Published online 16 June 2012 *Nucleic Acids Research, 2012, Vol. 40, No. 16* ***7653–7665***
doi:10.1093/nar/gks538

Prioritizing cancer-related key miRNA–target interactions by integrative genomics

Yun Xiao[1], Jinxia Guan[1], Yanyan Ping[1], Chaohan Xu[1], Teng Huang[1], Hongying Zhao[1], Huihui Fan[1], Yiqun Li[1], Yanling Lv[1], Tingting Zhao[2], Yucui Dong[3], Huan Ren[3] and Xia Li[1,*]

[1]College of Bioinformatics Science and Technology, [2]Department of Neurology, The Affiliated Hospital and [3]Department of Immunology, Harbin Medical University, Harbin, Heilongjiang 150086, China

Received December 28, 2011; Revised May 8, 2012; Accepted May 13, 2012

ABSTRACT

Accumulating evidence indicates that microRNAs (miRNAs) can function as oncogenes or tumor suppressor genes by controlling few key targets, which in turn contribute to the pathogenesis of cancer. The identification of cancer-related key miRNA–target interactions remains a challenge. We performed a systematic analysis of known cancer-related key interactions manually curated from published papers based on different aspects including sequence, expression and function. Known cancer-related key interactions show more miRNA binding sites (especially for 8mer binding sites), more reliable binding of miRNA to the target region, higher expression associations and broader functional coverage when compared to non-disease-related interactions. Through integrating these sequence, expression and function features, we proposed a bioinformatics approach termed PCmtI to prioritize cancer-related key interactions. Ten-fold cross-validation of our approach revealed that it can achieve an area under the receiver operating characteristic curve of 93.9%. Subsequent leave-one-miRNA-out cross-validation also demonstrated the performance of our approach. Using miR-155 as a case, we found that the top ranked interactions can account for most functions of miR-155. In addition, we further demonstrated the power of our approach by 23 recently identified cancer-related key interactions. The approach described here offers a new way for the discovery of novel cancer-related key miRNA–target interactions.

INTRODUCTION

MicroRNAs (miRNAs) are single-stranded RNAs consisting of ~22 nt. They play important roles in the post-transcriptional regulation of gene expression by translation repression and mRNA decay based on partially base-paring to the 3′ untranslated regions (UTRs) of their target messenger RNAs (mRNAs). During the last few years, many studies have highlighted the roles of miRNAs in many cancer-related processes including apoptosis, proliferation, survival and metastasis. Dysfunction of miRNAs leads to the abnormality of their downstream targets, which, in turn, can cause cancer development. Therefore, identifying cancer-related miRNA–target interactions is pivotal for understanding how miRNAs acting as oncogenes or tumor suppressor genes are involved in the pathogenesis of cancer.

Despite recent advances in identifying miRNAs associated with cancer (1) and developing corresponding bioinformatics methods (2), the discovery of the cancer-related miRNA–target interactions is still lagging. Experimental evidence indicates that the regulation of few key targets can largely explain the functions of individual miRNAs (3). For example, two studies have recently revealed that targeted mutagenesis of miR-155 binding sites in the 3′UTR of the AID gene could lead to the similar phenotypes of deletion of miR-155 itself (4,5). The miR-15a and miR-16-1 cluster, residing in the 13q14 chromosome region, was found to be frequently deleted in chronic lymphocytic leukemia (CLL). Further experiments demonstrated that the cluster can target an oncogene BCL2. Loss of the cluster in CLL leads to the over-expression of BCL2, which subsequently triggers the initiation of most CLL (6). Although many studies have demonstrated the cooperative effects of multiple miRNAs to 'fine-tune' gene expression (7,8), many-to-many regulatory relations are more difficult to be studied and

*To whom correspondence should be addressed. Tel: +86 451 86615922; Fax: +86 451 86615922; Email: lixia@hrbmu.edu.cn

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2012. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

experimentally proven than one-to-one functional relations. Furthermore, because miRNAs generally have many targets, experiments used for the discovery of these cancer-related key miRNA–target interactions can be time-consuming and laborious. Thus, there is a substantial need for a method of prioritizing cancer-related key miRNA–target interactions.

It is worth noting that very little is known about the properties of cancer-related key miRNA–target interactions. Because different types of seed matched sites, ranging in length from 6 nt to 8 nt (i.e. canonical targets), are corresponding to different site efficiencies, the type and the number of binding sites may provide important clues to identifying key interactions. For example, the lin-4 miRNA has been found to control the developmental timing of the *Caenorhabditis elegans* by regulating the expression of the protein-coding gene lin-14 (9,10). Although hundreds of targets are predicted for lin-4, genetic experiments showed that the lin-4:lin-14 is the most important interaction, because mutations of lin-4-binding sites in lin-14 phenocopy mutations of lin-4 (11). Target-prediction results showed that the target with the highest number of binding sites among all predicted targets of lin-4 is lin-14, and that all binding sites in the 3′UTR of lin-14 belong to the 8mer sites (12). In addition to canonical targets for miRNAs, recent studies also found that miRNAs have non-canonical targets that are not dependent on the seed sequence and generally show more extensive base pairing. However, these non-canonical targets only play modest roles in miRNA function (12).

The integration of miRNA and mRNA expression profiles has been widely used to improve miRNA-target detection (13), because expression correlations between miRNAs and their corresponding targets can partially reflect the efficiency of interactions (14). More importantly, many experiments revealed that miRNAs modulate the concentration of key target proteins in a dose-dependent manner (15). For example, a dose-dependent development block mediated by the transfection of miR-150 in mice was found to be mainly caused by the down-regulation of its key target, c-Myb (16). A recent study also showed that changes in the mRNA levels closely reflect the influence of miRNAs on gene expression (14). Therefore, expression relationships between miRNAs and their targets may also provide clues to finding key interactions. In addition, considering the fact that loss of binding sites in few key targets for a specific miRNA can phenocopy most aspects of the miRNA mutations, we suspected that the majority of the functions of the miRNA should depend on the interactions with these key targets, that is, the functions of these few targets are sufficient to capture most functions of the miRNA. Thus, functional associations of miRNAs with their targets can be an important factor for identifying key interactions.

In this study, we systematically analyzed sequence, expression and function features of known cancer-related key miRNA–target interactions manually curated from >3000 literatures. By integrating these different features, we proposed an approach, termed PCmtI, to prioritize cancer-related key miRNA–target interactions. Our method produced good predictions on the known cancer-related key interactions by 10-fold cross-validation and leave-one-miRNA-out cross-validation. We also demonstrated that prioritization using integrated features significantly outperforms those using individual features. Our results suggest that PCmtI can help biologist to find novel cancer-related key miRNA–target interactions. We made our approach freely available on the web at http://bioinfo.hrbmu.edu.cn/PCmtI.

MATERIALS AND METHODS

Data sources

Mature miRNA sequences, miRNA family and cluster data were obtained from miRBase (release 16) (17). The annotated 3′UTR of each transcript of a gene was downloaded from UCSC Genome Browser (hg18, http://genome.ucsc.edu/) (18), and then the longest 3′UTR of the gene was used to search for different types of binding sites of miRNAs. Predicted conserved and non-conserved targets of miRNAs were obtained using TargetScan (19). The atlas gene expression data from 79 normal human tissues was downloaded from Gene Expression Omnibus (GEO; GSE1133) (20). Four paired miRNA and mRNA expression data sets were downloaded from GEO (multiple myeloma: GSE17306 and prostate cancer: GSE25692) and The Cancer Genome Atlas (TCGA) (21–24). The human protein–protein interaction network was obtained from HPRD (25). Gene annotation information about molecular function, biological process and cellular component was obtained from gene ontology (GO) (26). Pathway information was downloaded from KEGG (27). We got 797 disease-related miRNAs from miR2Disease (28), HMDD (29) and dbDEMC (30), and 8995 disease-related genes from OMIM (31), GAD (32) and CGC (33).

Experimentally validated cancer-related key miRNA–target interactions were collected as positive interactions by manually curating >3000 literatures. To construct a *bona fide* set of negative interactions, we chose miRNA–target interactions in which both miRNAs and their targets are not associated with any disease. It should be clear that miRNAs with expression values in four paired miRNA–mRNA expression data sets, and genes with expression values in both the atlas and paired expression data sets and with GO, KEGG and network annotations were selected for the following analysis.

Feature extractions

For a given cancer-related key miRNA–target interaction, most of the abnormal phenotypes mediated by the dysfunction of the miRNA are caused by this miRNA–target interaction. The miRNA should effectively control the target in order to make the target highly responsive to expression changes of the miRNA. Most importantly, the key target should account for most functions of the miRNA, that is, the target may be a functional hub among all targets of the miRNA. Therefore, for a given miRNA–target interaction, we construct different types of features

based on sequence, expression and function information as follows.

Sequence features

Because different types of binding sites have different site efficiencies and the regions beyond the seed pairing also contribute to site efficiency, we thus considered sequence features including the numbers and types of binding sites in the 3′UTR, and site context. Our study focused on four types of binding sites: one is the 6mer site, which perfectly matches the 6-nt miRNA seed; another is the 7mer-m8 site, for which seed paring is supplemented by a Watson–Crick match to miRNA nucleotide 8; the third is the 7mer-A1 site, for which seed pairing is supplemented by an A across from miRNA nucleotide 1; the forth is the site with both the m8 and A1 match, which is called 8mer site. In general, 8mer sites are more effective than 7mer sites, which are more effective than 6mer sites (34). We calculated the numbers of different types of binding sites and a 'context score' obtained from TargetScan.

Recent experiments highlighted that miRNAs in the same family or in the same cluster can be cooperatively involved in many important biological processes (7,35,36). We reasoned that miRNAs belonging to the same family or cluster may help to control the key targets. Therefore, we recorded the family and cluster related to the miRNA. A miRNA cluster is defined as a set of miRNAs separated by less than 10 kb. Then, we calculated the total number of each type of binding site for all miRNAs belonging to the family, and the total number of each type of binding site for all miRNAs belonging to the cluster.

Expression features

Co-expression of genes has been widely used to predict gene functions based on an assumption that co-expressed genes tend to have similar functions (37). To characterize the functional essentiality of the target from the perspective of gene expression, we calculated the average Pearson correlation between this target and all other targets of the miRNA using the atlas expression data set that has been widely used to infer functional relationships between genes (38,39). In addition, expression relationships between miRNAs and their targets are often used to delineate the regulatory effects of miRNAs on their targets (40). Based on the four paired miRNA and mRNA expression data sets, we calculated the Pearson correlations between the miRNA and the target.

Function features

Like expression data, we used human protein interaction network, GO and KEGG resources to further characterize the functional essentiality of the target among all targets of the miRNA. Using the human protein interaction network, we calculated the average shortest distance between the target and all other targets of the miRNA. As for GO, we considered each of the three GO sub-ontologies (i.e. molecular function, biological process and cellular component). For each GO sub-ontology, we determined significantly over-represented GO terms in all targets of the miRNA through GO enrichment analyses based on the hypergeometric distribution test. Subsequently, we obtained GO terms annotated for the target from the GO database and then calculated the percentage of GO terms of the target in all enriched GO terms associated with the miRNA. Similarly, we determined the biological pathways significantly over-represented in all targets of the miRNA, and then calculated the percentage of pathways associated with the target in all enriched pathways.

Taken together, for a miRNA–target interaction, we obtained sequence features (including the numbers of 8mer, 7mer-m8, 7mer-A1 and 6mer sites as well as a context score for this interaction, the total number of each type of binding site in the miRNA family, the total number of each type of binding site in the miRNA cluster), expression features (including the average expression correlation between the target and all other targets, and expression correlations between the miRNA and the target), and function features (including the average shortest distance between the target and all other targets, and the functional coverage of the target among all targets based on three GO sub-ontologies and KEGG pathways). These features were used to construct a model for prioritization of cancer-related key interactions.

The PCmtI model

PCmtI integrates all of the sequence, expression and function features described above for prioritizing cancer-related key miRNA–target interactions. Due to the large size difference between the positive and negative interactions, we randomly constructed 1000 negative sets with the same number of interactions as in the positive set from the negative interactions. On the basis of these features, 1000 SVM classifiers were built using the positive set and the 1000 negative sets. We combined the outputs of these 1000 classifiers, and computed an average prediction score for a specific miRNA–target interaction. The prediction score was used to rank miRNA–target interactions. Interactions with high prediction scores would have higher possibility to be cancer-related key interactions.

Performance evaluation

To evaluate the performance of PCmtI using individual or all features, we applied the 10-fold cross-validation method. The positive and negative interactions were randomly and evenly divided into 10 groups, with each group having the same numbers of positive and negative samples. In each validation run, one random group was regarded as the testing set, and the rest nine were regarded as the training set. Prediction scores of interactions in the testing set were calculated using a PCmtI model created based on the training set. We plotted a receiver operating characteristic (ROC) curves and calculated the area under the ROC (AUC).

We also used the leave-one-miRNA-out cross-validation to further assess the performance. For each miRNA included in the positive interactions, all interactions associated with the miRNA were selected as the testing set. The positive and negative interactions not associated with the miRNA were used to construct a

PCmtI model. Using the PCmtI model, we calculated prediction scores for all interactions in the testing set. According to the prediction scores, we ranked the interactions in the testing set in a descending order, and retrieved the relative ranks of the known cancer-related key interactions.

3′UTR luciferase reporter assays

3′UTR fragments of PAK7, TCF4 or FOXO3 containing the putative binding sites for miR-155 were subcloned into pGL3 luciferase reporter vectors. Respective counterparts carrying mutated sequences in the complementary binding sites for the seed regions of miR-155 at the 3′UTRs of the above genes were also constructed (JIN SIRUI Inc. Nanjing, China). For luciferase reporter assays, human malignant glioma cells LN18 (ATCC) were seeded onto 12-well plates and co-transfected with optimized 40 pmol miR-155 (or control miRNA) and 1.6 μg one of the constructed vectors when the cells reached 60–70% confluence, using Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA). miR-155 (hsa-miR-155 mimics) and control miRNA mimics were synthesized and purchased from Invitrogen. Assays were performed on a Multi-Mode Microplate Readers (M5, Molecular Devices, Inc., Sunnyvale, CA, USA) 48 h after transfection using the Luciferase Reporter Assay System (Promega, Beijing, China) according to the manufacturer's protocol. Each set of experiments was repeated at least four times.

RESULTS

Features of cancer-related key miRNA–target interactions

We manually retrieved 210 high-confidence cancer-related key miRNA–target interactions involving 91 miRNAs and 122 genes from >3000 papers (Supplementary Table S1). These cancer-related key interactions, which have been experimentally demonstrated to play key roles in miRNA-mediated cancer development, were defined as the positive interactions. In order to obtain the negative ones, the interactions in which miRNAs and their targets are not found to be related to disease were selected. Then, 8433 negative interactions were obtained. For every interaction, we subsequently extracted different sequence, expression and function features (Figure 1).

Different types of seed-matched sites (i.e. 8mer, 7mer-m8, 7mer-A1 and 6mer site) and multiple matches to the same target are important for the efficiency of miRNA–target interactions. Comparing the numbers of different types of binding sites between the positive and negative interactions, we found significant differences in the numbers of 8mer and 7mer-A1, and the total number of all types of binding sites [Figure 2A; $P = 6.66e^{-16}$, 0.048 and $9.05e^{-05}$, respectively, two-side Kolmogorov Smirnov (KS) test]. The numbers of 7mer-m8 binding sites in the positive interactions were slightly different from those in the negative interactions ($P = 0.09$, two-side KS test). There was no significant difference in the numbers of 6mer binding sites. Specially, the positive interactions exhibit significantly more 8mer binding sites ($P = 3.13e^{-16}$, one-side KS test) compared with the negative interactions, suggesting strong site efficiency of cancer-related key interactions.

As expected, due to similar targets of miRNAs in the same family, the numbers of binding sites for the same family showed similar tendency: obvious differences for all types of binding sites excluding 6mer binding sites ($P < 0.001$, two-side KS test), and more 8mer binding sites ($P = 3.62e^{-14}$, one-side KS test; Figure 2B). By considering miRNAs belonging to the same cluster, we also revealed that the cancer-related key interactions tended to have more 8mer binding sites compared to the negative interactions (Figure 2C; $P = 0.002$, one-side KS test), which may be attributed to the fact that many miRNAs in the same cluster also belong to the same family, suggesting that these co-clustering miRNAs possibly sharing a common transcriptional unit can help to regulate the key target. In addition, for each interaction, we gained a context score, a metric proposed in (34), characterizing the binding site efficiency by combining the contribution of site context features, such as 3′ pairing contribution and local AU contribution. We revealed that the context scores of the positive interactions were significantly lower than those of the negative interactions ($P < 2.2e^{-16}$, one-side KS test), suggesting higher binding affinity in the cancer-related key interactions.

Expression of miRNAs and their targets can further enhance understanding of the cancer-related key interactions. We downloaded the atlas expression data (20), and four paired miRNA and mRNA expression data sets referring to different types of cancers [including glioblastoma (GBM), ovarian cancer, multiple myeloma and prostate cancer] from GEO and TCGA. For each interaction, the average expression correlation between the target and all other targets of the miRNA was calculated using the atlas expression data. We found a modest difference between the positive and negative interactions, although without statistical significance (Figure 2D; $P = 0.055$, two-side KS test). Using the four paired miRNA and mRNA expression data sets, we observed that the expression correlations of the positive interactions were significantly different from those of the negative interactions in two data sets ($P = 1.02e^{-08}$ and $3.34e^{-10}$ for GBM and ovarian cancer, respectively, two-side KS test), and slightly different in the expression data set of multiple myeloma ($P = 0.074$, two-side KS test; Figure 2D).

Subsequently, we used the human protein interaction network, GO and KEGG annotation information to further explore functional distinctions between the positive and negative interactions. For each interaction, we calculated the average shortest distance between the target and all other targets relevant to the corresponding miRNA using the human protein interaction network. The distances in the positive interactions were significantly shorter than those in the negative interactions (Figure 2E; $P < 2.2e^{-16}$, one-side KS test). With regard to GO, we used three separate sub-ontologies (i.e. molecular function, biological process and cellular component) to characterize the functional essentiality of miRNA–target interactions. For each interaction, we created a measure to

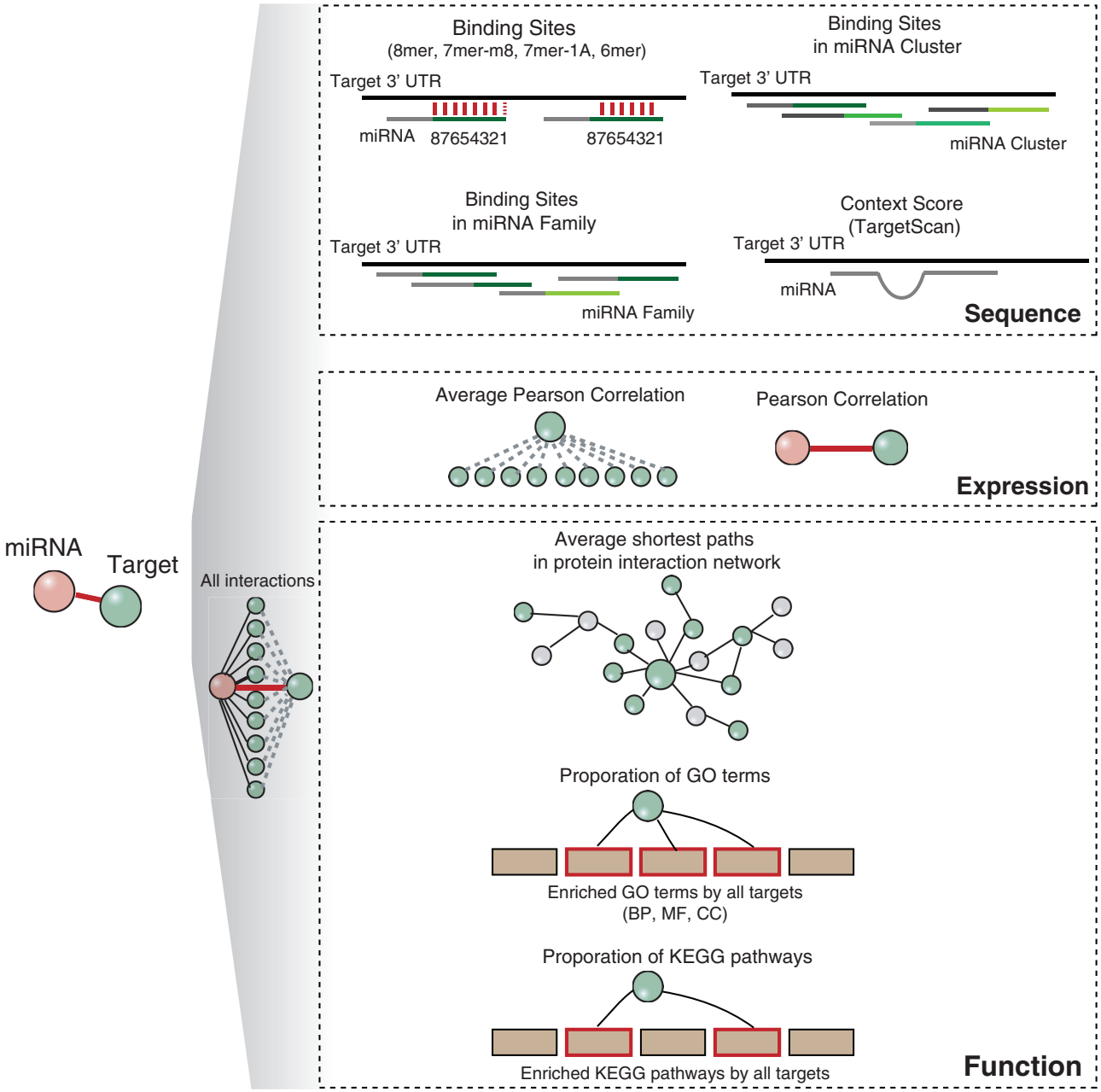

Figure 1. Sequence, expression and function features. Given a miRNA–target interaction, the numbers of different types of binding sites of this miRNA on the 3′UTR of the target were calculated. Similarly, the overall numbers of each type of binding site for its neighboring co-clustering miRNAs and miRNAs from the same family were separately calculated. The context score of the interaction was obtained from TargetScan algorithm. For expression features, Pearson correlations between the miRNA and the target were calculated using four matched miRNA and mRNA expression data sets. An average expression correlation between the target and all other targets associated with the miRNA was also calculated based on the atlas expression data set. Using the human protein interaction network, we calculated the average shortest distance between the target and all other targets associated with the miRNA. As for GO, the proportion of the GO terms related to the target among all GO terms significantly over-represented in all targets of the miRNA was calculated. Similarly, we calculated the proportion of the pathways related to the target using KEGG.

assess the functional coverage of the interaction (i.e. the extent to which the target can account for the functions of the miRNA). The GO-based score is represented by the percentage of GO terms annotated for the target in significantly enriched GO terms determined using all targets of the corresponding miRNA. The GO-based scores in the positive interactions were significantly higher than those in the negative interactions (Figure 2E; $P < 0.001$ for three GO sub-ontologies, one-side KS test). Likewise, we calculated a KEGG-based score for each interaction using a similar method used in GO analyses. The KEGG-based scores in the positive interactions were also significantly higher than those in the negative interactions (Figure 2E; $P < 2.2e^{-16}$, one-side KS test). Overall,

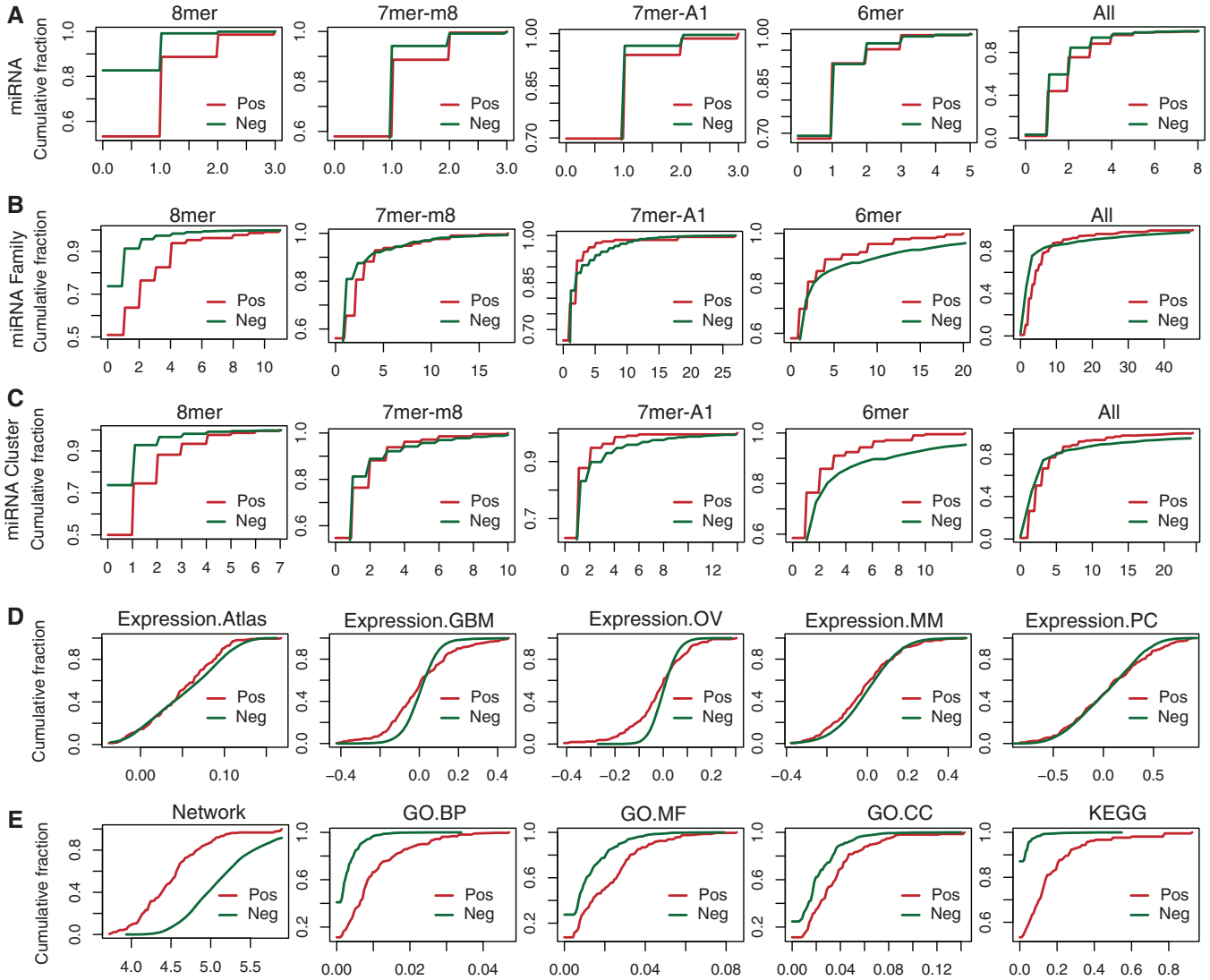

Figure 2. Comparisons of sequence, expression and function features between the positive and negative interactions. (**A**) Cumulative distributions of the numbers of different types of binding sites for the positive interactions (red) and the negative interactions (green). Cumulative distributions of the numbers of different types of binding sites when considering miRNAs in the same family (**B**) or cluster (**C**). (**D**) Cumulative distributions of the expression coherence calculated using the atlas expression data set and four paired miRNA and mRNA expression data sets for the positive interactions (red) and the negative interactions (green). (**E**) Cumulative distributions of the average shortest distance from the human protein interaction, and functional coverage from GO and KEGG for the positive interactions (red) and the negative interactions (green). GBM, OV, MM and PC represent GBM, ovarian cancer, multiple myeloma and prostate cancer, respectively. BP, MF and CC represent three GO sub-ontologies: biological process, molecular function and cellular component, respectively.

these results suggest that cancer-related key miRNA–target interactions harbor strong functional links.

PCmtI: an approach based on integrative genomics

We developed a method called PCmtI for prioritization of cancer-related key miRNA–target interactions. The 210 positive and 8344 negative interactions were used to train the PCmtI model. Considering the large size difference between the positive and negative interactions, we applied an integrated strategy for constructing the PCmtI model. From the negative interactions, we randomly chose 1000 sets with the same size as the positive interactions. Using the positive interaction set and each randomly selected negative set, a SVM classifier was generated. Ultimately, we constructed 1000 SVM classifiers, which were subsequently assembled to build a classifier cluster. The output of each classifier in the cluster was combined to generate an average prediction score representing the possibility of an interaction to be a cancer-related key interaction (see 'Materials and Methods' section for details).

Validation of PCmtI using individual and integrated features

For each type of feature, we assessed whether our approach is capable of prioritizing interactions known to be involved in cancer using 10-fold cross-validation. Obviously, using single features, PCmtI reached higher

AUC scores for the cancer-related key interactions than for randomly selected interactions (Figure 3A). Among all of these features, sequence-based features provided the highest AUC score of 82.4% (When excluding miRNA family and miRNA cluster related features, the AUC score was reduced to 73.5%). GO and network-based features also offered high AUC scores (80.2 and 81.5%, respectively). These results suggest that these features differ in their usefulness, and that combination of these features may further improve the performance.

To increase the performance of our approach, we integrated these sequence, expression and function features and re-evaluated the performance of our approach using 10-fold cross-validation. As expected, the AUC scores were 93.9% for cancer-related key interactions compared to 49.8% for randomly selected interactions. Obviously, using all of these features performed better than using single features (Figure 3B), suggesting that integration of multiple genomic features can be used to effectively prioritize cancer-related key interactions.

Leave-one-miRNA-out cross-validation

To further validate our approach, we evaluated the performance using leave-one-miRNA-out cross-validation. This cross-validation was performed for each miRNA included in the known cancer-related key interactions. Since different miRNAs have different numbers of interactions, the ranks were transformed into the relative ranks. If the known cancer-related key interaction was at the top of the list, it was assigned a relative rank of 1.0, and if at the bottom, it was assigned a relative of rank of 0.0.

A total of 91 prioritizations (referring to 91 miRNAs and 210 known cancer-related key interactions) were performed. The average relative rank is 0.86. In Figure 4, the distribution of relative ranks shows a strong right-leaning trend. About 74.8% of relative ranks of known cancer-related key interactions were at 0.8 to 1.0, and only 7.1% were less than 0.5.

Validation using miR-155 as a case

Using the positive and negative interactions, we trained a PCmtI model. Application of this model allows us to discovery novel cancer-related key interactions. To date, many miRNAs have been demonstrated to function as oncogenes or tumor suppressor genes. An interesting case is miR-155 that has been found to be highly expressed in lymphoma (41), CLL (42), acute myelogenous leukemia (43), lung cancer (44), pancreatic cancer (45) and breast cancer (44). High expression of miR-155 has also been reported to correlate with poor prognosis in non-small

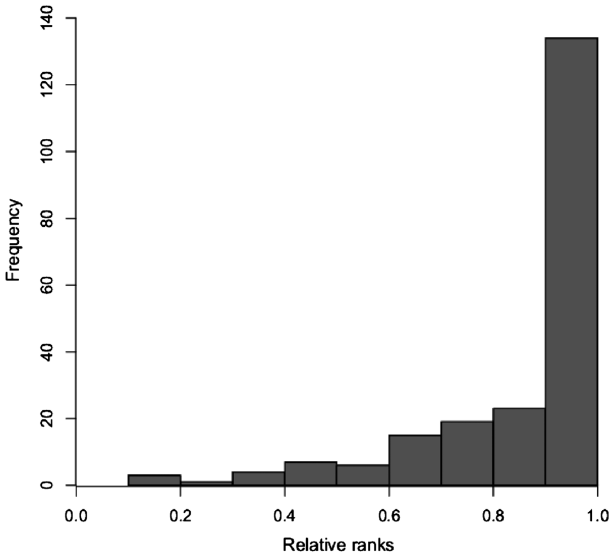

Figure 4. Distribution of the relative ranks of the known cancer-related key miRNA–target interactions.

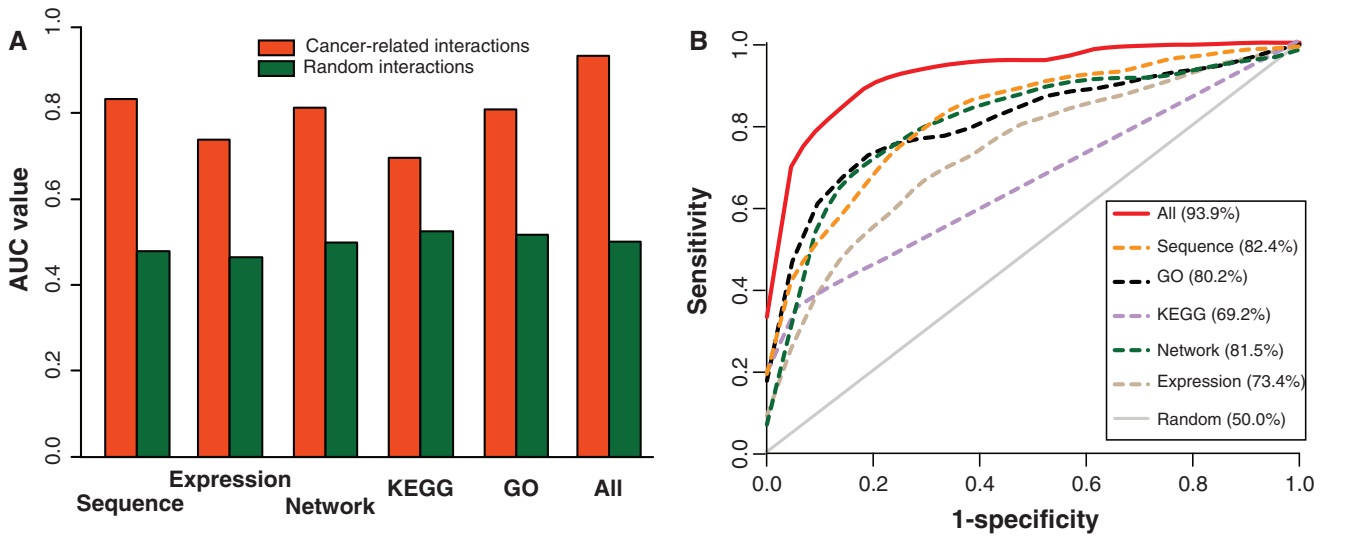

Figure 3. Ten-fold cross-validation results. (**A**) The AUC scores obtained using different types of features including sequence, expression, network, KEGG and GO were shown for prioritizations of the known cancer-related key interactions (red) and random interactions (green). By integrating all these features, the AUC score was calculated. (**B**) ROC curves corresponding to individual features or combination of features.

cell lung cancer (46). Importantly, genetically engineered mice with ectopic expression of miR-155 showed proliferation of polyclonal pre-B cells followed by leukemia or high-grade lymphoma (47). However, less is known about the mechanisms of dysregulation of miR-155 in cancer. Using the PCmtI model, we ranked all interactions relevant to miR-155 according to their prediction scores in a descending order. The top 20 interactions were selected as potential cancer-related key interactions (Table 1). Two of the three known cancer-related key interactions relevant to miR-155 were included in the list (i.e. miR-155:SMAD5, rank 13; miR-155:FOXO3, rank 18). In addition, several interactions were demonstrated to play crucial roles in certain biological processes, although without obvious evidence linking with cancer. For example, miR-155 was found to repress SMAD1 and SMAD5 expression, with consequence of inhibition of bone morphogenetic protein (BMP) signaling, which in turn reverses BMP-mediated cell growth inhibition (48). Induction of miR-155 in tumor-activated monocytes can suppress human CCAAT/enhancer-binding protein β (CEBPB) protein expression and cytokine production, and this effect can be mimicked by silencing of CEBPB (49). A recent study reported that miR-155 participated in the maturation of human dendritic cells and control of pathogen binding mostly through directly targeting the transcription factor SPI1 (50).

By literature mining, 11 biological processes have been demonstrated to be associated with miR-155, such as inflammation response, apoptosis and cell migration. We used these top 20 targets to capture the functions of miR-155 by function enrichment analysis (Benjamini–Hochberg corrected $P < 0.05$), and found that seven of the 11 known biological processes are identified (Supplementary Table S2). Using the top 30 targets, eight of the 11 were identified. The miRNA body map (51) provides a useful tool for predicting miRNA functions by integrating multi-level biological resources. We identified 15 GO biological process terms associated with miR-155 using the tool, and found seven out of the 15 GO biological process terms significantly over-represented in the top 20 targets, and 10 out of the 15 terms in the top 30 targets. These results suggest that the majority of functions of miR-155 can be characterized using the top ranked targets, underscoring the importance of the top ranked interactions for the function of miR-155.

Additionally, analyzing expression correlations of these top 20 interactions using the paired miRNA and mRNA expression data set from GBM showed two interactions showing significant negative correlations including miR-155:PAK7 ($r = -0.51$, $P < 2.2e^{-16}$) and miR-155:TCF4 ($r = -0.19$, $P = 1.82e^{-04}$). In order to investigate whether miR-155 directly targets the 3′UTRs of PAK7 and TCF4, we performed 3′UTR luciferase reporter assays for PAK7 and TCF4. A known target of miR-155, FOXO3, was also evaluated. Respective reporter plasmids harboring the wild-type versus mutated 3′UTR regions of FOXO3, PAK7 or TCF4 downstream of the luciferase coding region were constructed. When LN18 GBM cells were co-transfected with pGL3-FOXO3-3′UTR (or pGL3-TCF4-3′UTR) and the mature miR-155, we observed a significant decrease in relative luciferase activity, while such decrease was not observed with the control miRNA and with pGL3-PAK7-3′UTR. Furthermore, when we mutated the putative miR-155 binding sites at the 3′UTRs of FOXO3, TCF4 or PAK7, the relative luciferase activity between the miR-155 group and control was not significantly different (Figure 5). These results indicated that miR-155 directly targets TCF4 and FOXO3 rather than PAK7, suggesting that TCF4 may be another key target for miR-155. Consistent with our

Table 1. The top 20 interactions associated with miR-155

Rank	MiRNA	Gene	GBM[a]	OV[b]	MM	PC[d]	8mer	7mer-m8	7mer-A1	6mer	Predict score
1	miR-155	STAT1	0.26	0.27	0.10	−0.56	0	0	1	0	1.129
2	miR-155	SMAD1	0.31	0.20	0.01	−0.49	1	0	0	0	1.128
3	miR-155	CEBPB	0.48	0.15	0.20	0.74	1	0	0	0	1.113
4	miR-155	RPS6KB1	0.17	−0.05	0.04	−0.25	1	1	0	1	1.112
5	miR-155	PML	0.41	0.18	0.09	−0.53	0	1	0	0	1.097
6	miR-155	FOS	0.27	−0.08	0.15	−0.32	1	0	0	1	1.091
7	miR-155	CSF1R	0.29	0.30	0.25	−0.30	1	0	0	0	1.070
8	miR-155	BIRC3	0.40	0.29	0.05	0.27	0	1	0	0	1.052
9	miR-155	JAK2	0.16	0.21	−0.04	−0.29	0	1	0	0	1.043
10	miR-155	ETS1	−0.05	−0.04	0.12	0.53	2	0	0	1	1.017
11	miR-155	TFEC	0.32	0.43	0.21	0.26	0	0	1	0	1.014
12	miR-155	RB1	0.18	−0.03	−0.07	−0.60	0	0	1	1	0.995
13	miR-155	SMAD5	0.13	−0.14	−0.08	−0.42	0	1	1	0	0.983
14	miR-155	SPI1	0.31	0.33	−0.24	0.49	0	0	0	0	0.967
15	miR-155	TCF4	−0.19	−0.03	−0.41	0.38	2	2	0	0	0.964
16	miR-155	PAK7	−0.51	−0.15	−0.28	0.39	1	0	0	0	0.961
17	miR-155	RAC1	0.16	0.04	0.06	−0.35	0	1	0	1	0.954
18	miR-155	FOXO3	−0.09	−0.03	−0.24	−0.44	1	2	1	0	0.952
19	miR-155	HDAC9	0.04	0.09	−0.14	−0.50	0	1	1	0	0.951
20	miR-155	CTSS	0.42	0.37	−0.14	0.44	0	0	1	0	0.947

[a,b,c,d]Represent Pearson correlations calculated using four paired miRNA and mRNA expression data sets from GBM, ovarian cancer, multiple myeloma and prostate cancer, respectively.

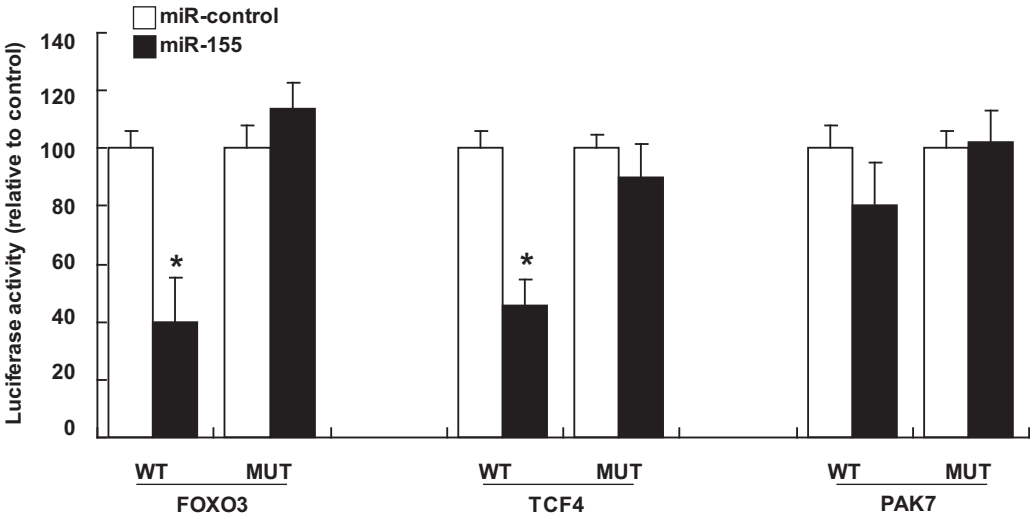

Figure 5. FOXO3 and TCF4 are targets for miR-155. Human GBM cells LN18 were co-transfected with the luciferase reporter construct carrying the 3′UTR sequence of the supposed target (WT for wild type sequences, MUT for mutated sequences) and miR-155, or miR-control. After 48 h, luciferase reporter assays were performed, and then relative luciferase activity was calculated. Each set of experiments was repeated at least four times. *Student's *t*-test, $P < 0.05$.

results, a most recent study (52) reported that miR-155 can directly suppress the expression of TCF4 that is an important regulator of epithelial-to-mesenchymal transition (EMT), and in turn reduces the aggressiveness of tumor cell dissemination.

Validation using recently identified cancer-related key miRNA–target interactions

In order to evaluate the performance in searching for novel cancer-related key miRNA–target interactions, we therefore examined recently published papers regarding dysfunction of miRNAs in tumorigenesis, and then obtained 23 novel cancer-related key miRNA–target interactions that are not included in the 210 positive interactions (Supplementary Table S3). We recorded the ranks of these 23 interactions based on a PCmtI model constructed using the positive and negative interactions. Of these 23 recently identified interactions, 15 were ranked in the top 20 of the interaction list for the corresponding miRNA, further indicating the superior performance of the approach.

The web-tool PCmtI

We developed a free-available web-tool PCmtI (http://bioinfo.hrbmu.edu.cn/PCmtI) for prioritization of cancer-related key miRNA–target interactions. The web-tool used these 210 positive and 8344 negative interactions to construct a model by integrating all sequence, expression and function features. PCmtI allows a user to input a miRNA name, and then displays a prioritization result page. In this result page, all interactions associated with the given miRNA are ranked according to prediction scores calculated using the model, and all features of these interactions are also provided. This web-tool may improve the chance of identifying cancer-related key miRNA–target interactions.

DISCUSSION

Identifying cancer-related key miRNA–target interactions helps to determine which genes are their downstream key targets and to further understand how miRNAs function as oncogenes or tumor suppressor genes involving in the development of cancer. With increasing experimentally validated cancer-related interactions, analyzing these interactions from different aspects (including sequence, expression and function) can reveal some unique properties, which can be used to discovery novel cancer-related key interactions. Here, we systematically explored sequence, expression and function features of known cancer-related key interactions by comparing to non-disease-related interactions.

Among all types of binding sites, 8mer binding sites provide the best basis for carrying out the dysfunction of miRNAs in cancer. Using seed-targeting 8mer locked nucleic acid oligonucleotides, a recent study successfully performed antagonism of miRNA function (53). We observed that cancer-related key interactions tend to have more binding sites, especially for 8mer binding sites, suggesting that cancer-related key interactions require high site efficiency. This is further supported by the observation that cancer-related key interactions have lower context scores.

With regard to expression, cancer-related key interactions show high expression correlations (including positive and negative correlations). Different patterns of expression correlations can result from complex regulatory circuits (54), which may be corresponding to different functional roles in the miRNA-mediated repression (55). High negative correlations generally reflect strong repression, such as controls of 'leaky' mRNAs in coherent feed-forward loops in which the miRNAs and their targets are inhibited and activated, respectively, by the same signals. High positive correlations can be caused by incoherent feed-forward loops, in which both the

miRNAs and their targets are co-activated (or co-repressed) by the same signals (54). In addition, recent studies (56) reported that miRNAs can even directly activate rather than repress their targets under certain conditions (57) or by binding to the 5′UTR (58). Thus, understanding expression relationships between miRNAs and their targets and understanding the involvement of miRNAs in complex regulatory circuits can offer insights into the molecular mechanisms of miRNAs involved in cancer development.

Interestingly, as shown in Table 1, only one miRNA–target interaction shows inverse expression correlations in all of these four paired miRNA and mRNA expression data sets from different types of cancer. The majority of miRNA–target interactions have inverse expression correlations only in certain types of cancer. One major explanation is that both positive and negative correlations are beneficial for identification of cancer-related key interactions, because we found that known cancer-related key interactions tend to show high positive and negative expression correlations. In addition, we further analyzed expression correlations of all experimentally validated miRNA–target interactions from miRTarBase (59) across these four paired miRNA and miRNA expression data sets. Surprisingly, we did not observe any tendency towards negative correlations (Supplementary Figure S1). Moreover, only 9.3% of experimentally validated interactions showed consistent negative correlations across these four data sets. These results suggest that miRNA–target interactions are heavily dependent on the specific cellular context, which may be attributed to tissue specificity.

When comparing expression correlations between cancer-related key interactions and negative interactions using four data sets, significant differences were only observed in two data sets. These results can be explained by the cellular context-dependence of miRNA–target interactions. Furthermore, due to only a minority of cancer-related key interactions identified to date, it may be insufficient to reveal the differences in some data sets. Even so, we believe that integration of features derived from these data sets can still provide advantages for prioritization of cancer-related key interactions.

By analyzing the GO-based functional essentiality of cancer-related key interactions, we demonstrated that the key targets in cancer-related interactions participate into more functions of their corresponding miRNAs when compared to those in non-disease-related interactions. Likewise, we observed a similar tendency using pathway annotations from KEGG. These findings suggest that the key targets in cancer-related interactions seem to be sufficient for reflecting partial or complete functions of their corresponding miRNAs. We next analyzed a network feature about cancer-related key interactions, and found significantly shorter average distances between key targets and other targets of their corresponding miRNAs, which indicates that these key targets in cancer-related interactions may highly connect with other targets. The key targets in cancer-related interactions may be involved in many functions of their corresponding miRNAs by cooperating with different targets, supporting their broad functional coverage.

It is well known that disease genes and non-disease genes have significant differences in a number of biological features, such as 3′UTR length. Significant differences between the positive and negative interactions may result from differences between disease genes and non-disease genes. To test this possibility, we randomly selected the same number of disease gene-related interactions as that in the positive set and the same number of non-disease gene-related interactions as that in the negative set, and then compared these two random selected interaction sets. We repeated the process 10 times, and found that all sequence features and most of the expression features do not exhibit significant differences while function features show significant differences (Supplementary Figure S2; $P < 0.05$, two-side KS test). When comparing the positive interactions with all disease gene-related interactions, we further found that the positive interactions show significantly higher functional coverage than disease gene-related interactions (Supplementary Figure S3; $P < 2.2e^{-16}$, one-side KS test). Our results suggest that these significant differences between the positive and negative interactions should be associated with cancer-related key miRNA–target interactions rather than just disease-related genes.

In order to investigate whether these significant differences also exist between other disease-related interactions and the negative interactions, we obtained 43 cardiovascular disease-related interactions by literature mining (Supplementary Table S4). Through comparing these 43 interactions with the negative interactions, we found similar results as those for the cancer-related key interactions—these cardiovascular disease-related interactions also show more 8mer binding sites and stronger functional links relative to the negative interactions (Supplementary Figure S4), suggesting that the differences may be extrapolated to other human diseases.

Recently, we witnessed the emergence of a number of methods for prioritizing disease genes by integrating different biological sources (60–63). These previous studies demonstrated the efficiency of prioritization methods for identifying novel disease genes and providing deep insights into the pathogenesis of disease (64). Inspired by these previous studies, we developed a method, named PCmtI, to prioritize cancer-related key miRNA–target interactions by using an ensemble SVM classifier model based on these sequence, expression and function features described above. Ten-fold cross-validation and leave-one-miRNA-out cross-validation showed the good performance of our approach. Prioritization of recently identified cancer-related key interactions also showed the ability of our approach for discovering novel cancer-related key interactions.

Recent studies found that miRNA oncogenes and miRNA tumor suppressors tend to regulate tumor suppressors and oncogenes, respectively (65). Therefore, we examined whether the top 20 interacting targets of some known miRNA oncogenes/tumor suppressors are significantly enriched in tumor suppressors/oncogenes. We selected five well-known miRNA oncogenes (including

miR-155, miR-107, miR-146a, miR-224 and miR-20a) and five miRNA tumor suppressors (including miR-143, miR-145, miR-34a, miR-200a and miR-195) from (65), and found that the top 20 targets of each miRNA oncogene/tumor suppressor are significantly enriched in tumor suppressors/oncogenes (*P*-values for miRNA oncogenes: miR-155, $5.24e^{-03}$; miR-107, $2.22e^{-05}$; miR-146a, $3.94e^{-04}$; miR-224, $2.22e^{-05}$; miR-20a, $4.98e^{-02}$; *P*-values for miRNA tumor suppressors: miR-143, $1.02e^{-04}$; miR-145, $1.30e^{-03}$; miR-34a, $1.20e^{-08}$; miR-200a, $1.26e^{-02}$; miR-195, $1.30e^{-03}$, Fisher's exact test). Similarly, we selected some well-known oncogenes/tumor suppressors (oncogenes: MYCN, WNT1, CDC25B, ERG and PDGFB; tumor suppressors: BRCA1, BRCA2, FOXO1A, RUNX3 and TCTA) to check whether their top 20 interacting miRNAs (ranked according to their prediction scores) are enriched in miRNA tumor suppressors/miRNA oncogenes, using a miRNA set analysis tool TAM (66). Our results showed that the top 20 miRNAs of most of these oncogenes/tumor suppressors are significantly enriched in miRNA tumor suppressors/miRNA oncogenes (*P*-values for oncogenes: MYCN, $1.25e^{-09}$; WNT1, $2.45e^{-05}$; CDC25B, $1.38e^{-12}$; ERG, $3.43e^{-03}$; PDGFB, $1.73e^{-11}$; *P*-values for tumor suppressors: BRCA1, $6.58e^{-03}$; BRCA2, 0.12; FOXO1A, $3.47e^{-2}$; RUNX3, $1.37e^{-09}$; TCTA, $1.45e^{-05}$).

Note that miRNAs in the same cluster tend to be co-expressed. To investigate whether this fact could affect our results, we selected three highly co-expressed miRNAs from the same cluster including miR-18a, miR-19a and miR-20a (36), which share 114 targets. We then prioritized interactions for each miRNA based on a model constructed using all positive and negative interactions. For each miRNA, the top 30 interactions were extracted. We did not find any common targets among these top 30 interaction sets of the three miRNAs (Supplementary Figure S5), suggesting that the fact may not influence our results.

Although the performance of our method is very encouraging, there is still much room for improvement. A recent study (51) compared several miRNA target databases by analyzing mass spectrometry protein expression data from miRNA perturbation experiments and demonstrated that MIRDB (67) outperforms TargetScan. Thus, we re-evaluated our model using predicted interactions from MIRDB, and found that the AUC score increased from 93.9% using TargetScan to 95.5% using MIRDB, indicating the improvement of performance by increasing the accuracy of miRNA target prediction. At present, a large number of mRNA and miRNA expression data sets are available. Integration of these expression data sets by meta-analysis (68) may further improve our approach. Additionally, with increased understanding of the cancer-related key interactions, more distinct features can be discovered. For example, recent evidence showed that the presence of single nucleotide polymorphisms in the 3′UTRs of targets can lead to gain or loss miRNA controls, which in turn contributes to many human diseases (69). RNA editing events were also reported to influence miRNA-mediated regulations (70). We expect that more features can be used to help improve the performance of our method.

In summary, we present a computational method called PCmtI for prioritization of cancer-related key miRNA–target interactions by combining sequence, expression and function features. We believe that PCmtI is a useful method for prioritizing cancer-related key interactions, which provides a new way of hypothesis generation that will help to reveal the molecular mechanism responsible for miRNA-associated cancer development.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–4 and Supplementary Figures 1–5.

FUNDING

The National Natural Science Foundation of China [61073136 and 61170154]; Specialized Research Fund for the Doctoral Program of Higher Education of China [20102307110022]; Science Foundation of Heilongjiang Province [JC200711, ZD200816-01 and D200834]; Science Foundation of Heilongjiang Province Education Department [1154h12, 11541137 and 1055HG009]. Funding for open access charge: The National Natural Science Foundation of China [61073136].

Conflict of interest statement. None declared.

REFERENCES

1. Croce,C.M. (2009) Causes and consequences of microRNA dysregulation in cancer. *Nat. Rev. Genet.*, **10**, 704–714.
2. Li,X., Wang,Q., Zheng,Y., Lv,S., Ning,S., Sun,J., Huang,T., Zheng,Q., Ren,H., Xu,J. *et al.* (2011) Prioritizing human cancer microRNAs based on genes' functional consistency between microRNA and cancer. *Nucleic Acids Res.*, **39**, e153.
3. Lai,E.C. (2005) miRNAs: whys and wherefores of miRNA-mediated regulation. *Curr. Biol.*, **15**, R458–R460.
4. Dorsett,Y., McBride,K.M., Jankovic,M., Gazumyan,A., Thai,T.H., Robbiani,D.F., Di Virgilio,M., Reina San-Martin,B., Heidkamp,G., Schwickert,T.A. *et al.* (2008) MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-Igh translocation. *Immunity*, **28**, 630–638.
5. Teng,G., Hakimpour,P., Landgraf,P., Rice,A., Tuschl,T., Casellas,R. and Papavasiliou,F.N. (2008) MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. *Immunity*, **28**, 621–629.
6. Cimmino,A., Calin,G.A., Fabbri,M., Iorio,M.V., Ferracin,M., Shimizu,M., Wojcik,S.E., Aqeilan,R.I., Zupo,S., Dono,M. *et al.* (2005) miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc. Natl Acad. Sci. USA*, **102**, 13944–13949.
7. Melton,C., Judson,R.L. and Blelloch,R. (2010) Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature*, **463**, 621–626.
8. Georges,S.A., Biery,M.C., Kim,S.Y., Schelter,J.M., Guo,J., Chang,A.N., Jackson,A.L., Carleton,M.O., Linsley,P.S., Cleary,M.A. *et al.* (2008) Coordinated regulation of cell cycle transcripts by p53-Inducible microRNAs, miR-192 and miR-215. *Cancer Res.*, **68**, 10105–10112.
9. Lee,R.C., Feinbaum,R.L. and Ambros,V. (1993) The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, **75**, 843–854.

10. Wightman,B., Ha,I. and Ruvkun,G. (1993) Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell*, **75**, 855–862.
11. Ambros,V. (2004) The functions of animal microRNAs. *Nature*, **431**, 350–355.
12. Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
13. Sales,G., Coppe,A., Bicciato,S., Bortoluzzi,S. and Romualdi,C. (2010) Impact of probe annotation on the integration of miRNA-mRNA expression profiles for miRNA target detection. *Nucleic Acids Res.*, **38(Suppl.)**, W352–W359.
14. Guo,H., Ingolia,N.T., Weissman,J.S. and Bartel,D.P. (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, **466**, 835–840.
15. Xiao,C. and Rajewsky,K. (2009) MicroRNA control in the immune system: basic principles. *Cell*, **136**, 26–36.
16. Xiao,C., Calado,D.P., Galler,G., Thai,T.H., Patterson,H.C., Wang,J., Rajewsky,N., Bender,T.P. and Rajewsky,K. (2007) MiR-150 controls B cell differentiation by targeting the transcription factor c-Myb. *Cell*, **131**, 146–159.
17. Kozomara,A. and Griffiths-Jones,S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.*, **39**, D152–D157.
18. Kent,W.J., Sugnet,C.W., Furey,T.S., Roskin,K.M., Pringle,T.H., Zahler,A.M. and Haussler,D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
19. Friedman,R.C., Farh,K.K., Burge,C.B. and Bartel,D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.
20. Su,A.I., Wiltshire,T., Batalov,S., Lapp,H., Ching,K.A., Block,D., Zhang,J., Soden,R., Hayakawa,M., Kreiman,G. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA*, **101**, 6062–6067.
21. Zhou,Y., Chen,L., Barlogie,B., Stephens,O., Wu,X., Williams,D.R., Cartron,M.A., van Rhee,F., Nair,B., Waheed,S. *et al.* (2010) High-risk myeloma is associated with global elevation of miRNAs and overexpression of EIF2C2/AGO2. *Proc. Natl Acad. Sci. USA*, **107**, 7904–7909.
22. Rajasekhar,V.K., Studer,L., Gerald,W., Socci,N.D. and Scher,H.I. (2011) Tumour-initiating stem-like cells in human prostate cancer exhibit increased NF-kappaB signalling. *Nat. Commun.*, **2**, 162.
23. Cancer Genome Atlas Research Network. (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, **455**, 1061–1068.
24. Cancer Genome Atlas Research Network. (2011) Integrated genomic analyses of ovarian carcinoma. *Nature*, **474**, 609–615.
25. Keshava Prasad,T.S., Goel,R., Kandasamy,K., Keerthikumar,S., Kumar,S., Mathivanan,S., Telikicherla,D., Raju,R., Shafreen,B., Venugopal,A. *et al.* (2009) Human Protein Reference Database: 2009 update. *Nucleic Acids Res.*, **37**, D767–D772.
26. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
27. Kanehisa,M. and Goto,S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.
28. Jiang,Q., Wang,Y., Hao,Y., Juan,L., Teng,M., Zhang,X., Li,M., Wang,G. and Liu,Y. (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.*, **37**, D98–D104.
29. Lu,M., Zhang,Q., Deng,M., Miao,J., Guo,Y., Gao,W. and Cui,Q. (2008) An analysis of human microRNA and disease associations. *PLoS One*, **3**, e3420.
30. Yang,Z., Ren,F., Liu,C., He,S., Sun,G., Gao,Q., Yao,L., Zhang,Y., Miao,R., Cao,Y. *et al.* (2010) dbDEMC: a database of differentially expressed miRNAs in human cancers. *BMC Genom.*, **11(Suppl. 4)**, S5.
31. Hamosh,A., Scott,A.F., Amberger,J., Bocchini,C., Valle,D. and McKusick,V.A. (2002) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **30**, 52–55.
32. Becker,K.G., Barnes,K.C., Bright,T.J. and Wang,S.A. (2004) The genetic association database. *Nat. Genet.*, **36**, 431–432.
33. Futreal,P.A., Coin,L., Marshall,M., Down,T., Hubbard,T., Wooster,R., Rahman,N. and Stratton,M.R. (2004) A census of human cancer genes. *Nat. Rev. Cancer*, **4**, 177–183.
34. Grimson,A., Farh,K.K., Johnston,W.K., Garrett-Engele,P., Lim,L.P. and Bartel,D.P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell*, **27**, 91–105.
35. Mestdagh,P., Bostrom,A.K., Impens,F., Fredlund,E., Van Peer,G., De Antonellis,P., von Stedingk,K., Ghesquiere,B., Schulte,S., Dews,M. *et al.* (2010) The miR-17-92 MicroRNA cluster regulates multiple components of the TGF-beta pathway in neuroblastoma. *Mol. Cell*, **40**, 762–773.
36. Xiao,Y., Xu,C., Guan,J., Ping,Y., Fan,H., Li,Y., Zhao,H. and Li,X. (2012) Discovering dysfunction of multiple microRNAs cooperation in disease by a conserved microRNA co-expression network. *PLoS One*, **7**, e32201.
37. Franke,L., van Bakel,H., Fokkens,L., de Jong,E.D., Egmont-Petersen,M. and Wijmenga,C. (2006) Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes. *Am. J. Hum. Genet.*, **78**, 1011–1025.
38. Kislinger,T., Cox,B., Kannan,A., Chung,C., Hu,P., Ignatchenko,A., Scott,M.S., Gramolini,A.O., Morris,Q., Hallett,M.T. *et al.* (2006) Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. *Cell*, **125**, 173–186.
39. Brown,K.R. and Jurisica,I. (2005) Online predicted human interaction database. *Bioinformatics*, **21**, 2076–2082.
40. Wang,X. (2006) Systematic identification of microRNA functions by combining target prediction and expression profiling. *Nucleic Acids Res.*, **34**, 1646–1652.
41. Metzler,M., Wilda,M., Busch,K., Viehmann,S. and Borkhardt,A. (2004) High expression of precursor microRNA-155/BIC RNA in children with Burkitt lymphoma. *Genes Chromosomes Cancer*, **39**, 167–169.
42. Kluiver,J., Poppema,S., de Jong,D., Blokzijl,T., Harms,G., Jacobs,S., Kroesen,B.J. and van den Berg,A. (2005) BIC and miR-155 are highly expressed in Hodgkin, primary mediastinal and diffuse large B cell lymphomas. *J. Pathol.*, **207**, 243–249.
43. Garzon,R., Volinia,S., Liu,C.G., Fernandez-Cymering,C., Palumbo,T., Pichiorri,F., Fabbri,M., Coombes,K., Alder,H., Nakamura,T. *et al.* (2008) MicroRNA signatures associated with cytogenetics and prognosis in acute myeloid leukemia. *Blood*, **111**, 3183–3189.
44. Volinia,S., Calin,G.A., Liu,C.G., Ambs,S., Cimmino,A., Petrocca,F., Visone,R., Iorio,M., Roldo,C., Ferracin,M. *et al.* (2006) A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc. Natl Acad. Sci. USA*, **103**, 2257–2261.
45. Greither,T., Grochola,L.F., Udelnow,A., Lautenschlager,C., Wurl,P. and Taubert,H. (2010) Elevated expression of microRNAs 155, 203, 210 and 222 in pancreatic tumors is associated with poorer survival. *Int. J. Cancer*, **126**, 73–80.
46. Yanaihara,N., Caplen,N., Bowman,E., Seike,M., Kumamoto,K., Yi,M., Stephens,R.M., Okamoto,A., Yokota,J., Tanaka,T. *et al.* (2006) Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell*, **9**, 189–198.
47. Costinean,S., Zanesi,N., Pekarsky,Y., Tili,E., Volinia,S., Heerema,N. and Croce,C.M. (2006) Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in E(mu)-miR155 transgenic mice. *Proc. Natl Acad. Sci. USA*, **103**, 7024–7029.
48. Yin,Q., Wang,X., Fewell,C., Cameron,J., Zhu,H., Baddoo,M., Lin,Z. and Flemington,E.K. (2010) MicroRNA miR-155 inhibits bone morphogenetic protein (BMP) signaling and BMP-mediated Epstein-Barr virus reactivation. *J. Virol.*, **84**, 6318–6327.
49. He,M., Xu,Z., Ding,T., Kuang,D.M. and Zheng,L. (2009) MicroRNA-155 regulates inflammatory cytokine production in tumor-associated macrophages via targeting C/EBPbeta. *Cell. Mol. Immunol.*, **6**, 343–352.
50. Martinez-Nunez,R.T., Louafi,F., Friedmann,P.S. and Sanchez-Elsner,T. (2009) MicroRNA-155 modulates the pathogen binding ability of dendritic cells (DCs) by down-regulation of DC-specific intercellular adhesion molecule-3 grabbing non-integrin (DC-SIGN). *J. Biol. Chem.*, **284**, 16334–16342.

51. Mestdagh,P., Lefever,S., Pattyn,F., Ridzon,D., Fredlund,E., Fieuw,A., Ongenaert,M., Vermeulen,J., De Paepe,A., Wong,L. *et al.* (2011) The microRNA body map: dissecting microRNA function through integrative genomics. *Nucleic Acids Res.*, **39**, e136.
52. Xiang,X., Zhuang,X., Ju,S., Zhang,S., Jiang,H., Mu,J., Zhang,L., Miller,D., Grizzle,W. and Zhang,H.G. (2011) miR-155 promotes macroscopic tumor formation yet inhibits tumor dissemination from mammary fat pads to the lung by preventing EMT. *Oncogene*, **30**, 3440–3453.
53. Obad,S., Dos Santos,C.O., Petri,A., Heidenblad,M., Broom,O., Ruse,C., Fu,C., Lindow,M., Stenvang,J., Straarup,E.M. *et al.* (2011) Silencing of microRNA families by seed-targeting tiny LNAs. *Nat. Genet.*, **43**, 371–378.
54. Inui,M., Martello,G. and Piccolo,S. (2010) MicroRNA control of signal transduction. *Nat. Rev. Mol. Cell Biol.*, **11**, 252–263.
55. Shkumatava,A., Stark,A., Sive,H. and Bartel,D.P. (2009) Coherent but overlapping expression of microRNAs and their targets during vertebrate development. *Genes Dev.*, **23**, 466–481.
56. Chekulaeva,M. and Filipowicz,W. (2009) Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells. *Curr. Opin. Cell Biol.*, **21**, 452–460.
57. Vasudevan,S., Tong,Y. and Steitz,J.A. (2007) Switching from repression to activation: microRNAs can up-regulate translation. *Science*, **318**, 1931–1934.
58. Jopling,C.L., Yi,M., Lancaster,A.M., Lemon,S.M. and Sarnow,P. (2005) Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science*, **309**, 1577–1581.
59. Hsu,S.D., Lin,F.M., Wu,W.Y., Liang,C., Huang,W.C., Chan,W.L., Tsai,W.T., Chen,G.Z., Lee,C.J., Chiu,C.M. *et al.* (2011) miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res.*, **39**, D163–D169.
60. Aerts,S., Lambrechts,D., Maity,S., Van Loo,P., Coessens,B., De Smet,F., Tranchevent,L.C., De Moor,B., Marynen,P., Hassan,B. *et al.* (2006) Gene prioritization through genomic data fusion. *Nat. Biotechnol.*, **24**, 537–544.
61. Furney,S.J., Calvo,B., Larranaga,P., Lozano,J.A. and Lopez-Bigas,N. (2008) Prioritization of candidate cancer genes: an aid to oncogenomic studies. *Nucleic Acids Res.*, **36**, e115.
62. Xiao,Y., Xu,C., Ping,Y., Guan,J., Fan,H., Li,Y. and Li,X. (2011) Differential expression pattern-based prioritization of candidate genes through integrating disease-specific expression data. *Genomics*, **98**, 64–71.
63. Kann,M.G. (2007) Protein interactions and disease: computational approaches to uncover the etiology of diseases. *Brief Bioinform.*, **8**, 333–346.
64. Tiffin,N., Adie,E., Turner,F., Brunner,H.G., van Driel,M.A., Oti,M., Lopez-Bigas,N., Ouzounis,C., Perez-Iratxeta,C., Andrade-Navarro,M.A. *et al.* (2006) Computational disease gene identification: a concert of methods prioritizes type 2 diabetes and obesity candidate genes. *Nucleic Acids Res.*, **34**, 3067–3081.
65. Wang,D., Qiu,C., Zhang,H., Wang,J., Cui,Q. and Yin,Y. (2010) Human microRNA oncogenes and tumor suppressors show significantly different biological patterns: from functions to targets. *PLoS One*, **5**, e13067.
66. Lu,M., Shi,B., Wang,J., Cao,Q. and Cui,Q. (2010) TAM: a method for enrichment and depletion analysis of a microRNA category in a list of microRNAs. *BMC Bioinformatics*, **11**, 419.
67. Wang,X. (2008) miRDB: a microRNA target prediction and functional annotation database with a wiki interface. *RNA*, **14**, 1012–1017.
68. Xiao,Y., Xu,C., Xu,L., Guan,J., Ping,Y., Fan,H., Li,Y., Zhao,H. and Li,X. (2012) Systematic identification of common functional modules related to heart failure with different etiologies. *Gene*, **499**, 332–338.
69. Ryan,B.M., Robles,A.I. and Harris,C.C. (2010) Genetic variation in microRNA networks: the implications for cancer research. *Nat. Rev. Cancer.*, **10**, 389–402.
70. Blow,M.J., Grocock,R.J., van Dongen,S., Enright,A.J., Dicks,E., Futreal,P.A., Wooster,R. and Stratton,M.R. (2006) RNA editing of human microRNAs. *Genome Biol.*, **7**, R27.