Identification of hub miRNA biomarkers for bladder cancer by weighted gene coexpression network analysis

Abstract: Bladder cancer (BC) is a common urinary system tumor with high aggressiveness, and it results in relatively high mortality due to a lack of precise and suitable biomarkers. In this study, we applied the weighted gene coexpression network analysis method to miRNA expression data from BC patients, and screened for network modules associated with BC progression. Hub miRNAs were selected, followed by functional enrichment analyses of their target genes for the most closely related module. These hub miRNAs were found to be involved in several functional pathways including pathway in cancer, regulation of actin cytoskeleton, PI3K-Akt signaling pathway, mitogen-activated protein kinase (MAPK) signaling pathway, Wnt signaling pathway, proteoglycans in cancer, focal adhesion and p53 signaling pathway via regulating target genes. Finally, their prognostic significance was tested using analyses of overall survival. A few novel prognostic miRNAs were identified based on expression profiles and related survival data. In conclusion, several miRNAs that were critical in BC initiation and progression have been identified in this study. These miRNAs, which may contribute to a comprehensive understanding of the pathogenesis of BC, could serve as potential biomarkers for BC prognosis or as new therapeutic targets.

Keywords: bladder cancer, miRNA expression, functional enrichment analysis, bioinformatics analysis

Introduction

Bladder cancer (BC) is one of the most common malignant cancers of the urinary system and is associated with an annual mortality of over 150,000.1,2 Bladder transitional cell carcinoma accounts for more than 90% of these deaths. Approximately 70%–75% of transitional cell carcinoma cases are non-muscle invasive bladder cancer and the remaining 25%–30% are muscle invasive bladder cancer. Approximately 70% of non-muscle invasive bladder cancer patients ultimately develop relapse within 5 years after undergoing transurethral resection of BC. Meanwhile, muscle invasive bladder cancer is likely to be linked with lethal metastasis at the time of diagnosis.3–5 To date, sensitive and specific markers for the diagnosis and treatment of BC have not been found. Therefore, it is crucial to search for biomarkers to predict the development and prognosis of BC, which can also be used as therapeutic targets.

miRNAs are constituted by a large number of naturally occurring endogenous small non-coding RNAs. miRNAs, with lengths of 22–25 nucleotides, can regulate gene expression through translational inhibition or mRNA cleavage. A variety of studies have shown that miRNAs play important regulatory roles in diverse physiologic processes, including cell proliferation, adhesion and migration.7,8 Furthermore,
miRNAs can target multiple mRNAs, which render them powerful regulators. For example, loss of miR-218, a tumor suppressor, enhances cancer cell migration and invasion in prostate cancer through direct regulation of LIM and SH3 domain protein 1 (LAP1), which is a member of the nebulin family of actin-binding proteins.\textsuperscript{9} miR-124/miR-9 plays a role in the neuronal development of primary neurons by regulating glycoprotein M6A (GPM6A), a transmembrane glycoprotein.\textsuperscript{10} Moreover, a study has shown that miR-1 inhibits cell proliferation, invasion, metastasis and further progression of esophageal squamous cell carcinoma by binding to its target gene, TAGLN2, which also belongs to the family of actin-binding proteins.\textsuperscript{11} In addition, abnormal miRNA expression may interfere with many cellular signaling pathways; moreover, it can profoundly affect the pathogenesis and progression of cancer. Abnormal expression of miRNAs has been observed in several kinds of human tumors, including BC.\textsuperscript{12-14} The role of miRNAs in BC is a promising area of research to better understand the mechanisms of BC initiation, development and metastasis.

The rapid growth of high-throughput microarray technology in recent years has provided new methods to identify BC progression-related genes. Previously, identification of these genes was accomplished through basic gene expression analysis methods. Molecular mechanisms can be partially understood by screening differentially expressed genes or differentially expressed miRNAs (DEMs).\textsuperscript{15,16} Moreover, recent bioinformatics advancements allow us to apply the weighted gene coexpression network analysis (WGCNA) algorithm to interpret expression profiles, which ensure more comprehensive identification of new biomarkers.\textsuperscript{17} WGCNA can detect highly coexpressed genes that are closely connected within the network based on similarities in sample expression profiles. These subnetwork regions are called modules and are usually composed of functionally related genes.\textsuperscript{17,18} Therefore, the WGCNA algorithm for BC microarray miRNA expression data was performed in order to construct a coexpression network based on interactions between DEMs. In the meantime, hub miRNAs potentially associated with the pathogenesis of BC were identified and validated as prognostic biomarkers. These miRNAs may further reveal the biologic mechanisms of BC initiation and progression and may also serve as potential therapeutic targets.

**Materials and methods**

**Acquisition of gene expression data and DEMs**

The raw miRNA expression data from 418 BC and 19 normal tissue samples were downloaded from the Cancer Genome Atlas (TCGA).\textsuperscript{19} DEMs between BC and normal samples were selected using the “edgeR” R package.\textsuperscript{20} miRNAs with a false discovery rate of <0.05 and | log₂ fold change | of >0.585 were identified as DEMs.

**WGCNA and the identification of clinically significant modules**

As reported above, the WGCNA package provided an approach to constructing miRNA modules within a network based on correlations between DEM expression profiles.\textsuperscript{21,22} First, a Pearson’s correlation matrix was computed for all pair-wise miRNAs. Subsequently, an adjacency matrix was constructed using a power adjacency function \( \alpha_{mm} = \text{Power} (S_{mn}^\beta) = |S_{mn}|^\beta \), \( \alpha_{mm} = \text{adjacency between two miRNAs}, S_{mn} = \text{Pearson’s correlations between two miRNAs}. \) \( \beta \) is a soft thresholding parameter allowing for a maximal correlation coefficient. The power of \( \beta = 3 \) (model fitting index \( R^2 = 0.83 \)) was chosen in accordance with the scale-free topology criterion. Next, the topological overlap matrix was calculated from the adjacency matrix to assess the interconnectedness of the miRNAs.\textsuperscript{23} Finally, the “cutreeStaticColor” function was applied to classify similar expression miRNAs into gene modules (minModuleSize = 20 and mergeCutHeight = 0.95). To identify the modules that were significantly associated with BC development, gene significance (GS) was calculated according to the formula: \( GS_i = -\log(P_i) \), where \( P_i \) was obtained from the value of the student’s \( t \)-test (two-tailed) to identify the differential expression between two groups. GS represented the correlation between gene expression and disease progression. The module significance was defined as the mean GS in a module, while a module with a high module significance value was considered to be the key module.\textsuperscript{21}

**Detection of candidate hub DEMs**

Hub miRNAs were defined based on intramodular connectivity and module membership (MM) in our study. The intramodular connectivity was calculated as follows: \( K_i = \sum_j a_{ij} \), where \( i, j \) belongs to Module \( q \), \( K_i = \text{intramodular connectivity of gene } i \) and \( a_{ij} = \text{adjacency between genes } i \) and \( j \).\textsuperscript{24} Intramodular connectivity reflects the correlation between miRNAs in the coexpression network. The MM was computed as originally described, namely, \( MM_i = |\text{cor}(x(i), \text{Eigene})| \), where the module eigengene stands for the first principal component of a given module. The MM ranged from 0 to 1, which reflected the Pearson’s correlation coefficient between gene expression and the module eigengene in the network.\textsuperscript{25} In general, miRNAs showing a high degree of connectivity and MM values within a module were defined as “hub miRNAs”, which were
assumed to be functionally significant.22,24,26 The top 25% of miRNAs with the highest connectivity and MM values in the blue module were selected based on the module sizes as hub miRNAs for further study.

Bioinformatics analysis of miRNA target genes
Potential target genes of hub DEMs were retrieved from the target prediction database TargetScan (http://www.targetscan.org/). In addition, genes with a cumulative weighted context++ score of ≤−0.4 in the software were selected.27 Functional Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis were performed using DAVID v6.8 (http://david.abcc.ncifcrf.gov/) bioinformatics online resources in order to uncover the biologic function of target genes. A P-value cutoff of <0.01 was set, and at least five genes were selected as the enriched function of target genes.

Human tissue specimens
Human tumor tissues and adjacent normal tissue specimens were obtained from the Department of Urology at Nanjing First Hospital, Nanjing Medical University. Ethical approval was obtained from the ethics committee of Nanjing First Hospital and written informed consent was obtained from all patients. The specimens were collected during surgery and frozen in liquid nitrogen.

Western blotting
Total protein was extracted from the tumor tissues and adjacent normal tissue specimens using radioimmunoprecipitation assay lysis buffer containing a protease inhibitor cocktail (Roche, Shanghai, China). The protein concentration was quantified using a bicinchoninic acid protein assay kit (Beyotime, Shanghai, China). Equivalent amounts of proteins were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis, transferred to polyvinylidene fluoride membranes (EMD Millipore, Billerica, MA, USA) and blocked with 5% non-fat milk in distilled water for 2 h at room temperature. Later, the membranes were incubated with diluted anti-TAGLN2 primary antibodies (Abcam, #ab121146) overnight at 4°C, followed by incubation with horseradish peroxidase-conjugated secondary antibody (YI FEI XUE BIOTECHNOLOGY, Nanjing, China) for 2 h at room temperature. β-Actin was used as an internal control.

Human protein atlas
Expression of the TAGLN2 protein in BC tissues and normal bladder tissues was determined from immunohistochemical data publicly available in the Human Protein Atlas (www.proteinatlas.org).28

Survival analyses
Survival analyses in this study were performed using the TCGA database. The miRNA expression profile and survival data from 406 BC patients were used in our study after eliminating patients without available clinical information. The prognostic values of signature candidate miRNA profiles were assessed using GraphPad Prism 7. BC samples were divided into two groups according to their median proposed miRNA expression. The survival of two patient cohorts with candidate miRNAs was compared using a Kaplan–Meier survival plot. Moreover, log-rank P-values and the hazard ratio (HR) with 95% CIs were also calculated.

Results
Identification of DEMs
Using TCGA miRNA expression data, DEM expression profiles from BC and normal tissue samples were acquired by the “edgeR” package in R, with $P<0.05$ and $\log_{2}(\text{fold change})>0.585$ being selected as the criteria. A total of 480 DEMs were identified after analysis, among which 350 genes were upregulated while 130 were downregulated (Figure 1; Table S1).
expression patterns were divided into modules adopting the average hierarchical clustering linkage. Modules containing at least 20 miRNAs were extracted, and 5 modules were eventually identified (Figure 2A). The blue module had the highest mean GS among these modules (Figure 2B), and the correlation coefficient ($r$) of this module indicated a significant correlation of intramodular connectivity with gene significance (Figure 2C). This suggested that the blue module was closely related to BC progression. The top 25% of the 54 miRNAs (Table 1) showing the highest connectivity and MM values in the blue module were chosen as candidate hub miRNAs.29

**Hub miRNA target prediction and functional annotation**

Target genes of hub miRNAs were predicted using TargetScan V7.1. Those with cumulative weighted context++ scores of $\leq-0.4$ in the software were selected (Table S2).27 The miRNA regulatory network was visualized using CytoScape software, and miRNAs with many interactions were graphically highlighted (Figure 3). In total, 1351 predicted target genes were selected for further analysis. Altogether, 27 enriched GO biologic process terms (Figure 4A) and 17 KEGG pathways (Figure 4B) were obtained using a threshold of $P < 0.01$ within the online software DAVID. GO analysis results suggested that the most represented GO terms in biologic processes were related to cell proliferation and regulation of growth. On the other hand, target genes were found to be significantly enriched in cancer-related pathways, such as pathways in cancer, regulation of actin cytoskeleton, PI3-K-Akt signaling pathway, mitogen-activated protein kinase (MAPK) signaling pathway, Wnt signaling pathway, proteoglycans in cancer, focal adhesion and p53 signaling pathway.

**Validation of the identified target genes**

TAGLN2, the target gene with the highest repeatability, was randomly selected for further validation in order to determine the clinical relevance of target gene expression. It was a common target gene for miR-1–1/miR-1–2, miR-133a-1/miR-133a-2 and miR-133b. Western blots showed significantly higher TAGLN2 expression levels in BC tissue than in adjacent normal tissue specimens (Figure 5A), which was consistent with immunohistochemistry from the human protein map (www.proteinatlas.org, Figure 5B). The TAGLN2 mRNA level was consistently higher in BC tissues than in adjacent normal tissue specimens (Figure 5A), which was consistent with immunohistochemistry from the human protein map (www.proteinatlas.org, Figure 5B). The TAGLN2 expression also significantly increased the risk of advanced TNM stages (Figure 5E), as shown in

**Table 1** Top 25% of the 54 miRNAs showing the highest connectivity and MM values in the blue module

<table>
<thead>
<tr>
<th>microRNA</th>
<th>Module</th>
<th>Connectivity</th>
<th>Module membership</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-miR-133a-1</td>
<td>Blue</td>
<td>14.52815127</td>
<td>0.942911887</td>
</tr>
<tr>
<td>hsa-miR-133a-2</td>
<td>Blue</td>
<td>14.13870062</td>
<td>0.93183002</td>
</tr>
<tr>
<td>hsa-miR-1–2</td>
<td>Blue</td>
<td>13.22743045</td>
<td>0.921363767</td>
</tr>
<tr>
<td>hsa-miR-1–1</td>
<td>Blue</td>
<td>13.21699224</td>
<td>0.920517623</td>
</tr>
<tr>
<td>hsa-miR-133b</td>
<td>Blue</td>
<td>12.77790952</td>
<td>0.91496915</td>
</tr>
<tr>
<td>hsa-miR-143</td>
<td>Blue</td>
<td>11.9099187</td>
<td>0.890547722</td>
</tr>
<tr>
<td>hsa-miR-195</td>
<td>Blue</td>
<td>11.52715541</td>
<td>0.86839224</td>
</tr>
<tr>
<td>hsa-miR-139</td>
<td>Blue</td>
<td>11.20058022</td>
<td>0.854359765</td>
</tr>
<tr>
<td>hsa-miR-6507</td>
<td>Blue</td>
<td>10.72163066</td>
<td>0.82522312</td>
</tr>
<tr>
<td>hsa-miR-145</td>
<td>Blue</td>
<td>10.39113951</td>
<td>0.801384187</td>
</tr>
<tr>
<td>hsa-miR-28</td>
<td>Blue</td>
<td>10.2439735</td>
<td>0.8227846</td>
</tr>
<tr>
<td>hsa-miR-3199–2</td>
<td>Blue</td>
<td>9.04028532</td>
<td>0.818183518</td>
</tr>
<tr>
<td>hsa-miR-548a</td>
<td>Blue</td>
<td>8.813461621</td>
<td>0.812107848</td>
</tr>
</tbody>
</table>

Abbreviation: MM, module membership.
hsa-miR-195, hsa-miR-6507, hsa-miR-145 and hsa-miR-548ba miRNAs were negative factors for the survival of BC patients.

Discussion

BC is a fatal form of cancer, and the identification of new precise and suitable targets is of vital importance. Circulating miRNAs may serve as potential targets due to their involvement in various biologic processes, including cell proliferation, differentiation, apoptosis, metabolism and tumorigenesis. The identification of miRNA biomarkers and targets was accomplished by analyzing the microarray-based miRNA expression data obtained from TCGA database. Initially, 480 DEMs were identified in each group, including 350 upregulated and 130 downregulated miRNAs. A weighted miRNA coexpression network was subsequently constructed and several miRNA modules were identified based on DEM expression profiles. The blue module of coexpressed miRNAs was found to be of more importance.
miR-133b, which is located in chromosomal region 6p12.2, could induce apoptosis in renal cell carcinoma cell lines. miR-133a were remarkably inhibited in renal cell carcinoma specimens, while miR-1 and miR-133a restoration could promote the proliferation and invasion of BC cells. Furthermore, the enriched GO terms and KEGG pathways of hub miRNA target genes were also characterized by functional enrichment analysis. These target genes were found to be enriched in pathways in cancer, regulation of actin cytoskeleton, PI3K-Akt signaling pathway, MAPK signaling pathway, Wnt signaling pathway, proteoglycans in cancer, focal adhesion, p53 signaling pathway, MAPK signaling pathway, Wnt signaling pathway, proteoglycans in cancer, focal adhesion, p53 signaling pathway, cell proliferation and migration and chemoresistance.35,36 Furthermore, the enriched GO terms and KEGG pathways of hub miRNA target genes were also characterized by functional enrichment analysis. These target genes were found to be enriched in pathways in cancer, regulation of actin cytoskeleton, PI3K-Akt signaling pathway, MAPK signaling pathway, Wnt signaling pathway, proteoglycans in cancer, focal adhesion, p53 signaling pathway, cell proliferation and migration and chemoresistance.35,36

Figure 4 Functional annotations of the target genes.
Notes: (A) The 27 enriched GO biologic process terms. (B) The 17 enriched KEGG pathways.
Abbreviations: GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.
Figure 5 TAGLN2 is upregulated in human BC specimens.

Notes: (A, B) TAGLN2 expression in normal bladder tissue and BC specimens. (C) TAGLN2 expression in the TCGA BC RNAseq dataset (N=408). (D) Oncomine data showing TAGLN2 expression in normal (n=48) vs tumor (n=81) bladder specimens. (E) One-way ANOVA analysis showed that TAGLN2 expression levels were significantly associated with TNM stage in the TCGA BC RNAseq dataset.

Abbreviations: ANOVA, analysis of variance; BC, bladder cancer; N, normal; TCGA, The Cancer Genome Atlas; T, tumor.

Table 2 Overall survival of the hub miRNAs in the blue module

<table>
<thead>
<tr>
<th>Hub miRNA</th>
<th>HR (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-mir-133a-1</td>
<td>1.372 (1.026–1.834)</td>
<td>0.0327</td>
</tr>
<tr>
<td>hsa-mir-133a-2</td>
<td>1.284 (0.961–1.716)</td>
<td>0.0921</td>
</tr>
<tr>
<td>hsa-mir-1–1</td>
<td>1.205 (0.902–1.610)</td>
<td>0.2069</td>
</tr>
<tr>
<td>hsa-mir-1–1–1</td>
<td>1.222 (0.914–1.633)</td>
<td>0.1759</td>
</tr>
<tr>
<td>hsa-mir-133b</td>
<td>1.530 (1.145–2.045)</td>
<td>0.0498</td>
</tr>
<tr>
<td>hsa-mir-143</td>
<td>1.597 (1.195–2.134)</td>
<td>0.0118</td>
</tr>
<tr>
<td>hsa-mir-195</td>
<td>1.452 (1.086–1.940)</td>
<td>0.0127</td>
</tr>
<tr>
<td>hsa-mir-139</td>
<td>1.133 (0.848–1.514)</td>
<td>0.3977</td>
</tr>
<tr>
<td>hsa-mir-6507</td>
<td>1.585 (1.186–2.118)</td>
<td>0.0020</td>
</tr>
<tr>
<td>hsa-mir-145</td>
<td>1.705 (1.276–2.279)</td>
<td>0.0004</td>
</tr>
<tr>
<td>hsa-mir-28</td>
<td>0.823 (0.615–1.100)</td>
<td>0.1878</td>
</tr>
<tr>
<td>hsa-mir-3199–2</td>
<td>1.107 (0.828–1.479)</td>
<td>0.5110</td>
</tr>
<tr>
<td>hsa-mir-548ba</td>
<td>1.408 (1.054–1.882)</td>
<td>0.0212</td>
</tr>
</tbody>
</table>

Note: Bold type indicates statistical significance.

Abbreviations: CI, confidence interval; HR, hazard ratio.

regulation of growth. In particular, many of these enriched functions regulated by hub miRNAs participate in the formation and progression of multiple tumors. For instance, the MAPK cascade is a critical pathway for human cancer cell survival, dissemination and resistance to drug therapy. Numerous stimuli, including DNA damage pathways and internal metabolic stress, as well as signaling from external growth factors, cell–matrix interactions and communication from other cells, can cause increased protein amplification. Moreover, these stimuli can alter the tumor microenvironment, thus over-activating the pathway and leading to tumor progression, as other studies have shown.37,38 Studies have demonstrated that miR-1/miR-133a regulates the MAPK pathway by inhibiting the phosphorylation of extracellular
regulated protein kinases, which are involved in the metastasis of colorectal carcinoma. The regulation of actin cytoskeleton signaling pathways is identified as a cancer-related factor in multiple cancer datasets. Many key proteins are involved in this pathway, including WASP family proteins, Arp2/3 complexes, LIM-kinases, coflin and cortactin. These proteins are regulated by miR-143 and overexpressed in multiple cancers, and can promote the migration, invasion and metastasis of cancer cells. The p53 gene plays an important role in various functional activities, such as avoiding the accumulation of damaged DNA, maintaining genomic stability, regulating cell differentiation by inducing cell cycle arrest, promoting apoptosis and DNA repair senescence. Cell division and proliferation will be uncontrolled if the p53 gene is damaged by cancer-induced factors. Increasing evidence has highlighted the important roles of miRNAs in mediating tumor suppression functions of p53. Sachdeva et al have reported that miR-145 is involved in the regulation of the dynamic balance between p53, a tumor suppressor, and c-Myc, a proto-oncogene, so as to maintain normal cell growth and proliferation. The downregulation of miR-145 expression will subsequently trigger cell malignancy.

Notably, it is found that many genes are targeted by multiple hub miRNAs. For example, GPM6A, LASP1 and TAGLN2 are targeted by at least four hub miRNAs (Table S3). GPM6A has been shown in previous studies to promote the aggressiveness of lymphoid leukemia by inducing cell migration. Zhao et al demonstrated that LASP1 might promote liver, gut and mesentery metastases by using an orthotopic xenograft mouse model of pancreatic ductal adenocarcinoma, the upregulation of which was closely related to poor prognosis for pancreatic ductal adenocarcinoma patients. In this study, TAGLN2 was randomly selected for further verification. The results suggested that TAGLN2 expression was significantly higher in tumor tissue than in adjacent normal bladder tissue, and increased TAGLN2 expression also significantly increased the risk of advanced TNM stages, indicating that these genes may be closely associated with the development of BC. Thus, further studies regarding BC initiation and progression should be conducted on these genes independently from the proposed miRNAs.

Finally, the potential prognostic value of hub miRNAs was validated based on miRNA expression profiles in BC patients, as well as their survival information in the TCGA database. Seven out of the 11 hub miRNAs (including hsa-miR-133a-1, has-miR-133b, has-miR-143, has-miR-195 and hsa-miR-145) were found to be closely associated with the overall survival for BC patients.

Conclusion
To summarize, a key module significantly associated with BC progression has been confirmed in this study. Hub miRNAs within this module may play vital roles in the initiation and progression of BC, and may therefore be candidates for functional studies. Moreover, they were found to regulate a variety of cancer-related pathways at the functional level, such as pathways in cancer, regulation of actin cytoskeleton, MAPK, Wnt, proteoglycans in cancer, focal adhesion and p53-associated cellular signaling pathways, as well as cell proliferation and various cancer-related genes. In addition, some miRNAs were also found to be associated with overall survival for BC patients. Our study not only provides new insights into the initiation and progression of BC, but also identifies new potential biomarkers and therapeutic targets. Nonetheless, further studies are still needed.

Acknowledgments
This study was supported by grants from the National Natural Science Foundation of China (81370853 and 81570613) to RPJ, and Research and Innovation Program for Graduates of Jiangsu Province (CXZZ13_0583), Jiangsu Provincial Medical Youth Talent (QNRC2016073), Key Program of Science and Technology Foundation from Nanjing Medical University (2016NJMUZD042), and Natural Science Foundation of Jiangsu Province (BK20170135) to YZG.

Disclosure
The authors report no conflicts of interests in this work.

References