Screening of feature genes of the renal cell carcinoma with DNA microarray


Abstract. – AIM: To investigate the underlying molecular mechanisms of renal cell carcinoma (RCC) by using the microarray expression profiles of normal kidney and RCC tissue for early diagnosis and treatment of RCC.

MATERIALS AND METHODS: The gene expression profile of GES781 was downloaded from Gene Expression Omnibus database, including nine tissue samples of RCC tissues removed from nine patients and eight adjacent normal renal tissue samples. We identified the differentially expressed genes (DEGs) by Multtest package in R software. The screened DEGs were further analyzed by bioinformatics methods. Firstly, the comparison of the DEGs expression degree was performed by cluster analysis. Secondly, DAVID was used to perform functional analysis of up- and down-regulated genes and the protein-protein interaction (PPI) networks were constructed by prePPI. Finally, the pathways of genes in PPI networks were discovered by WebGestalt.

RESULTS: Compared with the control, we screened 648 down-regulated and 681 up-regulated DEGs. And the down- and up-regulated DEGs with maximum expression degree were UMOD (uromodulin) and FABP7 (fatty acid binding protein 7), respectively. There was significant difference in the gene expression between the normal kidney and RCC tissue. The up-regulated DEGs in RCC tissue were significantly related to the immune responses and the down-regulated DEGs were significantly related to the oxidation reduction. The most significant pathway in the PPI network of UMOD was cytokine-cytokine receptor interaction.

CONCLUSIONS: The screened DEGs have the potential to become candidate target molecules to monitor, diagnose and treat the RCC, and might be beneficial for the early diagnosis and medication control of RCC.

Key Words: Renal cell carcinoma, Cluster analysis, Pathway analysis, Interaction network.

Introduction

Renal cell carcinoma (RCC) is a common malignant tumor of the highest mortality rate in the genitourinary cancers and the incidence of RCC has been steadily rising by 2-4% each year. Since the biological behavior of RCC is various and the molecular pathogenesis of RCC remains unclear, patients are picked up with advanced stages. Therefore, discovery of the underlying molecular mechanisms of RCC is urgently needed and would contribute to improving early diagnosis and patient therapy.

Many genes and signaling pathways are involved in RCC development. Two loci on chromosome 11 including WT1 (Wilms’ tumor 1, located at 11p13) gene and WT2 (Wilms’ tumor 2, located at 11p15) gene have been implicated in the genesis of a minority of Wilms tumors. An abnormal WT1 gene is present in patients with WAGR syndrome (Wilms tumor, aniridia, genitourinary abnormalities, mental retardation) or Drash syndrome (male pseudohermaphroditism, progressive glomerulonephritis); an abnormal WT2 gene is present in patients with Beckwith-Wiedemann syndrome or hemihypertrophy. Several studies have revealed altered expression of epidermal growth factor receptor (EGFR)-family members in RCC. And it has been reported that CXCR2 (chemokine receptors 2)/CXCR2 ligand biology is an important component of RCC tumor-associated angiogenesis and tumorigenesis. What’s more, Petrella et al reported that IL (interleukin)-1 induced tumor cell invasion of RCC cells through a process that was dependent on the activity of matrix metalloproteinases and was independent of migration rate. Fatty acid-binding proteins (FABP) are involved in the intracellular transport of fatty acids, and the level of brain-type FABP was over
expressed and the liver-type FABP appeared to be reduced in RCC, which are important in cell signaling, regulation of gene expression, cell growth, and differentiation. Furthermore, bone morphogenetic proteins (BMPs) are cytokines which are important for kidney homeostasis and play important roles in the RCC. Markic et al. found that the expression of BMP2, BMP4, BMP6, BMP7, BMPRIA, BMPRIB and BMPRII have stronger expression levels in RCC, especially the expression level of BMP2. Though an increasing number of molecular changes have been identified in RCC, the molecular mechanisms of RCC aren’t fully understood.

Since DNA microarray analysis is an established technique to simultaneously compare gene expression patterns between different conditions, it is one approach to identifying key molecular events and pathways involved in RCC. In this study, we downloaded the gene expression profiles of RCC tissues and patient-matched normal kidney tissues. The differentially expressed genes (DEGs) in RCC tissue were selected by using the expression profiling of normal kidney tissue and RCC tissue. In addition, the screened DEGs were further analyzed by bioinformatics methods. We anticipate that our work could improve the understanding of the underlying molecular mechanisms of RCC and could provide novel insights for the early diagnosis and medication control of RCC.

**Materials and Methods**

**Samples**

The transcription profile of GSE781 was downloaded from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) which was based on the Affymetrix Human Genome U133A Array. Total seventeen kidney tissue specimens were available for further analysis, including nine tissue samples of RCC tissues removed from nine patients during radical nephrectomy as well as eight adjacent normal renal tissue present in the same surgical samples. The annotation information of all probe sets was provided by Affymetrix Company where we downloaded the raw data files.

**Data Preprocessing and Differential Expression Analysis**

Data preprocessing and normalization were performed using the Support Vector Regression. Firstly, the probe-level data in CEL files were converted into expression measures. We used scoring methods to select a single representative probe set for each gene, thus creating a simple one-to-one mapping between gene and probe set. Then, the missing parts of data were imputed and the complete data were standardized with Support Vector Regression. The Multitest package in R software was used to identify DEGs in RCC tissue. In order to circumvent the multi-test problem which might induce too much false positive results, the Benjamini-Hochberg (BH) procedure was used to adjust the raw p-values into false discovery rate (FDR). The FDR < 0.05 and |logFC| > 1 were used as the cut-off criteria.

**Comparison of the Gene Expression Between the Normal Kidney and RCC Tissue**

The gene expression levels of the same tissue were significantly different in various disease states because of the specificity of gene expression in the same species under different conditions. The expression values of DEGs screened from RCC tissue were hierarchically clustered by Cluster to intuitively observe the differences in gene expression levels between the normal kidney tissue and RCC tissue.

**Functional Enrichment Analysis of DEGs**

Based on the deficiency of individual gene analysis, the gene set enrichment analysis could evaluate differential expression patterns of gene groups instead of those of individual genes to distinguish whether the biological functions and characteristics changed. The p value indicated the probability that a gene was endowed a Gene Ontology (GO) function randomly and it was usually used as the criterion for assigning a certain function to a module. The smaller the p value was, the more likely to prove that function of the module was not occurred randomly but for the purpose to accomplish a certain biological function, and it has unparalleled biological significance. DAVID (Database for Annotation, Visualization and Integrated Discovery) bioinformatics resources consists of an integrated biological knowledgebase and analytic tools aimed at systematically extracting biological meaning from large gene or protein lists. The functional enrichment analysis for the screened DEGs was performed by DAVID and the FDR < 0.05 was chosen as the cut-off criterion.
**Protein-Protein Interaction (PPI) Network Construction**

The protein-protein interactions (PPIs) research could reveal the functions of proteins at the molecular level and help discover the rules of cellular activities including growth, development, metabolism, differentiation and apoptosis. The identification of protein interactions in a genome-wide scale is an important step for the interpretation of the cellular control mechanisms. In this analysis, we screened the up- and down-regulated DEGs with the maximal expression levels and constructed the PPI networks by prePPI.

PrePPI is a prediction method of genome-wide protein interaction based on the three-dimensional information. This method based on the three-dimensional structures of proteins and could provide accurately predictive information with high coverage. The accuracy and coverage rate of prePPI are vastly superior to the method based on non-structure information, due to the use of homology model and combination of the geometrical relationship.

**Pathway Enrichment Analysis**

The pathway enrichment analysis of genes in the constructed PPI networks where the up- and down-regulated DEGs with maximal expression levels located was performed by using WebGestalt. The FDR < 0.05 was selected as the cut-off criterion.

**Results**

**Identification of DEGs**

The standardized expression profiling data after preprocessing were shown in Figure 1. For dataset GSE781, a total of 1329 DEGs were identified at the criteria of FDR < 0.05 and |logFC|>1, including 648 down-regulated genes and 681 up-regulated genes. The top ten down-regulated DEGs and up-regulated DEGs were listed in Table 1. And the down- and up-regulated DEGs with maximum expression degree were UMOD (uromodulin) and FABP7 (fatty acid binding protein 7), respectively.

**Comparison of the Gene Expression Between the Normal Kidney and RCC Tissue**

The expression values of DEGs screened from the RCC tissue were hierarchically clustered by Cluster, and the differences in gene expression between the normal kidney and RCC tissue were intuitively observed. The color contrast indicated that there was significant difference in the gene expression between the normal kidney and RCC tissue (Figure 2).

**Functional Enrichment Analysis of DEGs**

The functional enrichment analysis of all DEGs was performed by DAVID and the FDR < 0.05 was chosen as the cut-off criterion. The up-regulated DEGs in RCC tissue were significantly
Table I. The top ten down-regulated DEGs and up-regulated DEGs.

<table>
<thead>
<tr>
<th>ID</th>
<th>Gene</th>
<th>logFC</th>
<th>FDR</th>
<th>ID</th>
<th>Gene</th>
<th>logFC</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>206716_at</td>
<td>UMOD</td>
<td>-8.34201</td>
<td>3.98E-08</td>
<td>205029_s_at</td>
<td>FABP7</td>
<td>6.934455</td>
<td>0.000348</td>
</tr>
<tr>
<td>206054_at</td>
<td>KNG1</td>
<td>-7.20445</td>
<td>3.98E-08</td>
<td>218484_at</td>
<td>NDUF4L2</td>
<td>6.594476</td>
<td>2.4E-07</td>
</tr>
<tr>
<td>204704_s_at</td>
<td>ALDOB</td>
<td>-7.01503</td>
<td>7.33E-05</td>
<td>213479_at</td>
<td>PTPX2</td>
<td>4.971608</td>
<td>0.000188</td>
</tr>
<tr>
<td>206024_at</td>
<td>HPD</td>
<td>-7.0068</td>
<td>4.07E-06</td>
<td>206025_s_at</td>
<td>TNFAIP6</td>
<td>4.696365</td>
<td>2.95E-06</td>
</tr>
<tr>
<td>205244_s_at</td>
<td>SLC13A3</td>
<td>-6.77881</td>
<td>9.2E-07</td>
<td>216834_at</td>
<td>RGS1</td>
<td>4.652913</td>
<td>4.49E-05</td>
</tr>
<tr>
<td>220281_at</td>
<td>SLC12A1</td>
<td>-6.3331</td>
<td>1.35E-05</td>
<td>221870_at</td>
<td>EHD2</td>
<td>4.60127</td>
<td>9.2E-07</td>
</tr>
<tr>
<td>221298_s_at</td>
<td>SLC22A8</td>
<td>-6.33018</td>
<td>3.91E-06</td>
<td>204416_s_at</td>
<td>APOC1</td>
<td>4.535479</td>
<td>4.7E-06</td>
</tr>
<tr>
<td>219554_at</td>
<td>RHCG</td>
<td>-6.24629</td>
<td>3.98E-08</td>
<td>201313_at</td>
<td>ENO2</td>
<td>4.476422</td>
<td>1.35E-05</td>
</tr>
<tr>
<td>205892_s_at</td>
<td>FABP1</td>
<td>-6.22295</td>
<td>8.47E-05</td>
<td>213915_at</td>
<td>NKG7</td>
<td>4.395586</td>
<td>3.63E-06</td>
</tr>
<tr>
<td>209977_at</td>
<td>PLG</td>
<td>-6.20962</td>
<td>9.35E-05</td>
<td>221009_s_at</td>
<td>ANGPTL4</td>
<td>4.392976</td>
<td>1.78E-06</td>
</tr>
</tbody>
</table>

related to the immune responses and the down-regulated DEGs were significantly related to the oxidation reduction (Table II).

PPI Network Construction

The interactive objects of the UMOD gene and FABP7 gene were searched by prePPI software published in Nature in 2012, and the PPI networks were constructed. According to the three-dimensional structure information of proteins included in prePPI software, we found 207 and 44 interaction partners for UMOD gene and FABP7 gene respectively (Figure 3). The IL-8 and IGFBP7 (Insulin-like growth factor binding protein 7) were involved in the PPI network of FABP7.

Figure 2. The differences in expression values of DEGs between the normal kidney and RCC tissue.
Table II. The functional enrichment analysis for the identified up- and down-regulated DEGs.

<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Up-regulated genes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0006955~immune response</td>
<td>105</td>
<td>5.81E-29</td>
</tr>
<tr>
<td>GO:0006952~defense response</td>
<td>75</td>
<td>6.38E-14</td>
</tr>
<tr>
<td>GO:0045321~leukocyte activation</td>
<td>42</td>
<td>1.56E-11</td>
</tr>
<tr>
<td>GO:0002684~positive regulation of immune system process</td>
<td>40</td>
<td>2.21E-10</td>
</tr>
<tr>
<td>GO:0046649~lymphocyte activation</td>
<td>34</td>
<td>1.27E-08</td>
</tr>
<tr>
<td>GO:009611~response to wounding</td>
<td>58</td>
<td>4.96E-08</td>
</tr>
<tr>
<td>GO:0048584~positive regulation of response to stimulus</td>
<td>33</td>
<td>5.34E-06</td>
</tr>
<tr>
<td>GO:0009725~response to hormone stimulus</td>
<td>35</td>
<td>0.005852</td>
</tr>
<tr>
<td>GO:0016054~organic acid catabolic process</td>
<td>37</td>
<td>1.72E-20</td>
</tr>
<tr>
<td>GO:0046395~carboxylic acid catabolic process</td>
<td>37</td>
<td>1.72E-20</td>
</tr>
<tr>
<td>GO:0006732~coenzyme metabolic process</td>
<td>37</td>
<td>2.50E-15</td>
</tr>
<tr>
<td>GO:0044271~nitrogen compound biosynthetic process</td>
<td>37</td>
<td>3.61E-05</td>
</tr>
<tr>
<td>GO:0051186~cofactor metabolic process</td>
<td>40</td>
<td>3.66E-14</td>
</tr>
<tr>
<td>GO:0048878~chemical homeostasis</td>
<td>43</td>
<td>0.009707</td>
</tr>
<tr>
<td>GO:0006812~cation transport</td>
<td>47</td>
<td>0.002305</td>
</tr>
<tr>
<td>GO:0055085~transmembrane transport</td>
<td>48</td>
<td>0.001261</td>
</tr>
<tr>
<td>GO:0006901~generation of precursor metabolites and energy</td>
<td>49</td>
<td>7.88E-13</td>
</tr>
<tr>
<td>GO:010033~response to organic substance</td>
<td>56</td>
<td>0.001357</td>
</tr>
<tr>
<td>GO:0006811~ion transport</td>
<td>68</td>
<td>9.90E-07</td>
</tr>
<tr>
<td>GO:0055114~oxidation reduction</td>
<td>108</td>
<td>3.88E-36</td>
</tr>
</tbody>
</table>

**Pathway Enrichment Analysis**

The pathway enrichment analysis for genes in the PPI networks was analyzed by using Webgestalt. Two significant pathways in the PPI network of UMOD gene were obtained (Table III). The most significant pathway was hsa04060: cytokine-cytokine receptor interaction (FDR = 5.20E-11). The IL2RA, TNFRSF4, CXCL1, CXCL9, CXCR2, BMP2, BMPR2, BMP7 and IL-8 genes participated in this pathway were interacted with the UMOD gene. The second most important pathway was TGF-beta signaling pathway (FDR = 5.26E-05). And the BMP4, BMP2, ABMRPR2, TGF3, BMP7 and BMP6 were participated in this pathway. No significant pathway in the PPI network of FABP7 gene was discovered because of low number of genes related to FABP7.

**Discussion**

In this study, we investigated the underlying molecular mechanism of RCC by using bioinformatics methods. Total 1329 DEGs were identified by the gene expression profiles of normal kidney tissue and RCC tissue. The down- and up-regulated DEGs with maximum expression degree were UMOD gene and FABP7 gene, respectively. Furthermore, the screened DEGs in RCC tissue are closely related with immune responses, hormone responses and cytokine-cytokine receptors.

The FABP7 is a member of fatty acid binding protein family and FABPs play roles in fatty acid uptake, transport, and metabolism. Several types of carcinomas, including RCC, overexpress FABP7. Domoto et al. found that S100A10, annexin II and FABP7 performed well as RCC markers and FABP7 was the most specific to RCC. Furthermore, we found IL-8 and IGFBP7 were involved in the PPI network of FABP7. It has been reported that IGFBP7 is a potential transcription factor with a variable distribution along the renal tubular epithelium and it could be involved in the regulation of cell growth and differentiation. Koo et al. revealed that the IL8 gene is maximally hypomethylated in RCC tissue compared to normal tissue. Therefore, our results were consistent with the previous reports.

The UMOD gene is exclusively transcribed in the kidney and encodes Tamm Horsfall protein (THP), also known as uromodulin. Recent genome-wide association studies have identified common variants in the UMOD region associated with kidney function and chronic kidney disease. In this study, we found that the most significant pathway in the PPI network of UMOD was...
Screening of feature genes of the renal cell carcinoma with DNA microarray

**Figure 3.** The constructed protein-protein interaction networks of UMOD gene (A) and FABP7 gene (B).

**Table III.** The most significant pathways enriched in the PPI network of UMOD gene.

<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
<th>FDR</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa04060: Cytokine-cytokine receptor interaction</td>
<td>26</td>
<td>2.20E-11</td>
<td>CXCL1, ACVRL1, OSMR, BMPR2, TGFBR3, CXCL9, CNTFR, CXCR2, CCL5, TNFRSF4, TGFBR2, CSF3R, EGF, LTA, BMP2, FLT1, IL2RA, IL8, FLT3, TGFBR1, CCR9, INHBB, CCR6, TNFSF11, CCR3, BMP7 INHBB, BMP4, NOG, BMP2, ACVRL1, ACVRL1, TGFBR1, BMPR2, TGFBR3, BMP7, THBS1, BMP6, TGFBR2</td>
</tr>
<tr>
<td>hsa04350: TGF-beta signaling pathway</td>
<td>12</td>
<td>5.26E-05</td>
<td></td>
</tr>
</tbody>
</table>
hsa04060: cytokine-cytokine receptor interaction. And the IL2RA, TNFRSF4, CXCL1, CXCR2 and IL-8 were participated in this pathway. Liu et al. found that the elimination of TNP expression from mouse kidneys caused a marked elevation of circulating IFN-γ, IL1α, TNF-α, IL6, CXCL1, and IL13. Several immune response and inflammatory genes were found to be up-regulated in RCC, including IL2RA, IL2RB, IL7R, IL10RB as well as tumour necrosis factor family genes TNFAIP3, TNFAIP6, TNFAIP8, TNFRSF4 by Magdalena et al. And the BMP4, BMP2, ABMPR2, TGBF3, BMP7 and BMP6 were participated the second most important pathway: TGF-beta signaling. Wang et al. found that BMP-2 could inhibit the growth of RCC as well as cause induction of osseous bone formation.

Conclusions

As expected, the screened DEGs have the potential to become candidate target molecules to monitor, diagnose and treat the RCC. And we anticipate that our work could contribute to understanding the molecular mechanisms of RCC. Since our study was designed to identify genes that are differentially regulated in RCC, we did not analyze enough samples of different tumor grades. Therefore, the molecular mechanism and target therapy of RCC need to be further explored and researched. Furthermore, the results of our study need to be confirmed by experimental research.

Conflict of Interest

The Authors declare that there are no conflicts of interest.

References

Screening of feature genes of the renal cell carcinoma with DNA microarray


