binding site, secondary structure of the UTR, etc.).
29. Predictors for targeting should consider target multiplicity, distance of targets to the stop signal, and other information that could be integrated.
30. Current miRNA databases are largely dependent on miRBase for naming miRNAs and they are interlinked. Both are good, but an ontology integrating all information and allowing logical reasoning other the data would do much to improve on the current knowledge.
31. Enable the prediction of noncanonical mature miRNAs such as loop-miRs.
32. Establish whether sequence information is sufficient to predict the mature miRNA.

31. As pointed out in challenge (11), it is important to discover noncanonical mature miRNAs such as loop-miRs that are located in the loop part. Yet those mature miRNAs are ignored by current tools that predict the location of the mature miRNA [3]. Here the challenge is to develop an algorithm enabling the identification of noncanonical mature miRNAs. At the same time it is important to annotate miRNAs more precisely in existing databases to support algorithm development.
32. Recent studies show that it is possible to predict the microRNA precursor considering only sequence information, more

specifically, utilizing a vector of k-mers [27, 33, 34]. Studies considering the mature miRNAs' targets similarly led to some success when just employing sequence information [35, 36]. Here the challenge is to explore the idea of the prediction of the location of the mature part considering just the sequence information. This can also answer whether sequence information is a determining factor for the selection of mature miRNAs during RISC incorporation.

Points (24 to 30) bring us to miRNA function which (*see* query 25) is not a singular process void of interactions with other processes [12].

2.5 MicroRNA Function

MicroRNAs are under spatiotemporal control and so are messenger RNAs. Posttranscriptional control involving miRNAs can only occur when both the miRNA and at least one of its targets is coexpressed in the same space at the same time. Whether both are expressed in the same space can be confirmed relatively easily using RNA-seq. MicroRNAs can also be transported to their targets via exosomes [37]. Posttranscriptional regulation can only be measured on the transcriptional level for those regulations where the mRNA is degraded, however, it may be hard to differentiate among RNAs specifically degraded in response to miRNA regulation and other degradation events. For those miRNA regulatory events that modulate protein abundance without degrading the mRNA, protein abundance needs to be correlated with mRNA abundance and miRNA abundance and its overall occupancy status in RISC. Additionally, miRNAs with many coexpressed targets may not cause a measurable effect while the same miRNA coexpressed with only a few of its targets may cause strong regulation. Additionally, miRNA sponges (see below) can modulate miRNA regulation. Even for two slightly similar miRNAs (X and Y) a complex interaction network can unfold (Fig. 1). The hypothetical regulatory network in Fig. 1 also shows that some target sites may be shared among miRNAs while others are not. The location of target sites may also be important (exon mRNA D) or in 3′ UTRs (A–D). The lncRNA in the example will likely hide any gene regulatory effect of miR-X while not significantly affecting miR-Y (Fig. 1).

Therefore, all possible target sites of a miRNA need to be monitored. MicroRNAs can be derived from genes and, therefore, they are bound to the same regulation as the gene. To elucidate this complex situation, a number of statistics need to be defined.

33. How much RISC is available in a cell?
34. What is the association rate and dissociation rate of the mature sequence and RISC, that is, how long is a miRISC complex active before the mature miRNA is replaced or the miRISC is degraded?

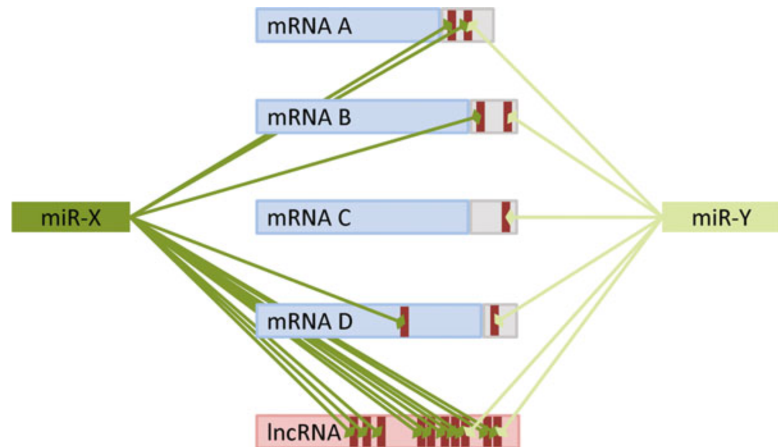

Fig. 1 Two miRNAs (green) and three of their mRNA targets (blue) as well as one long noncoding RNA acting as a sponge (red). Target sites are red boxes and connections show binding of the two miRNAs. 3′ UTRs are in gray

35. How many proteins are translated from one mRNA before it is degraded?
36. When miRISC binds an mRNA, but does not degrade it, will it dissociate from the mRNA and allow translation in another round or will it stay bound until either is degraded?
37. Do all miRNAs have the same chance to be incorporated in RISC or is there a structural bias?

33. while this seems trite at first glance, it cannot be expected that the RISC amount is similar for all cell types and, therefore, this needs to be investigated for various tissues. Perhaps the RISC amount also varies in response to stress and other stimuli. Initially (*see* 4), overall expression levels could be established mining publicly available mRNA and protein expression data.
34. Similarly to (17), it is important to understand whether RISC can be reused and whether the mature miRNA goes back into the pool of possible miRISC partners or is degraded. How fast do mature miRNAs associate and dissociate to RISC (if they do). Once loaded, is the miRISC only used to regulate one mRNA or is it reused and if it is how long does the mature miRNA stay associated with miRISC?
35. When miRISC inhibits translation; it is important to know whether the same mRNA would be translated multiple times. If so, does the inhibition affect all possible translation events of the mRNA?
36. Very similar to (33), assuming that there are multiple rounds of translation from on mRNA and that there is only translational

repression. Does miRISC dissociate from the mRNA after failed translation and bind another mRNA or is it bound to the mRNA and degraded together with it?

37. Similarly to processing with Drosha and Dicer, is there a sequence-based or energetical bias for incorporation of the mature miRNA into RISC? This can be investigated by mining publicly available expression data.

2.6 MicroRNA Sponges

MicroRNAs, which are important micromanagers of gene expression at the posttranscriptional level, are themselves subject to regulation both at the transcriptional and posttranscriptional levels. One of the most interesting posttranscriptional regulatory mechanisms that modulates intracellular miRNA abundance involves sponging by other transcripts that includes but is not limited to lncRNAs, circRNAs, and pseudogenes [38]. Although sponging, by definition, requires the presence of multiple miRNA recognition elements on the sponging transcript, there are numerous examples of sponging that involve a single binding site. Thus, the major challenge in this field is to distinguish among various potential miRNA-RNA interactions especially when there exists a single interaction between two transcripts. The best-characterized interaction includes miRNA-mediated silencing and/or translational inhibition of the target transcript. However, the miRNA binding to its target RNA can induce miRNA degradation as well. More importantly, miRNA-RNA interaction may simply sequester the miRNA away from its real target RNA without having any effect on the sponging RNA. The outcome can be predicted when multiple miRNA binding sites exist, especially on circRNAs. However, we do not yet have a sufficiently good grasp of sequence information and/or complementary rules for reliable predictions with bioinformatics methods only. Pressing questions in this field are:

38. What is the minimum number of miRNA binding sites for effective sponging?
39. What is the minimum amount of sequence complementarity required to induce the sponging mechanism without compromising the integrity of the miRNA and the bound RNA?
40. Which protein complexes are involved in the recognition of sponged miRNAs versus others?
41. What triggers the release of the sequestered or sponged miRNAs?
42. Are all miRNAs subject to sponging or is there a selective set of miRNAs targeted for sponging?
43. How do sponging, gene regulation, and miRNA regulation collaborate to form regulatory circuits?

Answering these (43) questions will significantly forward our understanding of miRNA-based regulation and will lead to large improvement for the computational prediction of functional miRNA interactions. All this will aid experimental studies and support novel questions such as stress adaptation.

2.7 Studying MicroRNAs and Extreme Stress Adaptation

The identification and characterization of both conserved and novel species-specific miRNAs in non–genome-sequenced organisms poses numerous challenges that can be circumvented via de novo transcriptome sequencing and assembly and/or through the development of complex bioinformatics pipelines [39]. To gain global insights into the unity of miRNA regulation in different environmental stresses will require expanding the current studies to more unique species [40]. Increasing the biodiversity of a comparative work to more uniquely adapted species is difficult when working on nonmodel organisms. This is due to the harsh environments these animals inhabit as well as due to the difficulty of performing downstream validation (for example siRNA and knock-out) experiments, in a scientific system that is centered on model organisms [41].

The field of miRNA research in nonmodel organisms is rapidly expanding and this can be helped with the adoption of more readily available next generation sequencing technologies. This expansion is paving the way for the generation of increasingly reliable and accurate biomarker profiles for different stresses that could be applied in biomedical contexts. Examples of this from extreme animal survival strategies include using the miRNAs implicated in protecting the brains of anoxia tolerant and for the development of stroke therapeutics and interventions, as well as the use of freeze tolerant responsive miRNAs to innovate organ and tissue preservation.

3 Conclusion

Much has been achieved in the past decades since the first account of miRNA regulation. MicroRNAs are now recognized as important regulators of gene expression and it has become clear that miRNA-based regulation and transcription factor-based gene regulation are intertwined and form complex regulatory networks. Unfortunately, the discourse about and involving miRNAs is suffering because some terms do not have the same meaning to all researchers. For example, what is a miRNA? Does this refer to the mature miRNA or a generic concept involving regulation? Does the term miRTron refer to a miRNA precisely excised during splicing or, more generally, to a miRNA encoded in an intron? Therefore, we want to add one more challenge here:

44. An ontology and a controlled vocabulary defining terms and actions in miRNomics is needed to ensure a common understanding of miRNA-based regulation.

It is clear that trying to investigate all miRNA-target interactions encoded in a genome is a futile endeavor. This is due to both miRNAs and their targets being under spatiotemporal control. Some miRNAs may be triggered only under special circumstances such as in response to stresses. This entails the need for computational prediction of miRNAs and their targets directly from genomes. Currently, eukaryotic genomes with millions of candidate hairpins are a formidable challenge. In order to alleviate the situation, we define (44) challenges in miRNomics which, when solved, will enhance the accuracy of computational predictions. We hope that this is an inspiration to researchers in the field of miRNomics to perhaps slightly modify their experiments to solve one of the challenges while pursuing their primary aims.

References

1. Ergin K, Çetinkaya R (2021) Regulation of microRNAs. In: Allmer J, Yousef M (eds) miRNomics: microRNA biology and computational analysis. Methods in molecular biology, vol 2257. Springer, New York
2. Olcum M, Tufekci KU, Genc S (2021) MicroRNAs in genetic etiology of human diseases. In: Allmer J, Yousef M (eds) miRNomics: microRNA biology and computational analysis. Methods in molecular biology, vol 2257. Springer, New York
3. Yildiz MT, Tutar L, Giritlioğlu NI, Bayram B, Tutar Y (2021) MicroRNAs and heat shock proteins in breast cancer biology. In: Allmer J, Yousef M (eds) miRNomics: microRNA biology and computational analysis. Methods in molecular biology, vol 2257. Springer, New York
4. Karagur ER, Akgun S, Akca H (2021) Computational and bioinformatics methods for microRNA gene prediction. In: Allmer J, Yousef M (eds) miRNomics: microRNA biology and computational analysis. Methods in molecular biology, vol 2257. Springer, New York
5. Nguyen LK, Dobrzyński M, Fey D, Kholodenko BN (2014) Polyubiquitin chain assembly and organization determine the dynamics of protein activation and degradation. Front Physiol 5:4. https://doi.org/10.3389/fphys.2014.00004
6. Lee RC, Feinbaum RL, Ambros V (1993) The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75:843–854. https://doi.org/10.1016/0092-8674(93)90529-Y
7. Uzuner E, Ulu GT, Gürler SB, Baran Y (2021) The role of MiRNA in cancer: pathogenesis, diagnosis, and treatment. In: Allmer J, Yousef M (eds) miRNomics: microRNA biology and computational analysis. Methods in molecular biology, vol 2257. Springer, New York
8. Hamid SM, Akgül B (2014) Master regulators of posttranscriptional gene expression are subject to regulation. Methods Mol Biol 1107:303–310
9. Yaylak B, Akgül B (2021) Experimental microRNA detection methods. In: Allmer J, Yousef M (eds) miRNomics: microRNA biology and computational analysis. Methods in molecular biology, vol 2257. Springer, New York
10. Erson-Bensan AE (2014) Introduction to MicroRNAs in biological systems. In: Yousef M, Allmer J (eds) miRNomics: MicroRNA biology and computational analysis, 1st edn. Humana Press, New York, pp 1–14
11. Mechtler P, Johnson S, Slabodkin H, Cohanim AB, Brodsky L, Kandel ES (2017) The evidence for a microRNA product of human DROSHA gene. RNA Biol 14:1508–1513. https://doi.org/10.1080/15476286.2017.1342934
12. Saçar Demirci MD, Yousef M, Allmer J (2019) Computational prediction of functional microRNA–mRNA interactions. Methods Mol Biol 1912:175–196

13. Yan K, Liu P, Wu C-A, Yang G-D, Xu R, Guo Q-H, Huang J-G, Zheng C-C (2012) Stress-induced alternative splicing provides a mechanism for the regulation of MicroRNA processing in *Arabidopsis thaliana*. Mol Cell 48:521–531. https://doi.org/10.1016/j.molcel.2012.08.032
14. Yang X, Zhang H, Li L (2012) Alternative mRNA processing increases the complexity of microRNA-based gene regulation in Arabidopsis: alternative splicing of microRNA binding sites. Plant J 70:421–431. https://doi.org/10.1111/j.1365-313X.2011.04882.x
15. Niu M, Tabari E, Ni P, Su Z (2018) Towards a map of cis-regulatory sequences in the human genome. Nucleic Acids Res 46:5395–5409. https://doi.org/10.1093/nar/gky338
16. Dweep H, Showe LC, Kossenkov AV (2021) Functional annotation of microRNAs using existing resources. In: Allmer J, Yousef M (eds) miRNomics: microRNA biology and computational analysis. Methods in molecular biology, vol 2257. Springer, New York
17. Winter J, Link S, Witzigmann D, Hildenbrand C, Previti C, Diederichs S (2013) Loop-miRs: active microRNAs generated from single-stranded loop regions. Nucleic Acids Res 41:5503–5512. https://doi.org/10.1093/nar/gkt251
18. Okamura K, Ladewig E, Zhou L, Lai EC (2013) Functional small RNAs are generated from select miRNA hairpin loops in flies and mammals. Genes Dev 27:778–792. https://doi.org/10.1101/gad.211698.112
19. Ghanbarian H, Yıldız MT, Tutar Y (2021) MicroRNA targeting. In: Allmer J, Yousef M (eds) miRNomics: microRNA biology and computational analysis. Methods in molecular biology, vol 2257. Springer, New York
20. Nachtigall PG, Bovolenta LA (2021) Computational detection of microRNA targets. In: Allmer J, Yousef M (eds) miRNomics: microRNA biology and computational analysis. Methods in molecular biology, vol 2257. Springer, New York
21. Akgül B, Erdoğan İ (2018) Intracytoplasmic re-localization of miRISC complexes. Front Genet 9:403. https://doi.org/10.3389/fgene.2018.00403
22. Cosacak M, Yiğit H, Kizil C, Akgül B (2018) Re-arrangements in the cytoplasmic distribution of small RNAs following the maternal-to-zygotic transition in drosophila embryos. Genes 9:82. https://doi.org/10.3390/genes9020082
23. Velandia-Huerto CA, Yazbeck AM, Schor J, Stadle PF (2021) Evolution and phylogeny of microRNAs—protocols, pitfalls, and problems. In: Allmer J, Yousef M (eds) miRNomics: microRNA biology and computational analysis. Methods in molecular biology, vol 2257. Springer, New York
24. Langenberger D, Bermudez-Santana C, Hertel J, Hoffmann S, Khaitovich P, Stadler PF (2009) Evidence for human microRNA-offset RNAs in small RNA sequencing data. Bioinformatics 25:2298–2301. https://doi.org/10.1093/bioinformatics/btp419
25. Shi W, Hendrix D, Levine M, Haley B (2009) A distinct class of small RNAs arises from pre-miRNA-proximal regions in a simple chordate. Nat Struct Mol Biol 16:183–189. https://doi.org/10.1038/nsmb.1536
26. Zhao J, Schnitzler GR, Iyer LK, Aronovitz MJ, Baur WE, Karas RH (2016) MicroRNA-offset RNA alters gene expression and cell proliferation. PLoS One 11:e0156772. https://doi.org/10.1371/journal.pone.0156772
27. Yousef M, Khalifa W, Acar E, Allmer J (2017) MicroRNA categorization using sequence motifs and k-mers. BMC Bioinformatics 18:170. https://doi.org/10.1186/s12859-017-1584-1
28. Odenthal L, Allmer J, Yousef M (2021) Ensemble classifiers for multiclass microRNA classification. In: Allmer J, Yousef M (eds) miRNomics: microRNA biology and computational analysis. Methods in molecular biology, vol 2257. Springer, New York
29. Dedeoğlu BG, Noyan S (2021) Experimental microRNA targeting validation. In: Allmer J, Yousef M (eds) miRNomics: microRNA biology and computational analysis. Methods in molecular biology, vol 2257. Springer, New York
30. Demirci MDS (2021) Computational detection of pre-microRNAs. In: Allmer J, Yousef M (eds) miRNomics: microRNA biology and computational analysis. Methods in molecular biology, vol 2257. Springer, New York
31. Yousef M, Parveen A, Kumar A (2021) Computational methods for predicting mature microRNAs. In: Allmer J, Yousef M (eds) miRNomics: microRNA biology and computational analysis. Methods in molecular biology, vol 2257. Springer, New York
32. de Amorim TS, Pedro DLF, Paschoal AR (2021) MicroRNA databases and tools. In: Allmer J, Yousef M (eds) miRNomics: microRNA biology and computational analysis. Methods in molecular biology, vol 2257. Springer, New York
33. Yousef M, Allmer J, Khalifa W (2016) Accurate plant MicroRNA prediction can be achieved

using sequence motif features. J Intell Learn Syst Appl 08:9–22. https://doi.org/10.4236/jilsa.2016.81002
34. Yousef M (2019) Hamming distance and K-mer features for classification of pre-cursor microRNAs from different species. In: Benavente-Peces C, Slama SB, Zafar B (eds) Proceedings of the 1st international conference on smart innovation, ergonomics and applied human factors (SEAHF). Springer International Publishing, Cham, pp 180–189
35. Yousef M, Khalifa W, Acar İE, Allmer J (2017) Distinguishing between MicroRNA targets from diverse species using sequence motifs and K-mers. In: Proceedings of the 10th international joint conference on biomedical engineering systems and technologies. SCITEPRESS - Science and Technology Publications, Setúbal, pp 133–139
36. Yousef M, Levy D, Allmer J (2018) Species categorization via MicroRNAs - based on 3'UTR target sites using sequence features: in: proceedings of the 11th international joint conference on biomedical engineering systems and technologies. SCITEPRESS - Science and Technology Publications, Funchal, Madeira, Portugal, pp 112–118
37. Tastan B, Tarakcioglu E, Birinci Y, Park Y, Genc S (2021) Role of exosomal microRNAs in cell-to-cell communication. In: Allmer J, Yousef M (eds) miRNomics: microRNA biology and computational analysis. Methods in molecular biology, vol 2257. Springer, New York
38. Alkan AH, Akgül B (2021) Endogenous miRNA sponges. In: Allmer J, Yousef M (eds) miRNomics: microRNA biology and computational analysis. Methods in molecular biology, vol 2257. Springer, New York
39. Zhang J, Hadj-Moussa H, Storey KB (2016) Current Progress of high-throughput MicroRNA differential expression analysis and random Forest gene selection for model and non-model systems: an R implementation. J Integr Bioinform 13:306–306
40. Hadj-Moussa H, Hawkins LJ, Storey KB (2021) Role of microRNAs in extreme animal survival strategies. In: Allmer J, Yousef M (eds) miRNomics: microRNA biology and computational analysis. Methods in molecular biology, vol 2257. Springer, New York
41. Hawkins LJ, Storey KB (2020) Advances and applications of environmental stress adaptation research. Comp Biochem Physiol A Mol Integr Physiol 240:110623. https://doi.org/10.1016/j.cbpa.2019.110623