RESEARCH   **Open Access**

Identification of candidate miRNA biomarkers from miRNA regulatory network with application to prostate cancer

Wenyu Zhang[1†], Jin Zang[2†], Xinhua Jing[1], Zhandong Sun[1], Wenying Yan[1], Dongrong Yang[3], Feng Guo[4*]
and Bairong Shen[1*]

Abstract

Background: MicroRNAs (miRNAs) are a class of non-coding regulatory RNAs approximately 22 nucleotides in length that play a role in a wide range of biological processes. Abnormal miRNA function has been implicated in various human cancers including prostate cancer (PCa). Altered miRNA expression may serve as a biomarker for cancer diagnosis and treatment. However, limited data are available on the role of cancer-specific miRNAs. Integrative computational bioinformatics approaches are effective for the detection of potential outlier miRNAs in cancer.

Methods: The human miRNA-mRNA target network was reconstructed by integrating multiple miRNA-mRNA interaction datasets. Paired miRNA and mRNA expression profiling data in PCa versus benign prostate tissue samples were used as another source of information. These datasets were analyzed with an integrated bioinformatics framework to identify potential PCa miRNA signatures. In vitro q-PCR experiments and further systematic analysis were used to validate these prediction results.

Results: Using this bioinformatics framework, we identified 39 miRNAs as potential PCa miRNA signatures. Among these miRNAs, 20 had previously been identified as PCa aberrant miRNAs by low-throughput methods, and 16 were shown to be deregulated in other cancers. In vitro q-PCR experiments verified the accuracy of these predictions. miR-648 was identified as a novel candidate PCa miRNA biomarker. Further functional and pathway enrichment analysis confirmed the association of the identified miRNAs with PCa progression.

Conclusions: Our analysis revealed the scale-free features of the human miRNA-mRNA interaction network and showed the distinctive topological features of existing cancer miRNA biomarkers from previously published studies. A novel cancer miRNA biomarker prediction framework was designed based on these observations and applied to prostate cancer study. This method could be applied for miRNA biomarker prediction in other cancers.

Keywords: miRNA biomarker, Gene expression, miRNA regulatory network, Prostate cancer

Background

MicroRNAs (miRNAs) are a class of small non-coding RNAs of approximately 22 nucleotides in length with the potential to regulate human genes through translation inhibition or mRNA cleavage [1]. Recent studies have shown that miRNAs are involved in a wide variety of biological processes such as cell proliferation [2], development [3], and apoptosis [4]. Abnormal expression of miRNAs has been implicated in various human cancers and may constitute a potential signature for cancer diagnosis [5-7]. However, limited data on cancer related miRNAs are available, and their regulatory mechanisms remain largely unknown.

Extensive research efforts have focused on the identification of potential cancer miRNA biomarkers [6-10]. The preliminary detection of differentially expressed (DE) miRNAs from large-scale miRNA expression profiling data and low-

* Correspondence: guofeng27@suda.edu.cn; bairong.shen@suda.edu.cn
†Equal contributors
[1]Center for Systems Biology, Soochow University, Suzhou 215006, China
[4]Central lab, the First Affiliated Hospital of Soochow University, Suzhou 215006, China
Full list of author information is available at the end of the article

throughput experimental validation for selected outlier miRNAs are the routine methods used in these studies. As the activities of outlier miRNAs are at least partially reflected in the aberrant expression of their target genes [11], systematic computational approaches that integrate miRNA regulatory data and gene expression profiling data were shown to be more effective to infer potential outlier miRNA activities in cancers [11,12].

MiRNAs are known to function in a multiple-to-multiple relationship with their target genes, and a concept referred to as miRNA regulation module was proposed based on this theory [13]. This idea was further explored in cancer studies, and attempts have been made to identify candidate abnormal miRNAs or miRNA regulatory modules in cancer [12,14-17]. The assumption that abnormal miRNAs associated with cancer show increased functional synergism because of their co-regulatory effects on the same genes [18] was the underlying foundation of these computational approaches.

In contrast to these analyses, the miRNA regulatory network was shown to follow power-law distribution in our study. This implies that more miRNAs tend to have fewer target genes and fewer genes have more miRNA regulators. Based on this analysis, we defined a novel out degree (NOD) measure to characterize the independent regulatory power of individual miRNAs, i.e., the number of genes uniquely regulated by one specific miRNA. Our analysis indicated that miRNAs with larger NOD values are statistically more likely to be candidate cancer biomarkers, which implies that abnormal miRNAs in cancer generally have independent regulatory power. Based on this observation, we proposed a novel pipeline to infer candidate cancer miRNA biomarkers, and applied this method to prostate cancer (PCa). The schematic workflow of our pipeline is shown in Figure 1. Paired miRNA and mRNA expression profiling data for PCa, and reliable miRNA-mRNA interaction data were integrated to generate a PCa conditional miRNA regulatory network, and then candidate PCa miRNA biomarkers in the above conditional network were prioritized according to their independent regulatory power. PubMed citation analysis of the known PCa abnormal miRNAs and in vitro q-PCR
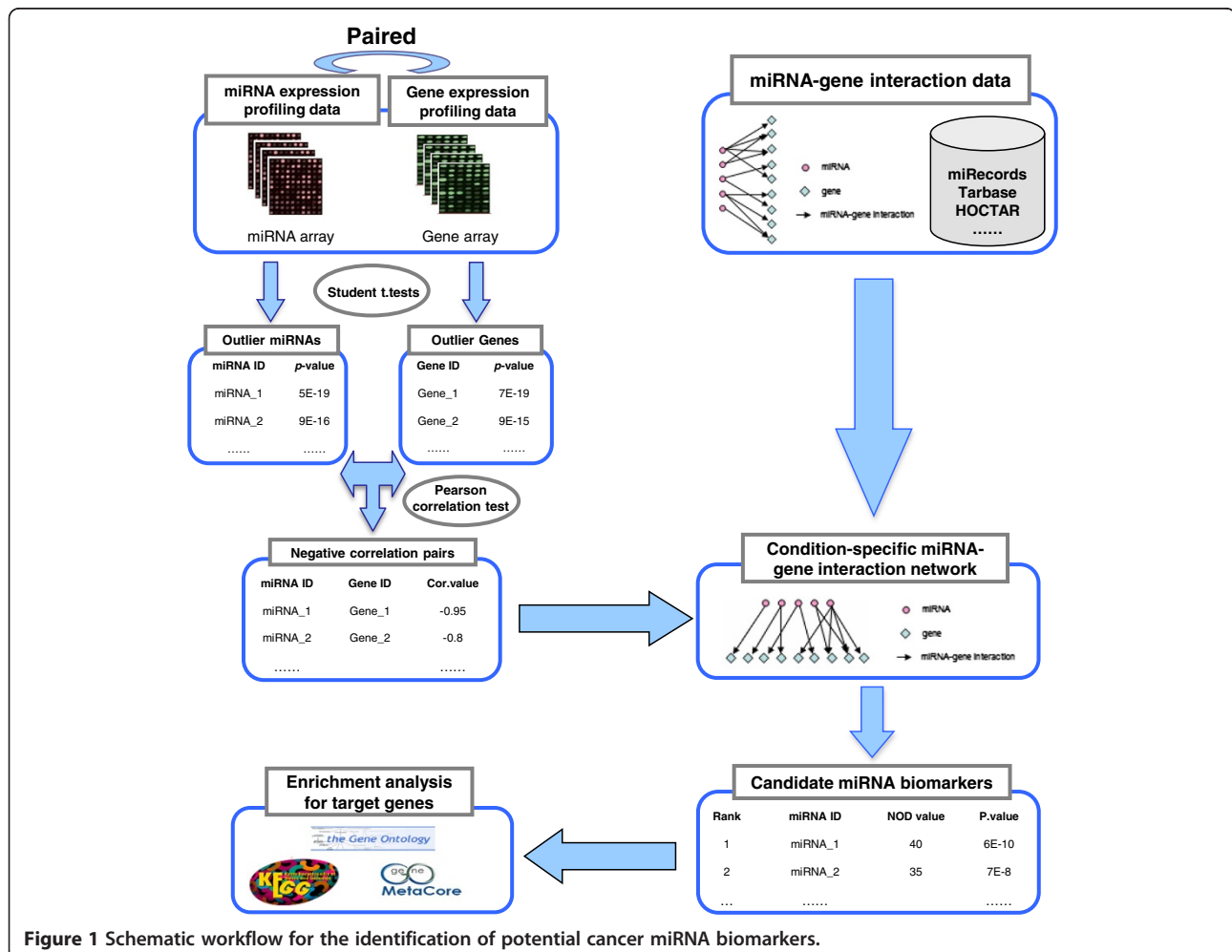

Figure 1 Schematic workflow for the identification of potential cancer miRNA biomarkers.

technology were used to verify the accuracy of our prediction results. Finally, systematic methods were applied to explore the relationship between PCa and the unique target genes of candidate PCa miRNA biomarkers. The results indicated that our method can detect potential miRNA-mRNA target relationships in specific cancers and can be applied for the identification of miRNA biomarkers in cancer.

Methods

Dataset collection

Expression profiles (GSE34933 from NCBI GEO) for PCa and benign prostate tissue (BPH) samples generated by Zhong and colleagues [19,20] were used. Eight available paired miRNA and mRNA expression profiles (each containing 4 samples for PCa and BPH) were selected for further analysis. Information on these profiles is provided in Additional file 1. Normalized miRNA and mRNA data were downloaded directly. For mRNA expression data, the average probe intensity was calculated and used as the gene expression level for genes with multiple probes. Finally, the profiles included information on the expression of 851 miRNAs and 19595 genes.

Another dataset used in this study was the miRNA-mRNA network. This dataset consisted of a combination of experimentally validated targeting data and computational prediction data. The experimentally validated data included information from miRecords [21], TarBase [22], miR2Disease [23], and miRTarBase [24], while the computational prediction data consisted of miRNA-mRNA target pairs residing in no fewer than 2 datasets from HOCTAR [25], ExprTargetDB [26], and starBase [27]. In total, there were 32739 regulatory pairs among 641 miRNAs and 7706 target genes.

Prostate cancer miRNA biomarker identification

We developed a novel approach to identify candidate miRNA biomarkers for PCa. The schematic workflow of our pipeline is described in Figure 1. Paired miRNA and gene expression, and miRNA-mRNA networks were integrated to predict outlier miRNAs associated with PCa progression. This procedure consisted of four separate stages. First, differentially expressed miRNAs and genes between PCa and BPH samples were detected using the two-sample t-test. Second, Pearson's correlation was used to detect negative correlations between the expression profiles of outlier miRNAs and outlier genes. In the third step, the intersection data of the negative correlations and miRNA-mRNA binding pairs were retrieved to identify miRNA regulatory networks related to PCa progression. In the fourth and final step, a new index designated as novel out-degree (NOD) was defined to measure the independent regulatory power of an individual miRNA, and used to prioritize novel PCa miRNA biomarkers.

Step 1: Detection of differentially expressed miRNAs and genes associated with prostate cancer

The detection of cancer-specific abnormal changes in miRNA and gene expression is the aim of cancer studies [28-31]. Here, we used two-sample t-tests to identify differently expressed miRNAs and genes associated with PCa progression on the basis of their expression profiles. The top 30% miRNAs (or genes) ranked by their statistical significance (p-value) were retrieved for further analysis. As a result, 256 miRNAs and 5878 genes were considered as candidate PCa outliers.

The threshold for the expression of outlier miRNAs and outlier genes is often arguable. A less stringent cut-off (top 40%) and a stricter cut-off (top 20%) were tested for candidate miRNA biomarker prediction. Details of the comparison between these predictions are listed in Additional file 2. The data indicated that the prediction results were highly conserved and only the number of candidate miRNAs changed with the different thresholds. Therefore, we adopted a moderate threshold (top 30%) in the present study.

Step 2: Acquisition of inverse correlation pairs

One major function of miRNAs is the cleavage of transcripts of its target genes at the post-transcriptional level. Thus, the inverse correlation of expression profiles should be one prerequisite for miRNAs and candidate targets. In the present study, the Pearson's correlation method was used to detect negative correlations between outlier miRNAs and outlier genes. The cut-off for the correlation coefficient was roughly chosen to be -0.6, as it has been used as a threshold in several correlation studies [32,33].

Step 3: Constructing a prostate cancer miRNA-mRNA binding network

According to the above miRNA-mRNA binding data from experimental validation and computational prediction databases, we identified possible human miRNA-mRNA target pairs. We further filtered these target pairs with the collected information on miRNA-mRNA negative correlations to generate a PCa miRNA regulatory network. As a result, the miRNA-mRNA target sub-network consisted of 136 miRNAs and 551 target genes.

Step 4: Prioritizing candidate prostate cancer miRNA biomarkers

Generally, we face two main challenges for the prediction of miRNAs related to cancer based on miRNA-mRNA regulatory data. First, for genes with abnormal expression that are regulated by more than one miRNA, it is difficult to discriminate which miRNA contributed to the deregulation of this gene. Second, besides miRNA regulation, other factors such as DNA methylation may

also result in abnormal expression of the studied gene. To overcome these problems, we defined a novel out-degree (NOD) index to measure the independent regulatory power of an individual miRNA, *i.e.,* the genes uniquely regulated by one specific miRNA. Based on the observation that miRNAs with greater independent regulatory power were more likely to be cancer biomarkers as described in the Results section, we prioritized candidate PCa miRNA biomarkers according to their NOD values, as calculated from the PCa miRNA regulatory network.

In summary, the number of uniquely regulated genes was first computed as a NOD value for each miRNA in the PCa miRNA regulatory network. These miRNAs were further ranked by their NOD values. The Wilcoxon signed-rank test was then applied to assign a statistical significance value (*p*-value) to each miRNA, which indicated whether the NOD value of an individual miRNA was significantly greater than the median level of all these candidate miRNAs. Herein, the threshold of the *p*-value was set at 0.01. Finally, 39 miRNAs were detected as potential PCa miRNA biomarkers in our study.

Performance comparison with other computational methods

To evaluate the accuracy of our method, we compared its performance with that of two other computational approaches, the miRNA expression fold-change based on the t-test method [34] and another method based on the cancer miRNA synergism theory [12]. The same numbers of top ranked miRNAs as in our prediction results were extracted from these two methods for comparison. The performance of each computational method was expressed as the percentage of known PCa abnormal miRNAs in their prediction results.

In vitro q-PCR confirmation of candidate prostate cancer miRNA biomarkers

When normal prostate tissue (NPT) samples are unavailable, benign prostatic hyperplasia (BPH) samples can be used as normal prostate samples for comparison with PCa samples [35,36]. The study group consisted of 25 Han Chinese patients with PCa and 20 Han Chinese individuals with BPH with ages ranging from 60 to 91. The PCa and BPH samples were part of a sample set collected for clinical diagnostic tests at the First Affiliated Hospital of Soochow University (Suzhou, China). No extra samples were collected from the study subjects; therefore, verbal consent was obtained from all participating individuals. The study procedure was approved by the ethics committee of Soochow University. The PCa and BPH tissues were snap-frozen in liquid nitrogen and stored at -80°C. Total RNA was extracted with the TRIzol reagent (Invitrogen, China). RNA quantity was measured on a Nanodrop 1000 Spectrophotometer (Thermo Scientific,

China). Universal reverse transcription of all the mature miRNAs was performed by enzymatic tailing of the miRNAs by using Poly(A) Polymerase. MiRNAs were first tailed and then reverse transcribed by using universal primers. The sequences of miRNAs were obtained from the miRNAMap database [37]. MiRNA specific primers were designed with Primer 3 software. Quantitative PCR was performed in a volume of 20 μl containing 2 μl of cDNA diluted 10 times, 10 μl of LightCycler® 480 SYBR Green I Master (Roche, China), and 200 nM of each primer. U6 expression was used as the internal control, and all quantitative PCR values were normalized to those of U6 RNA. Triplicates were performed for all reactions with a LightCycler® 480 System (Roche, China). Relative expression was analyzed by the Pfaffl method. All the statistical analyses were carried out on Graphpad Prism software.

Systematic analysis of the target genes of candidate prostate cancer miRNA biomarkers

The uniquely regulated genes associated with our prediction miRNAs from the PCa miRNA-mRNA target network were retrieved. Gene Ontology (GO) analysis and pathway analysis were performed to explore the relationships between these genes and PCa. The Database for Annotation, Visualization and Integrated Discovery (DAVID) [38] was used for GO annotation and KEGG pathway [39,40] analysis. Another pathway source, Meta-Core™ Database from GeneGo Inc., was used for GeneGo pathway mapping analysis. The highly significantly mapped pathways (*p*-value < 0.01) were further confirmed for their association with PCa via NCBI PubMed literature exploration.

Results
Global features of the miRNA-mRNA target network

Individual miRNAs can regulate multiple genes, and an individual target gene can be co-regulated by several miRNAs [41,42]. However, further exploration of the in-degree distribution of targets and out-degree distribution of miRNAs for the whole human miRNA-mRNA target network revealed that the multiple-to-multiple relationship between miRNA and mRNA is over-emphasized. Compared with the random network simulated with the same nodes and binding links, the in-degree distribution and out-degree distribution of the real miRNA-mRNA target network followed the power-law distribution (see Figure 2), with slopes of -2.37 and -0.71, respectively. This indicated that the distribution of the miRNA regulatory network was also scale-free and similar to other biological complex networks such as the protein interaction network [43].

The analysis of our reconstructed miRNA-mRNA network indicated that a large fraction (34.41%) of genes was exclusively regulated by an individual miRNA. Based
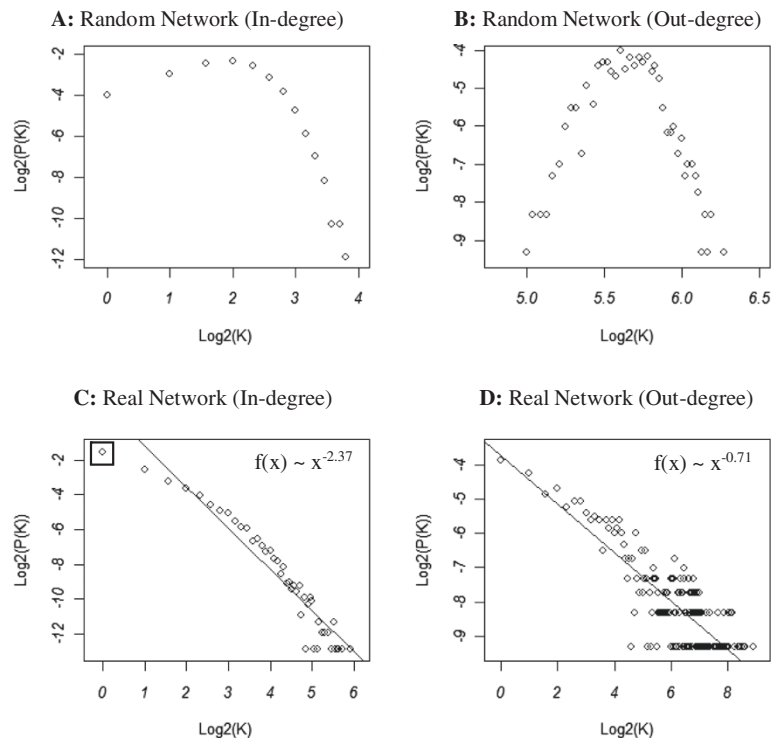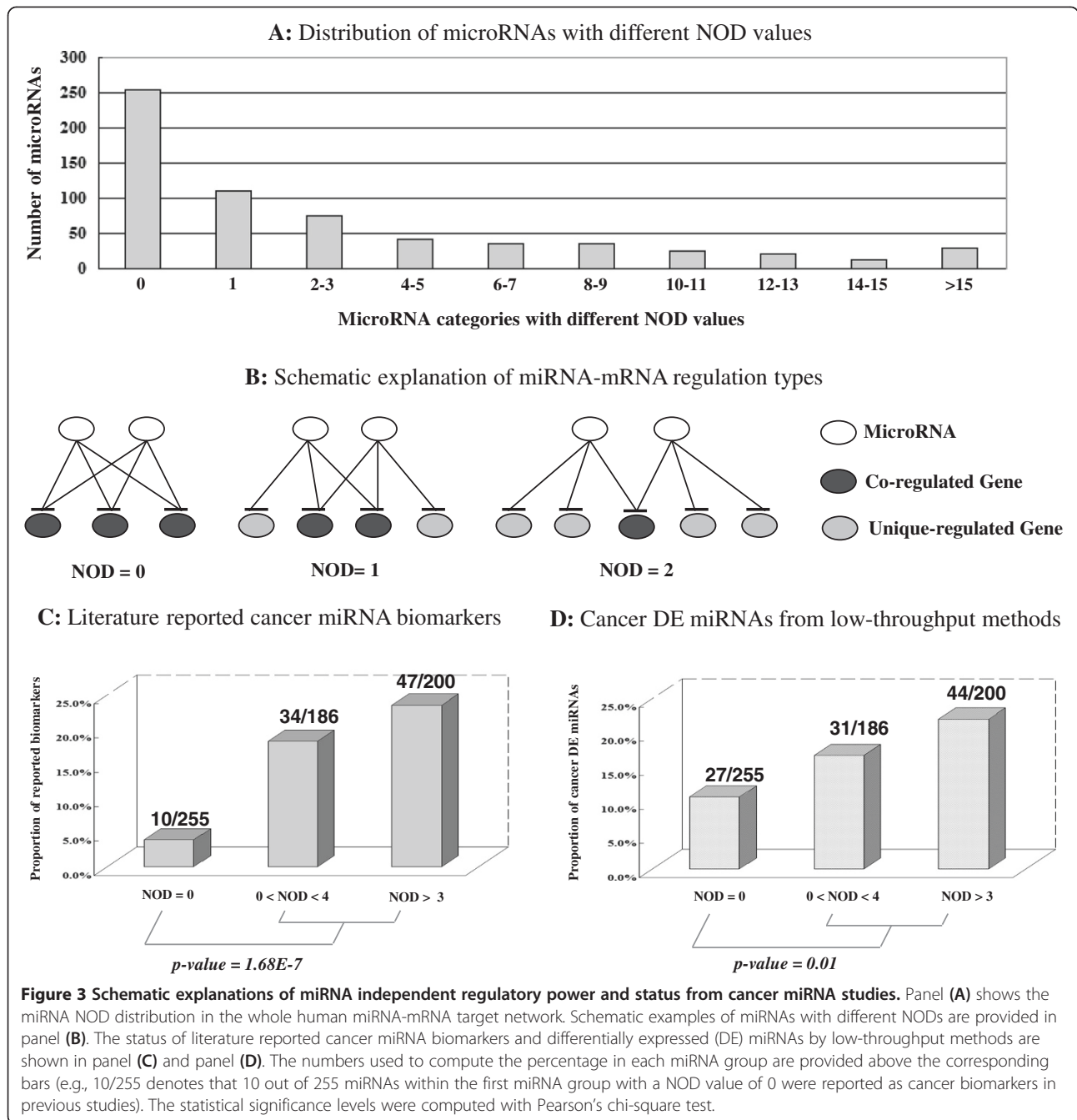
Figure 2 Degree distribution of the whole human miRNA-mRNA target network. Panels **(A)** and **(B)** respectively illustrate the In- and Out- degree distributions of the random network, whichwas simulated with the same number of miRNA and mRNA nodes and miRNA-mRNA target pairs. The x-axis represents log-2 based degrees, and the y-axis indicates the log-2 based frequencies of nodes with corresponding degrees. The in-degree and out-degree distributions for the random network were symmetrical while the panels **(C)** and **(D)** show the power law distributions with slopes of -2.37 and -0.71 for the in-degree and out-degree distributions of the real network, respectively. The circle inside the small pane in panel **(C)** represents genes (34.41% of total genes) that were uniquely targeted by an individual miRNA.

on this observation, we defined a new index designated as novel out degree (NOD) to measure the independent regulatory power of an individual miRNA (see Figure 3 and Methods). Interestingly, the distribution of miRNA NOD values was also shown to follow the power law. According to their NOD levels, we further classified these miRNAs into three categories: miRNAs with no independent regulatory power (NOD = 0); miRNAs with a small independent regulatory power (0 < NOD < 4); and miRNAs with a large independent regulatory power (NOD > 3). We then explored the current cancer related miRNAs by text mining of PubMed citations. Briefly, we first checked whether each miRNA had been previously claimed as a cancer biomarker by using the NCBI PubMed search engine. Then, to strengthen our findings, we used a previously described method [44] to further explore differentially expressed (DE) miRNAs in cancer detected by low-throughput technology in studies published before January 1st 2013. A statistically significant difference of their biomarker potentials could be observed between miRNAs without independent regulatory power and those with independent regulatory power. The results shown in Figure 3C and Figure 3D indicated that miRNAs

with larger independent regulatory power were more likely to be potential cancer miRNA biomarkers and show aberrant functions in cancer.

Prediction of candidate prostate cancer miRNA biomarkers

Based on the above observation that miRNAs with larger independent regulatory power are more likely to be cancer abnormal miRNAs, we developed a pipeline to infer candidate cancer miRNA biomarkers from cancer conditional miRNA regulatory networks, and then applied this pipeline for PCa study, as described in the Methods section. A total of 39 miRNAs were predicted to be candidate PCa miRNA biomarkers in our study. Among these miRNAs, 20 (51.3%) had previously been shown to be PCa aberrant miRNAs by low-throughput methods in early studies, and thus could be potential miRNA biomarkers in PCa. Among the remaining 19 candidates, 16 miRNAs had been reported to show outlier activities in other cancers, whereas the activities of the other three miRNAs had not been explored yet. These miRNAs were considered potential PCa miRNAs requiring further investigation. Detailed information about the prediction

Figure 3 Schematic explanations of miRNA independent regulatory power and status from cancer miRNA studies. Panel **(A)** shows the miRNA NOD distribution in the whole human miRNA-mRNA target network. Schematic examples of miRNAs with different NODs are provided in panel **(B)**. The status of literature reported cancer miRNA biomarkers and differentially expressed (DE) miRNAs by low-throughput methods are shown in panel **(C)** and panel **(D)**. The numbers used to compute the percentage in each miRNA group are provided above the corresponding bars (e.g., 10/255 denotes that 10 out of 255 miRNAs within the first miRNA group with a NOD value of 0 were reported as cancer biomarkers in previous studies). The statistical significance levels were computed with Pearson's chi-square test.

miRNAs and known PCa miRNAs used in this study can be found in Additional files 3 and 4, respectively.

We further evaluated the performance of our method by comparing our results with data on fold-changes in miRNA expression obtained using the t-test approach [34] and a method based on the cancer miRNA synergism theory [12] (see Methods). The results are shown in Figure 4. As a computational method based on the integration of expression fold-change and miRNA regulatory network feature information, our method performed better than fold-change of t-test based methods. The performance of our method was comparable to that of the method described by Xu et al. [12], based on the reported PCa miRNA information (Figure 4). Nevertheless, Xu's method is based on a SVM classifier, and is therefore highly dependent on the PCa miRNA training data, whereas our approach does not require any prior knowledge. Overall, these results indicated the satisfactory performance of our prediction method.
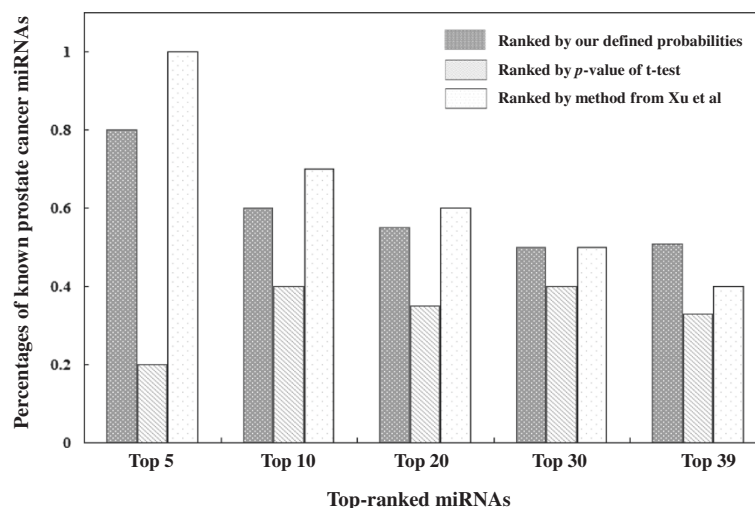
Figure 4 Performance comparison of the prediction of prostate cancer miRNAs. This figure summarizes the performance comparison of three PCa miRNA prediction methods – miRNA expression fold-change obtained by the t-test based method, the cancer miRNA synergism theory based method, and our integrative probability based method. The y-axis represents the percentage of known PCa abnormal miRNAs in the top ranked miRNA sets, and the x-axis indicates the top ranked miRNAs from these three approaches.

In vitro validation of candidate prostate cancer miRNA biomarkers

To further investigate the activities of PCa miRNA biomarkers predicted by our approach, we randomly selected three miRNAs (*miR-155*, *miR-648*, and *miR-197*) to detect differences in their expression between PCa tissues and benign prostatic hyperplasia tissues by q-PCR technology. The detection results are shown in Figure 5. Two out of 3 miRNAs (66.7%) were differentially expressed in the two groups of samples (*p*-value < 0.01).

Among these two miRNAs, *miR-155* downregulation in PCa was in agreement with the results of a previous report [45], whereas *miR-648* was identified as a novel PCa miRNA biomarker by our study. Although *miR-197* did not show significant outlier activity in PCa, it was previously proposed as a potential miRNA biomarker for lung cancer in another study [46]. The experimental analysis of the activities of outlier miRNAs verified the reliability of our method.

Functional analysis of target genes of candidate prostate cancer miRNA biomarkers

In our study, the predicted candidate PCa miRNA biomarkers, along with their uniquely regulated genes (see Additional file 3), provide potential miRNA-mRNA target pairs in PCa. The unique target genes regulated by these candidate miRNAs may also be involved in PCa, assuming that the predicted miRNAs are true PCa miRNAs. To validate our hypothesis, we retrieved the uniquely regulated genes of our predicted miRNAs from the PCa miRNA-mRNA target network, and explored their relationships with PCa through GO analysis and Pathway enrichment analysis (see Methods).

The Database for Annotation, Visualization, and Integrated Discovery (DAVID) [38] was applied for the Gene ontology (GO) annotation at three levels: molecular function, biological process, and cellular component. The ten most highly enriched items for each domain are shown in Figure 6. The results indicated that these genes are well mapped in several PCa associated biological processes, such as cell cycle [47] and regulation of apoptosis [48]. These results further confirmed the accuracy of the predicted PCa miRNA biomarkers to a certain extent.

We used the DAVID and the MetaCore™ Database from GeneGo Inc. to map these outlier genes in KEGG pathways and MetaCore™ pathways, respectively. Statistically significantly enriched pathways (p-value < 0.01) from these two datasets were retrieved. For KEGG pathways, we found that these target genes were only considerably enriched in one PCa related pathway termed as "Pathways in cancer". The top 10 highly enriched pathways from the GeneGo Database were plotted in Figure 7. Significantly enriched pathways, such as cytoskeleton remodeling [49], have been previously shown to be PCa related pathways. Text mining searches in NCBI PubMed were used to explore the relationships between enriched pathways and PCa. The results showed that 20 out of 24 (87.5%) pathways found in the GeneGo database were related to PCa. The PubMed citations regarding the association of these pathways with PCa are listed in Additional file 5. We also compared our enrichment pathways with novel PCa associated pathways detected using other systematic methods [50]. Of 24 enriched pathways from the
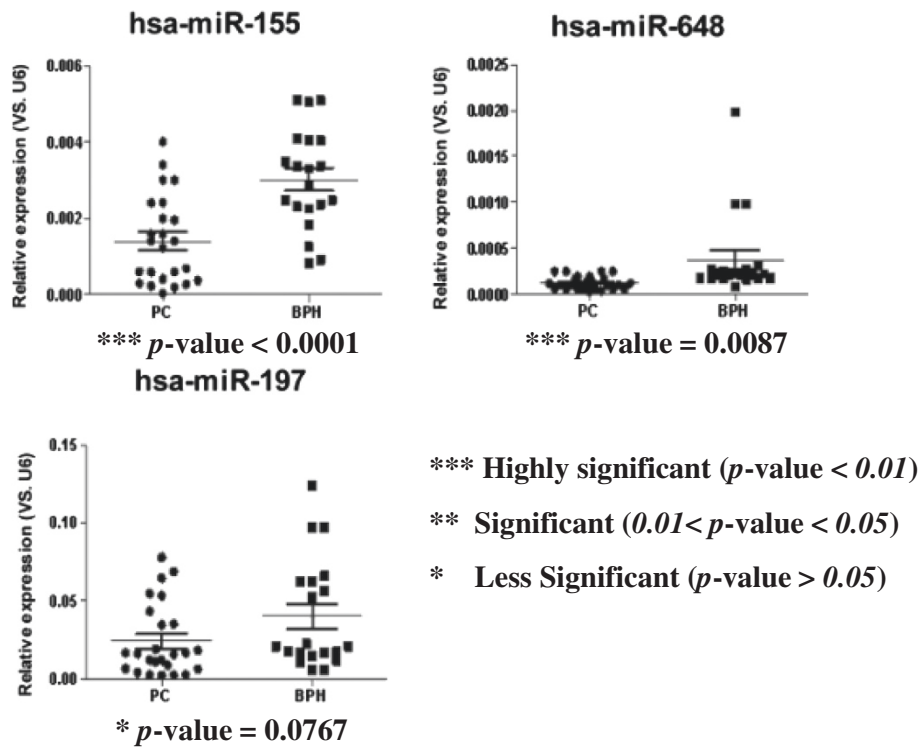
Figure 5 q-PCR results for selected candidate prostate cancer miRNA biomarkers. Three miRNAs were randomly selected from our prediction results to detect their outlier activity in PCa samples by using the q-PCR method. Endogenous U6 expression was used as control for normalization. The statistical significance of differences between groups was calculated using the Student's t-test.
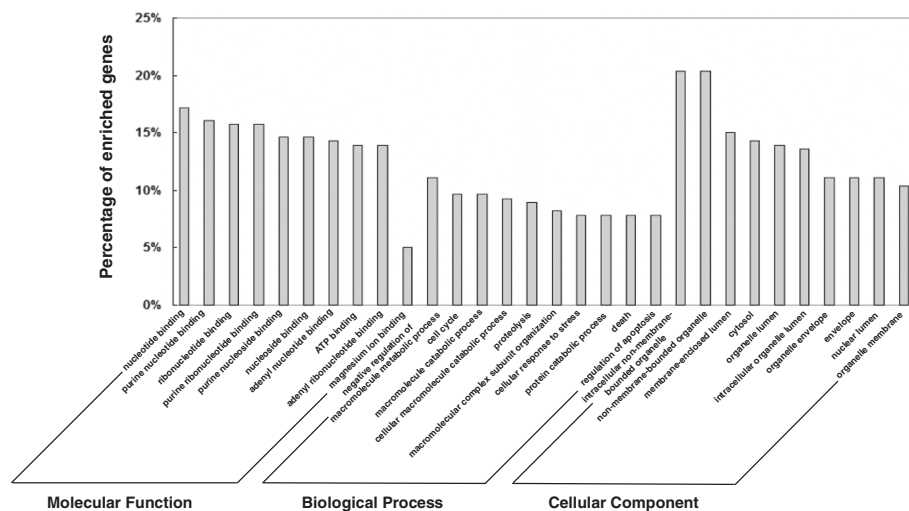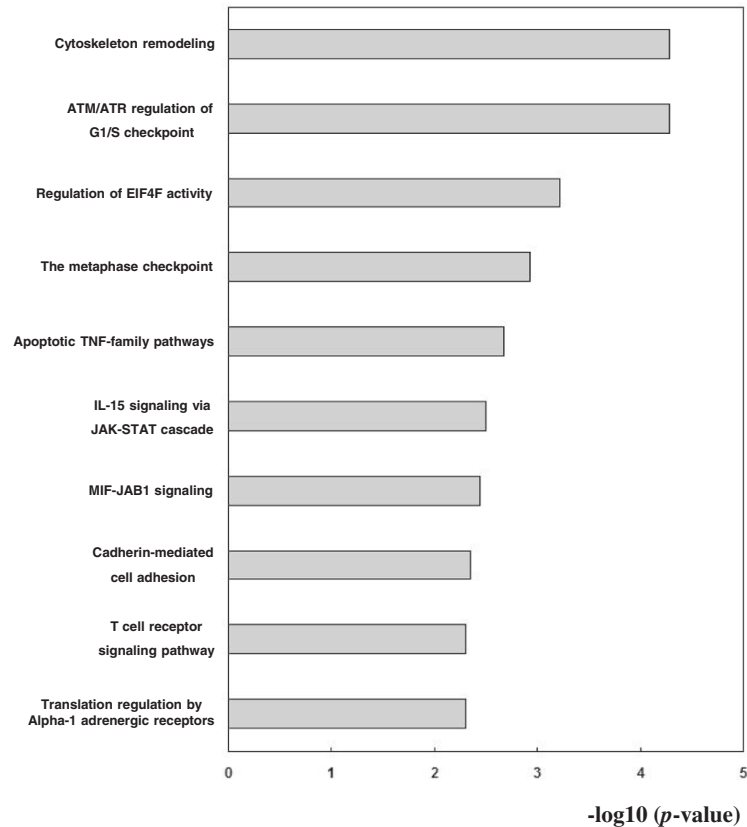

Figure 6 Gene ontology annotation for the uniquely regulated genes of the predicted prostate cancer miRNA biomarkers. The uniquely regulated genes of the predicted PCa miRNA biomarkers identified by our method were retrieved and annotated with DAVID tools in three domains of gene ontology: Molecular Function, Biological Process, and Cellular Component. The top 10 significantly enriched items for each domain are shown.

Figure 7 GeneGo pathway enrichment analysis for uniquely regulated genes of predicted prostate cancer miRNA biomarkers. The uniquely regulated genes of the predicted PCa miRNA biomarkers from our method were retrieved and mapped in the GeneGo Database. The statistical significance level (*p*-value) was negative 10-based log transformed. The top 10 significantly enriched pathways are shown.
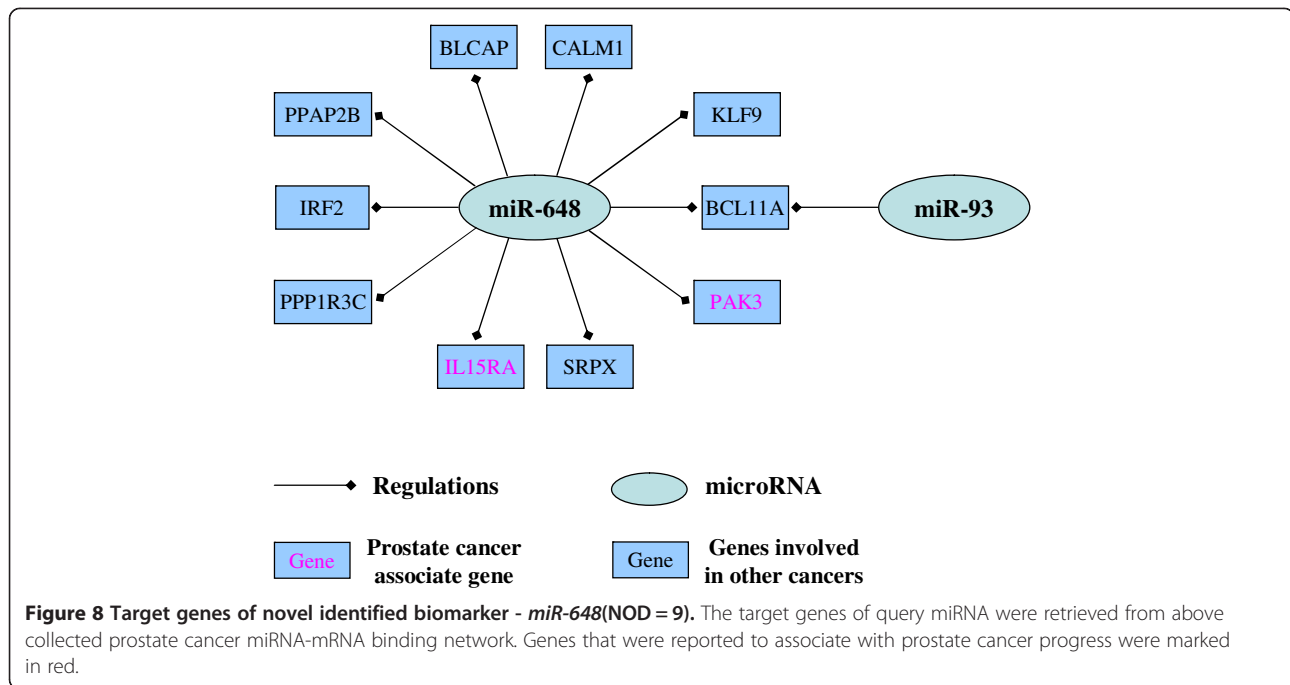
GeneGo database in our study, 18 (75%) were common pathways with those of previous studies. Taken together, these analyses confirmed the correlation between the target genes and PCa, and thus verified the reliability of our predicted PCa miRNAs.

Discussion

Previous studies have provided evidence of multiple-to-multiple relationships between miRNAs and their target genes [41,42,51]. From the average view of the miRNA-mRNA target network, that conclusion seems reasonable. Indeed, there are on average 51 target genes for each individual miRNA, and 4 co-regulator miRNAs for each gene in the whole miRNA-mRNA targeting network according to our analysis of the reconstructed network, as described in the Methods section. Based on this theory, numerous attempts have been made to predict cancer related miRNA regulatory modules [14], and cancer miRNAs were shown to have more synergism with their co-regulatory effects on the same genes [18].

In the present study, we conducted an in-depth exploration of this matter. Our results revealed the scale-free feature of the miRNA-mRNA target network. We introduced the NOD index to measure the independent regulatory power of an individual miRNA. Contrary to the functional synergism of cancer miRNAs, our data showed that miRNAs with greater independent regulatory power were more likely to be potential biomarkers in human cancers. Based on this evidence, we developed a novel integrative method to infer candidate cancer miRNA biomarkers from the miRNA regulatory network by linking paired miRNA and gene expression data, and highly reliable miRNA-mRNA target data. This pipeline was further applied to PCa. A total of 39 miRNAs were predicted as potential PCa miRNA signatures. Among these miRNAs, 20 have previously been reported to be PCa aberrant miRNAs by low-throughput methods, and 16 miRNAs were shown to be deregulated in other cancers. In vitro q-PCR experiments and functional analysis further verified the accuracy of our predictions and miR-648 was identified as a novel candidate PCa miRNA biomarker. The target genes of miR-648 are listed in Figure 8. With the exception of BCL11A, the other nine genes are exclusively regulated by miR-648.

Figure 8 Target genes of novel identified biomarker - *miR-648*(NOD = 9). The target genes of query miRNA were retrieved from above collected prostate cancer miRNA-mRNA binding network. Genes that were reported to associate with prostate cancer progress were marked in red.

Among these 10 genes, PAK3 and IL15RA were previously shown to be involved in PCa progression [52,53], and PAK3 was identified as an epigenetic biomarker for the prognostic diagnosis of PCa [52]. Despite the fact that no reports on the association of the remaining genes with PCa were found, these genes were shown to be associated with the progression of other cancers, such as breast cancer [54] pancreatic cancer [55] and leukemia [56]. These evidences support that miR-648 could be involved in PCa and act as a novel molecular biomarker for PCa diagnosis.

The miRNA-mRNA network reconstructed in this study consisted of experimentally validated data and computational predicted data. The data resources of the computational prediction databases used in this study, HOCTAR [25], ExprTargetDB [26], and starBase [27], were derived from the predictions based on gene expression information, such as microarray data and Next-generation sequencing data. Therefore, this predicted data should be more accurate than data predicted by programs merely based on sequence level, such as TargetScan [57] and RNAhybrid [58]. The reliable miRNA-mRNA targeting data could guarantee the accuracy of the predicted activities of outlier miRNAs.

The present results provide a basis for the development of algorithms for cancer miRNA biomarker identification. Indeed, two points require further improvement. Firstly, as the gene transcriptional expression data do not reflect changes in protein expression levels [11], the cancer miRNA activity cannot be predicted by our method for miRNAs that function through translational repression.

Secondly, the detailed outlier patterns (up-regulation or down-regulation) for the prediction of outlier miRNAs need to be further explored. The integration of protein expression data, transcription factor (TF) information and other omics data is a potential method to improve the prediction. This information will be incorporated in future studies aimed at further developing and refining our method.

Conclusions

The present analysis revealed novel distinctive features of cancer miRNA biomarkers. A novel bioinformatics framework was proposed to infer candidate cancer miRNA biomarkers from a miRNA regulatory network. The methodology may accelerate the discovery of novel miRNA signatures for cancer diagnosis and treatment, and should also be feasible for the study of other diseases.

Additional files

Additional file 1: MicroRNA and gene expression datasets. Paired miRNA and gene expression profiles from the same samples of four prostate cancer and four benign prostate hyperplasia individuals were used. Detailed information about these datasets is provided in this table.

Additional file 2: Comparison of the prediction prostate cancer miRNA biomarkers with different outlier miRNA (gene) thresholds. MicroRNAs with high statistical significance (*p*-value < 0.01, Wilcoxon signed-rank test) were considered as candidate prostate cancer miRNA biomarkers. MiRNAs common to three groups are highlighted in yellow, while miRNAs common to two groups are marked in red.

Competing interests

The authors declare they have no competing interests.

Authors' contributions

WZ designed the miRNA biomarker prediction pipeline, performed the statistical analysis and drafted the manuscript. FG and JZ carried out the in vitro validation experiments. XJ performed the data collection process. WY, ZS and DY participated in the functional enrichment analysis. BS conceived and coordinated the overall study and revised the manuscript. All authors read and approved the final manuscript.

Author details

[1]Center for Systems Biology, Soochow University, Suzhou 215006, China. [2]Department of Urology, the First Affiliated Hospital of Soochow University, Suzhou 215006, China. [3]Department of Urology, the Second Affiliated Hospital of Soochow University, Suzhou 215004, China. [4]Central lab, the First Affiliated Hospital of Soochow University, Suzhou 215006, China.

References

1. Bartel DP: **MicroRNAs: genomics, biogenesis, mechanism, and function.** *Cell* 2004, **116**:281–297.
2. Dar AA, Majid S, de Semir D, Nosrati M, Bezrookove V, Kashani-Sabet M: **miRNA-205 suppresses melanoma cell proliferation and induces senescence via regulation of E2F1 protein.** *J Biol Chem* 2011, **286**:16606–16614.
3. Kulozik AE: **Stay tuned: miRNA expression and nonsense-mediated decay in brain development.** *Mol Cell* 2011, **42**:407–408.
4. He JF, Luo YM, Wan XH, Jiang D: **Biogenesis of MiRNA-195 and its role in biogenesis, the cell cycle, and apoptosis.** *J Biochem Mol Toxicol* 2011, **25**:404–408.
5. Humeau M, Torrisani J, Cordelier P: **miRNA in clinical practice: pancreatic cancer.** *Clin Biochem* 2013, **46**:933–936.
6. Kojima S, Naya Y, Ichikawa T, Seki N: **MiRNA profiling in prostate cancer.** *Nihon Rinsho* 2011, **69**(Suppl 5):92–95.
7. Pignot G, Cizeron-Clairac G, Vacher S, Susini A, Tozlu S, Vieillefond A, Zerbib M, Lidereau R, Debre B, Amsellem-Ouazana D, Bieche I: **MicroRNA expression profile in a large series of bladder tumors: identification of a 3-miRNA signature associated with aggressiveness of muscle-invasive bladder cancer.** *Int J Cancer* 2013, **132**:2479–2491.
8. Lerebours F, Cizeron-Clairac G, Susini A, Vacher S, Mouret-Fourme E, Belichard C, Brain E, Alberini JL, Spyratos F, Lidereau R, Bieche I: **miRNA expression profiling of inflammatory breast cancer identifies a 5-miRNA signature predictive of breast tumor aggressiveness.** *Int J Cancer* 2013, **133**:1614–1623.
9. Chen J, Zhang D, Zhang W, Tang Y, Yan W, Guo L, Shen B: **Clear cell renal cell carcinoma associated microRNA expression signatures identified by an integrated bioinformatics analysis.** *J Transl Med* 2013, **11**:169.
10. Tang Y, Yan W, Chen J, Luo C, Kaipia A, Shen B: **Identification of novel microRNA regulatory pathways associated with heterogeneous prostate cancer.** *BMC Syst Biol* 2013, **7**(3):S6.
11. Madden SF, Carpenter SB, Jeffery IB, Bjorkbacka H, Fitzgerald KA, O'Neill LA, Higgins DG: **Detecting microRNA activity from gene expression data.** *BMC Bioinforma* 2010, **11**:257.
12. Xu J, Li CX, Lv JY, Li YS, Xiao Y, Shao TT, Huo X, Li X, Zou Y, Han QL, Li X, Wang LH, Ren H: **Prioritizing candidate disease miRNAs by topological features in the miRNA target-dysregulated network: case study of prostate cancer.** *Mol Cancer Ther* 2011, **10**:1857–1866.
13. Yoon S, De Micheli G: **Prediction of regulatory modules comprising microRNAs and target genes.** *Bioinformatics* 2005, **21**(Suppl 2):ii93–ii100.
14. Zhang W, Edwards A, Fan W, Flemington EK, Zhang K: **miRNA-mRNA correlation-network modules in human prostate cancer and the differences between primary and metastatic tumor subtypes.** *PLoS One* 2012, **7**:e40130.
15. Kim SJ, Ha JW, Zhang BT: **Constructing higher-order miRNA-mRNA interaction networks in prostate cancer via hypergraph-based learning.** *BMC Syst Biol* 2013, **7**:47.
16. Bonnet E, Michoel T, Van de Peer Y: **Prediction of a gene regulatory network linked to prostate cancer from gene expression, microRNA and clinical data.** *Bioinformatics* 2010, **26**:i638–i644.
17. Zhang S, Li Q, Liu J, Zhou XJ: **A novel computational framework for simultaneous integration of multiple types of genomic data to identify microRNA-gene regulatory modules.** *Bioinformatics* 2011, **27**:i401–i409.
18. Bandyopadhyay S, Mitra R, Maulik U, Zhang MQ: **Development of the human cancer microRNA network.** *Silence* 2010, **1**:6.
19. He HC, Zhu JG, Chen XB, Chen SM, Han ZD, Dai QS, Ling XH, Fu X, Lin ZY, Deng YH, Qin GQ, Cai C, Chen JH, Zhong WD: **MicroRNA-23b downregulates peroxiredoxin III in human prostate cancer.** *FEBS Lett* 2012, **586**:2451–2458.
20. Chen JH, He HC, Jiang FN, Militar J, Ran PY, Qin GQ, Cai C, Chen XB, Zhao J, Mo ZY, Chen YR, Zhu JG, Liu X, Zhong WD: **Analysis of the specific pathways and networks of prostate cancer for gene expression profiles in the Chinese population.** *Med Oncol* 2012, **29**:1972–1984.
21. Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T: **miRecords: an integrated resource for microRNA-target interactions.** *Nucleic Acids Res* 2009, **37**:D105–D110.
22. Sethupathy P, Corda B, Hatzigeorgiou AG: **TarBase: A comprehensive database of experimentally supported animal microRNA targets.** *RNA* 2006, **12**:192–197.
23. Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, Li M, Wang G, Liu Y: **miR2Disease: a manually curated database for microRNA deregulation in human disease.** *Nucleic Acids Res* 2009, **37**:D98–D104.
24. Hsu SD, Lin FM, Wu WY, Liang C, Huang WC, Chan WL, Tsai WT, Chen GZ, Lee CJ, Chiu CM, Chien CH, Wu MC, Huang CY, Tsou AP, Huang HD: **miRTarBase: a database curates experimentally validated microRNA-target interactions.** *Nucleic Acids Res* 2011, **39**:D163–D169.
25. Gennarino VA, Sardiello M, Avellino R, Meola N, Maselli V, Anand S, Cutillo L, Ballabio A, Banfi S: **MicroRNA target prediction by expression analysis of host genes.** *Genome Res* 2009, **19**:481–490.
26. Gamazon ER, Im HK, Duan S, Lussier YA, Cox NJ, Dolan ME, Zhang W: **Exprtarget: an integrative approach to predicting human microRNA targets.** *PLoS One* 2010, **5**:e13534.
27. Yang JH, Li JH, Shao P, Zhou H, Chen YQ, Qu LH: **starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data.** *Nucleic Acids Res* 2011, **39**:D202–D209.
28. Chen J, Qian F, Yan W, Shen B: **Translational biomedical informatics in the cloud: present and future.** *Biomed Res Int* 2013, **2013**:658925.

29. Chen J, Wang Y, Shen B, Zhang D: **Molecular signature of cancer at gene level or pathway level? Case studies of colorectal cancer and prostate cancer microarray data.** *Comput Math Methods Med* 2013, **2013**:909525.

30. Chen J, Zhang D, Yan W, Yang D, Shen B: **Translational bioinformatics for diagnostic and prognostic prediction of prostate cancer in the next-generation sequencing era.** *Biomed Res Int* 2013, **2013**:901578.

31. Jiang J, Cui W, Vongsangnak W, Hu G, Shen B: **Post genome-wide association studies functional characterization of prostate cancer risk loci.** *BMC Genomics* 2013, **14**(8):S9.

32. Roh YH, Noh JH, Kim W, Oh JH, Gong HS, Baek GH: **Cross-cultural adaptation and validation of the Korean version of the Oxford shoulder score.** *Arch Orthop Trauma Surg* 2012, **132**:93–99.

33. Wei K, Chen J, Chen Y, Wu L, Xie D: **Multiple-strategy analyses of ZmWRKY subgroups and functional exploration of ZmWRKY genes in pathogen responses.** *Mol Biosyst* 2012, **8**:1940–1949.

34. Kurahashi I, Fujita Y, Arao T, Kurata T, Koh Y, Sakai K, Matsumoto K, Tanioka M, Takeda K, Takiguchi Y, Yamamoto N, Tsuya A, Matsubara N, Mukai H, Minami H, Chayahara N, Yamanaka Y, Miwa K, Takahashi S, Nakagawa K, Nishio K: **A microarray-based gene expression analysis to identify diagnostic biomarkers for unknown primary cancer.** *PLoS One* 2013, **8**:e63249.

35. Stoss O, Werther M, Zielinski D, Middel P, Jost N, Ruschoff J, Henkel T, Albers P: **Transcriptional profiling of transurethral resection samples provides insight into molecular mechanisms of hormone refractory prostate cancer.** *Prostate Cancer Prostatic Dis* 2008, **11**:166–172.

36. Tamura K, Furihata M, Tsunoda T, Ashida S, Takata R, Obara W, Yoshioka H, Daigo Y, Nasu Y, Kumon H, Konaka H, Namiki M, Tozawa K, Kohri K, Tanji N, Yokoyama M, Shimazui T, Akaza H, Mizutani Y, Miki T, Fujioka T, Shuin T, Nakamura Y, Nakagawa H: **Molecular features of hormone-refractory prostate cancer cells by genome-wide gene expression profiles.** *Cancer Res* 2007, **67**:5117–5125.

37. Hsu PW, Huang HD, Hsu SD, Lin LZ, Tsou AP, Tseng CP, Stadler PF, Washietl S, Hofacker IL: **miRNAMap: genomic maps of microRNA genes and their target genes in mammalian genomes.** *Nucleic Acids Res* 2006, **34**:D135–D139.

38. da Huang W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nat Protoc* 2009, **4**:44–57.

39. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M: **KEGG: Kyoto Encyclopedia of Genes and Genomes.** *Nucleic Acids Res* 1999, **27**:29–34.

40. Liu G, Ding M, Chen J, Huang J, Wang H, Jing Q, Shen B: **Computational analysis of microRNA function in heart development.** *Acta Biochim Biophys Sin (Shanghai)* 2010, **42**:662–670.

41. Jayaswal V, Lutherborrow M, Ma DD, Yang YH: **Identification of microRNA-mRNA modules using microarray data.** *BMC Genomics* 2011, **12**:138.

42. Tran DH, Satou K, Ho TB: **Finding microRNA regulatory modules in human genome using rule induction.** *BMC Bioinforma* 2008, **9**(12):S5.

43. Delprato A: **Topological and functional properties of the small GTPases protein interaction network.** *PLoS One* 2012, **7**:e44882.

44. Yang Z, Ren F, Liu C, He S, Sun G, Gao Q, Yao L, Zhang Y, Miao R, Cao Y, Zhao Y, Zhong Y, Zhao H: **dbDEMC: a database of differentially expressed miRNAs in human cancers.** *BMC Genomics* 2010, **11**(4):S5.

45. Roa W, Brunet B, Guo L, Amanie J, Fairchild A, Gabos Z, Nijjar T, Scrimger R, Yee D, Xing J: **Identification of a new microRNA expression profile as a potential cancer screening tool.** *Clin Invest Med* 2010, **33**:E124.

46. Du L, Schageman JJ, Subauste MC, Saber B, Hammond SM, Prudkin L, Wistuba II, Ji L, Roth JA, Minna JD, Pertsemlidis A: **miR-93, miR-98, and miR-197 regulate expression of tumor suppressor gene FUS1.** *Mol Cancer Res* 2009, **7**:1234–1243.

47. Luo J, Liu X: **Polo-like kinase 1, on the rise from cell cycle regulation to prostate cancer development.** *Protein Cell* 2012, **3**:182–197.

48. Gurumurthy S, Vasudevan KM, Rangnekar VM: **Regulation of apoptosis in prostate cancer.** *Cancer Metastasis Rev* 2001, **20**:225–243.

49. Wang Q, Symes AJ, Kane CA, Freeman A, Nariculam J, Munson P, Thrasivoulou C, Masters JR, Ahmed A: **A novel role for Wnt/Ca2+ signaling in actin cytoskeleton remodeling and cell motility in prostate cancer.** *PLoS One* 2010, **5**:e10456.

50. Wang Y, Chen J, Li Q, Wang H, Liu G, Jing Q, Shen B: **Identifying novel prostate cancer associated pathways based on integrative microarray data analysis.** *Comput Biol Chem* 2011, **35**:151–158.

51. Friedman RC, Farh KK, Burge CB, Bartel DP: **Most mammalian mRNAs are conserved targets of microRNAs.** *Genome Res* 2009, **19**:92–105.

52. Rosenbaum E, Begum S, Brait M, Zahurak M, Maldonado L, Mangold LA, Eisenberger MA, Epstein JI, Partin AW, Sidransky D, Hoque MO: **AIM1 promoter hypermethylation as a predictor of decreased risk of recurrence following radical prostatectomy.** *Prostate* 2012, **72**:1133–1139.

53. Tourkova IL, Yurkovetsky ZR, Gambotto A, Makarenkova VP, Perez L, Balkir L, Robbins PD, Shurin MR, Shurin GV: **Increased function and survival of IL-15-transduced human dendritic cells are mediated by up-regulation of IL-15Ralpha and Bcl-2.** *J Leukoc Biol* 2002, **72**:1037–1045.

54. Gromova I, Gromov P, Kroman N, Wielenga VT, Simon R, Sauter G, Moreira JM: **Immunoexpression analysis and prognostic value of BLCAP in breast cancer.** *PLoS One* 2012, **7**:e45967.

55. Bommer GT, Jager C, Durr EM, Baehs S, Eichhorst ST, Brabletz T, Hu G, Frohlich T, Arnold G, Kress DC, Goke B, Fearon ER, Kolligs FT: **DRO1, a gene down-regulated by oncogenes, mediates growth inhibition in colon and pancreatic cancer cells.** *J Biol Chem* 2005, **280**:7962–7975.

56. Choo A, Palladinetti P, Holmes T, Basu S, Shen S, Lock RB, O'Brien TA, Symonds G, Dolnikov A: **siRNA targeting the IRF2 transcription factor inhibits leukaemic cell growth.** *Int J Oncol* 2008, **33**:175–183.

57. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP: **MicroRNA targeting specificity in mammals: determinants beyond seed pairing.** *Mol Cell* 2007, **27**:91–105.

58. Kruger J, Rehmsmeier M: **RNAhybrid: microRNA target prediction easy, fast and flexible.** *Nucleic Acids Res* 2006, **34**:W451–W454.