Exploration of the molecular mechanism of prostate cancer based on mRNA and miRNA expression profiles

Abstract: Prostate cancer, the second most common cancer in men, has been rarely explored by integrating mRNA and miRNA expression profiles. In this study, we combined two mRNA expression datasets and six documented miRNAs to uncover the comprehensive molecular mechanism of prostate cancer. Results showed that a total of 30 genes were significantly differentially expressed in 49 tumor samples by comparing with 22 normal samples. Importantly, all samples from the two datasets can be clearly classified into two groups, tumor group and normal group, based on the selected differentially expressed genes (DEGs). The miRNA–mRNA regulation network indicated that 4 out of 30 DEGs can be regulated by three miRNAs. In addition, prognostic performance validation using in silico databases showed that the DEGs can significantly differentiate between low-risk and high-risk prostate cancer. In summary, multiple biological processes are probably involved in the development and progression of prostate cancer. First, dysregulation of SV2 can regulate transporter and transmembrane transporter activity and then provide the necessary nutrients for tumor cell proliferation. Second, HOXD10 can induce cell proliferation via TCF-4. Finally, dysregulation of CACNA1D can further suppress tumor apoptosis in prostate cancer. The identification of critical genes and valuable biological processes can be useful for the understanding of the molecular mechanism of prostate cancer.

Keywords: mRNA, DEGs, miRNA, prognostic

Introduction
Prostate cancer, as the second most common cancer in men and fourth most common tumor type worldwide, is estimated to be diagnosed in 220,800 men in the USA, and almost 12% (27,540) died of their disease. The World Health Organization (WHO)'s report in 2012 shows that the incidence rates were estimated to be 10.5% and 4.5% in east and south-central Asia, respectively. In recent years, however, the incidence and mortality rate of prostate cancer have experienced a steady increase in some Asian countries. The primary prostate cancer has a 5-year survival rate of nearly 100%, and the distant metastasis has a 5-year survival rate of 28%. Various factors contribute to the incidence of prostate cancer including genetic, age, lifestyle and race.

In the current era, several prostate cancer biomarkers have been approved by the US Food and Drug Administration (FDA). One of these is PCA antigen 3 (PCA3) which is specifically upregulated in tumor tissues. Another one is PSA isoform (ProPSA) which can distinguish tumor tissues from normal tissues. In recent years, molecular and genetic profiles have been widely used for the screening of therapeutic targets. The fusion of androgen-regulated promoters with ERG and E26 transformation-specific (ETS) family members has been identified as the most common alteration...
in prostate cancer genome.5 Tomlins et al’s6 study shows that approximately 40%–50% of prostate cancers have TMPRSS2-ERG fusion. MALAT-1 fragment named as MD-miniRNA has been reported to differentiate prostate tumor and normal samples with the sensitivity of 56.6% and specificity of 84.8%.9 Somatic point mutations such as SPOP, TP53, FOXA1 and PTEN also have been frequently identified in prostate cancer.10 In addition, gene expression signatures of prostate cancer have been widely revealed based on oligonucleotide or cDNA microarray technology. One gene that encodes the serine protease hepsin was identified to be upregulated in malignant prostate tumor samples.11 The protein level of two genes, hepsin and pim-1, was significantly correlated with clinical outcome based on tissue microarray.12 In addition, the overexpression of genes that participated in multiple signal pathways especially cell cycle regulation, DNA replication and DNA repair was significantly related to proliferation rate of metastatic prostate cancer.13 Although various studies have been carried out based on microarray or next-generation sequencing (NGS) technologies, studies by integrating mRNA and miRNA expression profiles are rare in prostate cancer. More and more publicly available datasets were submitted to Gene Expression Omnibus (GEO) database with the development of technologies. In this study, to gain further insight into the molecular mechanism of prostate cancer, we comprehensively characterized the expression profiles of 49 prostate tumor samples. First, the significant common differentially expressed genes (DEGs) from two datasets were identified using advanced bioinformatic algorithms. Then, the common DEGs were subject to Cytoscape. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. Further, the regulation network between mRNA and six reported miRNAs was constructed. Finally, prognostic performance of the common DEGs was validated in silico using SurvExpress online database.

Materials and methods

mRNA and miRNA expression profiles download

A lot of publicly available gene expression profiles have been submitted to GEO or ArrayExpress database in recent years. In this study, we carefully searched the GEO database and downloaded two representative mRNA expression profiles of prostate cancer. One dataset, with GEO accession number GSE46602, was deposited by Mortensen et al,14 and this dataset consisted of 36 tumor tissues and 14 normal tissues. Another dataset named GSE55945 was submitted by Arredouani et al15 which contains 13 tumor tissues and eight normal tissues. The two datasets were obtained from Affymetrix Human Genome U133 Plus 2.0 Array. Detailed sample information and experiment design can be obtained from previous published papers.

Identification of DEGs

To explore the significant DEGs in prostate cancer, two datasets were carefully analyzed with in-house pipeline based on R language and public annotation databases. Briefly, mRNA expression values from the two datasets were subject to background correction to reduce noise. Then, normalization and log2 transformation of the expression values were carried out using GeneChip Robust Multi-array Analysis (GC-RMA) algorithm.16 In addition, control probe sets or uninformative probe sets were removed, and the mean expression value was calculated for the genes with multiple probes. Finally, DEGs between tumor and normal tissues were screened out using algorithm within Limma (linear models for microarray analysis) package.17 Criteria for DEG selection were set to adjust P-value ≤0.05 and absolute log2 fold change ≥2 which can significantly reduce false positive. In addition, common DEGs between the two datasets were figured out based on Venn diagram, and the expression pattern of the common DEGs in all samples was constructed by using heatmap.2 method within ggplot package.18

GO and KEGG pathway annotation

To explore biological functions of the DEGs, the identified common DEGs were further subject to GO and KEGG pathway enrichment using online tool of DAVID (Database for Annotation, Visualization and Integrated Discovery).19 All parameters were set as default, and the GO terms consisted of biological process (BP), cellular component (CC) and molecular function (MF).

mRNA–miRNA regulation network construction

More and more investigations have shown that miRNA can regulate tumor-related mRNA expression and play an important role in the whole process of tumor.20 In the study by Alhasan et al,21 six miRNAs, including miR-605-5p, miR-135a-3p, miR-495-5p, miR-433-5p, miR-371a-3p and miR-106a-5p, were demonstrated to be significantly up/down-expressed in prostate cancer. Blind studies also validated that the six miRNAs can differentiate prostate cancer patients from healthy people with 89% accuracy.22 Based on the six miRNAs, we constructed the regulation network using CyTargetLinker plugin in Cytoscape.23 In brief, target genes for the six miRNAs were predicted using microCosm, mirTarbase and TargetScan databases. Then, intersection between the common DEGs and target genes were selected and subject to Cytoscape.
Prognostic performance of the common DEGs
It is critical to validate the clinical outcomes of identified common DEGs for the diagnosis or treatment of prostate cancer. In this study, in silico validation was conducted using SurvExpress online tool that is based on publicly available mRNA expression profiles with clinical outcomes. Two datasets, GSE16560 and GSE40272, were used, and parameters were carefully selected with the instruction from the developer. In brief, we need to input our list of genes, select tissues, datasets, and set survival analysis options. The censored parameter was set to survival_month. All the samples were used as train set and test set for the estimation of beta coefficient using Cox model and the estimation of prognostic index. The optional weight parameter was estimated as the beta coefficient from the Cox model. Other parameters were left as default. Detailed sample information for the two datasets can be obtained from previous reports.

Results
Identification of DEGs in prostate cancer
To identify significant DEGs, the gene expression values were subject to preprocessing first. Figure 1 shows that the values are almost at the same level for the two datasets after

Figure 1 Boxplots for gene expression of each sample in GSE46602 (A) and GSE55945 (B).
normalization. Then, significant DEGs were figured out by applying robust empirical Bayesian method. Results showed that a total of 286 and 51 genes in GSE46602 and GSE55945, respectively, were significantly dysregulated in tumor samples. In GSE46602, 90 genes (31.5%) were upregulated and 196 genes were downregulated (68.5%), and in GSE55945, 32 genes (62.7%) were upregulated and 19 genes were downregulated (37.3%). In addition, 30 genes were found to be simultaneously differentially expressed in both GSE46602 and GSE55945 (Figure 2A), and the fold change and adjusted P-value are listed in Table 1. In addition, fold change correlation of the common DEGs was 0.78 (Figure 2B), which indicates that the expression level of the common DEGs was consistent between the two datasets. Further, the expression pattern of the common DEGs in all samples was constructed and subject to hierarchical clustering. Figure 3 shows that tumor (red) and normal (blue) samples can be clearly classified into two different subgroups except for Normal_18 and Normal_1. Error assignment for these two samples was probably due to tumor heterogeneity or variation in expression values.

**GO and KEGG pathway annotation**

To obtain the corresponding biological functions of the common DEGs, functional annotations were carried out using DAVID online tool that is based on GO and KEGG pathway database. As listed in Table 2, SLC2A5 and CACNA1D can be enriched into carbohydrate digestion and absorption KEGG pathway. GO annotation results indicated that the common DEGs were mainly related to the BP of xenobiotic metabolic process, anterior/posterior pattern specification, locomotory exploration behavior, proximal/distal pattern formation and behavioral fear response with P-value

---

**Figure 2** Venn diagram of the DEGs in GSE46602 and GSE55945 (A) and fold change correlation between the two datasets (B). **Abbreviation:** DEGs, differentially expressed genes.

**Table 1** The identified 30 common DEGs in GSE46602 and GSE55945

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>GSE46602</th>
<th>GSE55945</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adjusted P-value</td>
<td>Log FC</td>
</tr>
<tr>
<td>ACSM1</td>
<td>2.39E-06</td>
<td>2.94</td>
</tr>
<tr>
<td>AMACR</td>
<td>1.16E-05</td>
<td>2.88</td>
</tr>
<tr>
<td>AOC1</td>
<td>1.26E-09</td>
<td>-3.05</td>
</tr>
<tr>
<td>C1QTNF3-AMACR</td>
<td>9.23E-09</td>
<td>3.79</td>
</tr>
<tr>
<td>CACNA1D</td>
<td>0.000226</td>
<td>2.08</td>
</tr>
<tr>
<td>CCK</td>
<td>1.82E-07</td>
<td>-2.38</td>
</tr>
<tr>
<td>CD177</td>
<td>1.89E-12</td>
<td>-6.06</td>
</tr>
<tr>
<td>CYP3A5</td>
<td>1.31E-13</td>
<td>-3.73</td>
</tr>
<tr>
<td>DLX1</td>
<td>3.35E-08</td>
<td>4.41</td>
</tr>
<tr>
<td>DNAH5</td>
<td>2.61E-05</td>
<td>2.04</td>
</tr>
<tr>
<td>F5</td>
<td>0.00165576</td>
<td>3.00</td>
</tr>
<tr>
<td>HOXC4</td>
<td>1.51E-09</td>
<td>2.51</td>
</tr>
<tr>
<td>HOXC6</td>
<td>2.91E-08</td>
<td>3.10</td>
</tr>
<tr>
<td>HOXD10</td>
<td>8.37E-09</td>
<td>-2.30</td>
</tr>
<tr>
<td>HPN</td>
<td>2.97E-14</td>
<td>2.60</td>
</tr>
<tr>
<td>INS1M1</td>
<td>0.00204</td>
<td>2.15</td>
</tr>
<tr>
<td>KCNG3</td>
<td>0.00233</td>
<td>2.06</td>
</tr>
<tr>
<td>LINC00844</td>
<td>6.18E-12</td>
<td>-4.45</td>
</tr>
<tr>
<td>LOCI100287413</td>
<td>1.58E-08</td>
<td>3.19</td>
</tr>
<tr>
<td>LAMP</td>
<td>5.33E-11</td>
<td>-2.75</td>
</tr>
<tr>
<td>MS4AR</td>
<td>0.000133</td>
<td>2.04</td>
</tr>
<tr>
<td>OR51E2</td>
<td>0.0029</td>
<td>2.60</td>
</tr>
<tr>
<td>PCA3</td>
<td>1.07E-06</td>
<td>4.19</td>
</tr>
<tr>
<td>PDUMS</td>
<td>1.41E-08</td>
<td>2.49</td>
</tr>
<tr>
<td>PENK</td>
<td>2.25E-05</td>
<td>-2.26</td>
</tr>
<tr>
<td>PRAT47</td>
<td>0.00012</td>
<td>3.07</td>
</tr>
<tr>
<td>SIM2</td>
<td>3.80E-11</td>
<td>-3.52</td>
</tr>
<tr>
<td>SLC2A5</td>
<td>8.13E-11</td>
<td>-2.35</td>
</tr>
<tr>
<td>SV2B</td>
<td>6.19E-13</td>
<td>-3.16</td>
</tr>
<tr>
<td>TMEM45B</td>
<td>0.000505</td>
<td>2.42</td>
</tr>
</tbody>
</table>

**Abbreviations:** DEGs, differentially expressed genes; FC, fold change.
of <0.05 (Table 2). In addition, four MF terms including sequence-specific DNA binding, transcription factor activity, sequence-specific DNA binding and neuropeptide hormone activity were enriched with *P*-value of <0.05 (Table 2). CC annotation showed that HPN, SLC2A5, F5, etc mainly located on plasma membrane (Table 2).

**miRNA–mRNA regulation network**

Further dysregulation mechanism of the common DEGs was explored by constructing miRNA–mRNA regulation network. Prediction results showed that the six miRNAs can target 15, 1,252 and 2,046 genes in miRTarBase, TargetScan and MicroCosm databases, respectively. In addition, four out of the predicted target genes can be identified in the common DEGs. Based on the miRNA–mRNA pairings constructed using Cytoscape, we can find that miR-106a-5p can regulate the expression of PDLIM5 and SV2B (Figure 4). Also, miR-135a-3p and miR-371a-3p can regulate the expression of HOXD10 and CACNA1D, respectively (Figure 4).

**Prognostic performance of the common DEGs**

Clinical outcome validation is critical for the identification of biomarker in cancer research. In this study, two independent...
publicly available datasets, GSE16560 and GSE40272, were used to validate the prognostic performance of the identified common DEGs. Survival analysis based on clinical information from the two datasets showed that high-/low-risk prostate cancer group can be significantly differentiated in the two datasets with $P$-value $1.9\times 10^{-7}$ and $1.0\times 10^{-5}$, respectively (Figure 5). In addition, higher concordance index (CI) indicated that the common DEGs can be used for the prediction of prognostic performance (Table 3).

**Discussion**

The comprehensive analysis based on mRNA and miRNA expression profiles advanced the understanding of the molecular mechanism of prostate cancer. A total of 30 DEGs were identified in 49 tumor samples by comparing with normal samples. Importantly, all tumor and normal samples can be clearly classified into two groups based on the identified DEGs.

KEGG pathway annotation showed that SLC2A5 and CACNA1D participated in the carbohydrate digestion and absorption pathway. The protein encoded by SLC2A5 is a fructose transporter responsible for fructose uptake, and the protein encoded by CACNA1D is the alpha-1D subunit of calcium channel. Enhanced growth and proliferation of tumor cells place increased demand for nutrients for the synthesis of DNA and protein. The dysregulation of transporter and calcium channel probably increase the availability of nutrients.

The complex molecular mechanism of prostate cancer was further unveiled by miRNA–mRNA regulation network. The regulation network showed that miR-106a-5p can regulate the expression of SV2B (Figure 4). This gene encodes a member of the synaptic vesicle protein 2 family which participates in the regulation of vesicle trafficking and exocytosis. Expressed sequence tag (EST) analysis results showed that SV2B was differentially expressed in prostate cancer. GO annotations related to this gene included transporter activity and transmembrane transporter activity. Due to high demand of nutrient for tumor generation and metastasis, the dysregulation of SV2 in prostate cancer can provide necessary nutrient.

Another critical gene is HOXD10 which is the target of miR-135a-3p (Figure 4). This gene belongs to homeobox (Hox) superfamily and encodes transcription factors that control cell differentiation and morphogenesis during development. Documents showed that tumor growth
can be suppressed with ectopic expression of HOXB13 via downregulation of β-catenin–TCF pathway.\textsuperscript{35} Jung et al.\textsuperscript{34} study also demonstrated that HOXB13 can negatively regulate the expression of TCF-4 and then suppress cell proliferation in prostate cancer. Expression study also shows that HOXD10 is commonly downregulated in gastric tumor tissue compared with normal stomach tissue. Functional analysis revealed that HOXD10 can upregulate IGFBP3, which can activate caspase-3 and caspase-8 and subsequently induce cell apoptosis. All the results suggest that HOXD10 probably functions as a tumor suppressor in prostate cancer.

In addition, CACNA1D was significantly dysregulated and can be regulated by miR-371a-3p (Figure 4). This gene encodes the alpha-1D subunit of calcium channel. Calcium channel has been demonstrated to contribute to cell proliferation, differentiation and apoptosis. Calcium channel blocker has an important regulatory role and can affect malignant transformation in prostate cancer.\textsuperscript{36} It has also been documented that Ca2+ signaling participated in androgen-induced PSA secretion. Prostate cancer cell proliferation and androgen receptor-mediated gene expression are significantly suppressed with the addition of nifedipine and tetrandrine.\textsuperscript{37} In addition, knocking down of CACNA1D by siRNA can abrogate androgen receptor transactivation.\textsuperscript{37}

Apart from the abovementioned genes, several novel or rarely reported genes such as PDLIM5, HOXC4 and PENK were also identified in prostate cancer.

**Conclusion**

Multiple biological processes probably involved in the development and progression of prostate cancer. First, the dysregulation of SV2 can regulate transporter and transmembrane transporter activity and then provide necessary nutrient for tumor cell proliferation. Then, HOXD10 can induce cell proliferation via TCF-4. Finally, the dysregulation of CACNA1D can further suppress tumor apoptosis.

**Acknowledgments**

This work was supported by grants from the Natural Science Foundation of Jiangsu Province (BK20160481), the Project for Medical Young Talents of Jiangsu Province (QNRC2016360) and the Natural Science Foundation of Yangzhou (YZ2015111).

**Author contributions**

All authors contributed toward data analysis, drafting and critically revising the paper, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

**Disclosure**

The authors report no conflicts of interest in this work.

**References**