DNA AND CELL BIOLOGY
Volume 26, Number 4, 2007

Pp. 265–272
DOI: 10.1089/dna.2006.0566

Algorithms for Mapping of mRNA Targets for MicroRNA

ILYA IOSHIKHES,[1,3] SASHWATI ROY,[2,3] and CHANDAN K. SEN[2,3]

ABSTRACT

MicroRNAs (miRNAs) are involved in human health and disease as endogenous suppressors of the translation of coding genes. At this early point of time in miRNA biology, it is important to identify specific cognate mRNA targets for miRNA. Investigation of the significance of miRNAs in disease processes relies on algorithms that hypothetically link specific miRNAs to their putative target genes. The development of such algorithms represents a hot area of research in biomedical informatics. Lack of biological data linking specific miRNAs to their respective mRNA targets represents the most serious limitation at this time. This article presents a concise review addressing the most popular concepts underlying state-of-the-art algorithms and principles aimed at target mapping for specific miRNAs. Strategies for improvement of the current bioinformatics tools and effective approaches for biological validation are discussed.

INTRODUCTION

An estimated 30% of human genes may be regulated, in part, by a novel posttranscriptional mechanism involving microRNAs (miRNAs) (Rajewsky, 2006). miRNAs are small RNAs that regulate gene expression in animals primarily through translational repression (Ambros, 2001, 2004; Alvarez-Garcia and Miska, 2005). The biomedical significance of miRNAs has been addressed in other articles in this special issue on miRNAs. At present, the precise mechanisms of interaction of the miRNA with mRNA, resulting in miRNA binding to specific target sites, are largely unclear. Because experimental data regarding specific functional targets for particular miRNA molecules are scarce, development of reliable bioinformatics algorithms and tools for target prediction poses significant challenge. Although several algorithms for computational mapping of miRNA targets have been published (John *et al.*, 2006; Kim *et al.*, 2006; Rajewsky, 2006; Yoon and De Micheli, 2006), the quality of the predictions often falls short of expectations. The lack of efficient and reliable algorithms that would predict specific targets for specific miRNAs may be viewed as a major bottleneck in miRNA research. Such algorithms would lead to new insights of the molecular mechanisms that are implicated in the posttranscriptional regulation of numerous genes of immense biomedical significance. Here we review the salient features of the existing algorithms and shed light on future potentials.

miRNAs are an abundant class of endogenous non-protein-coding small (19–25 nucleotides in length) RNAs, which negatively regulate gene expression at the posttranscriptional level in many biological processes (Fig. 1). miRNAs regulate gene expression by inducing cleavage or translational inhibition of their target mRNAs through base pairing to partially complementary sites (Pasquinelli *et al.*, 2005). Although there is substantial interest in the biological significance of miRNA, the number of experimentally identified miRNA targets is very limited. Most of our knowledge regarding these sites comes from indirect data based on predictive algorithms (Lewis *et al.*, 2005).

miRNA TARGET PREDICTION: BIOCHEMICAL PRINCIPLES

Although the biological importance of miRNAs has become quite clear, the specific biochemical rules of recognition and regulation of target genes by miRNA remain much less understood (Bentwich, 2005). One of such rules was formulated with the notion of seed match (nearly perfect complementarity of the miRNA 5′ end with 3′ end of the mRNA binding site), which transformed during a time (see *MiRanda* and *TargetScan* descriptions below for details). The seed length may vary from 6 to 8 nucleotides, starting from the 2nd miRNA nucleotide. The first nucleotide is often mismatched, or starts with the U

Departments of Biomedical Informatics[1] and Surgery,[2] Davis Heart & Lung Research Institute,[3] The Ohio State University Medical Center, Columbus, Ohio.

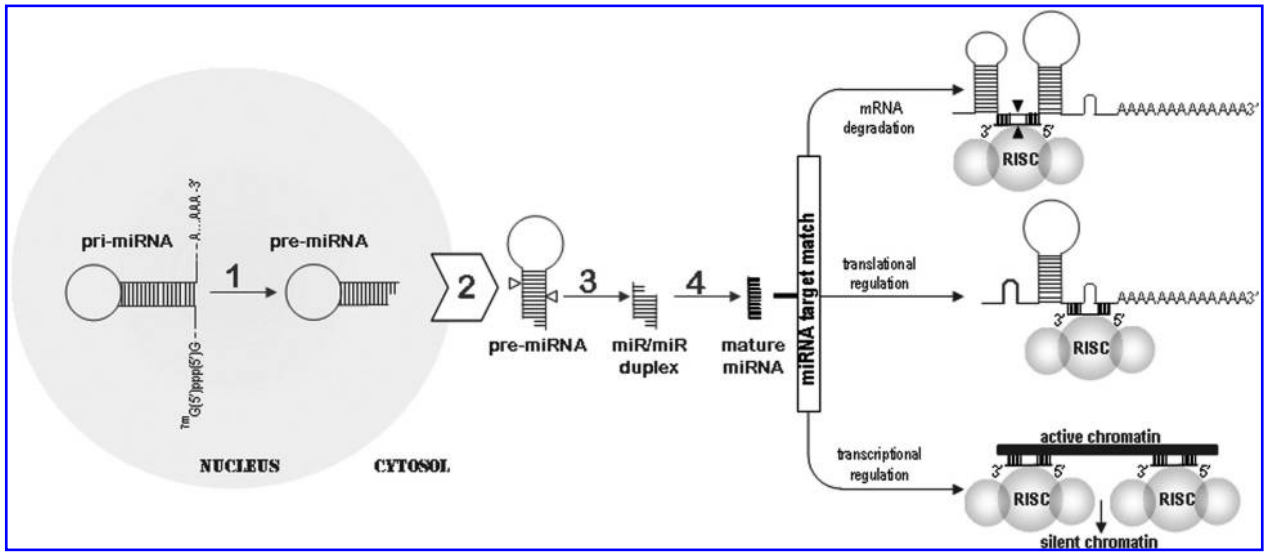

FIG. 1. Overview of the major mechanisms involved in the generation and function of miRNA. Primary transcripts of miRNA (pri-miRNA) are generated by polymerase II and possess a 5′ 7-methyl guanosine cap and are polyadenylated. Processing of pri-miRNA in the nucleus is mediated by a microprocessor complex (1) including Pasha and Drosha. Drosha is a RNase III endonuclease, which asymmetrically cleaves both strands of the hairpin stem at sites near the base of the primary stem loop thus releasing a 60- to 70-nucleotide *pre-miRNA* that has a 5′ phosphate and a 2-nucleotide 3′ overhang. Specific RNA cleavage by Drosha predetermines the mature miRNA sequence and provides the substrate for subsequent processing events. The pre-miRNAs are transported to the cytoplasm by Exportin-5 (2). Once in the cytosol, a second RNase III endonuclease, Dicer (3), cleaves the pre-miRNA. Dicer releases a 22-nucleotide mature double-stranded miRNA with 5′ phosphates and a 2-nucleotide 3′ overhangs. One strand of the miRNA duplex is subsequently incorporated into an effector complex termed RNA-induced silencing complex (4) or RISC that mediates target gene expression. For details, see Gregory *et al.* (2006) and Talmor-Neiman *et al.* (2006a, 2006b). miRNA-induced mRNA degradation is commonly seen in plants. Translational regulation by miRNA is common in animals. Regulation of transcription by miRNA is common in yeast and plants. This mechanism is possibly functional in animals as well.

(Bentwich, 2005). Specific seed parameters may also be flexibly defined by a user rather than being hardwired in any given algorithm, see e.g. *MovingTargets* (Burgler and Macdonald, 2005). The requirement of complementarity to the miRNA 3′ end is much looser, but it may compensate for imperfect complementarity at the 5′ end. Recently, Brennecke *et al.* (2005) systematically evaluated the minimal requirements for functional miRNA target duplexes *in vivo* and distinguished classes of target sites with different functional properties. Target sites have been grouped into the following 2 major categories:

1. 5′ dominant sites with sufficient complementarity to the miRNA 5′ end that may function with little or no support from pairing to the miRNA 3′ end, and
2. 3′ compensatory sites that have insufficient 5′ pairing and require strong 3′ pairing for function.

Examples and genome-wide statistical analyses maintain that both classes of sites are used in biologically relevant genes. In addition, many algorithms make a point on exploring the combinatorial nature of miRNA binding. Some miRNAs show preferential binding to coexisting close potential target sites rather than to single sites. On the other hand, silencing of certain genes may be governed by cooperative action of several different interacting miRNAs, rather than by only one of them. For instances of specific implementations, some algorithms, such as *MovingTargets*, may consider the number of miRNA targets as a specific predictive parameter (Burgler and Macdonald, 2005), while others, such as *TargetScan*, may implicitly derive it from energetic considerations giving higher score to multiple weak target sites than to fewer stronger sites (Lewis *et al.*, 2003). Recently, the combinatorial nature of miRNA regulation formed the basis for a high-throughput model, such as *GenMiR*. This model was based on combined mRNA and miRNA expression microarray data in the context of mRNA regulation by a variety of miRNAs (Huang *et al.*, 2005). Of note, the combinatorial nature of miRNA regulation has clear parallels with the regulation of mRNA by transcription factors (TF) combined in regulatory modules (Kel-Margoulis *et al.*, 2002; Remenyi *et al.*, 2004) with functional TF binding sites surrounded by multiple weaker motifs resembling rather degenerated functional binding motif (Zhang *et al.*, 2006). Whether such parallels imply hidden analogy of underlying molecular mechanisms of TF and miRNA binding remains to be established. Other miRNA mapping algorithms, such as *DianamicroT*, however, are strategically focused on the discovery of targets that have single functional miRNA, with alternative set of predictions (Kiriakidou *et al.*, 2004). The 2 sets may need to be combined to obtain a complete picture. Finally, many algorithms consider cross-species comparisons as a powerful tool to reliably identify mRNA targets for miRNA (see below).

COMPUTATIONAL STATE OF THE ART

Existing algorithms for identifying miRNA targets

Algorithm and software development for the computational recognition of the miRNA targets represents one of the most rapidly expanding areas of biomedical informatics. Several algorithms are available to date (Bentwich, 2005; Brown and Sanseau, 2005) and many more are being developed. It is difficult to recall another area of bioinformatics in which every new algorithm gets immediate attention, recognition, and publication in the leading journals, sometimes even without mandatory experimental verification of the results. Although the general miRNA paradigm is well described, there is a clear paucity of experimental data addressing specific miRNAs and their corresponding mRNA targets. Because most of the information relevant for the development of algorithm comes from indirect experimental data (Lewis *et al*., 2005), direct identification of mRNA targets for miRNA on the basis of experimental biology would represent a valuable resource for the design of reliable algorithms. At present, the development of such algorithm *per se* is at its starting phase (Burgler and Macdonald, 2005; Huang *et al*., 2005; Krek *et al.*, 2005; Robins *et al.*, 2005). The development of algorithms, more specifically considering biophysical interaction between the miRNA and genomic/cDNA sequence, represents the next necessary step in this area. Several of the most popular algorithms for miRNA target prediction are described below, followed by an attempt to outline the possible future prospects in the area.

MiRanda—*identifying miRNA targets in human and Drosophila genomic sequences [http://www .microrna.org/]*

The *miRanda* algorithm (John *et al.*, 2004; Hsu *et al.*, 2006a, 2006b) consists of 2 basic steps, supplemented by statistical and phylogenetic estimations to identify potential targets. At the beginning, *miRanda* reads RNA sequences (such as miRNA) from **file1** and genomic DNA/RNA sequences from **file2**. Any variety of relevant sequences in FASTA format is possible as the input. One or more miRNA sequences from **file1** are scanned against all sequences in **file2,** and potential target sites are reported. Potential target sites are identified using a 2-step strategy. First, a dynamic programming local alignment [essentially a modified Smith-Waterman alignment (Smith and Waterman, 1981)] is carried out between the query miRNA sequence and the potential genomic target sequence. The scores of the alignment are based on sequence complementarity and not on sequence identity. In other words, the algorithm looks for **A:U** and **G:C** matches instead of **A:A** and **G:G**. The **G:U** wobble pair is also permitted, but generally scores less than perfect matches (Fig. 2). Nonspecific penalties for mismatches, gap opening, and gap extension complement the initial scoring scheme. Importantly, miRNA positions 2–11 (modified to 2–8 according to later experimental findings) of the miRNA are in perfect complementary match with the reference sequence, as was confirmed experimentally. In *miRanda* a scaling factor doubles the initial scores for those positions. The value of the scaling factor at each position and the gap penalties are potentially adjustable parameters subject to optimization as more experimental information becomes available.

At the second stage of analysis, the Vienna package for RNA folding (Hofacker, 2003) is employed to estimate the thermodynamic properties of a predicted duplex. The miRNA and potential genomic target sequence are artificially joined into a single sequence by a linker of 8 artificial X bases that cannot base pair. The thermodynamic parameters related to an optimal (with minimum energy) folding of this artificial RNA allow scoring potential target sites by their folding energies, in addition to those by alignment scores. However, the folding procedure is computationally intensive and time consuming. miRNAs

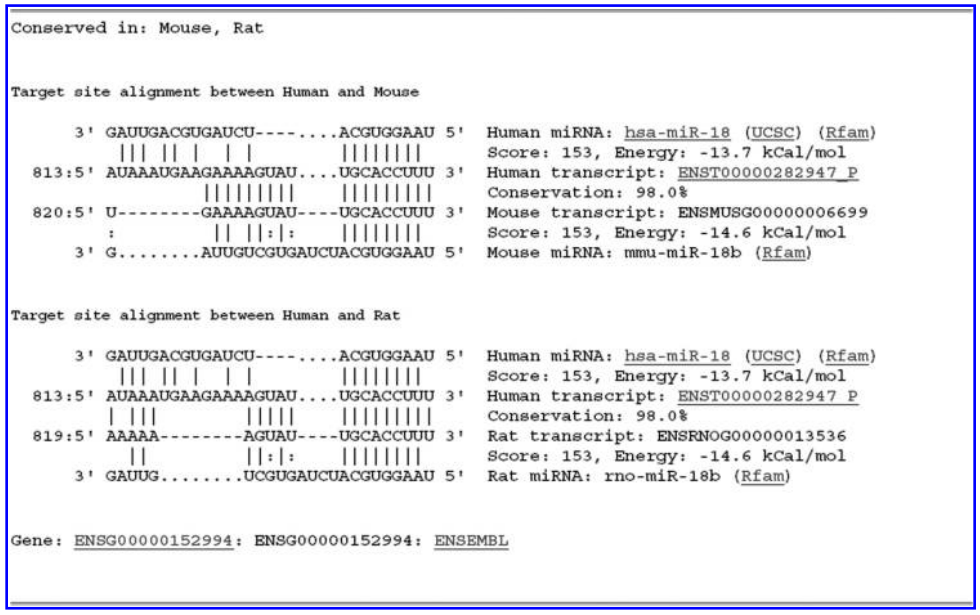

FIG. 2. miRNA-mRNA alignment by *miRanda.*

conserved across several species get higher priority in the search. The efficiency of the algorithm is evaluated by comparison of the number of sites predicted using natural sequences with those by shuffled sequences.

TargetScan *prediction of miRNA targets in mammalian genomes [http://www.TargetScan.org/]*

TargetScan has been named *TargetScanS* in recent versions (Lewis and Redrup, 2005; Lewis *et al*., 2005; Hsu *et al*., 2006b). This popular program (Lewis *et al*., 2003) combines thermodynamics-based modeling of miRNA/mRNA duplex interactions with comparative sequence analysis to predict miRNA targets conserved across multiple genomes. Given a miRNA conserved in multiple organisms and a set of orthologous 3′ untranslated region (UTR) sequences from these organisms, *TargetScan* searches the UTR for segments of perfect Watson-Crick complementarity to bases 2–8 of the miRNA numbered from its 5′ end (the so-called seed match reduced from 7 to 6 bp in a later version). The program then extends each seed match as far as possible in each direction, allowing **G:U** pairs, but stopping at mismatches, and uses the RNA-fold program of the Vienna package (Hofacker, 2003) to complete the alignment. The program produces the scores according to the site's binding energy and searches for conserved regions in other species. This approach explicitly relies on targets conserved across species for its predictions.

Diana-microT—*predicting human miRNA targets [http://diana.pcbi.upenn.edu/cgi-bin/micro_t.cgi]*

This program initially identifies putative miRNA/mRNA interactions based on binding energies between 2 RNAs paired imperfectly (Kiriakidou *et al*., 2004). For this, it uses a window of 38 nucleotides ''sliding'' over the mRNA sequence and calculates the minimum binding energy between the miRNAs and sequences in the human 3′ UTR database. The window is chosen based on the length of each miRNA (21–23 nucleotides) plus additional bases for loops and bulges. The same type of matches as in *miRanda* is allowed. In order to speed up the program, for every putative RNA duplex, the program first determines whether there are at least 3 consecutive canonical pairings between the 2 sequences. If so, a modified dynamic programming algorithm is applied, calculating a pairing between the 2 sequences, that yields the minimum free energy. The program uses 2 nucleotides at once to calculate the free energy between the 2 nucleotides of the miRNA paired with the 2 from the putative mRNA target. During this phase of the program, mismatches are allowed. The next step involves parsing the output of the dynamic programming and eliminating overlapping windows. The user defines an energy cut-off for the windows. Finally, the results are filtered according to certain experimental features. Our experience with *Diana-microT* shows that it frequently misses true targets, perhaps because of its focus on single targets as opposed to target clusters.

PicTar—*combinatorial miRNA target prediction [http://pictar.bio.nyu.edu/]*

This method is designed to reveal common targets of several miRNA, but is also applicable to map putative targets for specific miRNA. The approach heavily relies on cross-species comparisons through multiple sequence alignments of orthologous 3′ UTRs, where potential miRNA targets are located. It has been noted that the ability to confidently detect miRNA binding sites is tightly linked to the power of cross-species comparisons. Evolutionarily conserved sites are likely to be functional, and sequence conservation therefore serves as a filter to define likely target sites (Lall *et al.*, 2006). The *PicTar* algorithm implements probabilistic model estimating the likelihood for a given sequence segment to function as a binding site for a single miRNA or combination thereof and using general 3′ UTR sequences as a background. Similar to the aforementioned algorithms, *PicTar* considers a perfect match between 7-nucleotide mature miRNA ''nucleus'' or ''seed'' starting at the miRNA 5′ end and corresponding mRNA target as a sufficient (although not necessary) condition for a binding to a conserved 3′ UTR mRNA sequence segment. In accordance with the Brennecke rules (Brennecke *et al*., 2005), miRNA hits may also allow mismatches in the ''seed,'' which should be compensated by complementary binding of the miRNA's 3′ end. Imperfect sites are subject to passing a threshold of miRNA/mRNA duplex binding energy. The program first looks for the conserved 3′ UTR segments containing minimal number of perfect and imperfect matches for a given miRNA set specified by the user, and then derives a Hidden Markov Model (HMM)–based score for a given UTR to be targeted by the given miRNA. Although this algorithm may be considered as the next-generation algorithm comparing to those previously described (taking into account combinatorial and competitive binding by multiple miRNA), it nevertheless relies on the same type of the sequence alignment and energetic considerations as the previous algorithms. Remarkably, the latest *PicTar* (Lall *et al*., 2006) foresees several ways of further algorithm improvement based on novel experimental data, including improved scoring for sites with imperfect seed and accounting for the miRNA secondary structure, as well as incorporation of evolutionary considerations and miRNA and mRNA expression levels.

Some sort of dynamic programming alignment is involved in all these programs, usually combined with a computationally time-consuming RNA folding procedure or other estimates of the RNA duplex binding energy. Remarkably, program emulations available online for *miRanda* and *TargetScan* simply search for putative targets among precalculated data, instead of performing real interactive calculations. This makes it much harder to predict targets for novel miRNA sequences. Experimental constraints, including mandatory matches or bulges in certain positions, are usually added to an algorithm outside of the major mathematical frame. An important shortcoming is that most of the programs essentially search only for sites conserved across several species. Although this may lead to important discoveries (Xie *et al.*, 2005), it leaves room to miss the targets different between species.

DATABASES OF mRNA TARGETS FOR miRNA

Current data on miRNA targets, either experimentally verified or computationally predicted, are enlisted in several publicly available databases (Griffiths-Jones, 2004, 2006; Griffiths-Jones *et al.*, 2006; Hsu *et al*., 2006a; Megraw *et al.*, 2006; Sethupathy

et al., 2006). These databases have developed graphic-user interface and a variety of search options that enable user flexibility and rapid data retrieval. While the Sanger Center miRBase [better known as an miRNA Registry (Griffiths-Jones, 2004, 2006; Griffiths-Jones *et al.*, 2006; Hsu *et al.*, 2006a, 2006b; Huang *et al.*, 2006; Shahi *et al.*, 2006)] provides the largest and most comprehensive set of computationally predicted targets obtained by the existing software (primarily *miRanda* and *TargetScan*), the *TarBase* (Megraw *et al.*, 2006; Sethupathy *et al.*, 2006) is much smaller in size, yet contains experimentally verified data. miRNAMap (http://mirnamap.mbc.nctu.edu.tw) attempts to provide additional verification of the miRBase data and to use the verified data for more reliable computational predictions (Hsu *et al.*, 2006a). miRNAMap stores the known miRNA genes, the putative miRNA genes, the known miRNA cognate targets, and the putative miRNA targets.

EXISTING ALGORITHMS AND POSSIBLE FUTURE DIRECTIONS: A CRITICAL ANALYSIS

Currently existing algorithms for the miRNA target prediction have certain limitations in terms of both conceptual mathematical design and relevance to biological experimental data. Mathematically, the existing algorithms typically use combination of miRNA/mRNA sequence alignment with energetic miRNA/mRNA binding considerations. Conceptually, a typical sequence alignment procedure uses a modified Smith-Waterman alignment (Smith and Waterman, 1981) that roots in evolutionary considerations. A typical matrix used by the Smith-Waterman algorithm for DNA/DNA alignment should be modified according to the rules of nucleotide complementarity for the miRNA/mRNA alignment. In that case G/C and A/T(U) matches would be considered as perfect, unlike A/A, C/C, G/G, and T/T in regular case. G/U wobble pair with a score lower than for the perfect matches is usually also considered. Initially, the *miRanda* algorithm assigns score of + 5 for G/C and A/T pairs, + 2 for G/U wobble pair, and − 3 for mismatch pair, with gap-opening and gap-elongation parameters 8.0 and 2.0, respectively. The matrix is used to perform a Smith-Waterman alignment of miRNA versus mRNA (Fig. 3). This alignment matrix is crucial for the performance of the program.

As mentioned by Smith and Waterman themselves in their original paper (Smith and Waterman, 1981), the alignment is part of more general problem to measure the minimum number of events required to convert one sequence into another. That idea is applicable to the analysis of evolutionarily related sequences (genes, genomes, and proteins), yet its straightforward application to the evolutionary unrelated sequences (like miRNA/mRNA) is questionable. In any case, the Smith and Waterman approach should be further combined with some RNA folding algorithm to bring into consideration energetic parameters of the miRNA/mRNA binding. The latter step is time consuming. A proposal for such combination leaves room for criticism as well. The above-cited approach aims to take advantage on combining already existing algorithms instead of developing novel algorithm specifically designed to address a given problem. Successful development of novel algorithms initially often relies on ideas and mathematical apparatus from other scientific areas. This is necessary at the nascent stage when a new subspecialty is unfolding. The evolution of algorithm development for mapping the miRNA targets depends on vital resources, such as direct experimental data from experiments addressing miRNA/mRNA interactions.

So far the modified Smith-Waterman sequence alignment (Smith and Waterman, 1981) has been a necessary element for any miRNA target-mapping algorithm. The alternative is to simply rely on a perfect match of the seed 5′ miRNA segment with corresponding mRNA target. As discussed, the Smith-Waterman alignment was originally designed for comparing evolutionarily related sequences. miRNAs and their mRNA targets do not seem to fall in that category. The controversy was resolved by modifying the scoring matrix of the Smith-Waterman alignment. Although the process of the matrix adjustment probably will continue, hopefully some genuine approach for miRNA/mRNA sequence comparison will emerge. Usage of the computationally expensive Vienna RNA folding software package (Hofacker, 2003; Bernhart *et al.*, 2006; Obernosterer *et al.*, 2006; Tranzer and Stadler, 2006) and the likes should be replaced by more cost-efficient procedures allowing necessary direct energetic estimates of the miRNA/mRNA binding. An interesting attempt in that direction was already made at a relatively early stage of algorithm development. The *RNAhybrid* algorithm (Rehmsmeier *et al.*, 2004) finds most energetically favorable sites of miRNA/mRNA hybridization. *RNAhybrid* utilizes an original dynamic programming approach extending classical RNA secondary structure prediction algorithm (Zuker and Stiegler, 1981) to 2 sequences, instead of mimicking the miRNA/mRNA binding by folding a single RNA molecule. The algorithm eliminates an artifactual base pairing inside a single RNA strand. The suggested procedure also significantly increased algorithm speed, enabling search of large genomic databases in a reasonable time feasible.

miTarget was based on a different approach (Kim *et al.*, 2006). This algorithm implements miRNA target gene predictions using classification by Support Vector Machine. *miTarget* utilizes multiple structural, thermodynamic, and position-based

	C	G	A	T	U
C	-3	+5	-3	-3	-3
G	+5	-3	-3	+2	+2
A	-3	-3	-3	+5	+5
T	-3	+2	+5	-3	-3
U	-3	+2	+5	-3	-3

FIG. 3. Scoring matrix used by the *miRanda* algorithm.

features reflecting the mechanism of miRNA binding. Reliability of predictions offered by the algorithm depends on the availability of reliable experimental targets, as well as on the correct sequences that are able to serve as reliable negative control. Both issues remain to be resolved satisfactorily restricting the efficiency of *miTarget*.

New experimental data, particularly those addressing miRNA/mRNA interactions, represent the lifeline of any new algorithm seeking to successfully predict mRNA targets for given miRNA. The current principles and elements underlying the development of such algorithm need to be united in a single mathematical framework instead of being eclectically combined. New experimental data, in turn, would require optimization of specific mathematical parameters that define the algorithms. Some of such specific parameters and their significance are listed below:

1. *Scaling parameters.* There are different levels of stringency for the determination of miRNA/mRNA matches. Until recently, it was accepted that some miRNA nucleotides (miRNA ''seed'') perfectly match the mRNA sequence, whereas in other cases mismatches are possible. This consideration is utilized by algorithms such as *TargetScan, PicTar*, and the latest version of *miRanda* (http://www.microrna.org/mammalian/index.html). In *miRanda*, the above-said consideration is taken into account as a scaling factor, with scores in these positions multiplied by 2 versus the initial scores. The recently formulated Brennecke rules (Brennecke *et al*., 2005) are consistent with this notion for the 5′ dominant sites. However, another broad class of the 3′ dominant sites does not obey the seed-match principle mentioned here. Therefore, the scaling factors may vary depending on specific positions in the miRNA sequence. In the framework of the new algorithms, position-dependent scaling factors for sequence matches and gap penalties (see below) may be introduced to modify the scores obtained by the initial alignment matrix. The values of the scaling factors should be calculated upon the algorithm optimization.
2. *Gap parameters.* The gap parameters (Smith and Waterman, 1981) set at the beginning also may vary, for both the gap opening and the extension. Scaling, as described above, is also possible.
3. *Mismatch scores.* Smith-Waterman alignment (Smith and Waterman, 1981) usually includes a certain penalty for mismatches, as in *miRanda*. A score for most of the mismatches is set quite arbitrarily. The new algorithms should have an option to optimize the score. In addition, penalty for each mismatch may be different both position wise and base wise (i.e., being dependent on specific nucleotides and positions thereof), as it is also for different matches.
4. *Conserved targets. MiRanda, PicTar, TargetScan*, and other algorithms heavily rely on conservation of putative miRNA targets across genomes of different species (Bentwich, 2005; Bentwich *et al.*, 2005). Novel algorithms should explore that property as well in a consistent manner, but as an option only. That would allow users search for conserved or species-specific targets.

BIOLOGIC VALIDATION OF PREDICTED TARGETS

Only a small number of predicted targets have been experimentally validated. All miRNA target–finder algorithms return lists of candidate target genes. How valid is that output in a biological setting? One of the most common approaches of biological validation of predicted target involves tissue-culture assays using *reporter target gene constructs* fused to target sequences (Taganov *et al.*, 2006; Tsuchiya *et al.*, 2006; Wang *et al.*, 2006). Cells containing the specific mRNA in question are transfected with a reporter construct representing the predicted target gene (Lewis *et al*., 2003). If the predicted match is biologically valid, the miRNA will down-regulate the reporter. If the prediction is not accurate, expression of the reporter will not be affected by the miRNA. A more rigorous test asks whether *point mutation* of *target sites* in such reporters increases their activity, which might indicate relief from endogenous miRNA-mediated down-regulation (Meister *et al.*, 2004). Transcriptome analysis has proven to be a productive approach to determine mRNA targets for any given miRNA. Time course mRNA microarray experiments may reliably identify down-regulated genes in response to overexpression of specific miRNA (Wang and Wang, 2006). The approach may miss some miRNA targets that are principally down-regulated at the protein level. However, the high-throughput capacity of the assay makes it an effective tool to rapidly identify a large number of promising miRNA targets. Finally, *loss-* and *gain-of-function miRNA genetics* has the clear potential of being critical in evaluating the biological relevance of thousands of target genes predicted by bioinformatic studies, and for evaluating the degree to which miRNA-mediated regulation of any ''validated'' target functionally matters to the animal or plant. This will probably necessitate detailed studies of a broad range of biological processes, and potentially the analysis of multiple miRNA-mutant animals, or ones in which miRNA activity has been inhibited by chemical inhibitors (e.g., 2′ O-methylated oligonucleotides) (Hutvagner *et al.*, 2004).

While biological validation of predicted target is critical, failure to biologically validate the expression of a predicted miRNA does not necessarily imply that the bioinformatic prediction was incorrect. It is possible that the miRNA is not expressed in the examined tissues, or that the miRNA is expressed only in certain phase of cell cycle, or that the miRNA is expressed in low abundance, which escapes detection by the technique used. This latter cause is especially problematic for miRNA that shares a high degree of sequence homology with another miRNA. Expression of an abundant miRNA may therefore mask the expression of a rare one that is very similar in sequence, especially when using polymerase chain reaction amplification.

ACKNOWLEDGMENTS

The authors thank Professors Terry Elton and Michael Ostrowski for helpful discussion and Abhilash Mohan for assistance with Figure 3. This work was supported in part by NIH GM069589, GM077185 and HL073087 to CKS.

REFERENCES

ALVAREZ-GARCIA, I., and MISKA, E.A. (2005). MicroRNA functions in animal development and human disease. Development **132,** 4653–4662.

AMBROS, V. (2001). MicroRNAs: tiny regulators with great potential. Cell **107,** 823–826.

AMBROS, V. (2004). The functions of animal microRNAs. Nature **431,** 350–355.

BENTWICH, I. (2005). Prediction and validation of microRNAs and their targets. FEBS Lett **579,** 5904–5910.

BENTWICH, I., AVNIEL, A., KAROV, Y., AHARONOV, R., GILAD, S., BARAD, O., BARZILAI, A., EINAT, P., EINAV, U., MEIRI, E., SHARON, E., SPECTOR, Y., and BENTWICH, Z. (2005). Identification of hundreds of conserved and nonconserved human microRNAs. Nat Genet **37,** 766–770.

BERNHART, S.H., TAFER, H., MUCKSTEIN, U., FLAMM, C., STADLER, P.F., and HOFACKER, I.L. (2006). Partition function and base pairing probabilities of RNA heterodimers. Algorithms Mol Biol **1,** 3.

BRENNECKE, J., STARK, A., RUSSELL, R.B., and COHEN, S.M. (2005). Principles of microRNA-target recognition. PLoS Biol **3,** e85.

BROWN, J.R., and SANSEAU, P. (2005). A computational view of microRNAs and their targets. Drug Discov Today **10,** 595–601.

BURGLER, C., and MACDONALD, P.M. (2005). Prediction and verification of microRNA targets by MovingTargets, a highly adaptable prediction method. BMC Genomics **6,** 88.

GREGORY, R.I., CHENDRIMADA, T.P., and SHIEKHATTAR, R. (2006). MicroRNA biogenesis: isolation and characterization of the microprocessor complex. Methods Mol Biol **342,** 33–47.

GRIFFITHS-JONES, S. (2004). The microRNA Registry. Nucleic Acids Res **32,** D109–D111.

GRIFFITHS-JONES, S. (2006). miRBase: the microRNA sequence database. Methods Mol Biol **342,** 129–138.

GRIFFITHS-JONES, S., GROCOCK, R.J., VAN DONGEN, S., BATEMAN, A., and ENRIGHT, A.J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res **34,** D140–D144.

HOFACKER, I.L. (2003). Vienna RNA secondary structure server. Nucleic Acids Res **31,** 3429–3431.

HSU, P.W., HUANG, H.D., HSU, S.D., LIN, L.Z., TSOU, A.P., TSENG, C.P., STADLER, P.F., WASHIETL, S., and HOFACKER, I.L. (2006a) miRNAMap: genomic maps of microRNA genes and their target genes in mammalian genomes. Nucleic Acids Res **34,** D135–D139.

HSU, P.W., LIN, L.Z., HSU, S.D., HSU, J.B., and HUANG, H.D. (2006b) ViTa: prediction of host microRNAs targets on viruses. Nucleic Acids Res. **35,** D381–D385.

HUANG, H.Y., CHIEN, C.H., JEN, K.H., and HUANG, H.D. (2006). RegRNA: an integrated web server for identifying regulatory RNA motifs and elements. Nucleic Acids Res **34,** W429–W434.

HUANG, J.C., MORRIS, Q.D., and FREY, B.J. (2005). A computational high-throughput method for detecting miRNA targets PSI-Group Technical Report TR2005-026. www.psi.toronto.edu. University of Toronto.

HUTVAGNER, G., SIMARD, M.J., MELLO, C.C., and ZAMORE, P.D. (2004). Sequence-specific inhibition of small RNA function. PLoS Biol **2,** E98.

JOHN, B., ENRIGHT, A.J., ARAVIN, A., TUSCHL, T., SANDER, C., and MARKS, D.S. (2004). Human MicroRNA targets. PLoS Biol **2,** e363.

JOHN, B., SANDER, C., and MARKS, D.S. (2006). Prediction of human microRNA targets. Methods Mol Biol **342,** 101–113.

KEL-MARGOULIS, O.V., KEL, A.E., REUTER, I., DEINEKO, I.V., and WINGENDER, E. (2002). TRANSCompel: a database on composite regulatory elements in eukaryotic genes. Nucleic Acids Res **30,** 332–334.

KIM, S.K., NAM, J.W., RHEE, J.K., LEE, W.J., and ZHANG, B.T. (2006). miTarget: microRNA target gene prediction using a support vector machine. BMC Bioinformatics **7,** 411.

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.

KREK, A., GRUN, D., POY, M.N., WOLF, R., ROSENBERG, L., EPSTEIN, E.J., MACMENAMIN, P., DA PIEDADE, I., GUNSALUS, K.C., STOFFEL, M., and RAJEWSKY, N. (2005). Combinatorial microRNA target predictions. Nat Genet **37,** 495–500.

LALL, S., GRUN, D., KREK, A., CHEN, K., WANG, Y.L., DEWEY, C.N., SOOD, P., COLOMBO, T., BRAY, N., MACMENAMIN, P., KAO, H.L., GUNSALUS, K.C., PACHTER, L., PIANO, F., and RAJEWSKY, N. (2006). A genome-wide map of conserved microRNA targets in *C. elegans*. Curr Biol **16,** 460–471.

LEWIS, A., and REDRUP, L. (2005). Genetic imprinting: conflict at the Callipyge locus. Curr Biol **15,** R291–R294.

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.

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.

MEGRAW, M., SETHUPATHY, P., CORDA, B., and HATZIGEORGIOU, A.G. (2006). miRGen: a database for the study of animal microRNA genomic organization and function. Nucleic Acids Res **35,** D149–D155.

MEISTER, G., LANDTHALER, M., DORSETT, Y., and TUSCHL, T. (2004). Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing. RNA **10,** 544–550.

OBERNOSTERER, G., LEUSCHNER, P.J., ALENIUS, M., and MARTINEZ, J. (2006). Post-transcriptional regulation of microRNA expression. RNA **12,** 1161–1167.

PASQUINELLI, A.E., HUNTER, S., and BRACHT, J. (2005). MicroRNAs: a developing story. Curr Opin Genet Dev **15,** 200–205.

RAJEWSKY, N. (2006). MicroRNA target predictions in animals. Nat Genet **38 Suppl,** S8–S13.

REHMSMEIER, M., STEFFEN, P., HOCHSMANN, M., and GIEGERICH, R. (2004). Fast and effective prediction of microRNA/target duplexes. RNA **10,** 1507–1517.

REMENYI, A., SCHOLER, H.R., and WILMANNS, M. (2004). Combinatorial control of gene expression. Nat Struct Mol Biol **11,** 812–815.

ROBINS, H., LI, Y., and PADGETT, R.W. (2005). Incorporating structure to predict microRNA targets. Proc Natl Acad Sci USA **102,** 4006–4009.

SETHUPATHY, P., CORDA, B., and HATZIGEORGIOU, A.G. (2006). TarBase: a comprehensive database of experimentally supported animal microRNA targets. RNA **12,** 192–197.

SHAHI, P., LOUKIANIOUK, S., BOHNE-LANG, A., KENZELMANN, M., KUFFER, S., MAERTENS, S., EILS, R., GRONE, H.J., GRETZ, N., and BRORS, B. (2006). Argonaute—a database for gene regulation by mammalian microRNAs. Nucleic Acids Res **34,** D115–D118.

SMITH, T.F., and WATERMAN, M.S. (1981). Identification of common molecular subsequences. J Mol Biol **147,** 195–197.

TAGANOV, K.D., BOLDIN, M.P., CHANG, K.J., and BALTIMORE, D. (2006). NF-kappaB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses. Proc Natl Acad Sci USA **103,** 12481–12486.

TALMOR-NEIMAN, M., STAV, R., FRANK, W., VOSS, B., and ARAZI, T. (2006a) Novel micro-RNAs and intermediates of microRNA biogenesis from moss. Plant J **47,** 25–37.

TALMOR-NEIMAN, M., STAV, R., KLIPCAN, L., BUXDORF, K., BAULCOMBE, D.C., and ARAZI, T. (2006b) Identification of trans-acting siRNAs in moss and an RNA-dependent RNA polymerase required for their biogenesis. Plant J **48,** 511–521.

TRANZER, A., and STADLER, P.F. (2006). Evolution of microRNAs. Methods Mol Biol **342,** 335–350.

TSUCHIYA, Y., NAKAJIMA, M., TAKAGI, S., TANIYA, T., and YOKOI, T. (2006). MicroRNA regulates the expression of human cytochrome P450 1B1. Cancer Res **66,** 9090–9098.

WANG, B., LOVE, T.M., CALL, M.E., DOENCH, J.G., and NOVINA, C.D. (2006). Recapitulation of short RNA-directed translational gene silencing *in vitro*. Mol Cell **22,** 553–560.

WANG, X., and WANG, X. (2006). Systematic identification of microRNA functions by combining target prediction and expression profiling. Nucleic Acids Res **34,** 1646–1652.

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.

YOON, S., and DE MICHELI, G. (2006). Computational identification of microRNAs and their targets. Birth Defects Res C Embryo Today **78,** 118–128.

ZHANG, C., XUAN, Z., OTTO, S., HOVER, J.R., MCCORKLE, S.R., MANDEL, G., and ZHANG, M.Q. (2006). A clustering property of highly-degenerate transcription factor binding sites in the mammalian genome. Nucleic Acids Res **34,** 2238–2246.

ZUKER, M., and STIEGLER, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res **9,** 133–148.

Address reprint requests to:
Ilya Ioshikhes, Ph.D.
Department of Biomedical Informatics
Ohio State University
3172c Graves Hall
333 W. 10th Ave.
Columbus, OH 43210

E-mail: Ilya.Ioschikhes@osumc.edu

Received for publication December 21, 2006; received in revised form December 27, 2006; accepted January 12, 2007.

This article has been cited by:

1. Zetao Shen, Rui Zhou, Chen Liu, Yaofeng Wang, Wanqi Zhan, Ziyun Shao, Jian Liu, Feifei Zhang, Lijun Xu, Xinying Zhou, Lu Qi, Feng Bo, Yanqing Ding, Liang Zhao. 2017. MicroRNA-105 is involved in TNF-α-related tumor microenvironment enhanced colorectal cancer progression. *Cell Death & Disease* **8**:12. . [Crossref]
2. G Y He, J L Hu, L Zhou, X H Zhu, S N Xin, D Zhang, G F Lu, W T Liao, Y Q Ding, L Liang. 2016. The FOXD3/miR-214/MED19 axis suppresses tumour growth and metastasis in human colorectal cancer. *British Journal of Cancer* **115**:11, 1367-1378. [Crossref]
3. Wei Wang, Gang Ji, Xin Xiao, Xu Chen, Wei-Wei Qin, Fan Yang, Yu-Fang Li, Lin-Ni Fan, Wen-Jin Xi, Yi Huo, Wei-Hong Wen, An-Gang Yang, Tao Wang. 2016. Epigenetically regulated miR-145 suppresses colon cancer invasion and metastasis by targeting LASP1. *Oncotarget* **7**:42. . [Crossref]
4. Feifei Zhang, Yuhao Luo, Ziyun Shao, Lijun Xu, Xiaoxu Liu, Ya Niu, Jiaolong Shi, Xuegang Sun, Yawei Liu, Yanqing Ding, Liang Zhao. 2016. MicroRNA-187, a downstream effector of TGFβ pathway, suppresses Smad-mediated epithelial–mesenchymal transition in colorectal cancer. *Cancer Letters* **373**:2, 203-213. [Crossref]
5. Amitava Das, Chandan K. Sen. NutrimiRomics 53-60. [Crossref]
6. Minyu Wang, Jilin Wang, Xuan Kong, Huimin Chen, Yingchao Wang, Miao Qin, Yanwei Lin, Haoyan Chen, Jie Xu, Jie Hong, Ying-Xuan Chen, Weiping Zou, Jing-Yuan Fang. 2015. MiR-198 represses tumor growth and metastasis in colorectal cancer by targeting fucosyl transferase 8. *Scientific Reports* **4**:1. . [Crossref]
7. Tanu Sharma, Ryan Hamilton, Chandi C Mandal. 2015. miR-214: a potential biomarker and therapeutic for different cancers. *Future Oncology* **11**:2, 349-363. [Crossref]
8. Sara A MacLellan, Calum MacAulay, Stephen Lam, Cathie Garnis. 2014. Pre-profiling factors influencing serum microRNA levels. *BMC Clinical Pathology* **14**:1. . [Crossref]
9. Lijun Xu, Yue Zhang, Hui Wang, Guanhua Zhang, Yanqing Ding, Liang Zhao. 2014. Tumor suppressor miR-1 restrains epithelial-mesenchymal transition and metastasis of colorectal carcinoma via the MAPK and PI3K/AKT pathway. *Journal of Translational Medicine* **12**:1. . [Crossref]
10. B Wang, W Li, H Liu, L Yang, Q Liao, S Cui, H Wang, L Zhao. 2014. miR-29b suppresses tumor growth and metastasis in colorectal cancer via downregulating Tiam1 expression and inhibiting epithelial–mesenchymal transition. *Cell Death & Disease* **5**:7, e1335-e1335. [Crossref]
11. Hui Wang, Hongying An, Bin Wang, Qing Liao, Weidong Li, Xuejun Jin, Shuzhong Cui, Yajie Zhang, Yanqing Ding, Liang Zhao. 2013. miR-133a represses tumour growth and metastasis in colorectal cancer by targeting LIM and SH3 protein 1 and inhibiting the MAPK pathway. *European Journal of Cancer* **49**:18, 3924-3935. [Crossref]
12. David J. Welsh, Andrew J. Peacock. 2013. Cellular Responses to Hypoxia in the Pulmonary Circulation. *High Altitude Medicine & Biology* **14**:2, 111-116. [Abstract] [Full Text] [PDF] [PDF Plus]
13. YUK C. CHAN, JAIDEEP BANERJEE, SANG YONG CHOI, CHANDAN K. SEN. 2012. miR-210: The Master Hypoxamir. *Microcirculation* **19**:3, 215-223. [Crossref]
14. Hesan Luo, Jinjin Zou, Zhongyi Dong, Qin Zeng, Dehua Wu, Li Liu. 2012. Up-regulated miR-17 promotes cell proliferation, tumour growth and cell cycle progression by targeting the RND3 tumour suppressor gene in colorectal carcinoma. *Biochemical Journal* **442**:2, 311-321. [Crossref]
15. Olena Babenko, Andrey Golubov, Yaroslav Ilnytskyy, Igor Kovalchuk, Gerlinde A. Metz. 2012. Genomic and Epigenomic Responses to Chronic Stress Involve miRNA-Mediated Programming. *PLoS ONE* **7**:1, e29441. [Crossref]
16. JinHua Zuo, YunXiang Wang, HaiPing Liu, YuanZheng Ma, Zheng Ju, BaiQiang Zhai, DaQi Fu, Yi Zhu, YunBo Luo, BenZhong Zhu. 2011. MicroRNAs in tomato plants. *Science China Life Sciences* **54**:7, 599-605. [Crossref]
17. M. Hossein Radfar, Willy Wong, Quaid Morris. 2011. Computational Prediction of Intronic microRNA Targets using Host Gene Expression Reveals Novel Regulatory Mechanisms. *PLoS ONE* **6**:6, e19312. [Crossref]
18. Stephen A. Stanhope, Srikumar Sengupta, Johan den Boon, Paul Ahlquist, Michael A. Newton. 2009. Statistical Use of Argonaute Expression and RISC Assembly in microRNA Target Identification. *PLoS Computational Biology* **5**:9, e1000516. [Crossref]
19. James A. Williams, Aaron E. Carnes, Clague P. Hodgson. 2009. Plasmid DNA vaccine vector design: Impact on efficacy, safety and upstream production. *Biotechnology Advances* **27**:4, 353-370. [Crossref]
20. Anna Alisi, Andrea Masotti, Valerio Nobili. 2009. Profiling microRNA expression: A snapshot of nonalcoholic steatohepatitis and a recording of its pathogenesis. *Hepatology* **49**:2, 706-707. [Crossref]

21. Zujun Yin, Chunhe Li, Xiulan Han, Fafu Shen. 2008. Identification of conserved microRNAs and their target genes in tomato (Lycopersicon esculentum). *Gene* **414**:1-2, 60-66. [Crossref]

22. Chandan K. Sen, Sashwati Roy. 2007. miRNA: Licensed to Kill the Messenger. *DNA and Cell Biology* **26**:4, 193-194. [Abstract] [PDF] [PDF Plus]