Bioinformatics, 31(8), 2015, 1226–1234
doi: 10.1093/bioinformatics/btu811
Advance Access Publication Date: 12 December 2014
Original Paper

Gene expression

Identifying cancer-related microRNAs based on gene expression data

Xing-Ming Zhao[1,]*, Ke-Qin Liu[2,3], Guanghui Zhu[4], Feng He[1], Béatrice Duval[3], Jean-Michel Richer[3], De-Shuang Huang[1], Chang-Jun Jiang[1], Jin-Kao Hao[3,]* and Luonan Chen[5,]*

[1]School of Electronics and Information Engineering, Tongji University, Shanghai 201804, China, [2]Center for Bioinformatics and Systems Biology, Department of Radiology, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA, [3]LERIA, University of Angers, 49045 Angers Cedex 01, France, [4]Department of Mathematics, Shanghai University, Shanghai 200444, China and [5]Key Laboratory of Systems Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

*To whom correspondence should be addressed.

Associate Editor: Ziv Bar-Joseph

Received on July 2, 2014; revised on November 30, 2014; accepted on December 4, 2014

Abstract

Motivation: MicroRNAs (miRNAs) are short non-coding RNAs that play important roles in post-transcriptional regulations as well as other important biological processes. Recently, accumulating evidences indicate that miRNAs are extensively involved in cancer. However, it is a big challenge to identify which miRNAs are related to which cancer considering the complex processes involved in tumors, where one miRNA may target hundreds or even thousands of genes and one gene may regulate multiple miRNAs. Despite integrative analysis of matched gene and miRNA expression data can help identify cancer-associated miRNAs, such kind of data is not commonly available. On the other hand, there are huge amount of gene expression data that are publicly accessible. It will significantly improve the efficiency of characterizing miRNA's function in cancer if we can identify cancer miRNAs directly from gene expression data.

Results: We present a novel computational framework to identify the cancer-related miRNAs based solely on gene expression profiles without requiring either miRNA expression data or the matched gene and miRNA expression data. The results on multiple cancer datasets show that our proposed method can effectively identify cancer-related miRNAs with higher precision compared with other popular approaches. Furthermore, some of our novel predictions are validated by both differentially expressed miRNAs and evidences from literature, implying the predictive power of our proposed method. In addition, we construct a cancer-miRNA-pathway network, which can help explain how miRNAs are involved in cancer.

Availability and implementation: The R code and data files for the proposed method are available at http://comp-sysbio.org/miR_Path/

Contact: liukeq@gmail.com

Supplementary information: supplementary data are available at *Bioinformatics* online.

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

© The Author 2014. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com

1 Introduction

MicroRNAs (miRNAs), as a large family of gene regulators, are involved in various biological processes, such as cell development, proliferation, differentiation and apoptosis. Specifically, they play essential roles in regulating gene expression after the genes are transcribed (Bartel, 2004), where one miRNA may regulate multiple genes. It was found that more than 60% of the protein-coding genes in the human genome are regulated by miRNAs (Esteller, 2011).

Recently, accumulating evidences indicate that miRNAs are extensively involved in tumors (Calin and Croce, 2006). For example, miR-21 induces invasion and metastatic capacity in colon cancer cells (Asangani *et al.*, 2008), while miR-155 up-regulation and let-7a down-regulation can be used to predict the poor survival of lung cancer patients (Yanaihara *et al.*, 2006). In past years, the impact of miRNAs on various cancers has been determined with biological experiments (Croce, 2009; Hanahan and Weinberg, 2011). Unfortunately, it is not feasible to detect all cancer associated miRNAs in laboratory due to the wide range of biological processes in which the miRNAs are involved. Recently, some computational approaches have been proposed to predict cancer-related miRNAs. For example, some statistical approaches have been developed to detect miRNAs that are differentially expressed between normal and cancer samples, and these miRNAs are considered related to cancer (Kuo *et al.*, 2012; Oulas *et al.*, 2011). However, few of such miRNA expression data are available compared with the huge amount of mRNA expression data. Furthermore, the noise inherited in the expression data makes it difficult to detect the differentially expressed miRNAs. Considering miRNAs are regulators of gene expression and the aberrant gene expression might lead to certain diseases, some methods have been proposed to predict disease associated miRNAs based on the miRNA-gene regulation circuit (Chen *et al.*, 2012; Li *et al.*, 2011). Nevertheless, it is not an easy task to identify cancer-related miRNAs based only on their target genes as one miRNA may target hundreds or even thousands of genes while one gene may be regulated by multiple miRNAs (Lim *et al.*, 2005). Furthermore, the lack of context information about the regulation also degrades the performance of such approaches. Under the circumstances, some approaches have been developed to predict cancer-related miRNAs by investigating matched gene and miRNA expression data as well as miRNA-gene regulations (Wuchty *et al.*, 2013; Zhang *et al.*, 2014). Unfortunately, the scarceness of matched miRNA and gene expression datasets limits the application of this promising approach.

Compared with the rare matched miRNA and gene expression profiles, there are huge amounts of gene expression datasets that are publicly available, which can help provide insights into the functions of their miRNA regulators. In this work, we propose a novel computational framework to identify cancer associated miRNAs based on their target genes' expression profiles as well as the miRNA-gene regulation network. Assuming that miRNAs are related to cancers by regulating the pathways in which their target genes are located, we respectively identify sets of miRNAs related to lung cancer, colon cancer, breast cancer and gastric cancer. The benchmarking results demonstrate the higher accuracy of our proposed method compared with other popular approaches. The high consistency between results on independent distinct datasets for the same cancer type implies the robustness of our proposed method. Furthermore, part of our novel predictions are validated by both differentially expressed miRNAs and evidences from literature, indicating the predictive power of our approach. In addition, the pathways regulated by cancer associated miRNAs can help explain how miRNAs are involved in the initiation and progression of cancers.

2 Methods

2.1 Gene expression datasets

The gene expression datasets of four different types of cancers were downloaded from NCBI Gene Expression Omnibus (GEO) (Edgar *et al.*, 2002), including lung cancer, colon cancer, gastric cancer and breast cancer. In particular, to evaluate the efficiency and robustness of our proposed approach, for each cancer type, we collected two independent datasets generated under the same platform. Table 1 lists the detailed information corresponding to each dataset.

For each dataset, the probe-level gene expression data extracted from the CEL files were normalized with the robust multi-array average method (Irizarry *et al.*, 2003). The probes were then mapped to genes with the annotation files obtained from GEO. For the gene that is associated with multiple probes, we assigned it the probe with the largest variation across its expression profile.

2.2 miRNA target genes

We firstly collected the miRNA target genes predicted with different tools, including PicTar (Krek *et al.*, 2005), miRanda (version 3.0) (John *et al.*, 2004), microT (version 5.0) (Maragkakis *et al.*, 2009) and TargetScan (release 6.2) (Lewis *et al.*, 2005). All the miRNAs were downloaded from miRBase (version 16) (Griffiths-Jones *et al.*, 2006). Specifically, for one miRNA, we kept its target genes if they were predicted by at least two tools. Note that some miRNAs will not be considered if they do not have any target genes after the filtering procedure. Then, we extended the miRNA–gene interactions by including those deposited in TarBase (version 6.0) (Sethupathy *et al.*, 2006) that contains experimentally determined miRNA target genes. As a result, we obtained a set of miRNA–gene interactions involving 547 miRNAs and 6796 genes, where each miRNA targets around 12 genes on average.

2.3 Tissue-specific miRNA expression profiles

The expression profiles of 345 miRNAs across 40 normal human tissues were obtained from the supplementary files of the paper by Liang *et al.* (2007). The tissue specificity score of a miRNA m_i in tissue t_j was defined as follows (Lee *et al.*, 2008):

$$\mathrm{Ts}_{ij} = (n-1)E_{ij} \Bigg/ \sum_{k=1,k\neq j}^{n} E_{ik}, \qquad (1)$$

where E_{ij} denotes the expression of m_i in t_j and the same for E_{ik}, and n is the total number of tissues in which m_i is expressed. If Ts_{ij} is above a certain threshold (e.g. 1.5 here), the miRNA m_i will be regarded as specifically expressed in tissue t_j. In particular, if a miRNA was found to be specifically expressed in multiple tissues with the tissue specificity score defined above, only the one in which the miRNA has the highest abundance will be regarded as the tissue that the miRNA is specifically expressed in. In other words, each miRNA will be regarded as only specifically expressed in one tissue if its tissue specificity score is above the threshold.

2.4 Cellular pathways

The predefined biological pathways were obtained from the Molecular Signatures Database (Liberzon *et al.*, 2011), which is a large collection of annotated functional gene sets. We chose the canonical pathways from the curated (c2) gene sets in MsigDB V3.0, which contains 880 metabolic and signaling pathways collected from the online databases, such as BioCarta (www.biocarta.com), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa *et al.*, 2008) and Reactome (Matthews *et al.*, 2009).

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

Table 1. Eight gene expression datasets for four different types of cancers

Cancer	GEO accession number	Number of samples (disease/control)	Platform
Lung cancer	GSE7670 (Su *et al.*, 2007)	54 (27/27)	GPL96
	GSE10072 (Landi *et al.*, 2008)	107 (58/49)	GPL96
Breast cancer	GSE15852 (Pau Ni *et al.*, 2010)	40 (20/20)	GPL96
	GSE20437 (Graham *et al.*, 2010)	18 (9/9)	GPL96
Colon cancer	GSE9348 (Hong *et al.*, 2010)	82 (70/12)	GPL570
	GSE20916 (Skrzypczak *et al.*, 2010)	69 (45/24)	GPL570
Gastric cancer	GSE13911 (D'Errico *et al.*, 2009)	69 (38/31)	GPL570
	GSE19826 (Wang *et al.*, 2012)	27 (12/15)	GPL570

2.5 Identifying cancer-related miRNAs

Figure 1 depicts the flowchart of our proposed framework to identify cancer-related miRNAs based on gene expression profiles. First, for each miRNA, we identified its target gene sets that are possibly related to cancer. Second, the dysfunctional pathways that are associated with cancer were detected, and those miRNA target genes enriched pathways were picked up. Finally, the miRNAs regulating the above-identified pathways were ranked and those highly ranked miRNAs are more likely to be related to cancer. The details were addressed as follows.

In general, each miRNA regulates multiple genes, where the regulation relationships are largely dependent on temporal conditions. To reduce false positives, we assumed that given a miRNA, the expression of its target genes should fluctuate in a concert way. Therefore, given a miRNA, we clustered all its possible target genes into different groups according to their Pearson's correlation coefficients calculated based on their expression profiles. In the clustering procedure, the hierarchical clustering was employed, where genes with correlation coefficient above 0.7 were clustered into one group while those genes that cannot be grouped into any clusters were discarded. Specifically, we only kept those clusters with size larger than three. Note that it is possible to get different number of target gene clusters for distinct miRNAs. For each gene cluster, we further assessed its discriminative ability of separating the cancers from controls with support vector machines (SVMs), where the LIBSVM toolbox implemented in R e1071 package was adopted. The classification capacity of each cluster was evaluated by the area under ROC curve (AUC) score with 5-fold cross validation, where all samples were randomly split into five equal-size subsets without overlap and four subsets were used as training set while the rest one as test set. To evaluate the performance of each cluster in a robust way, the 5-fold cross-validation procedure was repeated for 100 times and the mean of the AUC scores was regarded as its final score, where the pROC package in R was utilized to calculate the AUC score. Consequently, those gene clusters with AUC scores larger than 0.8 were considered related to cancer, and one miRNA will be regarded as cancer miRNAs if at least one of its target clusters has good discriminative capability of separating cancers from controls. The approach identifying cancer miRNAs based on its target gene clusters was called miR_Clust here.

Generally, the cancers are caused due to the dysfunction of cellular pathways instead of the mutation of single genes (Liu *et al.*, 2012). To identify those dysregulated pathways underlying cancers, we defined the pathway's activity score P as follows:

$$P = \sum_{i=1}^{n} t_i \Big/ n, \qquad (2)$$

where n represents the number of genes in each pathway, t_i denotes the t-score associated with the ith gene within the pathway, and the t-score was obtained with the student's t-test by comparing the gene expression profiles for cancers against those for controls. Moreover, the statistical significance of each pathway was evaluated with permutation test, where the same number of genes as that in each pathway was randomly selected and the pathway activity score was calculated for this set of random genes. The permutation test was repeated for 1000 times and the ratio of random gene sets with activity score larger than the one for the pathway was regarded as the probability of observing the pathway randomly, i.e. P-value. As a result, the pathways with P-values < 0.01 were selected as the dysfunctional pathways in cancer. Given a dysfunctional pathway and a gene cluster targeted by one miRNA, the enrichment significance (P-value) of the pathway over the gene cluster can be calculated as follows:

$$P(X \geq x) = 1 - \sum_{k=0}^{x-1} \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, \qquad (3)$$

where N is the number of genes in the expression data, M is the number of genes in the pathway, n is the number of genes in one cluster, and k is the number of common entries between pathway and cluster. If P-value is <0.05, the dysfunctional pathway was considered regulated by the corresponding miRNA, and the miRNA is accordingly regarded as cancer miRNA. This procedure was repeated for all the target clusters of each miRNA. Note that, for each miRNA, we identified the enriched pathways for every target cluster instead of the pool of all target clusters as the genes in the same cluster tend to be co-expressed. It was well recognized that the genes belonging to the same pathway tend to be co-expressed and it is reasonable to assume the genes in one cluster are more likely to be within the same pathway compared with those not in the same cluster (Zhao *et al.*, 2008). Furthermore, the enrichment analysis on the pool of all genes together will cause bias toward large pathways. As a result, for each miRNA, we may obtain a set of dysfunctional pathways regulated by the miRNA. This approach for identifying cancer miRNAs based on the dysfunctional pathways was called miR_Path in this work.

In addition, we ranked the miRNAs by designing a score to evaluate the association between miRNAs and cancer based on their regulated pathways. For each miRNA, the score M_s was defined as follows:

$$M_s = \frac{\sum_{i=1}^{n} H_i * P_i}{n}, \qquad (4)$$

where $H_i = -\log(q_i)$, q_i denotes the P-value of pathway i obtained with the hypergeometric test, P_i denotes the pathway activity score, and n represents the number of all pathways regulated by the miRNA. In the case of one pathway enriched in multiple target clusters for one miRNA, the minimum P-value will be used for q_i.

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

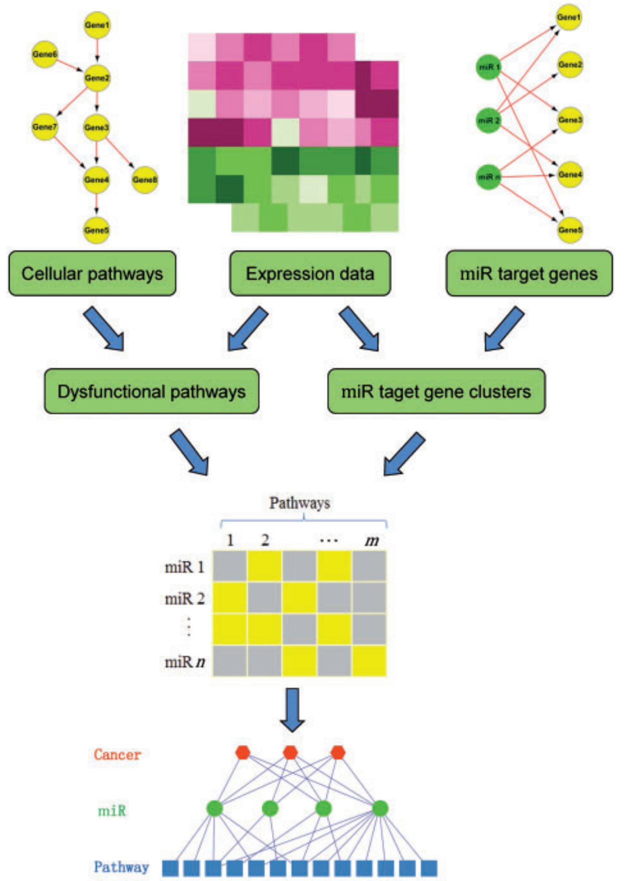

Fig. 1. The flowchart of the framework for identifying cancer-related miRNAs based on gene expression profiles

With the score defined above, we can rank miRNAs for each cancer type and those highly ranked miRNAs are supposed to be more likely related to cancer.

3 Results

3.1 Identification of cancer-related miRNAs

For the four different types of cancers considered here, we identified the possible miRNAs associated with them with miR_Path and miR_Clust. Furthermore, we also proposed one alternative and intuitive approach, namely miR_DG, to identify cancer miRNAs based on their target genes. In miR_DG, we assumed one miRNA to be associated with cancer if its target genes are significantly enriched in those that are differentially expressed between cancers and controls. The differentially expressed genes were detected with student's *t*-test with correction for multiple testing by Benjamini–Hochberg procedure (*P*-value cutoff of 0.05), and the hypergeometric test was used for enrichment analysis of miRNAs' target genes over those differentially expressed genes with a threshold of 0.1. In this work, the miR_DG approach was used as the baseline approach when comparing the performance of distinct approaches.

We also compared our methods with three other popular approaches developed to predict cancer miRNAs, where no miRNA expression data were considered. The IMRE approach inferred miRNA expression levels from mRNA expression data based on putative miRNA targets, and those up-regulated or down-regulated miRNAs in cancer were identified and regarded related to cancer (Kuo *et al.*, 2012). The FCS approach prioritized the cancer miRNAs by measuring functional similarity between miRNA target genes and known cancer genes (Li *et al.*, 2011). Zhang's pipeline identified cancer miRNAs based on the miRNA-gene regulations, where the miRNAs that regulated the top 30% genes differentially expressed between normal and control samples were regarded as cancer miRNAs (Zhang *et al.*, 2014). In this work, for fair comparison, the top ranked same number of miRNAs as those predicted by miR_Path were picked up for these three approaches. The detailed miRNAs predicted by the six approaches can be found in Supplementary Material S1.

Table 2 summarizes the performance of distinct approaches on datasets for lung cancer, colon cancer, gastric cancer and breast cancer, and the manually curated miRNA-disease associations obtained from the HMDD database(Lu *et al.*, 2008) (downloaded in January 2012) were used as the gold standard. From Table 2, we can see our proposed miR_Path and miR_Clust perform very well with comparable overall F1 score about 0.4 across distinct cancers, while miR_Path performs best with respect to precision. The most intuitive and widely used miR_DG approach performs worst possibly due to the inherited noise in the existing miRNA–gene interactions, whereas the clustering step in our approach can efficiently filter out those noisy interactions and uncover true signals. From the results, we can also see that our proposed miR_Path and miR_Clust significantly outperform the other three popular approaches, which illustrates the good predictive power of our approach. The good performance of miR_Path demonstrates that instead of their direct target genes, the pathways regulated by miRNAs can better characterize miRNAs' function as well as their relationship to cancer.

To evaluate the robustness of our proposed approach, for each cancer type, we identified their associated miRNAs in two independent datasets from different laboratories but generated under the same platform. From the results shown in Table 2, we can see that our proposed approach has similar performance on distinct datasets for the same cancer type, indicating the robustness of our proposed approach. Furthermore, for each cancer type, we compared the miRNAs identified in two different datasets as summarized in Table 3. We used the following measure C_m to evaluate distinct approaches, $C_m = m_a \cap m_b / \min\,(m_a, m_b)$, where m_a and m_b, respectively, represents the identified miRNAs from the two datasets of the same cancer type. As demonstrated by the results shown in Table 3, more than 60% of the miRNAs identified by miR_Path and miR_Clust were conserved between different datasets of the same cancer type, implying these miRNAs are indeed related to cancers. The miRNAs identified by miR_Clust are more conserved compared with those by miR_Path possibly due to the incompleteness of the pathway knowledge. Actually, the target gene clusters we identified here can be regarded as pathways to some extent. However, only a very small part of the miRNAs found by miR_DG were conserved across datasets. The overlap of our identified cancer miRNAs between distinct datasets also demonstrates that our approach is indeed robust for identifying cancer miRNAs.

Considering some miRNAs are more conserved and well studied, there is possible bias in the miRNA-gene regulations, which might lead to the high overlap of certain miRNAs in the above results. To see whether this is the case, we permutated the miRNA–gene interaction network for 1000 times while preserving the degrees of both miRNAs and genes, and we predicted cancer miRNAs based on the generated random interactions to see whether we can get cancer miRNAs and high conserved overlap by chance. We calculated *P*-value for random miRNAs as the probability of observing them with higher F1 score than our predicted miRNAs, and the *P*-value for conserved miRNAs as the probability of finding random miRNAs that have higher overlap between two datasets than our predicted

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

Table 2. The performance of different approaches over eight cancer datasets

Index	Methods	Lung cancer		Colon cancer		Gastric cancer		Breast cancer	
		GSE7670	GSE10072	GSE9348	GSE20916	GSE13911	GSE19826	GSE15852	GSE20437
Precision	miR_Path	**0.36**	**0.4**	**0.36**	**0.34**	**0.39**	**0.4**	0.45	**0.41**
	miR_Clust	0.35	0.39	0.33	0.33	0.37	0.36	0.26	0.25
	Zhang's	0.31	0.31	0.29	0.26	0.37	0.38	**0.51**	0.41
	IMRE	0.19	0.19	0.21	0.23	0.26	0.2	0.12	0.23
	FCS	0.28	0.27	0.29	0.25	0.37	0.37	0.3	0.34
	miR_DG	0.27	0.32	0.29	0.21	0.27	0.25	0.25	0.14
Recall	miR_Path	0.47	0.39	0.29	0.4	0.45	0.45	0.14	0.24
	miR_Clust	**0.54**	**0.45**	**0.33**	**0.47**	**0.48**	**0.49**	**0.63**	**0.63**
	Zhang's	0.39	0.29	0.23	0.31	0.43	0.42	0.12	0.2
	IMRE	0.24	0.18	0.17	0.27	0.3	0.22	0.04	0.15
	FCS	0.37	0.26	0.23	0.3	0.43	0.41	0.1	0.23
	miR_DG	0.09	0.07	0.05	0.03	0.05	0.09	0.01	0.02
F1	miR_Path	0.41	0.39	0.32	0.37	**0.42**	**0.43**	0.22	0.32
	miR_Clust	**0.43**	**0.42**	**0.33**	**0.39**	**0.42**	0.41	**0.42**	**0.42**
	Zhang's	0.34	0.3	0.25	0.28	0.4	0.4	0.2	0.27
	IMRE	0.21	0.19	0.19	0.25	0.28	0.21	0.06	0.18
	FCS	0.32	0.27	0.26	0.27	0.4	0.39	0.15	0.27
	miR_DG	0.13	0.12	0.08	0.06	0.09	0.13	0.01	0.04

Note. The numbers in bold denote the best ones with respect to corresponding indices.

miRNAs (details can be found in Supplementary Material S2). The low *P*-values for both cancer miRNAs and their conservation indicate the statistical significance of our proposed approach.

Since the common miRNAs predicted from two independent datasets for the same cancer type are more convincible, we will focus on those common miRNAs hereinafter. For each cancer type, we ranked their related miRNAs according to the miRNA scores (M_s). In particular, the miRNAs with M_s large than the mean of all predicted miRNAs were selected as the top ranked miRNAs (HRmiRs). Note that the HRmiRs are a subset of the common miRNAs identified from two datasets. The HRmiRs identified for the four cancer types can be found in Supplementary Material S3. Figure 2 shows the precision of HRmiRs, predicted miRNAs conserved between two independent datasets (OmiRs) and all identified miRNAs (AmiRs) by miR_Path for the four cancer types, where the cancer miRNAs from HMDD were used as gold standards. Not surprisingly, the HRmiRs achieve the highest precision among the three sets of miRNAs, indicating that the miRNAs with higher M_s are more likely to be related to cancers. Among the HRmiRs, we noticed that most of them are related to cancer although some of them are not specifically associated with certain cancer type. For example, among the 41 HRmiRs identified from the colon cancer datasets, 19 miRNAs are reported related to at least one cancer type according to HMDD except the 18 colon cancer miRNAs (Supplementary Material S3). In other words, about 90% (37/41) of the HRmiRs are known to be related to cancer. These results demonstrate that HRmiRs are indeed related to cancer and provide insights into the molecular underpinnings of cancer from another perspective.

From the results, we can clearly see that our proposed miR_path can identify cancer miRNAs with high precision even without the presence of miRNA expression data, confirming again the efficiency of the proposed approach.

Table 3. The number of identified miRNAs that are conserved across distinct datasets for the same cancer type

Cancer	Index	miR_Path	miR_clust	miR_DG
Lung cancer	#miRNA	88	120	5
	C_m	0.77	0.86	0.22
Colon cancer	#miRNA	83	110	1
	C_m	0.82	0.87	0.07
Gastric cancer	#miRNA	127	163	3
	C_m	0.72	0.79	0.14
Breast cancer	#miRNA	29	393	0
	C_m	0.62	1.00	0.00

3.2 Validation of predicted cancer miRNAs

Except for the gold stand cancer-miRNA associations deposited in HMDD, to validate our predictions, we retrieved four independent miRNA expression datasets for the four cancers, where the miRNAs that were differentially expressed between normal and cancer samples were considered related to cancer. If our predicted cancer miRNAs are also differentially expressed miRNAs (DemiRs), they are more likely related to cancer, which can validate our predictions to some extent. Table 4 summarizes the number of miRNAs identified by our approach and those DemiRs as well as their overlaps, where the DemiRs were detected with paired student's *t*-test (*P*-value cutoff of 0.05) and the predicted miRNAs are those overlapped between two datasets for the same cancer type. From the results, we can see that most of our predictions can be validated by DemiRs in lung cancer and breast cancer. For gastric cancer, about half of the DemiRs can be found by our approach. All these results demonstrate that our predictions are convincible, and the results also indicate the high precision of miR_path, which is consistent with the benchmark results described above.

Moreover, we also verified our predictions by querying PubMed, and found that some of our predictions have already been reported in literature. For example, for the GSE13911 and GSE19826 datasets of gastric cancer, we respectively identified 16 and 13 miRNAs that are not recorded related to gastric cancer according to HMDD but were DemiRs. By searching PubMed, we found that some of these miRNAs have already been reported to affect gastric cancer in literature. For example, miR-29c was found to be significantly down-regulated in advanced gastric carcinoma and play a role of tumor-suppresser through its target gene RCC2 to confer a growth

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

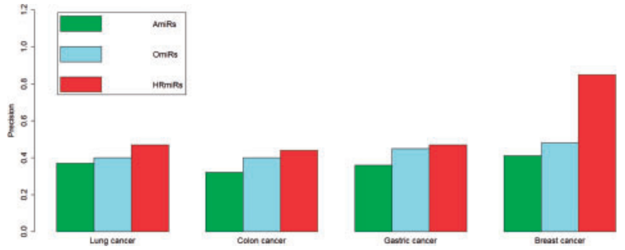

Fig. 2. Precision of miRNAs identified to be associated with different cancer types. HRmiRs: top ranked miRNAs; OmiRs: overlapping miRNAs identified from two datasets; AmiRs: all miRNAs identified to be related to certain cancer type

advantage on gastric tumor cells (Matsuo *et al.*, 2013), and miR-26a was strongly down-regulated and inhibited cell proliferation in gastric cancer (Deng *et al.*, 2013a, b).

These evidences imply that our novel predictions may be indeed related to gastric cancer.

3.3 Cancer-miRNA association network

Based on the HRmiRs identified for each cancer type, we further constructed a cancer-miRNA association network as shown in Figure 3, where a cancer type will be linked to one miRNA if this miRNA is predicted to be related to the cancer type by miR_Path. The association network gives a clear global view about the relationship between miRNAs and cancers, from which we can see that some miRNAs are associated with a specific type of cancer while some others are related to at least two cancer types.

Focusing on the miRNAs that are specifically associated with certain types of cancer, we verified these cancer type-specific miRNAs by checking whether these miRNAs are specifically expressed in the tissues that the specific cancer origins from. With the miRNA expression data across 40 normal human tissues (Liang *et al.*, 2007), we identified 16 miRNAs for colon, 15 for lung, 21 for stomach and 6 for breast (Supplementary Material S4), where these miRNAs are both specifically expressed in the corresponding tissues and predicted to be related to cancer. In addition, query results from PubMed indicate that most of these tissue-specific miRNAs have already been reported related to the corresponding cancer types. For example, mir-195 was found to be significantly down-regulated in gastric cancers and treatment with mir-195 strikingly suppressed the growth of gastric cancer cell (Deng *et al.*, 2013a, b), and mir-195 can be used as the prognosis biomarker of gastric cancer (Brenner *et al.*, 2011; Wu *et al.*, 2011). The abnormal expression of mir-137 has been found involved in the progression and metastasis of colorectal cancer and acts as a tumor suppressor in colon cancer (Balaguer *et al.*, 2010; Chen *et al.*, 2013). More evidence about the association between tissue-specific miRNAs and cancer types can be found in Table 5. The tissue specificity of cancer type-specific miRNAs implies that these miRNAs are indeed related to specific cancer types, which will otherwise not be found without the cancer-miRNA association network. The consistence between tissue-specific miRNAs and cancer type-specific miRNAs confirm again that our identified miRNAs are indeed related to cancers.

3.4 Cancer-miRNA-pathway association network

In the cancer-miRNA association network, there are some miRNAs that are associated with multiple cancers. These multiple-cancers associated miRNAs may regulate the common biological processes across diverse cancer types and therefore play important roles in cancer initiation and progression. Next, we focused on the 47 miRNAs that are associated with at least two cancer types as well as the pathways they regulate by extending the cancer-miRNA network to a cancer-miRNA-pathway network (CMP network) as shown in Figure 4, where only three outstanding miRNAs families (let-7, miR-200 and miR-29) and some related miRNAs (miR-16, miR-524-5p and miR-520d-5p) were shown for clearness. The complete network can be found in Supplementary Material S5. In Figure 4, we only showed the KEGG pathways that have been well categorized.

Table 4. The number of predicted cancer miRNAs that are differentially expressed

Cancer	miRNA dataset	miRNAs	Methods	
			miR_Path	miR_Clust
Lung cancer	GSE18692	Predicted	88	120
		DemiRs	162	162
		Overlap[a]	58 (67%)	92 (77%)
Colon cancer	GSE56350	Predicted	83	110
		DemiRs	104	104
		Overlap[a]	19 (23%)	23 (21%)
Gastric cancer	GSE28700	Predicted	127	163
		DemiRs	57	57
		Overlap[a]	25 (20%)	36 (22%)
Breast cancer	GSE45666	Predicted	29	393
		DemiRs	133	133
		Overlap[a]	24 (83%)	86 (22%)

[a]The number in the bracket denotes the percentage of predicted miRNAs that can be validated by DemiRs.

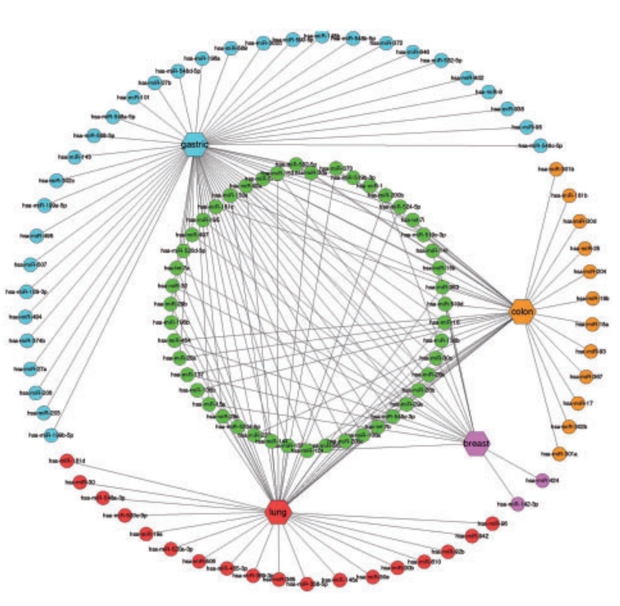

Fig. 3. The cancer-miRNA association network. In the network, the miRNAs in the inner circle are predicted related to at least two cancer types

In the network, the let-7 family is notably associated with multiple cancer types, and has been found to acts as tumor suppressors for lung and colon cancer (Akao *et al.*, 2006; Shell *et al.*, 2007; Takamizawa *et al.*, 2004). The members of the let-7 family are master regulators of cell proliferation pathways by targeting RAS and MYC oncogenes (Johnson *et al.*, 2005; Sampson *et al.*, 2007). It is found that the epithelial cells undergo abnormal cell divisions when

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

Table 5. Evidences for tissue-specific miRNAs that are related to the tissues in which the cancer initiates

Cancer	miRNAs	Evidence (PMID)
Lung cancer	miR-365	23507558, 22185756
	let-7a	23566834, 23349018, 21622546, 21097396
	miR-32	22349819
Colon cancer	miR-137	23275153, 23201162, 22895557, 20682795,19659786
	miR-181b	23719259, 18172508, 18079988
	miR-367	23393343
Gastric cancer	miR-195	23333942, 22046085, 21987613
	miR-196a	24527072
	miR-203	23790975, 21454413, 21063914
	miR-206	23751352, 23348698
	miR-372	23242208, 22027184
	miR-93	22842726
	miR-9	23383271, 22703336, 21931274, 21225631, 20102618
	miR-497	21258880
	miR-143	23932921, 21874264, 19439999
Breast cancer	miR-148b	25051376, 23233531
	miR-218	22705304, 21385904
	miR-223	25004125, 24400121
	miR-27a	25223182, 24632568, 24191129

let-7 is inactivated (Lu *et al.*, 2005), indicating the lower expression level of let-7 in tumor during cancer development. As shown in the CMP network, the let-7 family acts as tumor suppressor by regulating some important cancer-related pathways, such as P53 signaling pathway, E2F mediated regulation of DNA replication pathway, cell cycle mitotic pathway and DNA repair pathway. The miR-200 family was found to regulate the epithelial–mesenchymal transition and tumor metastasis by targeting the mRNA of E-cadherin transcriptional repressors ZEB1 and ZEB2, two important transcriptional repressors of the cell adherence and polarity (Aigner *et al.*, 2007; Comijn *et al.*, 2001). The three members, i.e. miR-29a, miR-29b and miR-29c of the miR-29 family are known as anti-oncomirs. They can impact epigenetic alteration in tumor cells (Fabbri *et al.*, 2007) by reverting DNA methylation through their direct targets, DNA methyltransferases 3A (Dnmt3a) and 3B (Dnmt3b), in lung cancer tissues. It has been found that the overexpression of miR-29a significantly reduces the invasiveness and proliferation phenotypes in lung cancer cell lines (Muniyappa *et al.*, 2009). Another study indicates that miR-29 family can up-regulate P53 by targeting P85 and CDC42 (Park *et al.*, 2009). In the CMP network, the miR-29 family regulates the PDGF pathway. PDGF regulates diverse differentiation and proliferation pathways and promotes tumor growth via autocrine stimulation of malignant cells, and PDGF receptors were found to be critical for the PI3K/AKT activation and negatively regulated by mTOR (Wu *et al.*, 2008). It can be seen that the pathways regulated by miR-29 family can help interpret how it is involved in the development of cancer.

Among the miRNAs, miR-16 was predicted to be associated with all four types of cancers. miR-16 is among the miRNAs that were firstly found to be associated with cancer development in chronic lymphocytic leukemia (Calin *et al.*, 2002), where miR-16 acts as a tumor suppressor and was found to be deleted or down-regulated in tumor cells (Calin *et al.*, 2005). As shown in the CMP network, miR-16 regulates the P53 signaling pathway and the E2F mediated regulation of DNA replication pathway, indicating the important roles of miR-16 in cancer. miR-524-5p was not only found to regulate the WNT pathway but also regulate two APC-related pathways in the CMP network, indicating the important role of miR-524-5p in tumor. miR-520d-5p was found to regulate the Notch signaling pathway which in turn regulates β-catenin to cooperate with the WNT pathway while tumor formation (Klaus and Birchmeier, 2008). It can be seen that the cross-talks between above miRNAs and pathways help us understand how the miRNAs are involved in the tumor process.

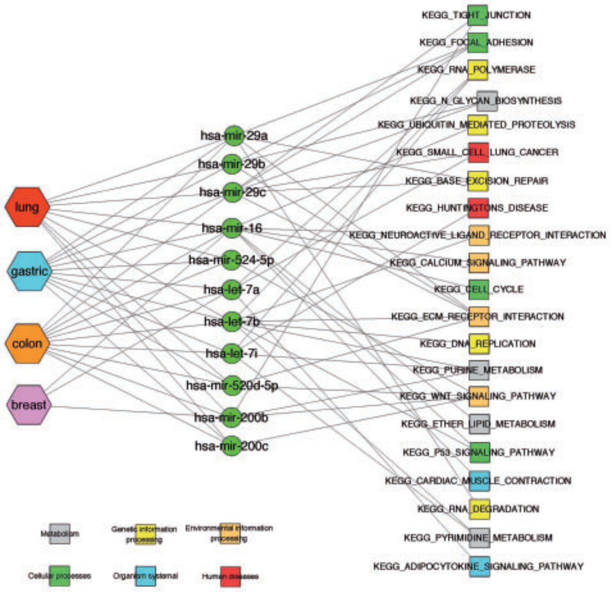
KEGG_TIGHT_JUNCTION
KEGG_FOCAL_ADHESION
KEGG_RNA_POLYMERASE
KEGG_N_GLYCAN_BIOSYNTHESIS
KEGG_UBIQUITIN_MEDIATED_PROTEOLYSIS
KEGG_SMALL_CELL_LUNG_CANCER
KEGG_BASE_EXCISION_REPAIR
KEGG_HUNTINGTONS_DISEASE
KEGG_NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION
KEGG_CALCIUM_SIGNALING_PATHWAY
KEGG_CELL_CYCLE
KEGG_ECM_RECEPTOR_INTERACTION
KEGG_DNA_REPLICATION
KEGG_PURINE_METABOLISM
KEGG_WNT_SIGNALING_PATHWAY
KEGG_ETHER_LIPID_METABOLISM
KEGG_P53_SIGNALING_PATHWAY
KEGG_CARDIAC_MUSCLE_CONTRACTION
KEGG_RNA_DEGRADATION
KEGG_PYRIMIDINE_METABOLISM
KEGG_ADIPOCYTOKINE_SIGNALING_PATHWAY
hsa-mir-29a
hsa-mir-29b
hsa-mir-29c
hsa-mir-16
hsa-mir-524-5p
hsa-let-7a
hsa-let-7b
hsa-let-7i
hsa-mir-520d-5p
hsa-mir-200b
hsa-mir-200c
lung
gastric
colon
breast
Metabolism
Genetic information processing
Environmental information processing
Cellular processes
Organism systemal
Human diseases

Fig. 4. The CMP network. The pathways are colored according to their biological categories annotated in KEGG

We also noticed that miR-548c-3p was predicted to be associated with three types of cancers although it is not reported to be related to any type of cancer in HMDD. In the CMP network, miR-548c-3p was found to regulate the cell cycle related pathways, DNA repair pathway and the stabilization of P53 pathway. Recently, miR-548c-3p was identified as a mediator of the expression of topoisomerase IIα (TOP2A) that plays critical roles in maintaining DNA topology after replication (Srikantan *et al.*, 2011). miR-548c-3p represses the expression of TOP2A by binding to its 3′ UTR region in a competition with Human antigen R (HuR) that is an RNA-binding protein stabilizing and modulating the translation of its target mRNAs. Recently, HuR was found to be commonly overexpressed in cancers (Abdelmohsen and Gorospe, 2010; Dixon *et al.*, 2001; Lopez de Silanes *et al.*, 2005). Therefore, we assumed that miR-548c-3p is related to cancer by regulating TOP2A through the competition with HuR, where the dysregulation of TOP2A will affect the DNA stabilization. Our prediction is also supported by a recent study that reports mir-548c-3p was significantly down-regulated in different pathological grades of hepatocellular carcinoma (Noh *et al.*, 2013).

In summary, the CMP network constructed here provides insights into the molecular underpinnings of cancer and can help to interpret how the miRNAs are involved in the pathogenesis of cancers through the cellular pathways that they regulate.

4 Discussion

The miRNAs are important regulators of gene expression, the aberrant function of which may drive the initiation and development of cancer. However, the global cancer-miRNA association landscape is far from complete and only a limited number of miRNAs are known

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

related to cancer, which hinders the development of miRNA based therapeutic strategies. In literature, some computational approaches have been developed to detect which miRNAs are associated with which cancer types, among which the integrative analysis of matched gene and miRNA expression profiles is one of the most promising approaches. Unfortunately, the matched gene and miRNA expression profiles are not commonly available. On the other hand, the large amount of public accessible gene expression data provides an alternative way to investigate the associations between miRNAs and cancers. In general, the existing approaches utilize the predicted miRNA-gene regulations to identify the cancer associated miRNAs, whereas the noise in the regulation relationship as well as the lack of the context information about these regulations may degrade the performance of those promising approaches.

In this work, we developed a novel framework to identify cancer miRNAs based on gene expression data and miRNA-gene regulations. Unlike existing approaches, we proposed a new technique to filter out those noisy and false regulations based on gene expression data, and detected cancer miRNAs. The benchmark results on four distinct cancer datasets demonstrate the predictive power of our proposed method, and the good performance on multiple independent datasets for the same cancer type indicates the robustness of the proposed approach. Furthermore, among our predicted cancer miRNAs, some of them have been reported in literature and some are found to be differentially expressed based on independent miRNA expression datasets, which validate our predictions to some extent. With a cancer-miRNA association network constructed here, we also detected some pan-cancer miRNAs and cancer type-specific miRNAs, thereby providing insights into the mechanism of cancers. In particular, we identified the pathways regulated by cancer miRNAs, which can help explain how miRNAs are involved in cancers.

The good performance of our proposed approach attributes to the removal of noisy interactions between miRNAs and genes. In this work, we detected the context specific target genes of miRNAs with the assumption that they will fluctuate together if they were regulated by the same miRNA(s). However, that is not the case if two genes are, respectively, regulated by two sets of miRNAs while share few miRNAs, where we can only detect the major regulators of these genes. The integrative analysis of matched miRNA and gene expressions has shown to be promising approach in determining the miRNA-gene regulations. We believe the performance of our proposed approach will be significantly improved if such kind of data could be integrated in our framework in the future. In addition, we only considered around 500 miRNAs here, but there are more than 2000 human mature miRNAs recorded in miRBase now. Our approach can be easily extended to the new miRNAs as long as their target genes are available. With more information about miRNAs as well as their target genes emerging, more cancer miRNAs are believed to be discovered in the future.

Funding

This work was partly supported by Strategic Priority Research Program of the Chinese Academy of Sciences [XDB13040700]; National Natural Science Foundation of China [91130032, 61103075]; Innovation Program of Shanghai Municipal Education Commission [13ZZ072] and Shanghai Pujiang Program [13PJD032].

Conflict of Interest: none declared.

References

Abdelmohsen,K. and Gorospe,M. (2010) Posttranscriptional regulation of cancer traits by HuR. *Wiley Interdiscip. Rev. RNA*, **1**, 214–229.

Aigner,K. *et al.* (2007) The transcription factor ZEB1 (deltaEF1) promotes tumour cell dedifferentiation by repressing master regulators of epithelial polarity. *Oncogene*, **26**, 6979–6988.

Akao,Y. *et al.* (2006) let-7 microRNA functions as a potential growth suppressor in human colon cancer cells. *Biol. Pharm. Bull.*, **29**, 903–906.

Asangani,I.A. *et al.* (2008) MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. *Oncogene*, **27**, 2128–2136.

Balaguer,F. *et al.* (2010) Epigenetic silencing of miR-137 is an early event in colorectal carcinogenesis. *Cancer Res.*, **70**, 6609–6618.

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

Brenner,B. *et al.* (2011) MicroRNAs as a potential prognostic factor in gastric cancer. *World J. Gastroenterol.*, **17**, 3976–3985.

Calin,G.A. and Croce,C.M. (2006) MicroRNA signatures in human cancers. *Nat. Rev. Cancer*, **6**, 857–866.

Calin,G.A. *et al.* (2002) Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc. Natl Acad. Sci. USA*, **99**, 15524–15529.

Calin,G.A. *et al.* (2005) A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N. Engl. J. Med.*, **353**, 1793–1801.

Chen,D.L. *et al.* (2013) Overexpression of paxillin induced by miR-137 suppression promotes tumor progression and metastasis in colorectal cancer. *Carcinogenesis*, **34**, 803–811.

Chen,X. *et al.* (2012) RWRMDA: predicting novel human microRNA–disease associations. *Mol. Biosyst.*, **8**, 2792–2798.

Comijn,J. *et al.* (2001) The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion. *Mol. Cell*, **7**, 1267–1278.

Croce,C.M. (2009) Causes and consequences of microRNA dysregulation in cancer. *Nat. Rev. Genet.*, **10**, 704–714.

D'Errico,M. *et al.* (2009) Genome-wide expression profile of sporadic gastric cancers with microsatellite instability. *Eur. J. Cancer*, **45**, 461–469.

Deng,H. *et al.* (2013a) MicroRNA-195 and microRNA-378 mediate tumor growth suppression by epigenetical regulation in gastric cancer. *Gene*, **518**, 351–359.

Deng,M. *et al.* (2013b) miR-26a suppresses tumor growth and metastasis by targeting FGF9 in gastric cancer. *PLoS One*, **8**, e72662.

Dixon,D.A. *et al.* (2001) Altered expression of the mRNA stability factor HuR promotes cyclooxygenase-2 expression in colon cancer cells. *J. Clin. Invest.*, **108**, 1657–1665.

Edgar,R. *et al.* (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, **30**, 207–210.

Esteller,M. (2011) Non-coding RNAs in human disease. *Nat. Rev. Genet.*, **12**, 861–874.

Fabbri,M. *et al.* (2007) MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. *Proc. Natl Acad. Sci. USA*, **104**, 15805–15810.

Graham,K. *et al.* (2010) Gene expression in histologically normal epithelium from breast cancer patients and from cancer-free prophylactic mastectomy patients shares a similar profile. *Br J Cancer* 2010, **102**, 1284–1293.

Griffiths-Jones,S. *et al.* (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140–D144.

Hanahan,D. and Weinberg,R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**, 646–674.

Hong,Y. *et al.* (2010) A 'metastasis-prone' signature for early-stage mismatch-repair proficient sporadic colorectal cancer patients and its implications for possible therapeutics. *Clin. Exp. Metastasis*, **27**, 83–90.

Irizarry,R.A. *et al.* (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4**, 249–264.

John,B. *et al.* (2004) Human MicroRNA targets. *PLoS Biol.*, **2**, e363.

Johnson,S.M. *et al.* (2005) RAS is regulated by the let-7 microRNA family. *Cell*, **120**, 635–647.

Kanehisa,M. *et al.* (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res.*, **36**, D480–D484.

Klaus,A. and Birchmeier,W. (2008) Wnt signalling and its impact on development and cancer. *Nat. Rev. Cancer*, **8**, 387–398.

Krek,A. *et al.* (2005) Combinatorial microRNA target predictions. *Nat. Genet.*, **37**, 495–500.

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

Kuo,T.Y. *et al.* (2012) Computational analysis of mRNA expression profiles identifies microRNA-29a/c as predictor of colorectal cancer early recurrence. *PLoS One*, **7**, e31587.

Landi,M.T. *et al.* (2008) Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS One*, **3**, e1651.

Lee,E.J. *et al.* (2008) Systematic evaluation of microRNA processing patterns in tissues, cell lines, and tumors. *RNA*, **14**, 35–42.

Lewis,B.P. *et al.* (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15–20.

Li,X. *et al.* (2011) Prioritizing human cancer microRNAs based on genes' functional consistency between microRNA and cancer. *Nucleic Acids Res.*, **39**, e153.

Liang,Y. *et al.* (2007) Characterization of microRNA expression profiles in normal human tissues. *BMC Genomics*, **8**, 166.

Liberzon,A. *et al.* (2011) Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, **27**, 1739–1740.

Lim,L.P. *et al.* (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, **433**, 769–773.

Liu,K.Q. *et al.* (2012) Identifying dysregulated pathways in cancers from pathway interaction networks. *BMC Bioinformatics*, **13**, 126.

Lopez de Silanes,I. *et al.* (2005) HuR: post-transcriptional paths to malignancy. *RNA Biol.*, **2**, 11–13.

Lu,J. *et al.* (2005) MicroRNA expression profiles classify human cancers. *Nature*, **435**, 834–838.

Lu,M. *et al.* (2008) An analysis of human microRNA and disease associations. *PLoS One*, **3**, e3420.

Maragkakis,M. *et al.* (2009) Accurate microRNA target prediction correlates with protein repression levels. *BMC Bioinformatics*, **10**, 295.

Matsuo,M. *et al.* (2013) MiR-29c is downregulated in gastric carcinomas and regulates cell proliferation by targeting RCC2. *Mol. Cancer*, **12**, 15.

Matthews,L. *et al.* (2009) Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.*, **37**, D619–D622.

Muniyappa,M.K. *et al.* (2009) MiRNA-29a regulates the expression of numerous proteins and reduces the invasiveness and proliferation of human carcinoma cell lines. *Eur. J. Cancer*, **45**, 3104–3118.

Noh,J.H. *et al.* (2013) MiR-145 functions as a tumor suppressor by directly targeting histone deacetylase 2 in liver cancer. *Cancer Lett*, **335**, 455–462.

Oulas,A. *et al.* (2011) Computational identification of miRNAs involved in cancer. *Methods Mol. Biol.*, **676**, 23–41.

Pau Ni,I.B. *et al.* (2010) Gene expression patterns distinguish breast carcinomas from normal breast tissues: the Malaysian context. *Pathol Res Pract*, **206**, 223–228.

Park,S.Y. *et al.* (2009) miR-29 miRNAs activate p53 by targeting p85 alpha and CDC42. *Nat. Struct. Mol. Biol.*, **16**, 23–29.

Sampson,V.B. *et al.* (2007) MicroRNA let-7a down-regulates MYC and reverts MYC-induced growth in Burkitt lymphoma cells. *Cancer Res.*, **67**, 9762–9770.

Sethupathy,P. *et al.* (2006) TarBase: a comprehensive database of experimentally supported animal microRNA targets. *RNA*, **12**, 192–197.

Shell,S. *et al.* (2007) Let-7 expression defines two differentiation stages of cancer. *Proc. Natl Acad. Sci. USA*, **104**, 11400–11405.

Skrzypczak,M. *et al.* (2010) Modeling oncogenic signaling in colon tumors by multidirectional analyses of microarray data directed for maximization of analytical reliability. *PLoS One*, **5**, e13091.

Srikantan,S. *et al.* (2011) Translational control of TOP2A influences doxorubicin efficacy. *Mol. Cell Biol.*, **31**, 3790–3801.

Su,L.J. *et al.* (2007) Selection of DDX5 as a novel internal control for Q-RT-PCR from microarray data using a block bootstrap re-sampling scheme. *BMC Genomics*, **8**, 140.

Takamizawa,J. *et al.* (2004) Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res.*, **64**, 3753–3756.

Wang,Q. *et al.* (2012) Upregulated INHBA expression is associated with poor survival in gastric cancer. *Med. Oncol.*, **29**, 77–83.

Wu,E. *et al.* (2008) Comprehensive dissection of PDGF-PDGFR signaling pathways in PDGFR genetically defined cells. *PLoS One*, **3**, e3794.

Wu,W.Y. *et al.* (2011) Potentially predictive microRNAs of gastric cancer with metastasis to lymph node. *World J. Gastroenterol.*, **17**, 3645–3651.

Wuchty,S. *et al.* (2013) Important miRs of pathways in different tumor types. *PLoS Comput. Biol.*, **9**, e1002883.

Yanaihara,N. *et al.* (2006) Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell*, **9**, 189–198.

Zhang,W. *et al.* (2014) Identification of candidate miRNA biomarkers from miRNA regulatory network with application to prostate cancer. *J. Transl. Med.*, **12**, 66.

Zhao,X.M. *et al.* (2008) Uncovering signal transduction networks from high-throughput data by integer linear programming. *Nucleic Acids Res.*, **36**, e48.

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016