

Computational Prediction of microRNAs and their Targets

Salim A and Vinod Chandra SS*

Computer Centre University of Kerala, Thiruvananthapuram, India

Abstract

microRNAs (miRNAs) are small single stranded RNAs, having an average length of 22 nucleotides. microRNAs are generated from endogenous hairpinshaped transcripts through post-transcriptional activity. They are found in diverse organisms and play critical roles in gene expression regulations of many essential cellular processes. Different computational methods are currently available for identifying microRNA genes in the genomes of different species, and predict exactly the functional part of the microRNA gene, namely the mature microRNA. The mature microRNA makes base pairing with mRNA, where complementarities exist between them and thus regulate gene expression. This article presents a brief description of microRNA related databases, different approaches for prediction and target gene identification. The methods employed, characteristics, and species addressed in each of publicly available tools are analyzed.

Keywords: miRNA; miRNA targets; miRNA prediction

Introduction

The microRNAs (miRNAs) are highly conserved, small, but endogenous noncoding RNAs that regulate gene expression. microRNAs are transcribed and processed to its mature form (22 nt) through a series of operations, where a hairpin shaped stem loop structure called pre microRNA with hundreds to thousands of nucleotides in length, and pri microRNA (70 nt) are its intermediate forms. microRNAs can interact with target mRNAs at specific sites, results in either cleavage and destruction of target mRNA or translational repression [1,2]. Regulatory functions of microRNAs are being identified, which includes mediating mRNA decay by directing rapid deadenylation of mRNAs, stabilizing mRNA, increasing gene expression, affecting cell functions like cell proliferation and differentiation [3,4].

Intense research are progressing in identification of biological functions of microRNAs, pathways affected by microRNA in different diseases, and modeling signal pathway networks using predicted targets [5-7]. Its interaction in protein transactional mechanism attenuate or progress protein synthesis, thereby involved in the regulatory role in human diseases, especially in diseases like lymphoma, leukemia and many cardiac problems [8,9]. To cite a few, the microRNA, miR203 act as a tumor (Glioblastoma) suppressor, as its expression reduces cell growth, migration and invasion [10]. miR208 plays role in progression of heart failure but therapeutic inhibition of miR208a could improve cardiac function and survival during heart failure [11]. Deciphering the exact nature of microRNA:mRNA interactions will unravel the body mechanism to greater extents and will pave way to better diagnosis and more effective drugs. Derek lemons et al. [7] discusses the suitability of using microRNAs as phenotypic screen and reiterate its essential requirement to exact identification of microRNA target sites.

The way microRNA and their targets interact in animals and plants are different in certain aspects. The microRNA target for plant microRNAs requires high complementarities for identification by genomewide scale searches [12-16]. The plant microRNA exhibits perfect or nearly perfect base pairing with the target but in the case of animals, the pairing is more imperfect. This makes the microRNA target identification problem in animals more complex compared to that in plants. Also microRNAs in plants bind to their targets within the coding regions cleaving at single sites, whereas most of the microRNA binding sites in animals are in the 3' UTR regions. In animal microRNA-mRNA interactions, multiplicity (one microRNA targeting

more than one gene) and cooperation (one gene targeted by several microRNAs) are very common but not in the case of plants [17,18]. All these make the approaches in microRNA target prediction in plants and animals different in many aspects [19,20].

microRNAs are initially identified by genetic screening technique [21,22]. Experimental approaches for microRNA identification include Northern blotting, Polymerase chain reaction (PCR), Rapid amplification of cDNA etc. The northern blot is a technique used in molecular biology research to study gene expression by detection of RNA or isolated mRNA in a sample, whereas PCR amplify a single or few copies of a piece of DNA across several orders of magnitude, generating millions or more copies of a particular DNA sequence. Experimental identification of microRNA is slow since some microRNAs are difficult to isolate (may contain other non-coding RNAs such as tRNA, rRNA, siRNA etc.), low expression of microRNAs, tissue specificity, and procedural delay in experiments [23,24]. So computational approaches have been extensively used in microRNA research to identify most probable candidates. Large number of algorithms and techniques has been devised both for microRNA prediction and for identification of microRNA target sites. Computational techniques mainly focus on the secondary structure, sequence characteristics, conservation among related organisms and folding free energy index of genomic segment [25]. Identification of novel microRNA genes and its targets are most imminent problems toward the understanding of post-transcription gene regulation. In this paper, existing computational methods for microRNA prediction, target prediction, databases of microRNAs and targets, and signaling pathways related with microRNAs are discussed.

microRNA Prediction

Quite a large number of computational microRNA prediction

*Corresponding author: Vinod Chandra SS, Computer Centre University of Kerala, Thiruvananthapuram, India, Tel: 914712305801; E-mail: vinodchandross@gmail.com

Received May 10, 2014; Accepted June 26, 2014; Published July 05, 2014

Citation: Salim A, Vinod Chandra SS (2014) Computational Prediction of microRNAs and their Targets. J Proteomics Bioinform 7: 193-202. doi:10.4172/jpb.1000320

Copyright: © 2014 Salim A, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

algorithms have been developed over past 11 years. They all rely in one way or other, the characteristics such as structural features of hairpin like structures, thermodynamic parameters, sequence conservation among multiple species, sequence specific (properties of consecutive nucleotides) parameters. The computational approaches have been successfully used to identify microRNA genes in various plants and animals including human, mouse, *C. elegans*, *Drosophila*, *A. thaliana*, *Oryza sativa* and many more. Table 1 shows computational approaches to microRNA identification or classification available in the public domain. The biggest advantage of such methods is prediction of potential candidate miRNAs which could subsequently verify directly or indirectly by experimental approaches. In the initial stages, computational approaches for prediction were utilizing genetic conservation between the species and hence comparative in nature. Later, many algorithms are developed irrespective of genetic conservation. All the tools were built based on data collected and properties extracted from experimental results on microRNA identification. The selection and validation of feature set and advances in machine learning techniques give way to the development algorithms with very high accuracy. Recently, studies are conducted exclusively to determine discriminant power of features selected in pre-microRNA identification [26]. A significant change arose due the advent of Next Generation Sequencing techniques, a quite a few algorithms based on enormous data output from sequencing techniques are also available. These sections discuss various computational methods for microRNA identification.

Comparative, sequence and structure based prediction

The different functional elements in the human genome have been systematically discovered by Comparative genomics [27]. Studies shows that sequence conservation in human and other vertebrates are roughly 38%, 3' UTR the region having widespread post transcriptional regulatory effects, shows extreme conservation in vertebrates, and is followed by 5'UTR region (which also shows post transcriptional regulation) [28]. When experimentally identified microRNAs across different species analyzed, the strong conservation enlightened the importance of restricting search for microRNAs to the conserved noncoding regions only. Otherwise, for example, Bentwich et al. [29] screened 11 million hairpins like structures in his non-comparative approach to identify microRNAs from human genome and validated microRNAs till the date is few hundreds only. If search space is limited to specific regions of the genome, it saves lots of computation and reduces false positive rates.

Among the computational tools listed in Table 1, the tools such as miRScan, Srnaloop, MIRcheck, and miRfinder are examples of tools that employ conservation somewhere in the algorithm. The other determinants are properties of 70 nt intermediate precursor of the transcript called pre-microRNA. Some algorithms taken into consideration even longer transcripts called pri-microRNA from which pre-microRNA has been derived. Both pri-microRNA and pre-microRNA form a hair pin like stem loop structure due to base pairing, and properties like length of stem, G:C content, A:U content, G:C ratio, number of bulges, free energy of structure, and many more are considered as determinants. To determine folding energy structure, the programs like RNA folding program Mfold [30,31] or Vienna RNA software package [32-34] or similar programs were being used.

MiRScan is a tool, described by Lim et al. in 2003 [35] and is used to identify microRNA genes in *C. elegans*. Since *C. elegans* and *C. briggsae* are evolutionary conserved species, *C. elegans* genome is scanned for hairpin structures that are conserved with that of *C. briggsae*. Around

36000 hairpins were found satisfying this minimum requirement. These hairpins are evaluated by similarity search with training set (data from known microRNAs). The similarity search are based on features like base pairing with microRNA (both at fold back region and rest of fold back region), conservation at 5' end, conservation at 3' end, initial 5 nucleotides of microRNA, and bulge symmetry etc. It is found that base pairing has major deciding role microRNA identification.

MiRAlign is a computational tool not based on strict sequence conservation, but loose mature microRNA conservation. Properties of secondary structure conservation are also considered [36-40]. The web interface of the tool provided the facility to input candidate sequence of length between 50 and 300 nucleotides, and to specify limits of other parameters such as Minimum Free Energy (MFE), and Minimum Sequence Similarity. The known microRNAs and precursors from six species were used to prepare training set. The algorithm works in 5 steps: initially, Minimum Free Energy (MFE) of the sequence and its reverse are calculated and retain those sequences with MFE less than 20 kcal. Second level of elimination is by a pair wise sequence alignment of candidate sequence with known microRNAs, where the tool CLUSTALW [41-54] is used to find sequence similarity score. In the third and fourth step, filtering is based on homology with known microRNAs. The position in the stem loop, the arm in which the microRNA can present, are restricted in third step, whereas pair wise structure alignment measure is used in fourth step. Finally, a Total similarity score(tss) with all homologs identified is calculated and sequence with highest score are identified as microRNA.

Many of the microRNA genes are clustered together in human, mouse and *Drosophila* [42,55-57]. Conservation pattern in clusters and properties of neighboring areas of microRNAs were analyzed for new predictions [58]. ERPIN is a profile based dynamic programming algorithm for identification small RNAs such tRNA, rRNA, microRNA and other motifs. The known animal microRNA precursor data from miRNA registry are grouped into 18 clusters, each containing varying number of similar microRNAs. The clustering is done by sequence alignment using the tool CLUSTALW, followed by visual inspection. Each cluster, the sequence and structure characteristics form the basis of pro le search. Another tool which utilizes clustering principles is miR-abela.

There are many tools especially designed based on conservation among plant genes. MIRcheck and findMiRNA are two examples. The tool, MIRcheck [36] identifies repetitive elements in genomic sequence between *Arabidopsis thaliana* and *Oryza sativa* and predicted successfully microRNAs in *Arabidopsis thaliana*.

Expressed Sequence Tag(EST) represent expressed portion of genome, and are base pairs generated randomly from selected cDNA clones. ESTs are useful for gene prediction and identification for species where draft genome sequence is not available. Zhang et al. proposed method for microRNA prediction by EST analysis [14,15,59] and they identified 338 new potential plant microRNAs from 60 different plant species. All known *Arabidopsis* microRNA sequences are compared with publicly available EST database, then closely matching sequence (n/n) , $(n-1)/n$, $(n-2)/n$ where n is length of previously known *Arabidopsis* microRNA) were obtained, and secondary structures were predicted using Zucker folding algorithm, mFold. The final prediction is based on combined score of free energy and helical score for each individual arm.

Non-comparative, sequence and structure based prediction

These category of algorithms are not rely on phylogenetic

No.	Tool	Species	Method in brief	Availability
1	ERPIN Lambert et al. [33,34]	Animal, Plants	Sequence alignment, dynamic programming, clustering	http://rna.igmors.u-psud.fr/Software/erpin.php (Stand-alone application)
2	MiRScan Lim et al. [35,36]	<i>C.elegans</i>	Evolutionary conservation and Sequence Similarity	http://genes.mit.edu/mirscan (Web application)
3	Srnaloop Grad et al. [37]	<i>C.elegans</i>	Conservation, sequence alignment by dynamic programming, folding energy, scoring based on base pairing, filter based on structural parameters	http://arep.med.harvard.edu/miRNA (Stand-alone application)
4	MiRCheck Bartel et al. [38]	<i>Arabidopsis</i>	Conservation, base pairing features of secondary structures	http://www.mybiosoftware.com/rna-analysis/11199 (Stand-alone application)
5	findMiRNA Adai et al. [39]	<i>Arabidopsis</i>	Predicts possible microRNAs corresponds to target sites by the use of sequence complementarity, free energy, conservation	http://sundarlab.ucdavis.edu/mirna (Stand-alone application)
6	miRAlign Wang et al. [40]	Animals and plants	Sequence alignment, Free energy	http://bioinfo.au.tsinghua.edu.cn/miralign (Web application)
7	TripletSVM Chenghai et al. [41]	Human	Support vector machines, Sequence characteristics(triplet)	http://bioinfo.au.tsinghua.edu.cn/software/mirnasvm (Stand-alone application)
8	miR-abela Sheng et al. [42]	Mammals	Support vector machines for valid stem loop filtering. Search limited to extended genomic regions of known microRNA clusters of human/mouse/rat.	http://www.mirz.unibas.ch/cgi/pred_miRNA_genes.cgi (Web application)
9	MiRFinder Huang et al. [43]	Human	Support vector machines with features extracted based on mutations of pre-microRNA secondary structure with that of pseudo microRNAs, free energy, base pairing of mature microRNAs	http://www.bioinformatics.org/mirfinder/ (Stand-alone application)
10	miRDeep Friedlander et al. [44]	All species	Align sequenced reads with genome, probabilistic scoring of secondary structure and signature	www.mdc-berlin.de (Stand-alone application)
11	ANN Classifier Chandra [45]	Human	Artificial neural networks, Structural and thermodynamic features	http://www.mca.cet.ac.in/research.htm (Stand-alone application)
12	MirExpress Wang et al. [46]	Human, Plants	Sequenced genome is not required, pre-processing on Deep sequencing data, alignment with known microRNAs and prediction using microRNA expression profile	http://mirexpress.mbc.nctu.edu.tw (Stand-alone application)
13	Mpred Chandra et al. [47]	Human	Hidden markov model, Artificial neural networks, structural and thermodynamic features	http://www.mca.cet.ac.in/research.htm (Stand-alone application)
14	MiRPara Wu et al. [48]	All species	Support vector machines, Properties of microRNAs, pre-microRNAs and pri-microRNAs	https://code.google.com/p/mirpara/ (Stand-alone application)
15	microPred Rukshan et al. [49]	Human	Support vector machines, Sequence and structure features	http://www.cs.ox.ac.uk/people/manohara.rukshan.batuwita/microPred.htm (Stand-alone application)
16	miRDeep* Jiyuan An et al. [50]	Human	Analysis on deep sequenced data- reads aggregation, sequence alignment, aggregate for potential microRNAs	http://www.australianprostatecentre.org/research/software/mirdeep-star (Stand-alone application)
17	mirTools Zhu et al. [51]	Human	Unique reads from deep sequencing data are identified, aligned with reference genome, classified to different categories, microRNAs are identified from 'unclassified' sequences using miRdeep tool	http://122.228.158.106/mr2_dev/download.php (Stand-alone application)
18	miRanalyzer Michael Hackenberg et al. [52]	Human	Unique reads from deep sequencing data are identified, aligned with known microRNAs, then non coding RNAs. Machine learning using Random forest algorithm (feature vectors :structure, sequence (triplet), energy parameters)	http://web.bioinformatics.cicbiogune.es/microRNA (both stand-alone and web application)
19	miR-BAG Ashwani Jha et al. [53]	All species	Structural, thermodynamic, positional Features of 21 nt wide window, global features of full length sequence and machine learning approach (BAGing(Bootstrap Aggregating) with SVM, Naive Bayes and Best First Decision Tree (BFS))	http://scbb.iht.res.in/presents/mirbag (Both standalone and web application)

Table 1 lists microRNA prediction software tools in the order of development and publication year. The table shows a brief description of method used, the type of availability (whether through web interface or to download and install locally), and species related.

Table 1: Available pre-microRNA classification and prediction Tools.

conservation. Hence initial reduction of search space for microRNA prediction is impossible. But, these algorithms are capable of finding microRNAs from species where close analogues species data are not

available. The nucleotide sequences and their characteristics along with other structural and thermodynamic parameters of stem loop structures are used prediction of microRNAs.

A local triplet sequence feature has been employed in tools, **TripletSVM**, **MiPred** and **MiRank** as critical parameters. For every three consecutive nucleotides from the secondary structure representation, a paired set is represented by a left brace '(' and unpaired by a dot '.', then "(((represents three consecutive pairs, while "... represents three consecutive nonpairs. The remaining six compositions are "(.", "(.", "(.", "(.", "(.", and "(.". The total triplet structure rises up to 32 when the middle nucleotide is fixed as either one of A, C, G, U. This triplet structure, on each hairpin is counted and used as input feature vector of Support Vector Machine(SVM) in the case TripletSVM. Positive training data has been collected from miRNA registry, while negative samples collected from protein coding regions. The criteria for selection of negative samples are fixed in such a way that collected samples are similar to the real pre-microRNAs in terms of widely accepted characteristics such as number of base pairs, maximum energy of hairpin, number of multiple loops etc. On the test data set prepared from human chromosome 19, the authors claim a specificity of 89%. Once trained with human microRNA data, TripletSVM successfully classified 'crossspecies'(another 11 species data from miRNA registry) data too. Other than triplet structure values, MiPred uses Minimum free energy(MFE) and P-value. MFE of secondary structure is obtained using Vienna RNA software [32]. The P-value is a measure of difference in MFE values of original sequence and its random variations. $P\text{-value} = R / (N+1)$, calculated by repeatedly measuring MFE of randomized sequence obtained from the original one while keeping dinucleotide distribution as constant, where R number of iterations where MFE less than that of the original sequence and N is the total number of iterations. MiPred is based on Random forest algorithm where machine learning techniques such as bagging [60] and random feature selection are utilized. The relative significance of each feature vector for microRNA prediction is also determined. MiPred claims that P-value, MFE, C..., U(((, and A(((are top five determinants of pre-microRNA like hairpins as real or pseudo microRNAs. The local triplet structure has been utilized by yet another tool, MiRank [61]. In addition to triplet structure, MiRank's feature selection includes normalized minimum free energy(MFE), normalized number base pairing and normalized loop length. Normalization factor is total length of pre-microRNA. The algorithm relies on random walks on a graph. The relationship between known microRNAs, putative microRNAs are modeled using a weighted graph $G(V,E)$, where each vertex V represent a microRNA and each edge represent relation between them. The weight of edge is proportional to the closeness of relationship between vertices. It is defined as $w_{ij} = \exp[-d(x_i, x_j)^2 / \sigma]$ where $d(x_i, x_j)$ is Euclidean distance between samples x_i and x_j , and is heat transfer parameter. Sequence conservation is not deciding factor at any stage of algorithm and hence it could identify microRNAs from genome sequences where a few microRNAs are available.

Support Vector Machine (SVM) as machine learning technique has been widely in microRNA predictions. TripletSVM, MirPara, PMirP [62], microPred [49], MiRFinder and miR-abela are example tools where SVM has been used. MirPara, predicts microRNA coding regions from genome scale sequences. This tool is a preferable choice to complement High Throughput Sequencing (HTS) experiments, by identifying putative pre-microRNA sequences. The training set consists of microRNA secondary structures of 6330 microRNA sequences from miRBase release 13.0. After the initial screening, 5576 pri-microRNA sequences belonging to animals (4886), plants (1215) and viruses (227) were selected. Unlike other tools, mirPara considers the parameters related to microRNA, pre-microRNA and pri-microRNA (as full length pri-microRNA sequences is not available on public database, algorithm

utilizes partial pri-microRNA sequences). A total of 42 parameters ranging broad categories such as size, sequence, stability, and structure were identified and corresponding data set is used to train the SVM. Three SVM models are created, one for animal data, second for plant data and third for the entire data. Among the 42 parameters, 10 top scoring parameters identified as seed parameters. These seeds are used to generate 10 highest scoring pairs, which in turn utilized to generate 10 highest triplets and so on. Caution has been taken so that negative data set is completely random or be closely related to positive data set (in both cases SVM may not be able to distinguish from the biological sequences). miRPara initially splits the genome sequences into 500 nt fragments with 200 nt overlap, and predicts hairpins in the fragments by using UNAFold. Next, it rejects hairpins having length less than 60 nt, checks for duplicates, revalidate using UNAFold then, classify using the SVM classifier. MiRFinder introduces a new secondary structure representation with paired, unpaired, insertion, deletion and bulge of bases. Other than minimum free energy and base pairing features of mature microRNA region, frequency of different two nucleotide combinations according the above representation constitute the complete feature vector.

High Throughput Sequencing (HTS)

Second generation sequencing platforms enable genome sequencing in a very faster rate, with drastically reduced cost, by massively parallel instruments capable of processing millions of reads. A read means one random fragment of nucleotide sequences processed individually. Later the sequenced representative reads are aligned with reference genome. Examples of these systems are Genome Sequencer FLX from Roche Applied Science (454 Sequencing), the SOLiD System from applied Biosystems and Illumina Genome Analyzer [63]. Recently, many tools have been developed to discover known and novel microRNAs from the data outputs of the above mentioned systems. All tools perform preprocessing to find representatives of reads, majority perform genomic alignments. miRDeep, miRAnalyzer, mirTools, and mirExpress examples of such computational tools, where miRAnalyzer and mirTools works based on its alignment known standard databses. The problems with this method are very low success rate [64] (maybe one microRNA identification out of 10000 reads) and conservative nature of algorithm.

miRDeep [50] detects novel microRNAs from RNA Sequencing(RNAseq) [65] data. RNASeq is a powerful method for discovering, and profiling RNA transcripts using Next Generation Sequencing technologies. miRDeep starts with aggregation of similar reads from the input data (RNASeq data output), filter the reads by retaining the ones having length between 18 to 23 nucleotides, followed by a genomic mapping. miRDeep algorithm is a probabilistic model based on position and frequency of sequenced RNA data with that of known microRNA precursors. miRDeep^{*} and miRDeep2 are variations of same algorithm with improved capabilities. Erl Zhu et al. developed a tool, mirTools, [51] which utilizes miRdeep and several other popular sequence databases. Again, the data from Deep Sequencing experiments are the input. Initially, low quality reads are filtered out and unique sequences are aligned against miRBase, RFam, and RepeatMasker, [66] and coding genes of reference genome, and hence classified into known microRNA, degradation fragments of non-coding RNA, generic repeats, and mRNA. The sequences which do not fall on above categories are used to identify novel microRNA using miRDeep program. miRAnalyzer, the algorithm reads the read-count input le and trim reads with less than 17 bases and longer than 26 bases. The resultant sequences are aligned with corresponding species data

retrieved from miRBase. This mapping is done in four steps aligning to mature, maturestar, unobserved maturestar and hairpin sequences. The sequences are then aligned with other transcriptome libraries: RefSeq and RFam, unmapped entries are used to predict novel microRNA using a Random Forest Algorithm. The prediction of new microRNAs reached ROC values of 97.9% and recall values of up to 75% on unseen data. mirExpress [46] perform the microRNA expression profile in three steps : preprocessing of raw data from Deep sequencing experiments, alignment of all sequencing reads against known mature microRNAs, and finally microRNA expression profile is constructed by computing the sum of read counts for each microRNA according to the alignment criteria.

microRNA Target Prediction

MicroRNAs direct post transcriptional regulation of gene expression by binding to complementary site on target mRNAs. The result of this binding either immediate degradation or repression of translation. The degradation or blocking of translation depends on the degree of complementary relation between microRNA sequence and its target sequence. It is identified that a single microRNA may regulate multiple genes and /or multiple genes may regulate a single target sites. The target genes that are regulated by microRNAs greatly help the study of microRNA function in animals and plants [67,68].

Thousands of microRNAs are identified in animals, plants and virus. But, targets of the majority of these microRNAs have not been identified due to the cost factor and delay experimental validation. The challenges in computational prediction of target sites, especially in animals, are complexity of microRNA: mRNA interaction and limited knowledge of exact rules governing the system. Table 2 shows current computational microRNA target prediction methods, which have been employed successfully to identify potential microRNA targets in mRNA sequences in many species. These algorithms taken different characteristics of microRNA binding sites such structural, base pairing, thermodynamic and positional properties of complementary sequence of microRNA whose target to identified, evolutionary conservation etc. Generally, algorithms use a subset of the following characteristic properties:

1. In plants, microRNAs are nearly or near-perfect complementarity to its target mRNAs with no bulges or gaps at the sites. In animals, the functional binding and sequence complementarity are related, and extensive studies resulted in accepted rule is that the 5' region of a miRNA from nucleotides 2 or 8 or 2 to 9 from seed region.
2. The folding free energy of RNA-RNA duplex better than about 12 kcal/mol [38,69,70].
3. Mature microRNA binding sites of the mRNA are highly conserved from species to species particularly within the kingdom, and the seed region is the most evolutionary conserved region [13,69,70].
4. Base pairing pattern: Base pairing patterns of with miRNA:mRNA duplex on the basis of different types of targets-e 5' dominant seed only targets, 5' dominant canonical targets, and the 3' complimentary targets [71].

Thermodynamic stability of miRNA:mRNA duplex are analyzed by calculation of free energy (ΔG). The minimum free energy or the Gibbs free energy ΔG describes the energy of molecules in aqueous solution. For an equilibrium process $\Delta G=0$, for unfavorable process $\Delta G>0$ and for a favorable process $\Delta G<0$. Hence, biomolecules in solution arrange themselves so as to minimize the free energy of the entire system.

Likewise, the G in folded nucleic acid structure tends to be the lowest, to ensure a stable structure.

Among the algorithms listed in Table 2 the tools MiRanda, TargetScan, PicTar, DIANA-microT, targetRank, and MTar used evolutionary conservation as part of prediction process, whereas RNA22, MicroTar, TargetSpy, and NBmiRTar have considered factors other than conservation. TargetScan, miRanda and PicTar perform extensive search in 3' UTR mRNAs for probable targets. All the three perform, seed match almost in similar ways, followed by free energy calculations. MiRanda focuses on sequence matching of miRNA:mRNA pairs, by estimating energy of physical interaction. It was initially developed for prediction of microRNA targets in Drosophila and later extended to find microRNA targets in mammals (human, mouse and rat) and Zebra fish [68]. A matching score of +5 for A:U, +5 for G:C, +2 for G:U and mismatch score of 3 are assigned. A gap opening is penalized by a score 8 whereas gap elongation penalty is only 2. The individual score is multiplied by a scaling factor for nucleotide positions 111 of the mRNA, hence a final score 'S' is obtained. The free energy G is calculated by RNAlib package [72]. Based on the cutoff score 'S', the free energy G the match is pursued further and resulting targets are ranked accordingly. Using this software, a large number of targets were identified including protein coding genes in Homo sapiens. The false positive rate was estimated to be 24%.

PicTar [70,73] predicts and analyze microRNA target in flies. It also checks whether targeting in flies is comparable that of vertebrates, whether miRNA:mRNA relationships are conserved between them. A Hidden Markov Model (HMM) was used to validate the target sites. The algorithm first finds candidates 3' UTR sites, called anchor sites (seeds), with specified minimum number of WatsonCrick base pairing. The imperfect sites which lacks minimum free energy (two-thirds of free energy of perfectly base paired miRNA:mRNA duplex) are filtered out. Anchors are then fed into a trained HMM to yield a score for 3' UTR in the multiple alignments. Separate scores were obtained for different organisms. Its false positive rate has been estimated to be around 30%.

Targetscan [74] is a tool used to predict microRNAs which bind to 3' UTRs of vertebrate transcriptomes. TargetScan could predict more than 451 human microRNA targets. This algorithm has three parts. First it searches for a conserved seed match between nucleotides in the positions 2 to 7 of the microRNA and a 6 nucleotide section of the 3'UTR of the mRNA in a multiple alignment of sequences of all organisms under test. Then it proceeds to the second step which searches for a 'conserved anchor' on the 3' UTR, just downstream of the seed match for all sequences. In the third step, it looks for a match to a seed region in microRNA with the 3' UTR of mRNA. Then mRNA is declared as a target provided a conserved match is found in second or third step or in both. TargetScanS, a modified version of TargetScan, omits multiple sites in each target and further filters the targets using thermodynamic stability criterion [17]. Using this modified method more than 5300 human genes were predicted as possible targets of microRNAs [17,74]. The false positive rate varies between 22% to 31%.

Target Spy, not rely on evolutionary constraints and not looking for presence of seed match, it allows detection of microRNA targets, especially 3' compensatory sites. However, TargetSpy intermediate results are filtered out on seed match and conservation. Initial screening of candidate targets is done by searching for areas in target sequence where predicted free energy is below a threshold value. In this study, around forty five features are identified for the machine learning purpose. But, a Correlation Feature Selection (CFS) technique, identifies seven best features which returns an Area Under Curve

No.	Tool	Species	Method in brief	Availability
1	MiRanda Enright et al. John et al. [68,75,76]	<i>Drosophila</i> , Vertebrates	Evolutionary conservation, binding energy, weighted score using base pairs, gap penalties	http://www.microma.org/microma/getDownloads.do (Stand-alone application)
2	TargetScan Lewis et al. [74]	Vertebrates	Seed match, Match outside seed region, Conservation of seed region	http://genes.mit.edu/tscan/targetscanS.html (Web application)
3	RNAhybrid Rehmsmeier [18]	<i>Drosophila</i>	Negative normalized minimum free energy values based on length of target sequence and length of microRNA, shows statistical significance of predicted targets	http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/dl_pre-page.html (Stand-alone application)
4	DIANA-microT Kirikidou et al. [77]	Human, Mouse	Search restricted 3' UTR of mRNAs, Minimum energy of potential sequence (38 nt) in comparison with 100 complementary sequence, filtering based on microRNA associated proteins	http://diana.cslab.ece.ntua.gr/microT/ (Web application, can download the software on request)
5	TargetScanS Lewis et al. [17]	Vertebrates	Simple version of TargetScan, perfect seed match, scoring based on dynamic programming	http://genes.mit.edu/tscan/targetscanS2005.html (Web application)
6	PicTar Kerk et al. [70,73]	<i>Drosophila</i> , Vertebrates	Search limited to conservative 3' UTR regions, scoring by HMM and multiple sequence alignment with eight vertebrate species	http://pictar.mdc-berlin.de/ (Web application)
7	RNA22 Huynh et al. [78] Phillipe L et al. [79]	<i>C.elegans</i> , Human, <i>Drosophila</i> , Mouse	Initial identification of putative sites by pattern match (with-out knowing targeting microRNAs), then associate microRNA with target(user defined parameters-minimum number of base pairs, maximum number of unpaired bases, maximum allowed free energy)	http://cbcsrv.watson.ibm.com/rna22.html (Web application)
8	MicroTar Rahul and Mortti [80]	<i>C.elegans</i> , <i>Drosophila</i> , mouse	No evolutionary conservation constraint, seed match and free energy(RNAIib) are deciding parameters	http://tiger.dbs.nus.edu.sg/microtar/ (Stand-alone application)
9	NBmiRTar Yousef et al. [81]	human, mouse, y, worm, ze- bra_sh	Do not require sequence conservation, based on seed match and outside features of microRNA:mRNA duplex, and Naive Bayes classifier	http://wotan.wistar.upenn.edu/NBmiRTar/ (Web application)
10	PiTA Kertesz et al. [82]	Human,Mouse,y <i>C.elegans</i> ,	Secondary structure, Seed region with single mismatch or G:U wobble, Free energy of miRNA:mRNA structure	http://genie.weizmann.ac.il/pubs/mir07/mir07_data.html (Stand-alone application)
11	targetRank Cydney et al. [83]	Human, mouse	Conservation, Sequence alignment with 16 vertebrate genome, Seed match	http://genes.mit.edu/targetrank/ (Web application)
12	miRDB Xiaowei [84,85]	Human, mouse, rat, dog, chicken	Target prediction by the tool MirTarget2	http://mirdb.org/miRDB/index.html (Web application)
13	MiRtif Yuchen et al. [86]	Worm, Mouse, Human, Fly	Support vector machine microRNA:mRNA interaction filter. predicted microRNA interactions from miRanda, PicTar, and TargetScanS are further filtered using SVM	http://bsal.yu.edu.tw/mirtif (Web application)
14	MTar Chandra et al. [71]	Human	Structural, positional, thermodynamic properties and machine learning with ANN	http://www.mca.cet.ac.in/research.htm (Stand-alone application)
15	targetSpy Martin et al. [87]	Human,mouse, rat, chicken.y	Evolutionary conservation and perfect seed match are not a criteria, based on ranked structural, base pairing properties (45 parameters) with machine learning	http://www.targetspy.org/ (Web application)
16	psRNATarget Xinbin Dai et al. [88]	Plant	Reverse complementary matching score between microRNA and mRNA, unpaired energy required to open secondary structure of mRNA	http://plantgmn.noble.org/psRNATarget (Web application)
17	MiRTar Hsu et al. [89]	Human	microRNA target are determined using TargetScan, miRanda, PITA, and RNAHybrid. Also find extend of biological function of microRNA by estimating over expression in KEGG pathways	http://mirtar.mbc.nctu.edu.tw/human/ (Web application)
18	comiR Claudia Coronello et al. [90]	Human	Single probabilistic score calculated from microRNA target score from four popular target finding tools,PITA, miRanda, TargetScan and mirSVR	http://www.benoslab.pitt.edu/comir/index2.php (Web application)
19	HomoTarget Hamed Ahmadi et al. [88]	Human	Pattern recognition neural network (PRNN) based classifier. Initial sequence alignment, then seed match based filtering, followed by PRNN	http://lbb.ut.ac.ir/Download/LBBsoft/homoTarget/ (Stand-alone application)
20	starBase V2.0 Jun-Hao Li et al. [91]	Human	microRNA targets are identified using prediction overlap from five tools, TargetScan, miRanda, Pictar2, PITA and RNA22, while finding ceRNA interactions	http://starbase.sysu.edu.cn/targetSite.php (Web application)

Table 2 lists microRNA target prediction software tools in the order of development and publication year. The table shows a brief description of method used, the type of availability (whether through web interface or to download and install locally), and species related.

Table 2: Available microRNA target prediction Tools.

(AUC) value of 0.79. These features are Compactness, G+C content ratio between microRNA and target site, length of longest stretch of consecutive base pairing, number of base pairing in 8 mer seed, binding asymmetry, G+C content of target site, and position of target site in 3' UTR. Binding asymmetry is the ratio between paired bases in 3' versus the amount paired bases in the 5' region. Thus, TargetSpy is a good tool for finding species specific targets, poorly conserved/low quality target sites.

MTar and HomoTarget are ANN based programs with positional, structural and thermodynamic parameters for classification. There is an initial sequence alignment, then filtering followed feature extraction, selection and classification. MTar classified three potential microRNA targets (5' seedonly, 5' dominant, and 3' canonical) interactions and reports 94.5% sensitivity and 90.5% specificity whereas HomoTarget claimed 99% specificity.

Different algorithms employ different rules for target prediction, resulting in different list of targets. Same algorithm, when used with data set collected from different sources, predictions are different. TargetScan with different data set of same species resulted only 47% overlap, whereas miRanda has given 65% overlap [75-92]. Researchers turned their attention to develop tools that employ multiple target finding programs, and obtain better result than what they could individually gained. One such tool is, comiR [90] a support vector machine based, where single probabilistic score has been calculated from ensemble of four microRNA target finding algorithms, PITA, miRanda, TargetScan and mirSVR. FermiDirac (FD) equation is used to find binding potential of microRNA:mRNA interaction from the thermodynamic models, PITA and miRanda. The other two, TargetScan and mirSVR are not thermodynamic models, where weighted sum of score (WSM) has been calculated. Authors claimed that both sensitivity and specificity in the combined model improved than individual values. Using a combined score all the four and support vector machine comiR is constructed, outperformed individually by an AUC value better by 12 to 17%. Starbase v2.0 is general online tool to find possible gene regulatory functions of non-coding RNAs, such microRNAs, long noncoding(lncRNAs), circularRNAs and pseudogenes. This tool predicts competing endogenous RNAs(ceRNAs) [93] by taking into consideration all possible targets of given mRNA and targets of other noncoding RNAs. Starbase V2.0 can be used to find probable targets, it employs five popular target finding algorithms, namely TargetScan, miRanda, Pictar2, PITA and RNA22. The most predicted targets when results combined from all five used to find ceRNA interactions.

microRNA and its Target Databases

Thousands of microRNA and its targets have been identified in animals, plants and viruses (Table 3). The miRBase database published a mature microRNA sequence of animals and virus along with their hairpin precursors and their discovery structure and functions [94,95]. Presently there are 24521 mature microRNA entries belonging to 206 species (release 20). This database also contains microRNA targets sequences.

A database of experimentally tested microRNA targets in human, mouse, fruit fly, worm, and zebra fish are listed in TarBase [96]. The database is functionally linked to several other useful databases such as Gene Ontology (GO) and UCSC Genome Browser. TarBase provide more than 65000 gene versus target interaction in its latest release. It also shows the experiments that were conducted to test microRNA versus target interactions, publications from which data are extracted.

miRecords is a resource for animal microRNA-target interactions.

miRecords keeps Validated and Predicted set of records. The validated set, curated from available literature, contains 2705 interactions between 644 microRNAs and 1901 target genes. The Predicted set is by integrating results from the leading target prediction tools such as DIANA-microT, MiRanda, MirTarget2, NBmiRTar, PicTar, PITA, RNA22, RNAhybrid, and TargetScan/ TargetScanS.

The present miRTarBase 4.0 database contains more than 50000 interactions between microRNA gene and target sites, collected from literature through data mining [97]. It is specified that interactions are experimentally confirmed by reporter assays, western blot, microarray experiments, pSILAC or qRT-PCR. They also included information like the species of microRNAs, species of target genes and experimental conditions.

PMRD is a plant database consisting of microRNA sequence and their target genes. The web server also provides secondary dimension structure and expression profiling [98]. There are 8433 microRNAs collected from 121 plant species in PMRD, including model plants and major crops.

miRNAMap is a web database contains experimental verified microRNAs and experimental verified microRNA target genes in human, mouse, rat, and other metazoan genomes [99]. This database contains 2241 microRNA and estimates average length of microRNA as 21.95 nt and average length of pre-microRNA as 88.38 nt. ViTa [100] and VirMir [101-105] are providing virus host interaction databases. ViTa investigate regulatory relationships between host microRNAs and related viruses to curate the known virus microRNA genes and the known/putative target sites of human, mice, rat and chicken microRNAs. VirMir is database of predicted viral microRNA candidates, examined the 2266 viral genome sequences for putative microRNA hairpins and identified 33691 hairpin candidates.

Databases in target pathways

microRNAs emerge as regulators of signaling pathways in diseases and other biological processes. HMDD V2.0 (the Human MicroRNA Disease Database) relates 572 microRNAs and 378 diseases. The disease versus microRNA associations are collected from 3511 articles. miR2Disease is a manually curated database which contains information about microRNA based deregulation in various human diseases. The miR2Disease documents 3273 relationships between 349 human microRNAs and 163 human diseases from published articles. PhenomiR is database of microRNA deregulation in diseases and biological processes, curated manually from published experiments. They evaluated 365 articles and database contains over 12,000 data points each one representing a deregulated microRNA in experiments.

KEGG (Kyro Encyclopedia of Genes and Genomes) is a collection of database of high level functions and biological systems, and KEGG pathway is one such system which contains pathway maps for molecular interactions for metabolism, genetic information processing, organizational systems, human diseases and drug development. DIANA-miPath v2.0 as a web server allows uploading a microRNA and results the specific KEGG pathways the given microRNA are related. The tool uses meta-analysis algorithm and predicted or validated microRNA targets sites, shows microRNA pathway interaction heat maps based on interaction levels.

Conclusion

The gene regulatory characteristics of microRNAs attracted researchers focus into its prediction from both plant and animal genes.

No	Database	Species	Availability
1	miRBase Griffith-Jones et al. [94,95]	animals, virus	http://www.mirbase.org
2	TarBase Sethupathy et al. [96]	Animals, virus	http://diana.imis.athena-innovation.gr/DianaTools
3	miRecords Xiao et al. [102]	Animals	http://mirecords.umn.edu/miRecords/
4	mirTarBase Hsu et al. [97]	Animals	http://mirtarbase.mbc.nctu.edu.tw/
5	miRNAMap Hsu et al. [99]	Animals	http://mirnamap.mbc.nctu.edu.tw
6	PMRD Zhenhai et al. [98]	Plants	http://bioinformatics.cau.edu.cn/PMRD
7	ViTa Hsu et al. [100]	Virus	http://vita.mbc.nctu.edu.tw/intr.php
8	Vir-Mir Li et al. [101]	Virus	http://alk.ibms.sinica.edu.tw/
9	miR2Disease Jiang et al. [103]	microRNA deregulation in diseases	http://www.mir2disease.org
10	PhenomiR Rueppert al. [104]	microRNA deregulation in diseases	http://mips.helmholtz-muenchen.de/phenomir
11	HMDD v2.0 Qinghua Cui et. al. [103]	microRNA expression in diseases	http://www.cuilab.cn/hmdd
12	DIANA-miPath v2.0 Ioannis S. Vlachos et. al. [105]	hierarchical clustering of miRNAs and pathways based on the levels of microRNA:mRNA interactions	http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=mirpath/index

Table 3 lists microRNA related databases- predicted and experimentally validated microRNAs, microRNA: mRNA interactions, microRNA related disease pathways.

Table 3: microRNA and its target Databases.

Even though, thousands of microRNAs are identified experimentally, for many of them, targets are not yet identified. MicroRNA can have more one target, and one can be identified from different genes, makes the process complicated. Existing methods on computational prediction of microRNAs and their targets utilize the evolutionary conservation information, secondary structure features, sequence based characteristics, thermodynamic parameters. A vast number of machine learning approaches, algorithms, and data set preparation techniques have been developed. Although all these in silico methods are perfectible, their predictions have to be systematically validated experimentally. Recent years more effective target identification, including the usage of multiple tools and modern approaches in machine learning are developed. Researchers attention are shifting towards precise identification regulated diseases, the set of proteins that may affected, and role of other non-coding RNAs and pseudogenes in microRNA:mRNA interactions. In Computational terms, all these requirements converge in a single point, the fine tuned identification of valid microRNA target sites.

References

- Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116: 281-297.
- Narry Kim V (2005) Small RNAs: Classification, biogenesis mechanism and function. *Mol Cell* 19: 1-15.
- Ambros V (2004) The functions of animal microRNAs. *Nature* 431: 350-355.
- Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. *Cell* 136: 215-233.
- Mendell JT, Olson EN (2012) MicroRNAs in stress signaling and human disease. *Cell* 148: 1172-1187.
- Baek D, Villén J, Shin C, Camargo FD, Gygi SP, et al. (2008) The impact of microRNAs on protein output. *Nature* 455: 64-71.
- Lemons D, Maurya MR, Subramaniam S, Mercola M (2013) Developing microRNA screening as a functional genomics tool for disease research. *Front Physiol* 4: 223.
- Esquela-Kerscher A, Slack FJ (2006) Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer* 6: 259-269.
- Iorio MV, Ferracin M, Liu CG, Veronese A, Spizzo R, et al. (2005) MicroRNA gene expression deregulation in human breast cancer. *Cancer Res* 65: 7065-7070.
- Dontula R, Dinasarapu A, Chetty C, Pannuru P, Herbert E, et al. (2013) MicroRNA 203 Modulates Glioma Cell Migration via Robo1/ERK/MMP-9 Signaling. *Genes Cancer* 4: 285-296.
- Montgomery RL, Hullinger TG, Semus HM, Dickinson BA, Seto AG, et al. (2011) Therapeutic inhibition of mir-208a improves cardiac function and survival during heart failure. *Circulation* 124: 1537-1547.
- Brennecke J, Stark A, Russell RB, Cohen SM (2005) Principles of microRNA-target recognition. *PLoS Biol* 3: e85.
- Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, et al. (2002) Prediction of plant microRNA targets. *Cell* 110: 513-520.
- Zhang B, Pan X, Cannon CH, Cobb GP, Anderson TA (2006) Conservation and divergence of plant microRNA genes. *Plant J* 46: 243-259.
- Zhang BH, Pan XP, Wang QL, Cobb GP, Anderson TA (2005) Identification and characterization of new plant microRNAs using EST analysis. *Cell Res* 15: 336-360.
- Dezulian T, R Emmert M, Palatnik JF, Weigel D, Huson DH (2006) Identification of plant microRNA homologs. *Bioinformatics* 22: 359-360.
- Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120: 15-20.
- Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R (2004) Fast and effective prediction of microRNA/target duplexes. *RNA* 10: 1507-1517.
- Kim JT, Gewehr JE, Martinetz T (2004) Binding matrix: a novel approach for binding site recognition. *J Bioinform Comput Biol* 2: 289-307.
- Moxon S, Moulton V, Kim JT (2008) A scoring matrix approach to detecting miRNA target sites. *Algorithms Mol Biol* 3: 3.
- 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.
- Wightman B, Ha I, Ruvkun G (1993) Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* 75: 855-862.
- Lai EC (2003) microRNAs: runts of the genome assert themselves. *Curr Biol* 13: R925-936.
- Wightman B, Ha I, Ruvkun G (2006) Posttranscriptional regulation of heterogenic

- gene lin-14 by lin-4 medicine temporal pattern-formation in *C. elegans*. *Cell* 30: 395-407.
25. Chen JH, Le SY, Shapiro B, Currey KM, Maizel JV (1990) A computational procedure for assessing the significance of RNA secondary structure. *Comput Appl Biosci* 6: 7-18.
26. de On Lopes I, Schliep A, de Lf de Carvalho AC (2014) The discriminant power of RNA features for pre-miRNA recognition. *BMC Bioinformatics* 15: 124.
27. Gumucio DL, Heilstedt-Williamson H, Gray TA, Tarlé SA, Shelton DA, et al. (1992) Phylogenetic footprinting reveals a nuclear protein which binds to silencer sequences in the human gamma and epsilon globin genes. *Mol Cell Biol* 12: 4919-4929.
28. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, et al. (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 15: 1034-1050.
29. Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, et al. (2005) Identification of hundreds of conserved and nonconserved human microRNAs. *Nat Genet* 37: 766-770.
30. Mathews DH, Sabina J, Zuker M, Turner DH (1998) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* 288: 911-940.
31. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31: 3406-3415.
32. Hofacker IL (2003) Vienna RNA secondary structure server. *Nucleic Acids Res* 31: 3429-3431.
33. Legendre M, Lambert A, Gautheret D (2005) Profile-based detection of microRNA precursors in animal genomes. *Bioinformatics* 21: 841-845.
34. Lambert A, Fontaine JF, Legendre M, Leclerc F, Pernal E, et al. (2004) The ERPIN server: an interface to profile-based RNA motif identification. *Nucleic Acids Res* 32: W160-165.
35. Lim LP, Glasner ME, Yekta S, Burge CB, Bartel DP (2003) Vertebrate microRNA genes. *Science* 299: 1540.
36. Lim LP, Lau NC, Weinstein EG, Abdelhakim A, Yekta S, et al. (2003) The microRNAs of *Caenorhabditis elegans*. *Genes Dev* 17: 991-1008.
37. Grad Y, Aach J, Hayes GD, Reinhart BJ, Church GM, et al. (2003) Computational and experimental identification of *C. elegans* microRNAs. *Mol Cell* 11: 1253-1263.
38. Jones-Rhoades MW, Bartel DP (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol Cell* 14: 787-799.
39. Adai A, Johnson C, Mlotshwa S, Archer-Evans S, Manocha V, et al. (2005) Computational prediction of miRNAs in *Arabidopsis thaliana*. *Genome Res* 15: 78-91.
40. Wang X, Zhang J, Li F, Gu J, He T, et al. (2005) MicroRNA identification based on sequence and structure alignment. *Bioinformatics* 21: 3610-3614.
41. Xue C, Li F, He T, Liu GP, Li Y, et al. (2005) Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine. *BMC Bioinformatics* 6: 310.
42. Sewer A, Paul N, Landgraf P, Aravin A, Pfeffer S, et al. (2005) Identification of clustered microRNAs using an ab initio prediction method. *BMC Bioinformatics* 6: 267.
43. Huang TH, Fan B, Rothschild MF, Hu ZL, Li K, et al. (2007) MiRFinder: an improved approach and software implementation for genome-wide fast microRNA precursor scans. *BMC Bioinformatics* 8: 341.
44. Friedländer MR, Chen W, Adamidi C, Maaskola J, Einspanier R, et al. (2008) Discovering microRNAs from deep sequencing data using miRDeep. *Nat Biotechnol* 26: 407-415.
45. Vinod C, Reshmi G (2009) A pre-microRNA classifier by structural and thermodynamic motifs. *World Congress on Nature and Biologically Inspired Computing, IEEE*.
46. Wang WC, Lin FM, Chang WC, Lin KY, Huang HD, et al. (2009) miRExpress: Analyzing high-throughput sequencing data for profiling microRNA expression. *BMC Bioinformatics* 10.
47. Reshmi G, Chandra SS, Babu VJ, Babu PS, Santhi WS, et al. (2011) Identification and analysis of novel microRNAs from fragile sites of human cervical cancer: computational and experimental approach. *Genomics* 97: 333-340.
48. Wu Y, Wei B, Liu H, Li T, Rayner S (2011) MiRPara: a SVM-based software tool for prediction of most probable microRNA coding regions in genome scale sequences. *BMC Bioinformatics* 12: 107.
49. Batuwita R, Palade V (2009) microPred: effective classification of pre-miRNAs for human miRNA gene prediction. *Bioinformatics* 25: 989-995.
50. An J, Lai J, Lehman ML, Nelson CC (2013) miRDeep*: an integrated application tool for miRNA identification from RNA sequencing data. *Nucleic Acids Res* 41: 727-737.
51. Zhu E, Zhao F, Xu G, Hou H, Zhou L, et al. (2010) mirTools: microRNA profiling and discovery based on high-throughput sequencing. *Nucleic Acids Res* 38: W392-397.
52. Han K (2011) Effective sample selection for classification of pre-miRNAs. *Genet Mol Res* 10: 506-518.
53. Ashwani J, Rohit C, Mrigaya M, Heikham RS, Ravi S (2012) miR-BAG: Bagging based identification of microRNA precursors. *PLOS one* 7.
54. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673-4680.
55. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T (2001) Identification of novel genes coding for small expressed RNAs. *Science* 294: 853-858.
56. Tanzer A, Amemiya CT, Kim CB, Stadler PF (2005) Evolution of microRNAs located within Hox gene clusters. *J Exp Zool B Mol Dev Evol* 304: 75-85.
57. Tanzer A, Stadler PF (2004) Molecular evolution of a microRNA cluster. *J Mol Biol* 339: 327-335.
58. Altuvia Y, Landgraf P, Lithwick G, Elefant N, Pfeffer S, et al. (2005) Clustering and conservation patterns of human microRNAs. *Nucleic Acids Res* 33: 2697-2706.
59. Smalheiser NR (2003) EST analyses predict the existence of a population of chimeric microRNA precursor-mRNA transcripts expressed in normal human and mouse tissues. *Genome Biol* 4: 403.
60. Breiman L (1996) Bagging predictors. *Machine Learning* 24.
61. Xu Y, Zhou X, Zhang W (2008) MicroRNA prediction with a novel ranking algorithm based on random walks. *Bioinformatics* 24: i50-58.
62. Zhao D, Wang Y, Luo D, Shi X, Wang L, et al. (2010) PMirP: a pre-microRNA prediction method based on structure-sequence hybrid features. *Artif Intell Med* 49: 127-132.
63. Mardis ER (2008) The impact of next-generation sequencing technology on genetics. *Trends Genet* 24: 133-141.
64. Szittyta G, Moxon S, Santos DM, Jing R, Fevereiro MP, et al. (2008) High-throughput sequencing of *Medicago truncatula* short RNAs identifies eight new miRNA families. *BMC Genomics* 9: 593.
65. Voelkerding KV, Dames SA, Durtschi JD (2009) Next-generation sequencing: from basic research to diagnostics. *Clin Chem* 55: 641-658.
66. Tarailo-Graovac M, Chen N (2009) Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics Chapter 4: Unit 4*.
67. Wang X, Wang X (2006) Systematic identification of microRNA functions by combining target prediction and expression profiling. *Nucleic Acids Res* 34: 1646-1652.
68. John B, Enright AJ, Aravin A, Tuschl T, Sander C, et al. (2004) Human MicroRNA targets. *PLoS Biol* 2: e363.
69. Lai EC (2004) Predicting and validating microRNA targets. *Genome Biol* 5: 115.
70. Grün D, Wang YL, Langenberger D, Gunsalus KC, Rajewsky N (2005) microRNA target predictions across seven *Drosophila* species and comparison to mammalian targets. *PLoS Comput Biol* 1: e13.
71. Chandra V, Girijadevi R, Nair AS, Pillai SS, Pillai RM (2010) MTar: a computational microRNA target prediction architecture for human transcriptome. *BMC Bioinformatics* 11 Suppl 1: S2.

72. Lorenz R, Bernhart SH, Höner Zu Siederdisen C, Tafer H, Flamm C, et al. (2011) ViennaRNA Package 2.0. *Algorithms Mol Biol* 6: 26.
73. Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, et al. (2005) Combinatorial microRNA target predictions. *Nat Genet* 37: 495-500.
74. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB (2003) Prediction of mammalian microRNA targets. *Cell* 115: 787-798.
75. Enright AJ, John B, Gaul U, Tuschl T, Sander C, et al. (2003) MicroRNA targets in *Drosophila*. *Genome Biol* 5: R1.
76. Betel D, Wilson M, Gabow A, Marks DS, Sander C (2008) The microRNA.org resource: targets and expression. *Nucleic Acids Res* 36: D149-153.
77. Kiriakidou M, Nelson PT, Kouranov A, Fitziev P, Bouyioukos C, et al. (2004) A combined computational-experimental approach predicts human microRNA targets. *Genes Dev* 18: 1165-1178.
78. Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, et al. (2006) A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* 126: 1203-1217.
79. Loher P, Rigoutsos I (2012) Interactive exploration of RNA22 microRNA target predictions. *Bioinformatics* 28: 3322-3323.
80. Thadani R, Tammi MT (2006) MicroTar: predicting microRNA targets from RNA duplexes. *BMC Bioinformatics* 7 Suppl 5: S20.
81. Yousef M, Jung S, Kossenkov AV, Showe LC, Showe MK (2007) Naïve Bayes for microRNA target predictions—machine learning for microRNA targets. *Bioinformatics* 23: 2987-2992.
82. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E (2007) The role of site accessibility in microRNA target recognition. *Nat Genet* 39: 1278-1284.
83. Nielsen CB, Shomron N, Sandberg R, Hornstein E, Kitzman J, et al. (2007) Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA* 13: 1894-1910.
84. Wang X, El Naqa IM (2008) Prediction of both conserved and nonconserved microRNA targets in animals. *Bioinformatics* 24: 325-332.
85. Wang X (2008) miRDB: a microRNA target prediction and functional annotation database with a wiki interface. *RNA* 14: 1012-1017.
86. Yang Y, Wang YP, Li KB (2008) MiRTif: a support vector machine-based microRNA target interaction filter. *BMC Bioinformatics* 9 Suppl 12: S4.
87. Sturm M, Hackenberg M, Langenberger D, Frishman D (2010) TargetSpy: a supervised machine learning approach for microRNA target prediction. *BMC Bioinformatics* 11: 292.
88. Dai X, Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res* 39: W155-159.
89. Hsu JB, Chiu CM, Hsu SD, Huang WY, Chien CH, et al. (2011) miRTar: an integrated system for identifying miRNA-target interactions in human. *BMC Bioinformatics* 12: 300.
90. Coronello C, Hartmaier R, Arora A, Huleihel L, Pandit KV, et al. (2012) Novel modeling of combinatorial miRNA targeting identifies SNP with potential role in bone density. *PLoS Comput Biol* 8: e1002830.
91. Li JH, Liu S, Zhou H, Qu LH, Yang JH (2014) starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res* 42: D92-97.
92. Thomson DW, Bracken CP, Goodall GJ (2011) Experimental strategies for microRNA target identification. *Nucleic Acids Res* 39: 6845-6853.
93. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP (2011) A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 146: 353-358.
94. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* 34: D140-144.
95. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res* 36: D154-158.
96. Sethupathy P, Corda B, Hatzigeorgiou AG (2006) TarBase: A comprehensive database of experimentally supported animal microRNA targets. *RNA* 12: 192-197.
97. Hsu SD, Lin FM, Wu WY, Liang C, Huang WC, et al. (2011) miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res* 39: D163-169.
98. Zhang Z, Yu J, Li D, Zhang Z, Liu F, et al. (2010) PMRD: plant microRNA database. *Nucleic Acids Res* 38: D806-813.
99. Hsu SD, Chu CH, Tsou AP, Chen SJ, Chen HC, et al. (2008) miRNAMap 2.0: genomic maps of microRNAs in metazoan genomes. *Nucleic Acids Res* 36: D165-169.
100. Hsu PW, Lin LZ, Hsu SD, Hsu JB, Huang HD (2007) ViTa: prediction of host microRNAs targets on viruses. *Nucleic Acids Res* 35: D381-385.
101. Li SC, Shiau CK, Lin WC (2008) Vir-Mir db: prediction of viral microRNA candidate hairpins. *Nucleic Acids Res* 36: D184-189.
102. Xiao F, Zuo Z, Cai G, Kang S, Gao X, et al. (2009) miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res* 37: D105-110.
103. Jiang Q, Wang Y, Hao Y, Juan L, Teng M, et al. (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res* 37: D98-104.
104. Ruepp A, Kowarsch A, Schmid D, Buggenthin F, Brauner B, et al. (2010) PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes. *Genome Biol* 11: R6.
105. Vlachos IS, Kostoulas N, Vergoulis T, Georgakilas G, Reczko M, et al. (2012) DIANA miRPath v.2.0: investigating the combinatorial effect of microRNAs in pathways. *Nucleic Acids Res* 40: W498-504.