Minireview

Prediction and validation of microRNAs and their targets

Isaac Bentwich*

Rosetta Genomics Ltd., 10 Plaut Street, Science Park, Rehovot 76706, Israel

Received 22 July 2005; revised 8 September 2005; accepted 13 September 2005

Available online 30 September 2005

Edited by Shou-Wei Ding

Abstract **MicroRNAs are short non-coding RNAs that inhibit translation of target genes by binding to their mRNAs, and have been shown to play a central role in gene regulation in health and disease. Sophisticated computer-based prediction approaches of microRNAs and of their targets, and effective biological validation techniques for validating these predictions, now play a central role in discovery of microRNAs and elucidating their functions.**

Keywords: MicroRNA; Target; Prediction; Algorithms; Validation

1. Introduction

MicroRNAs are short non-coding RNAs that suppress translation of target genes by binding to their mRNAs, and have been shown to play a central role in gene regulation in health and disease [1–4]. MicroRNAs' function as post-transcriptional inhibitors is based on a number of microRNAs, the function of which has been demonstrated biologically. Study of mutant phenotypes in worm led to the discovery of the first microRNAs, *lin-4* and *let-7*, which control developmental timing, by regulating translation of their respective targets [5,6]. Similar mutant studies, showed that *lsy-6* regulates left–right asymmetry in the nervous system in worm [2] and that *bantam* and *miR-14* control apoptosis in fly [7,8]. Ectopic expression of miR-181 led to hematopoietic differentiation [9], and pancreatic-islet-specific *miR-375* has been shown to regulate insulin secretion [4], in mouse.

Initially, most microRNAs were discovered by massive cloning and sequencing efforts [10–14], with informatics playing a limited role of verifying that the cloned sequences are part of a hairpin structure, typical of microRNA precursors [15]. It was apparent however, that these approaches are limited, especially in detecting low abundancy microRNAs, or ones which are tissue-specific, especially in tissues which are difficult to obtain and sequence. This led to development of increasingly sophisticated bioinformatic approaches for prediction of novel microRNAs, and sensitive biological validation techniques, needed to validate such predictions. Similarly, several bioinformatic approaches have evolved, which predict microRNA targets, and to a limited extent, methodologies which validate such target predictions.

Prediction of microRNAs in plants relies on principles similar to animal microRNA prediction, but takes into account features that are unique to plant microRNAs, such as longer and variable hairpin precursor length. Prediction of microRNA targets in plants is drastically simpler than in animals, as plant microRNAs typically bind their targets with near perfect complementarity. Several effective algorithms have been developed for prediction of microRNAs and their targets in plants, which are based on take these unique feature, and are not reviewed here.

2. MicroRNA prediction and validation

2.1. Principles of MicroRNA prediction

Bioinformatic prediction of microRNAs is based on machine learning techniques that use known microRNAs as a 'training-set', in order to 'train' a computer program, such that it is capable of identifying postulated novel microRNA sequences. Since microRNAs are derived from short ~60 nucleotide-long 'hairpin-shaped' precursors, a large group of such hairpin sequences randomly found in the genome, is typically used as a 'control group'. The vast majority of these randomly found hairpins are assumed not to be microRNA precursors.

The training-set is then studied for common *distinctive properties* of the known microRNAs, which set them apart from the control group of random hairpins. Once such distinctive properties are found, a computer algorithm is constructed, which scores sequences on their similarity to these distinctive properties, and accordingly, on their probability to be valid novel microRNAs. In general, distinctive properties include *structural features* such as hairpin length, hairpin-loop length, thermodynamic stability, base-pairing, bulge size and location, and distance of the microRNA from the loop of its hairpin precursor; and *sequence features* such as nucleotide content and location, sequence complexity, repeat elements and internal and inverted sequence repeats (see Fig. 1).

The resulting 'predictor' algorithm, is then iteratively checked and improved by training it on a *subset* of known microRNAs, and checking its scoring accuracy on a *separate* subset of known microRNAs, against a control group of random hairpins. The computer does not 'know' this second subset, and hence scores them as it would any unknown sequences. These scores may therefore be assessed for their sensitivity and specificity.

*Fax: +972 3 5480153.
E-mail address: bentwich@rosettagenomics.com (I. Bentwich).

0014-5793/$30.00 © 2005 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.febslet.2005.09.040

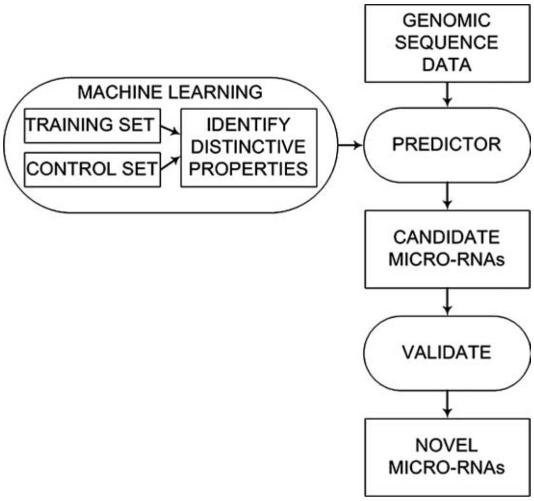

Fig. 1. *Machine learning prediction of microRNAs.* Machine learning algorithms are used to identify distinctive properties that differentiate between a training set of known microRNAs and a control set of genomic hairpins. Based on these, a predictor is used to identify candidate microRNAs from genomic sequence data. Finally, biological validation determines which of these candidates are valid novel microRNAs.

Most predictor algorithms depend on evolutionary conservation of microRNA sequences between different species. Such algorithms receive as input sequences that are homologous in two species, and use various approaches to detect microRNAs that are conserved in these two species. This approach allows filtering out many of the false-positive candidates, but is obviously limited to detecting conserved microRNAs.

Computerized identification of novel microRNAs is a difficult pattern-recognition challenge. No one property is sufficient for accurately detecting microRNAs, and in most cases rigid thresholds of property-values are also not sufficiently sensitive. Rather, it is the combination of multiple properties, with suitably different weighing of these different properties, that provides a more desirable accuracy. This is typically achieved by an iterative fine-tuning process, of and modifying the weight given to various distinctive properties, and testing its accuracy, as described above.

Finally, an attempt is made to validate expression of high-scoring predicted microRNAs in various tissues and/or cell cultures. This too is challenging, since failure to biologically validate the expression of a predicted microRNA, not necessarily implies that the bioinformatic prediction was incorrect: It may be that the microRNA is not expressed in the examined tissues, or is expressed only in certain cell-phases, or is expressed in low abundancy which escapes detection by the technique used. This latter cause is especially problematic with microRNAs, which are often very similar in sequence to one another. Expression of an abundant microRNA may therefore mask the expression of a rare one that is very similar in sequence, especially when using PCR amplification.

2.2. MicroRNA prediction algorithms

Several prominent computerized microRNA detection approaches have been developed and utilized. Lai et al. [16] identified 48 microRNA candidates in Drosophila, 24 of which were validated, using a computational microRNA detection program called *miRseeker*. This algorithm assesses the folding patterns of RNA sequences conserved between two *Drosophila* species using Mfold [17], in order to detect conserved hairpin structures having a nucleotide divergence characteristic of known microRNAs.

Lim et al. [3,18] identified 30 novel microRNAs in *C. elegans* and 38 novel human microRNAs using a sophisticated algorithm, called *MirScan*. This algorithm uses a different RNA folding algorithm, RNAFold (also known as Vienna Package) [19], to find hairpin structures in sequences that are evolutionarily conserved. Each conserved hairpin, is considered as a potential microRNA-precursor, and is then further assessed for the location of the microRNA within it. This is done by passing a 21-nucelotide window along the hairpin, and scoring each position for its similarity to known microRNAs. The algorithm is based on a training-set of 50 published microRNAs from *C. briggsae* and *C. elegans*. This approach successfully identified conserved microRNAs within the large number of conserved hairpins found in the genome (~35 000 hairpins conserved between *C. briggsae* and *C. elegans*; ~15 000 hairpins conserved between man, mouse and pufferfish). Grad et al. [20] used a similar approach to detect and validate 14 microRNAs in *C. elegans*.

Berezikov et al. [21] identified 16 novel human microRNAs, using a phylogenetic-based approach. Phylogenetic shadowing is a powerful genomic-conservation assessment technique, which determines the level of conservation of each nucleotide in a the assessed sequence [22]. Using this approach, Berezikov et al. found that nucleotides in the stem of microRNA hairpins precursors are significantly more conserved than in sequences flanking the hairpin, and in the hairpin's loops. They then used this distinctive property, in conjunction with other known properties of microRNAs, as described above, to identify novel microRNA candidates.

Recently, our group identified 89 novel human microRNAs, including 54 primate-specific microRNAs, using a novel integrated microRNA detection approach [23]. Unlike other techniques described above, this approach does not depend on sequence conservation, and was therefore capable of detecting a large number of microRNAs that seem to be unique to primates. Our goal was to create a broad 'funnel' which would allow us to scan as many candidate microRNAs as possible, and yet effectively 'zoom-in' on and validate the actual microRNA. We began by 'folding' non-coding regions of the entire human genome, using the RNAFold algorithm [19], yielding ~11 million hairpins. From these, we used our algorithm, *PalGrade*, to select a set of 5300 high-scoring candidates, which were subjected to microarray experiments using a microarray technique we developed [24]. 359 of these were shown to be expressed by microarray experiments, and were subjected to a novel sequencing technique we reported, yielding 89 novel validated human microRNA. This approach allowed detection of the largest cluster of microRNAs discovered to date, comprising 54 new predicted microRNAs, 43 of which we have biologically validated. Interestingly, while this cluster is located adjacent to three previously reported micro-RNAs, it is not conserved beyond primates, and so went undetected by other microRNA prediction algorithms, all of which depend on sequence conservation (see Fig. 2).

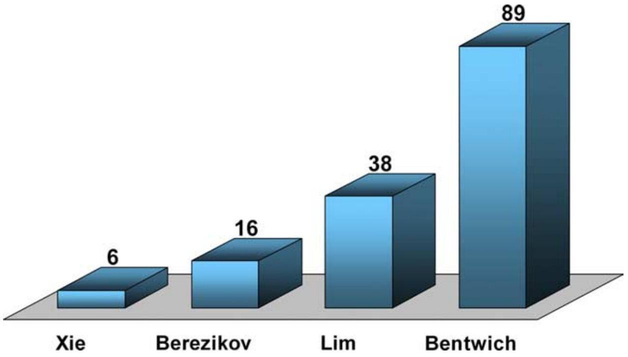

Fig. 2. *Predicted and validated microRNAs.* The number of microRNAs discovered informatically, and validated biologically, by different algorithmic approaches.

2.3. MicroRNA validation techniques

Validating expression of bioinformatically predicted microRNAs presents significant technical challenges. MicroRNAs are tiny in length (~22 nucleotides), are often expressed in low concentration, and in many cases are highly similar in sequence to other microRNAs. Traditional gene expression techniques are therefore not always suitable and sufficiently sensitive and specific, in identifying expression of rare, bioinformatically predicted microRNAs, especially ones that are similar in sequence to one another. Fortunately, the past several years have seen significant progress in development, implementation and refinement of several validation approaches, which adequately address these challenges.

Sequencing. Cloning and sequencing provide the highest level of validation for predicted microRNAs. Several cloning methodologies have been used, and are briefly described as follows.

Random cloning and sequencing of size-fractionated RNA, which has initially been the main approach for biological detection of microRNAs [13], was later used as in conjunction with informatic predictions, as an 'indirect' means for their validation [18]. According to this approach, informatic predictions are carried out in parallel to random cloning and sequencing, and the results are then compared. It does not allow validation of rare, bioinformatically predicted microRNAs.

Amplified partial sequencing is based on PCR amplification of adaptor-ligated cDNA clones using a primer with *partial coverage* of the predicted MIR sequence and an adaptor primer [18]. This method, in different variations, has been the main method for sequencing predicted microRNAs, which could not be found by random cloning and sequencing of microRNA enriched libraries. However, a major limitation of this approach, is that it allows actual independent sequencing of only a few nucleotides (typically 5–7 nucleotides), since the rest of the microRNA is 'fixated' by the primer. This is especially problematic in view of the significant sequence similarity between microRNAs.

Sequence-specific cloning and sequencing is a novel approach we have recently reported, which overcomes the abovementioned limitations and allows sequencing of full-length microRNAs. Based on the sequence of a predicted microRNA, a biotin-labeled oligonucleotide is designed and used to capture the homologous microRNA from a cDNA library enriched for small RNAs. The captured cDNA molecules are then cloned and sequenced.

Hybridization. Different hybridization assays provide important indirect validation for predicted microRNA sequences. Northern blots are successfully used to validate predicted microRNAs [25], and still are considered a golden standard. However, it is now clear that Northern blots are not always sufficiently sensitive and specific to validate expression of rare microRNAs [24]. Other hybridization essays include *RNase protection* [26], and a signal-amplifying *ribozyme* method [27].

High-throughput. Several high-throughput validation methods have been described, which may be used for validating predicted microRNAs. While methodologies listed below are technically based on hybridization, they allow highthroughput validation.

Membrane arrays using radioactive detection methods have been used as an inexpensive, effective method for detection of expression of microRNAs [28]. This method is probably less suited for sensitively monitoring expression of a large number of predicted microRNAs.

Microarrays have now been shown, by several independent groups, to be an effective, sensitive and specific means of high-throughput detection of expression microRNAs [24,29–35]. While microarrays are usually used for profiling expression of *known* genes, they may be used successfully to validate expression of postulated microRNAs, provided that the RNA is properly size-fractionated [23,24].

Bead-based profiling is a novel approach for profiling expression of microRNAs, which is significantly less expensive than traditional microarrays, is more flexible in its design, and which, based on initial data, seems to be sensitive and specific [36]. Capture probes that are complementary to the microRNAs of interest, are coupled to microscopic polystyrene beads that are impregnated with a dye (this is in contrast to traditional microarrays where capture probes are fixated on a glass slide). Multiple microRNAs (currently up to 100) may be tested simultaneously, by assigning a different dye to each microRNA. The beads are used to capture the microRNA from an amplified library, and flow cytometery is used to detect the type (dye color) and amount of microRNAs in the sample. *MirMASA* by Genaco, is another example of this approach [24].

3. MicroRNA target prediction and validation

Computational prediction of microRNA targets presents a significant challenge: (a) Unlike microRNA prediction, there does not exist a large enough group of known microRNA targets which can be used as a training set. (b) Validating microRNA target prediction is much more complex, no high throughput means available, only a small number of predictions have actually been validated [37].

Accordingly the approach taken with prediction microRNA targets is different from that of microRNA prediction, in that it is based on algorithms that are based on empiric evidence, rather than on machine learning algorithms. A set of studies, briefly reviewed below, have demonstrated different characteristics of the microRNA binding to its targets. These 'anecdotal' features serve as the basis for the basis of the various microRNA target prediction algorithms, which are reviewed below.

3.1. MicroRNA binding-site mechanics

Obligatory 5′-end 'seed', conserved, often flanked by adenosines. Elaborate single nucleotide mutation studies of several known microRNAs have been used to investigate the binding pattern of these microRNAs to their respective targets [38–41]. A clear conclusion from these different studies is the importance of the 5′ end segment of the microRNA, frequently referred to as its 'seed'. This seed, 6–8 nucleotides in length, has been shown to be critical, and at least in some cases sufficient, for microRNAs to suppress their targets. Its 5′ end is typically unpaired, or starts with a Uricil (i.e., its binding site ends with an Adenosine), and preferably does not contain G:U wobbles. A computerized analysis of conserved microRNA binding sites shows that the seed is often flanked by adenosines [42].

Compensatory 3′-end. While the 5′-end seed is clearly of central importance, there is significant evidence that the 3′-end of a microRNA's may compensate for insufficient base-pairing of its 5′ seed [39–41,43]. Several studies suggest that there are two types of microRNA binding-sites: *5′ dominant* sites (perfectly binding 5′ seed, with or without support of 3′ binding) and *3′ compensatory* sites (3′ binding compensates for imperfect 5′ seed binding) [39,41]. Many microRNA binding sites have bulges in their central or 3′-end sections, which in some cases have been demonstrated to be somewhat important for the binding [43]. The significance of these bulges is still not fully understood.

Multiple binding sites and their context. Mutation studies have been used to explore the role of multiple binding sites of microRNAs to the same mRNA target, and the context in which these sites are found. Such studies show that microRNAs function may depend on binding to these multiple binding-sites [38,43]. MicroRNAs have been shown to be capable of functioning in a collaborative, combinatorial manner: When any one of the two *let-7* binding-sites on its target *lin-41* is replaced by a miR221 binding-site, then both microRNAs are needed to inhibit this target [38]. There is currently contradicting evidence as to the significance of the context of microRNA binding-sites: modifying the 27 nucleotide sequence separating the two let-7 binding sites in *C. elegans* blocked the function of this microRNA [43], and yet a similar experiment by a different group in Zebrafish got contradictory results, and further showed that let-7 maintains functionality even when it binding sites are moved including into coding regions [38].

Target mRNA structure. Recent studies suggest that the 2-dimensional structure of microRNA binding-sites and their immediate mRNA vicinity must be sufficiently unstable, so as to be physically accessible to be bound by microRNA. These studies analyzed 2-dimensional structures of the mRNA of comprising known microRNA binding sites, observing frequently appearing patterns: a seed region of the binding site comprising a segment of at least three nucleotides, which is not bound (e.g., is not in a stem formation) [44]. A region surrounding a binding site that has low free energy, and does not contain stabilizing structures (e.g., stems), and does contains destabilizing structures [45].

3.2. MicroRNA target prediction algorithms

Stark et al. [46] used a target prediction algorithm to detect *Drosophila* microRNA targets, six of which were biologically validated. The algorithm is based on detecting complementary sequences of the 5′-end 8 nucleotide seed of the microRNA, that are evolutionarily conserved (preferably across more than two species), and uses MFold to calculate the thermodynamic stability of the binding. Multiple binding sites are required in order to achieve significant predictive power (targets having single binding sites would require biological validation). It does not filter out seeds containing G:U wobbles (which later turned out to be weaken the binding). The algorithm recovered and scored highly all previously known targets.

Rehmsmeier et al. [47] presented an improved RNA folding algorithm, called *RNAhybrid*, which provides improved free-energy assessment of hybridization of a short RNA to a long RNA (e.g., a microRNA to its target), and used it to predict *Drosophila* microRNA targets, thus overcoming a disadvantage (at the time) of the Mfold and RNAFold. They used the algorithm to seek microRNA targets in Drosophila, by forcing a match of a 6-nucleotide seed starting from the 2nd nucleotide from the 5′ end of the microRNA. The algorithm recovered some of the known targets, and suggested additional postulated targets.

Lewis et al. [37,42] used a sophisticated algorithm, called *TargetScan*, and its improved version *TargetScanS*, to identify mammalian microRNA targets, and were impressively successful in biologically validating 11 out of 15 predicted targets tested. TargetScan seeks a strong 7-nucleotide seed, starting from the 2nd nucleotide from the 5′ end, uses RNAFold to calculate the thermodynamic free-energy of the binding, and scores both single binding site and multiple binding-sites. TargetScanS is an improved algorithm that requires a shorter seed (6-nucleotides), which is preceded by an adenosine, and is located in a short 'island' of conservation, the surrounding of which are less conserved. It does not rely on free-energy calculation. The algorithm specifically recovers all known microRNA targets, and is estimated to have a 22–31% false positive rate (for targets conserved in mammals vs. conserved in mammals plus pufferfish, respectively).

Kiriakidou et al. [40] used an algorithm called *DIANA-MicroT*, which is trained to identify microRNA targets having a single binding-site, and have biologically validated 7 out of 7 such predicted human microRNA targets. This algorithm takes a different approach from those of other algorithms described above: (a) it focuses on single binding site targets, and (b) it seeks binding sites that have a typical central bulge, and require 3′ binding, beyond the obligatory 5′ seed. The algorithm successfully recovered all previously known prototypical *C. elegans* microRNA targets.

Enright and John et al. [48,49] used an algorithm called *miRanda* to identify microRNA targets in *Drosophila* and man. The algorithm uses a position-weighted matrix to emphasize binding of the microRNA's 5′-end segment more than its 3′-end segment, uses RNAFold for free-energy calculation, and relies on evolutionary conservation of the binding sites. The algorithm correctly recovered 9 out of 10 previously validated microRNA targets, and has an estimated 24–39% false positive rate (corresponding to 4–2 binding sites per microRNA, respectively).

Xie et al. [50] identified a large class of conserved, regulatory 8 nucleotide motifs, many of which are likely to be microRNA targets. While not a formal microRNA target algorithm, the authors report a large number of 3′UTR motifs, many of which are likely to be microRNA targets. The notion that these motifs are indeed microRNA binding sites is supported by the following striking differences between these motifs vs. other motifs: (a) strong directional bias with respect to DNA strand, (b) peak at 8-nucleotide length, and (c) end with an adenosine (i.e., an adenosine complementary to the 5′-end of a microRNA binding to that motif).

Krek and Grun et al. [51,52] used an advanced algorithm called *PicTar* to identify microRNA targets in vertabrates, *C. elegans* and *Drosophila*, and report extensive biological and informatic validation of its predictions. This algorithm is trained to identify both binding-sites targeted by a single microRNA, as well as those that are co-regulated by several microRNAs in a coordinated manner. It utilizes sophisticated pair-wise alignment to accurately filter for binding sites that are conserved across many species (7 *Drosophila* species; 8 vertebrates), and takes into account clustering and co-expression of microRNAs, and ontological information (matching microRNAs with potential targets that are expressed in the same cells and developmental phase). The authors have biologically validated 7 out of 13 microRNA targets predicted by the algorithm, and have further specifically recovered 8 out of 9 known targets with experimental in vivo evidence and 4 out of 10 targets having conservation of only the primary binding site. The algorithm is estimated to have ~30% false positive rate.

Using similar informatic methodologies to those described above, our group has been seeking targets for the 89 novel microRNAs which we have reported [23]. If we assume a signal to noise ratio of 2:1, in accord with previous studies [42,51], our initial data indicates that approximately 7250 genes are targeted by microRNAs, representing 49% of our gene set. These results are in accord with Lewis et al. [42] who estimated that 148 conserved microRNAs target 30% of all genes. Since genes are often regulated by multiple microRNAs [51], the additional 200 novel microRNAs that we have checked are expected to contribute a significant, although not linear, increase in the number of genes targeted by microRNAs.

3.3. MicroRNA target validation

Validating predictions of microRNA targets is much more challenging than validating predicted microRNAs. At present there does not exist a simple, high throughput method for biologically validating microRNA targets. Validation of microRNA target prediction algorithms therefore relies on a combination of informatic and biological validation strategies.

Informatic validation. MicroRNA target prediction algorithms may be informatically validated by a combination of two strategies. The first strategy is to evaluate an algorithm's success in correctly identifying known microRNA targets, i.e., targets that have already been biologically validated, and scoring them highly. The limitations of this strategy are twofold: (a) the number of validated targets is still small, and (b) the target prediction algorithms are to some extent based on these known targets.

The second strategy is to compare the number of postulated binding-sites that an algorithm finds for a real microRNA, with that found for a control group of artificially generated 'fictitious microRNAs'. In this approach, for each real microRNA tested, one or more artificial controls are created: artificial sequences in which the nucleotides of the microRNA have been shuffled, and which resemble the tested microRNA in various properties, such as frequency of appearance in the genome, dinucleotide composition, etc. It is then possible to compare the number of conserved binding-sites found for the real microRNA to those found for the artificial control sequences, and accordingly to calculate a signal to noise ratio, and an estimated false-positive rate. An algorithm is considered successful if (a) it has successfully identified and gave high scores to most of the known binding sites, and (b) has demonstrated a significant signal to noise ratio. The signal to noise ratio is useful in assessing the number of microRNA targets found in the genome in general, and for assessing the specificity of the algorithm's predictions. One should bear in mind, however, that this is a crude tool: the fact that a microRNA is found to have fewer binding sites than the 'noise' level, does necessarily mean that these predicted binding sites are not real.

Biologic validation. While the ultimate validation of predicted microRNA targets is biologic validation, the current biologic validation methodologies are still extremely labor intensive, and do not allow high-throughput target validation. The commonly used validation methodologies include: reporter-gene constructs [37,40,51,53,54], mutation studies [38–43,53], gene-silencing techniques [4,53,54], rescue assays [7], and classic genetic studies [2,5–8,54]. Overall, some 30 animal microRNA targets have been validated to date using these various techniques. The biological validation of predicted microRNA targets, albeit in small numbers, has nonetheless confirmed that various target prediction engines are indeed capable of identifying microRNA targets. Future development of high throughput target validation techniques will be necessary to raise the specificity and sensitivity of microRNA target prediction algorithms.

4. Conclusions

Prediction of microRNAs and their targets have come a long way in a few short years. From a secondary role, of checking that short cloned sequences reside within hairpins, to a leading role, of detecting hundreds of microRNAs that go undetected by biological means, and prediction of their potential targets. MicroRNA prediction algorithms, and validation techniques used in conjunction with these algorithms, are opening a door to an unfolding, previously unseen universe of gene regulation: From initial estimates in 2003 that no more than 33 human microRNAs remain to be detected [3], to two estimates earlier this year of 129 and 300 microRNAs remaining to be detected informatically (6 and 16 of which, respectively, were validated) [21,50], to a recent estimate of at least 680 microRNAs awaiting detection (89 of which were biologically validated) [23]. MicroRNA target prediction algorithms, while still in a maturation phase, have already established the notion that microRNAs regulate at least 30% of all human genes, possibly many more [42].

Prediction methods and validation techniques, of both microRNAs and their targets, are co-dependent. Sensitive biological validation techniques are key in fine-tuning informatic prediction algorithms. And yet, developing such biological techniques often depends on effective prediction algorithms. An integrated detection approach, which combines computational prediction together with high-throughput biological validation, has been most effective in discovery of microRNAs [23]. Arguably, development of a similarly integrated approach for detection and high throughput validation of microRNA targets could be instrumental.

Intriguing questions regarding microRNAs await further investigation: Why does the body need all these microRNAs? Why are they so heavily involved in differentiation and cancer?

What causes many of them to be so well conserved throughout evolution? I recently presented a theoretical model, which argues that microRNAs may be part of a genomic 'language' that participates in encoding cellular differentiation [55]. Effective methodologies for prediction and validation of microRNAs and their targets will be key in broadening our understanding of the roles and functions of this extraordinary group of genes.

Acknowledgements: I thank Eilon Sharon, Omer Barad, Ranit Aharonov, Ori Inbar and Shlomit Gilad, for input and help in preparing this mini-review, and the entire Rosetta Genomics team, whose creativity, dedication, support and friendship has made our microRNA discovery journey successful and enjoyable. I pray that the fruits of the scientific work on microRNAs be put to good use for the benefit of all being.

References

[1] Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116, 281–379.

[2] Johnston, R.J. and Hobert, O. (2003) A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. Nature 426, 845–849.

[3] Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B. and Bartel, D.P. (2003) Vertebrate microRNA genes. Science 299, 1540.

[4] Poy, M.N. et al. (2004) A pancreatic islet-specific microRNA regulates insulin secretion. Nature 432, 226–230.

[5] Lee, R.C., Feinbaum, R.L. and Ambros, V. (1993) The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75, 843–854.

[6] Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R. and Ruvkun, G. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. Nature 403, 901–906.

[7] Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B. and Cohen, S.M. (2003) Bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in *Drosophila*. Cell 113, 25–36.

[8] Xu, P., Vernooy, S.Y., Guo, M. and Hay, B.A. (2003) The *Drosophila* microRNA Mir-14 suppresses cell death and is required for normal fat metabolism. Curr. Biol. 13, 790–795.

[9] Chen, C.Z., Li, L., Lodish, H.F. and Bartel, D.P. (2004) MicroRNAs modulate hematopoietic lineage differentiation. Science 303, 83–86.

[10] Lau, N.C., Lim, L.P., Weinstein, E.G. and Bartel, D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science 294, 858–862.

[11] Lee, R.C. and Ambros, V. (2001) An extensive class of small RNAs in *Caenorhabditis elegans*. Science 294, 862–864.

[12] Lagos-Quintana, M., Rauhut, R., Lendeckel, W. and Tuschl, T. (2001) Identification of novel genes coding for small expressed RNAs. Science 294, 853–858.

[13] Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W. and Tuschl, T. (2002) Identification of tissue-specific MicroRNAs from mouse. Curr. Biol. 12, 735–739.

[14] Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A. and Tuschl, T. (2003) New microRNAs from mouse and human. RNA 9, 175–179.

[15] Ambros, V. et al. (2003) A uniform system for microRNA annotation. RNA 9, 277–279.

[16] Lai, E.C., Tomancak, P., Williams, R.W. and Rubin, G.M. (2003) Computational identification of *Drosophila* microRNA genes. Genome Biol. 4, R42.

[17] Mathews, D.H., Sabina, J., Zuker, M. and Turner, D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. 288, 911–940.

[18] Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B. and Bartel, D.P. (2003) The microRNAs of *Caenorhabditis elegans*. Genes Dev 17, 991–1008.

[19] Hofacker, I.L. (2003) Vienna RNA secondary structure server. Nucleic Acids Res. 31, 3429–3431.

[20] Grad, Y., Aach, J., Hayes, G.D., Reinhart, B.J., Church, G.M., Ruvkun, G. and Kim, J. (2003) Computational and experimental identification of *C. elegans* microRNAs. Mol. Cell 11, 1253–1263.

[21] Berezikov, E., Guryev, V., van de, B.J., Wienholds, E., Plasterk, R.H. and Cuppen, E. (2005) Phylogenetic shadowing and computational identification of human microRNA genes. Cell 120, 21–24.

[22] Boffelli, D., McAuliffe, J., Ovcharenko, D., Lewis, K.D., Ovcharenko, I., Pachter, L. and Rubin, E.M. (2003) Phylogenetic shadowing of primate sequences to find functional regions of the human genome. Science 299, 1391–1394.

[23] Bentwich, I. et al. (2005) Identification of hundreds of conserved and nonconserved human microRNAs. Nat. Genet. 37, 766–770.

[24] Barad, O. et al. (2004) MicroRNA expression detected by oligonucleotide microarrays: system establishment and expression profiling in human tissues. Genome Res. 14, 2486–2494.

[25] Sempere, L.F., Freemantle, S., Pitha-Rowe, I., Moss, E., Dmitrovsky, E. and Ambros, V. (2004) Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. Genome Biol. 5, R13.

[26] Lee, Y., Jeon, K., Lee, J.T., Kim, S. and Kim, V.N. (2002) MicroRNA maturation: stepwise processing and subcellular localization. EMBO J. 21, 4663–4670.

[27] Hartig, J.S., Grune, I., Najafi-Shoushtari, S.H. and Famulok, M. (2004) Sequence-specific detection of MicroRNAs by signal-amplifying ribozymes. J. Am. Chem. Soc. 126, 722–723.

[28] Krichevsky, A.M., King, K.S., Donahue, C.P., Khrapko, K. and Kosik, K.S. (2003) A microRNA array reveals extensive regulation of microRNAs during brain development. RNA 9, 1274–1281.

[29] Baskerville, S. and Bartel, D.P. (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. RNA 11, 241–247.

[30] Liu, C.G. et al. (2004) An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. PNAS 101, 9740–9744.

[31] Miska, E.A., Alvarez-Saavedra, E., Townsend, M., Yoshii, A., Sestan, N., Rakic, P., Constantine-Paton, M. and Horvitz, H.R. (2004) Microarray analysis of microRNA expression in the developing mammalian brain. Genome Biol. 5, R68.

[32] Thomson, J.M., Parker, J., Perou, C.M. and Hammond, S.M. (2004) A custom microarray platform for analysis of microRNA gene expression. Nature Meth. 1, 47–53.

[33] Nelson, P.T., Baldwin, D.A., Scearce, L.M., Oberholtzer, J.C., Tobias, J.W. and Mourelatos, Z. (2004) Microarray-based, high-throughput gene expression profiling of microRNAs. Nature Meth. 1, 155–161.

[34] Babak, T., Zhang, W., Morris, Q., Blencowe, B.J. and Hughes, T.R. (2004) Probing microRNAs with microarrays: tissue specificity and functional inference. RNA 10, 1813–1819.

[35] Lim, L.P. et al. (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature 433 (7027), 769–773.

[36] Lu, J. et al. (2005) MicroRNA expression profiles classify human cancers. Nature 435, 834–838.

[37] Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P. and Burge, C.B. (2003) Prediction of mammalian microRNA targets. Cell 115, 787–798.

[38] Kloosterman, W.P., Wienholds, E., Ketting, R.F. and Plasterk, R.H. (2004) Substrate requirements for let-7 function in the developing zebrafish embryo. Nucleic Acids Res. 32, 6284–6291.

[39] Brennecke, J., Stark, A., Russell, R.B. and Cohen, S.M. (2005) Principles of MicroRNA-target recognition. PLoS. Biol. 3, e85.

[40] Kiriakidou, M., Nelson, P.T., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z. and Hatzigeorgiou, A. (2004) A combined computational experimental approach predicts human microRNA targets. Genes Dev. 18, 1165–1178.

[41] Doench, J.G. and Sharp, P.A. (2004) Specificity of microRNA target selection in translational repression. Genes Dev. 18, 504–511.

[42] Lewis, B.P., Burge, C.B. and Bartel, D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are MicroRNA targets. Cell 120, 15–20.

[43] Vella, M.C., Reinert, K. and Slack, F.J. (2004) Architecture of a validated microRNA:target interaction. Chem. Biol. 11, 1619–1623.

[44] Robins, H., Li, Y. and Padgett, R.W. (2005) Incorporating structure to predict microRNA targets. PNAS 102, 4006–4009.

[45] Zhao, Y., Samal, E. and Srivastava, D. (2005) Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. Nature 436, 214–220.

[46] Stark, A., Brennecke, J., Russell, R.B. and Cohen, S.M. (2003) Identification of *Drosophila* MicroRNA targets. PLoS. Biol. 1, E60.

[47] Rehmsmeier, M., Steffen, P., Hochsmann, M. and Giegerich, R. (2004) Fast and effective prediction of microRNA/target duplexes. RNA 10, 1507–1517.

[48] Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C. and Marks, D.S. (2003) MicroRNA targets in *Drosophila*. Genome Biol 5, R1.

[49] John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C. and Marks, D.S. (2004) Human MicroRNA targets. PLoS. Biol. 2, e363.

[50] Xie, X., Lu, J., Kulbokas, E.J., Golub, T.R., Mootha, V., Lindblad-Toh, K., Lander, E.S. and Kellis, M. (2005) Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. Nature 434, 338–345.

[51] Krek, A. et al. (2005) Combinatorial microRNA target predictions. Nature Genet. 37, 495–500.

[52] Grun, D., Wang, Y., Langenberger, D., Gunsalus, K. and Rajewsky, N. (2005) microRNA Target predictions across seven *Drosophila* species and comparison to mammalian targets. PLoS. Comput. Biol. 1, e13.

[53] O'Donnell, K.A., Wentzel, E.A., Zeller, K.I., Dang, C.V. and Mendell, J.T. (2005) c-Myc-regulated microRNAs modulate E2F1 expression. Nature 435, 839–843.

[54] Johnson, S.M. et al. (2005) RAS Is Regulated by the let-7 MicroRNA Family. Cell 120, 635–647.

[55] Bentwich, I. (2005) A postulated role for microRNA in cellular differentiation. FASEB J. 19, 875–879.