Abstract. To explore the complex molecular mechanisms of bladder cancer, mRNA and miRNA expression profiles were combined for systematic analyses. A total of 18 common differentially expressed genes (DEGs) were identified from two mRNA expression datasets which consisted of 206 tumor and 74 normal tissues. Then, survival analysis based on the SurvExpress database showed that the common DEGs were able to significantly differentiate low- and high-risk groups in 4 public bladder cancer datasets (p<0.01). Notably, the tumor and normal samples were able to be almost clearly classified into 4 groups based on these identified common DEGs. In addition, 6 out of the 18 common DEGs, including ALDH1A1 and SRPX, are regulated by 6 reported miRNAs based on regulatory network analyses. Expression levels of the 6 DEGs were validated in 10 bladder cancer samples using RT-PCR, and the expression values were concordant with the microarray results. Collectively, our analyses indicated that various biological processes are involved in the development and progression of bladder cancer. Firstly, cell cycle checkpoints and DNA repair networks of cancer stem-like cells were regulated by high expression of ALDH1A1, and hence promoted tumor self-renewal or metastasis. Then, activation of HspB6 induced the angiogenesis process which provides necessary nutrition and oxygen for tumor cells. Moreover, downregulation of the expression of tumor-suppressor genes SRPX and FLNC further promoted apoptosis and metastasis. The identification of potential biological processes and genes can be helpful for the understanding of bladder cancer molecular mechanisms.

Introduction

Bladder cancer, as one of the most common types of cancer of the urinary tract, accounts for ~380,000 new cases and ~150,000 deaths every year, worldwide (1). Statistics show that the incidence of bladder cancer experience an increase of 19.4% adjusted for the increase in total world population from 1990 to 2010 (2). In the US, the 5-year survival rate is ~77% (3). Epidemiological research has shown that age is one of the most significant risk factor, and the average age at diagnosis of bladder cancer is ~70-year old (4). Apart from age, smoking is considered as the leading environmental risk factor, and more than half of bladder cancer cases can be attributed to smoking (4).

Considering the complexity of bladder cancer molecular mechanisms, extensive research has been carried out using microarray or next-generation sequencing technologies. In addition, several mRNAs and miRNAs have been identified to be differentially expressed based on RNA-seq data from 129 tumors. One of them is FGFR3 which is upregulated in the papillary-like bladder cancer cluster. Moreover, miR-99a and miR-100 that downregulate FGFR3 were found to be downregulated in bladder cancer (5). In addition, 32 significant somatic mutations including TP53, RB1, FGFR3, were identified in 130 paired tumor and normal samples using whole-exome sequencing (5). Researches also showed that several pathways were frequently dysregulated in bladder cancer such as cell cycle regulation, kinase and phosphatidylinositol-3-OH kinase and chromatin remodeling (5).

Although research has been carried to explore bladder cancer molecular mechanisms, research based on the integration of mRNA and miRNA expression profiles has not been widely explored. With the development of microarray and sequencing technology, more and more datasets have been submitted to the Gene Expression Omnibus (GEO) database, and hence re-analyses of the deposited datasets with advanced
bioinformatic methods can further promote the understanding of bladder cancer molecular mechanisms (6). In the present study, DEGs in bladder cancer were firstly identified based on two mRNA expression datasets from different laboratories. Then, the common DEGs were subjected to function and pathway analyses. Furthermore, mRNA and miRNA regulatory networks were constructed. Finally, several critical DEGs were validated using SurvExpress database and RT-PCR method.

Materials and methods

Acquisition of microarray data. Public available datasets were used in the present study. Gene expression profiles GSE13507 and GSE37815 were downloaded from the GEO database (http://www.ncbi.nlm.nih.gov/geo/). These two datasets were submitted by Kim et al in 2008 (7) and Kim et al in 2012 (8), respectively. Dataset GSE13507 consists of 188 tumor and 68 normal tissues, and dataset GSE37815 consists of 24 expression profiles including 6 normal and 18 tumor tissues. RNAs were extracted and hybridized to Illumina human-6 v2.0 expression beadchip according to the manufacturer's instructions. Detailed information concerning the experiment design and samples were documented in previous studies (8,9).

Identification of differentially expressed mRNAs. To identify DEGs, raw data were systematically analyzed using in-house R script. Firstly, mRNA expression values were subjected to log2 transformation, background correction and normalization using GeneChip Robust Multi-Array Analysis (GC-RMA) (10) algorithm with default parameters. Then, uninformative control probe sets were filtered out, and the average expression value was calculated for genes with multiple probes. Finally, DEGs were identified using Linear Models for Microarray Analysis (Limma) package (11) with criteria of adjusted p≤0.01 and |log2 fold-change (FC)| ≥2. Common DEGs between the two datasets were presented using Venn diagram. Heat map of DEGs was constructed using heatmap.2 method within ggplot package.

Function and pathway enrichment analysis. To illuminate the biology functions related to the identified DEGs, Gene Ontology (GO) and KEGG pathway enrichment for the construction of miRNA-mRNA pairings contributing to bladder cancer. Combination of the 6 miRNAs has been demonstrated to be the best predictor to distinguish patients with urothelial bladder cancer from normal controls (14). miRNA-mRNA regulatory network was constructed using CyTargetLinker plugin (15) of Cytoscape (16). The prediction of miRNA targets were based on microcosm, miRTarBase and TargetScan databases. Common target genes from the 3 databases were selected, and the interaction between the 6 miRNAs and the common DEGs were constructed.

Virtual validation and RT-PCR validation. Clinical outcome validation for the common DEGs were carried out using online tool SurvExpress, which is based on a cancer-wide gene expression database with clinical outcomes (17). Four bladder cancer datasets in this database were used for virtual validation. Parameter settings were carefully selected according to the manufacturer's instructions.

Six DEGs including ALDH1A1, FLNC, CNN1, SRPX, HSPB6 and FAM107A were selected for further RT-PCR validation based on the criterion that DEGs can be regulated by miRNAs and were identified in the SurvExpress analysis. Total RNAs were extracted from 10 bladder cancer and adjacent normal tissues with TRIzol reagent from ThermoFisher (Waltham, MA, USA). Extraction was based on the standard protocol and manufacturer's instructions. Then, cDNA were obtained with M-MLV Reverse Transcriptase from Promega (Madison, WI, USA). mRNA expression values were detected using 7500 Real-Time PCR System (ThermoFisher). The internal GAPDH mRNA expression was used for normalization and relative quantification was calculated using the 2^ΔΔCt method. Primer sequences for the 6 genes were as follows: HSPB6, 5'-TTGCTGTCAAGTTGTTGGGC-3' (forward) and 5'-CGGTAGCGGTTGGAACT-3' (reverse); SRPX, 5'-CCCACAGCCCGAAACCT-3' (forward) and 5'-TGCTCTATCCTGCCAATG-3' (reverse); CNN1, 5'-ACCCTCTGTGGCTTGGC-3' (forward) and 5'-AATGATGTGCGCCCTTCT-3' (reverse); FLNC, 5'-AGAGAAATGGAGAGAGAATACGTC-3' (forward) and 5'-ACCRCTATTATATCATACATAC-3' (reverse); ALDH1A1, 5'-CTTGGAGATTTCCCCTTG-3' (forward) and 5'-TTGCTTCTGCTTTTGACA-3' (reverse); FAM107A, 5'-AGCACAACGCCTCTGTGACT-3' (forward) and 5'-TGCGGGCTTATGCTA-3' (reverse).

Results

DEGs in bladder cancer. Background correction and normalization were applied to the two datasets, and the medians of the gene expression values were almost at the same level indicating that the data were appropriate for subsequent analysis (data not shown). After independent DEG analysis, a total of 21 and 72 DEGs were screened out for GSE13507 and GSE37815, respectively. In addition, 18 DEGs (Table I) were identified to be differentially expressed in the two datasets (Fig. 1A). Among these common DEGs, 8 DEGs were upregulated and 10 DEGs were downregulated. In addition, there was a high correlation ($R^2=0.97$) for the common DEG expression values between GSE13507 and GSE37815 (Fig. 1B).

Furthermore, hierarchical clustering of the tumor and normal tissues was carried out based on the common DEGs. As indicated in Fig. 2, the tumor and normal tissues were able to be almost clearly classified into different clusters. Error assignment of several samples may have been caused by tumor heterogeneity or smaller expression value variation.

GO and KEGG pathway enrichment analysis. To explore the biology functions of the common DEGs, GO and KEGG
pathway enrichment were conducted using DAVID online tool. Results showed that 3 KEGG pathways were identified. *ALDH1A1* participates in retinol metabolism pathway, *DCN* is involved in the TGF-β signaling pathway, and *FLNC* plays a role in the MAPK signaling pathway. In addition, the common DEGs were mainly related to the cellular component of myofibril (p=1.0E-04), contractile fiber part (p=1.0E-04) and Z-disc (p=1.2E-04) (Table II). Several common DEGs, such as *CRYAB*, *PDLIM3*, *SYNM*, *CNN1*, *COL16A1* and *FLNC*, were significantly enriched in the cytoskeletal protein binding function (p=2.4E-04), structural molecule activity function (p=0.007), and structural constituent of muscle function (p=0.02) (Table II). However, the biological processes of cytoskeleton organization (p=0.004) and response to heat (p=0.04) were also enriched (Table II).

**Table I.** The identified 18 common DEGs in GSE37817 and GSE13507.

<table>
<thead>
<tr>
<th>Gene</th>
<th>GSE37817</th>
<th>GSE13507</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P-value</td>
<td>FC</td>
</tr>
<tr>
<td>ALDH1A1</td>
<td>0.00122</td>
<td>3.57</td>
</tr>
<tr>
<td>C2orf40</td>
<td>0.002817</td>
<td>2.36</td>
</tr>
<tr>
<td>CNN1</td>
<td>0.00986</td>
<td>3.36</td>
</tr>
<tr>
<td>COL16A1</td>
<td>0.004753</td>
<td>-2.5</td>
</tr>
<tr>
<td>CPED1</td>
<td>0.001365</td>
<td>-2.5</td>
</tr>
<tr>
<td>CRYAB</td>
<td>0.004264</td>
<td>3.31</td>
</tr>
<tr>
<td>DCN</td>
<td>0.004807</td>
<td>2.58</td>
</tr>
<tr>
<td>FAM107A</td>
<td>0.000146</td>
<td>-3.17</td>
</tr>
<tr>
<td>FLNC</td>
<td>0.009383</td>
<td>-3.69</td>
</tr>
<tr>
<td>HSPB6</td>
<td>0.002658</td>
<td>-2.96</td>
</tr>
<tr>
<td>PCP4</td>
<td>0.000849</td>
<td>-4.01</td>
</tr>
<tr>
<td>PDLIM3</td>
<td>0.007189</td>
<td>2.12</td>
</tr>
<tr>
<td>PLAC9</td>
<td>0.000661</td>
<td>-2.81</td>
</tr>
<tr>
<td>PRUNE2</td>
<td>0.001365</td>
<td>3.44</td>
</tr>
<tr>
<td>SMOC2</td>
<td>0.007966</td>
<td>-2.39</td>
</tr>
<tr>
<td>SRPX</td>
<td>0.000761</td>
<td>-3.52</td>
</tr>
<tr>
<td>SYNM</td>
<td>0.002099</td>
<td>3.11</td>
</tr>
<tr>
<td>TCEAL2</td>
<td>0.000251</td>
<td>-3.06</td>
</tr>
</tbody>
</table>

DEGs, differentially expressed genes. FC, fold change.

**Table II.** GO enrichment analysis results for the common DEGs

<table>
<thead>
<tr>
<th>Category</th>
<th>ID</th>
<th>GO term</th>
<th>P-value</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP</td>
<td>GO:0007010</td>
<td>Cytoskeleton organization</td>
<td>0.00</td>
<td><em>CRYAB</em>, <em>PDLIM3</em>, <em>SYNM</em>, <em>CNN1</em></td>
</tr>
<tr>
<td></td>
<td>GO:0009408</td>
<td>Response to heat</td>
<td>0.04</td>
<td><em>HSPB6</em>, <em>CRYAB</em></td>
</tr>
<tr>
<td>CC</td>
<td>GO:0030016</td>
<td>Myofibril</td>
<td>0.00</td>
<td><em>CRYAB</em>, <em>PDLIM3</em>, <em>SYNM</em>, <em>FLNC</em></td>
</tr>
<tr>
<td></td>
<td>GO:0044449</td>
<td>Contractile fiber part</td>
<td>0.00</td>
<td><em>CRYAB</em>, <em>PDLIM3</em>, <em>SYNM</em>, <em>FLNC</em></td>
</tr>
<tr>
<td></td>
<td>GO:0030018</td>
<td>Z-disc</td>
<td>0.00</td>
<td><em>CRYAB</em>, <em>PDLIM3</em>, <em>FLNC</em></td>
</tr>
<tr>
<td>MF</td>
<td>GO:0008092</td>
<td>Cytoskeletal protein binding</td>
<td>0.00</td>
<td><em>CRYAB</em>, <em>PDLIM3</em>, <em>SYNM</em>, <em>CNN1</em>, <em>FLNC</em></td>
</tr>
<tr>
<td></td>
<td>GO:0005198</td>
<td>Structural molecule activity</td>
<td>0.01</td>
<td><em>CRYAB</em>, <em>PDLIM3</em>, <em>SYNM</em>, <em>COL16A1</em></td>
</tr>
<tr>
<td></td>
<td>GO:0008307</td>
<td>Structural constituent of muscle</td>
<td>0.03</td>
<td><em>PDLIM3</em>, <em>SYNM</em></td>
</tr>
</tbody>
</table>

GO, Gene Ontology; DEGs, differentially expressed genes; BP, biological process; CC, cellular component; MF, molecular function.

**miRNA-mRNA regulatory network.** To generate a catalog of miRNA-mRNA pairings, the predicted target mRNAs of 6 reported miRNAs were intersected with the defined common DEGs. Predication results showed that 4,279, 1,818 and 3,086 target mRNAs, in microcosm, miRTarBase and TargetScan database, respectively, were able to be regulated by the 6 miRNAs. Among the target mRNAs, 93 common mRNAs were identified after intersection between the
3 databases. In addition, 8 miRNA-mRNA pairings were identified with the criteria that the prediction of target mRNAs were also differentially expressed in the microarray analysis. The regulatory network for the 8 miRNA-mRNA pairings are displayed in Fig. 3. Based on Fig. 3, it is evident that miRNA hsa-miR-34a-5p regulates FAM107A and HSPB6, respectively. ALDH1A1 and CNN1 are regulated by hsa-miR-221-3p, hsa-miR-21-5p and hsa-miR-16-5p.

Virtual validation and RT-PCR validation. It is critical for the validation of multi-gene biomarkers in the study of the molecular mechanisms of cancer. To verify the prognostic performance of the identified common DEGs, we firstly employed the SurvExpress online tool which provides survival analysis and risk assessment. Results are shown in Fig. 4 and are summarized in Table III. Survival analysis using Kaplan-Meier curves indicated that the common DEGs were able to significantly differentiate low- and high-risk groups in the
4 datasets, and the p-values were 0.0028, 0.00056, 0.0011 and 0.021, respectively. A higher concordance index (CI) value (Table III; CI >50) also demonstrated that better prediction was achieved between low- and high-risk groups.

Furthermore, RT-PCR validation was carried out using 10 specimens from patients to validate the microarray analysis results. Based on the criterion that the DEGs can be regulated by miRNAs and identified in the SurvExpress analysis, a total of 6 DEGs were selected for RT-PCR validation including HSPB6, SRPX, CNN1, FLNC, ALDH1A1 and FAM107A. The mRNA expression values for those selected DEGs were measured in tumor and adjacent normal tissues. The results indicated that the expression levels of HSPB6, FLNC and SRPX were slightly lower in the tumor tissues when compared with these levels in the adjacent normal tissues (Fig. 5); whereas, mRNA expression levels for the remaining 3 DEGs were significant higher in the tumor tissues (Fig. 5). All these results were nearly concordant with the results of the microarray analysis.

Discussion

Microarray and next generation sequencing technologies have significantly advanced the understanding of the molecular
mechanisms of bladder cancer. Extensive research concerning mRNA or miRNA expression, and whole-genome or exome sequencing, has been widely carried out in the past few years. However, efforts that attempt to unveil the complex mechanisms of bladder cancer using the integration of different omic data are rare. In the present study, a total of 18 genes were found to be simultaneously differentially expressed in 206 tumor tissues, and 10 were upregulated and 8 were downregulated. Notably, the tumor and normal samples were able to be clearly classified into 3 groups based on these common DEGs. Several normal samples were assigned into the tumor cluster incorrectly, which was probably caused by tumor heterogeneity or sample quality.

Pathway enrichment analysis revealed that the common DEGs were involved in the TGF-β signaling pathway. Although the molecular mechanism between bladder cancer and the TGF-β signaling pathway is unclear, TGF-β1 production has been demonstrated to be significantly associated with the phenotype of bladder cancer (18). RT-PCR results showed that BMP2 and INHBB within the TGF-β signaling pathway were significantly associated with tumorigenicity (p=0.02) and invasiveness (p=0.04) (19). Another important MAPK pathway was also enriched based on the common DEGs. Jebar et al found that mutually exclusive mutation of FGFR3 or RAS (HRAS or KRAS) can activate the MAPK pathway, and the exact role of FGFR3 in activation of the MAPK pathway warrants further studies (20).

Further analyses based on the mRNA and miRNA regulatory network partially unveiled the complex mechanisms of bladder cancer. The development and progression of bladder cancer probably involve multi-path processes including cancer stem cell processes, downregulation of the expression of tumor-suppressor genes, and promotion of cancer cell migration and angiogenesis (Fig. 6).

Recently, extensive research has focused on the study of cancer stem cells, which are believed to contribute to tumor self-renewal and metastasis (21). In the present study, it was found that putative cancer stem cell biomarker ALDH1A1 probably contributes to the poor prognosis of bladder cancer. In addition, based on the mRNA regulation network, the upregulation of ALDH1A1 is likely to be the consequence of the downregulation of the expression of miR-221-3p and miR-21-5p. ALDH1A1 combined with CD44 has been identified as a promising cancer stem cell biomarker in various types of cancers (22-24). Keymoosi et al demonstrated that ALDH1A1 was highly expressed in almost 16% (25/159) of bladder cancer cases based on immunohistochemistry. In addition, high expression of ALDH1A1 was found to be significantly correlated with clinical characteristics such as tumor size (p=0.002), pathologic stage (T1, p=0.007 and T2, p-value<0.001) and high recurrence rate (p=0.013) (22).

In ovarian cancer, knockdown of ALDH1A1 lead to S and G2 phase cell accumulation via marked decrease in p21 (25). In addition, DNA damage was also increased after ALDH1A1 knockdown evidenced by induction of γ-H2AX and BAX-mediated apoptosis (25). All of these findings indicate that ALDH1A1 participates in the regulation of cell cycle checkpoints and DNA repair networks in cancer stem-like cells (25).

Moreover, downregulation of the expression of tumor-suppressor genes also plays an important role in tumor development and metastasis. In the present study, downregulation of the expression of SRPX and FLNC was identified, which probably was inhibited by miR-200c-5p or miR-16-5p and miR-205-5p, respectively. SRPX was firstly isolated as a novel transformation suppressor gene, and SRPX expression is downregulated by retroviral oncogenes such as v-src or v-ras (26). Research has demonstrated that SRPX expression is markedly reduced in various human cancer cell lines (26,27). Tambe et al documented that the C-terminal region and the 3 consensus repeats in the N-terminal region of SRPX are critical in SRPX-induced apoptosis. Low expression of SRPX can sequentially activate caspase-12, -9 and -3 rather...
Moreover, the downregulation of the expression of endogenous FLNC participates in the degradation of extracellular matrix and cancer cell proliferation and metastasis is angiogenesis which can provide adequate oxygen and nutrient supplies (38). In the present study, angiogenesis was activated by the high expression of HspB6. In addition, based on the miRNA regulation network, the upregulation of HspB6 was likely to be the consequence of miR-34a-5p. HspB6, also referred to as hsp20, has been reported to be activated in physiological or pathological stress (39,40) and non-small cell lung cancer (41). Wang et al showed that HspB6 can promote growth factor secretion including VEGF and bFGF and induce myocardial angiogenesis (42). The angiogenesis process may be promoted via activation of VEGFR2 by HspB6 based on protein binding assay and immunostaining results (43). However, the complex regulatory mechanism of HspB6 in bladder cancer remains to be explored.

In summary, the development and progression of bladder cancer are induced via various processes. Firstly, the cell cycle checkpoints and DNA repair networks in cancer stem-like cell are regulated by high expression of ALDH1A1, and hence promote tumor self-renewal or metastasis. Then, activation of HspB6 promotes angiogenesis which can provide necessary nutrients and oxygen for tumor cells. Finally, downregulation of the expression of SRPX and FLNC further promote apoptosis and metastasis.

Acknowledgements

The present study was supported by grants from the National Natural Science Foundation of China (nos. 81571395, 81371748 and 81373075).

References


