An Integrative Framework for Identifying Consistent MicroRNA Expression Signatures Associated with Clear Cell Renal Cell Carcinoma

Jiajia Chen[1,2,#], Daqing Zhang[1,#], Wenyu Zhang[1], Yifei Tang[1], Lingchuan Guo[3,*], Bairong Shen[1,*]
[1]Center for Systems Biology, Soochow University, Suzhou, China
[2]School of Chemistry and Biological Engineering, Suzhou University of Science and Technology, Suzhou, China
[3]Department of Pathology, Soochow University School of Medicine, Suzhou 215123, China
[#] Contribute equally
*Corresponding author: bairong.shen@suda.edu.cn; szglc@hotmail.com

***Abstract*—Clear cell renal cell carcinoma (ccRCC) is the most common and invasive renal-originated malignancy. Altered microRNA expression has been observed in many human cancers including ccRCC. Microarray is routinely used in labs worldwide for detecting cancer specific microRNA expression profiles, but no consistent conclusion could be drawn so far. The function of microRNAs in carcinogenesis of this tumor type is thereof largely unknown. In this study, we describe an integrative framework to improve the comparability of differentially expressed microRNAs (DE-miRNAs) from different experiments, and apply it to 4 publicly available microRNA expression datasets in ccRCC. The approach uses a novel statistic method for cancer outlier detection. The identified DE-miRNAs are then screened by POMA, an in-house developed predictor, for microRNAs with real regulatory activity in the disease. The proposed framework not only achieves high reproducibility across different datasets but also identifies a consistent set of 12 DE-miRNAs which could be putative biomarkers and therapeutic targets. The targets of DE-miRNAs in each dataset were then mapped to functional databases for enrichment analysis. Both novel and previously characterized microRNA-regulated molecular pathways are identified that are likely to contribute to the pathogenesis of ccRCC. Overlapping comparison suggests that independent ccRCC expression profiles are more consistent at pathway level than that at gene/microRNA level.**

***Keywords*—meta-analysis; microRNA expression; Clear Cell Renal Cell Carcinoma; pathway enrichment; GeneGO's database**

I. INTRODUCTION

Renal cell carcinoma (RCC) represents the most common urological neoplasm. It accounts for more than 90% of the malignancy arising from the kidney and the incidence is continuously increasing [1]. Clear cell renal cell carcinoma (ccRCC) is the most frequent histological subtype of RCC. The overall clinical outcome of ccRCC is poor for lack of efficacy in traditional chemotherapy and radiotherapy [2]. Development of new treatment strategies has been slowed by the lack of biomarkers for the disease. Therefore, detecting new diagnostic and prognostic biomarkers in ccRCC becomes an important topic in study. microRNAs are short non-coding RNA molecules that regulate gene expression by translational inhibition or mRNA degradation [3]. Recent evidence suggests that aberrant changes in their expression are associated with human malignancies [4, 5]. Consequently, identifying microRNA expression profiles in diseases have recently gained remarkable interest. Using microRNA microarray techniques, various studies have identified microRNAs that are differentially expressed between ccRCC lesions and adjacent normal tissues. Nevertheless, the DE-miRNA lists reported by different laboratories vary widely and a common ccRCC-specific microRNA signature is so far not available.

In light of this inconsistency, we describe herein an integrated approach to improve the inter-dataset reproducibility for microRNA expression signatures. We then applied the proposed approach to 4 microRNA expression microarray datasets associated with ccRCC and obtained a consistent microRNA expression signature. Furthermore, we examined the functions and pathways that are most likely to be affected by the deregulated microRNAs. We demonstrate that the expression signatures of independent datasets are more consistent at pathway level than at microRNA/gene level. We also have identified novel microRNA-regulated molecular pathways that are likely to contribute to the pathogenesis of ccRCC. Fig. 1 displays the pipeline of the whole procedure in this study.

II. RESULTS

A. Detection of Differentially Expressed MicroRNAs with a Novel Statistical Method

The t-statistics is the most popular method for differential gene expression detection in microarray studies. Recently, it's been realized that oncogenes have heterogeneous activation patterns in most cancer types. Some oncogenes show altered expressions only in a small subset of samples. The study of Tomlins *et al.* [6] showed that t-statistics has low detection power in this case. The problem associated with t-statistics has motivated a series of new analytical methods [7-9]. Through applications to public prostate cancer microarray datasets in our previous study, we have demonstrated that the newly developed statistics showed superior performance than

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

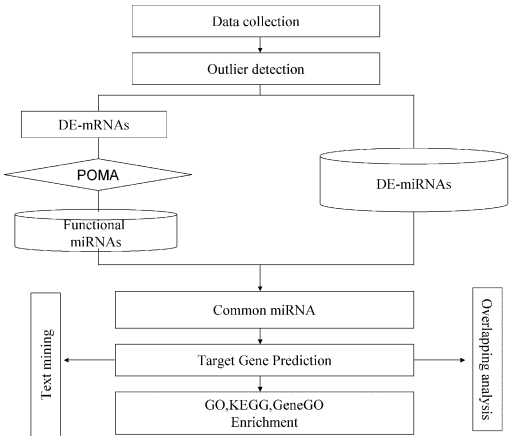

Figure 1. **Schematic diagram depicting the analysis pipeline in this study.**

traditional t-statistics in outlier detection, and therefore provided useful alternative to the t-statistics. In this study we used maximum ordered subset t-statistics (MOST) [9] for outlier detection. MOST seems to have superior performance when the number of activated samples is unknown.

We collected 4 publicly available microRNA profiling datasets. For each dataset we used MOST to identify differentially expressed outliers from subsets of cancer samples. The percentile of outliers is set to 0.05 to retrieve the top 5% and bottom 5% of the miRNA outliers.

B. Refinement of DE-miRNA Lists with the Pipeline of Outlier MicroRNA Analysis (POMA)

To further refine the microRNA outlier lists and remove the false positive discoveries, we applied POMA (Pipeline of Outlier microRNA Analysis), an in-house developed prediction model to the outlier lists generated by MOST. POMA is a simple model created in our previous study [10] to evaluate the relevance of microRNAs in a given disease condition. microRNAs with poor regulatory activity will be excluded from further analysis. The underlying hypothesis in POMA is that the microRNA activity can be reflected by the deregulated expression of its target genes; microRNAs having larger proportion of unique targeting outlier genes are more reliable to show regulatory activity. The stepwise procedure of POMA is as follows:

a) We first reconstructed the human microRNA-mRNA interaction network, based on the microRNA-mRNA interaction data extracted from 4 experimental validation databases (miRecords, TarBase, miR2Disease, miRTarbase) and 3 computational prediction databases (HOCTAR, ExprTargetDB, starBase).

b) By reanalyzing the publicly available gene expression data from Gene Expression Omnibus (GEO), we identified deregulated genes in ccRCC samples vs. normal samples.

c) The outlier genes detected above were then mapped to the microRNA-mRNA interaction network to construct a candidate microRNA-mRNA interaction sub-network in the ccRCC condition.

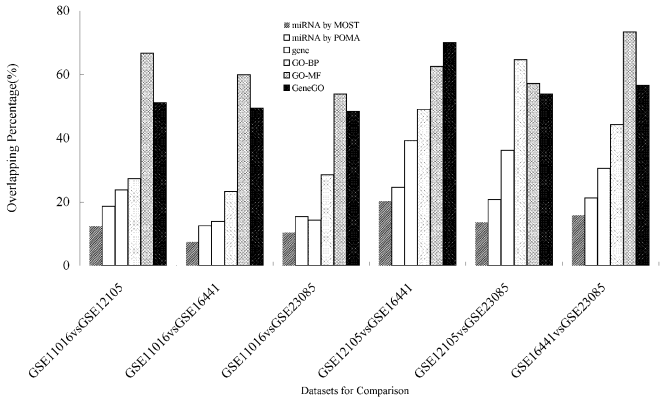

Figure 2. **Pair-wise comparison between 4 datasets at different levels.** X axis shows the 6 pair-wise comparison sets derived from 4 datasets. Y axis denotes the overlapping percentage at different levels.

d) A parameter was defined to measure the probability of microRNA having regulatory activity in ccRCC condition:

$$Z\_score = \alpha / \beta \quad (1)$$

α: Number of outlier genes exclusively targeted by a specific microRNA; β: Number of all the outlier genes targeted by a specific microRNA; (α, β >1).

Z_score is then calculated for each candidate microRNA in the sub-network. Using a threshold of 0.001, we identified a list of active microRNAs with potential regulatory role in ccRCC.

e) The predicted active microRNAs were then cross-matched with the DE-miRNA list of each dataset. The intersected microRNAs are retained for subsequent analysis (See additional file 1 for detailed list of intersected miRNAs).

C. Inter-dataset Consistency is Improved after POMA Filtration

After POMA filtration, the miRNA outlier list is reduced to a highly robust subset of direct functional microRNAs in ccRCC. By comparing the percent overlap of these lists, we observed a higher overlap between different datasets after POMA filtration. The overlapping percentage is illustrated in Fig. 2. The p-values for the difference in overlap were 4.38165E-05 by paired t-test, indicating the significance of the result. Although the number of outliers decreased, the consistency between different datasets is significantly enhanced. The improved consistency enables us to extract common microRNA expression signatures from these datasets. A set of 12 microRNAs shared by all the 4 datasets were retrieved, listed in Table I. Furthermore, to validate the relevance of these microRNAs in the regulation of ccRCC, we performed a literature search of the above identified microRNA markers. All of the 12 microRNAs within this set have been previously described for their roles in renal cell carcinoma. The number of supporting literatures for each microRNA is also listed in Table I. These “literature curetted”

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

microRNAs could serve as a focused but robust panel to distinguish clear cell renal carcinomas from normal kidney tissue.

TABLE I. DE-MIRNAS WITH OUTLIER ACTIVITY IN CCRCC PATHOGENESIS SHARED BY 4 DATASETS.

Human microRNA	Accession No.	Chr	Chromosomal location	microRNA cluster	Family	Pubmed citation No.
hsa-miR-210	MIMAT0000267	11	transcripts	0	210	15
hsa-miR-138-5p	MIMAT0000430	3	intergenic	0	138	5
hsa-miR-16-5p	MIMAT0000069	13	transcripts	1	15	2
hsa-miR-224-5p	MIMAT0000281	X	transcripts	1	224	9
hsa-miR-34a-5p	MIMAT0000255	1	intergenic	0	34	5
hsa-miR-184	MIMAT0000454	15	intergenic	0	184	6
hsa-miR-122-5p	MIMAT0000421	18	intergenic	1	122	6
hsa-miR-126-3p	MIMAT0000445	9	transcript	0	126	5
hsa-miR-155-5p	MIMAT0000646	21	transcript	0	155	14
hsa-miR-15b-5p	MIMAT0000068	3	transcript	1	15	3
hsa-miR-532-5p	MIMAT0002888	X	transcript	5	188	1
hsa-miR-660	MIMAT0003338	X	transcript	6	188	0

TABLE II. THE NUMBER OF VARIOUS ENRICHED BIOLOGICAL THEMES FOR DIFFERENT DATASETS.

Dataset	DE-miRNA	Target	GO-BP FDR < 0.05	GO-MF FDR < 0.05	KEGG FDR < 0.05	GeneGO FDR < 0.001
GSE11016	21	853	18	9	2	99
GSE 12105	22	764	54	11	7	152
GSE 16441	35	1136	110	15	11	149
GSE 23085	31	895	53	11	13	125
Shared	5	388	8	7	6	62

D. MicroRNA Targets Prediction and Functional Enrichment

For each dataset, we conducted a high-stringency target prediction to identify potential target genes of the microRNAs that are differentially expressed. Target genes were retrieved from both experimentally supported databases and prediction algorithms (See Method section for detailed information). Number of targets obtained for each dataset is listed in table II. Detailed list of target genes for each dataset could be found in additional file 2. To estimate the overall effect of the microRNAs on cellular functions, the targets from each individual dataset were mapped to several functional databases including GO [11], KEGG [12] and GeneGO (GeneGO, Inc.) Table II illustrates the number of various biological themes which are enriched with target genes for each dataset. Detailed lists of significantly enriched functional terms/pathways are available in Additional file 3.

E. Identification of ccRCC Related Functions and Pathways

Gene Ontology (GO) analysis suggested that the predicted targets of DE-miRNAs in ccRCC were significantly enriched for GO terms associated with transcription and cell cycle, including: regulation of cyclin-dependent protein kinase activity, transcription, DNA-dependent regulation of transcription, regulation of cell proliferation, sequence-specific DNA binding and transcription regulator activity.

The top KEGG pathways regulated by DE-miRNA converge on signaling pathways associated in cancer, such as Colorectal cancer, Cell cycle, Neurotrophin signaling pathway, Renal cell carcinoma, Prostate cancer, MAPK signaling pathway, and p53 signaling pathway.

We also identified multiple GeneGO pathways that are statistically enriched with the DE-miRNA targets, such as cell adhesion, cell cycle and cytoskeleton remodeling, many of which are known to be involved in tumor pathogenesis. We found 62 significantly enriched pathways that were overlapped by 4 datasets (see Additional file 3). To evaluate the relevance of these pathways in ccRCC, we summarized the published literatures describing the network objects constituting the pathway map. Among the 62 enriched GeneGO pathways, 36 pathways (58.1%) were found to be highly saturated with well-characterized RCC-related objects (ratio>0.15, p-value<0.0001, see Additional file 7 for the volcano plot). These pathways are considered to be potential pathways contributing to renal carcinogenesis. Table III lists the top 10 of the significant GeneGO pathways enriched with ccRCC-related objects. A further literature search in PubMed highlighted 22 out of the 36 putative ccRCC-related pathways with previous annotation in ccRCC carcinogenesis. The rest 14 pathways without literature support could be promising novel pathways which need more wet-lab validation (See details in Additional file 4).

Fig. 3 shows the most significant novel GeneGO pathway, TGF, WNT and cytoskeletal remodeling. This pathway is enriched with network objects previously found to associate with renal cell carcinoma, such as TCF, AKT, VEGF-A, WNT, Frizzled, TGF-beta, RhoA, Beta-catinin, c-Myc, Cyclin D1 and c-Jun. It's obvious in Fig. 3 that the pathway converges on WNT protein family and its downstream effectors. The WNT protein family has been implicated in oncogenesis and regulation of cell fate. Upon binding with the Axin-related protein, WNT regulates the stability of Beta-catenin, a downstream component of the Wnt signaling pathway. Beta-catenin then activates transcription factors of the TCF/LEF family which function in cell fate specification, leading to activation of WNT responsive genes, such as c-Myc, c-Jun and Cyclin D1. Both c-Myc and c-Jun have been

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

TABLE III. TOP 10 OF THE SIGNIFICANT GENEGO PATHWAYS ENRICHED WITH BOTH DE-MIRNA TARGETS AND CURATED RCC-RELATED GENES.

Pathway map	Pathway map category	Ratio of RCC related objects	P-value	Pubmed citation count
TGF, WNT and cytoskeletal remodeling	Cytoskeleton remodeling	25/111	2.41E-16	
AKT signaling	Signal transduction	17/43	5.00E-16	70
Brca1 as a transcription regulator	DNA damage	13/30	3.70E-13	
PTEN pathway	Signal transduction	15/46	7.79E-13	22
PIP3 signaling in cardiac myocytes	Development	15/47	1.12E-12	
Regulation of epithelial-to-mesenchymal transition (EMT)	Development	16/64	1.12E-11	12
Influence of Ras and Rho proteins on G1/S Transition	Cell cycle	14/53	1.14E-10	2
Cytoskeleton remodeling	Cytoskeleton remodeling	18/102	3.30E-10	2
Regulation of G1/S transition (part 1)	Cell cycle	11/38	7.43E-10	4
Receptor-mediated HIF regulation	Transcription	11/39	1.02E-9	7

suggested to play a role in cell cycle progression, apoptosis and cellular transformation. Cyclin D1 is also involved in the cell-cycle during G1/S transition. Activation of these genes has been reported during early RCC carcinogenesis. These findings are consistent with the observation that WNT signaling pathway is abnormally regulated during renal carcinoma development. Actin and a variety of actin-binding protein are also central in the pathway. Actins are involved in cell motility and maintenance of the cytoskeleton. Actin cytoskeletal disruption is found to be a characteristic event in early RCC development, which is required by morphological changes in transformed cells.

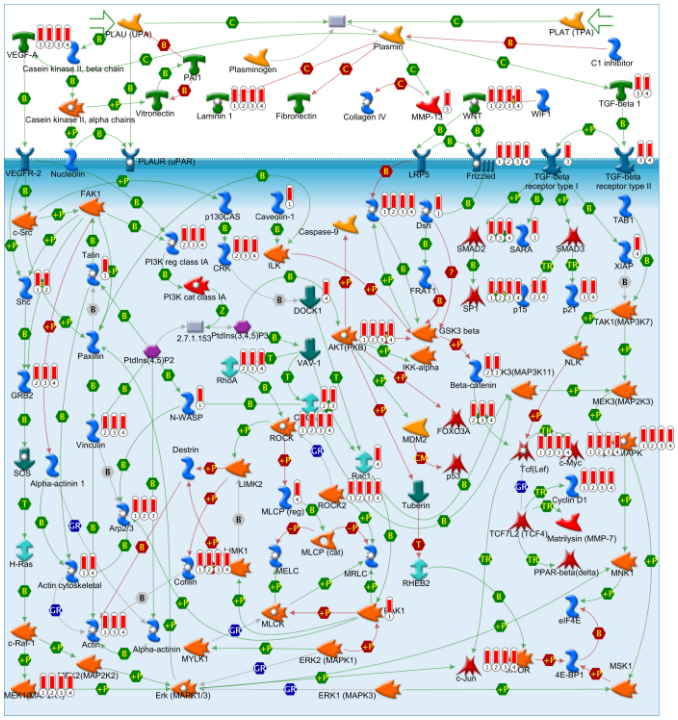

Figure 3. **Graphic illustration of Cytoskeleton remodeling_TGF, WNT and cytoskeletal remodeling pathway map.** Red thermometers show an object that can be regulated by a DE-miRNA. The numerical subscript corresponding to the datasets to which the gene belongs. See additional file 5 for the notation of each sign in this figure.

TGF, WNT and cytoskeletal remodeling pathway also contains some objects without previous annotation in ccRCC carcinogenesis, such as ROCK, MEK1, p38, MAPK, axin. These objects could be putative therapeutic targets for new treatment strategies.

F. Higher consistency can be reached at functional level than that at microRNA level

After mapping targets of DE-microRNAs to functional databases, we performed pair-wise comparison between four datasets at the various observation level including DE-microRNA, targets, GO-MF, GO-BP, and GeneGO pathway, respectively. For 4 different datasets, 6 pairwise sets are available for comparison. Fig. 2 shows the pairwise overlapping percentage between any two datasets. Then we performed paired t-test to decide whether differences observed between different levels are significant (see Table IV). We compared overlapping percentage at functional level (GO-MF, GO-BP and GeneGO pathway level) with that of DE-microRNAs and targets. According to the p-values shown in Table IV, it is clear that the overlapping percentages at functional level are significantly different from that at individual DE-microRNA or target level ($p<0.05$). The result of consistency analysis draws a conclusion that the expression signatures of independent datasets at GO or pathway level are significantly more consistent than at microRNA or target level.

III. DISCUSSION

In this study, we collected 4 publicly available ccRCC associated microRNA microarray datasets, and processed them by a meta-analysis procedure to obtain more reliable expression signatures. The proposed approach features a novel outlier detection method (MOST) and a functional miRNA prediction model (POMA), which could enhance the reproducibility of results across multiple datasets.

We first applied a new statistics to the identification of differentially-expressed microRNAs. Previous studies have

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

TABLE IV. SIGNIFICANCE OF COMPARISON OF OVERLAPPING PERCENTAGE AT DIFFERENT LEVELS

Levels		p-value
GO-BP	DE-microRNA	1.18E-02
	Target gene	1.21E-02
GO-MF	DE-microRNA	1.68E-05
	Target gene	4.77E-04
GeneGO pathway	DE-microRNA	9.70E-06
	Target gene	1.14E-04

shown that some oncogenes occur in only a minority of disease samples.This heterogeneous activation pattern is also one of the reasons for the inter-dataset inconsistency. Therefore, it is essential to examine the subset-specific cancer outlier genes. However, traditional analytical methods, for example, t-statistics would fail to find such oncogene expression profiles. As previous papers have indicated, new statistics perform better than the traditional t-statistical methods and are generally more appropriate for cancer microarray data analysis.

To further refine the list of DE-miRNAs and reduce false discovery, we used POMA to look for the subset of microRNAs with real regulatory activity in ccRCC condition. Those DE-miRNAs without regulatory activity in the disease were excluded from further analysis. POMA has already been successfully employed by our laboratory in the context of prostate cancer [10]. This model focuses not only on the profile of microRNAs, but also on transcripts with altered expression in ccRCC. After POMA filtration, final lists of DE-microRNAs is significantly reduced and substantial consistency is observed between the 4 independent miRNA microarray datasets.

Finally we obtained a list of 12 functional microRNAs that are differentially expressed in all 4 datasets. Literature mining confirmed that all of these microRNAs were previously found to be deregulated in renal cell carcinoma, lending credibility to the list. The DE-miRNA list identified here could serve as potential disease markers which await wet lab validation. Moreover, the general DE-miRNA detection procedure proposed here is not limited to ccRCC only, but could be extended to a wide range of other disease as well.

Consistency analysis at different levels provides solid evidence that cancer signatures at pathway level are more consistent across multiple datasets than at gene level. It's well recognized that cancer is a typical complex disease characterized with concerted molecular changes. There is a growing trend to understand the carcinogenesis as an intricate network of functionally correlated genes. As functionally related genes often display a coordinated expression to accomplish their roles in the cell, one would expect that the inconsistent microRNA lists across studies are functionally more consistent. In other words, the discrepancies at gene level would be less pronounced when the they are mapped to functional groups or biological pathways [13].Therefore we propose that reliable cancer signatures should be investigated in the context of biological pathways.

Functional analysis identified several biological pathways and processes which are preferentially targeted by the deregulated miRNAs. Looking initially at the top enriched GO terms, we see that many are involved in cell cycle regulation (e.g., G1/S transition). Deregulated cell cycle is a hallmark of neoplasia potentially caused by aberrant expression of cell cycle regulators. Recent studies showed that cell cycle checkpoint regulators such as cyclins and cyclin-dependent kinases, are coregulated by the deregulated microRNAs. For example, miR-16 family is reported to trigger a cell cycle arrest by coordinately suppressing multiple cell cycle regulatory genes [14]. Moreover, miR-16 happens to be among the deregulated microRNAs list identified in this study. All the evidences above corroborate the validity of the results of the present study.

To evaluate the relevance of the enriched GeneGO pathways in ccRCC, we performed text mining at both object level and pathway level. Many of the objects that constitute molecular pathways are known to play important roles in the renal carcinogenesis. In addition to the pathways previously found to be critical in RCC tumorigenesis, this study also identified 15 novel ccRCC related pathways. Drugs designed to inhibit these pathways might be useful in preventing tumor formation.

IV. CONCLUSIONS

In this study we have integrated a new outlier detection method MOST with POMA prediction model to create a framework by which to meta-analyze multiple microRNA microarray datasets. The methodology would hopefully improve the comparability of different microarray datasets. Functional analysis demonstrated that cancer signatures at pathway level are more consistent across multiple datasets than at gene level. Novel pathways involved in renal cancer identified in this study might be promising drug targets for ccRCC treatment.

V. MATERIALS AND METHODS

A. Dataset Collection

The publicly available datasets were retrieved from the Gene Expression Omnibus (GEO) as raw data files. Only profiles of ccRCC and non-matching normal renal tissues were extracted for analysis. The microRNA microarray expression datasets are summarized in Table V. The statistical methods used in the original articles for the identification of differentially expressed outliers are also listed. The gene expression files are given in additional file 6. Naming conventions changed somewhat between platforms, so microRNA probe sequences were mapped to latest Sanger miRBase version (release 18) [15] to unify the microRNA names. For the gene annotation, only genes with unique corresponding probe were retrieved, and those genes with multiple probes were removed.

B. Determination of the Differentially Expressed MicroRNAs

Signal intensities for each spot were calculated by subtracting median background intensities. After Quantile normalization, mean intensities were log2 transformed for statistical analysis. Outlier detection was performed with MOST, implemented in R scripts by Lian [9]. The percentile of outlier extraction was set as 0.05 (5%).

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

TABLE V. SUMMARY OF MICRORNA EXPRESSION DATASETS USED IN THIS STUDY.

GEO accession	Platforms	Probe Number	Number of samples		Statistics
			normal tissue	*cancer tissue*	
GSE11016	LC MRA-1001	835	17	17	t-test
GSE12105	Agilent Human microRNA Microarray	490	12	12	t-test
GSE16441	Agilent Human microRNA Microarray	851	8	8	SAM
GSE23085	LC MRA-1001	881	20	20	t-test

C. *Obtaining Target Genes for Differentially Expressed MicroRNAs*

Target mRNAs of the differentially expressed microRNAs were retrieved from databases with experimental evidence (miRecord), as well as from target prediction programs. Prediction of putative target genes was performed using three prediction programs: TargetScan, miRanda, and PicTar. Only targets found in at least two of the three prediction algorithms were retained in order to obtain a more solid result.

D. *Overlapping Analysis at Gene/Pathway level*

The percentage overlap between two datasets is calculated as follows:

$$\text{Overlapping percentage} = \frac{m}{n1+n2-m} \times 100\%,$$

where n1= the number of all the data in dataset 1; n2= the number of all the data in dataset 2; m= the number of overlapping data between two datasets.

E. *Functional enrichment of targets of differentially expressed microRNAs*

To study the function of the differentially expressed microRNAs, we mapped their target genes to GO, KEGG and GeneGO databases. GO and KEGG pathway analysis was performed the using David; GeneGO pathway enrichment analysis was performed using the MetaCore™ (GeneGO, St Joseph, MI, USA). P-values for significance analysis were determined by hypergeometric distribution to estimate the probability for a GO term or pathway that would be identified by chance alone. False Discovery Rate (FDR) adjustment was applied for multiple test correction.

ACKNOWLEDGMENT

This work was supported by the National Nature Science Foundation of China (31170795, 91029703), the Specialized Research Fund for the Doctoral Program of Higher Education of China (20113201110015). International S&T Cooperation Program of Suzhou (SH201120) and the National 973 Programs of China (2010CB945600).

REFERENCES

[1] R. Siegel, E. Ward, O. Brawley, and A. Jemal, "Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths," *CA Cancer J Clin,* vol. 61, pp. 212-236, Jul-Aug 2011.

[2] B. Escudier, "Advanced renal cell carcinoma: current and emerging management strategies," *Drugs,* vol. 67, pp. 1257-1264, 2007.

[3] D. P. Bartel, "MicroRNAs: genomics, biogenesis, mechanism, and function," *Cell,* vol. 116, pp. 281-297, Jan 23 2004.

[4] I. S. Lossos, D. K. Czerwinski, A. A. Alizadeh, M. A. Wechser, R. Tibshirani, et al., "Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes," *N Engl J Med,* vol. 350, pp. 1828-1837, Apr 29 2004.

[5] D. B. Allison, X. Cui, G. P. Page, and M. Sabripour, "Microarray data analysis: from disarray to consolidation and consensus," *Nat Rev Genet,* vol. 7, pp. 55-65, Jan 2006.

[6] S. A. Tomlins, D. R. Rhodes, S. Perner, S. M. Dhanasekaran, R. Mehra, et al., "Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer," *Science,* vol. 310, pp. 644-648, Oct 28 2005.

[7] Y. Wang and R. Rekaya, "LSOSS: Detection of Cancer Outlier Differential Gene Expression," *Biomark Insights,* vol. 5, pp. 69-78, 2010.

[8] J. W. MacDonald and D. Ghosh, "COPA--cancer outlier profile analysis," *Bioinformatics,* vol. 22, pp. 2950-2951, Dec 1 2006.

[9] H. Lian, "MOST: detecting cancer differential gene expression," *Biostatistics,* vol. 9, pp. 411-418, Jul 2008.

[10] W. Zhang and B. Shen, unpublished.

[11] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, et al., "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium," *Nat Genet,* vol. 25, pp. 25-29, May 2000.

[12] M. Kanehisa, M. Araki, S. Goto, M. Hattori, M. Hirakawa, et al., "KEGG for linking genomes to life and the environment," *Nucleic Acids Res,* vol. 36, pp. D480-484, Jan 2008.

[13] M. Zhang, C. Yao, Z. Guo, J. Zou, L. Zhang, et al., "Apparently low reproducibility of true differential expression discoveries in microarray studies," *Bioinformatics,* vol. 24, pp. 2057-2063, Sep 15 2008.

[14] Q. Liu, H. Fu, F. Sun, H. Zhang, Y. Tie, et al., "miR-16 family induces cell cycle arrest by regulating multiple cell cycle genes," *Nucleic Acids Res,* vol. 36, pp. 5391-5404, Sep 2008.

[15] A. Kozomara and S. Griffiths-Jones, "miRBase: integrating microRNA annotation and deep-sequencing data," *Nucleic Acids Res,* vol. 39, pp. D152-157, Jan 2011.

978-1-4673-4398-5/12/$31.00 ©2012 IEEE