Computational Methods for MicroRNA Target Prediction

Hamid Hamzeiy[1], Malik Yousef[2], and Jens Allmer[1]

[1]Molecular Biology and Genetics, Izmir Institute of Technology, Gulbahce, Urla, Izmir, Turkey

[2]The Galilee Society Institute of Applied Research, Shefa Amr, Israel

Key Words

Bioinformatics, Computational Biology, miRNA, MicroRNA, Target Prediction, Machine Learning

Running Head

Computational MicroRNA Target Prediction

Summary

MicroRNAs (miRNAs) are important players in gene regulation. The final and maybe the most important step in their regulatory pathway is the targeting. Targeting is the binding of the miRNA to the mature RNA (mRNA) via the RNA Induced Silencing Complex (RISC). Expression patterns of miRNAs are highly specific in respect to external stimuli, developmental stage or tissue. This is used to diagnose diseases such as cancer in which the expression levels of miRNAs are known to change considerably. Newly identified miRNAs are increasing in number with every new release of miRBase which is the main online database providing miRNA sequences and annotation. Many of these newly identified miRNAs do not yet have identified targets. This is especially the case in animals where the miRNA does not bind to its target as perfectly as it does in plants. Valid targets need to be identified for miRNAs in order to properly understand their role in cellular pathways. Experimental methods for target validations are difficult, expensive, and time consuming. Having considered all these facts it is of crucial importance to have accurate computational miRNA target predictions. There are many proposed methods and algorithms available for predicting targets for miRNAs but only a few have been developed to become available as independent tools and software. There are also databases which collect and store information regarding predicted miRNA targets. Current approaches to miRNA target prediction produce a huge amount of false positive and an unknown amount of false negative results and thus the need for better approaches is ever more evident. This chapter aims to give some detail about the current tools and approaches used for miRNA target prediction, provides some grounds for their comparison, and outlines a possible future.

1 Introduction

Initially identified two decades ago, miRNAs are now considered to have a central role in the RNA revolution. This has focused the scientific community's attention to these small RNAs and vigorous research efforts have resulted in the accumulation of a significant body of data related to miRNA biogenesis and function. This can be seen quite clearly in the super linear increase of miRBase (Kozomara & Griffiths-Jones, 2011) entries. Most of the 17,000 miRNA sequences currently available in the miRBase database are yet to have validated targets and thus there is a clear need for ever more precise and accurate miRNA target prediction.

A single miRNA has the potential to regulate hundreds of target mRNAs and multiple miRNAs may compete for the regulation of the same mRNA (Krek et al., 2005; Lewis, Burge, & Bartel, 2005; Wu et al., 2010). Having considered this fact it is not surprising to have more target genes than miRNAs (Figure 1). TarBase 6.0 (Vergoulis et al., 2012a) currently has more than 65,000 experimentally validated miRNA targets. It is estimated that as much as 90 percent of all human genes are somewhat regulated by miRNAs (Miranda et al., 2006). On average a single miRNA family is thought to have around 300 conserved targets which would mean that a large number of mammalian genes are miRNA regulated (Friedman, Farh, Burge, & Bartel, 2009). Self-regulatory pathways for miRNA biogenesis such as the inhibition of the synthesis of the Dicer protein which has an essential role in the miRNA biosynthetic pathway have also been identified (Johnson et al., 2007; Tokumaru, Suzuki, Yamada, Nagino, & Takahashi, 2008; Xie, Kasschau, & Carrington, 2003). This auto regulatory pathway leads to the establishment of a negative feedback system which could be exploited to control miRNA expression and thus miRNA mediated regulatory pathways.

Before miRNA target prediction tools were available, possible miRNA target sites were determined manually. These target sites were later confirmed by laborious and inefficient

techniques such as site-directed mutagenesis and other experimental methods (see Chapter 14 in this volume). The identification of the first targets for the let-7 and lin-4 miRNAs led to the idea that miRNAs have a pattern in targeting genes which could be used to develop target prediction algorithms (Mazière & Enright, 2007).

Gene targeting by miRNAs is generally believed to be the result of their binding to the 3′UTR of the target mRNA. Other studies (Forman, Legesse-Miller, & Coller, 2008; Ørom, Nielsen, & Lund, 2008; Place, Li, Pookot, Noonan, & Dahiya, 2008; Reczko, Maragkakis, Alexiou, Grosse, & Hatzigeorgiou, 2012a; Tay, Zhang, Thomson, Lim, & Rigoutsos, 2008a) have also confirmed gene regulation as a result of the binding of the miRNA to the coding region (commonly seen in plants (Jones-Rhoades & Bartel, 2004a) as well as the 5′UTR. Computational evidence suggests that regulation via the binding of the miRNA to the coding region differs in comparison to the binding pattern seen at the 3′UTR (Forman et al., 2008). It is suggested that miRNAs target the coding regions of mRNAs with short 3′UTRs (Reczko, Maragkakis, Alexiou, Grosse, & Hatzigeorgiou, 2012b). 3′UTRs are prone to change under different conditions which might result in the elimination of the target site (Selbach et al., 2008). This phenomenon presents an opportunity for the cell to regulate the function of the miRNA (see Chapter 18 in this volume for more details on miRNA regulation). Binding in the coding region on the other hand may present an evolutionary advantage for the cell as it could help in the preservation of the miRNA binding site (Lytle, Yario, & Steitz, 2007). Regulation of the miRNA function on this level may also be controlled by the inclusion or exclusion of the binding site as a result of alternative splicing (Gu, Jin, Zhang, Sarnow, & Kay, 2009a; Tay, Zhang, Thomson, Lim, & Rigoutsos, 2008b).

2 MicroRNA Target Prediction

Targeting patterns are different between plants and animals. Plants show a near perfect complement between their miRNA and their target mRNA and similar to the action of siRNAs, this could cause the cleavage of the double stranded RNA (dsRNA) (Rhoades et al., 2002; Vaucheret, 2006). This makes target prediction easier in plants, in comparison to animals, reucint the targeting problem to using computational methods for sequence similarity search (Zhang, 2005). On the other hand animal miRNA bind their targets with only partial complementarity (Figure 3). A region of about 6 to 8 nucleotides in length within the structure of the miRNA which is called the seed region is of crucial importance in the targeting. This seed sequence binds to the target mRNA leading to the regulation of the gene in question (Lewis et al., 2005; Lewis, Shih, Jones-Rhoades, Bartel, & Burge, 2003). Other than the seed region, two other regions namely the extended seed region and the delta seed region are also deemed important (Grimson et al., 2007; Liu, 2008). Binding at the 3′UTR is usually preferred over binding in the coding region or the 5′UTR but the reasons are yet to be unraveled and contradictory studies have made it difficult to reach a conclusion (Gu et al., 2009a; Lytle et al., 2007; Tay et al., 2008b). Binding in the coding region is known to be effective in plants (Gu et al., 2009a; Jones-Rhoades & Bartel, 2004b), but in animals it is proposed that coding region binding is only effective where there is a high degree of complementarity (similar to plants). This may lead to the disruption of the interaction of the transcript and the ribosome and thus to the inhibition of translation (Gu, Jin, Zhang, Sarnow, & Kay, 2009b).

2.1 Target Prediction Methodologies

Several different methods and approaches are currently in use for the prediction of miRNA targets (Rajewsky, 2006; Sethupathy, Megraw, & Hatzigeorgiou, 2006). The seed region is one of the most commonly used miRNA traits for miRNA target prediction and many studies

(Doench & Sharp, 2004; Lai, 2002; Lewis et al., 2005; Rajewsky & Socci, 2004) have pointed out the importance of binding between the seed region, located at the 5′ end of the miRNA, and its target mRNA. Other characteristics of the miRNA targeting pathway which are currently used for target prediction include the binding pattern of the seed region, the minimum free energy of the binding between the miRNA and its target mRNA, and the accessibility of the target site (Du & Zamore, 2005). Other studies (Brennecke, Stark, Russell, & Cohen, 2005; Yan et al., 2007) have also looked at base pairing between the miRNA and its target outside of the seed region. They suggest that binding beyond the seed region will compensate for weak binding of the seed region. Conserved sequences around the seed region (adenines for animals in particular (Lewis et al., 2005)) may also play a role in finding targets for miRNAs in different species. Even though this approach helps to eliminate a significant amount of false positive results, it may also result in loosing targets which are less conserved. Furthermore a study (Sethupathy, Corda, & Hatzigeorgiou, 2006) suggested that at least 30% of the experimentally validated target sites are non-conserved suggesting that the conservation of the miRNA target site alone is not enough.

2.1.1 Sequence-based Methods

The first thing that comes to mind when talking about miRNA targeting is the complementarity between the miRNA and its target. The small size of the miRNA transcript in respect to the genome rules out the possibility to rely solely on sequence complementarity for target predictions. This is because such approaches produce a huge number of potentially false positive hits. Even though complementarity is very important and useful in target prediction, other properties of this interaction such as bulges and mismatches complicate matters. The seed region is the main focus when sequence-based methods are considered (Lewis et al., 2005, 2003). Most tools look at the 3′UTR of the target gene when searching for complementarity, but others have

suggested looking at the 5′UTRs and the coding regions, too. Maybe the most important step in this method is the information regarding the sequence of the genome. The 3′UTRs for many mammal genomes are not well characterized. This complicates matters when searching for miRNA targets (Hubbard, 2002) within their bounds. When the boundaries of the 3′UTR are not properly defined they can be estimated by taking the downstream flanking sequence from the stop codon with an average corresponding to the 3′UTR length. Although this may partially solve the problem of undefined 3′UTRs, it is far from the precision needed for accurate predictions.

2.1.2 Structure-based Methods

Structure-based methods focus mostly on the thermodynamic stability of the miRNA:mRNA duplex. Several different programs are available for the prediction and analysis of the secondary structure and hybridization of miRNAs including Mfold (Zuker, 2003a) and the Vienna RNA Package (I. L. Hofacker, 2003). Some target prediction algorithms (Enright et al., 2003a; Krek et al., 2005; Lewis et al., 2003; Zuker, 2003b) use these tools to check for the thermodynamic stability of the predicted duplex using sequence complementarity. Other algorithms (Kiriakidou et al., 2004; Krüger & Rehmsmeier, 2006) on the other hand rely on thermodynamics as the initial factor in target prediction.

2.1.3 Homology-based Methods

As mentioned before looking at conserved targets within different species helps to reduce the number of false positive results (Enright et al., 2003b; Lewis et al., 2003; Stark, Brennecke, Russell, & Cohen, 2003) but this may also causes an increase in the number of false negatively identified targets. Homology-based methods usually focus on the seed region (Figure 4a). The

choice of genomes to look for conservation in this approach is very important and genomes which are very similar to each other should be avoided (Figure 4b). This is because at least 99% of the transcript will be conserved and maybe it would be better if the genomes were analyzed with larger evolutionary distance in mind.

2.2 Available Tools Overview

Currently there are more than a dozen algorithms (Table 1) which claim to predict miRNA targets by applying some of the features mentioned above. Among these are tools which combine experimental and computational methods hoping to achieve better predictions. An example for this approach would be the Diana-microT (Kiriakidou et al., 2004) which claims to be able to reproduce all known *C. elegance* miRNA targets. On the other hand programs like miRanda (John et al., 2004) rely on dynamic programming to find the most optimal complementation between a given miRNA and its target mRNA, and RNA secondary structure prediction algorithms like Mfold work by finding complementary regions. PicTar (Krek et al., 2005) was developed by performing multiple sequence alignments of the 3′UTR of eight vertebrates. PicTar uses a statistical approach and is emphasizing the importance of the conservation of the miRNA target site. A different approach based only on sequence information was applied by TargetBoost (O. Saetrom, Snøve, & Saetrom, 2005) which is essentially a machine learning algorithm. This approach set a trend towards applying machine learning algorithms to miRNA target predictions and other studies (Kim, Nam, Lee, & Zhang, 2005; Yan et al., 2007; Yousef, Jung, Kossenkov, Showe, & Showe, 2007a) later used this method. MicroTar (Thadani & Tammi, 2006) is another program which does not rely on the conservation of the miRNA target; instead it predicts miRNA targets by considering RNA duplex energies. Finally RNA22 (Miranda et al., 2006) aims to find miRNA targets by searching for patterns in the 3′UTR. In the following, TargetScanS

(Lewis et al., 2003) and RNAhybrid (Krüger & Rehmsmeier, 2006; Rehmsmeier, Steffen, Hochsmann, & Giegerich, 2004) will be discussed in more detail.

2.2.1 TargetScanS

TargetScanS is introduced as an extension to the TargetScan algorithm with some new features including the addition of two more species to the three which were originally in TargetScan. It predicts miRNA targets by looking at conserved target sequences between human, mouse, dog, rat, and chicken. This helps to reduce the number of false positive results and when tested, it was able to successfully identify targets for 5,300 human genes which were known to be targeted by miRNAs. The algorithm requires perfect binding in the seed region and then looks at binding beyond the seed region. The developers came to notice that the 8^{th} nucleotide of the target is usually an Adenosine and that the 8^{th} nucleotide often formed a Watson-Crick pair in the duplex. TargetScan tested the binding sites for their thermodynamic stability using RNAfold from the Vienna RNA Package but TargetScanS does not. The absence of the thermodynamic stability measure and the requirement for several hits in the 3′UTR for each miRNA helped to reduce the runtime for TargetScanS. TargetScanS results are available via their web server (http://genes.mit.edu/tscan/targetscanS2005.html).

2.2.2 RNAhybrid

RNAhybrid aims to predict potential targets for miRNAs by looking at the most energetically favorable hybridization sites between two separate RNA sequences and does not allow base pairings between the nucleotides of either of the two molecules. This feature sets it apart from tools such as Mfold and the Vienna RNA Package as they are only able to fold a single sequence. This means that when Mfold or the Vienna RNA Package are used for target prediction a linker

sequence would have to be introduced in between the miRNA and the target mRNA sequence which could easily lead to errors in folding and thus target prediction. Another feature of RNAhybrid which sets it apart from other methods is its robust statistical modeling. RNAhybrid claims to be able to predict multiple miRNA binding sites in larger RNAs and to be easy, fast, and flexible for the prediction of microRNA targets. For target prediction in humans RNAhybrid only looks at the 3′UTRs. Figure 2 shows a typical output of the program. Several different versions of the program are available for different platforms and are available for download from the Bielefeld Bioinformatics Server (http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/). The application is simple and comes with adequate documentation.

3 Methods for Filtering of Predicted Targets

As outlined before, gene regulation by miRNAs is often achieved by their targeting of the 3′UTR region of an mRNA (Figure 5). Recently, it has been shown that, at least for cyclin D1, only 7 of 45 predicted targets could be experimentally confirmed (Jiang, Feng, & Mo, 2009). From this it can be gathered that many of the assumptions that the miRNA target prediction algorithms are based on could be improved. This is even more supported by another more recent experimental study (Wu et al., 2010). Therefore, it is advisable to use these tools for guidance rather than accepting their results as ground truths.

There are several challenges regarding miRNA target prediction among which is the fact that a gene can be targeted by multiple miRNAs. However challenging this may be, it actually provides further criteria for discriminating true and false target predictions. For instance, if several miRNA target sites are found in a 3′UTR they would confirm each other and the resulting confidence would be raised. The location of the miRNA target site within the 3′UTR can also be

used for better target prediction. The target site should not be too close to the stop codon and it should also not be in the middle of the 3′UTR due to structural reasons. Figure 5 shows a target site (t1) which is close to the translation stop and may thus not be a good target. It further is within a secondary structure and can therefore not easily be accessed by the RISC complex bound mature miRNA. Figure 5 also has two other target sites one of which is fully accessible and is therefore a valid target while the last one (t3) is only partially accessible. In this case it would be important to calculate the minimum free energy (mfe) of the miRNA:mRNA duplex and compare it with the free energy of the present structure. The target is considered valid only if the mfe of the fold is higher than the mfe of the 3′UTR's structure. Below is a list of features which can potentially be used for discriminating true miRNA targets from false positive ones:

- Strong seed region pairing with minimal mismatches,
- The miRNA:mRNA duplex free energy should be minimal,
- Conserved adenosines around the seed region for animals (Lewis et al., 2005),
- Multiplicity control for a gene increases significance (Enright et al., 2003a),
- Proximity among target sites (Grimson et al., 2007; P. Saetrom et al., 2007),
- Target site secondary structure should be accessible (Du & Zamore, 2005),
- Gene expression profiles can validate regulation (Joung & Fei, 2009),
- Capping and polyadenylation can be useful (Barbato et al., 2009).

There is also growing evidence that targeting outside the 3′UTR is more common than expected and in the future target prediction algorithms need to take this into account (Kloosterman, Wienholds, Ketting, & Plasterk, 2004; Lytle et al., 2007). It may be beneficial to combine the output of several target prediction programs (Barbato et al., 2009) but since they are largely build on the same assumptions important targets may be missed nonetheless (Peter, 2010). Since many

of the tools in target prediction are based on machine learning algorithms which learn by example, it is clear that only results similar to known examples can be found.

Many target prediction algorithms have been described and implemented, many of which are listed in Table 1. Each of these algorithms uses one or several of the criteria listed above in order to find putative target sites and then to score the significance of the predictions. Many algorithms for predicting folding of RNA sequences have been written but the tools in Table 1 mostly use Mfold (Mathews, Sabina, Zuker, & Turner, 1999), RNAHybrid (Rehmsmeier et al., 2004), or the Vienna RNA package (I. Hofacker et al., 1994).

4 MicroRNA Target Databases

Predicted and identified targets and other miRNA related information need to be stored in a safe and easy to access environment for future use. Relevant databases have emerged by manual gathering of data from large numbers of experimentally validated miRNA targets and from high-throughput techniques. As such databases grow important issues such as the need for advanced searching and result filtering capabilities in order to accurately retrieve miRNAs or genes of specific interest become evident. Metadata and further enhancement of the currently available databases with added information from external sources will enable efficient data mining of available experimentally validated results. This is important as it will give way to producing useful novel observations (Vergoulis et al., 2012b). Currently miRTarBase (Hsu et al., 2011) provides a collection of miRNA-target interactions with experimental support. It has accumulated more than 3,500 miRNA targets by manually surveying the relevant literature. This is done after a systematic data mining step to filter research articles related to functional studies of miRNAs. Maybe the most comprehensive miRNA related database is miRBase which houses

information on both miRNA and target sequences along with predicted targets (for more information on databases pertaining to small RNAs please refer to Chapter 5 in this volume). TarBase on the other hand houses manually curated targets for different species with information on the target site and the miRNA:mRNA duplex. It also gives information about the type of experiment used for targeting and validation along with references to relevant publications. Argonaute (Shahi et al., 2006) contains information on mammalian miRNAs including their gene of origin and regulated target genes which are collected from literature and other databases. Animal miRNA targets and predictions from 11 different miRNA target prediction tools are stored in miRecords (Xiao et al., 2009a).

5 Conclusion

The number of computational methods for miRNA target prediction is increasing and new methods promise to deliver better results. Whether or not these methods are successful in keeping up with their promises or not is a subject for debate. One can expect to see better methods come by as our understanding of the miRNA regulatory pathway increases. The most important factor in developing such new algorithms and tools will be the accurate and precise computational modeling of the new scientific knowledge. This can range from better sequencing data and better classification of the 3′UTRs and splice sites, to the biosynthetic pathway of miRNAs and its regulators. This calls for extensive databases which can collect, store, and provide fast and efficient recalls of such scientific data. Whether existing databases are revised or updated, or new databases are designed, this may be one of the most important factors in the development of new and effective methods for miRNA target predictions.

6 Outlook

The current speed of advancements in miRNA related studies is staggering. In less than 20 years miRNAs have had a huge impact in biological sciences. If the advancements in target predictions keep up with the current pace one can predict that miRNA target predictions will be important player in many applications such as the development of new therapeutics. While predicting targets is possible on a per miRNA basis, genome wide studies are suffering from a large pool of possibilities and therefore we will see a trend towards incorporating all filtering mechanisms for miRNA target prediction, introduced in this work, and potentially further ones to increase the number of true positive identifications. Since no ground truth data is available, more and more small datasets (e.g: microarray data) providing a part of the truth will be incorporated in future studies.

7 Acknowledgements

HH would like to thank Associate Professor Dr. Jens Allmer for his kind guidance, encouragement, and advice and would also like to thank his parents for their ever increasing support. JA would like to thank his wife Açalya for the tolerance towards late hours spend on this and other work in this volume and his son Lukas Aren for providing fun distractions during the process.

8 References

Barbato, C., Arisi, I., Frizzo, M. E., Brandi, R., Da Sacco, L., & Masotti, A. (2009). Computational challenges in miRNA target predictions: to be or not to be a true target? *Journal of Biomedicine & Biotechnology*, *2009*, 803069. doi:10.1155/2009/803069

Brennecke, J., Stark, A., Russell, R. B., & Cohen, S. M. (2005). Principles of microRNA-target recognition. *PLoS biology*, *3*(3), e85. doi:10.1371/journal.pbio.0030085

Cheng, C., & Li, L. M. (2008). Inferring microRNA activities by combining gene expression with microRNA target prediction. *PloS one*, *3*(4), e1989. doi:10.1371/journal.pone.0001989

Creighton, C. J., Nagaraja, A. K., Hanash, S. M., Matzuk, M. M., & Gunaratne, P. H. (2008). A bioinformatics tool for linking gene expression profiling results with public databases of microRNA target predictions. *RNA*, *14*(11), 2290–2296. doi:10.1261/rna.1188208

Dai, X., & Zhao, P. X. (2011). psRNATarget: a plant small RNA target analysis server. *Nucleic acids research*, *39*(Web Server issue), W155–9. doi:10.1093/nar/gkr319

Darty, K., Denise, A., & Ponty, Y. (2009). VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics (Oxford, England)*, *25*(15), 1974–5. doi:10.1093/bioinformatics/btp250

Doench, J. G., & Sharp, P. A. (2004). Specificity of microRNA target selection in translational repression. *Genes & development*, *18*(5), 504–11. doi:10.1101/gad.1184404

Du, T., & Zamore, P. D. (2005). microPrimer: the biogenesis and function of microRNA. *Development (Cambridge, England)*, *132*(21), 4645–4652. doi:10.1242/dev.02070

Dweep, H., Sticht, C., Pandey, P., & Gretz, N. (2011). miRWalk--database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. *Journal of biomedical informatics*, *44*(5), 839–47. doi:10.1016/j.jbi.2011.05.002

Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., & Marks, D. S. (2003a). MicroRNA targets in Drosophila. *Genome biology*, *5*(1), R1. doi:10.1186/gb-2003-5-1-r1

Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., & Marks, D. S. (2003b). MicroRNA targets in Drosophila. *Genome biology*, *5*(1), R1. doi:10.1186/gb-2003-5-1-r1

Forman, J. J., Legesse-Miller, A., & Coller, H. A. (2008). A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(39), 14879–84. doi:10.1073/pnas.0803230105

Friedman, R. C., Farh, K. K.-H., Burge, C. B., & Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, *19*(1), 92–105. doi:10.1101/gr.082701.108

Gaidatzis, D., Van Nimwegen, E., Hausser, J., & Zavolan, M. (2007). Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics*, *8*, 69. doi:10.1186/1471-2105-8-69

Grimson, A., Farh, K. K.-H., Johnston, W. K., Garrett-Engele, P., Lim, L. P., & Bartel, D. P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Molecular Cell*, *27*(1), 91–105. doi:10.1016/j.molcel.2007.06.017

Gu, S., Jin, L., Zhang, F., Sarnow, P., & Kay, M. A. (2009a). Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. *Nature structural & molecular biology*, *16*(2), 144–50. doi:10.1038/nsmb.1552

Gu, S., Jin, L., Zhang, F., Sarnow, P., & Kay, M. A. (2009b). Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. *Nature structural & molecular biology*, *16*(2), 144–50. doi:10.1038/nsmb.1552

Hammell, M., Long, D., Zhang, L., Lee, A., Carmack, C. S., Han, M., Ding, Y., et al. (2008). mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. *Nature Methods*, *5*(9), 813–819. doi:10.1038/nmeth.1247

Hofacker, I., Fontana, W., Stadler, P., Bonhoeffer, L., Tacker, M., & Schuster, P. (1994). Fast Folding and Comparison of RNA Secondary Structures. *Monatshefte für Chemie*, *125*(2), 167–188. Retrieved from http://apps.isiknowledge.com/full_record.do?product=WOS&search_mode=GeneralSearch&qid=17&SID=Q1o4lPD8nJl7@1AN@ne&page=1&doc=1

Hofacker, I. L. (2003). Vienna RNA secondary structure server. *Nucleic Acids Research*, *31*(13), 3429–3431. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12824340

Hsu, S.-D., Lin, F.-M., Wu, W.-Y., Liang, C., Huang, W.-C., Chan, W.-L., Tsai, W.-T., et al. (2011). miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic acids research*, *39*(Database issue), D163–9. doi:10.1093/nar/gkq1107

Huang, J. C., Morris, Q. D., & Frey, B. J. (2007). Bayesian inference of MicroRNA targets from sequence and expression data. *Journal of Computational Biology*, *14*(5), 550–563. doi:10.1089/cmb.2007.R002

Hubbard, T. (2002). The Ensembl genome database project. *Nucleic Acids Research*, *30*(1), 38–41. doi:10.1093/nar/30.1.38

Jiang, Q., Feng, M.-G., & Mo, Y.-Y. (2009). Systematic validation of predicted microRNAs for cyclin D1. *BMC Cancer*, *9*, 194. doi:10.1186/1471-2407-9-194

John, B., Enright, A. J., Aravin, A., Tuschl, T., Sander, C., & Marks, D. S. (2004). Human MicroRNA targets. *PLoS Biology*, *2*(11), e363. doi:10.1371/journal.pbio.0020363

Johnson, C. D., Esquela-Kerscher, A., Stefani, G., Byrom, M., Kelnar, K., Ovcharenko, D., Wilson, M., et al. (2007). The let-7 microRNA represses cell proliferation pathways in human cells. *Cancer research*, *67*(16), 7713–22. doi:10.1158/0008-5472.CAN-07-1083

Jones-Rhoades, M. W., & Bartel, D. P. (2004a). Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Molecular cell*, *14*(6), 787–99. doi:10.1016/j.molcel.2004.05.027

Jones-Rhoades, M. W., & Bartel, D. P. (2004b). Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Molecular cell*, *14*(6), 787–99. doi:10.1016/j.molcel.2004.05.027

Joung, J.-G., & Fei, Z. (2009). Computational identification of condition-specific miRNA targets based on gene expression profiles and sequence information. *BMC bioinformatics*, *10 Suppl 1*, S34. doi:10.1186/1471-2105-10-S1-S34

Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., & Segal, E. (2007). The role of site accessibility in microRNA target recognition. *Nature Genetics*, *39*(10), 1278–1284. doi:10.1038/ng2135

Kim, S.-K., Nam, J.-W., Lee, W.-J., & Zhang, B.-T. (2005). A Kernel Method for MicroRNA Target Prediction Using Sensible Data and Position-Based Features. *2005 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology* (pp. 1–7). IEEE. doi:10.1109/CIBCB.2005.1594897

Kim, S.-K., Nam, J.-W., Rhee, J.-K., Lee, W.-J., & Zhang, B.-T. (2006). miTarget: microRNA target gene prediction using a support vector machine. *BMC bioinformatics*, *7*, 411. doi:10.1186/1471-2105-7-411

Kiriakidou, M., Nelson, P. T., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z., & Hatzigeorgiou, A. (2004). A combined computational-experimental approach predicts human microRNA targets. *Genes & Development*, *18*(10), 1165–1178. doi:10.1101/gad.1184704

Kloosterman, W. P., Wienholds, E., Ketting, R. F., & Plasterk, R. H. A. (2004). Substrate requirements for let-7 function in the developing zebrafish embryo. *Nucleic Acids Research*, *32*(21), 6284–6291. doi:10.1093/nar/gkh968

Kozomara, A., & Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic acids research*, *39*(Database issue), D152–7. doi:10.1093/nar/gkq1027

Krek, A., Grün, D., Poy, M. N., Wolf, R., Rosenberg, L., Epstein, E. J., MacMenamin, P., et al. (2005). Combinatorial microRNA target predictions. *Nature Genetics*, *37*(5), 495–500. doi:10.1038/ng1536

Krüger, J., & Rehmsmeier, M. (2006). RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Research*, *34*(Web Server issue), W451–454. doi:10.1093/nar/gkl243

Lai, E. C. (2002). Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nature genetics*, *30*(4), 363–4. doi:10.1038/ng865

Lewis, B. P., Burge, C. B., & Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, *120*(1), 15–20. doi:10.1016/j.cell.2004.12.035

Lewis, B. P., Shih, I., Jones-Rhoades, M. W., Bartel, D. P., & Burge, C. B. (2003). Prediction of mammalian microRNA targets. *Cell*, *115*(7), 787–798. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14697198

Lin, S., & Ding, J. (2009). Integration of ranked lists via cross entropy Monte Carlo with applications to mRNA and microRNA Studies. *Biometrics*, *65*(1), 9–18. doi:10.1111/j.1541-0420.2008.01044.x

Liu, J. (2008). Control of protein synthesis and mRNA degradation by microRNAs. *Current opinion in cell biology*, *20*(2), 214–21. doi:10.1016/j.ceb.2008.01.006

Lytle, J. R., Yario, T. A., & Steitz, J. A. (2007). Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5’ UTR as in the 3' UTR. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(23), 9667–9672. doi:10.1073/pnas.0703820104

Mathews, D. H., Sabina, J., Zuker, M., & Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of Molecular Biology*, *288*(5), 911–940. doi:10.1006/jmbi.1999.2700

Mazière, P., & Enright, A. J. (2007). Prediction of microRNA targets. *Drug discovery today*, *12*(11-12), 452–8. doi:10.1016/j.drudis.2007.04.002

Miranda, K. C., Huynh, T., Tay, Y., Ang, Y.-S., Tam, W.-L., Thomson, A. M., Lim, B., et al. (2006). A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell*, *126*(6), 1203–17. doi:10.1016/j.cell.2006.07.031

Nam, S., Kim, B., Shin, S., & Lee, S. (2008). miRGator: an integrated system for functional annotation of microRNAs. *Nucleic Acids Research*, *36*(Database issue), D159–164. doi:10.1093/nar/gkm829

Ørom, U. A., Nielsen, F. C., & Lund, A. H. (2008). MicroRNA-10a binds the 5’UTR of ribosomal protein mRNAs and enhances their translation. *Molecular cell*, *30*(4), 460–71. doi:10.1016/j.molcel.2008.05.001

Peter, M. E. (2010). Targeting of mRNAs by multiple miRNAs: the next step. *Oncogene*, *29*(15), 2161–2164. doi:10.1038/onc.2010.59

Place, R. F., Li, L.-C., Pookot, D., Noonan, E. J., & Dahiya, R. (2008). MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(5), 1608–13. doi:10.1073/pnas.0707594105

Rajewsky, N. (2006). microRNA target predictions in animals. *Nature Genetics*, *38 Suppl*, S8–13. doi:10.1038/ng1798

Rajewsky, N., & Socci, N. D. (2004). Computational identification of microRNA targets. *Developmental Biology*, *267*(2), 529–535. doi:10.1016/j.ydbio.2003.12.003

Reczko, M., Maragkakis, M., Alexiou, P., Grosse, I., & Hatzigeorgiou, A. G. (2012a). Functional microRNA targets in protein coding sequences. *Bioinformatics (Oxford, England)*, *28*(6), 771–6. doi:10.1093/bioinformatics/bts043

Reczko, M., Maragkakis, M., Alexiou, P., Grosse, I., & Hatzigeorgiou, A. G. (2012b). Functional microRNA targets in protein coding sequences. *Bioinformatics (Oxford, England)*, *28*(6), 771–6. doi:10.1093/bioinformatics/bts043

Rehmsmeier, M., Steffen, P., Hochsmann, M., & Giegerich, R. (2004). Fast and effective prediction of microRNA/target duplexes. *RNA*, *10*(10), 1507–1517. doi:10.1261/rna.5248604

Rhoades, M. W., Reinhart, B. J., Lim, L. P., Burge, C. B., Bartel, B., & Bartel, D. P. (2002). Prediction of plant microRNA targets. *Cell*, *110*(4), 513–520. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12202040

Saetrom, O., Snøve, O., & Saetrom, P. (2005). Weighted sequence motifs as an improved seeding step in microRNA target prediction algorithms. *RNA*, *11*(7), 995–1003. doi:10.1261/rna.7290705

Saetrom, P., Heale, B. S. E., Snøve, O., Aagaard, L., Alluin, J., & Rossi, J. J. (2007). Distance constraints between microRNA target sites dictate efficacy and cooperativity. *Nucleic Acids Research*, *35*(7), 2333–2342. doi:10.1093/nar/gkm133

Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., & Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature*, *455*(7209), 58–63. doi:10.1038/nature07228

Sethupathy, P., Corda, B., & Hatzigeorgiou, A. G. (2006). TarBase: A comprehensive database of experimentally supported animal microRNA targets. *RNA*, *12*(2), 192–7. doi:10.1261/rna.2239606

Sethupathy, P., Megraw, M., & Hatzigeorgiou, A. G. (2006). A guide through present computational approaches for the identification of mammalian microRNA targets. *Nature Methods*, *3*(11), 881–886. doi:10.1038/nmeth954

Shahi, P., Loukianiouk, S., Bohne-Lang, A., Kenzelmann, M., Küffer, S., Maertens, S., Eils, R., et al. (2006). Argonaute--a database for gene regulation by mammalian microRNAs. *Nucleic Acids Research*, *34*(Database issue), D115–118. doi:10.1093/nar/gkj093

Stark, A., Brennecke, J., Bushati, N., Russell, R. B., & Cohen, S. M. (2005). Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. *Cell*, *123*(6), 1133–1146. doi:10.1016/j.cell.2005.11.023

Stark, A., Brennecke, J., Russell, R. B., & Cohen, S. M. (2003). Identification of Drosophila MicroRNA targets. *PLoS Biology*, *1*(3), E60. doi:10.1371/journal.pbio.0000060

Tay, Y., Zhang, J., Thomson, A. M., Lim, B., & Rigoutsos, I. (2008a). MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature*, *455*(7216), 1124–8. doi:10.1038/nature07299

Tay, Y., Zhang, J., Thomson, A. M., Lim, B., & Rigoutsos, I. (2008b). MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature*, *455*(7216), 1124–8. doi:10.1038/nature07299

Thadani, R., & Tammi, M. T. (2006). MicroTar: predicting microRNA targets from RNA duplexes. *BMC bioinformatics*, *7 Suppl 5*(Suppl 5), S20. doi:10.1186/1471-2105-7-S5-S20

Tokumaru, S., Suzuki, M., Yamada, H., Nagino, M., & Takahashi, T. (2008). let-7 regulates Dicer expression and constitutes a negative feedback loop. *Carcinogenesis*, *29*(11), 2073–7. doi:10.1093/carcin/bgn187

Vaucheret, H. (2006). Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes & development*, *20*(7), 759–71. doi:10.1101/gad.1410506

Vergoulis, T., Vlachos, I. S., Alexiou, P., Georgakilas, G., Maragkakis, M., Reczko, M., Gerangelos, S., et al. (2012a). TarBase 6.0: capturing the exponential growth of miRNA

targets with experimental support. *Nucleic acids research*, *40*(Database issue), D222–9. doi:10.1093/nar/gkr1161

Vergoulis, T., Vlachos, I. S., Alexiou, P., Georgakilas, G., Maragkakis, M., Reczko, M., Gerangelos, S., et al. (2012b). TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. *Nucleic acids research*, *40*(Database issue), D222–9. doi:10.1093/nar/gkr1161

Wang, X. (2008). miRDB: a microRNA target prediction and functional annotation database with a wiki interface. *RNA (New York, N.Y.)*, *14*(6), 1012–7. doi:10.1261/rna.965408

Wu, S., Huang, S., Ding, J., Zhao, Y., Liang, L., Liu, T., Zhan, R., et al. (2010). Multiple microRNAs modulate p21Cip1/Waf1 expression by directly targeting its 3' untranslated region. *Oncogene*, *29*(15), 2302–2308. doi:10.1038/onc.2010.34

Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X., & Li, T. (2009a). miRecords: an integrated resource for microRNA-target interactions. *Nucleic acids research*, *37*(Database issue), D105–10. doi:10.1093/nar/gkn851

Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X., & Li, T. (2009b). miRecords: an integrated resource for microRNA-target interactions. *Nucleic acids research*, *37*(Database issue), D105–10. doi:10.1093/nar/gkn851

Xie, Z., Kasschau, K. D., & Carrington, J. C. (2003). Negative feedback regulation of Dicer-Like1 in Arabidopsis by microRNA-guided mRNA degradation. *Current biology : CB*, *13*(9), 784–9.

Yan, X., Chao, T., Tu, K., Zhang, Y., Xie, L., Gong, Y., Yuan, J., et al. (2007). Improving the prediction of human microRNA target genes by using ensemble algorithm. *FEBS letters*, *581*(8), 1587–93. doi:10.1016/j.febslet.2007.03.022

Yang, Y., Wang, Y.-P., & Li, K.-B. (2008). MiRTif: a support vector machine-based microRNA target interaction filter. *BMC bioinformatics*, *9 Suppl 12*, S4. doi:10.1186/1471-2105-9-S12-S4

Yousef, M., Jung, S., Kossenkov, A. V, Showe, L. C., & Showe, M. K. (2007a). Naïve Bayes for microRNA target predictions--machine learning for microRNA targets. *Bioinformatics (Oxford, England)*, *23*(22), 2987–92. doi:10.1093/bioinformatics/btm484

Yousef, M., Jung, S., Kossenkov, A. V, Showe, L. C., & Showe, M. K. (2007b). Naïve Bayes for microRNA target predictions--machine learning for microRNA targets. *Bioinformatics (Oxford, England)*, *23*(22), 2987–92. doi:10.1093/bioinformatics/btm484

Zhang, Y. (2005). miRU: an automated plant miRNA target prediction server. *Nucleic Acids Research*, *33*(Web Server issue), W701–704. doi:10.1093/nar/gki383

Zuker, M. (2003a). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research*, *31*(13), 3406–15.

Zuker, M. (2003b). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research*, *31*(13), 3406–15.

9 Figure Captions

Figure 1: A schematic representation of the interactions between miRNAs and their target genes.

Figure 2: The hypothetical secondary structures of the main types of miRNA:mRNA duplexes in animals drawn using VARNA which is a tool for drawing and visualization of RNA secondary structure (Darty, Denise, & Ponty, 2009). *a) Perfect complementarity at the 5' end of the miRNA (seed region) with a bulge and a mismatch towards the 3' end. b) The seed region contains a mismatch and a G-U wobble and the 3' end has two bulges. c) The seed region contains a bulge and the 3' end has a bulge and a mismatch.*

Figure 3: The typical output of RNAhybrid. The first line gives the name of the FASTA file of the target, the second line is its length, and the third line in the name of the FASTA file of the miRNA followed by its length. The mfe and the pvalue are then given along with a semi graphic representation of the hybridization.

Figure 4: a) Here are some of the possible different seed region types with a, b, c, and d representing different kinds of perfect complementation and e, h, and i showing different possible binding patterns with 1 mismatch in the middle and a G-U wobble can be seen in f and g. Analyses for the conservation of these seed regions in human, fly, worm, and zebra fish have suggested that perfect matches are more conserved than the G-U pair containing seed regions which are more conserved than the regions with mismatches (Gaidatzis, Van Nimwegen, Hausser, & Zavolan, 2007).

Figure 5: A highly simplified view of targeting. A RISC bound mature miRNA is displayed abstractly with available bonds symbolized by sticks. Three targets are displayed by bonds represented as sticks. Target 1 (t1) is close to the stop of the translation and inaccessible, t2 is freely accessible and t3 is partially accessible.

10 Tables

Table 1: The table below is a non-comprehensive list of miRNA targeting programs.

Name	**Summary**	**Clade**	**Link**
TargetScanS (Lewis et al., 2005)	Modeling of adenosines flanking the seed region. Similar to TargetScan.	Vertebrate	http://genes.mit.edu/tscan/targetscanS2005.html
TargetScan (Lewis et al., 2003)	5′seed sequence, homology, and thermodynamics based modeling.	Mammal, worm, fly	http://www.targetscan.org/
PicTar (Krek et al., 2005)	Stringent seed pairing for at least one target, target clustering, and duplex stability.	Vertebrate, fly, nematode	http://pictar.mdc-berlin.de/
miRanda (John et al., 2004)	Position specific complementarity, optimization, and interspecies conservation.	Vertebrate	www.microrna.org
EMBL (Stark, Brennecke, Bushati, Russell, & Cohen, 2005)	Finds anti-targets in the 3′UTR and miRNA binding sites.	Animal	N/A
DIANA-microT (Kiriakidou et al., 2004)	Experimental rule generation and duplex binding energy.	Human and mouse	http://diana.pcbi.upenn.edu/cgi-bin/micro_t.cgi
RNA22	Identifies clustered targets	Animal,	http://cbcsrv.watson.ibm.com/r

(Miranda et al., 2006)	from patterns and finds corresponding miRNAs.	worm, fly	na22.html
PITA Top (Kertesz, Iovino, Unnerstall, Gaul, & Segal, 2007)	Target site′s sterical accessibility energy model.	Animal, fly, worm	http://genie.weizmann.ac.il/pubs/mir07/
miRU (Zhang, 2005)	Sequence similarity with adjustable mismatch settings.	Plant	http://bioinfo3.noble.org/miRNA/miRU.htm
EIMMo (Gaidatzis et al., 2007)	Homology based Bayesian prediction and term enrichment.	Mammal, fly, worm, fish	http://www.mirz.unibas.ch/ElMMo3/
RNAhybrid (Krüger & Rehmsmeier, 2006)	Hybridization energy, no bifurcations, and no fixed seed region.	Animal	http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/
TargetBoost (O. Saetrom et al., 2005)	Determines position specific sequence motives using machine learning.	N/A	https://demo1.interagon.com/targetboost/
mirWIP (Hammell et al., 2008)	Structural accessibility, free energy of hybridization, and topology of seed pairing.	Worm	http://146.189.76.171/query.php
miRGator (Nam,	Integrates miRanda, PicTar and	Vertebrate	http://genome.ewha.ac.kr/miR

Kim, Shin, & Lee, 2008)	TargetScanS results with additional information.		Gator/
SigTerms (Creighton, Nagaraja, Hanash, Matzuk, & Gunaratne, 2008)	MS Excel based tool to simplify results of miRanda, PicTar, and TargetScan results.	Vertebrate	http://sigterms.sourceforge.net/
MiRTif (Yang, Wang, & Li, 2008)	Support vector machine (SVM) based filtering of predictions from other tools.	N/A	http://mirtif.bii.a-star.edu.sg/
TopKCEMC (Lin & Ding, 2009)	Integrates a number of other tools and evaluates the results statistically.	N/A	http://www.stat.osu.edu/~statgen/SOFTWARE/TopKCEMC/
N/A (Joung & Fei, 2009)	Gene expression profiles, SVM, and duplex base pairing.	Arabidopsis	http://www.biomedcentral.com/content/supplementary/1471-2105-10-S1-S34-S1.xls
GenMIR++ (Huang, Morris, & Frey, 2007)	Gene expression profiles, Bayesian inference, and uses TargetScanS predictions.	Vertebrate	http://www.psi.toronto.edu/genmir/
MIR (Cheng & Li, 2008)	Gene expression profiles, target enrichment, and binding	N/A	http://homes.gersteinlab.org/people/cc59/InferMiRNA/infermi

	energy.		r.html
psRNA Target (Dai & Zhao, 2011)	Extension and incorporation of new rules for miRU.	Plant	http://plantgrn.noble.org/psRNATarget/
NBmiRTar (Yousef, Jung, Kossenkov, Showe, & Showe, 2007b)	MiRanda score, folding energy, and Naive Bayes score.	Vertebrate	http://wotan.wistar.upenn.edu/NBmiRTar
miRecords (Xiao et al., 2009b)	Integrates the predictions of other tools.	Animal	http://mirecords.umn.edu/miRecords/
N/A (Stark et al., 2003)	Complementarity.	Drosophila	http://www.russell.embl.de/miRNAs
miRWalk (Dweep, Sticht, Pandey, & Gretz, 2011)	Complementarity and integration of 8 other prediction tools.	Human, mouse, rat	http://www.umm.uni-heidelberg.de/apps/zmf/mirwalk/index.html
miTarget (Kim, Nam, Rhee, Lee, & Zhang, 2006)	Support Vector Machine.	Human	http://cbit.snu.ac.kr/~miTarget
miRDB (Wang, 2008)	Support Vector Machine.	Human, mouse, rat,	http://mirdb.org

		dog, chicken	