11

MicroRNA Identification Based on Bioinformatics Approaches

Malik Yousef[1], Naim Najami[1,2] and Walid Khaleifa[1]
[1]*The Galilee Society Institute of Applied Research,*
[2]*Department of Biology, The Academic Arab College of Education, Haifa, Israel*

1. Introduction

One of the most fascinating aspects of RNA interference (RNAi) is the non-cell-autonomous nature of silencing. Seminal studies on RNAi focused on the ability of transgene silencing to propagate systemically throughout an organism, such as from a single Agrobacterium infiltrated leaf to other parts of the plant, or from a grafted silenced stock into a non-silenced scion[1, 2]

The discovery of RNAi was preceded first by observations of transcriptional inhibition by antisense RNA expressed in transgenic plants[3] and more directly by reports of unexpected outcomes in experiments performed by plant scientists in the U.S. and The Netherlands in the early 1990s[4] In an attempt to alter flower colors in petunias, researchers introduced additional copies of a gene encoding chalcone synthase, a key enzyme for flower pigmentation into petunia plants of normally pink or violet flower color. Soon after, a related event termed *quelling* was noted in the fungus *Neurospora crassa* [5], although it was not immediately recognized as related. Further investigation of the phenomenon in plants indicated that the downregulation was due to post-transcriptional inhibition of gene expression via an increased rate of mRNA degradation[6]. This phenomenon was called *co-suppression of gene expression*, but the molecular mechanism remained unknown.

Not long after, plant virologists working on improving plant resistance to viral diseases observed a similar unexpected phenomenon. While it was known that plants expressing virus-specific proteins showed enhanced tolerance or resistance to viral infection, it was not expected that plants carrying only short, non-coding regions of viral RNA sequences would show similar levels of protection. Researchers believed that viral RNA produced by transgenes could also inhibit viral replication[7]. The reverse experiment, in which short sequences of plant genes were introduced into viruses, showed that the targeted gene was suppressed in an infected plant. This phenomenon was labeled "virus-induced gene silencing" (VIGS), and the set of such phenomena were collectively called post transcriptional gene silencing [8][15].

The spread of RNA silencing is not limited to plants or viruses: the first reported experiments of RNAi in *Caenorhabditis elegans* (*C. elegans*) demonstrated a systemic silencing response induced by locally injected or ingested double-stranded RNA (dsRNA) molecules[9, 10]. In plants, as in *C. elegans*, the systemic silencing signal acts in a sequence-specific manner, invoking the involvement of an RNA component. Sequence-specific RNA

silencing that acts non-cellautonomously has tremendous implications, not only practically as an experimental tool but in biological processes as well. The long-distance movement of RNA silencing through the vasculature forms a crucial component of the antiviral defence system and has been implicated in microRNA (miRNA)-regulated stress responses[11] [12, 13] RNA dependent gene silencing can also move from cell to cell to elicit short-range signaling responses, such as in the patterning of leaves and roots[14, 15].

After these initial observations in plants, many laboratories around the world searched for the occurrence of this phenomenon in other organisms[16] [16]. Craig C. Mello and Andrew Fire's 1998 *Nature* paper reported a potent gene silencing effect after injecting double stranded RNA into *C. elegans* [9]. In investigating the regulation of muscle protein production, they observed that neither mRNA nor antisense RNA injections had an effect on protein production, but double-stranded RNA successfully silenced the targeted gene. Fire and Mello's discovery was particularly notable because it represented the first identification of the causative agent of a previously inexplicable phenomenon. Fire and Mello were awarded the Nobel Prize in Physiology or Medicine in 2006 for their work.

MicroRNAs are the most thoroughly characterized. These single-stranded RNAs are typically 19 to 25 nucleotides in length and are thought to regulate gene expression post-transcriptionally by binding to the 3′ untranslated regions (UTRs) of target mRNAs, inhibiting their translation[17]. Recent experimental evidence suggests that the number of unique miRNAs in humans could exceed 800 [18], though several groups have hypothesized that there may be up to 20,000[19] [20] noncoding RNAs that contribute to eukaryotic complexity.

RNA polymerase II transcribes miRNA genes, generating long primary transcripts (pri-miRNAs) that are processed by the RNase III–type enzyme Drosha, yielding hairpin structures (pre-miRNAs). Pre-miRNA hairpins are exported to the cytoplasm where they are further processed into unstable miRNA duplexes by the RNase III protein Dicer. The less stable of the two strands in the duplex is incorporated into a multiple-protein nuclease complex, the RNA-induced silencing complex (RISC), which regulates protein expression. In mammalian cells, these RISCs, guided by the miRNA, interact with the 3′ UTR of target mRNAs at regions exhibiting imperfect sequence homology, inhibiting protein synthesis by a mechanism that has yet to be fully elucidated.

Although hundreds of miRNAs have been discovered in a variety of organisms, little is known about their cellular function. Several unique physical attributes of miRNAs, including their small size, lack of polyadenylated tails, and tendency to bind their mRNA targets with imperfect sequence homology, have made them elusive and challenging to study.

Endogenously expressed miRNAs, including both intronic and intergenic miRNAs, are most important in translational repression and in the regulation of development, especially the timing of morphogenesis and the maintenance of undifferentiated or incompletely differentiated cell types such as stem cells[21]. The role of endogenously expressed miRNA in downregulating gene expression was first described in *C. elegans* in 1993 [25]. In plants this function was discovered when the "JAW microRNA" of *Arabidopsis* was shown to be involved in the regulation of several genes that control plant shape[22]. In plants, the majority of genes regulated by miRNAs are transcription factors [23]; thus miRNA activity is particularly wide-ranging and regulated entire gene networks during development by modulating the expression of key regulatory genes, including transcription factors as well as F-box_proteins[24]. In many organisms, including humans, miRNAs disruption have also been linked to the formation of tumors and dysregulation of the cell cycle. Here, miRNAs

can function as both oncogenes and tumor suppressors[25]. Another example, miRNAs are aberrantly expressed in: liver, pancreatic, oesophageal, stomach, colon, haematopoietic, ovarian, breast, pituitary, prostate, thyroid, testicular and brain cancers[26] [27] [28] [29] [30]; central nervous system disorders (e.g. schizophrenia and Alzheimer's disease) [31]; and cardiovascular disease[32] [33][36,37].

It is becoming clear that a comprehensive understanding of human biology must include both small and large non-coding RNAs, and that it is perhaps only through inclusion of these elements in the biomedical research agenda, including studies to determine the mechanistic basis of the causative variations identified by genome-wide association studies, that complex human diseases will be completely deciphered.

2. Computational methods

The discovery that microRNAs are synthesized as hairpin-containing precursors with many shared features has stimulated the development of several computational approaches to the discovery of new microRNA genes in various animal species. Many of these approaches rely heavily on conservation of sequence within and between species, while others emphasize machine learning methods to screen hairpin candidates for structural features shared by known microRNA precursors. The identification of animal microRNA targets is a particularly difficult problem because an exact match to the target sequence is not required. We discuss the most recently devised algorithms for microRNA and target discovery.

2.1 Machine learning approaches to miRNA discovery

Methods derived from the machine learning field have recently been applied to miRNA discovery with good success. Machine mearning depends on the development of algorithms and methods that allow a specific computer program to *learn* from data already collected on verified miRNAs. These algorithms require a training set for the learning process that consists of positive examples (that define the miRNA characteristics) and negative examples (the control set of non-miRNA sequences). The known microRNAs used as positive examples can be downloaded from the database miRBase [34, 35] and random sequences can be one choice of negative set. One of the most important tasks associated with the learning process is the identification of characteristics and the definition of the rules that define the positive class. This is especially important in this case as these characteristics are not always explicitly defined, Readers who wish to pursue machine learning in greater detail may consult a recent review [36].

Examples of supervised machine learning algorithms include, naïve Bayes, support vector machines (SVM), hidden Markov models (HMM), neural networks and the k-nearest neighbor algorithm. Naïve Bayes is a classification model obtained by applying a relatively simple method to a training dataset [37] A Naïve Bayes classifier calculates the probability that a given instance (example) belongs to a certain class. Support Vector Machines (SVMs) are widely used machine learning algorithms developed by Vapnik [38]. In this technique, the numbers describing each feature of a microRNA are combined into a single vector in an n-dimensional space. The algorithm compares the vectors from the positive class with those from the negative class, and finds a "hyperplane" which produces the best separation (margin) between the two classes. The "support vectors" are the samples from the two classes which are closest together but still separable--they "support" the separating

hyperplane, (See Figure 1). The performance of this algorithm, as compared to other algorithms, has proven to be particularly useful for the analysis of various classification problems, particularly when the two classes are closely related or non-uniform, and has recently been widely used in the bioinformatics field [39, 40].

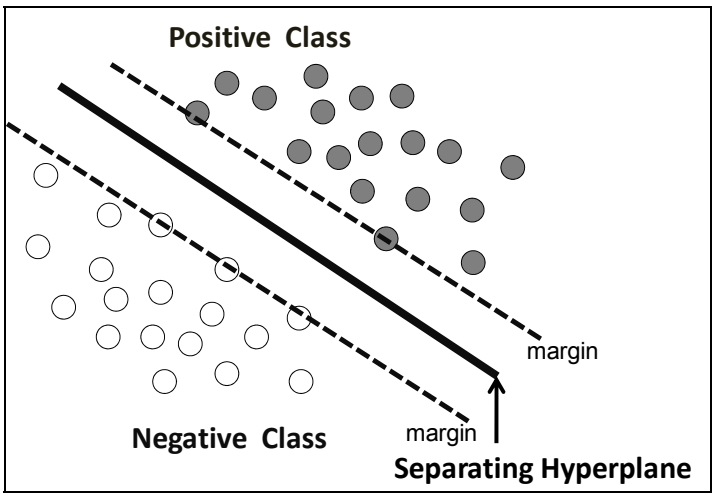


Fig. 1. The solid line is the Separating Hyperplane and the dashed lines are the margins for a SVM trained with samples from two classes. Samples (point) on the margin are called the support vectors

2.2 MicroRNA discovery tools

Numerous computational approaches (in addition to machine learning) have been implemented for miRNA gene prediction using methods based on sequence conservation and/or structural similarity[41]; [42],[43] ; [44]; [45]. Some of these tools are listed in Table 1. Lim and others [41] developed a program for identification of miRNAs, called MiRscan, with 70% specificity at a sensitivity of 50%. MiRscan uses seven miRNA features with associated weights to build a computational tool, which assigns scores to hairpin candidates. The weights are estimated using statistics based on the previously known miRNAs from *C.elegans*. Grad, et al., (2003), developed a computational method using sequence conservation and structural similarity to predict miRNAs in the *C.elegans* genome. Lai, et al., (2003) used similar ideas to develop a different computational tool for the Drosophila genome, called miRseeker. These efforts were previously reviewed by Bartel [46]. Others have used homology searches for revealing paralog and ortholog miRNAs ([42]; [47]; [48]; [49]; [50]). Additionally, Wang and others[51] developed a method based on sequence and structure alignment for miRNA identification.

ProMiR [52] is based on machine learning for miRNA discovery. ProMiR uses a highly specific probabilistic model (HMM) whose topology and states are handcrafted based on prior knowledge and assumptions, and whose exact probabilities are derived from the accumulated data. Pfeffer, et al., (2005) used support vector machines (SVMs) for predicting conserved miRNAs in herpesviruses. The features that defined the positive class were extracted from the sequence and structure features in the stem loop to form the

positive class. The negative class was generated from mRNAs, rRNAs, or tRNAs from human and viral genomes which should not include any miRNA sequences. The same approach was also applied to analysis of clustered miRNAs [53] using a tool named mir-abela, while Xue, et al.,(2005) developed a SVM classifier as a 2-class tool that does not rely on comparative genomic approaches. They defined a negative class called pseudo pre-miRNAs. The criteria for this negative class included a minimum of 18 paired bases, a maximum of -15 kcal/mol folding free energy and no multiple loops. The tool is called triplet-SVM. BayesMiRNAfind [54] is a machine learning approach based on the Naïve Bayes classifier for predicting miRNA genes. This method differs from previous efforts in two ways: 1) they generate the model automatically and identify rules based on the miRNA gene structure and sequence, allowing prediction of non-conserved miRNAs and 2) they use a comparative analysis over multiple species to reduce the false positive rate. This allows for a trade-off between sensitivity and specificity. The resulting algorithm demonstrates higher specificity and similar sensitivity to algorithms that use conserved genomic regions to reduce false positives [41, 43-45]. Grundhoff, et al.,(2006) have developed an approach to identify miRNAs that is based on bioinformatics and array-based technologies. The bioinformatics tool, VMir [55], does not rely on evolutionary sequence conservation. RNAmicro [56] is another miRNA prediction tool developed by Hertel and Stadler that relies mainly on comparative sequence analysis rather then structural features using two-class SVM.

Sheng, et al.,(2007) describe a computational method, mirCoS [57], that applies three support vector machine models, based on sequence, secondary structure, and conservation, sequentially to discover new conserved miRNA candidates in mammalian genomes.

Defining the negative class is a major challenge in developing machine learning algorithms for miRNA discovery. Two machine learning approaches have recently appeared for identifying microRNAs without the necessity of defining a negative class. Yousef, et al., (2008) presented a study using one-class machine learning for microRNA using only positive data to build the classifier (One-ClassMirnaFind [58]). Several different classifiers, including two classes SVM were used to compare the one-class approach to the corresponding two-class methods. Although the two-class procedure was generally found to be superior, it was more complex to implement.

Xu, et al., (2008) recently developed a tool called miRank. MiRank [59] is a novel ranking algorithm based on a random walk through a graph consisting of known miRNA examples and unknown candidate sequences. Each miRNA is a vertex connected to its neighbor by an edge which is weighted by its similarity of the miRNA features. The score or *relevance* of a vertex increases with the number of its connections. The vertices are then ranked by relevance score, and an arbitrary cutoff of the ranked list includes both the positive examples and the most similar of the predicted unknowns. The strength of miRank is its ability to identify novel miRNAs in newly sequenced genomes where there are few annotated miRNAs (positive examples). The authors found miRank to be superior to SVM classifiers, and attribute its success to the fact that it structures the list and ranks the candidate examples as well as the query sequences during the training and classification steps.

We should note in passing that high-throughput methods for sequencing isolated small RNAs provide a new tool for discovering new microRNA species [60] and a new method for amplifying low-concentration microRNAs allows easier testing of predictions [61] .

Algorithm	Web link	References
MiRseeker		Lai et al., 2003
MiRscan	http://genes.mit.edu/mirscan/	Lim et al., 2003a,b
miRank	*MiRank is programmed in Matlab*	Xu, et al.,2008
ProMiR II	http://cbit.snu.ac.kr/~ProMiR2/	Nam et al., 2005
PalGrade		Bentwich et al., 2005
mir-abela	http://www.mirz.unibas.ch/cgi/pred_miRNA_genes.cgi	Sewer et al., 2005
triplet-SVM	http://bioinfo.au.tsinghua.edu.cn/mirnasvm/	Xue, et al., 2005
Vmir	*http://www.hpi-hamburg.de/fileadmin/downloads/VMir.zip*	Grundhoff et al., 2006
RNA micro	http://www.bioinf.uni-leipzig.de/~jana/software/index.html	Hertel and Stadler 2006
mirCoS	Based on LIBSVM library package [62]	Sheng et al., 2007
BayesMiRNAfind	https://bioinfo.wistar.upenn.edu/miRNA/miRNA/login.php	Yousef et al., 2006,
One-ClassMirnaFind	http://wotan.wistar.upenn.edu/OneClassmiRNA/	Yousef et al., 2008

Table 1. Summary information about computational tools for miRNA predictions.

3. Target identification

Although recent findings [63] suggest MicroRNAs may affect gene expression by binding to either 5′ or 3′ untranslated regions of messenger RNA, most studies have found that microRNA mark their target mRNAs for degradation or suppress their translation by binding to the 3′-untranslated region (3′UTR) and most target programs search there . These studies have suggested that the microRNA seed segment which includes 6-8 nucleotides at the 5′ end of the mature miRNA sequence is very important in the selection of the target site (see Figure 2). Thus, most of the computational tools developed to identify mRNA target sequences depend heavily on complementarity between the miRNA seed sequence and the target sequence. Diana-microT [64] was one of the first computational tools for target prediction that identified specific interaction rules based on bioinformatics and experimental approaches. The tool successfully recovered all validated *C. elegans* miRNA targets

Several additional methods for the prediction of miRNA targets have been subsequently developed. These methods mainly use sequence complementarities, thermodynamic stability calculations, and evolutionary conservation among species to determine the likelihood of a productive miRNA:mRNA duplex formation [46, 65]. John et al., (2004) developed the miRanda [66] algorithm for miRNA target prediction. MiRanda uses dynamic programming to search for optimal sequence complementarities between a set of mature microRNAs and a given mRNA MicroRNA.org (http://www.microrna.org) [67] is a comprehensive resource of microRNA target predictions and miRNA expression profiles. Target predictions are based on the miRanda algorithm while miRNA expression profiles are derived from a comprehensive sequencing project of a large set of mammalian tissues and cell lines of normal and disease origin. Another algorithm RNAhybrid [68] [69] is similar to a RNA secondary structure prediction algorithm like the Mfold program [70] but it determines the most favorable hybridization site between two sequences.

Bennecke and others [71] have recently suggested that the 3′ out-seed segment of the miRNA:mRNA duplex can compensate for imperfect base pairing of the target with the seed segment and a recent computational approach [72] has considered the contributions of both seed and the out-seed miRNA segments in target identification. Using sequence

conservation reduces false positive predictions but as a result some less-conserved target-sites may be missed. This presents a dilemma, which is how to avoid rejection of these less highly conserved target sites while still reducing the very large numbers of predictions that are found when seed region conservation in the target is not required. In order to reduce the false positive predictions inherent in methods that heavily weight specific target sequence conservation, Lewis, et al.,(2005) developed TargetScanS [73]. TargetScanS scores target sites based on the conservation of the target sequences between five genomes (human, mouse, rat, dog and chicken) as evolutionarily conserved target sequences are more likely to be true targets. In testing, TargetScanS was able to recover targets for all 5300 human genes known at the time to be targeted by miRNAs.

PicTar [74] is a computational method to detect common miRNA targets in vertebrates, *C. elegans*, and *Drosophila*. PicTar is based on a statistical method applied to eight vertebrate genome-wide alignments (multiple alignments of orthologous nucleotide sequences (3' UTRs)). PicTar was able to recover validated miRNA targets at an estimated 30% false-positive rate. In a separate study PicTar was applied to target identification in *Drosophila melanogaster* [75] . These studies suggest that one miRNA can target 54 genes on average and that known microRNAs are projected to regulate a large fraction of all *D. melanogaster* genes (15%). This is likely to be a conservative estimate due to the incomplete input data.

TargetBoost [76] is a machine learning algorithm for miRNA target prediction using only sequence information to create weighted sequence motifs that capture the binding characteristics between microRNAs and their targets. The authors suggest that TargetBoost is stable and identifies more of the already verified true targets than do other existing algorithms.

Sung-Kyu, et al., (2005), also reported the development of a machine learning algorithm using SVM. The best reported results [77] were 0.921 sensitivity and 0.833 specificity. More recent Yan and others, used a machine learning approach that employs features extracted from both the seed and out-seed segments [72]. The best result obtained was an accuracy of 82.95% but it was generated using only 48 positive human and 16 negative examples, a relatively small training set to assess the algorithm.

In 2006, Thadani and Tammi [78] launched MicroTar, a novel statistical computational tool for prediction of miRNA targets from RNA duplexes which does not use sequence homology for prediction. MicroTar mainly relies on a quite novel approach to estimate the duplex energy. However, the reported sensitivity (60%) is significantly lower than that achieved using other published algorithms. At the same time, a microRNA pattern discovery method, RNA22 [79] was proposed to scan UTR sequences for targets . RNA22 does not rely upon cross-species conservation but was able to recover most of the known target sites with validation of some of its new predictions.

More recently, Yousef, et al.,(2007) described a target prediction method, (NBmiRTar [80]) using instead machine learning by a Naïve Bayes classifier. NBmiRTar does not require sequence conservation but generates a model from sequence and miRNA:mRNA duplex information derived from validated target sequences and artificially generated negative examples. In this case, both the seed and "out-seed" segments of the miRNA:mRNA duplex are used for target identification. NBmiRTar technique produces fewer false positive predictions and fewer target candidates to be tested than miRanda [66]. It exhibits higher sensitivity and specificity than algorithms that rely only on conserved genomic regions to decrease false positive predictions.

Algorithm	Web link	References
TargetScanS	http://genes.mit.edu/targetscan	Lewis, et al., 2005
miRanda	http://www.microma.org	John, et al., 2004
PicTar	http://pictar.bio.nyu.edu	Krek, et al. 2005
RNAhybrid	http://bibiserv.techfak.uni-bielefeld.de/rnahybrid	Rehmsmeier, et al., 2004
Diana-microT	http://www.diana.pcbi.upenn.edu/cgi-bin/micro_t.cgi	Kiriakidou, et al. 2004
Target Boost	https://demo1.interagon.com/demo	SaeTrom, et al. 2005
Rna22	http://cbcsrv.watson.ibm.com/rna22_targets.html	Miranda, et al. 2006
MicroTar	http://tiger.dbs.nus.edu.sg/microtar/	Thadani and Tammi 2006
NBmiRTar	http://wotan.wistar.upenn.edu/NBmiRTar	Yousef, et al. 2007
miRecords	http://mirecords.umn.edu/miRecords/	Xiao, et al., 2009

Table 2. MicroRNA Target prediction tools

In a 2004 review Lai [65] noted that there is almost no overlap among the predicted targets identified by the various methods and suggested that each tool captures a subset of the entire target class as a function of the specific features they have incorporated into their prediction models. More recently, Sethupathy, et al., (2006) conducted a comparison of the 5 most used tools for mammalian target prediction. This study indicated that 30% of the experimentally validated target sites are nonconserved, supporting the need for the development of different or complementary computational approaches to capture new target sites. Furthermore, the large number of predictions that each of these tools is producing suggests that the heavy reliance on homology or comparative sequence analysis is not sufficient to generate accurate predictions with a high sensitivity and there are yet to be identified recognition parameters that must be considered.

4. Databases for microRNA and targets

There is a variety of very useful databases that provide a significant amount of information on miRNA and Target predictions,(Table 3). The most extensive database for both miRNA and target sequences is miRBase[34]. MiRBase contains both miRNA mature sequences, hairpin sequences of precursors and associated annotation. Release 12.0 of the database contains 8619 entries representing hairpin precursor miRNAs, expressing 8273 mature miRNA products, in primates, rodents, birds, fish, worms, flies, plants and viruses. MiRBase also contains predicted miRNA target genes in miRBase Targets, and provides a gene naming and nomenclature function in the miRBase Registry. The miRNA target genes are predicted by the miRanda tool [66] and not necessarily experimentally validated.

TarBase [81] contains a set of experimentally supported targets in different species that are collected manually from the literature. TarBase version 5 has more than 1300 experimentally supported miRNA target interactions. The database has information about the target site described by the duplex of miRNA and gene. It also includes information on the experiments that were conducted to test the target, the sufficiency of the site to induce translational repression and/or cleavage, and a reference to the paper used to extract the information.

Argonaute [82] is a compilation of comprehensive information on mammalian miRNAs, their origin and regulated target genes in an exhaustively curated database. The source information of Argonaute is from both literature and other databases.

The most recently released database, miRecords [83], is an integrated resource for animal miRNA–target interactions. miRecords stores predicted miRNA targets produced by 11 established miRNA target prediction programs.

DataBase	Web Link
MiRBase	http://microrna.sanger.ac.uk/
TarBase	http://diana.cslab.ece.ntua.gr/tarbase/
Argonaute	http://www.ma.uni-heidelberg.de/apps/zmf/argonaute/
miRecords	http://mirecords.umn.edu/miRecords/

Table 3. Databases for microRNA and Targets

5. References

[1] Voinnet, O. and D.C. Baulcombe, *Systemic signalling in gene silencing.* Nature, 1997. 389(6651): p. 553-553.

[2] Palauqui, J.-C., et al., *Systemic acquired silencing: transgene-specific post-transcriptional silencing is transmitted by grafting from silenced stocks to non-silenced scions.* EMBO J, 1997. 16(15): p. 4738-4745.

[3] Napoli, C., C. Lemieux, and R. Jorgensen, *Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans.* The Plant Cell Online, 1990. 2(4): p. 279-289.

[4] Romano, N. and G. Macino, *Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences.* Molecular Microbiology, 1992. 6(22): p. 3343-3353.

[5] Van Blokland, R., et al., *Transgene-mediated suppression of chalcone synthase expression in Petunia hybrida results from an increase in RNA turnover.* The Plant Journal, 1994. 6(6): p. 861-877.

[6] Covey, S.N., et al., *Plants combat infection by gene silencing.* Nature, 1997. 385(6619): p. 781-782.

[7] Ratcliff, F., B.D. Harrison, and D.C. Baulcombe, *A Similarity Between Viral Defense and Gene Silencing in Plants.* Science, 1997. 276(5318): p. 1558-1560.

[8] Guo, S. and K.J. Kemphues, *par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed.* Cell, 1995. 81(4): p. 611-620.

[9] Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.* Nature, 1998. 391(6669): p. 806-811.

[10] Timmons, L. and A. Fire, *Specific interference by ingested dsRNA.* Nature, 1998. 395(6705): p. 854-854.

[11] Schwach, F., et al., *An RNA-Dependent RNA Polymerase Prevents Meristem Invasion by Potato Virus X and Is Required for the Activity But Not the Production of a Systemic Silencing Signal.* Plant Physiology.

[12] Pant, B.D., et al., *MicroRNA399 is a long-distance signal for the regulation of plant phosphate homeostasis.* The Plant Journal, 2008. 53(5): p. 731-738.

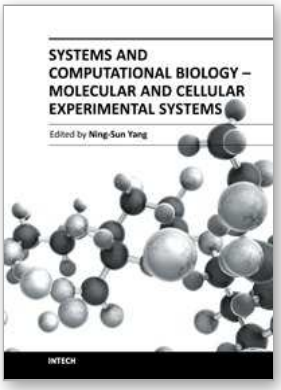
[13] Buhtz, A., et al., *Identification and characterization of small RNAs from the phloem of Brassica napus.* The Plant Journal, 2008. 53(5): p. 739-749.

[14] Chitwood, D.H., et al., *Pattern formation via small RNA mobility.* Genes & Development, 2009. 23(5): p. 549-554.

[15] Ecker, J.R. and R.W. Davis, *Inhibition of gene expression in plant cells by expression of antisense RNA.* Proceedings of the National Academy of Sciences, 1986. 83(15): p. 5372-5376.
[16] Pal-Bhadra, M., U. Bhadra, and J.A. Birchler, *Cosuppression in Drosophila: Gene Silencing of Alcohol dehydrogenase by white-Adh Transgenes Is Polycomb Dependent.* Cell, 1997. 90(3): p. 479-490.
[17] Ambros, V., *The functions of animal microRNAs.* Nature, 2004. 431(7006): p. 350-355.
[18] Bentwich, I., et al., *Identification of hundreds of conserved and nonconserved human microRNAs.* Nat Genet, 2005. 37(7): p. 766-770.
[19] Okazaki, Y.e.a., *Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.* Nature, 2002. 420(6915): p. 563-573.
[20] Imanishi, T., et al., *Integrative Annotation of 21,037 Human Genes Validated by Full-Length cDNA Clones.* PLoS Biol, 2004. 2(6): p. e162.
[21] Carrington, J.C. and V. Ambros, *Role of MicroRNAs in Plant and Animal Development.* Science, 2003. 301(5631): p. 336-338.
[22] Lee, R.C., R.L. Feinbaum, and V. Ambros, *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14.* Cell, 1993. 75(5): p. 843-854.
[23] Zhang, B., et al., *Plant microRNA: A small regulatory molecule with big impact.* Developmental Biology, 2006. 289(1): p. 3-16.
[24] Jones-Rhoades, M.W., D.P. Bartel, and B. Bartel, *MicroRNAs AND THEIR REGULATORY ROLES IN PLANTS.* Annual Review of Plant Biology, 2006. 57(1): p. 19-53.
[25] Zhang, B., et al., *microRNAs as oncogenes and tumor suppressors.* Developmental Biology, 2007. 302(1): p. 1-12.
[26] Visone, R. and C.M. Croce, *MiRNAs and Cancer.* Am J Pathol, 2009. 174(4): p. 1131-1138.
[27] Pang, J., et al., *Oncogenic role of microRNAs in brain tumors.* Acta Neuropathologica, 2009. 117(6): p. 599-611.
[28] Voorhoeve, P.M., et al., *A Genetic Screen Implicates miRNA-372 and miRNA-373 As Oncogenes in Testicular Germ Cell Tumors.* Cell, 2006. 124(6): p. 1169-1181.
[29] Khoshnaw, S.M., et al., *MicroRNA involvement in the pathogenesis and management of breast cancer.* Journal of Clinical Pathology, 2009. 62(5): p. 422-428.
[30] Novakova, J., et al., *MicroRNA involvement in glioblastoma pathogenesis.* Biochemical and Biophysical Research Communications, 2009. 386(1): p. 1-5.
[31] Kocerha, J., S. Kauppinen, and C. Wahlestedt, *microRNAs in CNS Disorders.* NeuroMolecular Medicine, 2009. 11(3): p. 162-172.
[32] Barringhaus, K.G. and P.D. Zamore, *MicroRNAs: Regulating a Change of Heart.* Circulation, 2009. 119(16): p. 2217-2224.
[33] Sen, C.K., et al., *Micromanaging Vascular Biology: Tiny MicroRNAs Play Big Band.* Journal of Vascular Research, 2009. 46(6): p. 527-540.
[34] Griffiths-Jones, S., et al., *miRBase: tools for microRNA genomics.* Nucl. Acids Res., 2008. 36(suppl_1): p. D154-158.
[35] Griffiths-Jones, S., *The microRNA Registry.* Nucleic Acids Res, 2004. 32(90001): p. D109-111.
[36] Larranaga, P., et al., *Machine learning in bioinformatics.* Brief Bioinform, 2006. 7(1): p. 86-112.
[37] Mitchell, T., *Machine Learning*1997: McGraw Hill.
[38] Vapnik, V., *The Nature of Statistical Learning Theory*1995: Springer.

[39] Haussler, D., *Convolution kernels on discrete structures*, 1999, Baskin School of Engineering, University of California: Santa Cruz. p. Technical Report UCSCCRL - 99-10.
[40] Pavlidis, P., et al. *Gene functional classification from heterogeneous data* in *Proceedings of the fifth annual international conference on Computational biology* 2001 Montreal, Quebec, Canada ACM Press.
[41] Lim, L.P., et al., *Vertebrate MicroRNA Genes.* Science, 2003. 299(5612): p. 1540.
[42] Weber, M.J., *New human and mouse microRNA genes found by homology search.* FEBS Journal, 2005. 272(1): p. 59-73.
[43] Lim, L.P., et al., *The microRNAs of Caenorhabditis elegans.* Genes Dev., 2003. 17(8): p. 991-1008.
[44] Lai, E., et al., *Computational identification of Drosophila microRNA genes.* Genome Biology, 2003. 4(7): p. R42.
[45] Grad, Y., et al., *Computational and Experimental Identification of C. elegans microRNAs.* Molecular Cell, 2003. 11(5): p. 1253-1263.
[46] Bartel, D.P., *MicroRNAs: Genomics, Biogenesis, Mechanism, and Function.* Cell, 2004. 116(2): p. 281-297.
[47] Lagos-Quintana, M., et al., *Identification of Novel Genes Coding for Small Expressed RNAs.* Science, 2001. 294(5543): p. 853-858.
[48] Lau, N.C., et al., *An Abundant Class of Tiny RNAs with Probable Regulatory Roles in Caenorhabditis elegans.* Science, 2001. 294(5543): p. 858-862.
[49] Lee, R.C. and V. Ambros, *An Extensive Class of Small RNAs in Caenorhabditis elegans.* Science, 2001. 294(5543): p. 862-864.
[50] Pasquinelli, A.E., et al., *Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA.* Nature, 2000. 408(6808): p. 86.
[51] Wang, X., et al., *MicroRNA identification based on sequence and structure alignment.* Bioinformatics, 2005. 21(18): p. 3610-3614.
[52] Nam, J.-W., et al., *Human microRNA prediction through a probabilistic co-learning model of sequence and structure.* Nucleic Acids Res, 2005. 33(11): p. 3570-3581.
[53] Sewer, A., et al., *Identification of clustered microRNAs using an ab initio prediction method.* BMC Bioinformatics, 2005. 6(1): p. 267.
[54] Yousef, M., et al., *Combining multi-species genomic data for microRNA identification using a Naive Bayes classifier.* Bioinformatics, 2006. 22(11): p. 1325-1334.
[55] Grundhoff, A., C.S. Sullivan, and D. Ganem, *A combined computational and microarray-based approach identifies novel microRNAs encoded by human gamma-herpesviruses.* RNA, 2006. 12(5): p. 733-750.
[56] Hertel, J. and P.F. Stadler, *Hairpins in a Haystack: recognizing microRNA precursors in comparative genomics data.* Bioinformatics, 2006. 22(14): p. e197-202.
[57] Sheng, Y., P. Engstrom, G., and B. Lenhard, *Mammalian MicroRNA Prediction through a Support Vector Machine Model of Sequence and Structure.* PLoS ONE, 2007. 2(9): p. e946.
[58] Yousef, M., et al., *Learning from positive examples when the negative class is undetermined-microRNA gene identification*, 2008. p. 2.
[59] Xue, C., et al., *Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine.* BMC Bioinformatics, 2005. 6(1): p. 310.
[60] Glazov, E.A., et al., *A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach*, 2008. p. gr.074740.107.
[61] Berezikov, E., et al., *Many novel mammalian microRNA candidates identified by extensive cloning and RAKE analysis*, 2006. p. 1289-1298.

[62] Chang, C.-C. and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001.
[63] Lytle, J.R., T.A. Yario, and J.A. Steitz, *Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5²€ UTR as in the 3²€ UTR.* Proceedings of the National Academy of Sciences, 2007. 104(23): p. 9667-9672.
[64] Kiriakidou, M., et al., *A combined computational-experimental approach predicts human microRNA targets.* Genes & Development, 2004. 18(10): p. 1165-1178.
[65] Lai, E., *Predicting and validating microRNA targets.* Genome Biology, 2004. 5(9): p. 115.
[66] John, B., et al., *Human MicroRNA Targets.* PLoS Biology, 2004. 2(11): p. e363.
[67] Betel, D., et al., *The microRNA.org resource: targets and expression.* Nucl. Acids Res., 2008. 36(suppl_1): p. D149-153.
[68] Rehmsmeier, M., et al., *Fast and effective prediction of microRNA/target duplexes.* RNA, 2004. 10(10): p. 1507-1517.
[69] Kruger, J. and M. Rehmsmeier, *RNAhybrid: microRNA target prediction easy, fast and flexible.* Nucl. Acids Res., 2006. 34(suppl_2): p. W451-454.
[70] Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction.* Nucleic Acids Res, 2003. 31 (13): p. 3406-3415.
[71] Brennecke, J., et al., *Principles of microRNA-target recognition.* PLoS Biol., 2005. 3: p. e85.
[72] Yan, X., et al., *Improving the prediction of human microRNA target genes by using ensemble algorithm.* FEBS Letters, 2007. 581(8): p. 1587.
[73] Lewis, B.P., et al., *Prediction of mammalian microRNA targets.* Cell, 2003. 115: p. 787.
[74] Krek, A., et al., *Combinatorial microRNA target predictions.* Nat Genet, 2005. 37(5): p. 495-500.
[75] Grun, D., et al., *microRNA Target Predictions across Seven Drosophila Species and Comparison to Mammalian Targets.* PLoS Computational Biology, 2005. 1(1): p. e13.
[76] SaeTrom, O.L.A., O.J. Snove, and P.A.L. SaeTrom, *Weighted sequence motifs as an improved seeding step in microRNA target prediction algorithms.* RNA, 2005. 11(7): p. 995-1003.
[77] Sung-Kyu, K., et al. *A Kernel Method for MicroRNA Target Prediction Using Sensible Data and Position-Based Features*. in *Computational Intelligence in Bioinformatics and Computational Biology*. 2005. Proceedings of the 2005 IEEE Symposium
[78] Thadani, R. and M. Tammi, *MicroTar: predicting microRNA targets from RNA duplexes.* BMC Bioinformatics, 2006. 7(Suppl 5): p. S20.
[79] Miranda, K.C., et al., *A Pattern-Based Method for the Identification of MicroRNA Binding Sites and Their Corresponding Heteroduplexes.* 2006. 126(6): p. 1203-1217.
[80] Yousef, M., et al., *Naive Bayes for microRNA target predictions machine learning for microRNA targets*, 2007. p. 2987-2992.
[81] Sethupathy, P., B. Corda, and A.G. Hatzigeorgiou, *TarBase: A comprehensive database of experimentally supported animal microRNA targets.* RNA, 2006. 12(2): p. 192-197.
[82] Shahi, P., et al., *Argonaute--a database for gene regulation by mammalian microRNAs.* Nucl. Acids Res., 2006. 34(suppl_1): p. D115-118.
[83] Xiao, F., et al., *miRecords: an integrated resource for microRNA-target interactions.* Nucl. Acids Res., 2009. 37(suppl_1): p. D105-110.



Systems and Computational Biology - Molecular and Cellular Experimental Systems

Edited by Prof. Ning-Sun Yang

ISBN 978-953-307-280-7
Hard cover, 332 pages
Publisher InTech
Published online 15, September, 2011
Published in print edition September, 2011

Whereas some â€œmicroarrayâ€ or â€œbioinformaticsâ€ scientists among us may have been criticized as doing â€œcataloging researchâ€, the majority of us believe that we are sincerely exploring new scientific and technological systems to benefit human health, human food and animal feed production, and environmental protections. Indeed, we are humbled by the complexity, extent and beauty of cross-talks in various biological systems; on the other hand, we are becoming more educated and are able to start addressing honestly and skillfully the various important issues concerning translational medicine, global agriculture, and the environment. The two volumes of this book presents a series of high-quality research or review articles in a timely fashion to this emerging research field of our scientific community.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Malik Yousef, Naim Najami and Walid Khaleifa (2011). MicroRNA Identification Based on Bioinformatics Approaches, Systems and Computational Biology - Molecular and Cellular Experimental Systems, Prof. Ning-Sun Yang (Ed.), ISBN: 978-953-307-280-7, InTech, Available from: http://www.intechopen.com/books/systems-and-computational-biology-molecular-and-cellular-experimental-systems/microrna-identification-based-on-bioinformatics-approaches

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821