World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:3, No:4, 2009

A study of Cancer-related MicroRNAs through Expression Data and Literature Search

Chien-Hung Huang[1], Chia-Wei Weng[2], Chang-Chih Chiang[2], Shih-Hua Wu[1], Chih-Hsien Huang[1], Ka-Lok Ng[2*]

Open Science Index, Computer and Information Engineering Vol:3, No:4, 2009 publications.waset.org/3278/pdf

***Abstract*—**MicroRNAs (miRNAs) are a class of non-coding RNAs that hybridize to mRNAs and induce either translation repression or mRNA cleavage. Recently, it has been reported that miRNAs could possibly play an important role in human diseases. By integrating miRNA target genes, cancer genes, miRNA and mRNA expression profiles information, a database is developed to link miRNAs to cancer target genes. The database provides experimentally verified human miRNA target genes information, including oncogenes and tumor suppressor genes. In addition, fragile sites information for miRNAs, and the strength of the correlation of miRNA and its target mRNA expression level for nine tissue types are computed, which serve as an indicator for suggesting miRNAs could play a role in human cancer. The database is freely accessible at http://ppi.bioinfo.asia.edu.tw/mirna_target/index.html.

***Keywords*—**MicroRNA, miRNA expression profile, mRNA expression profile, cancer genes, oncogene, tumor suppressor gene

I. Introduction

MICRORNAS (miRNAs) are a class of small non-coding RNAs. Experimental studies suggested that miRNA could induce gene silencing [1], [2], [3]. Gene silencing may be due to either mRNA repression or degradation. Several recent studies indicated that miRNA could regulate mRNA expression on a very large scale [4], [5]. It has been reported [6], [7], [8] that miRNAs could possibly involved in human diseases, and it could play the role of an oncogene (OCG) and tumor suppressor gene (TSG).

By integrating heterogeneous types of data, that is miRNA targets (TarBase [9]), miRNA and mRNA expression data (NCI-60 [10], [11]), cancer genes (tumor-associated gene database (TAG) [12]), and literatures data (PubMed), to investigate the regulatory relationship between miRNAs and cancer genes.

In the materials and procedures section, we gave a description of the input data and the methods used in this paper. In the result section, results for the correlation coefficient of miRNA and mRNA expression profile for nine tissues (three different chips and three different normalization methods) are reported. The final section is the conclusion section.

[1] Chien-Hung Huang, Shih-Hua. Wu, and Chih-Hsien Huang, is with the Department of Computer Science and Information Engineering, National Formosa University, Taiwan 632.

[2] Chia-Wei Weng, Chang-Chih Chiang and Ka-Lok Ng, is with the Department of Bioinformatics, Asia University, Taiwan 41354.

* Ka-Lok Ng is the corresponding author (phone: 886-423323456 ext. 1856; fax: 886-23321019;, e-mail: ppiddi@asia.edu.tw).

II. Materials and Procedures

A. Input datasets

TarBase is a manually curated collection of experimentally tested miRNAs, which is owned by Alexander Fleming Biomedical Science Research Center. The database is functionally linked to several other useful databases such as Gene Ontology (GO) and UCSC Genome Browser. It can be accessed at http://diana.cslab.ece.ntua.gr/tarbase/. The present work employed the TarBase v4 dataset.

MiRBase [13] is a database which collect miRNAs' precusors, mature miRNAs, FASTA sequences, and their target genes information. The database can be accessed at http://microrna.sanger.ac.uk/sequences/.

The NCI-60 is a set of 60 human cancer cell lines derived from diverse tissues. These cell lines include breast cancer, central neural system cancer, colon cancer, leukemia, melanoma, non-small cell lung cancer, ovarian cancer, prostate cancer, and renal cancer. The database can be accessed at http://discover.nci.nih.gov/cellminer/. Four publicly available datasets of gene expression profiles are selected in this study, including the miRNA expression, Affymetrix® U95(A-E), U133A, and U133B RNA expression datasets. Each of the RNA expression dataset uses three types of normalization methods, including GCRMA, MAS5 and RMA. Therefore, a total of ten datasets, including one miRNA dataset and nine RNA expression datasets, are used in this study.

The TAG database compiled a list of OCG and TSG through text-mining approach from the PubMed database. At current stage, TAG database contains information for 198 OCG, 167 TSG and 151 genes related to oncogenesis. Each of the TAG record was known to be involved in at least one cancer type. The database can be accessed at http://www.binfo.ncku.edu.tw/TAG/GeneDoc.php.

MiRNA genes are frequently located at fragile sites. Fragile sites information for miRNAs are obtained from Calin et al., work [14], and Debacker & Kooy work [15].

B. Procedures

In the first step, human miRNA entries are selected based on the goal of this study and 648 miRNA-target gene pairs are left. TarBase miRNA IDs are mapped to miRBase pre-miRNA IDs

Open Science Index, Computer and Information Engineering Vol:3, No:4, 2009 publications.waset.org/3278/pdf

through a BLAST alignment procedure.

In the second step, NCI-60 miRNA expression data IDs are converted to miRBase pre-miRNA IDs. After the processing, 635 entries of NCI-60 miRNA expression data are recorded. We also checked mRNA's gene IDs in NCI-60 Affymetrix RNA expression datasets and converted these gene IDs to NCBI official symbols by using Gene Name Service [16]. We obtained 44855 entries for U95(A-E) dataset, 20169 entries for U133A dataset, and 16441 entries for U133B dataset after preprocessing the data.

In the last step, the relationship between miRNA and it's target gene is quantified by computing the Pearson correlation coefficient (PCC), which is given by,

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (1)$$

where x_i is the expression intensity of miRNA; y_i is the expression intensity of miRNA's target gene; $\bar{x}$ and $\bar{y}$ is the mean expression intensity of the miRNA and its target gene respectively; and n is the total number of entries of expression data.

III. Results

The miRNAs' identifiers in the TarBase and NCI-60 expression datasets are checked in order to standardize with miRBase's miRNA IDs and NCBI's official gene symbol. After standardization the TarBase data is reduced from 648 entries to 420 entries. This is because some of the mature miRNA IDs documented by TarBase do not match with miRBase's mature miRNA IDs. The number of entries in NCI-60 datasets is also reduced, this is a result of the second step reported in the procedures section.

PCC values of miRNA expression level and their target mRNA expression level for nine tissue types are computed. A web site is set up for providing human miRNA target genes information, the PCC of miRNA and its target gene expression level for nine tissue types. For example, user can input a TarBase miRNA ID, such as hsa-miR-16, the platform returns two target genes, BCL2 and CCNT2. Average of the PCC values of hsa-miR-16 and BCL2 in leukemia, using the U95(A-E), U133A and U133B chips with the GCRMA, MAS5 and RMA is given in Table 1.

TABLE 1 AVERAGE PCC VALUES OF HSA-MIR-16 AND BCL2 FOR THE U95(A_E), U133A AND U133B CHIPS FOR LEUKEMIA

Dataset	Normalization	PCC
U95(A_E)	GCRMA	-0.694
	MAS5	-0.632
	RMA	-0.609
U133A	GCRMA	-0.762
	MAS5	-0.797
	RMA	-0.913
U133B	GCRMA	-0.658
	MAS5	-0.707
	RMA	-0.640

It is note that the PCC average values are smaller than -0.6 in all cases, which suggested that hsa-miR-16 could possible play a role in regulating the BCL2 gene in leukemia [17] and gastric [18].

The expression level of correlation between miRNA and it's target gene in cancer cell lines could be understood as the following. It is known that the biological processing of miRNA is to repress and/or cleavage the mRNA by incomplete or complete complementary binding with the mRNA. If miRNA and it's experimentally target gene which are retrieved in TarBase dataset are direct interaction, our computation result of PCC should reveal significant negative correlation. Table 1 results suggested that hsa-miR-16 could possibly play a role in regulating the cancer gene BCL2.

Fig. 1 shows the PCC results of BCL2 targeted by hsa-miR-16.

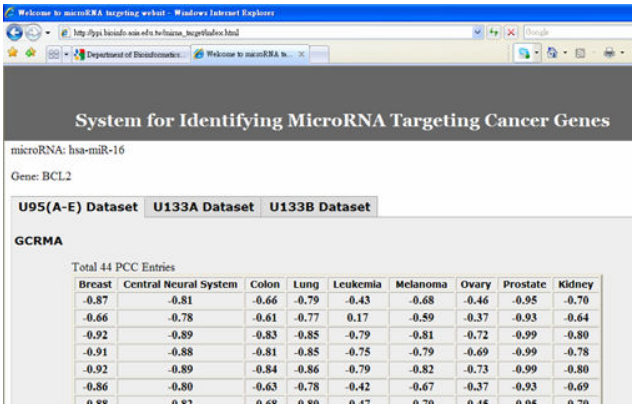


Fig. 1 PCC values of hsa-miR-16 and one of its target gene, BCL2 for nine tissue types, using the U95(A-E) chip with the GCRMA normalization method.

Through the literature search and text mining a total of 76 cancer related miRNAs and mRNAs are retrieved. Part of the results is given in Table 2. Table 2 lists 10 records of the miRNAs, their target genes (using gene symbol annotation), and the cancer types.

TABLE II CANCER RELATED MIRNAS AND MRNA

miRNA	Target gene	Cancer
hsa-let-7c	MYC	Burkiit lymphoma
miR-1	MET	lung cancer
miR-1	PIM1	lung cancer
miR-1	HDAC4	lung cancer
miR-1	FOXP1	lung cancer
miR-106a	E2F1	colon cancer
miR-106b	E2F1	gastric cancer
miR-125b	BAK1	prostate cancer

miR-125b	EIF4EBP1	prostate cancer
miR-126*	EGFL7	prostate cancer

These information would be very useful for investigating the regulatory role of miRNAs in cancer causing.

IV. Conclusion

It is an important issue to know whether miRNA could possibly play the role of an OCG and TSG. A platform is set up for providing experimentally verified human miRNA targeting genes, including OCG and TSG information. In addition, fragile sites information for miRNA, and the strength of the correlation of miRNA and its targeting mRNA expression level for nine tissue types are computed. These two pieces of information are very useful for identifying cancer-related miRNAs. We are in the process of enriching more information in the database, such as the Gene Ontology, and DNA methylation datasets, hence it would supply more and stronger evidence for the present study. In summary, the database provides, and an easy means of investigating the regulatory role for cancer study.

Acknowledgment

Drs. Chien-Hung Huang and Ka-Lok Ng works are supported by the National Science Council of R.O.C. under the grants of NSC 97-2221-E-150-063 and NSC 97-2221-E-468-010 respectively.

References

[1] Eric A Miska, How microRNAs control cell division, differentiation and death. *Current Opinion in Genetics & Development*. 15(5), 563-568, 2005.

[2] V. Narry.Kim, MicroRNA biogenesis: coordinated cropping and dicing. *Nature Reviews. Molecular Cell Biology* 6(5), 376-385, 2005.

[3] Erno Wienholds, and Ronald H.A. Plasterk,. MicroRNA function in animal development. *FEBS Letters,* 579(26), 5911-5922, 2005.

[4] L.P. Lim, N.C. Lau, P. Garrett-Engele, A.Grimson, J.M Schelter., J. Castle, D.P. Bartel, P.S. Linsley, J.M. Johnson. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, 433, 769-73, 2005.

[5] L.F. Sempere, S. Freemantle, I. Pitha-Rowe, E. Moss, E. Dmitrovsky, V. Ambros. Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. *Genome Biol*., 5, R13. 2004.

[6] L. He, J.M. Thomson, M.T. Hemann, E Hernando-Monge., D Mu., S. Goodson, S. Powers, C. Cordon-Cardo, S.W. Lowe, G.J. Hannon, S.M. Hammond. A microRNA polycistron as a potential human oncogene. *Nature,* 435, 828-33, 2005.

[7] A. Esquela-Kerscher & F.J. Slack Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer* 2006, 6(4):259-269, 2006.

[8] B Zhang, X Pan, GP Cobb, TA Anderson.. microRNAs as oncogenes and tumor suppressors. *Dev Biol.* 302(1):1-12, 2007.

[9] P. Sethupathy, B. Corda and A.G Hatzigeorgiou., TarBase: A comprehensive database of experimentally supported animal miRNA targets, *RNA*, Vol. 12, pp.192-197, 2005.

[10] Shankavaram et al., Transcript and protein expression profiles of the NCI-60 cancer cell panel: an integromic microarray study, *Molecular Cancer Therapeutics*, Vol. 6, No. 3, pp.820-832, 2007

[11] Blower et. al., MiRNA expression profiles for the NCI-60 cancer cell panel, *Molecular Cancer Therapeutics*, Vol. 6, No. 5, pp.1483-1491, 2007.

[12] Hsiang-Han.Chan and H. Sunny Sun, Identification of novel tumor-associated gene (TAG) by bioinformatics analysis. Master Thesis, Institute of Molecular Medicine, National Cheng Kung University, 2005.

[13] S Griffiths-Jones., Russell J. Grocock, Stijn van Dongen, Alex Bateman, Anton J.Enright, miRBase: miRNA sequences, targets and gene nomenclature, *Nucleic Acids Research*, Vol. 34, pp.140-144, 2006.

[14] Calin et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci* U S A, 101(9):2999-3004, 2004.

[15] Debacker Kim and R.Frank Kooy. Fragile sites and human disease. *Human Molecular Genetics*, 16(2), R150-158, 2007.

[16] Kuan-Ting Lin, Chia-Hung Liu, Jen-Jie Chiou, Wen-Hsien Tseng, Kuang-Lung Lin, Chun-Nan Hsu, Gene Name Service: No-Nonsense Alias Resolution Service for Homo Sapiens Genes, *Proceedings of the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology-Workshops*, pp.185-188, 2007.

[17] A Cimmino et al.,. miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc Natl Acad Sci.*, 102(39):13944-9, 2005. Erratum in: *Proc Natl Acad Sci* 14;103(7):2464, 2006..

[18] L Xia, D Zhang, R Du, Y Pan, L Zhao, S Sun, L Hong, J Liu, D Fan. miR-15b and miR-16 modulate multidrug resistance by targeting BCL2 in human gastric cancer cells. *Int J Cancer*. 123(2):372-9, 2008.

Open Science Index, Computer and Information Engineering Vol:3, No:4, 2009 publications.waset.org/3278/pdf