

Integrated Analysis of miRNA and Gene Networks on Cancer Expression Data

Yuping LI

College of Computer Science and Technology, Key
Laboratory of Symbolic Computation and Knowledge
Engineering of Ministry of Education
Jilin University
Changchun, China
e-mail: lypjlu@163.com

Wei DU*

College of Computer Science and Technology, Key
Laboratory of Symbolic Computation and Knowledge
Engineering of Ministry of Education
Jilin University
Changchun, China
e-mail: weidu@jlu.edu.cn

Lihua HE

College of Computer Science and Technology, Key
Laboratory of Symbolic Computation and Knowledge
Engineering of Ministry of Education
Jilin University
Changchun, China
e-mail: helh522@163.com

Zhongbo CAO

College of Computer Science and Technology, Key
Laboratory of Symbolic Computation and Knowledge
Engineering of Ministry of Education
Jilin University
Changchun, China
e-mail: zbcao66@163.com

Yanchun LIANG*

College of Computer Science and Technology, Key Laboratory of Symbolic Computation and Knowledge Engineering of
Ministry of Education, Jilin University
Changchun, China
Zhuhai Laboratory of Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education,
Zhuhai College of Jilin University
Zhuhai, China
e-mail: ycliang@jlu.edu.cn

Abstract—MicroRNAs (miRNAs) are one type of small non-coding RNAs and play important roles in regulating wide range of biological processes, such as stem cell maintenance, tissue development and cell metabolism. By modulating these key cellular processes, miRNAs can simultaneously regulate both oncogenes and tumor suppressor genes in cancer. With the development of high-throughput RNASeq technology, aberrant miRNA profiles and gene profiles have been detected in many cancers. And numerous studies indicate that vast miRNAs are involved in tumorigenesis and development of many cancers. They may act as either oncogenes or tumor suppressor genes in cancer and can be used as potential biomarkers for diagnosis, therapy and prognosis. To explore carcinogenic mechanism, it is critical to discover the aberrantly expressed miRNAs and their target genes. In this article, we propose an improved multiply-step approach to identify the target relationship between miRNA and gene pairs. An improved minimum redundancy maximum relevance (mRMR) feature selection method is used to get the differentially expressed miRNAs and genes. Then, we determine the significantly negative correlation between miRNA and gene pairs based on the expression level of selected miRNAs and genes in tumor and their corresponding adjacent normal tissues. To reduce the false positive rate of our approach, we check the relationship between miRNAs and their target genes by multiple miRNA target prediction

databases.

Keywords—MicroRNAs; Target genes; miRNA-target gene networks; Feature selection

I. INTRODUCTION

MicroRNAs (miRNAs) are one class of small non-coding RNAs which present in a variety of eukaryotic organisms. They play several important roles in regulating wide range of biological processes, such as stem cell maintenance, development, metabolism, proliferation, differentiation and apoptosis [1]. By modulating these key cellular processes, miRNAs can simultaneously regulate both oncogenes and tumor suppressor genes in cancer [2, 3]. MiRNAs regulate target gene expression by two ways: mRNA degradation or translation inhibition, through binding to complementary sites of target mRNAs in 3'-UTR [4]. And an individual miRNA may potentially bind and regulate many target genes with related function, such as those belonging to a cell signaling pathway [5], whilst a given target gene may be regulated by several different miRNAs [6]. It is estimated that ~30% of human coding genes are regulated by miRNAs [7].

With the development of high-throughput RNASeq

technology, the expression levels for thousands and even tens of thousands of miRNAs or genes can be measured simultaneously now. A number of projects have applied this technology to study the differences between diseased and normal tissues and disease [8]. MicroRNA (miRNA) encoded by miR genes, plays a crucial role in regulating gene expression of a considerable part of plant's and animal's genome. And numerous studies indicate that miRNAs are involved of tumorigenesis and development in many cancers, act as either oncogenes or tumor suppressors. For examples, miR-155 overexpress in many cancer types including hematopoietic cancers, breast, lung and colon cancer [9]. Overexpressed miRNAs in cancers, such as mir-17-92 cluster, may function as oncogenes and promote cancer development by negatively regulating tumor suppressor genes, which control cell differentiation or apoptosis [10]. Underexpressed miRNAs in cancers, such as let-7, function as tumor suppressor genes and may inhibit cancers by regulating oncogenes that control cell differentiation or apoptosis [10]. miRNA expression profiles may become useful biomarkers for cancer diagnostics [11]. To explore carcinogenic mechanism, it is critical to discover the aberrantly expressed miRNAs and their target genes.

There are many databases explored to predict the targets of miRNAs either by seed sequence complementary or thermodynamic algorithm [12-14]. However, these databases only provide the possibility of predicting direct target relationship between miRNAs and genes, and they cannot predict disease-specific pairs in specific diseases. And in recent years, more and more researchers pay attention to the effect of paired samples. The experiments which contain cancer tissue samples and their corresponding control tissues were designed by many researchers to make the results of expression experiments to be much better generalized to clinical research. In paired datasets, each cancer sample has one corresponding control sample from the same patient as paired data. It is natural to consider using of paired measurements originated from the same test object due to the heterogeneous nature of the cancer genetics [15]. There are also some studies focus on the essential improvement in the identification of differentially expressed genes when paired designs were used to compare independent sample designs [16, 17]. So in this article, the expression profiles of both miRNA and gene in the same cohort of cases of tumor and their corresponding adjacent normal tissues are considered. There is a multiply-step approach for the identification of miRNA-gene regulatory modules introduced by constructing the paired miRNA-gene expression profiles and predicting the putative target genes of miRNAs [18, 19]. And Jihong Fu et al. applied this multiply-step approach in clinical matched tumor and normal mucosa samples to identify the colorectal cancer (CRC) specific miRNAs and their target genes [20]. The algorithm use Student's t-test and Significant Analysis of MicroArray (SAM) to identify aberrantly expressed miRNAs and genes. And then do remain multiply steps on the selected miRNAs and genes. The method used by Jihong Fu et al. is sensitive to small perturbations of the data and the results have unstable signatures. So we need to use a more stability feature selection method to select the differentially

expressed miRNAs and genes, and do the remained multiply steps on these selected miRNAs and genes to get miRNA-target gene networks.

In this article, we use a more stability feature selection method called ensemble minimum redundancy maximum relevance by paired random selection (emRMR_prs) to identify significantly dysregulated miRNAs and genes, and do the remain multiply steps on these selected miRNAs and genes to get miRNA-target gene networks. Firstly, we use emRMRprs method to select the differentially expressed miRNAs and genes. Secondly, since miRNAs act as negative regulators, up-regulated miRNAs resulted in down-regulated target genes, and vice versa. Pearson's correlation analysis was applied to calculate the significantly negative correlation between miRNA and gene pairs, based on the expression level of selected miRNAs and genes which are aberrantly expressed in tumor and their corresponding adjacent normal tissues. Thirdly, only the target relationship miRNA-gene pairs, which were also predicted by one or more miRNA target prediction databases, are retained in final miRNA-target gene networks. In addition, a relevant web tool for the analysis of the aberrantly miRNAs and genes, and the miRNA-target gene networks construction based on miRNA and gene expression data has been proposed.

II. METHODS

As shown in Fig.1, a novel multiply-step approach was adopted to identify the abnormally expressed miRNA-gene pairs in cancers. Firstly, the significantly differentially expressed miRNAs and genes were selected by an improved minimum redundancy maximum relevance (mRMR) feature selection method [21]. Secondly, Pearson Correlation Coefficient is used to calculate the relationship between miRNAs and genes, and we can get the significantly negative correlation between miRNA and gene pairs. Thirdly, only the

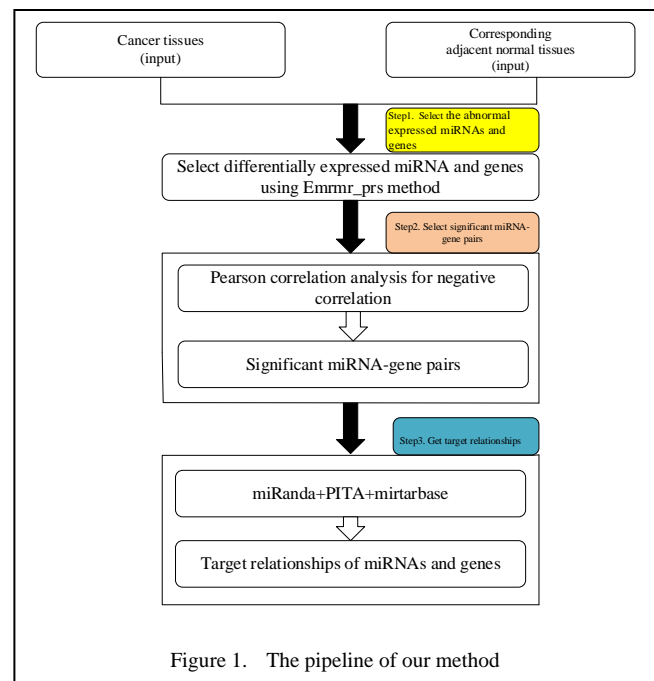


Figure 1. The pipeline of our method

target relationship miRNA-gene pairs, which were also predicted by databases mirtarbase [22], PITA [23] or miRanda [14], are retained.

A. Datasets

The Cancer Genome Atlas (TCGA) Data Portal provides a platform for researchers to search, download and analyze data sets [24]. It contains clinical information, genomic characterization data, and high level sequence analysis of the tumor genomes. There are various available cancer types in the database, and each type of cancer includes different number of cases. We select four common cancers, including Kidney renal clear cell carcinoma (KIRC), Liver hepatocellular carcinoma (LIHC), Prostate adenocarcinoma (PRAD) and Lung adenocarcinoma (LUAD). The details of the used datasets are shown in TABLE I.

B. Differential Expression Analysis

The abnormal expressed miRNAs and genes are focused and the discovery of the target-relationship of these differentially expressed miRNAs and genes are meaningful. We can select the different expressed miRNAs and genes from the original expression data by using the feature selection methods.

In the past few years, many feature selection methods had been proposed. However, most feature selection methods simply grade the features according to some rules and rank the features based on their scores, and then choose the top-ranked features. However, the selected features may be highly correlated, and the redundancy of the combined feature set will reduce the broadness of the feature set. So, a filter feature selection framework which contains the step of reducing redundancy was proposed by using minimum redundancy-maximum relevance (mRMR) [25, 26]. mRMR approach requires the features to be maximally dissimilar to each other in order to expand the representative power of the feature set. And it has been proved effective and frequently used to analyze the importance of different features [27, 28]. However, most of expression datasets only have few samples with high dimensions of genes. It is noticed that the feature selection results are influenced much by the expression data of samples, they may vary with even a single-case difference in the expression data set, and the outliers may cause the undesirable feature selection results [29]. In order to increase the stability of mRMR approach, an ensemble feature selection method based on mRMR was proposed to identify signatures in paired expression data [21]. Firstly, in order to increase the stability of the original method, the improved method uses an ensemble strategy to generate diverse subsets from the original dataset. Then, the mRMR method is used to obtain multiple feature lists on different subsets. Finally, a rank aggregation strategy is adopted to decide the final list of the selected features.

Before establishing the relationships between miRNAs and target genes, we use fold-change analysis on differentially expressed miRNAs and genes. We process each selected miRNA and gene expression data from each paired of tumor and normal samples by using fold-change, and calculate the average of the fold-change values of

miRNAs or genes. If the average fold-change value is higher than FC, the miRNA or gene is provided to be high expression, and if the average fold-change value is lower than 1/FC, it will be regarded as low expression, where FC is set as 1.5.

C. Establishing the Relationships between Mirnas and Target Genes

Since miRNAs act as negative regulators, up-regulated miRNAs resulted in down-regulated target genes, and vice versa. In order to measure these relationships accurately, Pearson Correlation Coefficient is used to calculate each relationship. The formula of Pearson Correlation Coefficient is defined as in (1):

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

Where x_i and y_i are the expression values of sample i in miRNA x and gene y . \bar{x} and \bar{y} are the average of expression values in miRNA x and gene y respectively, and n is the number of samples. If the correlation coefficient of one miRNA and one gene is less than -0.4, we consider there is a regulatory relationship between the miRNA and the gene.

To eliminate false positive rates of the target prediction databases, only the miRNA-gene pairs simultaneously predicted by mirtarbase and/or PITA and/or miRanda were considered as target relationship miRNA-mRNA pairs in our study.

D. Developing a Web Tool

A web tool with the novel multiply-steps approach for the analysis of the target predictions based on miRNA and gene expression data has been developed. There are three parts in this web tool. In the first part, the query section allows users search the results of miRNA target predictions based on different databases (PITA, miRanda and miRtarbase). In the second part, the feature selection section enable users get selected features by Emrmr_prs method, and users can download the result. In the last part, the analysis section contains a multistep procedure: (i) select a feature selection method to get the differently expressed features; (ii) use Pearson Correlation Coefficient or Mutual Information method and a selected threshold to measure the relationships of genes and miRNAs; (iii) a boolean combination of the target prediction databases are chosen, and only the target-relationship miRNA-gene pairs which are predicted by the combination of the databases are remained to build network.

III. RESULT

We applied this new multiply-step method on four cancer datasets, including KIRC (Kidney renal clear cell carcinoma) datasets, LIHC (Liver hepatocellular carcinoma) datasets, PRAD (Prostate adenocarcinoma) datasets and LUAD (Lung

adenocarcinoma) datasets. The details of these used datasets are shown in TABLE I.

TABLE I. DETAILS OF SELECTED CANCER TYPES

Name of dataset	Number of genes	Number of miRNAs	Number of cancer samples	Number of normal samples
KIRC	20502	923	22	22
LIHC	20502	959	50	50
PRAD	20502	919	53	53
LUAD	20502	948	50	50

TABLE II. THE SELECTED miRNAs AND GENES IN DIFFERENT DATABASES

Datasets Name	Number of up-regulation miRNAs	Number of down-regulation miRNAs	Number of up-regulation genes	Number of down-regulation genes
KIRC	80	71	489	504
LIHC	35	82	426	469
PRAD	115	2	352	431
LUAD	108	33	420	547

TABLE III. THE RESULTS OF CORRELATION BETWEEN DIFFERENTIAL EXPRESSION GENES AND miRNAs

Datasets Name	Number of correlated miRNAs	Number of correlated genes
LUAD	7	142
KIRC	15	131
LIHC	25	147
PRAD	13	105

A. Differential Expression Genes and miRNAs

As shown in Fig.1, the first step of the novel multiply-step approach aims to select the significantly differentially expressed miRNAs and genes. In this part, we carry out the process on the four datasets shown in TABLE I. The maximum number of selected miRNAs is set to 200 and the number of selected genes is set to 1000. According to the process of differential expression analysis, we can obtain the results of differential expression genes and miRNAs. TABLE II gives the results of differentially expressed miRNAs and genes.

B. Correlation between Differential Expression Genes and miRNAs

In this part, we apply the method mentioned in Section C of Method part on four cancer datasets and the results are shown in TABLE III. From TABLE III, we can see that the correlations between differential expression genes and miRNAs can be selected by applying our method on the datasets. In the results, we also can find one single miRNA has hundreds of target genes, and several approaches also show that miRNAs can have many targets [30].

Lung adenocarcinoma (LUAD) is a common histological form of lung cancer, and it is the most common type of lung

cancer. We set the Pearson coefficient threshold to -0.7, and the results of correlation between differential expression genes and miRNAs of LUAD are shown in Fig.2. The

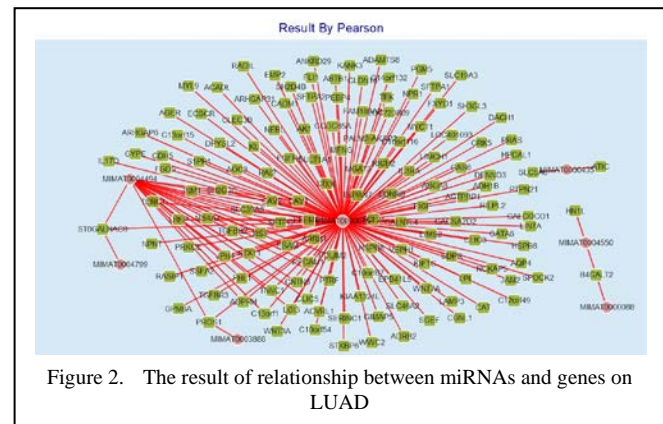


Figure 2. The result of relationship between miRNAs and genes on LUAD

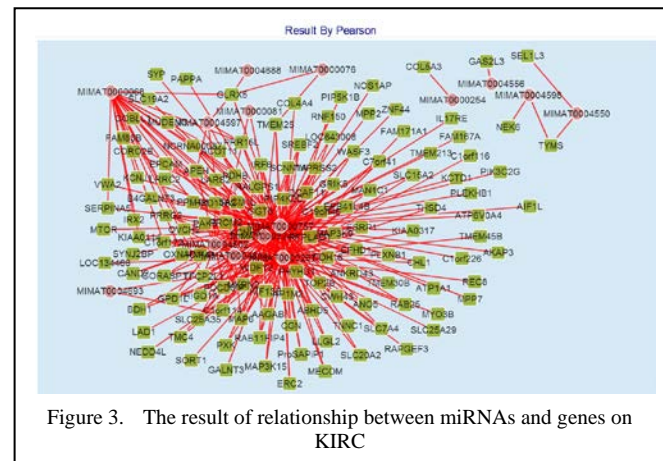


Figure 3. The result of relationship between miRNAs and genes on KIRC

correlations contain 168 miRNA-gene pairs, and there are 163 pairs of up-regulated miRNAs and down-regulated genes and 5 pairs of down-regulated miRNAs and up-regulated genes. The significant miRNA MIMAT0000076 (hsa-mir-21) can provide a link between inflammation and cancer and it was reported to be overexpressed in lung cancer [31].

Kidney renal clear cell carcinoma (KIRC) is a common group of chemotherapy-resistant diseases and can be distinguished by histopathological features and underlying gene mutations [32]. We set the Pearson coefficient threshold to -0.78 for KIRC and the detail of results is shown in Fig.3. The results contain 322 miRNA-gene pairs, including 316 pairs of up-regulated miRNAs and down-regulated genes, and 6 pairs of down-regulated miRNAs and up-regulated genes. The significant miRNA MIMAT0000068 (hsa-miR-15a) is reported to be upregulated in urine of patients with renal cell carcinoma and undetectable in oncocytoma, other tumors, and urinary tract inflammation.

Liver hepatocellular carcinoma (LIHC) is considered to be one of the most highly malignant and lethal cancers of the world. The multi-factorial molecular pathogenesis of LIHC is contributed by various genes and microRNAs [33]. We set

the Pearson coefficient threshold to -0.6 and the detail of results on LIHC is shown in Fig.4. The results contain miRNA-gene 439 pairs, including 51 pairs of up-regulated miRNAs and down-regulated genes, and 388 pairs of down-regulated miRNAs and up-regulated genes. The significant miRNA MIMAT0000278 (hsa-miR-221) has contribution to improve hepatocellular carcinoma through modulation of apoptosis [34].

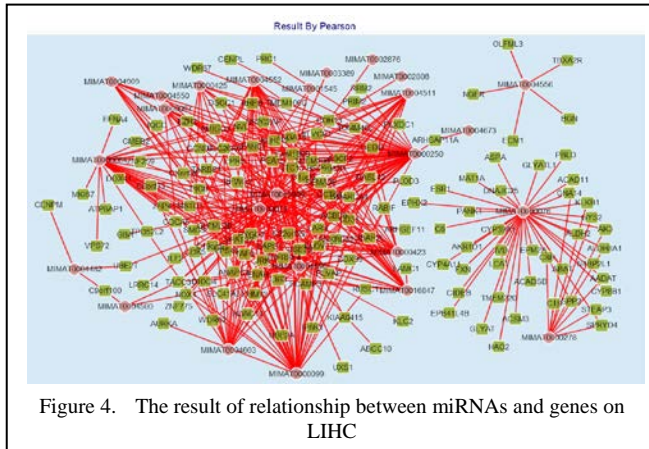


Figure 4. The result of relationship between miRNAs and genes on LIHC

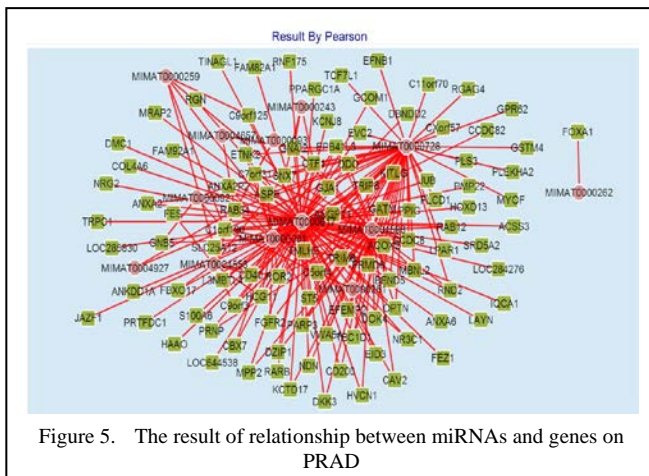


Figure 5. The result of relationship between miRNAs and genes on PRAD

Prostate carcinoma (PRAD) is the most common solid malignancy and the second most common cause of cancer-related death in men. The detail of results on PRAD is shown in Fig.5. We set the Pearson coefficient threshold to -0.65 and the results contain 259 pairs, including 258 pairs of up-regulated miRNAs and down-regulated genes and only one pair of down-regulated miRNAs and up-regulated genes. The significant miRNA MIMAT0000728 (hsa-miR-375) has been found upregulated in prostate tumors compared with normal prostate tissue [35].

C. The Filtered Relationship by miRNA Target Databases

In this section, for reducing the false positive rate of target prediction databases, only the miRNA-gene pairs predicted by miRTarBase, miRanda or PITA database are included in final results. In the analysis of the four datasets, the number of selected miRNAs, the number of selected

genes and the Pearson coefficient threshold are set to the same as which set in Section B of Result part. To eliminate false positive rates of the target predictions, we union the target-relationship miRNA-gene pairs that are predicted by all the databases.

In the result analyzed by the method on the LUAD dataset, there are 26 miRNA-gene pairs which contain one different miRNAs and 26 different genes. TABLE IV shows top ten results with the significant pearson score. Among the genes, the expression of SFTPA1 was associated with an increased frequency of lung cancer [36].

In the result analyzed by the method on the KIRC dataset, there are 11 miRNA-gene pairs which contain four different miRNAs and 11 different genes. TABLE V shows top ten results with the significant pearson score. It can be seen that miRNAs, miR-21-3p (MIMAT0000267) had an upregulated expression level in Clear Cell Renal Cell Carcinoma [37].

In the result analyzed by the method on the LIHC dataset, there are 29 miRNA-gene pairs which contain nine different miRNAs and 24 different genes. TABLE VI shows top ten results with the significant pearson score. Among the miRNAs, downregulation of MIMAT0000099 (hsa-miR-101-3p) promoted cancer cell growth and migration, and Rab5a is one of the target genes of miR-101-3p in LIHC [38].

In the result analyzed by the method on the PRAD dataset, there are 35 miRNA-gene pairs which contain 6 different miRNAs and 29 different genes. TABLE VII shows top ten results with the significant pearson score. Among the miRNAs, miR-200c (MIMAT0000617) and miR-141 are members of the miR-200 family and are important regulators of the epithelial to mesenchymal transition (EMT). In addition to the role of miR-200c and miR-141 in the phenotypic conversion of normal cells, dysregulation of normal patterns of miR-200c expression occurs in multiple types of cancer cells and is linked to tumor progression [39].

In summary, the improved multiply-step method for miRNAs and Genes Integrated Analysis can identify the target relationship miRNA-gene pairs, and those miRNA-gene pairs not only contain the pairs provided by the miRNA target prediction databases further comprise the pairs calculated based on expression data. The miRNA-gene pairs provided by the miRNA target prediction database have been examined, and we can get new miRNA-gene pairs using the proposed method. The new miRNA-gene pairs are meaningful in the study of cancers, and the new miRNA-gene pairs that are not provided by the miRNA target prediction database can be made available to other researchers to verify.

IV. CONCLUSIONS AND DISCUSSIONS

With the development of the expression technology, a large amount of expression datasets of miRNAs/genes can be obtained from the Internet. And a lot of studies indicate that miRNAs and genes which differentially expressed in tumor tissues and corresponding normal tissues play an important role in cancer development and growth. And miRNAs can regulate target gene expression, so it is critical to discover

the aberrantly expressed miRNAs and their genes to explore carcinogenic mechanism. In this article, we propose an improved multiply-step approach to identify the target relationship miRNA-gene pairs, and also propose an available web tool for the integrative analysis of target prediction, using different relatedness measures and

integration methods to identify target predictions and reconstruct miRNA-gene networks. miRNA and gene expression profiles are not straightforward for researchers, however the web tool is useful and will facilitate users in the recognition of the regulatory networks and in discovery of biological markers.

TABLE IV. DETAILS OF THE RESULT ANALYZED BY THE METHOD ON THE LUAD DATASET

miRNAAccession	geneSymbol	mirtarbase	pitaScore	miranda	pearsonScore
MIMAT0000076	TGFBR2	+	-9.61	+	-0.758
MIMAT0000076	WWC2	+	-	+	-0.739
MIMAT0000076	CADM1	-	3.34	+	-0.732
MIMAT0000076	ADPRH	-	-	+	-0.767
MIMAT0000076	CCDC85A	-	-	+	-0.759
MIMAT0000076	SFTPA1	-	-	+	-0.758
MIMAT0000076	KL	-	-	+	-0.731
MIMAT0000076	PTPN21	-	-	+	-0.733
MIMAT0000076	ARHGAP31	-	-	+	-0.738
MIMAT0000076	PALM2-AKAP2	-	-	+	-0.734

TABLE V. Details of The Result Analyzed by The Method on The KIRC Dataset

miRNAAccession	geneSymbol	mirtarbase	pitaScore	miranda	pearsonScore
MIMAT0000267	GPD1L	+	-	+	-0.817
MIMAT0000267	WDFY2	-	-	+	-0.814
MIMAT0000267	TMEM30B	-	-	+	-0.803
MIMAT0000757	CHL1	-	-7.96	-	-0.799
MIMAT0000081	PDHB	-	-	+	-0.783
MIMAT0000068	PRR15L	-	-	+	-0.833
MIMAT0000068	ACOT11	-	-	+	-0.803
MIMAT0000068	OXNAD1	-	-	+	-0.798
MIMAT0000068	RNF43	-	-	+	-0.79
MIMAT0000068	SACM1L	-	-	+	-0.801

TABLE VI. DETAILS OF THE RESULT ANALYZED BY THE METHOD ON THE LIHC DATASET

miRNAAccession	geneSymbol	mirtarbase	pitaScore	miranda	pearsonScore
MIMAT0000076	EPM2A	+	-	+	-0.692
MIMAT0000076	N4BP2L1	-	-	+	-0.639
MIMAT0000099	UBE2A	-	-9.0	+	-0.644
MIMAT0000099	SMARCA4	-	-8.62	+	-0.664
MIMAT0000099	PYGO2	-	-1.94	+	-0.644
MIMAT0000099	ANAPC7	-	-	+	-0.686
MIMAT0000099	C19orf61	-	-	+	-0.638
MIMAT0000423	ELOVL1	-	-5.79	+	-0.63
MIMAT0000064	EZH2	-	-	+	-0.631
MIMAT0000064	FAM189B	-	-	+	-0.693

TABLE VII. DETAIL OF THE RESULT ANALYZED BY THE METHOD ON THE PRAD DATASET

miRNAAccession	geneSymbol	mirtarbase	pitaScore	miranda	pearsonScore
MIMAT0000728	VWA5A	-	-	+	-0.7
MIMAT0000617	NR3C1	-	-4.05	+	-0.703
MIMAT0000617	ROR2	-	-2.14	+	-0.718
MIMAT0000617	C9orf3	-	-	+	-0.704
MIMAT0000617	FRMD6	-	-	+	-0.714
MIMAT0000617	ANGPT1	-	-	+	-0.709
MIMAT0000617	MBNL2	-	-	+	-0.714
MIMAT0000617	GATM	-	-	+	-0.704
MIMAT0000617	GJA1	-	-	+	-0.723
MIMAT0000617	ACOX2	-	-	+	-0.727

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (61472158, 61402194, 61572228, 61272207) and China Postdoctoral Science Foundation (2014T70291).

REFERENCES

- [1] Kartha, Reena V., and S. Subramanian. *MicroRNA Control of Apoptotic Programs in Cancer*. Trends in Stem Cell Proliferation and Cancer Research. 2013:503-530.
- [2] Manuela, Ferracin, and N. Massimo. *MicroRNAs and Their Role in Cancer*. eLS. John Wiley & Sons, Ltd, 2012.
- [3] Volinia, Stefano, et al. *A microRNA expression signature of human solid tumors defines cancer gene targets*. Proceedings of the National academy of Sciences of the United States of America 103.7 (2006): 2257-2261.
- [4] Gregory R I, Shiekhattar R. *MicroRNA biogenesis and cancer*. [J]. Cancer Research, 2011, 65(65):3509-12.
- [5] Webster, R.J., et al., *Regulation of epidermal growth factor receptor signaling in human cancer cells by microRNA-7*. Journal of Biological Chemistry, 2009. 284(9): p. 5731-5741.
- [6] Wu, W., Sun, M., Zou, G. M., & Chen, J. (2007). *MicroRNA and cancer: Current status and prospective*. International Journal of Cancer, 120(5), 953-960.
- [7] Lewis B P, Burge C B, Bartel D P. *Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets*[J]. cell, 2005, 120(1): 15-20.
- [8] Bo, T. and I. Jonassen, *New feature subset selection procedures for classification of expression profiles*. Genome biology, 2002. 3(4): p. 1-0017.11.
- [9] Faraoni, I., Antonetti, F. R., Cardone, J., & Bonmassar, E. (2009). *miR-155 gene: a typical multifunctional microRNA*. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 1792(6), 497-505.
- [10] Zhang, B., Pan, X., Cobb, G. P., & Anderson, T. A. (2007). *microRNAs as oncogenes and tumor suppressors*. Developmental biology, 302(1), 1-12.
- [11] Calin, G.A. and C.M. Croce, *MicroRNA-cancer connection: the beginning of a new tale*. Cancer research, 2006. 66(15): p. 7390-7394.
- [12] Maziere, P. and A.J. Enright, *Prediction of microRNA targets*. Drug discovery today, 2007. 12(11): p. 452-458.
- [13] Satoh, J. and H. Tabunoki. "Comprehensive analysis of human microRNA target networks." *BioData Mining* 4.1(2011):1-13.
- [14] Betel, D., Wilson, M., Gabow, A., Marks, D. S., & Sander, C. (2008). *The microRNA. org resource: targets and expression*. Nucleic acids research, 36(suppl 1), D149-D153.
- [15] Haney, S., M. Kam, and L. Hrebien. *Benefits of using paired controls for analyzing gene expression of prostate cancer*. in *Bioinformatics and BioEngineering, 2008. BIBE 2008. 8th IEEE International Conference on*. 2008. IEEE.
- [16] Tan, Q., M. Thomassen, and T.A. Kruse, *Feature selection for predicting tumor metastases in microarray experiments using paired design*. Cancer informatics, 2007. 3: p. 213.
- [17] Toh, S.H., Prathipati, P., Motakis, E., Kwoh, C.K., Yenamandra, S.P. and Kuznetsov, V.A., 2011. *A robust tool for discriminative analysis and feature selection in paired samples impacts the identification of the genes essential for reprogramming lung tissue to adenocarcinoma*. BMC genomics, 12(3), p.1.
- [18] Liu, Huiqing, et al. *Identifying mRNA targets of microRNA dysregulated in cancer: with application to clear cell Renal Cell Carcinoma*. BMC systems biology 4.1 (2010): 1.
- [19] Peng, Xinxia, et al. *Computational identification of hepatitis C virus associated microRNA-mRNA regulatory modules in human livers*. BMC genomics 10.1 (2009): 1.
- [20] Fu, Jihong, et al. *Identifying microRNA-mRNA regulatory network in colorectal cancer by a combination of expression profile and bioinformatics analysis*. BMC systems biology 6.1 (2012): 68.
- [21] Lihua HE, Zhongbo CAO, Yan WANG, Wei DU, Yanchun LIANG. *An Ensemble Feature Selection Method Based on mRMR for Paired Microarray Data*. Journal of Computational Information Systems 10: 11 (2014) 4875-4882.
- [22] Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X., & Li, T. (2009). *miRecords: an integrated resource for microRNA-target interactions*. Nucleic acids research, 37(suppl 1), D105-D110.I. S.
- [23] Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., & Segal, E. (2007). *The role of site accessibility in microRNA target recognition*. Nature genetics, 39(10), 1278-1284.
- [24] Hudson, Thomas J., et al. *International network of cancer genome projects*. Nature 464.7291 (2010): 993-998.
- [25] Ding, C. and H. Peng, *Minimum redundancy feature selection from microarray gene expression data*. Journal of bioinformatics and computational biology, 2005. 3(02): p. 185-205.
- [26] Peng, H., F. Long, and C. Ding, *Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005. 27(8): p. 1226-1238.
- [27] Cai, Y., Huang, T., Hu, L., Shi, X., Xie, L., & Li, Y. (2012). *Prediction of lysine ubiquitination with mRMR feature selection and analysis*. Amino acids, 42(4), 1387-1395.
- [28] Li, B. Q., Hu, L. L., Chen, L., Feng, K. Y., Cai, Y. D., & Chou, K. C. (2012). *Prediction of protein domain with mRMR feature selection and analysis*. PLoS One, 7(6), e39308.
- [29] Yousef, M., Jung, S., Showe, L. C., & Showe, M. K. (2007). *Recursive Cluster Elimination (RCE) for classification and feature selection from gene expression data*. BMC bioinformatics, 8(1), 1.
- [30] Friedman RC, Farh KK, Burge CB, Bartel DP; Farh; Burge; Bartel (January 2009). *Most mammalian mRNAs are conserved targets of microRNAs*. Genome Res. 19 (1): 92-105.
- [31] Raffaele Baffa , Fassan M, Volinia S, et al. *MicroRNA expression profiling of human metastatic cancers identifies cancer gene targets* †[J]. Journal of Pathology, 2009, 219(2):214-221.
- [32] Linehan W M, Srinivasan R, Schmidt L S. *The genetic basis of kidney cancer: a metabolic disease*[J]. Nature Reviews Urology, 2010, 7(5):277-85.
- [33] Wei L, Lian B, Zhang Y, et al. *Application of microRNA and mRNA expression profiling on prognostic biomarker discovery for hepatocellular carcinoma*[J]. BMC genomics, 2014, 15(1): 1.
- [34] Gramantieri L, Fornari F, Ferracin M, et al. *MicroRNA-221 targets Bmf in hepatocellular carcinoma and correlates with tumor multifocality*[J]. Clinical Cancer Research, 2009, 15(16): 5073-5081.
- [35] Zehentmayr F, Hauser-Kronberger C, Zellinger B, et al. *Hsa-miR-375 is a predictor of local control in early stage breast cancer*[J]. Clinical epigenetics, 2016, 8(1): 1.
- [36] Nathan N, Giraud V, Picard C, et al. *Germline SFTPA1 mutation in familial idiopathic interstitial pneumonia and lung cancer*[J]. Human molecular genetics, 2016, 25(8): 1457-1467.
- [37] Osanto S, Qin Y, Buermans H P, et al. *Genome-wide microRNA expression analysis of clear cell renal cell carcinoma by next generation deep sequencing*[J]. PloS one, 2012, 7(6): e38298.
- [38] Sheng Y, Li J, Zou C, et al. *Downregulation of miR-101-3p by hepatitis B virus promotes proliferation and migration of hepatocellular carcinoma cells by targeting Rab5a*[J]. Archives of virology, 2014, 159(9): 2397-2410.
- [39] Vrba L, Jensen T J, Garbe J C, et al. *Role for DNA methylation in the regulation of miR-200c and miR-141 expression in normal and cancer cells* [J]. PloS one, 2010, 5(1): e869.