Ruguo key genes and tumor driving factors identification of bladder cancer based on the RNA-seq profile

Minglei Zhang¹
Hongyan Li²
Di Zou³
Ji Gao²

¹Department of Orthopedics, Division of Tumor and Trauma Surgery, ¹Department of Urology, China–Japan Union Hospital of Jilin University, ¹Department of Nephrology, The First Affiliated Hospital of Changchun University of Chinese Medicine, Changchun, People’s Republic of China

Aim: This study aimed to select several signature genes associated with bladder cancer, thus to investigate the possible mechanism in bladder cancer.

Methods: The mRNA expression profile data of GSE31614, including ten bladder tissues and ten control samples, was downloaded from the Gene Expression Omnibus. The differentially expressed genes (DEGs) in bladder cancer samples compared with the control samples were screened using the Student’s t-test method. Functional analysis for the DEGs was analyzed using the Database for Annotation, Visualization, and Integrated Discovery from the Gene Ontology database, followed by the transcription function annotation of DEGs from Tumor-Associated Gene database. Motifs of genes that had transcription functions in promotor region were analyzed using the Seqpos.

Results: A total of 1,571 upregulated and 1,507 downregulated DEGs in the bladder cancer samples were screened. ELF3 and MYBL2 involved in cell cycle and DNA replication were tumor suppressors. MEG3, APEX1, and EZH2 were related with the cell epigenetic regulation in bladder cancer. Moreover, HOXB9 and EN1 that have their own motif were the transcription factors.

Conclusion: Our study has identified several key genes involved in bladder cancer. ELF3 and MYBL2 are tumor suppressors, HOXB9 and EN1 are the main regulators, while MEG3, APEX1, and EZH2 are driving factors for bladder cancer progression.

Keywords: bladder cancer, differentially expressed genes, tumor driving factor, function analysis

Introduction
Bladder cancer is one of the most common malignancies in epithelial and mesenchymal tissues of human’s urinary system and contributes to the increasing morbidity and mortality.¹ Statistics data refer that approximately 74,000 new cases were diagnosed as bladder cancer in 2014.² Treatment methods such as surgery, chemotherapy, and drug use play significant role in improving quality of life of patients with bladder cancer.³,⁴ However, prognosis of bladder cancer is unsatisfactory, with the poor 5-year survival rate due to difficulty in diagnosing and easy metastasis.⁴ Therefore, it will be of great significance to explore several signature genes and biomarkers for the target treatment and diagnosis of bladder cancer.

Recently, research on the signature genes and transcription factors (TFs), which are closely associated with bladder cancer, is becoming the hot spot. For instance, Dyrsjkot et al¹ investigated that 12 gene signatures were associated with nonmuscle invasive bladder cancer prognosis, such as UBE2C, COL18A1, SKAP2, and NEK1. According to Riester et al,⁵ several tumor metastasis, tumor stage, and progression genes such as FN1,
NNMT, POSTN, and SMAX6 are the prognostic signatures for high-risk bladder cancer. Dancik et al identified that genetic alterations of TP53, HRAS, KEM64, and FGFR3 were the diagnostic signature genes that will drive the muscle invasive bladder cancer. Besides, increasing evidences suggest that motifs with the conserved short sequences in the upstream regions of gene promoters play crucial roles in gene transcription process in many cancers, such as bladder cancer, prostate cancer, and colorectal cancer. Rachakonda et al reported that mutation of the telomerase reverse transcriptase (TERT) promoter was a potential prognostic target for bladder cancer in clinical application. Despite many studies on the exploration of the prognostic biomarkers in bladder cancer, the mechanism for poor diagnosis and prognoses of bladder cancer still remain largely unknown.

In this study, bioinformatics approaches were used to screen the differentially expressed genes (DEGs) and driving factors that may be helpful for the bladder cancer treatment and prognosis based on the RNA sequencing profile data. The significant functions of the DEGs and the motifs of crucial genes in bladder cancer were analyzed to predict the key genes and driving factors for bladder cancer. This study aimed to explore several signature genes for the target treatment of bladder cancer and to investigate the mechanism of bladder cancer. Our study may provide theoretical basis for the diagnostic and therapeutic exploration for bladder cancer treatment.

Methods
Data resources and preprocessing
The mRNA expression profile data of GSE31614 was downloaded from the Gene Expression Profile in National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/geo) based on the platform of GPL9115 Illumina Genome Analyzer II (Homo sapiens). The study contains the mRNA expression of cancer and matched adjacent tissues of seven testicular germ cell tumors and ten transitional cell carcinomas of bladder. In this study, the ten bladder cancer samples matched with ten matched normal samples were chosen for subsequent analysis.

Reads alignment and transcript expression analysis
The reads of genes sequences in the genome of patients with bladder cancer were aligned with the hg19 genome sequence deposited in the University of California Santa Cruz Genome Browser Database using the Bowtie software. Read with only one alignment result and with the base mismatch no more than two was selected. In addition, the expression values of genes in each transcript in each sample were calculated using the Cufflinks tool, based on the fragments per kilobase per million mapped fragments method.

Screening of DEGs
The DEGs in bladder cancer samples compared with matched normal samples were identified using the Student’s t-test method in R. P<0.05 and |log 2 fold change| ≥2 were chosen as the threshold.

Function annotation of DEGs
The gene ontology functions of the selected DEGs, including biological process (BP) and cellular component, were analyzed using the Database for Annotation, Visualization, and Integrated Discovery software. The false discovery rate <0.005 was chosen as the cutoff criterion. Besides, the genes that have the transcription regulatory functions were analyzed and selected based on the information of transcript database. Finally, genes that functions as oncogenes or tumor suppressors were selected based on the information of Tumor-Associated Gene database (http://sourceforge.jp/projects/sfnet_tdt).

Prediction of the upstream regulatory elements of DEGs
The region in one gene from upstream 1 kb to the downstream 0.5 kb of the transcription start site was considered as the promoter region in this study. The motif of the selected DEGs in the promoter region was selected using the Seqpos. P<0.00001 and z score <-7 were chosen as the threshold.

Results
DEGs screening
With P<0.05 and |log 2 fold change| ≥2 as the threshold, a total of 1,571 upregulated and 1,507 downregulated DEGs in the bladder cancer samples were compared with matched normal samples using Student’s t-test method. Besides, 57 upregulated DEGs and 79 downregulated DEGs were analyzed as the TFs based on the transcript database analysis. Three upregulated genes such as E74-like factor 3 (ELF3), v-myb avian myeloblastosis viral oncogene homologue-like 2 (MYBL2), and MYCN were the known oncogenes, while three downregulated genes, including EHF, FNACB, and PML were the known tumor suppressors.

Functional analysis for the DEGs
The selected downregulated DEGs significantly participated in the BP terms such as cell adhesion, extracellular matrix organization, response to endogenous stimulus, and chemical homeostasis (Table 1A). Also, the upregulated genes were mainly involved in BP terms related to cell cycle, such as
mitotic cell cycle, cell division, DNA replication, and cell cycle checkpoint (Table 1B).

In addition, the heat maps of DEGs showed that there were 35 upregulated genes and 138 downregulated genes in tumor tissues that were associated with the cell surface and extracellular matrix (Figure 1). Genes, such as ITGA6, LAMB3, MMP1, TGFA, VEGFA, WNT10A, and WNT7A were all involved in the cancer-related pathways, suggesting these genes might be involved in the bladder cancer progression.

Finally, we selected 12 genes that were associated with the cell epigenetic regulation in bladder cancer tissues (Figure 2). Genes such as maternally expressed 3 (MEG3), PLD6, SMCHD1, and ZFP36 were downregulated, while APEX nuclease multifunctional DNA repair enzyme 1 (APEX1), APOBEC3C, APOBEC3F, ATF7IP, CHEK1, CNOT6, enhancer of zeste homolog 2 Drosophila (EZH2), and GATAD2A were upregulated in the cancer tissues.

### Table 1 The significantly enriched GO terms by the differentially expressed genes in bladder cancer

<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
<th>P-value</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A: downregulated</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0007155 – cell adhesion</td>
<td>128</td>
<td>6.08E-20</td>
<td>1.11E-16</td>
</tr>
<tr>
<td>GO:00030198 – extracellular matrix organization</td>
<td>32</td>
<td>4.15E-11</td>
<td>7.55E-08</td>
</tr>
<tr>
<td>GO:0001501 – skeletal system development</td>
<td>61</td>
<td>1.62E-10</td>
<td>2.94E-07</td>
</tr>
<tr>
<td>GO:0009719 – response to endogenous stimulus</td>
<td>68</td>
<td>3.96E-09</td>
<td>7.21E-06</td>
</tr>
<tr>
<td>GO:0009725 – response to hormone stimulus</td>
<td>62</td>
<td>1.67E-08</td>
<td>3.04E-05</td>
</tr>
<tr>
<td>GO:0019226 – transmission of nerve impulse</td>
<td>60</td>
<td>1.70E-08</td>
<td>3.10E-05</td>
</tr>
<tr>
<td>GO:0007517 – muscle organ development</td>
<td>43</td>
<td>1.71E-08</td>
<td>3.11E-05</td>
</tr>
<tr>
<td>GO:0007167 – enzyme linked receptor protein signaling pathway</td>
<td>58</td>
<td>4.49E-08</td>
<td>8.17E-05</td>
</tr>
<tr>
<td>GO:0006873 – cellular ion homeostasis</td>
<td>61</td>
<td>8.13E-08</td>
<td>1.48E-04</td>
</tr>
<tr>
<td>GO:0035295 – tube development</td>
<td>42</td>
<td>1.71E-07</td>
<td>3.11E-04</td>
</tr>
<tr>
<td>GO:0030182 – neuron differentiation</td>
<td>66</td>
<td>4.55E-07</td>
<td>8.28E-04</td>
</tr>
<tr>
<td>GO:0042127 – regulation of cell proliferation</td>
<td>101</td>
<td>9.09E-07</td>
<td>1.66E-03</td>
</tr>
<tr>
<td>GO:0001944 – vasculature development</td>
<td>44</td>
<td>9.58E-07</td>
<td>1.74E-03</td>
</tr>
<tr>
<td>GO:0007242 – intracellular signaling cascade</td>
<td>145</td>
<td>1.62E-06</td>
<td>2.94E-03</td>
</tr>
<tr>
<td>GO:0048878 – chemical homeostasis</td>
<td>72</td>
<td>1.76E-06</td>
<td>3.21E-03</td>
</tr>
<tr>
<td><strong>B: upregulated</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0007049 – cell cycle</td>
<td>123</td>
<td>1.54E-14</td>
<td>2.77E-11</td>
</tr>
<tr>
<td>GO:0000278 – mitotic cell cycle</td>
<td>75</td>
<td>2.06E-14</td>
<td>3.74E-11</td>
</tr>
<tr>
<td>GO:0000887 – M phase of mitotic cell cycle</td>
<td>54</td>
<td>1.13E-13</td>
<td>2.05E-10</td>
</tr>
<tr>
<td>GO:0051301 – cell division</td>
<td>59</td>
<td>3.02E-11</td>
<td>5.47E-08</td>
</tr>
<tr>
<td>GO:0006259 – DNA metabolic process</td>
<td>77</td>
<td>1.51E-08</td>
<td>2.73E-05</td>
</tr>
<tr>
<td>GO:0006260 – DNA replication</td>
<td>40</td>
<td>1.56E-08</td>
<td>2.82E-03</td>
</tr>
<tr>
<td>GO:0000075 – cell cycle checkpoint</td>
<td>23</td>
<td>1.22E-06</td>
<td>2.21E-03</td>
</tr>
<tr>
<td>GO:0051726 – regulation of cell cycle</td>
<td>52</td>
<td>1.72E-06</td>
<td>3.11E-03</td>
</tr>
<tr>
<td>GO:0007059 – chromosome segregation</td>
<td>21</td>
<td>2.60E-06</td>
<td>4.71E-03</td>
</tr>
</tbody>
</table>

**Abbreviations:** FDR, false discovery rate; GO, gene ontology; DNA, deoxyribonucleic acid.

We selected the DNA motif for upregulated genes, such as CAATAAAA for homeobox B9 (HOXB9), TAAATTA for engrailed homeobox 1 (EN1), and AGGAAGGGAAAGGA for PRDM1 (Figure 3A), while the motif for downregulated genes were TTTCT for NR3C1, TGTTCT for E2F6, and ACATGCAGCATG for TP53 (Figure 3B), suggesting their potential tumor driving roles in bladder cancer (Figure 4).

### Discussion

Bladder cancer is one of the most common malignancies of the human urinary system, with increasing morbidity and mortality due to difficulty in diagnosing and easy metastasis. To explore several signature genes and biomarkers for the target treatment, diagnosis of bladder cancer will be of great significance. In this study, we analyzed the mRNA expression profile of bladder cancer to screen the signature genes and key driving factors for bladder cancer diagnosis and target treatment based on the RNA sequencing. The upregulated ELF3 and MYBL2 involved in cell cycle and DNA replication were the tumor suppressors. Downregulated MEG3 and upregulated APEX1 and EZH2 were involved in the cell epigenetic regulation in bladder cancer. Moreover, HOXB9 and EN1 that have their own motifs were the TFs in bladder cancer samples.
ELF3 is a member of the E26 transformation-specific (ETS) family that can bind and transactivate ETS sequences containing the consensus nucleotide core sequence GGA. Seth and Watson proved that ETS TFs could activate or repress the expression of genes that are involved in the cell proliferation, differentiation, and development and apoptosis of tumor cells in many cancers. Overexpression of ELF3 induces the endogenous transforming growth factor beta (TGF-β) type II receptor expression in human breast cancer cells, and the TGF-β type II receptor is necessary for mediating the effects of TGF-β on tumor cell growth inhibition. Thus, ELF3 may be associated with tumor cell growth. Also, ELF3 has been reported as the candidate transcription regulator involved in human urinary cydiodifferentiation. In this study, the upregulated ELF3 was selected as the tumor suppressor gene, indicating the inhibitory roles in bladder cancer. Meanwhile, MYBL2 is a member of the MYB family, which is a nuclear protein involved in cell cycle progression. Sala proved that MYBL2 regulated the cell cycle of cancer cells to affect the tumorigenesis. Also, Wu et al reported that genetic variation in cell cycle control genes were related to the...
increased risk of bladder cancer. Our data showed that the upregulated MYBL2 was involved in the cell cycle function in bladder cancer samples, suggesting that MYBL2 might be the tumor suppressor gene for bladder cancer risk through cell cycle.

Our findings showed that the upregulated HOXB9 and EN1 that have their own motifs were the TFs in bladder cancer samples. HOXB9 is a Abd-B homeobox family protein with a homeobox DNA-binding domain on chromosome 17, while EN1 has a role in controlling development during the central nervous system. However, it has been reported that overexpression of HOXB9 promoted the cell proliferation and angiogenesis that was related to breast cancer. Also, high level of HOXB9 in breast cancer induces the expressions of several angiogenic factors such as interleukin-8, vascular endothelial growth factor, and ErbB to activate their respective pathways, leading to the lung cancer metastasis. Besides, Bell et al demonstrated that significant hypermethylation of EN1 at the transcriptional start site was observed, suggesting it may be a novel biomarker for poor prognosis of human salivary gland adenoid cystic carcinoma. Therefore, HOXB9 may contribute to tumorigenesis and EN1 may be related to cancer diagnosis. On the other hand, E2F1 is the candidate gene that binds to the promoter region of HOXB9 to induce its overexpression and then accelerated breast cancer progression. EN1 has the Pax5 binding site on the promoter region and Pax2 could regulate the enhancer of Pax5 at the midbrain–hindbrain boundary. Also, inhibited Pax2 results in cell death of prostate cancer. Based on our results, we speculate that HOXB9 with the conserved motif CAATAAAA may promote bladder cancer development, while EN1 with DNA sequence TAATTA might be related to bladder cancer diagnosis.
Meanwhile, our results displayed that the downregulated MEG3 and upregulated APEX1 and EZH2 were the genes involved in the cell epigenetic regulation in bladder cancer samples compared with the controls, suggesting their important roles in bladder cancer. Ying et al\(^{34}\) proved that downregulation of MEG3 activated autophagy and increased cell proliferation of bladder cancer by affecting the chromosome. Role of APEX1 and EZH2 in bladder cancer have not been fully reported in previous researches. However, Kim et al\(^{35}\) proved that APEX1 was the driving factor for colon cancer progression, while Varambally et al\(^{36}\) referred that EZH2 was involved in prostate cancer progression. Based on our study, we speculate that MEG3, APEX1, and EZH2 may be the tumor driving factors for bladder cancer progression.

**Conclusion**

Our findings indicate that ELF3, MYBL2, HOXB9, EN1, MEG3, APEX1, and EZH2 are identified as key genes involved in bladder cancer progression or development. ELF3 and MYBL2 are the key genes that play crucial roles in suppressing bladder cancer, while MEG3, APEX1, and EZH2 are the driving factors for bladder cancer progression. Also, HOXB9 and EN1 are the TFs that play key roles in bladder cancer development and progression. Our study may provide theoretical basis for the future bladder cancer investigation. However, there were still some limitations in the current study. First, the sample size enrolled in our study was small. Second, in vivo and in vitro experimental verification were not performed to validate the functions of crucial genes in bladder cancer samples. Further studies based on a larger sample size and experiments, such as expression validation or knockdown assay in bladder cancer cell lines, are still needed to confirm our results.

**Disclosure**

Hongyan Li and Di Zou are regarded as co-second authors. The authors report no conflicts of interest in this work.
References


