Journal of Theoretical Biology 286 (2011) 79–84

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Human microRNA target identification by RRSM

Wan J. Hsieh, Hsiuying Wang*

Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan

ARTICLE INFO

Article history:
Received 26 November 2010
Received in revised form
26 May 2011
Accepted 17 June 2011
Available online 29 June 2011

Keywords:
Microarray expression
miRNA
Relative R^2 method
Regression model
Correlation

ABSTRACT

MicroRNAs (miRNAs) are small endogenously expressed non-coding RNAs that regulate target messenger RNAs in various biological processes. In recent years, there have been many studies concentrated on the discovery of new miRNAs and identification of their mRNA targets. Although researchers have identified many miRNAs, few miRNA targets have been identified by actual experimental methods. To expedite the identification of miRNA targets for experimental verification, in the literature approaches based on the sequence or microarray expression analysis have been established to discover the potential miRNA targets. In this study, we focus on the human miRNA target prediction and propose a generalized relative R^2 method (RRSM) to find many high-confidence targets. Many targets have been confirmed from previous studies. The targets for several miRNAs discovered by the HITS-CLIP method in a recent study have also been selected by our study.

1. Introduction

MicroRNAs (miRNAs) are endogenous and single-stranded ~23 nt RNAs that play crucial gene regulatory roles in animals and plants by pairing to the 3′ untranslated regions (UTRs) of the target messenger RNAs (mRNAs) of protein -coding genes to direct their post-transcriptional repression (Carrington and Ambros, 2003; Bartel, 2004; Mattick and Makunin, 2006). Extensive research has revealed the existence of more than 700 different human miRNAs (Griffiths-Jones et al., 2008). Griffiths-Jones et al. (2008)and several studies have demonstrated the importance of miRNA-mediated regulation in a wide range of basic biological processes, such as proliferation, apoptosis, cellular identity and pathogen–host interactions (Pillai et al., 2007; Carthew and Sontheimer, 2009).

The discovery of many miRNAs in various multi-cellular species has raised many questions, such as how these small non-coding RNAs function in cells. The key to answering this particular question is to explore their regulatory targets. The most general feature of miRNA regulation is the recognition of sequence motifs complementary to the 3′UTR of target mRNAs (Lewis et al., 2003; Grimson et al., 2007).

Several target prediction computational algorithms for motifs complementary predictions have been developed, for example, miRanda (John et al., 2004), TargetScan (Lewis et al., 2003; Lewis et al., 2005) and PicTar (Krek et al., 2005), but they show poor overlap between their predicted results, which might be caused by a number of false-negative and probably also false positive predictions (Bartel, 2009).

In addition to sequence motifs complementary predictions, gene expression profiling can also provide useful information for studying the biological functions of miRNAs. Therefore expression data analysis has been used as a complementary method for discovering miRNA targets (Lim et al., 2005). However, it can become computationally complicated when considering multiple miRNAs and their effects across multiple tissues. To overcome this difficulty, Huang et al. (2007b) and Wang and Li (2009b) proposed statistical methods to build up a network of associations between the miRNAs and their target mRNAs.

Huang et al. (2007b) established a method, GenMiR++, using Bayesian variation analysis to explore miRNA targets. However, it is complicated and requires extensive calculations. In order to provide a more effective approach, Wang and Li (2009b) proposed the relative R^2 method to select high-confidence targets of miRNAs from prediction targets, which is easy to interpret and less computationally expansive. This method successfully obtained many high-confidence targets for mouse miRNA in Wang and Li (2009b). In this study, we generalize the relative R^2 method to a more flexible form and called it as RRSM. We also establish program codes for performing RRSM for different original data and normalized data.

RRSM has several virtues for discovering high-confidence targets. Although the paired correlation analysis between miRNA and their targets has been discussed (Ritchie et al., 2009; Wang and Li, 2009a; Liu et al., 2010), observing several confirmed targets in the literature indicates that for many miRNAs, the correlation coefficient of the microarray expression of a miRNA and that of its confirmed target is nearly zero. The discussion and comparison of RRSM and the existing correlation analysis methods (Ritchie et al., 2009; Wang and Li, 2009a; Liu et al., 2010; Wang et al., in press) are given in Section 3.

When the correlation coefficient is not high, it is hard to use any standard statistical approaches to explore miRNA targets because

* Corresponding author. Tel.: +886 3 5712121x56813; fax: +886 3 5728745.
E-mail address: wang@stat.nctu.edu.tw (H. Wang).

0022-5193/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2011.06.022

there are no significant statistical evidence for a relationship between a miRNA and its true targets in terms of the conventional statistical methods. In contrast, since RRSM is derived from a relative instead of an absolute statistical viewpoint, it can provide an efficient way to identify the correct targets.

Wang and Li (2009b) demonstrated a great improvement for analyzing mouse miRNAs (Babak et al., 2004; Huang et al., 2007b). In this study, we focus on human miRNAs target prediction (Huang et al., 2007a). The analysis results clearly show that more interactions occur as verified by TarBase (Papadopoulos et al., 2009) and dataset mimiRNA (Ritchie et al., 2010) obtained from the high-confidence targets selected by RRSM than from those selected by GenMiR++ in Huang et al. (2007a).

Recently, the HITS-CLIP method, an approach relying on purifying RNA-binding proteins (RNABPs), has been developed to directly identify protein–RNA interactions in living tissues in a genome-wide manner. The unbiased nature of this platform has the potential for new discoveries, including the elucidation of preferred binding sequences and the identification of regulated RNA substrates (Jensen and Darnell, 2008; Licatalosi et al., 2008; Chi et al., 2009).

For comparison with the HITS-CLIP method, we also show that targets identified by the HITS-CLIP method can be identified by RRSM for targets appearing in both datasets (Huang et al., 2007a; Chi et al., 2009). The results reveal that RRSM can provide an appropriate means to discover correct human miRNA targets.

In this study, we explore 1559 high-confidence targets (Table S1) for human miRNAs and verify that many selected targets have been confirmed through previous studies. The RRSM methods and codes are provided on a website for readers to explore high-confidence miRNA targets. An R code user manual for running the RRSM code is given in the website to help biologists using the codes.

2. Results

RRSM is established based on a relative instead of an absolute statistical point of view and it provides an efficient approach

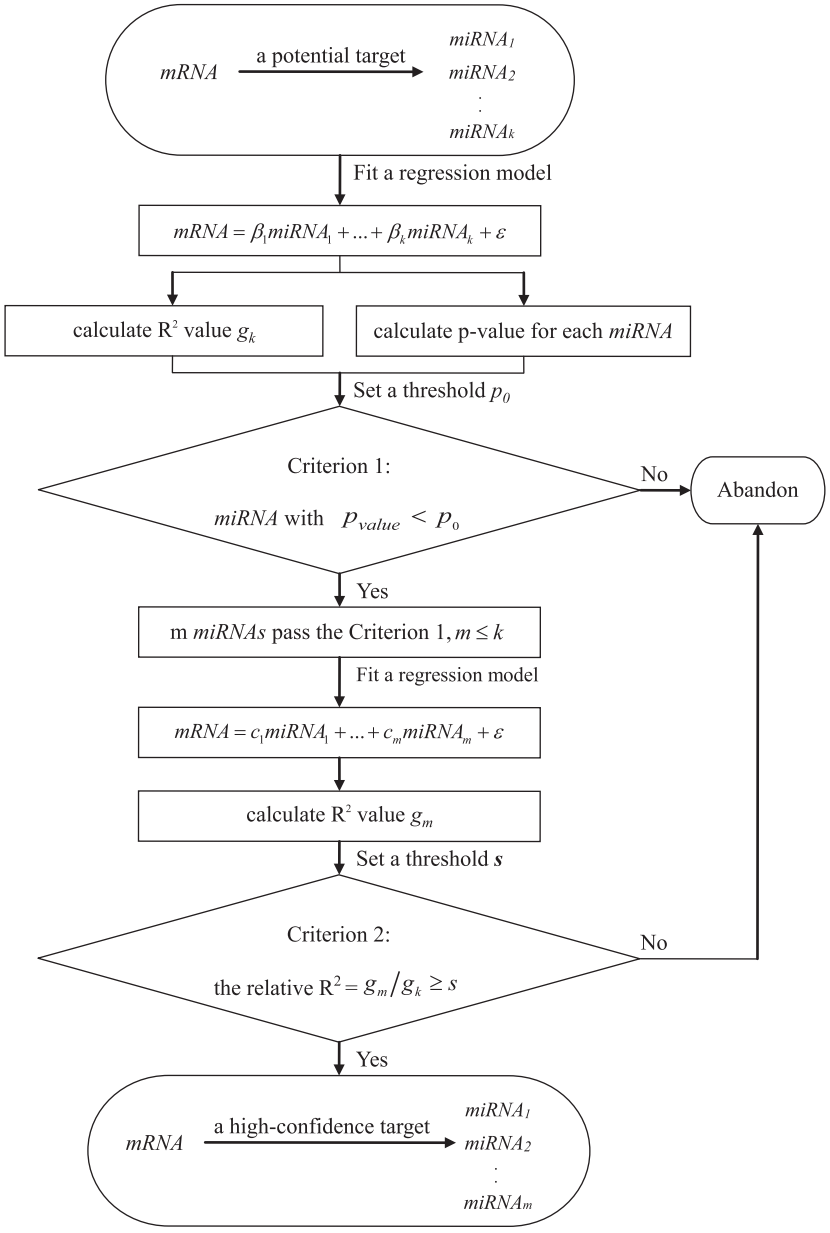

Fig. 1. The flowchart of the procedure for RRSM.

for miRNA target identification. In this study, in contrast to the mouse miRNA target analysis, we found that the approach using the original data can lead to a more satisfactory result than using the normalized expression profile, which was adopted in Wang and Li (2009b) for mouse miRNA target analysis. Therefore, we adopt the original data format in this study. Since we can select different transformation form for the miRNA expression data and mRNA data expression data, we propose a more generalized form for the relative R^2 method. The formula of the RRSM is given in Section 5.

Software for RRSM is available on http://www.stat.nctu.edu.tw/~hwang/website_wang%20new.htm

The steps in this method are briefly described in the flowchart (Fig. 1).

3. Data analysis

3.1. RRSM

The main aim of this study is to use RRSM to select high-confidence targets for human miRNAs and compare with other methods. We consider the miRNA and mRNA expression data for 114 human miRNAs and 16,063 mRNAs across a mixture of 88 normal and cancerous tissue samples common to the two datasets used in Huang et al. (2007a). A dataset was filtered from the data to include 6387 potential target pairs, covering 890 unique mRNAs, because some miRNAs have the same mRNAs as their potential targets (Huang et al., 2007a). The purpose of this study is to select high-confidence targets from the potential targets. Huang et al. (2007a) applied the Bayesian variation method for analyzing this dataset, but this method is complicated and has a high computational load. In this current study, we use the RRSM to select high-confidence targets from the potential targets and compare the results with Huang et al. (2007a).

We focus on the 6387 miRNA–mRNA potential target pairs, determining each miRNA and its target, and use the results corresponding to the microarray expression $16{,}063 \times 88$ data matrix and 114×88 data matrix to fit the regression model.

In order to select about one-fourth of the targets from the 6387 potential targets, we set $p_0=0.77$ and $s=0.995$ in RRSM, resulting in 1559 high-confidence targets being selected. Furthermore, there are many other choices of setting p_0 and s such that about 1600 targets could be selected by RRSM. Table 1 shows that we can alter the values of p_0 and s to accommodate our requirements.

To compare the performance of RRSM with the method of Huang et al. (2007a), GenMiR++, we examine the accuracy of both methods by exploring the confirmed targets appearing in TarBase.

We exhaustively searched the confirmed targets for the 6387 potential targets in TarBase and found that there are only 24 common interactions in the 6387 potential targets and TarBase (Papadopoulos et al., 2009). Table S2 shows the 24 TarBase interactions, which are the targets of 8 miRNAs, including miR-16, miR-1, miR-15b, miR-29c, miR-26a, miR-23a, miR-21 and miR-155, among 114 miRNAs.

For comparison, we list the numbers of interactions of the two methods. Using $p_0=0.77$ and $s=0.995$, we obtain 1559 high-confidence targets by RRSM, containing 10 of the 24 interactions. For comparison with the results of GenMiR++, we also exhaustively searched the interactions between TarBase and the results of Huang et al. (2007a) and found that there are only 4 interactions.

Table 1
Different choices of p_0 and s such that the number of potential targets is about 1600.

s	0.995	0.99	0.95	0.9	0.875	0.85
p_0	0.77	0.73	0.63	0.58	0.56	0.55

Table 2
Interaction numbers in TarBase and mimiRNA of the relative R^2 method (RRSM) and GenMiR++.

	Number of high-confidence targets	Number of interactions in TarBase	Number of interactions in mimiRNA
GenMiR++	1597	4	25
RRSM			
$s=0.995$, $p_0=0.77$	1559	10	43
$s=0.995$, $p_0=0.75$	1342	9	34
$s=0.990$, $p_0=0.72$	1485	8	31
$s=0.950$, $p_0=0.60$	1388	8	35
$s=0.900$, $p_0=0.57$	1519	8	33

Further comparisons with RRSM and GenMiR++ are presented in Table 2, where we consider five different thresholds for RRSM such that the number of high-confidence targets selected by these thresholds is near 1600.

For these thresholds there are at least 8 interactions in TarBase. The number is significantly larger than the interaction number 4 obtained from GenMiR++. This reveals that RRSM is more powerful than GenMiR++ (Huang et al., 2007a) for detecting high-confidence targets.

Besides comparing RRSM with GenMiR++ through the number of interactions in TarBase, we also make the comparison through the database mimiRNA (Ritchie et al., 2010) and Table S3 lists 118 interactions in 6387 potential targets appearing in mimiRNA. The "p-value cut off" and "Integrate with data" in the mimiRNA tool are selected to be "0.01" and "none", respectively. Table 2 presents the number of interactions of the methods. There are 25 interactions in mimiRNA among the 1597 targets selected by GenMiR++. There are at least 33 interactions in mimiRNA among the targets selected by RRSM. In both databases, the numbers of targets selected by RRSM are larger than those selected by GenMiR++. It shows there are more validations of the high-confidence targets selected by RRSM than those selected by GenMiR++, revealing the RRSM is a more effective method in predicting high-confidence targets.

In addition to comparing RRSM with GenMiR++, we also demonstrate its feasibility for selecting high-confidence targets of human miRNAs by comparing randomly selected results. As mentioned above, using RRSM to select the number of about one-fourth targets in 6387 targets enables selecting 10 interactions in the 24 interactions in TarBase, which is about $10/24(=0.417)$, making it larger than one-fourth. The larger proportion means that RRSM performs well in selecting the correct targets for human miRNAs. Fig. S1 shows that the proportion of interactions derived by RRSM is greater than the proportions of interactions obtained by a random selection.

This discussion shows that RRSM outperforms GenMiR++ and the randomly selecting methods for different thresholds of s and p-value. RRSM consists of two important criteria, the s value and the p-value. In this method, the threshold selection for the s value is the main criterion and the threshold selection for p-value is an ancillary criterion. Basically, we prefer a strict selection for the s value that may be greater than 0.9 and allows a relax p-value selection that may be less than 0.9.

In addition, we also compare the results with those from the HITS-CLIP method in Chi et al. (2009) and other previous studies. Fig. 6 in Chi et al. (2009) reveals Ago HITS-CLIP targets for miR-124, miR-9 and miR-125, respectively, which are shown in the most significant pathways (neuronal differentiation/cytoskeleton regulation).

There are a total of two targets, RAf1 and IQGAP1, for miR-124 shown in Figure 6 of Chi et al. (2009) appearing in the human miRNA dataset we used in this study. We conduct RRSM in our dataset and find that RAF1 can be selected using $p_0=0.69$ and $s=0.99$, while IQGAP1 can be selected using $p_0=0.85$ and $s=0.999$.

We also examine the mouse miRNA data used in Wang and Li (2009b). There is a total of two targets APC and VCL, for miR-125b and miR-124a, respectively, as shown in Figure 6 of Chi et al. (2009) appearing in the mouse miRNA dataset used in Wang and Li (2009b). We conduct RRSM in our dataset and find that APC can be selected using $p_0=0.87$ and $s=0.999$ and VCL can be selected using $p_0=0.39$ and $s=0.999$.

This study shows that confirmed mRNA targets interacting over Ago–miRNA–mRNA ternary maps can also be selected by RRSM, which demonstrates the validity of RRSM for detecting the relationship between the miRNAs and the mRNAs. The threshold values and proportion of selecting targets are shown in Table 3. Note that there is a total of 1770 targets in the mouse data used in Wang and Li (2009b).

3.2. Correlation analysis

To show the superiority of the proposed method over the standard statistical method for selecting the true targets, which does not show substantial evidence in statistical correlation coefficient analysis, we now examine correlation coefficients for miRNA and their confirmed targets.

In our earlier discussion of the mRNA and miRNA expression data, there are 6387 potential targets. We found that there are 3219 targets with the absolute correlation coefficients less than 0.1, 6174 targets with the absolute correlation coefficients less than 0.3 and 6380 targets with absolute correlation coefficients less than 0.5. For all of the potential targets, the maximal absolute correlation coefficient is about 0.5533. This clearly shows that the correlation coefficients of the miRNAs and their potential targets are not large. Fig. S2(A) summarizes the investigation results. In this case, we also calculate the rates of targets with positive correlation and the negative correlation, respectively, among the 6387 targets, which are presented in Fig. S2(B). Previous studies have pointed out that miRNA expression may be widely down-regulated at its target mRNAs (Calin et al., 2002; Lim et al., 2005; Ruby et al., 2007). But for the data we used, the proportion of negative correlation is not significantly large. The evidence shows that using only the correlation analysis to select miRNA targets might not lead to satisfactory results.

We now apply RRSM to the data using the criteria, $p_0=0.77$ and $s=0.995$, resulting in 1559 high-confidence targets selected, and using $p_0=0.72$ and $s=0.99$, resulting in 1485 high-confidence targets selected.

To verify the agreement between the analysis from RRSM and the down-regulation argument, we demonstrate that using the RRSM to select targets can guarantee that there are a larger proportion of negative correlation targets being selected, as shown in Fig. S3. There are 17 miRNAs with negative correlation coefficient targets proportion greater than 0.7 using the original 6387 targets.

Table 3
Threshold values for RRSM used in select targets in the HITS-CLIP method.

	miRNA	Target gene	p_0	s	Ratio of high-confidence targets to potential targets
Mouse	miR-125b	APC	0.87	0.999	1140/1770=0.644
	miR-124a	VCL	0.39	0.999	234/1770=0.132
Human	miR-124a	RAF1	0.69	0.99	1152/6387=0.18
		IQGAP1	0.85	0.999	2113/6387=0.33

We consider two target sets selected by RRSM for $p_0=0.72$ and $s=0.99$ and $p_0=0.77$ and $s=0.995$. The numbers of miRNAs with the proportion of negative correlation targets greater than 0.7 in these two sets are 28 and 30, respectively.

The comparison shown in Fig. S2(C) reveals that the high-confidence targets selected by RRSM have larger proportions of miRNAs with negative correlation targets, agreeing with the fact that the miRNA usually down-regulates its target. Furthermore, we list the miRNAs with the proportions of negative correlation targets larger than 0.7 in Fig. S4 for the targets selected by RRSM for $p_0=0.77$ and $s=0.995$.

In addition to the above numerical argument used to verify our results, we also find some confirmed targets from the literature in the targets selected by RRSM. Table 4 summarizes the miRNAs and their targets and represents the correlation and related studies. This shows that the correlation analysis is not an effective approach to select targets because most of the confirmed targets do not have high correlation with their corresponding miRNAs, whereas, these confirmed targets can be successfully selected by RRSM.

3.3. Existing correlation analysis methods

For each miRNA/mRNA pair, Ritchie et al. (2009) suggested to calculate a correlation coefficient for human and another for mouse data. Each pair was considered to be a conserved negative correlation (CNC) pair if the correlation coefficient in both human and mouse was below −0.3.

This type of interaction could be detected by miRNA/mRNA pairs that show significant negative correlations in expression in Ritchie et al. (2009). We apply this method to 6387 potential targets of the human data. There are only 65 pairs with a correlation coefficient below −0.3 and none of these 65 targets are interactions in TarBase. Thus, Ritchie et al. (2009) may not be suitable for analyzing the dataset (Huang et al., 2007a, 2007b).

Table 4
The literature of confirmed targets and correlations between the targets and the corresponding miRNAs selected by RRSM $p_0=0.77$ and $s=0.995$.

MiRNA	Target	Correlation	Reference
miR-1	ANXA2	−0.1374	TarBase
	TAGLN2	−0.4129	
	SFRS9	0.0021	
	AP3D1	0.1528	
	H3F3B	0.0092	
miR-23a	CXCL12	0.1485	TarBase
miR-29c	COL1A1	−0.0412	TarBase
	COL1A2	0.1149	
miR-16	BCL2	−0.0039	TarBase; Raveche et al. (2007), Calin et al. (2007), Guo et al. (2009) and Tsang et al. (2009).
miR-15b	BCL2	−0.1625	TarBase; Guo et al. (2009).
miR-15a	BCL2	0.0365	Calin et al. (2007), Garzon et al. (2007).
miR-181a	PCAF	−0.0266	Pichiorri et al. (2008).
miR-181b		0.0978	
miR-106b		−0.0142	
miR-25		−0.0005	
miR-32		−0.2240	
miR-223	LMO2	0.1445	Felli et al. (2009).
miR-21	JAG1	−0.0337	Hashimi et al. (2009).
miR-145	KLF5	0.1993	Cheng et al. (2009).
miR-124a	RAf1	0.4339	Chi et al. (2009) (HITS-CLIP).
	IQGAP1	−0.1177	

Note that the target IQGAP1 for miR-124a can be also selected if we relax the criteria for p_0=0.85 and s=0.999.

In addition, Wang and Li (2009a) and Liu et al. (2010) both apply the correlation analysis on NCI 60 cell lines to investigate the rates of targets with negative or positive correlation. It reveals the significance of the expression profiles between miRNAs and their targets in terms of the correlation analysis from these papers. NCI 60 data are all cancer cell lines. Although the previous studies show the feasibility of using correlation analysis on this dataset, we cannot guarantee the appropriateness of the correlation analysis approach to other dataset. Furthermore, Wang and Li (2009a) mainly compare the proportions of negative correlations of the predicted miRNA–mRNA interactions from TargetScan4.1 and miRBase using NCI 60 data. We apply the method of Wang and Li (2009a) to TaregetScan-predicted interactions in our dataset and discover that the proportion of negative correlations is 57.5%, which is not very significantly larger than the proportion of positive correlations. Based on the result and the aim of this study, which is to predict high-confidence interactions, but not to compare the correlations of interactions from TargetScan4.1 and miRBase, predicts the high-confidence miRNA–mRNA interactions, we did not present the comparison in our paper.

Especially for the dataset used in this study and other studies, Huang et al. (2007a), Huang et al. (2007b) and Wang and Li (2009b) reveal that the correlation analysis cannot show significant result for these datasets and a more involved approach is necessary to be developed for these datasets. We believe that the effect of an approach can be affected by the characteristic of a dataset.

4. Discussion

RRSM has successfully discovered many high-confidence human miRNA targets from the microarray expression data of the miRNAs and the mRNA. It is worth mentioning that compared with GenMiR++ (Huang et al., 2007a), the number of targets obtained from RRSM, which has been verified from TarBase and the previous studies, is significantly larger than that obtained by GenMiR++. A total of 1559 high-confidence targets (Table S1) were discovered in this study and we list targets associating with the corresponding p-values. In the statistical viewpoint, a small p-value indicates the significance of a discovery. There are 269 high-confidence targets with p-value less than 0.1 which can be ranked to be more potential targets than the other 1290 selected targets. In addition, Table 4 shows the 21 selected high-confidence targets are verified through previous studies to be true targets. Furthermore, we found that using the original data format in RRSM can provide a more accurate target prediction than using the normalized data format, which was adopted for mouse data in RRSM. The R codes and MATLAB codes for performing RRSM are established and available in http://www.stat.nctu.edu.tw/~hwang/website_wang%20new.htm.

5. Methods

5.1. Relative R^2 method (RRSM)

We generalize the relative R^2 method proposed in Wang and Li (2009b) to a more general form in this section.

First, suppose we have microarray expression data of n miRNAs, $z_1,\ldots,z_n$, across l tissues, $t_1,\ldots,t_l$, where the expression levels of the n miRNAs in tissue t_j are denoted as $z_{1j},\ldots,z_{nj}$. By prediction methods, such as TargetScan and microarray analyses, potential targets for each of these n miRNAs can be predicted.

RRSM is used to select high-confidence miRNA targets from the set of the predicted miRNA targets using microarray expression data. For each mRNA in the target set, we can find the miRNAs, say $z_1,\ldots,z_k$, such that each of the miRNAs has this mRNA as its potential target. We fit the microarray expression data of the mRNA in terms of the microarray expression of the k miRNAs using the regression model that is written as

$$f(y_j)=b_0g(z_{0j})+b_1g(z_{1j})+b_2g(z_{2j})+\cdots+b_kg(z_{kj})+\varepsilon_j, j=1,\ldots,l, \quad (1)$$

where ε_j is the error term and $f(t)$ and $g(t)$ are functions of t.

If we do not have any preference of choosing functions $f(\cdot)$ and $g(\cdot)$, we can just set $f(\cdot)$ and $g(\cdot)$ to be the identity functions. To select better transformations $f(\cdot)$ and $g(\cdot)$, we can selected several commonly used functions as $f(\cdot)$ or $g(\cdot)$ and derive the results based on different combinations of $(f(\cdot), g(\cdot))$. Finally, we can select a combination of $(f(\cdot), g(\cdot))$ such that the model (1) associated with this combination has the highest number of targets selected. In this miRNA targets study, the functions $f(t)$ and $g(t)$ are select to be the identity functions.

Under the model (1), the least squared estimator of $\beta=(b_0,b_1,\ldots,b_k)^T$ is $\hat{\beta}=(\hat{b}_0,\ldots,\hat{b}_k)^T=(Z^TZ)^{-1}Z^TY$ where $Y=(f(y_1),\ldots f(y_l))^T$, $Z=(w_{ij})_{l\times k}$ and $w_{ij}=g(z_{ij})$. Let $\hat{f}(y_i)=(Z\hat{\beta})_i$. Define $SS_{total}=\sum_i(f(y_i)-\hat{f}(y_i))^2$ and $SS_{reg}=\sum_i(\hat{f}(y_i)-\bar{f})^2$, where $\bar{f}$ denotes the mean of $f(y_1),\ldots f(y_l)$.

The R^2 is defined as SS_{reg}/SS_{total}, which is used as an indication of the fitness of the linear regression model. The number of R^2 lies between 0 and 1 and the larger the value means the model fits better.

We use the R^2 value of fitting an mRNA in terms of the k miRNAs, say g_k, as a baseline to select the high-confidence targets. The method is to select m miRNAs among the k miRNAs such that the R^2, say g_m, for the regression model based on the m miRNAs can satisfy $g_m/g_k\geq s$, where s is a given threshold. The value g_m/g_k is defined as the relative R^2 value. The smaller the m value means the better the results because we want to find small proportion of the high-confidence targets from the potential targets.

The steps of selecting m miRNAs are first to rank the miRNAs based on their p-values under the framework of testing if their corresponding coefficient b_j is equal to 0. The smaller p-value represents the more significant level. The p-value of miRNA z_i is defined as the following:

$$P(|W|\geq|\hat{b}_i|/\sqrt{Var(\hat{b}_i)})$$

where W denotes the standard normal variable. Note here we can set a threshold p_0 for the p-value such that the p-values of the selected miRNAs must be less than the threshold. Combining the above results, we need to set two thresholds, s and p_0, by applying RRSM. Basically, we can select the p_0 and s values based on the proportion of high-confidence targets that we intend to obtain from the set of potential targets.

In this study, we propose a flexible criterion to select a suitable transformation function to build an appropriate regression model. The regression model form can be adjusted by the characteristic of a dataset. We did not find the significant result by applying the correlation analysis to the dataset. With the significant result using the proposed method compared with the correlation analysis, we believe the new method is a potential tool in predicting targets for other datasets.

Acknowledgments

This study was supported by National Science Council and National Center for Theoretical Sciences, Taiwan.

Appendix A. Supplementary Materials

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.yjtbi.2011.06.022.

References

Babak, T., Zhang, W., Morris, Q., Blencowe, B.J., Hughes, T.R., 2004. Probing microRNAs with microarrays: tissue specificity and functional inference. RNA 10, 1813–1819.

Bartel, D.P., 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116, 281–297.

Bartel, D.P., 2009. MicroRNAs: target recognition and regulatory functions. Cell 136, 215–233.

Calin, G.A., et al., 2002. Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proceeding of the National Academy Sciences USA 99, 15524–15529.

Calin, G.A., Pekarsky, Y., Croce, C.M., 2007. The role of microRNA and other non-coding RNA in the pathogenesis of chronic lymphocytic leukemia. Best Practice & Research Clinical Haematology 20, 425–437.

Carrington, J.C., Ambros, V., 2003. Role of microRNAs in plant and animal development. Science 301, 336–338.

Carthew, R.W., Sontheimer, E.J., 2009. Origins and mechanisms of miRNAs and siRNAs. Cell 136, 642–655.

Cheng, Y., Liu, X., Yang, J., Lin, Y., Xu, D.-Z., et al., 2009. MicroRNA-145, a novel smooth muscle cell phenotypic marker and modulator, controls vascular neointimal lesion formation. Circulation Research 105, 158–166.

Chi, S.W., Zang, J.B., Mele, A., Darnell, R.B., 2009. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature 460, 479–486.

Felli, N., Pedini, F., Romania, P., Biffoni, M., Morsilli, O., et al., 2009. MicroRNA 223-dependent expression of LMO2 regulates normal erythropoiesis. Haematologica 94, 479–486.

Garzon, R., Pichiorri, F., Palumbo, T., Visentini, M., Aqeilan, R., et al., 2007. MicroRNA gene expression during retinoic acid-induced differentiation of human acute promyelocytic leukemia. Oncogene 26, 4148–4157.

Griffiths-Jones, S., Saini, H.K., van Dongen, S., Enright, A.J., 2008. miRBase: tools for microRNA genomics. Nucleic Acids Research 36, D154–158.

Grimson, A., et al., 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Molecular Cell 27, 91–105.

Guo, C.-J., Pan, Q., Li, D.-G., Sun, H., Liu, B.-W., 2009. miR-15b and miR-16 are implicated in activation of the rat hepatic stellate cell: an essential role for apoptosis. Journal of Hepatology 50, 766–778.

Hashimi, S.T., Fulcher, J.A., Chang, M.H., Gov, L., Wang, S., et al., 2009. MicroRNA profiling identifies miR-34a and miR-21 and their target genes JAG1 and WNT1 in the coordinate regulation of dendritic cell differentiation. Blood 114, 404–414.

Huang, J.C., et al., 2007a. Using expression profiling data to identify human microRNA targets. Nature Methods 4, 1045–1049.

Huang, J.C., Morris, Q.D., Frey, B.J., 2007b. Bayesian inference of microRNA targets from sequence and expression data. Journal of Computational Biology 14, 550–563.

Jensen, K.B., Darnell, R.B., 2008. CLIP: crosslinking and immunoprecipitation of in vivo RNA targets of RNA-binding proteins. Method Molecular Biology 488, 85–98.

John, B., et al., 2004. Human microRNA targets. PLoS Biology 2, e363.

Krek, A., et al., 2005. Combinatorial microRNA target predictions. Nature Genetics 37, 495–500.

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, 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, 787–798.

Licatalosi, D.D., et al., 2008. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature 456, 464–469.

Lim, L.P., et al., 2005. Microarray analysis shows that some microRNAs down-regulate large numbers of target mRNAs. Nature 433, 769–773.

Liu, H., et al., 2010. mRNA and microRNA expression profiles of the NCI-60 integrated with drug activities. Molecular Cancer Therapeutics 9, 1080–1091.

Mattick, J.S., Makunin, I.V., 2006. Non-coding RNA. Human Molecular Genetics 15, R17–29.

Papadopoulos, G.L., Reczko, M., Simossis, V.A., Sethupathy, P., Hatzigeorgiou, A.G., 2009. The database of experimentally supported targets: a functional update of TarBase. Nucleic Acids Research 37, D155–158.

Pichiorri, F., Suh, S.-S., Ladetto, M., Kuehl, M., Palumbo, T., et al., 2008. MicroRNAs regulate critical genes associated with multiple myeloma pathogenesis. Proceeding of the National Academy Sciences USA 105, 12885–12890.

Pillai, R.S., Bhattacharyya, S.N., Filipowicz, W., 2007. Repression of protein synthesis by miRNAs: how many mechanisms? Trends Cell Biology 17, 118–126.

Raveche, E.S., Salerno, E., Scaglione, B.J., Manohar, V., Abbasi, F., et al., 2007. Abnormal microRNA-16 locus with synteny to human 13q14 linked to CLL in NZB mice. Blood 109, 5079–5086.

Ritchie, W., Flamant, S., Rasko, J.E.J., 2010. mimiRNA: a microRNA expression profiler and classification resource designed to identify functional correlations between microRNAs and their targets. Bioinformatics 26, 223–227.

Ritchie, W., Rajasekhar, M., Flamant, S., Rasko, J.E.J., 2009. Conserved expression patterns predict microRNA targets. PLoS Computational Biology 5, e1000513.

Ruby, J.G., Jan, C.H., Bartel, D.P., 2007. Intronic microRNA precursors that bypass Drosha processing. Nature 448, 83–86.

Tsang, W.P., Kwok, T.T., 2010. Epigallocatechin gallate up-regulation of miR-16 and induction of apoptosis in human cancer cells. Journal of Nutritional Biochemistry 21, 140–146.

Wang, H., Li, W.-H., 2009a. Increasing microRNA target prediction confidence by the relative R2 method. Journal Theoretical Biology 259, 793–798.

Wang, H., Wang, Y.H., Wu, W.S. Yeast cell cycle transcription factors identification by variable selection criteria. Gene, in press. doi:10.1016/j.gene.2011.06.001.

Wang, Y.-P., Li, K.-B., 2009b. Correlation of expression profiles between microRNAs and mRNA targets using NCI-60 data. BMC Genomics 10, 218.