GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Predicting the target genes of microRNA based on microarray data

B. Cao[1], T. Ji[1], B. Zhou[1], J. Zou[2] and G.Q. Jiao[2]

[1]Department of Gastrointestinal Surgery,
Taixing Hospital Affiliated in Yangzhou University, Yangzhou, China
[2]Department of Cardiothoracic Surgery,
Wuxi People's Hospital Affiliated in Nanjing Medical University, Wuxi, China

Corresponding author: G.Q. Jiao
E-mail: jiaoguoqingjgq@hotmail.com

Genet. Mol. Res. 12 (4): 6059-6066 (2013)
Received April 9, 2012
Accepted September 7, 2013
Published December 2, 2013
DOI http://dx.doi.org/10.4238/2013.December.2.4

ABSTRACT. MicroRNAs (miRNAs) are small non-coding RNAs of approximately 22 nucleotides in length, which play important roles in regulating gene expression post-transcriptionally. Several computational methods and algorithms have been developed to predict miRNA targets. In this study, we described a method that can be used to integrate miRNA target prediction data from multiple sources and gene expression data to predict target genes of particular miRNAs. We used hsa-miR-375 as an example to test the feasibility of our method. A total of 5645 target genes of hsa-miR-375 were identified from five prediction programs, and among them, 2440 target genes were shared by at least 2 of these 5 programs. By using our method, the number was further reduced to 149 and 5 of the 149 target genes had been validated by previous study. This is a simple yet highly effective approach.

Key words: MicroRNA; Target genes; Prediction program; Gene expression microarray

INTRODUCTION

MicroRNAs (miRNAs) are small non-coding RNAs of approximately 22 nucleotides in length that were first discovered in the early 1990s in *Caenorhabditis elegans* (Lee et al., 1993). miRNAs bind to partially complementary sites in the 3'-untranslated region (UTR) of mRNAs and regulate gene expression post-transcriptionally by translational attenuation/repression or cleavage of target mRNAs (Wightman et al., 1993; Bartel, 2004; Bagga et al., 2005). miRNAs are involved in various biological processes, such as cell differentiation, apoptosis, immune responses, and carcinogenesis (Calin et al., 2004; Xu et al., 2004; Cheng et al., 2005). According to the latest miRBase release, more than 1000 mature miRNAs have been identified in the human genome (Griffiths-Jones et al., 2008), and further discovery is driving for the increasing use of high-throughput sequencing.

Since miRNAs, play important roles of post-transcriptional regression by targeting mRNAs, identification of their putative target genes is an important step to study miRNAs. In the last few years, many computational methods and algorithms have been developed to predict miRNA targets, such as TargetScan, PicTar, miRanda, MirTif, and NBmiRTar (John et al., 2004; Krek et al., 2005; Yang et al., 2008; Garcia et al., 2011). Most of these methods or algorithms were based on the following criteria: 1) strong Watson-Crick base-pairing of the 5' seed of the miRNA to a complementary site in the 3'-UTR of the mRNA, 2) conservation of the miRNA binding site, and 3) a local miRNA-mRNA interaction with a positive balance of minimum free energy (Barbato et al., 2009). Due to the low complementary rate between miRNA and mRNA, each algorithm has a definite rate of both false-positive and false-negative predictions (Rajewsky, 2006). To make reliable predictions about a particular gene or a specific miRNA, more than one algorithm is used in common practice.

DNA microarray enables investigators to study the gene expression profile and gene activation of thousands of genes and sequences simultaneously. DNA microarrays have been widely used to profile gene expression changes between two conditions and most of these changes are likely to be due to differences in transcriptional activity (Madden et al., 2010). The effects of miRNA-directed mRNA degradation may be detectable through changes in the expression of miRNA target genes. However, when the number of genes under study is on the order of several hundreds or thousands, a gene-by-gene search of miRNA targets of interest becomes impractical.

In this study, we described a new method to predict reliable target genes that can be used to link gene expression data and miRNA target predictions from multiple programs. Our method is conducted in the following procedure (Figure 1): Initially, we predicted the target genes of a particular miRNA by integrating the predictions of five algorithms. Secondly, we analyzed the differentially expressed genes before and after transfection of this miRNA. Thirdly, we performed further screening of the target genes from the first step by using the microarray data. Finally, text mining of the predicted target genes in our method was performed to verify the feasibility of our method.

MATERIAL AND METHODS

Predict target genes of hsa-miR-375 by algorithms

Five different miRNA target prediction programs were used to retrieve the target

genes of hsa-miR-375, TargetScan (Garcia et al., 2011), miRanda (John et al., 2004), RNA22 (Miranda et al., 2006), miRDB (Wang, 2008), and PicTar5 (Krek et al., 2005). Each of these programs uses both miRNA sequences and 3'-UTR of protein-coding mRNA sequences as input files generally in FASTA format and determines their binding ability by calculating the minimum free energy for hybridization. The miRNA target prediction data were downloaded from the TargetScan website (http://www.targetscan.org/), miRBase for miRanda (http://microrna.sanger.ac.uk/sequences/), the RNA22 website (http://cbcsrv.watson.ibm.com/rna22.html), the miRDB website (http://mirdb.org/miRDB/) and from the UCSC genome browser tract for PicTar5 (http://genome.usuc.edu). After the prediction of putative target genes by each tool with the default parameters, we extracted those target genes shared by at least 2 of these 5 tools for further analysis.

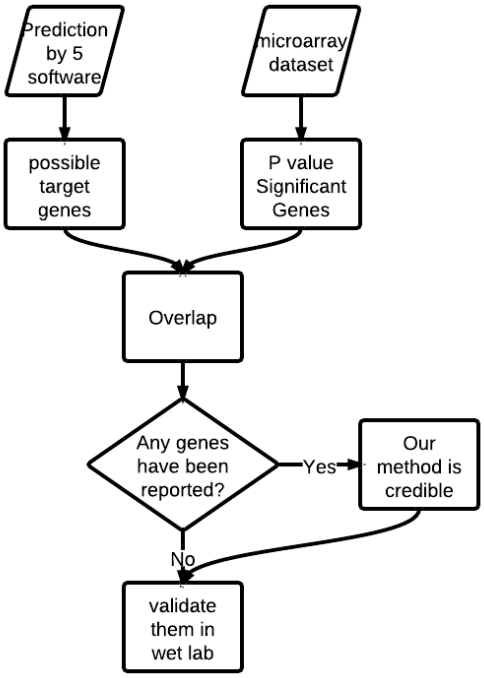

Figure 1. The workflow of our method to predict the target genes of a specific miRNA.

Predict target genes of hsa-miR-375 by gene expression data

We downloaded the gene expression profile from the study of Tsukamoto et al. (2010). In that study, they investigated expression profiles of miRNA in gastric carcinomas by use of a miRNA microarray platform covering a total of 470 human miRNAs. Further, they investigated the target genes of miR-375 by analyzing the gene expression profiles in cells transfected with precursor miR-375 (pre375) and negative control precursor miRNA#2 (preNeg). We extracted the differentially expressed genes with the cut-off criteria of fold-change ≤ 0.5 or fold-change ≥ 2 and P value <0.05.

RESULTS

Obtainment of target genes for hsa-miR-375 based on prediction algorithms

In this study, we chose hsa-miR-375, which has been implicated in a number of different cancers (de Souza Rocha et al., 2010; Ding et al., 2010) for applying our method. Five databases or algorithms were used to predict the target genes of hsa-miR-375. Different algorithms provide different predictions, and the degree of overlap between different lists of predicted targets is sometimes poor or null (Sethupathy et al., 2006). After the prediction of putative miRNA target genes by each tool with the default parameters, we obtained a total of 5645 target genes. Apparently, these predictions were noisy with a high false-positive and false-negative rate. To obtain reliable predictions, we extracted those target genes shared by at least 2 of these 5 tools and finally obtained a total of 2440 target genes for further analysis. Table 1 showed part of these target genes.

Table 1. The top 15 target genes of hsa-miR-375 predicted by 5 databases.

MicroRNA	Gene	miRanda	miRDB	PicTar	RNA22	TargetScan	Sum
hsa-miR-375	MTDH	1	1	0	1	1	4
hsa-miR-375	ELAVL4	1	1	1	0	1	4
hsa-miR-375	TRPS1	1	0	1	0	1	3
hsa-miR-375	SENP2	1	0	0	1	1	3
hsa-miR-375	SLC26A2	1	0	0	1	1	3
hsa-miR-375	SLC38A2	1	0	0	1	1	3
hsa-miR-375	RGS1	1	0	0	1	1	3
hsa-miR-375	PIAS1	1	0	1	0	1	3
hsa-miR-375	TXNDC10	1	0	0	1	1	3
hsa-miR-375	GRIN2B	1	0	0	1	1	3
hsa-miR-375	LIMD1	1	0	0	1	1	3
hsa-miR-375	VPS41	1	0	0	1	1	3
hsa-miR-375	TDRD6	1	0	0	1	1	3
hsa-miR-375	RASL12	1	0	0	1	1	3
hsa-miR-375	GLS	1	0	0	1	1	3

"1" represents the gene was predicted by the database and "0" represents the gene was not predicted by the database. "Sum" represents the predicted frequency of the 5 databases; its value area is 1-5.

Obtainment of target genes for hsa-miR-375 based on microarray data

We used the gene expression profile from a previous study and analyzed the differentially expressed genes between cells transfected with pre375 and preNeg. A total of 400 differentially expressed genes were identified with the cut-off criteria of fold-change ≤ 0.5 or fold-change ≥ 2 and P value <0.05. Table 2 listed the top 15 of the 400 differentially expressed genes. Expressions of these genes were significantly altered after transfected with pre375, suggesting that they may regulated by miR-375.

Combination of the results of prediction algorithm and gene expression microarray

Next, we integrated the target genes of the above two results and selected the overlapping genes. A total of 149 genes were identified to be overlapped (Table 3). These 149 genes may be the target genes of hsa-miR-375 with low false-positive or false-negative rate.

Table 2. The top 15 significant target genes of hsa-miR-375 predicted by gene expression microarray.

Gene symbol	Fold-change		P value	
	NUGU3	AZ521	NUGU3	AZ521
SUB1	0.06	0.13	1.59E-05	2.69E-07
TBC1D5	0.06	0.17	7.99E-07	1.23E-07
KLK10	0.07	0.20	6.22E-06	1.36E-05
KRT8	0.09	0.15	1.50E-06	6.51E-07
PSIP1	0.10	0.16	6.48E-06	1.84E-07
NPPB	0.11	0.05	7.26E-06	3.11E-05
NUPL1	0.12	0.11	1.05E-05	3.20E-07
YWHAZ	0.13	0.23	1.61E-05	6.18E-07
U69195	0.13	0.21	2.76E-06	1.76E-07
RBPJ	0.14	0.21	2.11E-06	4.02E-07
WDR51B	0.14	0.08	3.43E-05	1.65E-06
SETD7	0.14	0.21	1.16E-05	3.60E-06
ANKRD1	0.15	0.29	1.72E-06	3.32E-06
EBPL	0.16	0.28	4.23E-05	1.70E-06
WWC2	0.17	0.19	2.57E-06	9.74E-08

"NUGU3" and "AZ521" are two gastric carcinoma cell lines.

Table 3. Overlapping target genes of prediction algorithm and gene expression microarray.

ZFP91	TNS3	RPUSD4	OSBPL8	KIAA1191	ENAH	CNN3
ZFP36L2	TMEM55A	RPN1	OPA1	KIAA0152	EIF4G3	CLDN1
ZC3HAV1	TMEM106B	RLF	OGFOD1	KCTD10	EIF2S1	CKAP2
ZBTB8	TMED5	RBPJ	NXT2	KCNN4	EFNB2	CHSY1
YAP1	TIMM17A	RB1CC1	NT5C2	JAG1	DUSP6	CDR2
XPR1	TCF12	RAPGEF1	NOLC1	INSIG2	DPYSL3	CDKN2B
WWC2	SYNCRIP	RAP2C	NLE1	IL1RAP	DPY19L1	CDCA7L
WDR51B	STX6	RAB6A	NIN	IFT80	DNMT3B	CCDC32
VASN	STK38L	PTPMT1	NCOA7	HRSP12	DNAH11	CARD8
UXS1	SPRED1	PTP4A2	MYBL1	HN1	DIP2C	CALU
USP46	SPAG9	PSIP1	MTDH	HK2	DGKD	C3orf34
USP31	SLC7A6	PRKX	MRE11A	HEY1	DDEF2	C1QBP
USP1	SLC35B4	PRKCA	MPP5	HERPUD1	DCUN1D4	C1GALT1
UNC13B	SIPA1L3	PRDX1	MED8	GDAP1	CYP20A1	C15orf41
UBLCP1	SH3BP4	PLEKHA3	MBD2	GATAD1	CUL5	C14orf159
UBE3A	SET	PHLDA1	MAT2B	GATA6	CTGF	C11orf54
TXNDC10	SERTAD2	PDGFC	LIMD2	FZD4	CSTF2	BSN
TTC7B	SERP1	PAQR8	LHFPL2	FAM82C	CRIM1	BDH2
TSC22D2	SEC23A	PAPD4	LDHB	EXOC6	CORO1C	BCL10
TOR1AIP2	SAMD11	PAFAH1B1	KPNA4	ENOPH1	COG2	B3GALNT1
AZI2	APPL2	AMMECR1	ALMS1	AKAP7	AGPAT5	AEBP2
ACSL3	ABI2					

Validation of the predicted target genes

To validate the reliability of the 149 target genes, we searched the experimental verified target genes of hsa-miR-375 in publicly published papers. A total of 10 target genes of hsa-miR-375 that had been validated by RT-PCR or Western blot were screened (Table 4). Among the 10 validated genes, 5 genes were also identified by our methods: LDHB, C1QBP, USP1, MTDH, and YAP1 (Figure 2).

Table 4. Target genes predicted to be associated with hsa-miR-375.

Gene	Article	Author
JAK2	MiR-375 frequently downregulated in gastric cancer inhibits cell proliferation by targeting JAK2.	Ding et al., 2010
MTPN	Combinatorial microRNA target predictions.	Krek et al., 2005
LDHB	Tumor suppressive microRNA-375 regulates lactate dehydrogenase B in maxillary sinus squamous cell carcinoma.	Kinoshita et al., 2012
C1QBP	Combinatorial microRNA target predictions.	Krek et al., 2005
USP1	Combinatorial microRNA target predictions.	Krek et al., 2005
MTDH	Significance of dysregulated metadherin and microRNA-375 in head and neck cancer.	Hui et al., 2011
YAP1	miR-375 is activated by ASH1 and inhibits YAP1 in a lineage-dependent manner in lung cancer.	Nishikawa et al., 2011
ELAVL4	miR-375 inhibits differentiation of neurites by lowering HuD levels.	Abdelmohsen et al., 2010
RASD1	Epigenetically deregulated microRNA-375 is involved in a positive feedback loop with estrogen receptor alpha in breast cancer cells.	de Souza Rocha et al., 2010
TIMM8A	A distinct microRNA signature for definitive endoderm derived from human embryonic stem cells.	Hinton et al., 2010

The target genes that are identified by the miRNA microarray are highlighted in bold.

Figure 2. VENN graph displays our results. A total of 149 target genes of hsa-miR-145 were predicted by our methods and 5 of them had been validated by previous studies.

DISCUSSION

In this study, we describe a method for inferring the target genes of a particular miRNA by integrating the information provided by miRNA target prediction programs with mRNA gene expression data. This is a simple yet highly effective approach. By using this method, the number of target genes of hsa-miR-375 was reduced from 2440 to 149. Among the 149 predicted target genes, 5 had been validated by previous study.

Previous studies had predicted that up to 30% of mammalian genes are regulated by miRNAs (Lewis et al., 2003; Xie et al., 2005), and many regulatory patterns are likely to be regulated by miRNAs (Li et al., 2008). However, due to the increasing use of high-throughput sequencing, the number of target genes of a particular miRNA is increasing as well, rendering harder the search of miRNA targets of interest (Barbato et al., 2009). Our method enables to efficiently combine gene expression studies (mRNA profiles) with miRNAs expression profiles. Therefore, the results are more credible, although a disadvantage of this method is that it relies on the gene expression data of study.

However, the gene expression may be affected by many factors, such as tissue specific expression and 3'-UTR splice variants. Nevertheless, we have shown that the current method provides high-efficient results, better than traditional predictions. The method was able to predict miRNAs, which had been experimentally verified as showing significant differential expression.

In conclusion, we describe a method that: 1) can be used to integrate miRNA target prediction data from multiple sources and gene expression data to predict target genes of particular miRNAs, and 2) can increase either the efficiency and reliability of predicted target genes.

REFERENCES

Abdelmohsen K, Hutchison ER, Lee EK, Kuwano Y, et al. (2010). miR-375 inhibits differentiation of neurites by lowering HuD levels. *Mol. Cell Biol.* 30: 4197-4210.

Bagga S, Bracht J, Hunter S, Massirer K, et al. (2005). Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell* 122: 553-563.

Barbato C, Arisi I, Frizzo ME, Brandi R, et al. (2009). Computational challenges in miRNA target predictions: to be or not to be a true target? *J. Biomed. Biotechnol.* 2009: 803069.

Bartel DP (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116: 281-297.

Calin GA, Sevignani C, Dumitru CD, Hyslop T, et al. (2004). Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc. Natl. Acad. Sci. U. S. A.* 101: 2999-3004.

Cheng AM, Byrom MW, Shelton J and Ford LP (2005). Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis. *Nucleic Acids Res.* 33: 1290-1297.

de Souza Rocha SP, Breiling A, Gupta N, Malekpour M, et al. (2010). Epigenetically deregulated microRNA-375 is involved in a positive feedback loop with estrogen receptor alpha in breast cancer cells. *Cancer Res.* 70: 9175-9184.

Ding L, Xu Y, Zhang W, Deng Y, et al. (2010). MiR-375 frequently downregulated in gastric cancer inhibits cell proliferation by targeting JAK2. *Cell Res.* 20: 784-793.

Garcia DM, Baek D, Shin C, Bell GW, et al. (2011). Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. *Nat. Struct. Mol. Biol.* 18: 1139-1146.

Griffiths-Jones S, Saini HK, van Dongen S and Enright AJ (2008). miRBase: tools for microRNA genomics. *Nucleic Acids Res.* 36: D154-D158.

Hinton A, Afrikanova I, Wilson M, King CC, et al. (2010). A distinct microRNA signature for definitive endoderm derived from human embryonic stem cells. *Stem. Cells Dev.* 19: 797-807.

Hui AB, Bruce JP, Alajez NM, Shi W, et al. (2011). Significance of dysregulated metadherin and microRNA-375 in head and neck cancer. *Clin. Cancer Res.* 17: 7539-7550.

John B, Enright AJ, Aravin A, Tuschl T, et al. (2004). Human MicroRNA targets. *PLoS Biol.* 2: e363.

Kinoshita T, Nohata N, Yoshino H, Hanazawa T, et al. (2012). Tumor suppressive microRNA-375 regulates lactate dehydrogenase B in maxillary sinus squamous cell carcinoma. *Int. J. Oncol.* 40: 185-193.

Krek A, Grun D, Poy MN, Wolf R, et al. (2005). Combinatorial microRNA target predictions. *Nat. Genet.* 37: 495-500.

Lee RC, Feinbaum RL and Ambros V (1993). The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75: 843-854.

Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, et al. (2003). Prediction of mammalian microRNA targets. *Cell* 115: 787-798.

Li J, Musso G and Zhang Z (2008). Preferential regulation of duplicated genes by microRNAs in mammals. *Genome Biol.* 9: R132.

Madden SF, Carpenter SB, Jeffery IB, Bjorkbacka H, et al. (2010). Detecting microRNA activity from gene expression data. *BMC Bioinformatics* 11: 257.

Miranda KC, Huynh T, Tay Y, Ang YS, et al. (2006). A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* 126: 1203-1217.

Nishikawa E, Osada H, Okazaki Y, Arima C, et al. (2011). miR-375 is activated by ASH1 and inhibits YAP1 in a lineage-dependent manner in lung cancer. *Cancer Res.* 71: 6165-6173.

Rajewsky N (2006). microRNA target predictions in animals. *Nat. Genet.* (Suppl 38): S8-13.

Sethupathy P, Megraw M and Hatzigeorgiou AG (2006). A guide through present computational approaches for the identification of mammalian microRNA targets. *Nat. Methods* 3: 881-886.

Tsukamoto Y, Nakada C, Noguchi T, Tanigawa M, et al. (2010). MicroRNA-375 is downregulated in gastric carcinomas and regulates cell survival by targeting PDK1 and 14-3-3zeta. *Cancer Res.* 70: 2339-2349.

Wang X (2008). miRDB: a microRNA target prediction and functional annotation database with a wiki interface. *RNA* 14: 1012-1017.

Wightman B, Ha I and Ruvkun G (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell* 75: 855-862.

Xie X, Lu J, Kulbokas EJ, Golub TR, et al. (2005). Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* 434: 338-345.

Xu P, Guo M and Hay BA (2004). MicroRNAs and the regulation of cell death. *Trends Genet.* 20: 617-624.

Yang Y, Wang YP and Li KB (2008). MiRTif: a support vector machine-based microRNA target interaction filter. *BMC Bioinformatics* 9 (Suppl 12): S4.