COMPUTATIONAL PREDICTION OF MicroRNA PRECURSORS AND MicroRNA TARGETS

bv

BRAM SEBASTIAN

(Under the Direction of Samuel E. Aggrey)

ABSTRACT

MicroRNA is small 22 nucleotides long non-coding RNA which regulates genes by targeting mRNA especially the 3' UTR region. Identification of miRNAs and their targets by laboratory experiment has had limited success especially for lowly expressed therefore computational prediction approaches are needed. In this study grouping technique for miRNA precursor prediction is introduced. Compared with global alignment, grouping miRNA by classes yield a better sensitivity with very high specificity for pre-miRNA prediction even when a simple positional based secondary and primary structure alignment are used.

The program TarSpec was developed to predict miR-1a and miR-124 targets based on common features of miRNA and target binding characteristics observed from alignment between miRNA and the 3'UTR targets. TarSpec obtained 78% and 77% sensitivity for miR-1a and Mir-124 targets, and 98% specificity for both. TarSpec was used to scan the Platypus 3'UTR regions. This approach predicted 734 novel potential target of miR-1a in Platypus 3'UTR regions where 98 of them are in the well annotated chromosomal region and 124 novel potential target of miR-124 where 32 of them are in the well annotated chromosomal region.

Some miRNAs are derivations of transposable elements (TE). In human these TE derived miRNAs have a potential to regulate thousands of human genes. Therefore TEs as potential miRNA targets were investigated using an L2 derived miRNA miR-28. Three different miRNA target prediction programs miTarget, miRanda, and RNAhybrid were used to predict a potential miRNA-28 targets in human L2 transposable elements. It was demonstrated that the human TE is also a potential target for miRNAs; subsequently 1,094 of potential target were predicted in human L2.

INDEX WORDS: miRNA, miRNA targets, platypus, human, transposable elements, chicken, human genome, chicken genome, microRNA precursor, RNA secondary structure

COMPUTATIONAL PREDICTION OF MIRNA PRECURSORS AND MIRNA TARGETS

by

BRAM SEBASTIAN

BS, Iowa State University, 2004

MS, University of Georgia, 2007

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

© 2010

Bram Sebastian

All Rights Reserved

COMPUTATIONAL PREDICTION OF MicroRNA PRECURSORS AND MicroRNA

TARGETS

by

BRAM SEBASTIAN

Major Professor: Committee: Samuel E. Aggrey Liming Cai Shaying Zhao

Electronic Version Approved:

Maureen Grasso Dean of the Graduate School The University of Georgia December 2010

DEDICATION

I dedicate this dissertation to my family and friends who has supported me through the time of my study.

ACKNOWLEDGEMENTS

I thank my major professor and all professors serving in my committee member for all of the help and support that they have provided for me to achieve my degree.

TABLE OF CONTENTS

	Pag	;e
ACKNOWLED	GEMENTS	v
CHAPTER		
1 INTE	RODUCTION	1
2 LITE	ERATURE REVIEW	4
R	EFERENCES	7
3 MIR-	EXPLORE : PREDICTING MIRNA PRECURSORS BY CLASS	
GRO	UPING AND SECONDARY STRUCTURE POSITIONAL	
ALIC	GNMENT	.8
А	BSTRACT4	.9
Π	NTRODUCTION	0
Ν	1ATERIALS AND METHODS	2
R	ESULTS AND DISCUSSIONS	6
С	CONCLUSION	9
А	CKNOWLEDGEMENT	9
R	EFERENCES	0
Т	ABLES	5
F	IGURES	6
S	UPLEMENTARY DATA	7

4	POTENTIAL MIR-1a and MIR-124 TARGETS IN THE 3'	
	UNTRANSLATED REGIONS (3'UTRs) OF PLATYPUS (Ornithorhys	nchus
	anatinus)	77
	ABSTRACT	
	INTRODUCTION	79
	MATERIALS AND METHODS	82
	RESULTS AND DISCUSSION	84
	REFERENCES	87
	TABLES	90
	FIGURES	94
5	POTENTIAL TARGET OF MIR-28 IN HUMAN TRANSPOSABLE	
	ELEMENT L2	96
	ABSTRACT	97
	INTRODUCTION	98
	MATERIALS AND METHODS	100
	RESULTS AND DISCUSSION	102
	REFERENCES	104
	TABLES	107
	FIGURES	119
6	CONCLUSION	120
Al	PPENDIX 1 JAVA CODE FOR MIR-EXPLORE	121
Al	PPENDIX 2 PYTHON CODE FOR TARSPEC	135

CHAPTER 1

INTRODUCTION

MicroRNAs (MiRNAs) are small non-coding RNAs with 22 nucleotides long. They are expressed in a wide variety of organisms including viruses, plants, and animals (Bartel & Bartel 2003; Lim et al. 2003a; Pfeffer et al. 2004) and have a significant role in post transcriptional control of eukaryotes genome by degrading the mRNA transcripts or blocking translation (Lau et al. 2001; Lee & Ambros 2001). MiRNA genes are commonly transcribed by RNA-polymerase-II to primary miRNA (pri-miRNA) (Lee et al. 2004). The pri-miRNA is then further processed by RNAse-III endonuclease Drosha into precursor miRNA (pre-miRNA) which has a length of 60–100 nucleotides on average and forms a stem loop secondary structure (Denli et al. 2004). The pre-miRNA is then transported out from the nucleus to the cytoplasm and further processed to form the small double stranded RNAs by the enzyme Dicer which also initiates the formation of the RNA induced silencing complex (RISC) (Bernstein et al. 2001). One strand of the double stranded RNA is incorporated with the RISC and targets the mRNA transcripts to block gene expression (Griffiths-Jones et al. 2006).

It has been predicted that more than one third of human genes are conserved miRNA targets; only few miRNA functions have been fully identified because predicting and confirming miRNA targets are difficult (Xia et al. 2009). Some miRNAs have also been known to be derived from transposable elements (TE). MiRNA has been thought to be conserved among organisms, nevertheless non-conserved miRNAs has also been

found (Bentwich et al. 2005). Transposable Elements (TE) have been known to contribute to the origin and evolution of human miRNAs (Piriyapongsa et al. 2007). The distinction between miRNA and other small interfering RNAs (siRNAs) is that siRNAs are transcribed from some TEs and miRNAs are transcribed from loci other than any TEs and other genes (Bartel 2004). However, this claim has been disputed by the discovery of miRNAs derived from TEs (Smalheiser & Torvik 2005; Borchert et al. 2006; Piriyapongsa et al. 2007). Scientists often underestimate the contribution of TEs, which often are excluded from the computational search of novel miRNAs (Bentwich et al. 2005; Lindow & Krogh 2005; Nam et al. 2005; Li et al. 2006). This is because of the tendency that TEs are often treated as genomic parasites which do not play any functional roles in the host species (Doolittle & Sapienza 1980). However, more recent studies has proven that TEs can provide the functional roles to the host such as addition to a coding sequence (Volff 2006) and several regulatory sequences derived from TEs (Jordan et al. 2003; van de Lagemaat et al. 2003).

The tendency of excluding TEs in the novel miRNA search is extended to the miRNA target search. So far, the focus of miRNA target search is more on searching the 3' UTRs for a potential miRNA targets. If miRNA can be derived from TEs, it is likely that TEs can also being potential miRNA targets. In this study, we are going to use available bioinformatics tools to predict a potential target of miRNAs in TEs.

The objectives of this study are:

- 1. Develop a new approach to predict miRNA precursor.
- 2. Develop a microRNA target prediction program based on the observation of miRNA and target interaction features in a two dimensional level.
- Utilize available miRNA target prediction bioinformatics tools to predict novel miRNA targets in human transposable elements.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction to miRNA

MicroRNA (miRNA) is a short endogenous RNA with average length of 22 nucleotides that plays important roles in gene regulations. In animals, miRNA is complementarily paired to the target genes either imperfectly or perfectly to repress translation of or to degrade mature messenger RNA (mRNA)(Bartel 2004; Bartel 2009). There are more than 1,000 miRNAs encoded by human genome, and these miRNAs regulates approximately 30% of human genes (Bentwich et al. 2005; Friedman et al. 2009). Each miRNAs can target hundreds of mRNAs and one mRNA can be targeted by more than one miRNA (Brennecke et al. 2005).

MiRNA was first identified in 1993 in Victor Ambros Laboratory during the study of lin-4 gene in nematode *C. elegans*, but it was not categorized as miRNA (Wightman et al. 1993). They found that lin4 gene encodes a small RNA that regulates another gene which is lin-14 gene. They observed that a 61-nucleotide precursor from the lin-4 gene was processed to a 22 nucleotides long small RNA and this small RNA was paired imperfectly to the 3' UTR of the lin-14 mRNA and blocked the translation of the lin-14 gene. It was later found that let-7 encode a small RNA which repressed the expression of lin-41, lin-14, lin28, lin42, and daf12 during the developmental stage of *C. elegans*, and since then, this small RNA has been recognized as miRNA (Reinhart et al. 2000).

MiRNA genes have their own promoter and regulatory units and are mostly found in the intergenic region or the opposite strand of some genes (Lau et al. 2001), although some of miRNA genes are also found in the introns or even exons which will be regulated by the same regulator of the particular host genes (Rodriguez et al. 2004). MiRNAs are mostly transcribed by RNA Polymerase II (Pol II) (Lee & Kim 2004; Lee et al. 2004). The binding site of the Pol II is often found in the promoter near the DNA sequence that will be encoded as the hairpin loop of the miRNA precursor (pre-miRNA) (Zhou et al. 2007). The resulting transcript of the miRNA genes transcription is a primary miRNA (pri-miRNA) which is capped in the 5' end and poly adenylated in the 3' end (Cai et al. 2004). Pri-miRNA can be hundreds or thousands nucleotides long and contain one or more pre-miRNAs (Lee et al. 2004). A pri-miRNA can also be a coding gene if the 3' UTR region contain pre-miRNA (Cai et al. 2004). Some miRNAs are transcribed by RNA Pol III instead of Pol II. This miRNA has Alu, transfer RNA (tRNA) and mammalian wide interspersed repeat (MWIR) promoter (Faller & Guo 2008).

After transcription, pri-miRNA is then being processed to become pre-miRNA which has a hairpin like structure. This hairpin loop structures are approximately 70 nucleotides long in average. The double stranded part of the hairpin structure is recognized by a nuclear enzyme DiGeorge Syndrome Critical Region 8 (DGCR8) (Bartel 2004). DGCR8 forms a complex with an enzyme Drosha and cut the hairpin from the primiRNA by cleaving the RNA about 11 nucleotides away from the hairpin base (2 helical RNA turns into the stem) and the resulting hairpin (pre-miRNA) has two nucleotides hanging at the 3' end (Faller & Guo 2008). There are some pre-miRNAs which are spliced directly out of introns caused by the same complex (DGCR8 and Drosha), and

these pre-miRNAs are called "mirtrons". Mirtrons are originally thought to exist only in *Drosophila* and *C. elegans*, but they have also been found in mammals (Berezikov et al. 2007).

Once pre-miRNA is formed, it is transported out of the nucleus by an enzyme Exportin-5. This enzyme recognizes the 2 nucleotides hanging at the end of the 3' end of the pre-miRNA and mediate the transport of pre-miRNA to the cytoplasm (Murchison & Hannon 2004). In the cytoplasm, pre-miRNA is cut by an RNAse III enzyme Dicer. This enzyme cut away the loop that joins the 3' and 5' arms, resulting in an imperfect double stranded small RNA with an average length of 22 nucleotides (Ji 2008), one strand interact with RNA induced silencing complex (RISC) to become functional mature miRNA and another strand which is usually called miRNA star is degraded.

Mature miRNA is part of RISC which contains Dicer and many other proteins, and the complex of mature miRNA and RISC along with some other proteins is widely known as miRNP (Schwarz & Zamore 2002). When Dicer cut the stem of the hairpin loop structure, it also unwinds the RNA. Only one of the stems is incorporated into the RISC and the selection is based on the thermodynamics instability and weaker basepairing relative to the other strand (Khvorova et al. 2003; Krol & Krzyzosiak 2004). In some cases, both strands of the stems of the hairpin loop of pre-miRNA can be functional mature miRNAs (Okamura et al. 2008).

MiRNA functions as gene regulator, hence miRNA is complementary to one or more part of mRNA. In animal, miRNA is usually complementary to the 3' UTR region, while in plants miRNA is usually complementary to the coding region (Murchison & Hannon 2004). In plants, miRNA is more than likely to pair perfectly to the target to

degrade the mRNA (Wang et al. 2004). Whereas, in animal, miRNA is more than likely to imperfectly pair to its target and block translation of the mRNA and promote faster deadenylation to degrade mRNA faster than it is suppose to be (Williams 2008; Eulalio et al. 2009). For imperfectly paired miRNA, there is a region called "seed region" which is nucleotide position 2 to 7. This region need to pair perfectly to the target for the miRNA to be able to function properly (Martin et al. 2007; Maziere & Enright 2007). As a gene regulator, miRNA has been linked with some diseases including cancer and heart disease such as *hyperthropic cardiomyopathy* a thickening of the heart muscle (Fontana et al. 2008; Wang et al. 2010b). One example of the involvement of miRNA in cancer growth was in an experiment in mice where a genetic engineering was used to overexpressed miRNA that was found in lymphoma cells. The disease developed in 50 days and caused death 2 weeks later, while mice with normal level of this miRNA lived for over 100 days (He et al. 2005).

The identification of miRNA has been done by direct cloning (Xu et al. 2009), Quantitative Real Time PCR (RT-PCR) (Li et al. 2009), and micro array analysis (Li & Ruan 2009). Many miRNAs have been discovered using those techniques, but these biological techniques have limitation on detecting miRNA with highly constrained tissue and time specific expression, and low level of expression (Li et al. 2010a). Hence computational prediction of miRNA and its potential mRNA targets is needed to assist biologist to further study miRNA and possibly confirmed more novel miRNAs in the genome.

2.2 Computational Approaches for miRNA prediction

Computational prediction of miRNAs relies on a basic principal of known characteristics of miRNA and uses those to search in the genome. In animals, the prediction relies on the conservation of miRNAs across related species, formation of the stable stem loop structure of the pre-miRNA and the presence of mature miRNA in the stem but not in the loop. However, in virus and in plants, the computational prediction is more challenging since they have less conservation across species and the length of the hairpin structures are more varied (Chaudhuri & Chatterjee 2007). There are two major approaches in this computational method. They are comparative and non comparative.

Comparative method relies on the folding of pre-miRNA to the hairpin secondary structure and conservation of closely related genome to predict novel pre-miRNA. The corresponding location of the conserved sequence can then be identified as the candidate for new pre-miRNA. One program that uses comparative method is miRScan (Lim et al. 2003a; Lim et al. 2003b). MiRScan was developed to identify miRNA genes conserved in more than one genome. This program use RNAFold (Schuster et al. 1994) to obtain the hairpin structures in the sequences that are evolutionary conserved in *C. elegans* and *C. griggsae*. The program slide a window with 110 nt width to both C. elegans genome strands and fold the window using RNAFold to identify stem loop structure greater than 25 bp and a folding energy of at least -25 kcal/mol. Each hairpin is scanned using a window with 21 nt width assigning a log-likelihood score to each position for similarity to known miRNA. This program used 50 published miRNAs from *C. elegans* and *C. briggsae*. MiRScan has been able to identify 35 new miRNAs in *C.elegans* where 16 of them have been experimentally verified. When it is used to scan the hairpin retrieved

from human, mice, and puffer fish (approximately 15,000 hairpins) the program had sensitivity of 0.74 (Lim et al. 2003a). MiRScan has been used to identify more novel miRNA which has a high conservation level among organisms. Another comparative method is miRseeker (Lai et al. 2003) analyzed sequences of *D. melanogaster* and *D. pseudoobscura* for conserved stem loop structured sequences using Mfold (Zuker 2003). This program has 75% sensitivity and has been able to identify 48 novel miRNA candidates strongly conserved in more distant insect, nematode, or vertebrate genomes. Comparative methods achieve better prediction of miRNAs that have been conserved among organisms. This method relies on the known miRNAs therefore it potentially cannot predict a novel miRNA that has not been discovered previously, unless it is within the same class.

In order to predict completely novel miRNAs, non comparative method is needed. Since this method does not rely on conservation, it will have capability to predict miRNAs which are non-conserve and species specific. This method still relies on the premiRNA structure which is a hairpin. The purpose of the non-comperative method is to distinguish between the real pre-miRNA hairpin and non pre-miRNA hairpin which usually called pseudo hairpin. There are several methods using Support Vector Machine (SVM), such as triplet-SVM (Xue et al. 2005) which was further improved by using random forest algorithm and renamed to miPred (Jiang et al. 2007), and miR-abela (Sewer et al. 2005). These programs have been able to detect microRNAs with no conservation, but their sensitivity and specificity are lower compared to comparative methods. Besides SVM, Hidden Markov Model (HMM) has also been used to do noncomparative prediction, for example in ProMir (Nam et al. 2005) and ProMir II (Nam et

al. 2006). ProMiR and ProMiR II use both pre-miRNA and mature miRNA to create HMM profile based on those two plus the signal cleavage of Drosha. Yet again the specificity and sensitivity is not as high as the comparative methods. Besides those two methods, SCFG method and its derivation has also been used. It is difficult to formulate a good non comparative prediction program. It is also hard to compare the available programs, since they use different training data and test data to fit their own need. Their specificity and sensitivity number cannot be compared side by side unless independent test is done by using the same test data for all the programs.

2.3 Some SCFG based method for miRNA prediction

2.3.1 CID-miRNA

In 2008 (Tyagi et al. 2008) created CID-miRNA. This program focuses specifically on humans. In general, they used SCFG and structural filters to predict new miRNAs in human genome sequences. The SCFG quickly searches for a region that possibly contains pre-miRNA. In other words, the SCFG searches for stem loop structure similar to the pre-miRNA stem loop. This becomes the candidates that are going to be screened later. The SCFG modeling was done according to Eddy's book Biological Sequence Analysis (Sakakibara et al. 1994). Sakakibara's group divided the grammar into three parts. First is the alphabet, in RNA case, the alphabet are A, U, G, and C. Second is the finite set N of non terminal symbol S1 Sn, with a special symbol S0 (starting symbol). The third is a set P of production rules that specify how sequences containing non-terminals can be rewritten by expanding those non-terminals to new

subsequences. The language was a set of all sequences of terminal symbols that can be derived from S0 by applying repeatedly the production from P.



Figure 2.1. Example of production rules, derivation, parse tree, and secondary structure

Figure 2.1a An example of a production rules representing a set of RNA sequence having a common structure. This production rules can be classified into 4 types. $S \rightarrow aSa$ represent a base pairs (for example $S \rightarrow gSc$), $S \rightarrow aS$ and $S \rightarrow a$ represent unpaired base, $S \rightarrow S$ called a skip production representing a deletion, and finally $S \rightarrow SS$ representing a branched secondary structure (for example $S3 \rightarrow S4$ S9 produce a branched secondary structure (Figure 2.1d). A derivation is a rewriting process of the sequence based on the production rules. A simple example can be seen in Figure 2.1b. For every derivation there is a parse tree representing the syntactic structure produced by the derivation (Figure 2.1c). In the case of RNA, this corresponds to the secondary structure which we can see in Figure 2.1d.

The grammar of the SCFG used by CID-miRNA was generated by using a set of experimentally known miRNA (474 human miRNA precursors obtained from the following website http://microrna.sanger.ac.uk/sequences/. The grammar was then converted into Chomsky normal form. CFG is in a Chomsky normal form when all of the production rules are in the form of A \rightarrow BC, A \rightarrow a, or S $\rightarrow \varepsilon$, where A, B, and C are terminal symbols, a is the terminal symbols, S is the start symbol and ε is the empty string. B and C cannot be the start symbol. The grammar was then trained using the positive data and inside and outside algorithm. The positive data is then trimmed so that there are no sequences with close homology. This was to make sure that each sequence will be represented exactly once. Inside algorithm is an algorithm that calculates the probability score of a sequence given an SCFG. The outside algorithm uses the result from the inside algorithm to calculate the probability of a complete parse tree rooted at the start non-terminal for sequence x. They defined 50 to be the adequate number of data to train the model. CYK algorithm was then used to screen the positive and negative data to determine a score. The negative data was taken from 300 human rRNA having a hairpin structure assuming there are no miRNAs precursors in there. CYK algorithm is an algorithm used to calculate the maximum probability of a parse tree for a sequence. In other words, this algorithm provides a probability score which is the most probable tree for a particular sequence by aligning the tree with the sequence and trace back with a three dimensional dynamic programming providing the most probable parse tree. A normalized score was then calculated to fix the different length of the sequences. The inside outside algorithm takes a sequence of bases and returns total probability of all the derivations to generate that sequence. The CYK takes a sequence of bases as an input and

output the most likely secondary structure and its corresponding likelihood score. A machine learning toolkit WEKA (http://www.cs.waikato.ac.nz/ml/weka/) was then used to create a classification tree to separate real miRNA precursor and pseudo miRNA precursor. WEKA is a data mining software containing machine learning algorithms that have tools for classification, regression, data preprocessing and visualization. It can also develop a new machine learning scheme. The CID-miRNA was compared to miRAlign (Wang et al. 2005) and SVM based technique (Helvik et al. 2007). From the test result, CID-miRNA has 81% sensitivity and 99% specificity, miRAlign has 57% sensitivity and 98% specificity, while SVM based method has 96% sensitivity with a very low 33% specificity. Specificity was calculated as TN / (TN + FP), and sensitivity was calculated as TP / (FN + TP) where TN is true negative, FP is false positive, and FN is false negative. The result of CID-miRNA was used to scan human chromosome 3. They identified two putative pre-miRNAs that are yet to be annotated. The second novel premiRNA matched and expressed sequence tag (EST) which confirmed a strong candidacy for a true novel pre-miRNA, but yet to be validated by biological experiment.

The web server of CID-miRNA is available on a Linux base and provides capability to predict pre-miRNA structure in a given input primary sequence. The default parameter is set to be for optimum prediction for human miRNA. This can be change to optimize searches other than human (limited to apes and monkeys for better result). The maximum size of files that can be uploaded is 200kb.

2.3.2 RNAPromo

The RNAPromo (Rabani et al. 2008) is a program based on the SCFG representation of RNA motifs to identify short RNA motifs from an unaligned sequence

input. For each input, primary structure and secondary structure are specified. There are two parts in this technique. The first part is to identify the motif candidates, and the second part is using expectation maximization (EM) algorithm to refine the motif candidates.

As a validation sets, they used 204 Rfam (Griffiths-Jones et al. 2003) seed alignments with 8 instances, and extracted the complete sequence of the features such as 3'UTR, 5'UTR etc which contains the motif. Sequences were limited to 4000 bp, longer sequence will be truncated to this length centered on the motif. They then divide the sets that they have to 4 groups based on the sequence similarity (30%, 50%, 70% and 90%). Both ViennaRNA (Hofacker et al. 1994) and CONTRAFold (He et al. 2005) were used to provide the secondary structure prediction from these sequences. Random validation set was taken by taking the sequences contained in the Rfam validation set of specific sequence similarity level and partition them randomly into sets of similar sizes. The premiRNA sequences was taken from all organisms that has 10 or more known pre-miRNA in the miR-Base (Griffiths-Jones et al. 2006; Griffiths-Jones et al. 2008) and additional data with at least 20bp and overall size of 50 - 200 bases was taken from genomic transcripts of each organism randomly. These sequences were folded using ViennaRNA.

This program takes a set of RNA sequences assumed to share some motif and their secondary structures as an input. The algorithm identifies specific and relatively short candidate structures that appear in as many inputs as possible. These candidates are then filtered further by the EM algorithm. The score was calculated based on the loglikelihood score. They use standard receiver operating characteristics curve (ROC) and the area under the curve (AUC) to see the significance of the RNA's input likelihood

scores compared to the shuffle (random validation) sequences. Sequences with similarity higher than 90% were left out, since these sequences will provide high AUC score while not having functional motif.

The similarity on the sets was calculated based on Needleman-Wunsch global alignment algorithm. The similarity was the percentage of matches between the two sequences over the reported align region (including gap). They were using two sets of similarity value, which are the maximum similarity and the average similarity score. They used a limit of maximum similarity score of 90%, since they did not want to have sequences with high AUC score while not having any motif signal. The folding was done by running a 200 bp window with 100 bp overlap to cut running time and to avoid low accuracy of folding long sequences.

The statistical of the result was evaluated using basic cross validation test to assess the quality of the biological signal in the input. Each input was divided into k set and in this case they used k = 5. The motif was then learned from each of these k possible combinations of k-1 sets. A likelihood score that each motif assigned to the set of RNAs that were left out when learning was calculated. This likelihood score was than compared to the likelihood score of the random set of sequences with the same di-nucleotide distribution as the input sequence. Higher likelihood score to the left out sequence compared to the random sequence suggest that there is a biological signal in that specific left out sequence. They were using standard ROC curve and its associated area under the curve (AUC) to evaluate the significance of the likelihood score. The ROC curve is a plot of sensitivity vs 1 – specificity for a binary classifier system. The AUC is between 0 and 1. Values higher than 0.5 indicated that the left out sequence had a greater likelihood

value compare to the random sequence. Each AUC score was assigned a p-value using the AUC score distribution in a collection of random set with size similar to the tested set and comes from the same organism. The p-value defined the level of significance of the particular AUC score. For the motif, they took the top 10 motif candidates and trained these to the rest of the input sequences. The motif with the highest statistical score was declared a predicted motif. They did not calculate sensitivity or specificity of detecting the pre-miRNA, but by using this program they are able to identify some pre-miRNA features and distinguish pre-miRNAs of plants, animals, and viruses. They found that plants pre-miRNA had a small loop (average 5.6 bases) and long stems (average 38.1 bp) while animals had longer loop (average 10.9 bases) and shorter stems (32.8 bp). The overall lengths of plants pre-miRNA on average was 160 bases while animals was 88. They also observed that the stem length of pre-miRNA of animal viruses is 9.4 bp on average, which was closer to the average length of animal pre-miRNA stems. They also suggested that Drosha might recognize the cleavage signal by measuring the loop of the pre-miRNAs in animals. This was done by applying RNApromo to the pri-miRNA. In plants, randomly selected stem loop generally had similar loop length compared to the stem loop of pre-miRNA while in animals randomly selected stem loops had significantly different length of loop compared to pre-miRNA stem loop.

2.3.3 EvoFold

EvoFold (Pedersen et al. 2006) was not created specifically for miRNA prediction. But they tested this program in a miRNA data besides other noncoding RNAs. This program take multiple alignment and phylogenetic tree as inputs and provide specific secondary RNA structure with the corresponding folding potential score (fps) as

the output. Fps is length normalized likelihood ratio score which tends to emphasize the ratio of paired to unpaired bases. The phylogenetic tree included branches length estimate which specified the evolutionary relationship between the sequences in the multiple alignment. This method was based on two phylo-SCFGs. One was a functional RNA (fRNA) model that describes region potentially contain fRNA and other was a background model that describes region which did not contain fRNA. The scoring method was the log-likelihood ratio under these two models. The program is a linux based and available for download in (http://www.cbse.ucsc.edu/jsp/EvoFold).

Phylo-SCFG is a combination of the ability of SCFG to model RNA secondary structure with the phylogenetic model which can describe the substitution process along the branches of the tree. The advantage of this model is that it can align multiple sequences and weigh their information in a way that reflects the phylogeny. Two types of phylogenetic model are used in phylo-SCFG. The two models are single nucleotide and di-nucleotide (Figure 2.2e). The single nucleotide describes the non-pairing region of the RNA secondary structure and the di-nucleotide describes the substitution process of the stem pairing nucleotide.



Figure 2.2. The explanation diagram of EvoFold steps

There are two components in the phylo-SCFG, which are structural and nonstructural. The structural part describes the structural regions whose first and last bases are paired. This region can correspond to a single hairpin or a more complex structure and is referred as fold (Figure 2.2d). This component contain both nucleotide and dinucleotide models. The non-structural component describes the regions outside folds and contains only the single nucleotide phylogenetic model. The model for the fRNA contains both the structural and non-structural and the background model contain only the non-structural.

The program uses the fRNA model to assign a specific RNA secondary structure prediction to an input alignment (Figure 2.2c). The most probable structure given the multiple alignments is then predicted. Prediction without any structure is possible due to the presence of the non-structural model in the fRNA data set. All predicted fold that passed the fold elimination was included as the candidate set. The fps measure the

tendency of the alignment to contain any fRNA which was calculated as a log-odd score of the likelihood of seeing the alignment under fRNA model and the background model. The background model was designed to eliminate problem of alignment in GC rich regions. The fps score was length-normalized and the score was used to rank the fold.

This program was tested using the human miRNA taken from the 2004 microRNA registry (Griffiths-Jones 2004). From the test the program yielded 86% sensitivity. The false positive rate was calculated to be 5% when the calculation focused on the top ranked fold prediction. This calculation was not only for miRNA but for all dataset in the fRNA database. False positive rate also depended on the sequence conservation, number of bulges found in the stems, the genomic location, and also the overall shape of the secondary structure.

2.3.3.1 Explanation of Phylo-SCFG

The input of this algorithm is an alignment of RNA sequences, and the output is a single common structure for the particular alignment. The model consists of two parts which are the SCFG and the evolutionary part. The basis of the model is a simple SCFG with the following production rules: $S \rightarrow LS \mid L, F \rightarrow dFd \mid LS, L \rightarrow s \mid dFd$. Here, s represents the single base in a single string, and d represent bases that paired up in a stem. The non-terminal S produces loops while the non-terminal F produces stems, while L decides whether a specific loop would be a single base or a start of a new stem (Figure 2.3). Stems can have any length and loops have at least 2 positions length. Since F produces LS instead of just S, then the start of a new stem is also a position. This means that the two positions loop can be a base and a start of a new stem (bulges), two bases (paired) or two new stems (bifurcation). The probabilities of each production rule

determine the previous distribution of secondary structures, where each structures has a certain probability given by the SCFG. The SCFG probability was estimated from a training set of folded RNAs. SCFG needs to be in a Chomsky normal form and the computational time to run SCFG is proportional to the cubed of the sequence length.

Figure 2.3 Production of the RNA structure by the Grammar. (a) The rules being used starting from S. (b) The corresponding structure. (c) The production of a stem with a bulge

i,

. .

a)
$$S \rightarrow LS \rightarrow LLLLLLS \rightarrow LLLLLLL$$

 $\rightarrow ssLsssss \rightarrow ssdFdsssss$
 $\rightarrow ssdddFdddsssss$
 $\rightarrow ssdddLSdddsssss$
 $\rightarrow ssdddLLLLdddsssss$
 $\rightarrow ssdddssssdddsssss$

-

b)
$$s^{ss}s$$

 $d-d$
 $d-d$
 $ssd-d_{sssss}$

c) $F \rightarrow dFd \rightarrow ddFdd \rightarrow ddLSdd$ $\rightarrow ddLLdd \rightarrow ddLsdd \rightarrow dddFdsdd$

The model was implemented in four steps. The first one is a creation of a database of sequences with known structures. Second is the estimation of single base and base pairs frequency. Third is the estimation of the mutation rate. The fourth one is the estimation of the parameter probabilities from the SCFG production rules. The second step was done simply by calculating the single base and base paired in the training sequence (Figure 2.4). The third step is done by pairing a number of sequences in the data base of step 1. All possible ordered pairs were made with 85% sequence similarity. Mutation rate were calculated for single bases and base pairs. For the fourth step, the secondary structure of the database from step 1 along with the inside – outside algorithm were used to give each production rules a probability value.

Stem		Loop		Overall	
AU/UA	35.6%	A	36. <mark>4</mark> %	А	26.8%
GC/CG	53.4%	С	15.1%	с	21.4%
UG/GU	9.8%	G	21.2%	G	26.7%
Other	1.2%	U	27.3%	U	25.1%
Total	: 52.6%	Total:	47.4%		

Figure 2.4 The calculation of single base and base pair in the training sequence

They used the data obtained from those four steps to calculate the final likelihood score. Using all of those four parameters, a MAP likelihood score was calculated. This was to find the most likely secondary structure. MAP is Maximum a posteriori estimation method, similar to maximum likelihood method but uses augmented optimization objective that incorporates a prior distribution over the quantity is to be estimated. The parameter from the third step can be obtained by having a phylogenetic tree. The mutation rate illustrates the relatedness among individuals. Hence this method is called Phylo-SCFG which is a combination of both phylogenetic and SCFG approach.

2.4 Prediction of miRNA Target

2.4.1 The Bioinformatics Approach

Most computational miRNA target programs score and rank known miRNA target (Xia et al. 2009). To predict miRNA targets in plant is easy because the miRNA would mostly perfectly pair with the target, therefore using only complementary scanning of the target mRNA in plants is enough to get a high accuracy prediction program (Rhoades et al. 2002). However, the prediction of animal miRNA target has been a great challenge and as a result most miRNA target predictions have focused specifically on animal miRNA target (Xia et al. 2009).

The challenge of predicting a highly probable miRNA target for animal is the availability of the training data. There are not many currently known miRNA target genes and their exact target sites. This makes the design of algorithms difficult to achieve good specificity and sensitivity. Besides, the biological identification method for miRNA target has not been able to produce high throughput results, and thus make it even more difficult to obtain good training data and constraints for prediction (Stark et al. 2003; Enright et al. 2004; Xia et al. 2009).

However, there are some common characteristic features among miRNA target prediction algorithms. (1) The complementarity of miRNA and the targets, (2) MiRNA target sites conservation among organisms, (3) The stability of the binding energy between the miRNA and the mRNA targets, (4) Lack of complex secondary structure near the miRNA targets, (5) The binding ability of the 5' end of miRNA is stronger than the 3' end. Besides these common features each of the current miRNA target programs

have their own constraints and restrictions imposed to achieve the prediction accuracy that they desire.

2.4.1.1 miRanda

miRanda (Enright et al. 2004) was one of the first miRNA target prediction program. The program basically scanned the potential 3' UTR miRNA targets using complementary matching, thermal stability of the bound, and conservation of the target sites. miRanda scanned 3' UTR region of *Drosophila Melanogaster* genes for the complementary matching, and gave +5 for GC pair and AU pair, +2 for GU pair and -3 for others. Beside of the scoring, they also used what is called empirical rules for the scanning. There are 4 empirical rules which are (1) No mismatch in position 2 to 4, (2) less than five mismatches between position 3 to 12, (3) At least one mismatch between position 9 and L – 5 where L is the total length of the alignment, (4) Less than two mismatch for the rest of the alignment. The positioning is based on the starting position of the 5' end of the miRNA. The calculation of the binding energy is calculated using the RNAFold program (Hofacker et al. 1994). Based on those rules, they selected the top 10 genes as the candidate of miRNA target. These potential targets had the highest score and lowest free energy.

In more detail, miRanda used the initial *Drosophila* miRNA from the RFAM database (Griffiths-Jones et al. 2003). From this database, they are able to obtain 73 unique miRNA sequences. These miRNAs were then put into the FASTA format and were used as their miRNA target test. The target sites were obtained from the *Drosophila melanogaster* 3'UTR available in the Drosophila Genome project from Berkeley University. 3'UTR sequences were available for 14, 287 transcripts, which were

representations for 9,805 *D. melanogaster* genes. Another set of 3'UTR sequences from *D. pseudoobscura* was built from the genome project of this organism in Baylor College of Medicine. Each of the *D. melanogaster* 3'UTR sequences and peptides from the corresponding genes was mapped to the *D. pseudoobscura* using NCBI BLASTn and tBLASTn (Waterman & Eggert 1987). The results from the scan was used to get the 2000 bp candidates of the 3'UTR sequence orthologues of *D. melanogaster* presents in the *D. pseudoobscura*. AVID a global alignment tool (Bray et al. 2003) was used to align the actual *D. melanogaster* 3'UTR sequences to the candidate sequences from the *D. pseudoobscura*. After the alignment, they obtained 12,416 transcripts 8,282 genes from *D. pseudoobscura* mapped to the *D.melanogaster* 3'UTR. Hereafter, a set of 3'UTR from *Anopheles gambiae* was taken from the ENSEMBL database (Ashburner et al. 2000). The orthologues of 3'UTR of *D. melanogaster* and *A. gambiae* were obained by searching the ENSEMBL peptides using BLASTp. There were 9.823 genes of *A. gambiae* identified using this technique.

The algorithm of miRanda is almost the same with the Smith-Waterman alignment algorithm (Smith & Waterman 1981) except that instead of aligning with the same nucleotides, it aligns AU, GC, and GU. The scoring of the alignment and the empirical rules that the algorithm followed has been mentioned above, and the threshold of the scoring was $S \ge 80$ and the minimum energy is $\Delta G \le -14$ kcal/mol. The hits were considered to be conserved if they were also found in *D. pseudoobscura* and *A. gambiae* in addition to being found in the *D. melanogaster*. Conserved target site needed to have at least 80% sequence similarity to *D. pseudoobscura* and at least 60% sequence similarity to *A. gambiae*. All results from this scanning were then scored based on the conservation

and the top 10 is selected as a candidate of miRNA target genes. MiRanda does not allow overlap which means multiple miRNAs that were found to be bound to the same target sites would be selected using greedy algorithm and the miRNA with the highest scores and lowest free energy is the candidate that would be chosen. The negative control of the algorithm was the randomized sequences generated from the 73 miRNA from *D*. *melanogaster*. They generate 100 randomized sequences. The Z-score are then generated from the actual miRNA target count, the averaged random miRNA target counts and their standard deviation. From this algorithm, 535 potential miRNA target genes were identified and 264 of these genes have been validated to have a functional annotation (Enright et al. 2004).

2.4.1.2 TargetScan and TargetScanS

This program was developed for vertebrate miRNA target prediction (Lewis et al. 2003). The program used binding energy and alignment to predict conserve target sites among different species. TargetScan search for perfect complementary match in position 2 to 8 of the miRNA with the target UTRs. The position is counted from the 5' end of the miRNA. This perfect complementary match is then called the "miRNA seed". This seed is then extended as far as possible by allowing GU pairs and stop at a mismatch. The energy of the binding is then measured and a signal noise ratio was introduced to measure the accuracy of the prediction. The number of predicted target is called the "noise".

TargetScan was then further developed into TargetScanS (Lewis et al. 2005). The new program required a conservation of target sites among five organisms which are human, mouse, rat, dog, and chicken. Another improvement was the requirement of the

first position of the complementary match to be adenine. In 2007, another additional constraint was added to TargetScanS (Grimson et al. 2007). The additional constraints was to account for AU rich region near the target sites, the nucleotides pairing in position 13 to 16 in the 3' UTR, at least 15 nucleotides from the stop codon.

TargetScan uses miRNA from human that are also conserved in both mouse and fugu. They identified 121 human miRNA that has perfect homologs in mouse and equally conserved in fugu. From these miRNAs, they eliminate some redundancy by selecting only one representation of miRNAs that have the same seed region. They reduced the number of miRNA to 79 and assigned these miRNAs to a collection of non-redundant miRNA data from human which has mouse homologes (nrMamm). There were 55 miRNAs in nrMamm that has perfect conservation in fugu, hence these miRNAs are assigned to the non redundant miRNA data set, nrVert. The 3' UTR were obtained from all of the human genes and their orthologs to rat, mouse and fugu by using the database from http://www.ensembl.org/EnsMart. The annotated 3'UTR regions available were only 45% of rat genes, and none from the fugu genes. From the 45% of rat genes, 14% of them were less than 50 nucleotides long. Hence they extend these by 2 kb of the 3' flanking sequences.

The 3'UTRs were then searched for the complementary match against the miRNAs seed region. The seed was separated into two regions, the seed on the 5'end of the miRNA and the seed at the 3' end of the miRNA. For the seed at the 5'end of the miRNA, they retrieved 35 nucleotides flanking the seed match and for the seed at the 3' end, 5 nucleotides flanking the match were retrieved. Target sites in the 35 nucleotides flanking region that had a masked basepairs because of the previously performed repeat

masking and the seed match occurred less than 20 nucleotides downstream of a previous seed matches were not used. The basepairing between the miRNA seed and the target was extended with additional flanking basepairs as far as possible in both directions by allowing GU pairs but not Gaps. The rest of the basepairing patterns were predicted using RNAFold on an artificial stem loop attached to the extended matching seed. The artificial stem loop was 5'-GGGCCCGGGULLLLLLACCCGGGCCC-3 where L is anonymous unpaired loop character. Once the RNAFold alignment was done, the stem loop was removed and the free energy of the miRNA and target duplex was estimated using RNAeval (Hofacker et al. 1994). Once this process was done, a Z score is calculated based on the number of seed matches in the UTR and free energy. UTR with no seed matches was given Z score of 1. The formula for Z is $\sum_{k=1}^{n} e^{-Gk/T}$ where n is the number of seed matches in the UTR, Gk is the free energy of the kth evaluated target, and T is a parameter influences the weighting of UTRs with fewer high affinity target sites to the larger number of high affinity target sites. T is assigned by the author along with Rc and Zc which is the cut off value of both R and Z score. R is the ranking of the predicted sequences based on the Z score. Therefore, the final predicted miRNA target sites were all sequences that had higher ranking than the cut off value and higher Z score.

2.4.1.3 RNAhybrid

This program relies on the secondary structure of the complementary binding between the miRNA and the mRNA target (Rehmsmeier et al. 2004). This program is the better version of RNA secondary structure prediction program. It can calculate the binding free energy of the binding of small RNA to the larger RNA faster and accurately. They used *Drosophila* as their test species. The program scanned the input sequence for
the most free energy favorable sites for the miRNA but disallowed binding among the target nucleotides or among miRNA nucleotides. This program was improved in 2006 by allowing the user to choose their own parameters such as the length of "miRNA seed", the allowance of GU pairs, and a seed match speed up (Kruger & Rehmsmeier 2006).

The RNAhybrid program was an extension of Zuker's classic RNA folding prediction algorithm (Zuker & Stiegler 1981). The miRNA and the target were hybridized based on the minimum energy parameter by Mathew et al (Mathews et al. 1999). All possible Minimum Free Energy (MFE) of the miRNA and target was calculated using dynamic programming. Maximum loop allowed was 15 nucleotides. The space required for this algorithm is O (mn) and the time required is O (c²mn) where m is the length of the target, n is the length of the miRNA and c is the maximum length of the loop in either sequence. Additional suboptimal or optimal hybridization calculated the same way but masking the previously found hybridized nucleotides, which was not included in the next calculated sequences.

2.4.1.4 DIANA-microT

The DIANA-microT identifies miRNA targets in animals by using both bioinformatics and biological approach (Kiriakidou et al. 2004). To search for potential target sites, DIANA-microT uses the information from the 5' end seed, central bulge, and the binding of the 3' end. The program calculates binding energy by dynamic programming algorithm to test the binding ability of the miRNA to the target. In addition, it used information of the surrounding protein in the binding sites that would affect the central bulge which in turn would affect the binding ability of the miRNA to the target.

DIANA-microT was limitedly used to scan human 3'UTR retrieved from RefSeq database (Pruitt et al. 2003). They were able to extract 14,180,360 basepairs from 16,759 mRNAs. Repeat element such as Alu was filtered out before the search for the target. The initial search started with 10 miRNAs which were randomly chosen besides being conserved in human and mice. There were two factors that were hypothesized to affect miRNA and target interaction. The first one is binding energy. To address this, they design an algorithm that could identify the putative miRNA target interactions based on imperfectly paired RNA energy. Dynamic programming was used to calculate the free energy by allowing GU pairs. A window of size 38 nucleotides was slid through the 3'UTR for the scanning of potential target. After the scanning, they found the corresponding mouse orthologs of the 3'UTR target found in human.

The second factor that they believe to influence miRNA and target interaction was the miRNA associated protein that influences the length and position of the loop between the miRNA and its target. Argonoute protein family is the potential protein that guides such interaction between the miRNA and the target (Nelson et al. 2004). They determined the constraints of how many nucleotides long a bulge should be to be able to have an effective binding affinity between miRNA and its target. From this experiment, they found that near perfect complementarity between the first 9 nucleotides of miRNA to the target is important for miRNA function, but the pair of the most 5' end of the miRNA is not required. They called this region the proximal region. In contrast to the proximal region, the pairing of the nucleotides at the 3' end of the miRNA to the target (distal region) does not have to be perfect. The activity of the tested target which has a single nucleotide mismatch away from the central bulge and close to the 3' end did not

affect the function of the particular miRNA. The optimal length of the central bulge determined in this experiment was 5 nucleotides. Central bulge with length of 2–4 nucleotides is going to reduce the activity of the miRNA, but central bulge with length of more than 5 nucleotides will erase the activity of the miRNA.

The combination of both computational and the experimental approaches were used to build the DIANA-MicroT and this program was able to predict 94 human miRNA targets from the 10 miRNA selected using a cut off of -30 kcal/mol energy, where 9 of the predicted target were conserved in mouse. They randomized these 10 miRNAs 10 times to yield100 randomized sequences for the same search. There are 371 targets predicted from this randomized sequence, 13 of which were also conserved in mouse. By normalizing the number of the 10 miRNAs, they are able to conclude that the number of predicted target from the randomize sequence in human was 3.7 for each miRNA and for the real miRNAs was 9.4 for each miRNA.

2.4.1.5 PicTar

This algorithm uses probabilistic approach. PicTar stands for probabilistic identification of combinations of target sites (Krek et al. 2005). The algorithm has to principal components: the seed region and free energy binding, however, PicTar allows for imperfect match in the seed region. They divided the seed into two, perfect seed and imperfect seed. A perfect seed is where nucleotides from positions 2 through 7 of the miRNA 5' end match. A seed is imperfect when it contains GU pairs and the energy value is not increased. The energy cut-off value is set to the 33% of the perfectly matched miRNA / mRNA pairs for the perfect seed and 66% of the optimal energy for the imperfect seed. The second part of this algorithm is a probability model. The algorithm

scores the probability that miRNAs target one particular target or one miRNA target several sites. The false positive rate of this program was approximately 30%. This program showed that one miRNA can possibly target 200 mRNAs. The 3'UTR scanned for the target prediction was obtained from the UCSC genome browser using 8 vertebrates which included human, mouse, chimpanzee, rat, dog, chicken, fugu, and zebrafish.

2.4.1.6 RNA22

This program was used to identify miRNA binding sites and their corresponding miRNA: mRNA heteroduplexes (Miranda et al. 2006). Unlike other programs, this program does not rely on cross species sequence conservation as a constraint thereby allowing for the discovery of miRNA target sites that are not present in closely related organisms. This program identifies the potential target first and then finds the corresponding targeting miRNA. RNA22 has been shown to be able to identify most of the currently known heteroduplexes, and suggested that mammalian genome is likely to have much more miRNA targets and that miRNA targets not only 3'UTRs. This algorithm started by a pattern discovery of miRNAs with the Teiresias algorithm (Rigoutsos & Floratos 1998). They believed that miRNAs bind to the target by hybridization of complementary basepairs. The pattern collection that had ascertained by the first step captured sequence features of miRNA that are conserved, hence the reverse complement of such pattern allowed for the identification of conserved sequence in the untranslated region or in this case the putative miRNA binding sites. The region in the UTR that contained a lot of hits of such patterns was called target islands. The target islands as a candidate miRNA targets were then matched with the miRNA that targets

them. They paired each available miRNA with each generated target island for all possible matches. Given a putative miRNA and target island, they inserted a linker GCGGGGACGC (Stark et al. 2003) between the two sequence to form a heteroduplex. They used the Vienna package (Hofacker et al. 1994) to calculate the minimum free energy of the binding and to predict the secondary structure of the heteroduplex. There were three user parameters: M, G, and E, where M is the minimum required number of basepairs between the target and miRNA, G is the maximum number of allowed mismatch and E is the minimum free energy of the binding. The default setting was M=14, G=1 and E=-20 kcal/mol. In rna22, the recognition of binding sites was not by searching a conserved region nor the famous miRNA seed region but the binding sites was recognized by the presence of multiple distinct patterns that had been found by processing known miRNA sequences.

There are different algorithms for miRNA target prediction. These algorithms have some similarities. (1) The design based mainly on the complementarity between the miRNA and the target sequences. (2) Sequence conservation of the target among organisms was important for most programs although some miRNA targets are not conserved. (3) Binding free energy of the miRNA and the target features was considered in the determination of the potential mRNA targets. (4) Secondary structure of the flanking sequence of the target sites can be useful information to do the miRNA target prediction. Those prediction programs mentioned above are summarized in figure 2.5 and have been known as prediction programs using the basic rules of miRNA target prediction. These prediction programs complement each other and with assistance from future experimental data, prediction programs can be improved.

Method	Website	Organism	Algorithm Features
miRanda	http://www.microrna.org/	Human, drosophila, zebrafish	sequence match, conservation and thermal stability
TargetScan/TargetScanS	http://www.targetscan.org/	Human, mouse, rat, dog	the "seed" concept
RNAhybrid	http://bibiserv.techfak.uni-bielefeld.de/mahybrid/	Mammals	calculate the free energy rapidly and accurately
DIANA-microT	http://www.diana.pcbi.upenn.edu/	Human, mouse, rat, drosophila	predict mRNA targets containing single miRNA-recognition element
PicTar	http://pictar.bio.nyu.edu/	Vertebrate	divided the sequence into "perfect seed" and "imperfect seed"
RNA22	http://cbcsrv.watson.ibm.com/ma22.html	Mammals	doesn't consider cross-species sequence conservation

Figure 2.5 miRNA target prediction in animal (Xia et al. 2009)

As experimental data becomes available new characteristics would be added to the miRNA target prediction algorithms. It is important that the interaction between the miRNA and the target sites is well understood. This will enable the prediction algorithms to apply more strict constraints and hence yield a more accurate prediction approach. The interaction of the miRNA and the target mainly studied based on the 5' end seed, the bulge in the middle, and the 3' end. The role of the central bulge in miRNA target interaction was applied in the development of a program called FindTar (Ye et al. 2008). They divided the loops into three categories based on the position of the loop in the miRNA target interaction sites. The first one was the loop starting at position 9 and 11 (position count from the 5' end of the miRNA) which was called the standard loop. The second one is Type I decentered loop which located before position 9, and the third one is Type II decentered loop which started after position 11. After they defined the loop categories, they studied these loop characteristics in known miRNA target interactions and found that the standard loop is the most common in miRNA target interaction. Later, they introduced the central loop rules for miRNA prediction as the following: Loop coefficient was 1 for standard loop; 0.75 for Type II decentered loop; 0 for loop starting before position 7; 0.5 for Type I decentered loop. Loop size was 10 for loop with a size of 1bp; 20 for 2bp; 25 for 3bp. The total loop score was the multiplication of the loop

location coefficient and the loop size coefficient. The addition of this score increased the specificity of the prediction but slightly decreased the sensitivity.

The non-binding site region has some influence on miRNA target interaction. In 2008, Didiano et al (Didiano & Hobert 2008) found that miRNA seed matching is not a reliable predictor for a functional miRNA target interaction. In a nematode experiment they found that the *lsy-6* miRNA cannot regulate the mutated cog-1 3'UTR and heterologous 3'UTR which were predicted to be regulated by *lys-6*. Although they failed to regulate the target genes, the miRNA had the seed region bind nicely to the target. There are 13 different genes that have a match to the *lys-6* but not regulated by *lys-6*. The target gene cog-1 has two binding regions in the 3' UTR that can be recognized by *lys-6*. Position change in this region caused loss of regulation by the miRNA. Hence mutation by insertion or deletion that can shift the position of the 3'UTR would affect the function of the miRNA that target this region. With this study, they concluded that not only the seed matching is important, but the whole sequence of the 3'UTR is also important for the miRNA to have a functional role in the regulation of the potential target.

The secondary structure of the mRNA 3'UTR is also important in the miRNA and target interaction (Zhao et al. 2005). The investigation of the 70bp flanking sequence from the miRNA binding sites concluded that miRNA prefer 3'UTR with less complicated secondary structure as this is more accessible for the miRNA to bind (Zhao et al. 2005). Long et al further studied the influence of the secondary structure to the miRNA target by using the SFold program (Long et al. 2007). According to their analysis, miRNA first binds with four consecutive unpaired nucleotides to form the core basepairs, and the hybrid elongation followed to disturb the local target secondary

structure to form the miRNA and target duplex. The main investigation was the free energy of the original secondary structure disruption, the formation of the miRNA and target pairs, and the final product of the miRNA and target interaction. The data was gathered from the activity of *lin-41* 3'UTR in *C. elegans* and many more experimental data mainly from *C. elegans* and *D. melanogaster*. From this investigation, Long et al concluded that the secondary structure of the target gene complexity in the flanking sequence of the target can disturb the miRNA function. Hence this parameter should be considered for the miRNA prediction program development in the future.

2.4.2 The Biological Approach

2.4.2.1 MiRNA search based on mRNA level

The strategy to search for miRNA target gene was to transfect the mature miRNA duplex into the cells to overexpress miRNA, and the level of mRNA expression would be observed using microarray analysis for potential miRNA target genes (Lim et al. 2005; Nakamura 2005). The limitation of this technique was that it can only detect miRNA that would degrade mRNA but it will not be able to detect miRNA that blocked translation (Xia et al. 2009). Another possible approach was to use Ago protein can bind to both miRNA and mRNA. Ago protein is a protein that is very important for the RISC and miRNA interaction. This protein is needed for miRNA induced silencing. In 2007, Beitzinger et al (Beitzinger et al. 2007) used specific monoclonal antibodies against the protein Ago1 and Ago2 to co-immunoprecipitate Ago bound mRNAs. After identifying these mRNAs, they selected 6 randomly from the Ago1 bound mRNAs and used luciferase based reporter to analyze for possible miRNA existence. From this analysis,

they found that five out of six is miRNA targets while most used computational program (Miranda, TargetScan, PicTar) identified only 60% of Ago-bound mRNAs as miRNA target. It is an indication that the computational method to predict miRNA target still need a lot of improvement. However, Beitzinger et al mention stated that only high throughput validation of isolated mRNAs was able to provide conclusive comparison between the miRNA target identification approaches (Beitzinger et al. 2007). There was a similar work by Easow et al (Easow et al. 2007) but instead of analyzing Ago protein, they analyzed the association of the miRNP (miRNA and RISC complex) and the binding sites. In 2008, Liu et al (Liu et al. 2008) used a reverse screening method to identify miRNA target genes more precisely. They inserted *Drosophila* 3'UTR of the *ccnd1* gene into a luciferase reporter gene vector and transferred it together with a miRNA expression library into HepG2 cells. The function of the miRNA that reduces the luciferase expression was validated *in vivo* and from this experiment, they showed that miR-16 regulate *ccnd1* gene.

2.4.2.2 MiRNA search based on protein level

Some miRNA function as the repressor of translation instead of degrading mRNA directly. In 2008, Selbach et al and Baek et al introduce the technique to identify miRNA by observing the protein level using proteomic assays (Baek et al. 2008; Selbach et al. 2008). They overexpressed human miRNA by transfection of miRNA into the HeLa cells and measure the protein level by mass spectrometry. This revealed that a single miRNA can repress the production of hundreds of protein without changing the mRNA levels. The ability to undertake miRNA target search using proteomic technique would contribute more biological data to develop better miRNA target prediction algorithm.

REFERENCES

ASHBURNER, M. BALL, C. A. BLAKE, J. A. BOTSTEIN, D. BUTLER, H. CHERRY, J. M.

DAVIS, A. P. DOLINSKI, K. DWIGHT, S. S. EPPIG, J. T. HARRIS, M. A. HILL, D. P. ISSEL-

TARVER, L. KASARSKIS, A. LEWIS, S. MATESE, J. C. RICHARDSON, J. E. RINGWALD, M. RUBIN, G. M. SHERLOCK, G. & GENE ONTOLOGY, C. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**(1), 25-29.

BAEK, D. VILLEN, J. SHIN, C. CAMARGO, F. D. GYGI, S. P. & BARTEL, D. P. (2008). The impact of microRNAs on protein output. *Nature* **455**(7209), 64-U38.

BARTEL, D. P. (2004). MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **116**(2), 281-297.

BARTEL, D. P. (2009). MicroRNAs: Target Recognition and Regulatory Functions. *Cell* **136**(2), 215-233.

BEITZINGER, M. PETERS, L. ZHU, J. Y. KREMMER, E. & MEISTER, G. (2007). Identification of human microRNA targets from isolated argonaute protein complexes. *Rna Biology* **4**(2), 76-84.

BENTWICH, I. AVNIEL, A. KAROV, Y. AHARONOV, R. GILAD, S. BARAD, O. BARZILAI, A. EINAT, P. EINAV, U. MEIRI, E. SHARON, E. SPECTOR, Y. & BENTWICH, Z. (2005). Identification of hundreds of conserved and nonconserved human microRNAs. *Nature Genetics* **37**(7), 766-770.

BEREZIKOV, E. CHUNG, W. J. WILLIS, J. CUPPEN, E. & LAI, E. C. (2007). Mammalian mirtron genes. *Molecular Cell* **28**(2), 328-336.

BRAY, N. DUBCHAK, I. & PACHTER, L. (2003). AVID: A global alignment program. *Genome Research* **13**(1), 97-102.

BRENNECKE, J. STARK, A. RUSSELL, R. B. & COHEN, S. M. (2005). Principles of MicroRNA-target recognition. *Plos Biology* **3**(3), 404-418.

CAI, X. Z. HAGEDORN, C. H. & CULLEN, B. R. (2004). Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *Rna-a Publication of the Rna Society* **10**(12), 1957-1966.

CHAUDHURI, K. & CHATTERJEE, R. (2007). MicroRNA detection and target prediction: Integration of computational and experimental approaches. *DNA and Cell Biology* **26**(5), 321-337.

DIDIANO, D. & HOBERT, O. (2008). Molecular architecture of a miRNA-regulated 3 ' UTR. *Rna-a Publication of the Rna Society* **14**(7), 1297-1317.

EASOW, G. TELEMAN, A. A. & COHEN, S. M. (2007). Isolation of microRNA targets by miRNP immunopurification. *Rna-a Publication of the Rna Society* **13**(8), 1198-1204.

ENRIGHT, A. J. JOHN, B. GAUL, U. TUSCHL, T. SANDER, C. & MARKS, D. S. (2004).

MicroRNA targets in Drosophila. Genome Biology 5(1).

EULALIO, A. HUNTZINGER, E. NISHIHARA, T. REHWINKEL, J. FAUSER, M. & IZAURRALDE,

E. (2009). Deadenylation is a widespread effect of miRNA regulation. *Rna-a Publication* of the Rna Society **15**(1), 21-32.

FALLER, M. & GUO, F. (2008). MicroRNA biogenesis: there's more than one way to skin a cat. *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms* **1779**(11), 663-667.

FONTANA, L. SORRENTINO, A. CONDORELLI, G. & PESCHLE, C. (2008). Role of

microRNAs in haemopoiesis, heart hypertrophy and cancer. Biochemical Society

Transactions 36, 1206-1210.

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

GRIFFITHS-JONES, S. BATEMAN, A. MARSHALL, M. KHANNA, A. & EDDY, S. R. (2003). Rfam: an RNA family database. *Nucleic Acids Research* **31**(1), 439-441.

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

GRIFFITHS-JONES, S. SAINI, H. K. VAN DONGEN, S. & ENRIGHT, A. J. (2008). miRBase: tools for microRNA genomics. *Nucleic Acids Research* **36**, D154-D158.

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.

HE, L. THOMSON, J. M. HEMANN, M. T. HERNANDO-MONGE, E. MU, D. GOODSON, S.

POWERS, S. CORDON-CARDO, C. LOWE, S. W. HANNON, G. J. & HAMMOND, S. M. (2005).

A microRNA polycistron as a potential human oncogene. *Nature* **435**(7043), 828-833.

HELVIK, S. A. SNOVE, O. & SAETROM, P. (2007). Reliable prediction of Drosha

processing sites improves microRNA gene prediction. *Bioinformatics* 23(2), 142-149.

HOFACKER, I. L. FONTANA, W. STADLER, P. F. BONHOEFFER, L. S. TACKER, M. &

SCHUSTER, P. (1994). FAST FOLDING AND COMPARISON OF RNA SECONDARY

STRUCTURES. Monatshefte Fur Chemie 125(2), 167-188.

JI, X. H. (2008). The mechanism of RNase III action: How Dicer dices. In *Rna Interference* pp. 99-116.

JIANG, P. WU, H. WANG, W. MA, W. SUN, X. & LU, Z. (2007). MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features. *Nucleic Acids Research* **35**, W339-W344.

KHVOROVA, A. REYNOLDS, A. & JAYASENA, S. D. (2003). Functional siRNAs and rniRNAs exhibit strand bias. *Cell* **115**(2), 209-216.

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.

KREK, A. GRUN, D. POY, M. N. WOLF, R. ROSENBERG, L. EPSTEIN, E. J. MACMENAMIN,

P. DA PIEDADE, I. GUNSALUS, K. C. STOFFEL, M. & RAJEWSKY, N. (2005). Combinatorial microRNA target predictions. *Nature Genetics* **37**(5), 495-500.

KROL, J. & KRZYZOSIAK, W. J. (2004). Structural aspects of microRNA biogenesis. *Iubmb Life* **56**(2), 95-100.

KRUGER, J. & REHMSMEIER, M. (2006). RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Research* **34**, W451-W454.

LAI, E. C. TOMANCAK, P. WILLIAMS, R. W. & RUBIN, G. M. (2003). Computational identification of Drosophila microRNA genes. *Genome Biology* **4**(7).

LAU, N. C. LIM, L. P. WEINSTEIN, E. G. & BARTEL, D. P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science* **294**(5543), 858-862.

LEE, O. & KIM, V. N. (2004). Evidence that microRNA genes are transcribed by RNA polymerase II. *Cell Structure and Function* **29**, 68-68.

LEE, Y. KIM, M. HAN, J. J. YEOM, K. H. LEE, S. BAEK, S. H. & KIM, V. N. (2004). MicroRNA genes are transcribed by RNA polymerase II. *Embo Journal* 23(20), 4051-4060.

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.

LEWIS, B. P. SHIH, I. H. JONES-RHOADES, M. W. BARTEL, D. P. & BURGE, C. B. (2003). Prediction of mammalian microRNA targets. *Cell* **115**(7), 787-798.

LI, G.-X. NING, X.-M. LI, X.-J. WU, Z.-S. & YANG, G.-S. (2009). Identification of Suitable Reference for Quantitative RT-PCR Assays of miR-103 in Pig Tissues. *Zhongguo Shengwu Huaxue yu Fenzi Shengwu Xuebao* **25**(12), 1149-1154.

LI, L. XU, J. Z. YANG, D. Y. TAN, X. R. & WANG, H. F. (2010). Computational approaches for microRNA studies: a review. *Mammalian Genome* **21**(1-2), 1-12.

LI, W. & RUAN, K. C. (2009). MicroRNA detection by microarray. *Analytical and Bioanalytical Chemistry* **394**(4), 1117-1124.

LIM, L. P. GLASNER, M. E. YEKTA, S. BURGE, C. B. & BARTEL, D. P. (2003a). Vertebrate MicroRNA genes. *Science* **299**(5612), 1540-1540.

LIM, L. P. LAU, N. C. GARRETT-ENGELE, P. GRIMSON, A. SCHELTER, J. M. CASTLE, J.

BARTEL, D. P. LINSLEY, P. S. & JOHNSON, J. M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**(7027), 769-773.

LIM, L. P. LAU, N. C. WEINSTEIN, E. G. ABDELHAKIM, A. YEKTA, S. RHOADES, M. W. BURGE, C. B. & BARTEL, D. P. (2003b). The microRNAs of Caenorhabditis elegans. *Genes & Development* **17**(8), 991-1008.

LIU, Q. FU, H. J. SUN, F. ZHANG, H. M. TIE, Y. ZHU, J. XING, R. Y. SUN, Z. X. & ZHENG, X. F. (2008). miR-16 family induces cell cycle arrest by regulating multiple cell cycle genes. *Nucleic Acids Research* **36**(16), 5391-5404.

LONG, D. LEE, R. WILLIAMS, P. CHAN, C. Y. AMBROS, V. & DING, Y. (2007). Potent effect of target structure on microRNA function. *Nature Structural & Molecular Biology* **14**(4), 287-294.

MARTIN, G. SCHOUEST, K. KOVVURU, P. & SPILLANE, C. (2007). Prediction and validation of microRNA targets in animal genomes. *Journal of Biosciences* **32**(6), 1049-1052.

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.

MAZIERE, P. & ENRIGHT, A. J. (2007). Prediction of microRNA targets. *Drug Discovery Today* **12**(11-12), 452-458.

MIRANDA, K. C. HUYNH, T. TAY, Y. ANG, Y. S. TAM, W. L. THOMSON, A. M. LIM, B. & RIGOUTSOS, I. (2006). A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. *Cell* **126**(6), 1203-1217.

MURCHISON, E. P. & HANNON, G. J. (2004). MiRNAs on the move: miRNA biogenesis and the RNAi machinery. *Current Opinion in Cell Biology* **16**(3), 223-229.

NAKAMURA, H. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Scientist* **19**(6), 25-25.

NAM, J. W. KIM, J. KIM, S. K. & ZHANG, B. T. (2006). ProMiR II: a web server for the probabilistic prediction of clustered, nonclustered, conserved and nonconserved microRNAs. *Nucleic Acids Research* **34**, W455-W458.

NAM, J. W. SHIN, K. R. HAN, J. J. LEE, Y. KIM, V. N. & ZHANG, B. T. (2005). Human microRNA prediction through a probabilistic co-learning model of sequence and structure. *Nucleic Acids Research* **33**(11), 3570-3581.

NELSON, P. T. HATZIGEORGIOU, A. G. & MOURELATOS, Z. (2004). miRNP : mRNA association in polyribosomes in a human neuronal cell line. *Rna-a Publication of the Rna Society* **10**(3), 387-394.

OKAMURA, K. PHILLIPS, M. D. TYLER, D. M. DUAN, H. CHOU, Y. T. & LAI, E. C. (2008). The regulatory activity of microRNA star species has substantial influence on microRNA and 3 ' UTR evolution. *Nature Structural & Molecular Biology* **15**(4), 354-363.

PEDERSEN, J. S. BEJERANO, G. SIEPEL, A. ROSENBLOOM, K. LINDBLAD-TOH, K. LANDER,

E. S. KENT, J. MILLER, W. & HAUSSLER, D. (2006). Identification and classification of conserved RNA secondary structures in the human genome. *Plos Computational Biology* **2**(4), 251-262.

PRUITT, K. D. TATUSOVA, T. & MAGLOTT, D. R. (2003). NCBI Reference Sequence Project: update and current status. *Nucleic Acids Research* **31**(1), 34-37.

RABANI, M. KERTESZ, M. & SEGAL, E. (2008). Computational prediction of RNA structural motifs involved in posttranscriptional regulatory processes. *Proceedings of the National Academy of Sciences of the United States of America* **105**(39), 14885-14890.

REHMSMEIER, M. STEFFEN, P. HOCHSMANN, M. & GIEGERICH, R. (2004). Fast and effective prediction of microRNA/target duplexes. *Rna-a Publication of the Rna Society* **10**(10), 1507-1517.

REINHART, B. J. SLACK, F. J. BASSON, M. PASQUINELLI, A. E. BETTINGER, J. C. ROUGVIE, A. E. HORVITZ, H. R. & RUVKUN, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**(6772), 901-906.

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.

RIGOUTSOS, I. & FLORATOS, A. (1998). Combinatorial pattern discovery in biological sequences: the TEIRESIAS algorithm. *Bioinformatics* **14**(1), 55-67.

RODRIGUEZ, A. GRIFFITHS-JONES, S. ASHURST, J. L. & BRADLEY, A. (2004).

Identification of mammalian microRNA host genes and transcription units. *Genome Research* **14**(10A), 1902-1910.

SAKAKIBARA, Y. BROWN, M. HUGHEY, R. MIAN, I. S. SJOLANDER, K. UNDERWOOD, R. C.
& HAUSSLER, D. (1994). STOCHASTIC CONTEXT-FREE GRAMMARS FOR
TRANSFER-RNA MODELING. *Nucleic Acids Research* 22(23), 5112-5120.
SCHUSTER, P. FONTANA, W. STADLER, P. F. & HOFACKER, I. L. (1994). FROM
SEQUENCES TO SHAPES AND BACK - A CASE-STUDY IN RNA SECONDARY
STRUCTURES. *Proceedings of the Royal Society of London Series B-Biological Sciences* 255(1344), 279-284.

SCHWARZ, D. S. & ZAMORE, P. D. (2002). Why do miRNAs live in the miRNP? Genes & Development 16(9), 1025-1031.

SELBACH, M. SCHWANHAUSSER, B. THIERFELDER, N. FANG, Z. KHANIN, R. & RAJEWSKY,
N. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature*455(7209), 58-63.

SEWER, A. PAUL, N. LANDGRAF, P. ARAVIN, A. PFEFFER, S. BROWNSTEIN, M. J. TUSCHL, T. VAN NIMWEGEN, E. & ZAVOLAN, M. (2005). Identification of clustered microRNAs using an ab initio prediction method. *Bmc Bioinformatics* **6**.

SMITH, T. F. & WATERMAN, M. S. (1981). IDENTIFICATION OF COMMON

MOLECULAR SUBSEQUENCES. Journal of Molecular Biology 147(1), 195-197.

STARK, A. BRENNECKE, J. RUSSELL, R. B. & COHEN, S. M. (2003). Identification of Drosophila MicroRNA targets. *Plos Biology* **1**(3), 397-409.

TYAGI, S. VAZ, C. GUPTA, V. BHATIA, R. MAHESHWARI, S. SRINIVASAN, A. & BHATTACHARYA, A. (2008). CID-miRNA: A web server for prediction of novel miRNA precursors in human genome. *Biochemical and Biophysical Research Communications*

372(4), 831-834.

WANG, L. WANG, H. & MENG, X.-M. (2010). MicroRNAs and Cardiac Diseases. *Zhongguo Shengwu Huaxue yu Fenzi Shengwu Xuebao* **26**(1), 9-14.

WANG, X. J. REYES, J. L. CHUA, N. H. & GAASTERLAND, T. (2004). Prediction and identification of Arabidopsis thaliana microRNAs and their mRNA targets. *Genome Biology* **5**(9).

WANG, X. W. ZHANG, J. GU, J. HE, T. ZHANG, X. G. LI, Y. D. & LI, F. (2005). MicroRNA identification based on sequence and structure alignment. *Bioinformatics* **21**(18), 3610-3614.

WATERMAN, M. S. & EGGERT, M. (1987). A NEW ALGORITHM FOR BEST
SUBSEQUENCE ALIGNMENTS WITH APPLICATION TO TRANSFER RNARIBOSOMAL-RNA COMPARISONS. *Journal of Molecular Biology* 197(4), 723-728.
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(5), 855-862.
WILLIAMS, A. E. (2008). Functional aspects of animal microRNAs. *Cellular and Molecular Life Sciences* 65(4), 545-562.

XIA, W. CAO, G. J. & SHAO, N. S. (2009). Progress in miRNA target prediction and identification. *Science in China Series C-Life Sciences* **52**(12), 1123-1130.

XU, G. ZHANG, Y. JIA, H. LI, J. LIU, X. ENGELHARDT, J. F. & WANG, Y. (2009). Cloning and identification of microRNAs in bovine alveolar macrophages. *Molecular and Cellular Biochemistry* **332**(1-2), 9-16.

XUE, C. H. LI, F. HE, T. LIU, G. P. LI, Y. D. & ZHANG, X. G. (2005). Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine. *Bmc Bioinformatics* **6**.

YE, W. B. LV, Q. WONG, C. K. A. HU, S. FU, C. HUA, Z. CAI, G. P. LI, G. X. YANG, B. B. & ZHANG, Y. (2008). The Effect of Central Loops in miRNA:MRE Duplexes on the Efficiency of miRNA-Mediated Gene Regulation. *Plos One* **3**(3).

ZHAO, Y. SAMAL, E. & SRIVASTAVA, D. (2005). Serum, response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Circulation* 112(17), 340.

ZHOU, X. F. RUAN, J. H. WANG, G. D. & ZHANG, W. X. (2007). Characterization and identification of microRNA core promoters in four model species. *Plos Computational Biology* **3**(3), 412-423.

ZUKER, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**(13), 3406-3415.

ZUKER, M. & STIEGLER, P. (1981). OPTIMAL COMPUTER FOLDING OF LARGE RNA SEQUENCES USING THERMODYNAMICS AND AUXILIARY INFORMATION. *Nucleic Acids Research* **9**(1), 133-148.

CHAPTER 3

MIR-EXPLORE : PREDICTING MicroRNA PRECURSORS BY CLASS GROUPING AND SECONDARY STRUCTURE POSITIONAL ALIGNMENT¹

¹Sebastian, B., S.E. Aggrey. To be submitted to Journal of Bioinformatics and Biology Insights.

ABSTRACT

MicroRNAs (miRNAs) are small non-coding RNAs which regulate gene expressions by targeting the mRNAs especially in the 3'UTR regions. The identification of miRNAs has been done by biological experiment and computational prediction. The computational prediction approach has been done by two major methods, comparative and non-comparative. Comparative method is dependent on the conservation of the miRNA sequences and secondary structure. Non-comparative method in the other hand does not rely on conservation. We hypothesized that each miRNA class has its own unique set of features therefore, grouping miRNA by classes before using them as training data will improve sensitivity and specificity. The average sensitivity was 88.62% for Mir-Explore which rely on within miRNA class alignment, and 70.82% for mir-Abela with relies on global alignment. Compared with global alignment, grouping miRNA by classes yield a better sensitivity with very high specificity for pre-miRNA prediction even when a simple positional based secondary and primary structure alignment are used.

Keywords : miRNA, computational prediction, mir-explore

INTRODUCTION

Mature miRNAs are small 22 nucleotides long non-coding RNAs. They are expressed in a wide variety of organisms including viruses, plants, and animals (Bartel & Bartel 2003; Lim et al. 2003a; Pfeffer et al. 2004). They have significant role in post transcriptional control of eukaryotes genome causing degradation of the mRNA transcripts or blocking translation (Lau et al. 2001; Lee & Ambros 2001). In most of the eukaryotic genome, miRNA genes are transcribed by RNA-polymerase-II to primary miRNA (pri-miRNA) (Lee et al. 2004). This pri-miRNA is then further processed by RNAse-III endonuclease Drosha into precursor miRNA (pre-miRNA) which is 60 – 100 nucleotides long in average and forms a stem loop secondary structure (Denli et al. 2004). The pre-miRNA is then processed to form the small double stranded RNAs by the endonuclease Dicer which also initiates the formation of the RNA induced silencing complex (RISC) (Bernstein et al. 2001). One strand of the double stranded RNA is incorporated with the RISC and targets the mRNA transcripts to block gene expression (Griffiths-Jones et al. 2006).

MiRNAs have been known to be involved in some critical diseases including heart disease (Shen et al. 2010) and cancer (Zhao et al. 2010). In cancer, miRNAs can function as both oncogene and suppressor (Shenouda & Alahari 2009). Several classes of miRNAs such as miR-15, let-7, miR-16, miR-342, miR-223, and miR-107 have been reported to be involved in acute promyelocytic leukemia (APL) (Garzon et al. 2007). The oncogene YES and STAT1 which are responsible in the proliferation of the colon cancer cells are targeted by miR-145, thereby making this particular miRNA class a colon cancer suppressor(Gregersen et al. 2010). In lung cancer, miR-34 has been shown to have an

important role on the PRIMA-1 regulation, which is a small molecule that restores the cancer cell suppression function (Duan et al. 2010). The identification of the involvement of miRNAs in several cancers has assisted researcher to develop some therapeutics treatment for a possible cancer cure (Wiggins et al. 2010). The significant role of miRNAs in human health has made the continuing discovery of novel miRNAs in the genome important.

Laboratory experiment has been conducted to discover miRNAs by direct cloning and short RNA sequencing. However, it is difficult to identify miRNAs with low level of expression using only laboratory experiment technique (Lai et al. 2003). Hence, computational predictions are needed to support the identification of novel miRNAs. There are two major computational approaches for predicting miRNAs. They are comparative and non-comparative methods. Comparative method relies on the conservation of the miRNAs across the organism genome. Some examples of computational method using comparative method are miRScan (Lim et al. 2003a; Lim et al. 2003b), miRAlign (Wang et al. 2005), ERPIN (Legendre et al. 2005) and microHarvester (Dezulian et al. 2006). For miRNAs that do not have conserve sequences, non-comparative methods are needed. Some examples of non-comparative method are triplet-SVM (Xue et al. 2005), miPred (Jiang et al. 2007), ProMir (Nam et al. 2005) and miR-abela (Sewer et al. 2005). The main idea of the non-comparative method is to have the computer to learn and distinguish between real pre-miRNA and pseudo pre-miRNA. Hence most of the non-comparative methods are based on the utilization of Support Vector Learning Machine (SVM) and Hidden Markov Model (HMM). Many of the miRNA prediction algorithms were developed based on broad assumptions (Zhang et al.

2006). A study by Bram and Aggrey (Sebastian & Aggrey 2008) showed marked variability in sensitivity and specificity in predicting chicken pre-miRNA across classes using ERPIN, ProMir and miR-abela.

In this study, we create a comparative method to show that grouping chicken miRNAs by classes increase sensitivity and specificity of the prediction method even if only a very simple direct positional secondary structure alignment is used. The basic idea of our method is to create a consensus structure of the pre-miRNA for each miRNA classes and use this consensus structure to do the alignment with the query sequence.

MATERIALS AND METHODS

Training Data

A set of data was taken from the known pre-miRNA in chicken, human, mouse, zebra fish, fugu, worms, frog, chimpanzee, gorilla, platypus, pig, fruit fly, and buffalo which are available in the microRNA database miRBase (Griffiths-Jones et al. 2006). The data selected was based on the available miRNA classes in chicken. For example, if chicken has miR-1 class, then all available known miR-1 class pre-miRNA in the other organisms will be selected in this data set. The training data is taken from 80% of the sequences in the data set that best represents each miRNA classes. In order to give a good consensus structure, only classes that has 5 or more known pre-miRNA will be taken as the training data.

Positive Data and Negative Data

The remaining 20% of the sequences that was not used as the training data will be used as our positive data set.

The negative data was taken from all coding sequences in human and mouse taken from the UCSC genome browser (Kuhn et al. 2009). These coding sequences were scanned for the hairpin-like structures as a final negative data set. There were 33,932 hairpin-like structures in mouse and 36,662 hairpin-like structures in human. The hairpinlike structures scanning was done by using the program written by Dr. Alain Sewer for miR-abela (Sewer et al. 2005). Negative data for comparison with ProMir was taken from 5,000 randomly picked sequences of length 300 nucleotides from coding genes, tRNA, and rRNA of chicken taken from the UCSC genome browser.

The programming and testing

The training data set was aligned using MARNA (Siebert & Backofen 2005) to get the consensus structure with 80% primary and secondary structure identity. The identity percentage is defined by how many of the nucleotides are conserved among all of the training sequences in a particular column of the secondary structure. These consensus structures along with their primary sequence information were then used for alignment with the query sequence. The query sequence was limited to a minimum length of 300 nucleotides. The minimum length of 300 nucleotides is to allow our program to start the alignment process. The average length of the consensus sequence is 150 nucleotides and our program cannot take a query sequence that is shorter than the length of the consensus structure. The alignment was done based on the information of the position of the secondary and primary structure. An alignment data from MARNA provided the consensus secondary structure on the first row along with the primary structure information in the rest of the rows (Figure 1). A sliding window was created with a length the same as the number of characters including gaps and unpaired nucleotides in

the output of the MARNA alignment. The sliding window was used to scan the query sequence starting from position one. The consensus structures are made up of some gaps and hairpin structures which consist of stem and loop. The stem has their corresponding nucleotide pairs in a particular position. In Figure 1, the stem starts in position 1 and 75. The query sequence is a primary sequence without any gaps. Hence the alignment inserts gaps to the query sequence and shifting the gaps to align with the consensus structure and its corresponding nucleotide pairs in the stem. If an alignment matches the secondary structure and the corresponding stem nucleotides in the exact position shown in the MARNA alignment, then it was counted as one, otherwise it is zero. A match is defined as the same nucleotides in a particular column of a consensus secondary structure of the query sequence and the training sequence. As it has been mentioned above, our scoring method is a simple 0 and 1 to represent a match or not a match. The maximum score that an alignment can get is the same as the number of pair in the consensus secondary structure. Because each classes of miRNA have their own number of pairs in the secondary structure, the score of each miRNA classes will be different. This leads to the need of normalizing the score so that every miRNA classes will have a standardized score. The standardized score is calculated as follow.

Standardized Score = $(NMA / NMC) \times 100\%$

NMA = Number of matched secondary structure and nucleotide stems in the alignment.

NMC = Number of nucleotide pair in the stem of the consensus structure.

The same scoring system was used for both negative and positive data. Since we are using 80% identity in the construction of the consensus structure, and this approach is a direct position to position exact comparison, hence the program will consider a hit if the

standardized score from a query sequence is at least 80%. We compared the current approach (Mir-Explore) with miR-abela (Sewer et al. 2005) and ProMir (Nam et al. 2005). MiR-abela was tested using the same positive and negative data, while ProMir was tested using the negative data from 5,000 randomly selected sequences of length 300 nucleotides from chicken coding genes, tRNA and rRNA; The positive data for ProMir test was taken from the same data set used to test the other two programs, except that only chicken data was used. We tested ProMir with only chicken positive data because this program requires sequences to be inputted individually, which is laborious and time consuming. Therefore, we tested ProMir against Mir-Explore and miR-abela using only chicken data, and Mir-Explore was tested against miR-abela using data set across species.

The sensitivity and specificity were calculated as follow:

Sensitivity = TP / (TP+FN), where

TP: True Positive, the number of pre-miRNA predicted as pre-miRNA

FN: False Negative, the number of pre-miRNA predicted as non pre-miRNA Specificity = TN / (TN+FP)

TN: True Negative, the number of non pre-miRNA predicted as non pre-miRNA FP: False Positive, the number of non pre-miRNA predicted as pre-miRNA

For accuracy measurement, we use Mathews Correlation Coefficient (MCC) (Matthews 1975). MCC is a measure of the quality of binary classification. It takes in to account true and false positive and negative. The value of MCC ranges between -1 and +1 where +1 represents a perfect prediction, 0 represents a random prediction and -1 represents completely inverted predictions. The formula to calculate the MCC is as

follow, where TP is True Positive, TN is True Negative, FP is False Positive and FN is False Negative.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

RESULTS AND DISCUSSION

The sensitivity and specificity for miR-Explore, miR-abela, and ProMir using only chicken data is shown in Table 1. The specificity for miR-Explore and miR-abela was 99.99% and that of ProMir was 99.00%. The specificity of these programs was high using chicken data. However, there was marked variability in their sensitivity. Mir-Explore had 88% sensitivity, whereas, mir-Abela and ProMir had 78% and 53% sensitivity, respectively. The differences in sensitivity could be due to that fact that both miR-abela and ProMir were developed using mostly human and other mammalian training data whereas Mir-Explore included chicken as part of the training data. The training data in the secondary structure profiling can affect the sensitivity of a program that is dependent on the program. In principle, ProMir, a probabilistic co-learning method based on hidden Markov model (Nam et al. 2005) was developed to identify both close and distant homologs. Our results demonstrate that either the chicken is too distant from human or some of generalized assumptions under which ProMir was develop do not hold. ProMir scans the stem of the stem loop candidates to determine the signal of the Drosha cleavage site. However, multiple factors govern the Drosha cleavage site (Zeng et al. 2005). Therefore, it is possible that the ProMir did not capture most of the factors affecting Drosha cleavage in chicken thereby limiting its sensitivity. Whereas, miR-abela is a program that is created based on Support Vector Machine (SVM). This program takes both real pre-miRNA and pseudo pre-miRNA as the training data and used general feature of pre-miRNA to let the computer learn and distinguish between the two.

In general, there are some global features of pre-miRNAs that can be used as data for pre-miRNA prediction programs. These features include but not limited to the nucleotide length, number of bulges, and minimum free energy. In the most recent miRNA prediction programs these features have been being generalized (Hou et al. 2008). The advantage of using these generalized miRNA features is that they are able to predict novel miRNAs that belong to previously unknown classes. Despite of the advantage of using the generalized miRNA features, from the result shown in Table 1 suggest that using global features without adequate training data that can represent all organisms will have reduced sensitivity in identifying new miRNA for species that are not well represented in the training data.

The test result of comparing miR-Explore and miR-abela using expanded set of positive and negative data is also shown in Table 1. The detailed predictive results of different organisms for Mir-Explore and mir-Abela are provided in Supplementary 1 and 2, respectively. The central hypothesis for this study was that each miRNA class has its own unique set of features therefore, grouping miRNA by classes before using them as training data will improve sensitivity and specificity. Both Mir-Explore and mir-Abela are highly proficient in detecting true negative miRNA. The sensitivity of Mir-Explore was higher than mir-Abela for every species compared including humans. The average sensitivity was 88.62% for Mir-Explore, and 70.82% for mir-Abela. The calculation of MCC for miR-Explore yield a coefficient of 0.90 whereas miR-abela has a coefficient of 0.75. Compared with global alignment, grouping miRNA by classes yield a better

sensitivity with very high specificity for pre-miRNA prediction even when a simple positional based secondary and primary structure alignment are used. The grouping technique also yield a higher MCC coefficient compared to the program using generalized features of pre-miRNA. It can be argued that as much as each class has its own unique features, there are other features that are conserved across classes and species. The ability to predict miRNA with increased sensitivity will depend on the amount of conserved elements captured in the secondary structure. The training data for mir-21 were from human, chicken and mouse, yet Mir-Explore had 100% sensitivity in predicting mir-21 in other species that were not used in the training data. The features of mir-21 could have been conserved well across species. On the other hand, mir-125 could be least conserved even within the class and across species, and any predictive algorithm based on global alignment would yield relatively poor sensitivity.

There are some limitations to miR-Explore. First, this approach is dependent on previously known miRNA precursor class and can only predict novel miRNA with that particular class. Second, MiR-Explore rely on the conservation of miRNAs within a class and cross species and the availability of known miRNA data. Where there are not enough known miRNA data that can be used to build the consensus structure, the sensitivity may decline. Third, the alignment method of this approach is very simple. We used the positional information of the secondary structure along with the corresponding primary structure nucleotide, yet we were able to achieve a high sensitivity and specificity. Improving the alignment algorithm could improve the sensitivity and specificity.

CONCLUSION

It is difficult to develop a perfect miRNA prediction program, so to detect miRNA computationally, more than one program may be needed (Jiang et al. 2007). In this study we have shown that grouping miRNA by classes before using them as training data will improve sensitivity and specificity. However, this approach can only predict pre-miRNA of known classes. Even though ProMir and miR-abela may not have good sensitivity as miR-Explore, as an ab initio method, they have the potential to predict a novel miRNA on an unknown class. Mir-Explore can be used to confirm novel candidate miRNA predicted by the ab initio methods from known classes. Each program has its own strengths and limitations that can complement each other. MiR-Explore is a new technique that can contribute to the future discovery of novel miRNA.

ACKNOLEDGEMENT

We thank Dr. Alain Sewer from Université de Lausanne, Switzerland who has provided us with the program to scan for the hairpin-like structures in the human and mouse coding sequences.

REFERENCES

BARTEL, B. & BARTEL, D. P. (2003). MicroRNAs: At the root of plant development? *Plant Physiology* **132**(2), 709-717.

BERNSTEIN, E. CAUDY, A. A. HAMMOND, S. M. & HANNON, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**(6818), 363-366.

DENLI, A. M. TOPS, B. B. J. PLASTERK, R. H. A. KETTING, R. F. & HANNON, G. J. (2004). Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**(7014), 231-235.

DEZULIAN, T. REMMERT, M. PALATNIK, J. F. WEIGEL, D. & HUSON, D. H. (2006). Identification of plant microRNA homologs. *Bioinformatics* **22**(3), 359-360.

DUAN, W. R. GAO, L. WU, X. WANG, L. NANA-SINKAM, S. P. OTTERSON, G. A. &

VILLALONA-CALERO, M. A. (2010). MicroRNA-34a is an important component of

PRIMA-1-induced apoptotic network in human lung cancer cells. *International Journal of Cancer* **127**(2), 313-320.

GARZON, R. PICHIORRI, F. PALUMBO, T. VISENTINI, M. AQEILAN, R. CIMMINO, A. WANG,
H. SUN, H. VOLINIA, S. ALDER, H. CALIN, G. A. LIU, C. G. ANDREEFF, M. & CROCE, C.
M. (2007). MicroRNA gene expression during retinoic acid-induced differentiation of
human acute promyelocytic leukemia. *Oncogene* 26(28), 4148-4157.
GREGERSEN, L. H. JACOBSEN, A. B. FRANKEL, L. B. WEN, J. KROGH, A. & LUND, A. H.

(2010). MicroRNA-145 Targets YES and STAT1 in Colon Cancer Cells. *PLoS One* **5**(1), Article No.: e8836.

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

HOU, Y.-Y. YING, X.-M. & LI, W.-J. (2008). Computational approaches to microRNA discovery. *Yichuan* **30**(6), 687-696.

JIANG, P. WU, H. WANG, W. MA, W. SUN, X. & LU, Z. (2007). MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features. *Nucleic Acids Research* **35**, W339-W344.

KUHN, R. M. KAROLCHIK, D. ZWEIG, A. S. WANG, T. SMITH, K. E. ROSENBLOOM, K. R.

RHEAD, B. RANEY, B. J. POHL, A. PHEASANT, M. MEYER, L. HSU, F. HINRICHS, A. S.

HARTE, R. A. GIARDINE, B. FUJITA, P. DIEKHANS, M. DRESZER, T. CLAWSON, H. BARBER,

G. P. HAUSSLER, D. & KENT, W. J. (2009). The UCSC Genome Browser Database:

update 2009. Nucleic Acids Research 37, D755-D761.

LAI, E. C. TOMANCAK, P. WILLIAMS, R. W. & RUBIN, G. M. (2003). Computational identification of Drosophila microRNA genes. *Genome Biology* **4**(7).

LAU, N. C. LIM, L. P. WEINSTEIN, E. G. & BARTEL, D. P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science* **294**(5543), 858-862.

LEE, R. C. & AMBROS, V. (2001). An extensive class of small RNAs in Caenorhabditis elegans. *Science* **294**(5543), 862-864.

LEE, Y. KIM, M. HAN, J. J. YEOM, K. H. LEE, S. BAEK, S. H. & KIM, V. N. (2004). MicroRNA genes are transcribed by RNA polymerase II. *Embo Journal* 23(20), 4051-4060. LEGENDRE, M. LAMBERT, A. & GAUTHERET, D. (2005). Profile-based detection of microRNA precursors in animal genomes. *Bioinformatics* **21**(7), 841-845.

LIM, L. P. GLASNER, M. E. YEKTA, S. BURGE, C. B. & BARTEL, D. P. (2003a). Vertebrate MicroRNA genes. *Science* **299**(5612), 1540-1540.

LIM, L. P. LAU, N. C. WEINSTEIN, E. G. ABDELHAKIM, A. YEKTA, S. RHOADES, M. W.

BURGE, C. B. & BARTEL, D. P. (2003b). The microRNAs of Caenorhabditis elegans. *Genes & Development* 17(8), 991-1008.

MATTHEWS, B. W. (1975). COMPARISON OF PREDICTED AND OBSERVED SECONDARY STRUCTURE OF T4 PHAGE LYSOZYME. *Biochimica Et Biophysica Acta* **405**(2), 442-451.

NAM, J. W. SHIN, K. R. HAN, J. J. LEE, Y. KIM, V. N. & ZHANG, B. T. (2005). Human microRNA prediction through a probabilistic co-learning model of sequence and structure. *Nucleic Acids Research* **33**(11), 3570-3581.

PFEFFER, S. ZAVOLAN, M. GRASSER, F. A. CHIEN, M. C. RUSSO, J. J. JU, J. Y. JOHN, B. ENRIGHT, A. J. MARKS, D. SANDER, C. & TUSCHL, T. (2004). Identification of virusencoded microRNAs. *Science* **304**(5671), 734-736.

SEBASTIAN, B. & AGGREY, S. E. (2008). Specificity and sensitivity of PROMIR, ERPIN and MIR-ABELA in predicting pre-microRNAs in the chicken genome. *In Silico Biology* **8**, 377-381.

SEWER, A. PAUL, N. LANDGRAF, P. ARAVIN, A. PFEFFER, S. BROWNSTEIN, M. J. TUSCHL, T. VAN NIMWEGEN, E. & ZAVOLAN, M. (2005). Identification of clustered microRNAs using an ab initio prediction method. *Bmc Bioinformatics* **6**. SHEN, E. DIAO, X. H. WEI, C. WU, Z. H. ZHANG, L. & HU, B. (2010). MicroRNAs target gene and signaling pathway by bioinformatics analysis in the cardiac hypertrophy. *Biochemical and Biophysical Research Communications* **397**(3), 380-385.

SHENOUDA, S. K. & ALAHARI, S. K. (2009). MicroRNA function in cancer: oncogene or a tumor suppressor? *Cancer and Metastasis Reviews* **28**(3-4), 369-378.

SIEBERT, S. & BACKOFEN, R. (2005). MARNA: multiple alignment and consensus structure prediction of RNAs based on sequence structure comparisons. *Bioinformatics* **21**(16), 3352-3359.

WANG, X. W. ZHANG, J. GU, J. HE, T. ZHANG, X. G. LI, Y. D. & LI, F. (2005). MicroRNA identification based on sequence and structure alignment. *Bioinformatics* **21**(18), 3610-3614.

WIGGINS, J. F. RUFFINO, L. KELNAR, K. OMOTOLA, M. PATRAWALA, L. BROWN, D. & BADER, A. G. (2010). Development of a Lung Cancer Therapeutic Based on the Tumor Suppressor MicroRNA-34. *Cancer Research* **70**(14), 5923-5930.

XUE, C. H. LI, F. HE, T. LIU, G. P. LI, Y. D. & ZHANG, X. G. (2005). Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine. *Bmc Bioinformatics* **6**.

ZENG, Y. YI, R. & CULLEN, B. R. (2005). Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *Embo Journal* **24**(1), 138-148.

ZHANG, B. H. PAN, X. P. WANG, Q. L. COBB, G. P. & ANDERSON, T. A. (2006). Computational identification of microRNAs and their targets. *Computational Biology and*

Chemistry **30**(6), 395-407.
ZHAO, H. F. WANG, D. H. DU, W. T. GU, D. S. & YANG, R. C. (2010). MicroRNA and leukemia: Tiny molecule, great function. *Critical Reviews in Oncology Hematology* **74**(3), 149-155.

TABLE

Mir-Abela
Mir-Explore
ProMir

Average⁴

Others

DRE

UMM

HSA

GGA

Average²

Others

DRE

MMU

HSA

GGA

GGA

Species¹

71.00

65.00

82.00

74.00

78.00

78.00

88.00

86.00

95.00

89.00

92.00

91.00

53.00

Sensitivity

99.99

99.99

99.99

99.99

99.99

99.99

99.99

99.99

99.99

99.99

99.99

99.99

90.00

Specificity

Table 3.1 Sensitivity and specificity of ProMir, Mir-Explore and Mir-Abela in predicting pre-miRNA

GGA=chicken; HSA=human; MMU=Mouse; DRE=zebra fish; Others= fugu, worms, frog, chimpanzee, gorilla, platypus, pig,

drosophila and buffalo

²Average was calculated as the ratio between the total number of predicted pre-miRNA and the total number of pre-miRNA in the test

data.

FIGURES

Figure 3.1. An example of alignment output from MARNA, '(' and ')' indicates the stem

in the secondary structure, '.' is a mismatch and '—' is a gap.

(((((((((((((((())())((((())))	_))) .))))))) .)))))))
ACAUACUUCUUUAU-GUACCC-AU-AU-G-AAC-AUA-CAAU-GC	-UAUGGA-AUGUAAAGAAGUAUGU
ACAUACUUCUUUAU-AUGCCC-AU-AU-G-AAC-CUG-GCAA-UC	-UAUGGA-AUGUAAAGAAGUAUGU
ACAUACUUCUUCAU-AUGCCC-AU-AU-G-GAG-UCG-GCCGGCGU	-UAUGGA-AUGUUAAGAAGUAUGU
GAAUGGAAAGCAGU-AUGUAUUUUAUG-UAAA	AUAGAUUUGCAUUAUUAUUU
GCAUACUUCCUUAC-AUGCCC-AU-AC-U-AUA-UCA-UAAAUGGA	-UAUGGA-AUGUAAAGAAGUAUGU

SUPPLEMENTARY DATA

Supplementary 3.1 Sensitivity and specificity of miR-Explore in predicting pre-miRNA

across species

		Sensitivity							
					Zebra				
Class	Specificity	Chicken	Human	Mouse	Fish	Others	Average		
let-7	99.75%	2,2	2,2	2,2	2,2	38,39	97.87%		
mir1	100.00%	2,2	2,2	2,2	2,2	12,12	100.00%		
mir7	100.00%	1,2	1,2	2,2	1,2	13,16	75.00%		
mir9	100.00%	2,2	2,2	2,2	2,2	14,19	81.48%		
mir10	100.00%	2,2	2,2	2,2	2,2	10,14	81.82%		
mir15	100.00%	2,2	2,2	2,2	2,2	9,10	94.44%		
mir16	100.00%	2,2	2,2	2,2	2,2	8,8	100.00%		
mir17	100.00%	1,1	1,1	1,1	2,2	5,5	100.00%		
mir18	100.00%	1,1	1,1	1,1	2,2	7,7	100.00%		
mir19	99.96%	2,2	2,2	2,2	2,2	10,13	85.71%		
mir20	100.00%	2,2	2,2	2,2	2,2	6,6	100.00%		
mir21	100.00%	1,1	1,1	1,1	2,2	4,4	100.00%		
mir22	100.00%	1,1	1,1	1,1	2,2	4,5	90.00%		
mir23	100.00%	1,1	2,2	2,2	2,2	9,9	100.00%		
mir24	100.00%	1,1	2,2	1,2	2,2	6,6	92.31%		
mir26	100.00%	1,1	2,2	2,2	2,2	6,6	100.00%		
mir27	100.00%	1,1	2,2	2,2	2,2	9,9	100.00%		

mir29	99.98%	2,2	2,2	2,2	2,2	14,17	88.00%
mir30	99.99%	1,2	2,2	1,2	2,2	16,18	84.62%
mir31	99.99%	1,1	0,1	1,1	2,2	11,11	93.75%
mir32	100.00%	1,1	1,1	1,1	N/A	3,3	100.00%
mir33	100.00%	2,2	1,1	1,1	N/A	9,9	100.00%
mir34	100.00%	1,2	2,2	2,2	0,1	14,14	90.48%
mir92	99.81%	0,1	0,2	0,2	2,2	17,21	67.86%
mir99	100.00%	1,1	2,2	2,2	2,2	4,5	91.67%
mir100	100.00%	1,1	1,1	1,1	2,2	8,8	100.00%
mir101	100.00%	2,2	2,2	2,2	2,2	7,8	93.75%
mir103	100.00%	2,2	2,2	2,2	1,1	7,7	100.00%
mir106	100.00%	1,1	2,2	2,2	N/A	6,6	100.00%
mir107	100.00%	1,1	1,1	1,1	1,1	4,4	100.00%
mir122	100.00%	2,2	1,1	1,1	1,1	3,3	100.00%
mir124	99.99%	2,2	2,2	2,2	2,2	15,15	100.00%
mir125	99.97%	0,1	0,2	0,2	0,2	0,16	0.00%
mir126	100.00%	1,1	1,1	1,1	1,1	3,3	100.00%
mir128	100.00%	2,2	2,2	2,2	2,2	6,6	100.00%
mir130	100.00%	2,2	2,2	2,2	2,2	7,7	100.00%
mir133	100.00%	2,2	2,2	2,2	2,2	11,14	86.36%
mir135	100.00%	2,2	2,2	2,2	2,2	9,10	94.44%
mir137	100.00%	1,1	1,1	1,1	2,2	5,6	90.91%
mir138	100.00%	1,2	2,2	1,2	1,1	1,4	54.55%

mir140	100.00%	1,1	1,1	1,1	1,1	3,3	100.00%
mir142	100.00%	1,1	1,1	1,1	1,1	4,4	100.00%
mir144	100.00%	1,1	1,1	1,1	1,1	3,3	100.00%
mir146	100.00%	2,2	2,2	2,2	2,2	5,5	100.00%
mir147	100.00%	2,2	2,2	1,1	N/A	3,3	100.00%
mir148	100.00%	1,1	2,2	2,2	1,1	4,4	100.00%
mir153	100.00%	1,1	2,2	1,1	2,2	6,6	100.00%
mir155	100.00%	0,1	0,1	0,1	0,1	0,2	0.00%
mir181	100.00%	2,2	2,2	2,2	2,2	15,18	88.46%
mir183	100.00%	1,1	1,1	1,1	1,1	2,5	66.67%
mir184	100.00%	1,1	1,1	1,1	1,1	9,9	100.00%
mir187	100.00%	1,1	1,1	1,1	1,1	3,3	100.00%
mir190	99.99%	1,1	1,1	1,1	1,1	7,8	91.67%
mir193	100.00%	2,2	1,1	1,1	2,2	3,6	75.00%
mir194	100.00%	1,1	2,2	2,2	2,2	5,5	100.00%
mir196	100.00%	2,2	2,2	2,2	2,2	9,9	100.00%
mir199	100.00%	2,2	2,2	2,2	2,2	8,9	94.12%
mir200	100.00%	2,2	2,2	2,2	2,2	9,9	100.00%
mir202	100.00%	1,1	1,1	1,1	1,1	3,3	100.00%
mir203	100.00%	1,1	1,1	1,1	1,1	3,3	100.00%
mir204	100.00%	2,2	1,1	1,1	2,2	5,5	100.00%
mir205	100.00%	2,2	1,1	1,1	1,1	4,4	100.00%
mir206	100.00%	1,1	1,1	1,1	1,1	3,3	100.00%

mir211	100.00%	1,1	1,1	1,1	N/A	2,2	100.00%
mir214	100.00%	1,1	1,1	1,1	1,1	4,4	100.00%
mir215	100.00%	1,1	1,1	1,1	N/A	3,3	100.00%
mir216	99.99%	2,2	1,1	1,1	2,2	4,8	85.71%
mir217	100.00%	1,1	1,1	1,1	1,1	4,4	100.00%
mir218	100.00%	2,2	2,2	2,2	2,2	8,8	100.00%
mir219	100.00%	1,1	1,1	1,1	1,1	8,10	85.71%
mir221	100.00%	1,1	1,1	1,1	1,1	4,4	100.00%
mir222	100.00%	1,1	1,1	1,1	1,1	3,3	100.00%
mir223	100.00%	1,1	1,1	1,1	1,1	3,4	87.50%
mir301	100.00%	1,1	1,1	1,1	1,1	5,5	100.00%
mir302	100.00%	0,2	0,2	0,2	N/A	0,8	0.00%
mir365	100.00%	2,2	2,2	2,2	2,2	3,5	84.62%
mir367	100.00%	1,1	1,1	1,1	N/A	2,2	100.00%
mir375	100.00%	1,1	1,1	1,1	1,1	5,5	100.00%
mir383	100.00%	1,1	1,1	1,1	N/A	3,3	100.00%
mir429	100.00%	1,1	1,1	1,1	1,1	3,5	77.78%
mir449	100.00%	0,1	0,1	0,1	N/A	0,4	0.00%
mir451	100.00%	1,1	1,1	1,1	1,1	2,2	100.00%
mir454	100.00%	1,1	N/A	N/A	1,1	2,2	100.00%
mir455	100.00%	1,1	1,1	1,1	1,1	2,2	100.00%
mir460	100.00%	N/A	N/A	N/A	N/A	1,1	100.00%
mir466	100.00%	0,1	N/A	0,2	N/A	1,4	14.29%

mir489	100.00%	1,1	1,1	1,1	1,1	2,2	100.00%
mir490	100.00%	N/A	N/A	N/A	N/A	2,2	100.00%
mir499	100.00%	1,1	1,1	N/A	N/A	3,3	100.00%
mir551	100.00%	1,1	1,1	1,1	N/A	3,3	100.00%
mir1306	100.00%	N/A	N/A	N/A	N/A	1,1	100.00%

	SENSITIVITY MIR-ABELA								
				Zebra					
Class	Chicken	Human	Mouse	Fish	Others	Average			
let-7	2,2	0,2	1,2	2,2	26,39	65.96%			
mir1	2,2	2,2	1,2	2,2	10,12	85.00%			
mir7	2,2	2,2	2,2	2,2	14,16	91.67%			
mir9	2,2	2,2	2,2	2,2	16,19	88.89%			
mir10	2,2	2,2	1,2	2,2	11,14	81.82%			
mir15	1,2	1,2	0,2	2,2	3,9	41.18%			
mir16	2,2	2,2	2,2	1,2	7,8	87.50%			
mir17	0,1	0,1	0,1	0,2	1,5	10.00%			
mir18	2,2	0,1	1,1	2,2	3,7	61.54%			
mir19	2,2	2,2	2,2	2,2	8,13	76.19%			
mir20	1,2	1,2	1,2	1,2	3,6	50.00%			
mir21	1,1	1,1	1,1	2,2	3,4	88.89%			
mir22	1,1	1,1	1,1	2,2	3,5	80.00%			
mir23	1,1	2,2	2,2	2,2	5,9	75.00%			
mir24	0,1	1,2	1,2	2,2	1,6	38.46%			
mir26	1,1	2,2	2,2	2,2	3,6	100.00%			
mir27	0,1	1,2	1,2	1,2	4,9	43.75%			
mir29	2,2	2,2	2,2	2,2	13,17	84.00%			
mir30	1,2	2,2	2,2	2,2	11,18	69.23%			

Supplementary 3.2 Sensitivity of miR-abela in predicting pre-miRNA across species

mir31	1,1	1,1	1,1	0,2	8,11	68.75%
mir32	1,1	1,1	1,1	N/A	2,3	83.33%
mir33	2,2	1,1	1,1	N/A	8,9	92.31%
mir34	1,2	1,2	1,2	1,1	14,14	85.71%
mir92	1,1	2,2	2,2	2,2	17,21	92.86%
mir99	1,1	1,2	1,2	1,2	0,5	33.33%
mir100	0,1	0,1	0,1	2,2	5,8	53.85%
mir101	2,2	2,2	2,2	2,2	5,8	81.25%
mir103	2,2	1,2	2,2	1,1	1,7	50.00%
mir106	0,1	1,2	0,2	N/A	4,6	45.45%
mir107	0,1	0,1	1,1	1,1	0,4	25.00%
mir122	2,2	1,1	1,1	1,1	2,3	87.50%
mir124	2,2	2,2	1,2	1,2	11,15	73.91%
mir125	1,1	2,2	1,2	2,2	6,16	52.17%
mir126	1,1	1,1	1,1	1,1	2,3	85.71%
mir128	2,2	2,2	2,2	2,2	3,6	78.57%
mir130	1,2	2,2	1,2	2,2	5,7	73.33%
mir133	2,2	2,2	2,2	2,2	9,14	77.27%
mir135	2,2	2,2	2,2	0,2	9,10	83.33%
mir137	1,1	1,1	1,1	2,2	5,6	90.91%
mir138	0,2	1,2	0,2	0,1	1,4	18.18%
mir140	1,1	1,1	1,1	1,1	3,3	100.00%
mir142	1,1	1,1	0,1	1,1	3,4	75.00%

mir144	1,1	1,1	1,1	1,1	2,3	85.71%
mir146	2,2	1,2	2,2	2,2	4,5	84.62%
mir147	0,2	1,2	0,1	N/A	2,3	37.50%
mir148	0,1	1,2	2,2	0,1	1,4	40.00%
mir153	1,1	2,2	1,1	2,2	5,6	91.67%
mir155	0,1	0,1	1,1	1,1	1,2	50.00%
mir181	1,2	1,2	1,2	1,2	5,18	34.62%
mir183	0,1	0,1	0,1	0,1	3,5	33.33%
mir184	1,1	1,1	1,1	1,1	7,9	84.62%
mir187	1,1	0,1	0,1	0,1	0,3	14.29%
mir190	1,1	1,1	1,1	1,1	6,7	90.91%
mir193	1,2	1,1	0,1	2,2	4,6	66.67%
mir194	1,1	1,2	1,2	2,2	4,5	75.00%
mir196	2,2	2,2	2,2	2,2	6,9	100.00%
mir199	2,2	2,2	1,2	2,2	3,9	58.82%
mir200	2,2	2,2	2,2	2,2	7,9	88.24%
mir202	1,1	1,1	1,1	1,1	3,3	100.00%
mir203	1,1	1,1	1,1	1,1	3,3	100.00%
mir204	2,2	1,1	1,1	1,2	4,5	81.82%
mir205	2,2	1,1	1,1	1,1	3,4	88.89%
mir206	1,1	1,1	1,1	1,1	2,3	85.71%
mir211	1,1	1,1	1,1	N/A	2,2	100.00%
mir214	1,1	1,1	1,1	1,1	4,4	100.00%

mir215	1,1	1,1	1,1	N/A	3,3	100.00%
mir216	0,2	1,1	0,1	0,2	5,8	42.86%
mir217	1,1	1,1	1,1	1,1	3,4	87.50%
mir218	2,2	2,2	2,2	2,2	5,8	81.25%
mir219	1,1	1,1	1,1	1,1	5,10	64.29%
mir221	1,1	1,1	1,1	1,1	3,4	87.50%
mir222	1,1	1,1	1,1	1,1	3,3	100.00%
mir223	1,1	1,1	1,1	1,1	4,4	100.00%
mir301	1,1	0,1	0,1	0,1	0,5	11.11%
mir302	1,2	2,2	1,2	N/A	8,8	85.71%
mir365	1,2	1,2	2,2	1,2	3,5	61.54%
mir367	0,1	1,1	1,1	N/A	0,2	40.00%
mir375	1,1	0,1	0,1	1,1	3,5	55.56%
mir383	0,1	0,1	0,1	N/A	0,3	0.00%
mir429	1,1	0,1	1,1	1,1	4,5	77.78%
mir449	1,1	1,1	1,1	N/A	2,4	71.43%
mir451	1,1	1,1	1,1	1,1	2,2	100.00%
mir454	1,1	N/A	N/A	1,1	1,2	75.00%
mir455	1,1	1,1	1,1	1,1	2,2	100.00%
mir460	N/A	N/A	N/A	N/A	1,1	100.00%
mir466	0,1	N/A	1,2	N/A	4,4	71.43%
mir489	1,1	1,1	1,1	1,1	2,2	100.00%
mir490	N/A	N/A	N/A	N/A	0,2	0.00%

mir499	1,1	1,1	N/A	N/A	2,3	80.00%
mir551	1,1	1,1	1,1	N/A	3,3	100.00%
mir1306	N/A	N/A	N/A	N/A	0,1	0.00%

CHAPTER 4

POTENTIAL MIR-1a and MIR-124 TARGETS IN THE 3' UNTRANSLATED

REGIONS (3'UTRs) OF PLATYPUS (Ornithorhynchus anatinus)¹

¹Sebastian, B., S.E. Aggrey. To be submitted to Journal of Evolutionary Biology

ABSTRACT

MiRNAs are small non-coding RNAs which regulate gene expressions by targeting the mRNAs. The identification of miRNA targets has been done by biological experiment and computational prediction in both plants and mammals. The prediction of miRNA targets in animals has been a great challenge. MiRNAs of platypus (*Ornithorhynchus anatinus*) have been identified, verified, and stored in the miRBase (Griffiths-Jones 2004) but none of the miRNA targets are annotated. In this study we used common features of observed miRNA and target interaction to predict the potential target of miR-1a and miR-124 in the 3' UTR regions of Platypus (*Ornithorhynchus anatinus*). This approach predicted 734 novel potential target of miR-1a in Platypus 3'UTR regions where 98 of them are in the well annotated chromosomal region and 124 novel potential target of miR-124 where 32 of them are in the well annotated chromosomal region.

INTRODUCTION

MicroRNAs (miRNAs) are single stranded 22 nucleotides long small RNAs. They target mRNAs to repress translation or mediate mRNA degradation (Bartel 2004). In animals, miRNA is always related with the regulation of some important processes such as apoptosis, metabolism, and development (Kloosterman & Plasterk 2006). Hundreds of miRNAs has been identified from various organisms (Griffiths-Jones et al. 2008). Lee et al (1993) discovered a 22 nucleotides long small RNA *lin-4* which is observed to have antisense complementarity to the *lin-14* gene 3'UTR (Lee et al. 1993). This short RNA *lin-4* is now known to be the pioneer of the discovery of miRNAs (Lagos-Quintana et al. 2001). miRNAs are usually transcribed by RNA polymerase II as a primary miRNA which has cap and poly adenylated tail. The primary miRNA transcript is further cleaved by Drosha and Pasha RNAse III enzyme complex to become miRNA precursor (premiRNA). Pre-miRNA is then transported out to the cytoplasm by exportin-5 enzyme and further process by Dicer to generate mature miRNA, which will then be associated with RNA-induced silencing complex (RISC) to target mRNA (Bartel 2004).

It has been predicted that more than one third of human genes are conserved miRNA targets, but not many miRNA functions have been fully identified because predicting and confirming miRNA targets are difficult (Xia et al. 2009). Hence the study of miRNA target prediction and verification play significant roles on more understanding of miRNA functions. The miRNA prediction in animals is difficult because of the lack of known miRNA target genes and precise target sites (Xia et al. 2009). This makes it difficult for the bioinformatics approaches to have a good training data to achieve high sensitivity and specificity (Enright et al. 2004). Despite of these difficulties, there are

some features of miRNA and target interaction those are commonly used as computational prediction program: (1) the complementarity between the miRNA and the target; (2) the conservation of miRNA target across organisms; (3) the thermal stability of the miRNA and target duplex; (4) there is no complicated secondary structure at the target sites; and (5) the capability of the 5' end of the miRNA to bind stronger to the target compared to the 3' end (Xia et al. 2009).

MiRanda (Enright et al. 2004) was the pioneer of the computational prediction of miRNA target. The algorithm of miRanda has three phases which are sequence matching to determine whether two sequences are complementary and possibly bind; calculation of free energy of the binding; and conservation of the sequences across organisms. For the sequence matching, they used the "Smith-Waterman" alignment algorithm (Smith & Waterman 1981), but instead of pairing the same nucleotides together, the scores were based on complementarity of AU, GC and wobbly GU pairs. Besides of the alignment, they also created rules based on their observation of known miRNA and target mRNA interaction. First, no mismatches are allowed at the positions 2 to 4; second; second, less than five mismatches between positions 3 to 12; third, at least one mismatch between position 9 to L - 5 (where L is length of the alignment); fourth, less than two mismatches in the last five positions of the alignment. Positions were counted from the 5' end of the miRNA.

In Chapter 3 (this dissertation) it is shown that grouping miRNAs according to their classes increases sensitivity and specificity of pre-miRNA prediction. It is therefore hypothesized that each miRNA class has its own characteristics of miRNA and target interaction. The objective of this study was to use the observed common features in the

miRNA and target interactions to predict the targets of miRNA miR-1a and miR-124 and further test the efficacy of using unique rules for predicting miR-1a and miR-124 targets to predict mir-15a and miR-16 in platypus (*Ornithorhynchus anatinus*). Since miRNA targets have been annotated in platypus, the results of this study would also assist in the future annotation of miRNA targets in the platypus genome.

MATERIALS AND METHODS

The Platypus 3' UTRs

Four thousand, two hundred and nine platypus 3' UTRs region sequences were obtained from the UCSC genome browser database (Rhead et al. 2010) and used as input sequences for obtaining potential miRNA targets.

Positive and negative data

The positive data was obtained from the verified human miR-1 and miR-124 targets available from the miRecords database (Xiao et al. 2009). There were 102 verified human miR-1 targets and 181 verified miR-124 targets that were used as positive tests data. The negative data was 10,000 randomly generated sequences of length 150 nucleotides.

Positive data for comparative study

Sixty-five human miR-15a verified targets and 75 human miR-16 verified targets from the same database were also used as positive data to test whether rules developed for predicting miR1a and miR-124 targets could be used to predict mir-15a and mir-16. These miRNA were chosen because they have only 5 nucleotides in their seed compared to miR-1a and miR-124 which have 7 nucleotides each in their respective seeds.

Programming

A program TarSpec was created to undertake alignment based on the observed characteristics of miRNA and target interactions. Rules were developed for the miRNA target predictions in order to obtain minimum sensitivity of 75% and specificity and at least 98%. The input for this program was mature miRNA sequences and the query sequences, both in the FASTA format. The program utilized a sliding window with the length of the mature miRNA sequence. Direct alignment position by position inside the range of the window length with the potential target sequences was done. The number of matches and mismatches from each position was recorded. Sequences that met the rules were selected and considered as hits or counted as detected.

Rules:

All position is calculated from the 5' end of the miRNA. For miR-1a targets, positions 1-8 should have least 7 matches where position 2-5 needs to be a perfect match, and there should be at least 6 AU pairs and 3 GC pairs in the entire alignment. For miR-124 targets, positions 1-8 should have at least 7 matches where positions 2-5 need to be a perfect match, and there should be at least 4 AU pairs and 4 GC pairs in the entire alignment. The rules of TarSpec are similar to TargetScan (Lewis et al. 2003), but there are also differences. TargetScan allows a perfect match in position 2-8, while TarSpec allows 1 mismatch in any position between positions 1-8, except position 2-5. The AU rich region in TargetScan is outside of the miRNA and target binding sites, while with TarSpec the AU rich or GC rich region is inside the binding sites of the miRNA and target.

Testing:

The miR-1a and miR-124 known targets were screened with their respective rules to obtain their sensitivity. The rules for miR-1a and miR-124 were used to screen the known miR-15a and miR-16 targets. The 10,000 negative data was scanned to obtain the specificity of the program. All positive and negative data were also screened using the MiRanda and miTarget programs, except that in the case of miTarget, specificity was not tested because the program requires manual input.

Sensitivity and specificity were calculated as follow:

Sensitivity = Number of detected targets in the positive data/Total number of verified targets in positive data.

Specificity = 1 - (Number of detected targets in negative data/Total number of random sequences in the negative data).

Minimum binding energy

The minimum binding energy of miRNA and its targets was obtained to in order to test whether it could be used to distinguish between the real miRNA target and false positive for the miR-1 and miR-124 targets. The minimum binding energy of the entire length of the miRNA and target interactions, as well as the binding energy for only position 9 to the end of the miRNA. The Vienna package (RNAduplex) was used to calculate the minimum binding energy (Mueckstein et al. 2006).

RESULTS AND DISCUSSION

Mir-1 is one of the very important microRNAs involve in human myocardial disease (Cheng et al. 2010). It targets the cytoskeleton regulatory protein twinfilin-1 to cause cardiac hypertrophy (Li et al. 2010b). Meanwhile, miR-124 is also involves in

pancreatic disease and neural stem cell study (Rutter & Pullen 2010; Wang et al. 2010a). Since the functional roles of both miR-1a and miR-124 are well established, their evolutionary history cannot be overstated. Platypus is among the most interesting organism for evolutionary study. This animal is categorized as mammals, but has some features of avian and reptiles. Platypus undergoes lactation, yet lay eggs, and has snake similar poison in their spurs. They have 52 chromosomes most of which are small chromosomes. They have multiple chromosomes with some homology with bird's Z chromosome. The males have 5X and 5Y but the exact number of the chromosome dosage in males or females is still under investigation.

The miRNA in platypus has been annotated and stored in the miRNA database, but there is no annotation for RNA targets. In this study we used the observed characteristics of miRNA and target interactions to predict a possible novel Platypus miR-1a and miR-124 target candidates in the 3'UTR region. There are two annotated miR-1 in Platypus namely miR-1a and miR-1b. We chose miR-1a because the mature miRNA sequence is 100% identical to the human miR-1. The sensitivity and specificity of predicting miR-1a and miR-124 targets are shown in Table 4.1. The sensitivity of TarSpec for miR-1a target was 78%, and that of miR-124 was 77%. The specificity for both miR-1a and miR-124 was 98%. The sensitivity and specificity of miRanda for predicting miR-1a were 70% and 98%, respectively. MiRanda also predicted miR-124 with 79% sensivity and 99% specificity. The TarSpec rules predicted miR-1a with a better sensitivity than MiRanda, and specificity was comparable. The sensitivity and specificity of TarSpec was comparable to MiRanda in predicting miR-124. On the other hand, miTarget had 8% sensitivity in predicting miR-1a and 41% in predicting miR-124.

The differences in sensitivity of the different algorithms in predicting the miR targets depend on how best the program captures the interactions between the particular miRNA and the respective targets. Detailed biological information on such interactions is limiting and as a result, algorithms based on generalized biological information may have limited success in predicting miRNA targets.

The distributions of the minimum binding energy of the entire length of miRNA and target interactions for the true and false positive data, and that from positions 9 to end of the miRNA length are shown in Figures 4.1 and 4.2, respectively. These figures clearly demonstrate that minimum binding energy cannot be used to distinguish between true andfalse positive data. However, on most studies where minimum free energy had been used in prediction of miRNA targets, it has been used to distinguish between true positives and true negatives (Enright et al. 2004)

When the rules developed for miR-1a and miR-124 were used to predict miRNA targets of miR-15a and miR-16, they both had a sensitivity of 28% and specificity of 98% (Table 4.2). These results strongly suggest that each miRNA class has its own unique miRNA and target interactions, and as a result different rules may be required for the prediction of each miRNA class targets. Therefore, developing a global program for predicting miRNA targets of all classes will invariable be limiting in terms of sensitivity. To improve sensitivity, miRNA target program rules need to be tailored to suit the particular miRNA class. This claim is supported by the sensitivity obtained by predicting miR-15a and miR-16 using MiRanda and miTarget. The sensitivity of MiRanda in predicting targets of miR-15a and miR-16 were 23% and 0%, respectively (Table 4.3). The sensitivity of miTarget for miR-15a and miR-16 were 14% and 16%, respectively

(Table 4.4). Both MiRanda and miTarget were developed to predict targets of all classes of miRNA, and it is obvious that such generalized programs may have poor sensitivity in predicting miRNA targets of some classes despite having high specificity of 98%. The specificity of miTarget was not obtained because it requires inputting the negative data individually which is laborious.

Based on the sensitivity and specificity for predicting miR-1a and miR-124 based on the rules developed, platypus 3'UTRs from Platypus obtained from the UCSC genome database (Rhead et al. 2010) were scanned. From 4,209 UTRs sequences, 734 potential novel miR-1a targets were predicted (Table 4.5) of which 98 are annotated. One hundred and twenty-four miR-124 targets were also predicted (Table 4.6) of which 32 are annotated.

It is expected that the predicted targets would contain some false positives because the rules used to scan did not yield 100% sensitivity and 100% specificity. Therefore, biological experiments are needed to confirm these predicted targets, which will pave the way for their functions or evolutionary roles to be studied. Biological verifications would also provide more information needed to build efficient algorithms for target predictions.

REFERENCES

BAEK, D. VILLEN, J. SHIN, C. CAMARGO, F. D. GYGI, S. P. & BARTEL, D. P. (2008). The impact of microRNAs on protein output. *Nature* **455**(7209), 64-U38.

BARTEL, D. P. (2004). MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **116**(2), 281-297.

BEITZINGER, M. PETERS, L. ZHU, J. Y. KREMMER, E. & MEISTER, G. (2007). Identification of human microRNA targets from isolated argonaute protein complexes. *Rna Biology* **4**(2), 76-84.

BETEL, D. WILSON, M. GABOW, A. MARKS, D. S. & SANDER, C. (2008). The

microRNA.org resource: targets and expression. Nucleic Acids Research 36, D149-D153.

CHENG, Y. H. TAN, N. YANG, J. A. LIU, X. J. CAO, X. P. HE, P. C. DONG, X. L. QIN, S. S.

& ZHANG, C. X. (2010). A translational study of circulating cell-free microRNA-1 in acute myocardial infarction. *Clinical Science* **119**(1-2), 87-95.

ENRIGHT, A. J. JOHN, B. GAUL, U. TUSCHL, T. SANDER, C. & MARKS, D. S. (2004). MicroRNA targets in Drosophila. *Genome Biology* **5**(1).

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

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

GRIFFITHS-JONES, S. SAINI, H. K. VAN DONGEN, S. & ENRIGHT, A. J. (2008). miRBase: tools for microRNA genomics. *Nucleic Acids Research* **36**, D154-D158.

KLOOSTERMAN, W. P. & PLASTERK, R. H. A. (2006). The diverse functions of MicroRNAs in animal development and disease. *Developmental Cell* **11**(4), 441-450. LAGOS-QUINTANA, M. RAUHUT, R. LENDECKEL, W. & TUSCHL, T. (2001). Identification of novel genes coding for small expressed RNAs. *Science* **294**(5543), 853-858. LEE, R. C. FEINBAUM, R. L. & AMBROS, V. (1993). THE C-ELEGANS HETEROCHRONIC GENE LIN-4 ENCODES SMALL RNAS WITH ANTISENSE COMPLEMENTARITY TO LIN-14. *Cell* **75**(5), 843-854.

LI, Q. SONG, X. W. ZOU, J. WANG, G. K. KREMNEVA, E. LI, X. Q. ZHU, N. SUN, T.

LAPPALAINEN, P. YUAN, W. J. QIN, Y. W. & JING, Q. (2010). Attenuation of microRNA-1 derepresses the cytoskeleton regulatory protein twinfilin-1 to provoke cardiac hypertrophy. *Journal of Cell Science* **123**(14), 2444-2452.

LIM, L. P. LAU, N. C. GARRETT-ENGELE, P. GRIMSON, A. SCHELTER, J. M. CASTLE, J.

BARTEL, D. P. LINSLEY, P. S. & JOHNSON, J. M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**(7027), 769-773.

NAKAMURA, H. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Scientist* **19**(6), 25-25.

RHEAD, B. KAROLCHIK, D. KUHN, R. M. HINRICHS, A. S. ZWEIG, A. S. FUJITA, P. A.

DIEKHANS, M. SMITH, K. E. ROSENBLOOM, K. R. RANEY, B. J. POHL, A. PHEASANT, M.

MEYER, L. R. LEARNED, K. HSU, F. HILLMAN-JACKSON, J. HARTE, R. A. GIARDINE, B.

DRESZER, T. R. CLAWSON, H. BARBER, G. P. HAUSSLER, D. & KENT, W. J. (2010). The

UCSC Genome Browser database: update 2010. Nucleic Acids Research 38, D613-D619.

SELBACH, M. SCHWANHAUSSER, B. THIERFELDER, N. FANG, Z. KHANIN, R. & RAJEWSKY,
N. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature*455(7209), 58-63.

SMITH, T. F. & WATERMAN, M. S. (1981). IDENTIFICATION OF COMMON
MOLECULAR SUBSEQUENCES. *Journal of Molecular Biology* 147(1), 195-197.
XIA, W. CAO, G. J. & SHAO, N. S. (2009). Progress in miRNA target prediction and identification. *Science in China Series C-Life Sciences* 52(12), 1123-1130.

TABLES

	MiR-1a		Mir-124	
Programs	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
TarSpec	78	98	77	98
MiRanda	70	98	79	99
MiTarget	8	ND^1	41	ND

Table 4.1 Sensitivity and specificity for predicting Mir-1a and Mir-124 targets in humans

¹ND=Not determined

Table 4.2 Sensitivity and specificity of predicting miR-15a and miR-16 targets using rules for predicting miR-1a and miR-124 targets, MiRanda and MiTarget programs in humans.

MiR-15a		Mir-16	
Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
28	98	28	98
23	99	0	98
14	ND^1	16	ND
	MiR Sensitivity (%) 28 23 14	MiR-15a Sensitivity (%) Specificity (%) 28 98 23 99 14 ND ¹	MiR-15a Min Sensitivity (%) Specificity (%) Sensitivity (%) 28 98 28 23 99 0 14 ND ¹ 16

¹ND=Not determined

Table 4.3 Predicted targets of miR-1a in the Platypus 3'UTR in completely annotated chromosomal region

Chromosome	Position		
1	7446824-7447442	26147307-26147449	
	12448719-12450498	28552456-28553419	
	2600899-2601041	33132947-33133087	
	7249012-7249158	42443213-42443355	
	11318545-11318702	42716107-42716248	
	16553411-16553876	44481522-44481664	
	17170542-17170683	47081283-47081423	
	20656173-20657545		
2	2775358-2775963	27796816-27796910	
	18473044-18475201	30505539-30505679	
	22153924-22154192	32266513-32266600	
	36603078-36603434	49040885-49041030	
	11836510-11836657	51148779-51148919	
	24871766-24872058	51160372-51160512	
3	8031919-8035110	24178283-24178423	
	18104092-18104607	25719215-25719355	
	2174130-2174270	25864585-25865143	
	7060120-7060655	26210561-26210701	
	9225667-9225809	39885344-39886013	
	15695537-15695677	43305139-43305281	
	20778837-20781244		
4	1007515-1008801	20607251-20608257	
	9216131-9216272	20662358-20663763	
	11061498-11061638	21410906-21411149	
	12672495-12672636	33064871-33066861	
	18781932-18782073	49932945-49933090	
	20025607-20025692	53065205-53065534	
5	653843-655678	10684746-10684944	
	4066806-4066946	18115168-18115309	
	6073070-6073968		
6	10243485-10243608	7801510-7810754	
7	5347427-5347969	25755847-25755987	
	7684686-7686597	31204086-31207973	
	14582494-14582634	37616369-37616512	
10	6390067-6390971	4612835-4613242	
12	10177927-10179536	11397372-11397514	
	11393098-11393240	12721102-12721244	

15	1338462-1338699	
17	1103073-1103192	
18	4832534-4832979	
20	611318-611693	
X1	2367415-2367720	12130646-12130786
	796883-797025	18482893-18483038
	2109597-2109883	19009110-19009251
	4273355-4273498	22245135-22245277
	5767318-5767459	25797085-25798353
	6135366-6135506	28648749-28650974
	8933610-8933757	34964391-34964536
	9052334-9052475	40552967-40554371
	10114687-10114834	40798534-40798685
	10116865-10117012	41741568-41741912
X2	5411459-5412180	
X5	26098853-26098996	26567606-26567763

Table 4.6. Predicted Targets of miR-124 in the Platypus 3'UTR in completely annotated

Chromosome	Position	
1	13685345-13685630	28552456-28553419
2	166872-167869	1241644-1242016
3	8031919-8035110	22337334-22337621
	4446566-4447221	39038934-39040492
	8747501-8748417	42718269-42718443
	20778837-20781244	
4	10641800-10642713	48152968-48153736
	48152968-48153556	
5	653843-655678	16121690-16122025
6	3702199-3702968	3993594-3994997
	2022112-2022835	7812390-7813358
7	7684686-7686597	
10	9614890-9615523	9855094-9856435
11	6770159-6771403	
12	14205132-14206053	8503651-8503806
	14205132-14206053	
X1	15059668-15060330	28648749-28650974
	16192456-16193190	29292974-29293064
X2	5411459-5412180	

FIGURES

Figure 4.1 Distribution of the minimum binding energy value of the entire miRNA and target



Figure 4.2 Distribution of the minimum binding energy value of the tail (position 9 to the end of the miRNA length) of miRNA and target



CHAPTER 5

POTENTIAL TARGET OF MICRORNA MIR-28 IN HUMAN TRANSPOSABLE

ELEMENT L2¹

¹Sebastian, B., S.E. Aggrey. To be submitted to Journal of Insilico Biology

ABSTRACT

MicroRNAs (miRNAs) are small noncoding RNAs that has important roles on gene regulations. Some miRNAs are derivation of transposable elements (TE). In human these TE derived miRNAs has a potential to regulate thousands of human genes. In a sequence comparison study, it is proven that the TE derived miRNAs are less conserved compared to the non-TE derived miRNAs but there are some TE derived miRNAs which are conserved across species. The objective of this study was to investigate whether TE are also potential miRNA targets. We utilized an L2 derived miRNA miR-28 and used three different miRNA target prediction programs miTarget, miRanda, and RNAhybrid to predict a potential miRNA targets in human L2 transposable elements. We demonstrated that the human TE is also a potential target for miRNAs and subsequently predicted 1,094 of potential target in L2.

INTRODUCTION

Micro RNAs (miRNA) was first discovered in *Caenorhabditis elegans (C. elegans*) in 1993 by the analysis of a mutation in the developmental involved genes (Lee et al. 1993). The same regulatory system was then found seven years later in C. elegans let-7 genes(Reinhart et al. 2000). In the same year after the discovery of let-7 the homologs of this particular genes were found in the phylogenetically diverse animals (Pasquinelli et al. 2000). MiRNAs generally known as a short 22 nucleotides non coding RNA that regulates gene expression (Ambros 2004). From DNA, miRNA is transcribed into primary miRNA (pri-miRNA) by RNA polymerase II, processed by Drosha to yield a 70 to 90 nucleotides long hairpin structured miRNA precursor (pre-miRNA) and cleaved by Dicer to yield double stranded mature miRNA (Bartel 2004). Inside the nucleus, Drosha cleaved the primary miRNA near the base of the hairpin to yield the premiRNA. The pre-miRNA is then transported out of the nucleus to the cytoplasm to be cleaved by the Dicer to yield the mature double stranded miRNA. One of this duplex is rapidly degraded and another one is associated with RNA induced silencing complex (RISC) to act together targeting the mRNA for regulation of some genes by blocking translation or degrading the mRNA. MiRNA target specificity is determined by the partial complementary pairing between the miRNA and the target mRNA 3' UTRs (Bartel 2004).

Although miRNA characterization and discovery is continuing to develop, the evolutionary origin of miRNA has never been well studied (Piriyapongsa et al. 2007). MiRNA has been thought to be conserved among organisms, but yet non-conserve miRNAs has also been found (Bentwich et al. 2005). The contribution of the

Transposable Elements (TE) to the origin and evolution of human miRNAs has been evaluated (Piriyapongsa et al. 2007). Prior to this discovery, the distinction between miRNA and other small interfering RNAs (siRNAs) had been that siRNAs are transcribed from some TEs and miRNAs are transcribed from loci distinct from any TEs and other genes (Bartel 2004). However, this claim has been disputed by the discovery of miRNAs derived from TEs (Smalheiser & Torvik 2005; Borchert et al. 2006; Piriyapongsa et al. 2007). The contributions of TEs are often underestimated. They are often excluded from the computational search of novel miRNAs (Bentwich et al. 2005; Lindow & Krogh 2005; Nam et al. 2005; Li et al. 2006) because TEs are often treated as genomic parasites that do not play any functional roles in the host species (Doolittle & Sapienza 1980). However, more recent studies has proven that TEs can provide a functional roles to the host such as addition to a coding sequence (Volff 2006) and several regulatory sequences derived from TEs (Jordan et al. 2003; van de Lagemaat et al. 2003).

The tendency of excluding TEs in the novel miRNA search is extended to the miRNA target search. Thus far, the focus of miRNA target search has concentrated more on searching the 3' UTRs for a potential miRNA targets. Since miRNA can be derived from TEs it is very likely that TEs can also be potential miRNA targets. The objective of this study was to use available bioinformatics tools to predict a potential target of miRNAs in human TEs.
MATERIALS AND METHODS

Positive test data and human L2 Transposable Elements

The goal was to predict targets for MiR-28 which a TE derived miRNA, however, miR-28 has no biologically verified known targets in the database to be used as positive data, and as a result, data on miR-124 was used as positive data to test the efficacy of three miRNA target prediction tools. One hundred and eighty one biologically verified targets of miR-124 were downloaded from miRecords (Xiao et al. 2009). The negative data was a set of 1,000 randomly generated sequences of length 150 nucleotides. Fifty-six thousand, seven hundred and eighty-five human L2 TE were also downloaded from the UCSC genome browser (Rhead et al. 2010).

To begin the bioinformatics prediction, we tested a positive and negative data using three miRNA target prediction tools miTarget, miRanda, and RNAHybrid. Since mir-28 does not have adequate biologically verified known target, the positive data was taken from 181 known human mir-124 miRNA target which has been experimentally verified, downloaded from miRecords (Xiao et al. 2009).

miRNA Target Prediction Tools

MiRanda (Enright et al. 2004), miTarget (Kim et al. 2006) and RNA Hybrid (Rehmsmeier et al. 2004) were chosen to screen both positive and negative data in order to determine their individual and combined sensitivity and specificity. The programs were selected based on their ease of use (both online and downloadable), and have different distinct algorithms for prediction. MiRanda uses dynamic programming alignment, empirical rules based on their observation of known miRNA and target interactions, and minimum free energy to calculate a score that later determine if a sequence is a potential miRNA targets or not. MiTarget uses support vector machine (SVM) to distinguish miRNA targets and non miRNA target sequence by studying miRNA positive data and negative data set and using SVM to classify them as potential target or non potential target. RNAHybrid uses mainly minimum free energy to determine miRNA targets. This program is an extension of the RNA secondary structure prediction program with a restriction on disallowing intra-molecular binding (i.e., binding between miRNA and itself). RNAHybrid calculates the probability (p-value) for each prediction. If one miRNA is expected out of 1,000 query sequences, the program will generate 1,000 output sequences, each with a p-value. The p-value cut off can be set by the user based on the expected false positives. For example, if one miRNA and 1,000 query target sequences, with a p-value of 0.01 implies, the user is willing to accept 10 false positives.

Scanning of the Test Data and Human L2 Transposable Element

The 181 miR-124 positive data and the 1,000 negative data were screened using MiRanda and miTarget for initial sensitivity and specificity. The RNAHybrid program was used for final screening. We did not use RNAHybrid for the initial sensitivity and specificity determination but to further the prediction of the other two programs for the L2 TEs. The sequence tested for initial sensitivity and specificity was considered to be predicted when both miRanda and miTarget predicted the particular query sequence. The sensitivity and specificity is calculated as follow:

Sensitivity = (Predicted sequence from positive data / total number of known sequence tested) * 100%

Specificity = (1 - Predicted sequence from negative data / total number of negative data) * 100%

Hereafter, the human L2 TEs were also scanned for possible miR-28 targets. We selected L2 because it is the TE in human which has more miRNA derivative compared to other TEs. MiR-28 was selected because it was also derived from L2.

RESULTS AND DISCUSSION

The result of sensitivity and specificity test are shown in Figures 5.1 and 5.2, respectively. The initial sensitivity of miRanda and miTarget in predicting miR-124 targets are 79% and 41% respectively. The sensitivity declines with joint predictions from the two programs to 40%. Even though the joint sensitivity of the joint programs is low, it is expected that when used in ab-initio screening, the likelihood of false positives will be low. On the other hand, it is also expected that a reasonable amount of true positive will be missed.

The specificity test is conducted in the same manner as the sensitivity test. A negative sequence is considered to be predicted when it is predicted by both miRanda and miTarget as negative. The specificity for miRanda was 98.7%, and that for miTarget was 97.5%. The joint specificity of both programs was 99.2%. To enforce a better specificity, RNAHybrid was used to scan the predicted targets by miRanda and miTarget in the human L2 TEs. Based on the calculation of the expected false positive, we decided to select p-value of less than 0.05 for a sequence to be considered predicted.

After scanning all the TEs using the miRanda and miTarget jointly, we obtained 1,094 potential miR-28 targets (Table 1). We further scanned the predicted results in Table 1 using RNAHybrid with a p-value of less than 0.05. Based on the amount of

102

targets that is scanned (1094) we and p-value of 0.05, we expect 95% specificity. After the RNAHybrid scan, we further reduced the predicted miR-28 targets in human L2 TEs into 111 potential targets (Table 2). Even though miR-28 targets in TEs may be grossly under predicted using this approach, it is also amply demonstrates that TEs are potential miRNA targets that need to be explored. MiRNA target predictions. To date the focus of miRNA target predictions has been in the 3' UTRs but the full role of miRNAs can be studied when target identification is extended to include TEs. Even though these novel miRNA targets in TEs are predicted with high probability, there is the need not only to biologically verified them, but their functional roles studied as well.

REFERENCES

AMBROS, V. (2004). The functions of animal microRNAs. *Nature* **431**(7006), 350-355. BARTEL, D. P. (2004). MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **116**(2), 281-297.

BENTWICH, I. AVNIEL, A. KAROV, Y. AHARONOV, R. GILAD, S. BARAD, O. BARZILAI, A. EINAT, P. EINAV, U. MEIRI, E. SHARON, E. SPECTOR, Y. & BENTWICH, Z. (2005). Identification of hundreds of conserved and nonconserved human microRNAs. *Nature Genetics* **37**(7), 766-770.

BORCHERT, G. M. LANIER, W. & DAVIDSON, B. L. (2006). RNA polymerase III transcribes human microRNAs. *Nature Structural & Molecular Biology* **13**(12), 1097-1101.

DOOLITTLE, W. F. & SAPIENZA, C. (1980). Selfish Genes, The Phenotype, Paradigm and Genome Evolution. *Nature* **284**(5757), 601-603.

ENRIGHT, A. J. JOHN, B. GAUL, U. TUSCHL, T. SANDER, C. & MARKS, D. S. (2004). MicroRNA targets in Drosophila. *Genome Biology* **5**(1).

JORDAN, I. K. ROGOZIN, I. B. GLAZKO, G. V. & KOONIN, E. V. (2003). Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends in Genetics* **19**(2), 68-72.

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.
LEE, R. C. FEINBAUM, R. L. & AMBROS, V. (1993). The C-Elegans Heterochronic Gene
LIN-4 Encodes Small RNAs With Antisense Complementarity To LIN-14. *Cell* 75(5),
843-854.

LI, S. C. PAN, C. Y. & LIN, W. C. (2006). Bioinformatic discovery of microRNA precursors from human ESTs and introns. *Bmc Genomics* **7**.

LINDOW, M. & KROGH, A. (2005). Computational evidence for hundreds of nonconserved plant microRNAs. *Bmc Genomics* **6**.

NAM, J. W. SHIN, K. R. HAN, J. J. LEE, Y. KIM, V. N. & ZHANG, B. T. (2005). Human microRNA prediction through a probabilistic co-learning model of sequence and structure. *Nucleic Acids Research* **33**(11), 3570-3581.

PASQUINELLI, A. E. REINHART, B. J. SLACK, F. MARTINDALE, M. Q. KURODA, M. I.

MALLER, B. HAYWARD, D. C. BALL, E. E. DEGNAN, B. MULLER, P. SPRING, J.

SRINIVASAN, A. FISHMAN, M. FINNERTY, J. CORBO, J. LEVINE, M. LEAHY, P. DAVIDSON,

E. & RUVKUN, G. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408**(6808), 86-89.

PIRIYAPONGSA, J. MARINO-RAMIREZ, L. & JORDAN, I. K. (2007). Origin and evolution of human microRNAs from transposable elements. *Genetics* **176**(2), 1323-1337.

REHMSMEIER, M. STEFFEN, P. HOCHSMANN, M. & GIEGERICH, R. (2004). Fast and

effective prediction of microRNA/target duplexes. *Rna-a Publication of the Rna Society* **10**(10), 1507-1517.

REINHART, B. J. SLACK, F. J. BASSON, M. PASQUINELLI, A. E. BETTINGER, J. C. ROUGVIE, A. E. HORVITZ, H. R. & RUVKUN, G. (2000). The 21-nucleotide let-7 RNA regulates

developmental timing in Caenorhabditis elegans. Nature 403(6772), 901-906.

RHEAD, B. KAROLCHIK, D. KUHN, R. M. HINRICHS, A. S. ZWEIG, A. S. FUJITA, P. A.

DIEKHANS, M. SMITH, K. E. ROSENBLOOM, K. R. RANEY, B. J. POHL, A. PHEASANT, M.

MEYER, L. R. LEARNED, K. HSU, F. HILLMAN-JACKSON, J. HARTE, R. A. GIARDINE, B.

DRESZER, T. R. CLAWSON, H. BARBER, G. P. HAUSSLER, D. & KENT, W. J. (2010). The UCSC Genome Browser database: update 2010. *Nucleic Acids Research* **38**, D613-D619. SMALHEISER, N. R. & TORVIK, V. I. (2005). Mammalian microRNAs derived from genomic repeats. *Trends in Genetics* **21**(6), 322-326.

VAN DE LAGEMAAT, L. N. LANDRY, J. R. MAGER, D. L. & MEDSTRAND, P. (2003).

Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends in Genetics* **19**(10), 530-536.

VOLFF, J. N. (2006). Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *Bioessays* **28**(9), 913-922.

XIAO, F. F. ZUO, Z. X. CAI, G. S. KANG, S. L. GAO, X. L. & LI, T. B. (2009). miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Research* **37**, D105-D110.

TABLES

Table 5.1. Predicted miR-28 Targets in human L2 Transposable Elements using miRanda

and miTarget

	Location in the		
Chromosome		chromosome	
1	88144-88823	66157905-66158777	178841975-178842085
	1802396-1802599	69248485-69249265	185500105-185500509
	3280150-3280211	75174397-75174885	186709536-186710508
	4635351-4635635	76464805-76465090	186904135-186904546
	4783168-4783984	80235478-80235821	187210448-187211134
	5050132-5050586	81036617-81036859	187649674-187649874
	5298440-5298994	81368861-81369622	195299948-195300480
	6588362-6588879	85171259-85173150	197187793-197188219
	7196002-7197364	86477251-86477332	198825509-198826615
	8745887-8746039	93269772-93270003	200684478-200684564
	12573071-12573471	93373547-93374457	201883691-201884683
	15198409-15198544	94772655-94772918	203264387-203264631
	15657190-15657728	105955690-105956948	208026767-208026866
	16303110-16303421	108267769-108267878	210055112-210055859
	22121262-22121486	109156165-109156780	210461580-210461975
	23010821-23011036	110985557-110985871	213141256-213141700
	23379142-23379465	114060165-114060405	214118875-214119286
	28251496-28251710	117084130-117084619	215664068-215664362
	29683701-29684075	117691814-117692161	222043041-222043328
	33869286-33869815	118120669-118120937	223512108-223512820
	35903575-35903705	142711690-142711846	225144630-225146873
	37199973-37201091	142828300-142828456	225271143-225271409
	38086835-38087474	143380029-143380185	225814949-225815257
	38087638-38088042	143535135-143535779	226183712-226183936
	38166464-38167187	150002098-150002722	227631284-227631645
	39152758-39153452	154325975-154326719	229986953-229987117
	41999313-41999814	154693137-154693557	233599928-233600588
	44385076-44385310	160444350-160444937	235005051-235005580
	45264314-45264497	163308806-163309023	236598387-236598573
	45686693-45686805	163313233-163313415	236988936-236989362
	46294069-46294286	163946257-163947053	238071150-238071749
	47107050-47107780	164112716-164113414	239762314-239762881
	49180000-49180413	165522071-165522920	239950403-239950933
	50899780-50899922	172584388-172585318	244245720-244246200
	55168474-55169017	172625882-172626164	246544071-246544458

		17000(507 17000(057	24(5020(1)24(504257
	55535815-55536678	1/2926597-1/2926957	246593861-246594357
	56928816-56929125	1/30138//-1/3014//3	248001048-248001348
	6283391/-62834269	1/4990016-1/4991092	248001915-248002171
	64161566-64161802	1/628510/-1/628558/	249153687-249154549
2	1681974-1682112	63374235-63374593	158481556-158481837
	2719177-2719729	63891248-63891323	159043766-159044241
	2729124-2730362	67141149-67141332	161721381-161721810
	3456377-3456504	69416661-69417508	164878765-164879308
	3884852-3885103	70525692-70526143	176638502-176638646
	5119676-5120893	73384973-73385612	189447524-189447906
	5655292-5656248	73818970-73819199	191307041-191307651
	5676645-5677014	74375489-74375767	194662836-194662966
	7373233-7374315	74515629-74516275	196007309-196007526
	7376307-7376498	74770052-74770679	198994931-198995887
	8004326-8004695	75707926-75708058	200072881-200073033
	9318063-9318211	78506868-78507252	201178230-201179382
	9889481-9889895	80970137-80970589	202821378-202821480
	11465679-11467698	86647620-86648081	210144629-210145459
	12269710-12271157	95479567-95479715	218853746-218854270
	17941515-17941608	97578701-97578872	220257534-220257778
	19333541-19333724	104535480-104535639	223362742-223362829
	19841425-19841847	105733677-105734237	226275874-226276700
	23544777-23545067	106540947-106541386	227253031-227254178
	28350093-28350441	107186970-107187338	227724392-227724946
	33715219-33715896	108183179-108183641	229246541-229246743
	34180084-34181316	108646722-108647005	230397114-230397281
	36200951-36201577	110031050-110031251	230549067-230549342
	37044034-37044613	118358503-118359319	232102622-232103058
	42249236-42249584	121018683-121020128	232857230-232857947
	43244951-43245649	128739694-128740646	233912268-233912874
	44997738-44997806	129899997-129900321	237414305-237415102
	45144966-45146067	130513729-130514085	237766961-237767070
	45355919-45357533	131088486-131088867	238192927-238193714
	45814853-45815548	141604526-141605029	239143802-239144734
	47869813-47870214	143929854-143930578	240488129-240489049
	55249384-55250032	144461561-144462178	210100129 210109019
	61040499-61040733	149821382-149821841	
3	30015139-30015897	48840140-48840577	154193058-154193216
5	315503-315785	49838504-49839558	154363253-154363612
	3321434_3321671	54708394-54709505	157481085_157482607
	7413823-7413062	75908407-75000052	158070854_158071122
	7650071,7651502	75706407-75909052	158518107 158518551
	000071-7001000	70274574 70275670	160105612 160105002
	0052557-0052021	19214314-19213019	100103012-100103992

	10280764-10281110	79729550-79730902	161352089-161352286
	10852558-10852952	80602934-80603315	162171553-162171762
	14427000-14427262	104607081-104607396	165355984-165356264
	23424403-23424639	114726339-114727193	173087732-173087889
	26232231-26233552	116809324-116809794	179283776-179284162
	26529398-26529817	126879847-126880510	180036946-180037584
	36812026-36813023	129657449-129657639	181711269-181711475
	39258655-39259389	130036235-130037421	182142394-182143026
	41581246-41581400	132525281-132525437	193310098-193310389
	42029028-42029570	133351916-133352352	193524203-193524834
	42432678-42433359	142204585-142204954	196742610-196742801
	45110828-45111358	142790436-142790709	
	48676333-48676479	146859042-146859520	
	48712464-48712921	150452916-150454249	
4	3538159-3538515	56953601-56954109	154462709-154462807
	4326369-4326587	59466071-59466675	155087336-155087819
	6087679-6088239	63185625-63186045	156417359-156417711
	6295594-6295826	66515282-66515983	161661026-161661353
	6295594-6295826	68956770-68957406	163857530-163858370
	6648823-6649254	71504383-71504485	165101992-165102323
	17143789-17144439	76624166-76624577	168546320-168547393
	27468617-27468799	79050859-79051154	169540404-169540672
	32768595-32769057	86834171-86834766	171790581-171791320
	33081560-33081757	103507245-103507658	175420564-175421111
	33406611-33408077	103535165-103535947	176201410-176201709
	34108901-34109290	105140125-105141494	177738964-177739042
	37900888-37901070	105924143-105924220	185107603-185107866
	40473299-40473432	106468748-106469532	185450680-185450956
	42252051-42252236	114478216-114478641	185545896-185546438
	45486877-45488750	125116529-125118000	186148746-186149537
	49502942-49503098	135759486-135762005	186683410-186683793
	56614343-56615687	140710656-140710909	
	56785277-56785731	147698938-147699261	
5	4409953-4410383	54226356-54226537	102758193-102758454
	4658593-4658942	55950658-55950817	103374689-103375257
	5983148-5983484	56924235-56924687	103376332-103376696
	10994933-10996302	59409437-59409674	106441886-106442315
	12522322-12522912	60522402-60522650	110718705-110718992
	14711717-14712881	61284750-61284981	114359935-114359985
	17154947-17155939	64340586-64341335	114370664-114370898
	19532698-19534049	65281453-65282180	116412596-116413272
	19849262-19849573	71871936-71872136	117131619-117132556
	22549958-22550997	77620194-77620860	117207199-117207723

	24550000 24551225	70216474 70217000	110410957 110411510
	24330808-24331323	/82104/4-/821/089	11941085/-119411510
	29044245-29044457	79203337-79203708	119/204/4-119/20932
	22806005 22807224	01001020 01002100	120210002-120210009
	25490775 25400171	81981939-81982108	132332343-132332330
	35489775-35490171	838/4229-838/4300	144143228-144143818
	39543226-39543720	858/491/-858/5139	151809000-151810696
	40093388-40093636	91084481-91084696	152593335-15259345/
	4389//69-4389/99/	9/59/408-9/59/584	160319156-160319388
	44633954-44634802	985/59/0-985//295	1/3/2050/-1/3/2088/
	50544345-50544729	101815127-101815574	1///19855-1///210/6
6	52/25466-52/25889	102211//9-102212230	100(0(470 100(07117
6	1941430-1942030	82845999-82846564	129696472-129697117
	3104787-3104957	83122453-83123179	129884236-129885073
	4028867-4028921	84572281-84572507	13135/4/8-13135/542
	102/2432-102/328/	84892971-84894809	138064095-138064459
	11824078-11824262	8/5310/8-8/532448	14126/085-14126/491
	17871829-17872081	8/641/41-8/642035	142039869-142040419
	18800575-18801650	89607080-89607357	142503334-142504268
	21378162-21378365	93845261-93845817	142976636-142976971
	29515594-29516280	99027618-99027736	145327045-145327312
	30183070-30183231	101065384-101065930	146009229-146009934
	40958200-40958546	107062201-107062724	147215571-147216490
	46115526-46115644	109612949-109613698	150007580-150008225
	46724337-46724710	113979325-113980748	150951302-150952168
	48923035-48923944	114586347-114586799	152200027-152200642
	51974181-51974606	115070360-115070733	152344143-152344431
	52081608-52081905	116354755-116355251	152941266-152941890
	56275031-56276119	118574709-118575248	157235424-157236015
	62668429-62669084	121016316-121016775	159604471-159604874
	71095268-71095977	121540215-121540533	161429499-161429586
	71181451-71181901	123870736-123871842	161882226-161883131
	73039471-73040154	124110365-124110523	163551587-163551819
	76524745-76525248	124869772-124870228	168875036-168875331
	76546599-76546846	126264195-126264447	
	76546599-76546846	127413613-127414096	
7	739398-739643	81279443-81279697	124085319-124085607
	4394252-4394663	81773670-81773775	130614105-130614250
	12255816-12257248	84319791-84320219	130740412-130741069
	12680730-12682561	87695517-87696289	134497335-134499361
	14138668-14139140	88127710-88128233	135915469-135915612
	15069538-15069750	91442201-91442877	138658150-138658379
	15071160-15071564	93328805-93329077	141936428-141936658
	24579155-24579431	96087528-96089216	146846846-146847443

	24608614-24608807	96723764-96723824	149567740-149567882
	28331631-28331796	98787501-98787679	150043361-150044094
	36942433-36942669	102708300-102708875	150807483-150808014
	37304558-37305321	103133418-103134329	152812426-152812603
	43728702-43728754	106357847-106358084	157033117-157033838
	54813715-54814769	111461559-111461838	158725212-158725800
	55622279-55622637	117794956-117795927	159118995-159119209
	67146257-67146619	123309024-123309653	
	72276222-72276621	124014715-124015062	
8	6641602-6642213	61119283-61119637	113757118-113757287
	11374511-11374921	66646214-66646731	113956250-113957652
	15771411-15772350	67302377-67302946	115627356-115627601
	15828453-15828770	68115634-68116878	117063374-117063798
	16595797-16596625	68899778-68899930	120011846-120013485
	17441502-17442149	70364862-70365055	123141133-123141548
	23391963-23392128	73596106-73596776	123895247-123895497
	25415766-25416206	74619361-74619934	123897141-123897387
	29747081-29747560	79992210-79993058	127284128-127284853
	32747733-32748088	81613844-81615072	129380061-129380223
	34995967-34996133	86473905-86474000	130385543-130386441
	35353751-35354118	93506864-93507188	131085656-131085952
	37102384-37102761	94756276-94756403	138455275-138455501
	38735632-38735992	96235919-96236221	138673354-138673505
	39435524-39435896	98819787-98820029	139355934-139356261
	41646307-41646501	99231858-99232027	141056391-141056451
	42193327-42194146	100447916-100448406	141831086-141831319
	50889249-50889614	103737616-103738088	143215615-143215741
	53738496-53739453	105145734-105146281	144916867-144917273
	53887381-53887615	106837572-106838057	145062074-145062270
	55310131-55310852	107438110-107438268	
	56212183-56212574	109545446-109546213	
	60523725-60524133	111050326-111051303	
9	5374209-5374539	67688176-67688533	115436950-115437495
	11530031-11530531	67969974-67970122	118734082-118734195
	11779894-11780376	69425243-69425391	122640331-122640628
	19516076-19516122	72263583-72264345	122909867-122910040
	23607435-23607560	74537170-74537763	123432943-123433437
	24095180-24095469	77383063-77383713	127459323-127459429
	26758146-26758504	84380852-84381464	128696997-128697444
	30625422-30625827	86978051-86978270	128847832-128848475
	34432324-34432719	87545241-87545753	128974928-128975400
	34433389-34433585	88640139-88640292	129929074-129929492
	34721074-34721396	90864762-90865176	131291534-131291835

	35324532-35325002	91584042-91584708	134363954-134364568
	38938763-38939371	98306342-98306873	134365207-134365750
	40163392-40164000	98956696-98957143	134488819-134488951
	42411544-42411692	100553460-100553711	138069150-138069240
	43090143-43090291	100554158-100556128	140497526-140497995
	44075098-44075246	103856211-103856648	140806511-140807097
	45103669-45104025	106789116-106790091	
	46279203-46279561	110568079-110568247	
	47160875-47161233	111671915-111672027	
	65793077-65793685	113893720-113894448	
10	3316678-3316970	29345326-29345718	80066360-80066846
	4305652-4305983	29519374-29519707	81215550-81215916
	5324077-5324674	33627860-33629156	87336735-87338132
	5505322-5505957	34590238-34590948	88545561-88546571
	5911427-5911631	44178286-44178835	88706629-88707023
	6013976-6014190	44703678-44703904	94164726-94165093
	8382275-8383018	45935294-45935510	101888513-101888856
	11599635-11600116	49478647-49479246	106231999-106232832
	11722833-11723499	53479593-53479840	112383580-112383784
	16018957-16020296	55622668-55623886	119149659-119150259
	18398896-18399905	56551259-56551679	124126290-124126599
	21276994-21277842	59502133-59502957	126055253-126055676
	28185795-28186545	59688834-59689473	128013644-128014008
	28500697-28500896	69115620-69117064	130232379-130232866
	28873318-28873505	72455653-72456550	
	29344982-29345034	74000868-74001329	
11	21102589-21103368	57737271-57737461	91789089-91789821
	5531514-5532659	62698817-62699623	91818968-91819411
	7284719-7286222	64736859-64737129	95237409-95237659
	8665802-8666942	66608377-66608485	95424930-95425839
	12295399-12295711	67126635-67126736	95637325-95637538
	14057546-14058660	68673014-68673428	97453848-97454183
	15998306-16000000	70291217-70292039	101282247-101282696
	19541947-19542136	72388922-72389296	101522383-101522827
	24217964-24218414	72936857-72937322	121092262-121092362
	32511893-32512196	74458806-74459401	121093347-121094624
	34158823-34159950	76134036-76134192	122339527-122340044
	39311363-39311830	78123938-78124645	124389892-124390141
	42330887-42331036	82152452-82152736	126068128-126068377
	44823773-44824088	83356058-83356283	129543670-129544222
	46760292-46760500	86196540-86197365	
	47696603-47696751	86490520-86490924	
	49349686-49350139	89112750-89113201	

12	2233938-2234441	26496507-26496830	87805575-87806114
	3170667-3171596	27007403-27007800	88147268-88147453
	3621240-3621835	29334420-29334782	92219632-92220023
	4760835-4761843	33510243-33510474	92923136-92923953
	9788419-9788779	39156959-39157324	93910613-93910892
	12640519-12640866	39893243-39893673	94421882-94422168
	13339812-13340022	43431307-43432273	96237017-96237058
	13883789-13884113	49560301-49560699	99450163-99450395
	14582866-14583393	52327886-52328835	101620323-101620792
	15282567-15282883	52329148-52329374	110503346-110503739
	16118575-16119013	59920770-59921038	119132609-119132833
	16373087-16373439	65658769-65659115	119348347-119348860
	18742218-18742408	66550015-66550588	119853107-119853533
	23073524-23073654	76030037-76030418	132412193-132412429
	24535354-24535921	77627150-77627472	
	24660953-24661585	78563390-78564304	
13	19408268-19408415	57663787-57663988	93633374-93633506
	21582198-21582289	57968439-57969880	93948387-93948678
	22844360-22844968	60243391-60244121	94300325-94300438
	32513966-32514383	60357031-60357730	97722321-97723292
	34734442-34734721	66102371-66102996	100598692-100599838
	36734559-36734739	73178524-73178638	103579657-103580224
	42099077-42099595	74750036-74750928	104749352-104749806
	45569339-45569531	77162736-77163351	106418171-106418300
	46592140-46592803	77623063-77623335	109184600-109184859
	46976802-46976928	81364768-81365019	111664002-111664285
	50665991-50666526	82258153-82258446	111818434-111818526
	52476095-52476289	88926215-88926416	114960383-114960886
	55382095-55383032	90862111-90862368	
14	23017411-23017592	56441945-56442059	83972500-83972822
	23359240-23359834	58100631-58100810	85762827-85762925
	24912385-24913002	62381086-62381744	91707760-91708025
	26963225-26963592	63941412-63941800	92161947-92163855
	29957347-29957799	67237735-67238083	93192809-93192924
	30783196-30783804	69013043-69013467	94806807-94807662
	34309733-34310776	69020157-69020436	95902156-95902412
	36439377-36439482	70331139-70331327	98560047-98561640
	37235596-37235880	70882374-70882501	99226318-99227003
	45850285-45850662	77282750-77284703	101970273-101970974
	54842180-54842290	79046708-79046840	
	54954444-54954858	80886758-80887838	
15	85458739-85459180	57875115-57875459	91540588-91540968
	30971669-30971836	63923496-63923728	93227837-93228124

	31343701-31344413	64955240-64955907	94664555-94664794
	32018262-32018434	69063827-69064903	95168559-95169043
	32083896-32084760	70600617-70601287	97642655-97643494
	41637948-41638236	74223648-74223805	99248507-99248650
	43623863-43624423	75970995-75971680	99362165-99362446
	53611982-53612188	80799019-80799406	100815533-100815725
	53651154-53651422	81123437-81123690	101979214-101979352
	55709188-55709489	82538812-82540223	102036952-102038251
	56127905-56128311	84764273-84765646	
	57112346-57113386	87861198-87861600	
16	1623392-1623681	29829657-29829889	53809285-53809612
	1958722-1959084	30423306-30423489	66035374-66035574
	2253804-2253898	31279927-31280035	67586447-67586715
	2437094-2437431	34632188-34632732	68814687-68814945
	6483049-6483272	50456779-50456938	80917451-80918427
	20913447-20913843	50868213-50868587	84064011-84064315
	23757454-23757685	53237228-53237428	
17	4168609-4168715	20383315-20383396	50793345-50793559
	5925253-5925721	25405044-25405653	51556639-51557696
	6121390-6122082	25765852-25765933	53441596-53442202
	7354316-7354411	30452216-30452531	61302406-61303168
	7776568-7777919	31088271-31088836	63600704-63600829
	9030319-9030738	34174955-34175530	66990317-66991742
	10824671-10826361	40523023-40523369	67861798-67862480
	12512129-12512417	43775972-43776154	71575032-71575612
	15559678-15559999	46378383-46378655	79546698-79547022
	18367622-18367703	47509241-47509491	
	18369322-18369728	48864121-48864568	
18	1674955-1675194	29595234-29595489	56172482-56173389
	2704765-2705045	30944262-30945008	57605997-57606528
	9383744-9384021	31670393-31670891	67042114-67042546
	10521748-10522066	34504193-34504908	70241736-70243040
	11453550-11454116	38721162-38722946	74528113-74528461
	13126705-13126805	39188784-39189772	
	21231814-21232101	49112669-49113252	
	26211813-26212599	49289483-49289530	
19	8396109-8396169	37151350-37151511	58218552-58219002
	17844709-17844847	37552191-37552351	58950441-58950800
	28462551-28462794	38534476-38535129	59045045-59045158
	29787870-29788889	40245569-40246678	
	33009506-33010005	45658708-45658978	
	34528107-34528302	56402159-56402588	
	36717009-36717172	56726838-56726984	

	37017830-37018076	57052208-57053000	
20	3013429-3014259	32421053-32421514	49013619-49014034
	3094801-3095050	33133520-33133922	50795084-50795143
	4114875-4115566	33812558-33812628	51215782-51215913
	4662352-4662645	37318196-37318328	52214814-52215099
	5387145-5387802	38502067-38502729	52216238-52216595
	5511835-5512652	39085421-39085515	55944662-55945676
	10523815-10524448	39220844-39221338	
	16589946-16590862	39789139-39790140	
	18290405-18290635	43351415-43351854	
	30023503-30023641	46018594-46018678	
21	9770877-9771033	27463427-27464299	42320528-42320742
	18904622-18904917	34540418-34540816	44499587-44499747
	19017712-19018180	39351926-39352713	
	22483773-22484174	39900924-39901489	
22	17265745-17265898	31265824-31265945	44362877-44363711
	18772120-18773011	32962343-32963296	44905633-44905963
	21573367-21573998	34677963-34678483	46694568-46695679
	24144211-24144534	37346872-37348259	47345872-47346289
	25013119-25013944	39979955-39980687	48156516-48156807
	28015555-28015749	40784453-40785314	49677387-49678170
	28659778-28660040	43781182-43781383	
Х	4445948-4446458	52965013-52965248	128636099-128637656
	7854275-7855286	54644357-54644874	130136702-130136949
	8665367-8665639	57758284-57758971	131314354-131315327
	12049850-12051771	68401010-68401450	132679154-132679398
	13674026-13674271	71766715-71767467	133172653-133172853
	14019049-14019200	71928879-71929097	136148524-136148570
	16247812-16248438	92629663-92630198	141137088-141137281
	19350201-19350483	96299433-96300088	142540102-142540238
	20975047-20976375	97757695-97758308	146178128-146178466
	21192800-21193181	100924218-100924600	148406093-148406382
	22538710-22539106	104282834-104283126	150991959-150993149
	23099976-23100195	105282039-105282523	152875025-152875230
	24221195-24221445	106473217-106473630	154272491-154272591
	32445275-32445599	108287946-108288518	154767081-154767510
	40399461-40399876	108858149-108858447	
	40504004-40504647	109406774-109408199	
	43546921-43547510	111161266-111162280	
	51572124-51572372	114727041-114727831	
	52948969-52949204	115540418-115541229	
Y	13437944-13438588	17341474-17341858	

6_apd_hap1	819081-819769		
6_cox_hap2	1034399-1035087	1695171-1695332	
6_dbb_hap3	818940-819628		
6_mcf_hap5	818784-819470		
6_qbl_hap6	818887-819575		
6_mann_hap4	818571-819259		
6_ssto_hap7	856131-856817	1513482-1513643	
chr17_ctg5_hap1	1091123-1091345		
Un_gl000211	150760-150917		
Un_gl000212	93174-93805		
Un_gl000214	70001-70157		
Un_gl000221	16828-16984		
Un gl000246	12431-12584		

Table 5.2. Predicted miR-28 Targets in human L2 Transposable Elements usingmiRanda, miTarget and RNAHybrid with cutoff p-value 0.05

Chromosome	Position in the	e chromosome
1	3280150-3280211	187649674-187649874
	15198409-15198544	200684478-200684564
	44385076-44385310	213141256-213141700
	142828300-142828456	225144630-225146873
	143380029-143380185	229986953-229987117
2	3456377-3456504	104535480-104535639
	7376307-7376498	161721381-161721810
	11465679-11467698	200072881-200073033
	44997738-44997806	223362742-223362829
	63891248-63891323	237766961-237767070
3	7413823-7413962	48712464-48712921
	8832537-8832621	196742610-196742801
	48676333-48676479	
4	49502942-49503098	140710656-140710909
	56785277-56785731	147698938-147699261
	105924143-105924220	154462709-154462807
	135759486-135762005	
5	4409953-4410383	81981939-81982108
	4658593-4658942	114359935-114359985
	52725466-52725889	128216082-128216339
6	3104787-3104957	107062201-107062724
	30183070-30183231	124110365-124110523
	46115526-46115644	131357478-131357542
	84572281-84572507	163551587-163551819
	99027618-99027736	
7	43728702-43728754	134497335-134499361
	96723764-96723824	159118995-159119209
8	23391963-23392128	141056391-141056451
	70364862-70365055	143215615-143215741
9	19516076-19516122	122909867-122910040
	26758146-26758504	131291534-131291835
	118734082-118734195	
10	5505322-5505957	29519374-29519707
	29344982-29345034	53479593-53479840
11	57737271-57737461	97453848-97454183
	67126635-67126736	121092262-121092362

12	13339812-13340022	24660953-24661585
	23073524-23073654	96237017-96237058
13	21582198-21582289	82258153-82258446
	46976802-46976928	106418171-106418300
	73178524-73178638	
14	26963225-26963592	79046708-79046840
	36439377-36439482	85762827-85762925
	58100631-58100810	91707760-91708025
	70331139-70331327	93192809-93192924
	70882374-70882501	
15	30971669-30971836	63923496-63923728
	32018262-32018434	100815533-100815725
16	6483049-6483272	50456779-50456938
	31279927-31280035	
17	7354316-7354411	46378383-46378655
18	21231814-21232101	34504193-34504908
	31670393-31670891	49289483-49289530
19	8396109-8396169	56726838-56726984
20	30023503-30023641	50795084-50795143
21	9770877-9771033	44499587-44499747
22	17265745-17265898	48156516-48156807
	28015555-28015749	
X	24221195-24221445	152875025-152875230
	136148524-136148570	

FIGURES

Figure 5.1 Sensitivity of miRanda and miTarget in predicting human miR-28 targets



Figure 5.2 Specificity of miRanda and miTarget in predicting human miR-28 targets



CHAPTER 6

CONCLUSION

MiRNA is a small RNA that has very important role for gene regulations. In-vivo detection of miRNA has its limitations and computational prediction of miRNA is needed to assist biologist. The current study has provided a new approach by grouping the miRNA by classes to yield a sensitivity and specificity in predicting miRNA pre-cursor. Grouping miRNA by classes could also be a unique way of improving the prediction of miRNA targets. The relationships of miR-1a and miRNA-124 and their respective targets were studied and use to develop unique prediction algorithms (TarSpec) for these miR classes. These classes have noticeable different in the amount of AU and GC pairs. The sensitivity of TarSpec was reduced when used to predict mir-15a and miR-16. The sensitivity of TarSpec was better than existing programs like MiRanda and MirTarget in predicting Mir-1a, mir-15a and miR-16 targets. The sensitivity of TarSpec is comparable with MiRanda in predicting miRNA-24 targets. The TarSpec program was used to scan platypus 3' UTR region for possible miRNA targets. From 4,209 UTRs sequences, 734 potential novel miR-1a targets were predicted of which 98 are annotated. One hundred and twenty-four miR-124 targets were also predicted of which 32 are annotated. In human these transposable elements (TE) derived miRNAs have the potential to regulate thousands of human genes. We utilized an L2 derived miRNA miR-28 and three different miRNA target prediction programs miTarget, miRanda, and RNAhybrid to predict a potential miRNA targets in human L2 transposable elements. It was shown that the human TE is also a potential target for miRNAs and subsequently predicted 1,094 of potential target in L2.

121

APENDICES

Apendix A

JAVA CODE FOR MIR-EXPLORE

//The program is called miR-Explore. It will predict miRNA-precursor

//This program is written by Bram Sebastian

```
double score=0; // this is the variable to initialize the score of the scanning
                                                                                                                                                             System.out.println("Please specify the input, query and output files...");
                                                                                                                                                                                                                                                                                                                                                                                                                                                                     int count3=0; // this will help in the scanning of the pairPosition
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               double maxScore=0; // this is to store the maximum score
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          int maxGap=0; //# of gap in the loop
                                                                                                        public static void main (String[]args) {
                                                                                                                                    if (args.length != 3) {
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 int countGapOut=0;
                                                                                                                                                                                                                                                                         int wordlength1=0;
                                                                                                                                                                                                                                                                                                    int wordlength2=0;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          int windowSize=0;
                                                                                                                                                                                          System.exit(0);
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     int countGap=0;
                                                                                                                                                                                                                                                                                                                                int control=0;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              int posMax=0;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                int count4=0;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           int count5=0;
                                                                                                                                                                                                                                                                                                                                                                                                                int count1=0;
                                                                                                                                                                                                                                                                                                                                                                                                                                           int count2=0;
                                                                                 public class ScanAdd {
                                                                                                                                                                                                                                                 String s;
                                                                                                                                                                                                                                                                                                                                                           int i=0;
                                                                                                                                                                                                                                                                                                                                                                                      int j=0;
                       import java.util.*;
import java.io.*;
```

```
char[] chrInput = null; //this will be an array of the input chrm.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                ArrayList<Integer> pairPositionHelpI = new ArrayList<Integer>();
int chrLength=0,// this is to initialize the length of the chrm
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               ArrayList<Integer> pairPositionI = new ArrayList<Integer>();
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    &rrayList<Integer> pairPositionJ = new ArrayList<Integer>();
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      &rrayList<Integer> gapPositionI = new ArrayList<Integer>();
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           ArrayList<Matrix> collectMatrix = new ArrayList<Matrix>();
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   ArrayList<String> nucleotide = new ArrayList<String>();
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         String[]chrInput1=null;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    word1 = null;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                count____=0;
word = null;
                                                                                                                                                                                                                                                                                                                                                                                                                  countU =0;
count A=0;
count G=0;
count C=0;
count U=0;
                                     int fn=windowSize;
                                                       double countAA=0;
                                                                                                                                                                                                                                                                                                                                                                               double countUC=0;
                                                                          countAG=0;
                                                                                                             countAU=0;
                                                                                                                               double countA =0;
                                                                                                                                                                                                        countGU=0;
                                                                                                                                                                                                                            countG =0;
                                                                                                                                                                                                                                                                                 countCC=0;
                                                                                                                                                                                                                                                                                                                        countC_=0;
                                                                                                                                                                                                                                                                                                                                                            countUG=0;
                                                                                                                                                                                                                                                                                                                                                                                                countUU=0;
                                                                                                                                                                                                                                                                                                    countCU=0;
                                                                                                                                                                                                                                                                                                                                          countUA=0;
                                                                                            double countAC=0;
                                                                                                                                                  countGA=0;
                                                                                                                                                                     countGG=0;
                                                                                                                                                                                      countGC=0;
                                                                                                                                                                                                                                              countCA=0;
                                                                                                                                                                                                                                                               countCG=0;
                   int maxLoop=0;
                                                                                                             double
                                                                          double
                                                                                                                                                   double
                                                                                                                                                                     double
                                                                                                                                                                                      double
                                                                                                                                                                                                         double
                                                                                                                                                                                                                            double
                                                                                                                                                                                                                                                                                                     double
                                                                                                                                                                                                                                                                                                                                          double
                                                                                                                                                                                                                                                                                                                                                             double
                                                                                                                                                                                                                                                                                                                                                                                                double
                                                                                                                                                                                                                                                                                                                                                                                                                                                                           double
                                                                                                                                                                                                                                               double
                                                                                                                                                                                                                                                                                    double
                                                                                                                                                                                                                                                                                                                        double
                                                                                                                                                                                                                                                                                                                                                                                                                     double
                                                                                                                                                                                                                                                                                                                                                                                                                                       double
                                                                                                                                                                                                                                                                                                                                                                                                                                                         double
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              double
                                                                                                                                                                                                                                                                   double
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 double
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  char[]
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      char[]
```

<pre>ArrayList<integer> gapPositionJ = new ArrayList<integer>(); Matrix m1 = new Matrix (1,1);// this matrix is the matrix of the input converted into numbers Matrix mp = new Matrix (5,5);// this matrix is the matrix of the pairs String inpt = args[0]; trv{</integer></integer></pre>
<pre>BufferedReader in = new BufferedReader (new FileReader (inpt)); while ((s=in.readLine())!=null){ // this loop is just to be able to create the ml</pre>
<pre>count1++; word = s.toCharArray();</pre>
<pre>} ml = new Matrix (count1,word.length+10); windowSize = word.length; BufferedReader inn = new BufferedReader (new FileReader (inpt)); while ((s=inn.readLine())!=null){</pre>
<pre>count2++; word = s.toCharArray();</pre>
<pre>tor(int a=U;a<word.length;a++) (word[a]="-")="" if="" m1.add(count2-1,a,5);="" pre="" {="" word[a]="." }<=""></word.length;a++)></pre>
<pre>} else if (word[a]=='A') { m1.add(count2-1,a,1); </pre>
<pre>} else if (word[a]=='G') { m1.add(count2-1,a,2);</pre>
<pre>} else if (word[a]=='C') { m1.add(count2-1,a,3);</pre>
<pre>} else{ m1.add(count2-1, a, 4); }</pre>
<pre>//creating the matrices plus calculate the score for each pairs in the training data wordlength2=word.length-1; for(i=wordlength1;i<word.length 2;i++){<="" pre=""></word.length></pre>

matrix

124

```
if(m1.get(0,i)==1 && m1.get(0,j)==1){
    mp = new Matrix (5,5); // create matrix for pairs
    mp.fill(5,5,0); // fill in the matrix
                                                                                                                                                                                                                                                                                                                                                                                                            else if (ml.get(z,i)==1 && ml.get(z,j)==2){
    countAG++;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             else if (ml.get(z,i)==1 && ml.get(z,j)==5){
    countA_++;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 else if (ml.get(z,i) == 2 \& \& ml.get(z,j) == 4) {
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              else if (ml.get(z,i)==1 && ml.get(z,j)==3) {
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               else if (ml.get(z,i)==1 && ml.get(z,j)==4){
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              else if (ml.get(z,i)==2 && ml.get(z,j)==1){
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                else if (m1.get(z,i) == 2 \& m1.get(z,j) == 2) {
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                else if (ml.get(z,i) == 2 \& ml.get(z,j) == 3) 
                                                                                                                                                                                                                                                                                                                              if(m1.get(z,i)==1 && m1.get(z,j)==1){
wordlength2=word.length-control-1;
                                                                                                                                                                                                                                                for (int z=1;z<count1;z++) {</pre>
                                                                                                                                                               pairPositionHelpI.add(i);
                                                                                                                                    pairPositionI.add(i);
                                                                                                                                                                                         pairPositionJ.add(j);
                                                                                                                                                                                                                                                                       //ml.show(z,i);
                                                                                                                                                                                                                                                                                                    //ml.show(z,j);
                                                                                                                                                                                                                                                                                                                                                            countAA++;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            countGA++;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              countGG++;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            countGC++;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            countAC++;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               countGU++;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           countAU++;
                          (j=wordlength2;j>i;j--){
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          ــــ
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         ~
                            for
```

```
else if (ml.get(z,i)==3 && ml.get(z,j)==5) {
    countC_++;
else if (ml.get(z,i)==2 && ml.get(z,j)==5){
    countG_++;
                                                                                    else if (ml.get(z,i) == 3 &  ml.get(z,j) == 1) {
                                                                                                                                                                        else if (ml.get(z,i)==3 && ml.get(z,j)==2) {
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            else if (ml.get(z,i) ==4 && ml.get(z,j) ==1) {
    countUA++;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                else if (m1.get(z,i) ==4 \& m1.get(z,j) ==2) {
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    else if (ml.get(z,i) ==4 \&\& ml.get(z,j) ==3) {
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        else if (m1.get(z,i)==4 && m1.get(z,j)==4){
    countUU++;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            else if (ml.get(z,i)==4 && ml.get(z,j)==5){
    count_A++;
                                                                                                                                                                                                                                                              else if (ml.get(z,i) == 3 & ml.get(z,j) == 3) {
                                                                                                                                                                                                                                                                                                                                                  else if (ml.get(z,i)==3 && ml.get(z,j)==4){
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  else if (ml.get(z,i)==5 && ml.get(z,j)==1) {
    count_G++;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        else if (ml.get(z,i)==5 && ml.get(z,j)==2){
                                                                                                                                                                                                                                                                                                                                                                            countCU++;
                                                                                                                   countCA++;
                                                                                                                                                                                                                                                                                           countCC++;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            countUG++;
                                                                                                                                                                                                       countCG++;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  countUC++;
```

```
\ensuremath{\ensuremath{\mathcal{N}}} adding the value of the probability of each pair to the matrix mp
                               else if (m1.get(z,i)==5 && m1.get(z,j)==3) {
    count_U++;
                                                                                    else if (m1.get(z,i)==5 && m1.get(z,j)==4) {
    countU_++;
count_C++;
                                                                                                                                                          count_++;
                                                                                                                                                                                                                                 if (countAA!=0)
                                                                                                                                                                                                                                                  mp.add(0,0,1);
                                                                                                                                                                                                                                                                    if (countAG!=0)
                                                                                                                                                                                                                                                                                    mp.add(0,1,1);
                                                                                                                                                                                                                                                                                                                                                                          if(countA_!=0)
mp.add(0,4,1);
                                                                                                                                                                                                                                                                                                                                                                                                                              mp.add(1,0,1);
if(countGG!=0)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          if (countg_!=0)
mp.add(1,4,1);
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              if (countCG!=0)
mp.add(2,1,1);
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  if (countCC!=0)
                                                                                                                                                                                                                                                                                                      if (countAC!=0)
                                                                                                                                                                                                                                                                                                                                          if (countAU!=0)
                                                                                                                                                                                                                                                                                                                                                          mp.add(0,3,1);
                                                                                                                                                                                                                                                                                                                                                                                                              if (countGA!=0)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     if (countGC!=0)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        if (countGU!=0)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         mp.add(1,3,1);
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             if (countCA!=0)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             mp.add(2,0,1);
                                                                                                                                                                                                                                                                                                                                                                                                                                                                 mp.add(1,1,1)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     mp.add(1,2,1)
                                                                                                                                                                                                                                                                                                                      mp.add(0,2,1)
                                                                                                                                          else{
                                                                       \dot{}
```

collectMatrix.add(mp); if (count_C!=0)
mp.add(4,2,1);
if (count_U!=0)
mp.add(4,3,1);
if (count_!=0)
mp.add(4,4,1); if (countC_!=0)
mp.add(2,4,1);
if (countUA!=0)
mp.add(3,0,1);
if (countUG!=0) if (countU_!=0) mp.add(3,4,1); if(count_A!=0)
mp.add(4,0,1); if(count_G!=0)
mp.add(4,1,1); mp.add(2,3,1); mp.add(3,2,1);
if(countUU!=0) mp.add(3,3,1); mp.add(2,2,1); if (countCU!=0) mp.add(3,1,1); if (countUC!=0) countG_=0; countCA=0; countAA=0; countAU=0; countGU=0; countAG=0; countAC=0; countA_=0; countGA=0; countGG=0; countGC=0; countCG=0;

```
for(int d=1+pairPositionI.get(pairPositionI.size()-
                                                                                                                                                                                                                                                                                                                                                                                                                                                                           BufferedReader in = new BufferedReader (new FileReader (inpt));
                                                                                                                                                                                                                                                                                                                                                                                                                                      //this try catch is to count the gap in the loop
                                                                                                                                                                                                                                                                              control=word.length-j;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              1);d<pairPositionJ.get(pairPositionJ.size()-1);d++) {
    if(word1[d]=='-'){</pre>
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              while ((s=in.readLine())!=null){
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       countGap++;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       maxGap=countGap;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               if (maxGap<countGap) {
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   word1 = s.toCharArray();
                                                                                                                                     countU =0;
count A=0;
count G=0;
count C=0;
count U=0;
count =0;
                                       countC_=0;
countUA=0;
countCC=0;
                   countCU=0;
                                                                                                                  countUU=0;
                                                                             countUG=0;
                                                                                               countUC=0;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        if(count4>1){
                                                                                                                                                                                                                                                                                                ; 0=0;
                                                                                                                                                                                                                                                                                                                                                                                               catch (IOException e) {
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     count4++;
                                                                                                                                                                                                                                                                                                                                                          ~
                                                                                                                                                                                                                                                                                                                                                                                                                                                          try{
```

```
129
```

```
for(int d=pairPositionJ.get(0);d>pairPositionJ.get(pairPositionI.size()-1);d-
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           maxLoop=pairPositionJ.get(pairPositionJ.size()-1)-pairPositionI.get(pairPositionI.size()-1)-1;
// This is the start of the scanning
String inpt1 = args[1];
                                                                                                                                                                                                                                                                                                                                                                                               for(int d=pairPositionI.get(0);d<pairPositionI.get(pairPositionI.size()-</pre>
                                                                                                                                                                                                               //this try catch is to count the gap outside of the loop in the consensus structure
                                                                                                                                                                                                                                                                         BufferedReader in = new BufferedReader (new FileReader (inpt));
while ((s=in.readLine())!=null){
                                                                                                                                                                                                                                                                                                                                                                                                                                                                           gapPositionI.add(d);
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         if(word1[d] == '_') {
    gapPositionJ.add(d);
                                                                                                                                                                                                                                                                                                                                                                                                                                                         if(word1[d] == ' ') {
                                                                                                                                                                                                                                                                                                                                    word1 = s.toCharArray();
countGap=0;
                                                                                                                                                                                                                                                                                                                                                                 if(count5==0){
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     String output = args[2];
                                                                                                                         catch (IOException e) {
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   catch (IOException e) {
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               }
count5++;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          ~
                                                                                                                                                                                                                                                try{
                                                                                                                                                                                                                                                                                                                                                                                                                                1);d++){
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             }
(
```

```
130
```

<pre>BufferedReader in = new BufferedReader (new FileReader (inpt1)); PrintWriter q = new PrintWriter(new FileOutputStream(output), true); while ((s=in.readLine())!=null) { chrInput1 = s.split(" "); //chrInput1[0] = chrInput1[0].replaceAll("\n"," "); nucleotide.add(chrInput1[0]); //System.out.println(chrInput1[0]); //System.out.println(chrInput1[0]); for(int y=0;y<nucleotide.size();y++) chrinput="nucleotide.get(y).toCharArray();" pre="" {="" }="" }<=""></nucleotide.size();y++)></pre>	<pre>for (int c=0;c<chrinput.length;c++) (chrinput[c]="t")="" ;="" chrinnut[c]="Character" chrinput[c]="u" if="" pre="" toiinnercase(chrinnut[c]).<="" {="" chrinput[c]="T" }=""></chrinput.length;c++)></pre>	<pre>//System.out.println (maxGap); //System.out.println (maxGap); //System.out.println (maxGap); for(int a=maxGap;a>=0;a){// this is to shift the gap in the loop for(int shiftI=gapPositionI.size();shiftI>=0;shiftI){ //this is to shift the I</pre>	<pre>for(int hi=0;hi<pairpositioni.size();hi++) if(gappositioni.size()!="0)" if(pairpositioni.get(hi)="" {="">gapPositionI.get(gapPositionI.size()-1)) pairPositionI.set(hi,pairPositionI.get(hi)-shiftI); }</pairpositioni.size();hi++)></pre>	<pre>//System.out.println(pairPositionI); for(int shiftJ=gapPositionJ.size();shiftJ>=0;shiftJ){//this is to shift the J</pre>	<pre>for(int z=0;z<chrinput.length-windowsize;z++){ &&*="" (st);="" (z);="" *st<(z+fn)="" count3="0;" for(int="" if(="" if(count3<pairpositioni.size())="" st="pairPositionI.get(count3)+z){</pre" system.out.println=""></chrinput.length-windowsize;z++){></pre>
try			or gap		ել չ ո ւ
			arm f	ب بر ر	

```
else if(chrInput[st]=='A'&& chrInput[fn]=='_'){
    //System.out.println(collectMatrix.get(count3).get(0,4));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  //System.out.println(collectMatrix.get(count3).get(0,3));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            //System.out.println(collectMatrix.get(count3).get(1,0));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      //System.out.println(collectMatrix.get(count3).get(1,2));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               //System.out.println(collectMatrix.get(count3).get(1,1));
                                                                                                                                                                                                                                                                                                                                                                                 //System.out.println(collectMatrix.get(count3).get(0,0));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        //System.out.println(collectMatrix.get(count3).get(0,1));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           //System.out.println(collectMatrix.get(count3).get(0,2));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             score = score +(collectMatrix.get(count3).get(0,1));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                score = score +(collectMatrix.get(count3).get(0,2));
                                                                                                                                                                                                                                                                                                                                                                                                                                          score = score +(collectMatrix.get(count3).get(0,0));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       score = score +(collectMatrix.get(count3).get(0,3));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              score = score +(collectMatrix.get(count3).get(0,4));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             score = score +(collectMatrix.get(count3).get(1,0));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           score = score +(collectMatrix.get(count3).get(1,2));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        score = score + (collectMatrix.get(count3).get(1,1));
                                                                                                             //System.out.println (pairPositionI.get(count3));
                                                 fn=pairPositionJ.get(count3)+z-a-shiftJ-shiftI;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   else if(chrInput[st]=='A'&& chrInput[fn]=='G'){
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      else if(chrInput[st]=='A'&& chrInput[fn]=='C'){
                                                                                                                                                              //System.out.println (pairPositionJ.get(count3));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       else if(chrInput[st]=='G'&& chrInput[fn]=='G'){
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             else if(chrInput[st]=='A'&& chrInput[fn]=='U'){
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       else if(chrInput[st]=='G'&& chrInput[fn]=='A'){
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              else if(chrInput[st]=='G'&& chrInput[fn]=='C'){
                                                                                                                                                                                                                                                                                                                                if(chrInput[st]=='A'&& chrInput[fn]=='A'){
                                                                                                                                                                                                                      //System.out.println(chrInput[st]);
                                                                                                                                                                                                                                                                               //System.out.println(chrInput[fn]);
//System.out.println ("yo");
```

```
132
```

```
//System.out.println(collectMatrix.get(count3).get(1,3));
                                                                                                                                                                                                                                                                   //System.out.println(collectMatrix.get(count3).get(1,4));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                //System.out.println(collectMatrix.get(count3).get(2,0));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             //System.out.println(collectMatrix.get(count3).get(2,1));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             //System.out.println(collectMatrix.get(count3).get(2,2));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          //System.out.println(collectMatrix.get(count3).get(2,3));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             //System.out.println(collectMatrix.get(count3).get(2,4));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           //System.out.println(collectMatrix.get(count3).get(3,0));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        //System.out.println(collectMatrix.get(count3).get(3,1));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     //System.out.println(collectMatrix.get(count3).get(3,2));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                score = score +(collectMatrix.get(count3).get(2,1));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 score = score +(collectMatrix.get(count3).get(2,2));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             score = score +(collectMatrix.get(count3).get(2,3));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       score = score +(collectMatrix.get(count3).get(2,4));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       score = score +(collectMatrix.get(count3).get(3,0));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    score = score +(collectMatrix.get(count3).get(3,1));
                                                                                                      score = score +(collectMatrix.get(count3).get(1,3));
                                                                                                                                                                                                                                                                                                                      score = score +(collectMatrix.get(count3).get(1,4));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   score = score +(collectMatrix.get(count3).get(2,0));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          if(chrInput[st]=='C'&& chrInput[fn]=='G'){
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    else if(chrInput[st]=='C'&& chrInput[fn]==' '){
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 else if(chrInput[st]=='U'&& chrInput[fn]=='G'){
else if(chrInput[st]=='G'&& chrInput[fn]=='U'){
                                                                                                                                                                                                                if(chrInput[st]=='G'&& chrInput[fn]==' '){
                                                                                                                                                                                                                                                                                                                                                                                                                             if(chrInput[st]=='C'&& chrInput[fn]=='A'){
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          else if(chrInput[st]=='C'&& chrInput[fn]=='C'){
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       else if(chrInput[st]=='C'&& chrInput[fn]=='U'){
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    else if(chrInput[st]=='U'&& chrInput[fn]=='A'){
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  else if(chrInput[st]=='U'&& chrInput[fn]=='C'){
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 else
                                                                                                                                                                                                                                                                                                                                                                                                                                    else
                                                                                                                                                                                                                   else
```

```
else if(chrInput[st]=='U'&& chrInput[fn]=='_'){
    //System.out.println(collectMatrix.get(count3).get(3,4));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   else if(chrInput[st]=='_&& chrInput[fn]=='G'){
    //System.out.println(collectMatrix.get(count3).get(4,1));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      else if(chrInput[st]=='_`&& chrInput[fn]=='U'){
    //System.out.println(collectMatrix.get(count3).get(4,3));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                //System.out.println(collectMatrix.get(count3).get(4,0));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            //System.out.println(collectMatrix.get(count3).get(4,2));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               //System.out.println(collectMatrix.get(count3).get(4,4));
                                                                                                                                         //System.out.println(collectMatrix.get(count3).get(3,3));
                                                                                                                                                                                                                                                                                                                                                                               score = score +(collectMatrix.get(count3).get(3,4));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          score = score +(collectMatrix.get(count3).get(4,0));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  score = score +(collectMatrix.get(count3).get(4,1));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        score = score +(collectMatrix.get(count3).get(4,3));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               score = score +(collectMatrix.get(count3).get(4,4));
score = score +(collectMatrix.get(count3).get(3,2));
                                                                                                                                                                                       score = score +(collectMatrix.get(count3).get(3,3));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             score = score +(collectMatrix.get(count3).get(4,2));
                                                                                                                                                                                                                                                                                                                                                                                                                                                                           else if(chrInput[st]==' ' && chrInput[fn]=='A'){
                                                                                           else if(chrInput[st]=='U'&& chrInput[fn]=='U'){
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 else if(chrInput[st]==' '&& chrInput[fn]=='C'){
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    else if(chrInput[st]==' ' && chrInput[fn]==' '){
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          //System.out.println(chrInput[z+fn]);
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         //System.out.println(chrInput[st]);
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  score = score + 0;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              count3++;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 //else {
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           //fn--;
```

```
for(int hi=0;hi<pairPositionI.size();hi++){
    pairPositionI.set(hi,pairPositionHelpI.get(hi));</pre>
//System.out.println (score);
//System.out.println();
                             if (maxScore<score) {</pre>
                                                                                                               //fn=windowSize;
                                           maxScore=score;
                                                                                                                                                                                                                               q.println(maxScore);
                                                                       //score=0;
                                                         posMax=z;
                                                                                                   score=0;
                                                                                                                                                                                         maxScore=0;
                                                                                                                                                                                                                                                                        }
catch (IOException e){
                                                                                                                                                                                                                     ~
                                                                                                                                <u>~</u> ~
                                                                                                                                                                                                      ~
                                                                                                                                                                                                                                                              ~
                                                                                                                                                                                                                                                                                                        ~ ~
```

~
Apendix B

PYTHON CODE FOR TARSPEC

#!/usr/bin/env python import sys #This script was written by to scan an input sequence in FASTA format #for possible miRNA target candidates in Platypus #Written by Bram Sebastian, UGA IOB.

#class declaration for FASTA class
class Fasta:
 def __init__(self, name, sequence):
 #the following line will store sequence name
 self.name = name
 #the following will store the sequence
 self.sequence = sequence

#the following function will receive the fasta files
#and store it in an array (both name and sequence are stored)
def read_fasta(file):
 #this items array will store the sequence
 items =[]
 index = 0
 for line in file:
 if line.startswith(">>"):
 if line.startswith(">"):
 if index >= 1
 index += 1

```
seq = "
aninstance = Fasta(name,seq)
else:
    seq += line[:-1]
    aninstance = Fasta(name,seq)
items.append(aninstance)
return items
```

fastafile2 = open('Plat3primeUTRAll.txt','r'').readlines()#This line is to open the input sequence file fastafile1 = open("hsaMIR124.txt","r").readlines()#This line is to open the miRNA data file output = open('outPredictedOANTarget124.txt',"w")#This line is to create the output file

#Declare variable to store the fasta object from the two files above miRsequences = read_fasta(fastafile1) taRsequences = read_fasta(fastafile2)

count2to5=0 #number of match in position 1 to 4 starting from the 5' end of the mature miRNA #Initiate variable to count classification based on what has been used by miRANDA tarLength = len(i.sequence) #this line define the length of the target mRNA count9to22=0 #number of match in position 9 to L-5 of the mature miRNA count1to8=0 #number of match in position 3 to 12 of the mature miRNA #Start scanning the input sequence for j in miRsequences: for i in taRsequences: countTargets=0 countGC = 0countAU = 0countGU=0 inarow = 0miss = 0

length = len(j.sequence) #this line define the length of the miRNA

if j.sequence[a].upper()=='A' and (i.sequence[a+b].upper()=='T' or i.sequence[a+b].upper()=='U'): if j.sequence[a].upper()=='U' and (i.sequence[a+b].upper()=='A'): elif j.sequence[a].upper()=='G' and i.sequence[a+b].upper()=='C': if j.sequence[a].upper()=='G' and (i.sequence[a+b].upper()=='C'): if j.sequence[a].upper()=='U' and (i.sequence[a+b].upper()=='A'): if j.sequence[a].upper()=='G' and (i.sequence[a+b].upper()=='C'): elif j.sequence[a].upper()=='C' and i.sequence[a+b].upper()=='G': if j.sequence[a].upper()=='C' and i.sequence[a+b].upper()=='G': if j.sequence[a].upper()=='C' and i.sequence[a+b].upper()=='G': if $a \ge \text{length-8}$ and $a \le \text{length-1}$: for b in range (0,(tarLength-length)+1): if a \geq length-5 and a \leq length-2: if $a \ge 0$ and $a \le length-1$: for a in range (0, length): count9to22 +=1 count9to22 +=1 count2to5 +=1 countAU +=1 count2to5 +=1 count1 to 8 +=1count2to5 +=1 count1 to 8 +=1count1 to 8 +=1count2to5 +=1 count1 to 8 += 1countAU +=1 countGC +=1 #if a == length-9:

count9to22 +=1
countGC +=1
countGC +=1
elif j.sequence[a].upper()=='U' and i.sequence[a+b].upper()=='A':
countAU +=1
countAU +=1
elif j.sequence[a].upper()=='U' and i.sequence[a+b].upper()=='G':
countGU += 1
countGU += 1
elif j.sequence[a].upper()=='G' and i.sequence[a+b].upper()=='U':
countGU += 1
elif j.sequence[a].upper()=='G' and i.sequence[a+b].upper()=='U':
fif i.sequence[a].upper()=='G' and i.sequence[a+b].upper()=='G':
fif i.sequence[a].upper()=='G':
fif i.sequence[a].upper()=='G':
fif i.sequence[a+b].upper()=='G':
fif i.

#The following is the rules, and can be change accordingly to best fit a particular miRNA Class

if count2to5 ≥ 4 and count1to8 ≥ 7 and countAU ≥ 4 and countGC ≥ 4 : output.write (i.sequence+"\n\n") #output.write (j.name+"\n") output.write (i.name+"\n") countTargets +=1 count9to22 = 0count1to8 = 0count2to5 = 0countGU = 0countAU = 0countGC=0

print countTargets