Introduction

Clear cell renal cell carcinoma (ccRCC) is a type of RCC developed in adults (1). It has been reported that ccRCC is the most aggressive subtype of RCC (2). Surgery by radical or partial nephrectomy is the main choice for treatment of ccRCC. Although chemotherapy and immunotherapy could be applicable in patients with metastatic ccRCC, the outcome is unsatisfactory (3). Even some of ccRCC are along with the worst prognosis among the common epithelial tumors of the kidney (4). Therefore, exploring molecular biomarkers serving as diagnostic and therapeutic targets, when used alone or in combination with other clinical parameters, are urgently required for better clinical management.

Accumulating evidences suggest that certain differentially expressed genes (DEGs) are closely related to disease progression. Duns et al. (5) showed that histone methyltransferase gene SET domain containing 2 (SETD2) is a novel tumor suppressor gene in the process of ccRCC. A chromatin-remodeling gene AT-rich interaction domain 1A (ARID1A) is considered to be a new prognostic marker in ccRCC (6). Actually, these genes often play important roles in ccRCC progression via certain function or specific pathways. A recent study indicated that frequent methylation of Kelch like ECH associated protein 1 (KEAP1) gene promoter was vital for ccRCC development via KEAP1/ nuclear factor erythroid-2 related factor (NRF2) pathway (7). Moreover, some microRNAs (miRNAs) are abnormally expressed in ccRCC and contribute to tumorigenesis. Urinary miRNAs have been proved to be the predictors of tumor metastasis in ccRCC (8). Yang et al. (9) showed that miR-506 was down-regulated in ccRCC and inhibited cell growth and metastasis via targeting flotillin 1. Although there have been many researches to find genetic biomarkers for ccRCC, characterization progress of the genetic events associated with this cancer is not fully clear yet.

In this study, meta-analysis was used to detect potential DEGs between ccRCC and normal samples based on three microarray datasets. Moreover, functional and pathway enrichment analyses were carried out for these DEGs. Then, protein-protein interaction (PPI) network was investigated. Furthermore, miRNA-target gene interaction network was constructed. We hoped to explore...
the underlying molecular mechanisms of ccRCC and provide candidate target genes for ccRCC gene therapy.

Materials and Methods

Microarray data and preprocessing

This is a bioinformatics-based study. Microarray datasets were downloaded from Gene Expression Omnibus database based on the data quantity, sample grouping, microarray platform (Affymetrix) and number of citation. Finally, three datasets were selected: GSE6344, GSE781 and GSE53000. The reasons for selection are as follows: i. Large data quantities, ii. Clear grouping of the experiment (tumor vs. normal), iii. Common microarray platforms (Affymetrix), iv. Consistent sample types (tissue samples). In details, 10 ccRCC and 10 normal tissue samples sequencing on the platform of GPL96 [HG-U133A] Affymetrix Human Genome U133A Array were selected for analysis in microarray dataset GSE6344 (10). For GSE781 (11), 12 ccRCC and 5 normal tissue samples sequenced on the platform of GPL96 [HG-U133A] Affymetrix Human Genome U133A Array were selected. In addition, all samples in GSE53000 (12) (56 ccRCC and 6 normal tissue samples) sequencing on the platform of [HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version] were used for analysis. Specially, GSE53000 included two samples of lymph node metastasis and one sample of venous thrombus metastasis. Thus, principal component analysis (PCA) was carried out for the 56 ccRCC tissue samples and 6 normal tissue samples. As shown in Figure S1 (Supplementary Online Information at www.celljournal.org), two samples from lymph node metastasis and one sample from venous thrombus metastasis were obviously clustered together, while separated from normal samples. So they could be unified as tumor group for the following analysis. Sample information of the three datasets was shown in Tables S1-S3 (Supplementary Online Information at www.celljournal.org). In order to eliminate the expression value heterogeneity of each gene in different platforms, CEL format of files of three datasets were pooled together and preprocessed using Affyl software, including background adjustment, quantiles normalization, summarization and log2 fold change (FC) transformation using Robust Multi-array Average (RMA) algorithm in Affy package (13). The probe identities were converted to gene symbols based on the annotation files downloaded from different platforms. Probes that did not correspond to gene symbols were discarded. For different probes matched to one gene, the average value of different probes was used as the final expression value.

Differentially expressed genes identification

DEGs of ccRCC and normal samples were separately screened based on multiple experimental datasets using MetaDE package in R software (14). The heterogeneity test was performed according to the expression values of each gene under different experimental platforms with the statistical parameters of tau2, Qvalue and Qpval. The tau2=0 [estimated amount of (residual) heterogeneity] and Qpval=0.05 (P values for the test of heterogeneity) represented significant homogeneity. Finally, Benjami-Hochber adjusted P value (fdr) <0.05, tau2=0 and Qpval >0.05 were considered as the cut-off criteria for DEGs selection. Furthermore, the log2 FC of ccRCC vs. normal >0 represented up-regulated DEGs, while log2 FC of ccRCC vs. normal <0 represented that DEGs were down-regulated.

Functional annotation and pathway enrichment analysis of differentially expressed genes

The clusterProfiler is an online tool applied for enrichment analysis (15). Gene Ontology (GO) functional annotation was used to analyze functions assembled with the up- and down-regulated genes by clusterProfiler. GO functions include molecular function (MF), biological process (BP) and cellular component (CC). To better understand pathways of the involved DEGs, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was performed using clusterProfiler. A P value (the significance threshold of the hypergeometric test) of <0.05 and count (the number of enriched genes) of >2 were used as the cut-off criteria for this analysis.

Constructing protein-protein interaction network and modules analyses

PPI plays a key role in the completion of cellular functions, while they are usually correlated to each other in the form of a PPI network. The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) is a biological database of predicted and known PPIs. According to this database, the PPI network of DEG-encoding proteins in each group was constructed with the criterion of combined score (medium confidence) >0.4, and it was then visualized by the Cytoscape (version 3.2.0) software (National Institute of General Medical Sciences, USA). Score of nodes in the current network was analyzed using degree centrality, a topology property index. Higher node score presented more important node in the network, suggesting that is more likely the hub node in this network. Furthermore, the MCODE tool (National Institute of General Medical Sciences, USA) (16) in Cytoscape was used to screen the modules from the network.

miRNA-target gene regulatory network construction

The potential ccRCC related miRNAs were explored based on Enrichr database. The miRNA-target gene regulatory network was constructed with the miRNA associating with up- and down-regulated gene based on Cytoscape software (National Institute of General Medical Sciences, USA).

Results

Differentially expressed genes investigation between clear cell renal cell carcinoma and control groups

With fdr <0.05, tau2=0 and Qpval >0.05, 955 DEGs
A Bioinformatics Study of ccRCC

were identified in ccRCC group compared to that of the normal controls, including 511 up-regulated and 444 down-regulated genes.

**Gene ontology function and Kyoto Encyclopedia of Genes and Genomes pathway enrichment analyses**

Using clusterProfiler, GO functional enrichment analysis was performed and the results showed that the up-regulated DEGs were significantly enriched in functions like immune system process (GO_BP, \( P=5.02E^{-18} \)), intracellular (GO_CC, \( P=5.86E^{-18} \)) and protein binding (GO_MF, \( P=1.09E^{-18} \)). Meanwhile, the down-regulated genes were mainly enriched in functions like small molecule metabolic process (GO_BP, \( P=1.39E^{-39} \)), cytoplasm (GO_CC, \( P=1.01E^{-06} \)) and binding (GO_MF, \( P=3.81E^{-16} \)) (Fig.1).

Pathway enrichment analysis showed that the up-regulated genes were enriched in pathways like Toll-like receptor signaling pathway (\( P=1.94E^{-05} \)), while the down-regulated genes were enriched in pathways like metabolic pathways (\( P=1.27E^{-28} \)). The top five enriched pathways are listed in Table 1.

![Fig.1: GO functional enrichment analysis for DEGs between ccRCC and normal samples. A. Up- and down-regulated genes presented in top none biological processes. B. Up- and down-regulated genes presented in top eight cellular functions. C. Up- and down-regulated genes presented in top eight molecular functions. GO; Gene ontology, ccRCC; Clear cell renal cell carcinoma, and DEGs; Differentially expressed genes.](image)

<table>
<thead>
<tr>
<th>Category</th>
<th>Pathway ID</th>
<th>Pathway</th>
<th>Count</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up</td>
<td>hsa04620</td>
<td>Toll-like receptor signaling pathway</td>
<td>13</td>
<td>1.94E-05</td>
</tr>
<tr>
<td></td>
<td>hsa05133</td>
<td>Pertussis</td>
<td>11</td>
<td>2.13E-05</td>
</tr>
<tr>
<td></td>
<td>hsa04145</td>
<td>Phagosome</td>
<td>14</td>
<td>3.15E-04</td>
</tr>
<tr>
<td></td>
<td>hsa05150</td>
<td>Staphylococcus aureus infection</td>
<td>8</td>
<td>3.18E-04</td>
</tr>
<tr>
<td></td>
<td>hsa04666</td>
<td>Fc gamma R-mediated phagocytosis</td>
<td>10</td>
<td>5.19E-04</td>
</tr>
<tr>
<td>Down</td>
<td>hsa01100</td>
<td>Metabolic pathways</td>
<td>113</td>
<td>1.27E-28</td>
</tr>
<tr>
<td></td>
<td>hsa01200</td>
<td>Carbon metabolism</td>
<td>26</td>
<td>9.72E-16</td>
</tr>
<tr>
<td></td>
<td>hsa00190</td>
<td>Oxidative phosphorylation</td>
<td>26</td>
<td>5.04E-14</td>
</tr>
<tr>
<td></td>
<td>hsa05012</td>
<td>Parkinson’s disease</td>
<td>26</td>
<td>2.56E-13</td>
</tr>
<tr>
<td></td>
<td>hsa00280</td>
<td>Valine, leucine and isoleucine degradation</td>
<td>15</td>
<td>1.04E-11</td>
</tr>
</tbody>
</table>

\( P<0.05 \) was considered to be significantly different. KEGG; Kyoto encyclopedia of genes and genomes.
Protein-protein interaction network and modules investigation

To dig out more effective information, regarding the DEGs mentioned above, PPI network was constructed on the basis of interaction relationship among the proteins. With score=0.4, a total of 2483 PPI pairs and 643 DEG-encoded proteins were identified. According to the score of degree centrality, TYRO protein tyrosine kinase binding protein (TYROBP, degree=68, up-regulation), cathepsin S (CTSS, degree=53, up-regulation), colony stimulating factor 1 receptor (CSF1R, degree=52, up-regulation), Fc fragment of IgE receptor Ig (FCER1G, degree=43, up-regulation), protein tyrosine phosphatase, receptor type C (PTPRC, degree=43, up-regulation), mitogen-activated protein kinase 1 (MAPK1, degree=43, up-regulation), CD53 molecule (CD53, degree=42, up-regulation), Ras-related C3 botulinum toxin substrate 2 (RAC2, degree=42, up-regulation), cluster of differentiation 14 (CD14, degree=41, up-regulation), cytochrome C, and somatic (CYCS, degree=40, down-regulation) were the top 10 proteins encoded by DEGs.

Fig.2: The modules obtained from protein-protein interaction network. A. The module “a” was constructed by 33 nodes and 267 interactions. B. The module “b” was constructed by 14 nodes and 91 interactions. C. The module “c” was constructed by 7 nodes and 21 interactions. D. The module “d” was constructed by 36 nodes and 114 interactions. E. The module “e” was constructed by 8 nodes and 20 interactions. F. The module “f” was constructed by 6 nodes and 14 interactions. G. The module “g” was constructed by 5 nodes and 10 interactions. H. The module “h” was constructed by 20 nodes and 39 interactions. Green node represents down-regulated gene; red node represents up-regulated gene.
The MOCDE software showed eight modules from PPI network. The detail information was showed in Figure 2. To further investigate crucial pathways involved in the process of ccRCC, the KEGG pathway analysis was performed on DEGs in these modules. Results showed that chemokine signaling pathway \((P=4.88E^{-05})\), Parkinson’s disease \((P=3.17E^{-24})\), protein digestion and absorption \((P=3.31E^{-08})\) and ribosome \((P=5.00E^{-08})\) were the most significant pathways enriched by DEG-encoded proteins in the respectively modules a, b, c and d. Meanwhile, NOTCH signaling pathway \((P=3.99E^{-02})\), oxidative phosphorylation \((P=7.89E^{-10})\), oxidative phosphorylation \((P=5.74E^{-08})\) and adipocytokine signaling pathway \((P=2.18E^{-02})\) were the most significant pathways enriched respectively by modules e, f, g and h. The top two KEGG pathways in each module are listed in Table 2.

**miRNAs-target gene regulatory network analyses**

Here, miRNAs targeted up- and down-regulated genes were investigated based on Enrichr software. Then, the regulatory network was constructed using Cytoscape software. Results showed that there were four miRNAs (including miR-145, miR-199B, miR-199A and miR-412), 94 up-regulated genes [such as interferon regulatory factor 7 \((IRF7)\), TYRO protein tyrosine kinase binding protein \((TYROBP)\) and \(CD14\)], as well as 45 down-regulated genes [such as PPARG coactivator 1 alpha \((PPARGC1A)\), dystroglycan 1 \((DAG1)\) and Klotho \((KL)\)] in the present network (Fig.3).

### Table 2: Top two KEGG pathways in modules enriched by the differentially expressed genes of clear cell renal cell carcinoma

<table>
<thead>
<tr>
<th>Module ID</th>
<th>Pathway ID</th>
<th>Pathway name</th>
<th>Count</th>
<th>P value</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>hsa04062</td>
<td>Chemokine signaling pathway</td>
<td>6</td>
<td>4.88E^{-05}</td>
<td>CXCL9, RAC2, CXCL16, HCK, GNG2...</td>
</tr>
<tr>
<td></td>
<td>hsa0460</td>
<td>Cytokine-cytokine receptor interaction</td>
<td>6</td>
<td>3.17E^{-04}</td>
<td>CXCL9, IL10RA, CXCL16, TNFSF13B, CSF1R...</td>
</tr>
<tr>
<td>b</td>
<td>hsa05012</td>
<td>Parkinson’s disease</td>
<td>14</td>
<td>3.17E^{-24}</td>
<td>NDUFB2, NDUFA3, UQCRFS1, NDUFA9, NDUFB8...</td>
</tr>
<tr>
<td></td>
<td>hsa00190</td>
<td>Oxidative phosphorylation</td>
<td>14</td>
<td>3.97E^{-24}</td>
<td>NDUFB2, NDUFA3, UQCRFS1, NDUFA9, NDUFB8...</td>
</tr>
<tr>
<td>c</td>
<td>hsa04974</td>
<td>Protein digestion and absorption</td>
<td>4</td>
<td>3.31E^{-08}</td>
<td>COL4A2, OLA1A1, COLA1A1, COL15A1</td>
</tr>
<tr>
<td></td>
<td>hsa04512</td>
<td>ECM-receptor interaction</td>
<td>3</td>
<td>1.15E^{-05}</td>
<td>COL4A2, COL1A1, COL4A1</td>
</tr>
<tr>
<td>d</td>
<td>hsa03010</td>
<td>Ribosome</td>
<td>7</td>
<td>5.00E^{-08}</td>
<td>RPL35A, RPS24, RPS15A, RPS16, RPL30...</td>
</tr>
<tr>
<td></td>
<td>hsa05150</td>
<td>Staphylococcus aureus infection</td>
<td>4</td>
<td>6.74E^{-05}</td>
<td>ITGAM, C1QA, C3AR1, C1QC</td>
</tr>
<tr>
<td>e</td>
<td>hsa04330</td>
<td>Notch signaling pathway</td>
<td>1</td>
<td>3.99E^{-02}</td>
<td>DTX3L</td>
</tr>
<tr>
<td></td>
<td>hsa04623</td>
<td>Cytosolic DNA-sensing pathway</td>
<td>1</td>
<td>3.99E^{-02}</td>
<td>IRF7</td>
</tr>
<tr>
<td>f</td>
<td>hsa00190</td>
<td>Oxidative phosphorylation</td>
<td>6</td>
<td>7.89E^{-10}</td>
<td>SDHB, COX5A, COX6A1, SDHC, ATP5G3...</td>
</tr>
<tr>
<td></td>
<td>hsa05012</td>
<td>Parkinson’s disease</td>
<td>5</td>
<td>9.97E^{-08}</td>
<td>DHB, COX5A, COX6A1, SDHC, ATP5G3</td>
</tr>
<tr>
<td>g</td>
<td>hsa00190</td>
<td>Oxidative phosphorylation</td>
<td>5</td>
<td>5.74E^{-08}</td>
<td>ATP6VD1D1, ATP6V0C, PPA2, ATP5C1, ATP6V1H...</td>
</tr>
<tr>
<td></td>
<td>hsa05110</td>
<td>Vibrio cholerae infection</td>
<td>3</td>
<td>3.94E^{-05}</td>
<td>ATP6VD1D1, ATP6V0C, ATP6V1H</td>
</tr>
<tr>
<td>h</td>
<td>hsa04920</td>
<td>Adipocytokine signaling pathway</td>
<td>3</td>
<td>2.18E^{-02}</td>
<td>MTOR, PPARGC1A, PPARA</td>
</tr>
</tbody>
</table>

*KEGG; Kyoto Encyclopedia of Genes and Genomes.*
Discussion

ccRCC is one of the most common RCC identified in adults presenting the worst prognosis among the common epithelial tumors of kidney. In this study, total 955 DEGs were identified in ccRCC from normal samples. GO analysis showed that these DEGs, including up-regulated \textit{TYROBP} and \textit{IRF7}, and down-regulated \textit{PPARGC1A}, were mainly enriched in functions like immune system and small molecule metabolic processes. KEGG enrichment analysis demonstrated that these DEGs were significantly enriched in the pathways like Toll-like receptor signaling pathway as well as metabolic pathways. Furthermore, \textit{IRF7} was predicted to be targeted by miR-412 and \textit{PPARGC1A} could be targeted by miR-199B.

Immune system protects host body against disease. Aberration of immune system could lead to inflammatory diseases, autoimmune diseases and cancer (17). Importantly, ccRCC has shown extensive responses to immune checkpoint blockade therapies (18). Based on an in-depth immune profiling study, Chevrier et al. (19) identified that \textit{CD38} and \textit{CD204} involved in the tumor microenvironment modulate cancer progression. Previous study has presented that orchestration of immune checkpoints has prognostic value in ccRCC (20), but detailed mechanism of the checkpoints in ccRCC is yet unclear.

\textit{TYROBP} is the encoding gene of a transmembrane signaling polypeptide, which possesses an immunoreceptor tyrosine-based activation motif in its cytoplasmic domain (21). Previous study reported that TYROBP associated with the killer cell immunoglobulin-like receptor family and it served as an activating signal transduction element in cells (22). \textit{TYROBP} is reported to be related to Alzheimer’s pathology (21). Deficiency of \textit{TYROBP} is neuroprotective in a mouse model with early Alzheimer’s pathology. Meanwhile, the bone remodeling and brain function also depend on the integrity of \textit{TYROBP} signal (23). In addition, it has been documented that TYROBP presents a general function in inflammatory responses in microglia via the Jun NH2-terminal kinase (JNK)
signaling pathway (24). However, very rare studies focus on TYROBP correlation with ccRCC. In the current study, functional enrichment analysis showed that TYROBP was significantly enriched in the BP of immune system, indicating that TYROBP might also play critical role in the pathogenic inflammatory response of ccRCC. However, further analysis of TYROBP is still required to reveal the exact mechanism involved in pathogenesis of ccRCC.

IRF7, encoded interferon regulatory factor, is a member of type I interferon regulatory transcription factor family, which plays a critical role in the innate immunity (25). IRF7 silencing promotes bone metastasis of breast cancer via immune escape way (26). Coit et al. (27) have documented that IRF7 plays critical role in renal involved in lupus via regulating demethylation of DNA in naive CD4+ T cells. It was also reported that IRF7 can serve as a therapeutic target against renal tissue damage caused by bacterial infection (28). In the current study, functional enrichment analysis revealed that IRF7 was significantly enriched in the Toll-like receptor signaling pathway and immune response biological process. Toll-like receptor signaling pathway has been reported to play crucial roles in the inflammation, infection and cancer (29). Moreover, IRF7 was predicted to be targeted by miR-412 in ccRCC. miR-412 is a mature type of miR-142, contributing to important functions in inflammatory and immune response among different diseases (30). Taken together, we speculated that miR-412 might present a critical role in the immune response of ccRCC via targeting IRF7 in Toll-like receptor signaling pathway.

PPARGCA1 is the encoding gene of PPARG co-activator 1α, which is a co-regulator of mitochondrial biogenesis and oxidative phosphorylation. Cho et al. (31) have documented that genetic variation of PPARGCA1 regulates diet-associated inflammation in colorectal cancer. Moreover, polymorphisms of PPARGCA1 and lower expression of PPARGCA1 are risk factors in the pathogenesis of non-alcoholic fatty liver disease, characterized by abnormal inflammatory response during etiology (32). In the current study, down-regulation of PPARGCA1 was identified in ccRCC samples and functional enrichment showed that PPARGCA1 was significantly enriched in the immune response biological process. Further analysis showed that RRARGCA1 was regulated by miR-199b. miR-199b is down-regulated in several solid tumors (33). Hou et al. (34) have documented that down-regulation of miR-199b strongly correlated with poor survival of HCC patients. In addition, miR-199b associated with poor survival of ovarian cancer patients, and loss of miR-199b results in hypermethylation in ovarian cancer (35). These evidences indicated that miR-199b might contribute to the pathogenesis of ccRCC via regulating PPARGCA1 expression.

Conclusion

Abnormal immune response might be an important mechanism of ccRCC. TYROBP, IRF7 and PPARGCA1 might play important roles in ccRCC via taking part in the immune system. Furthermore, miR-199b and miR-412 might serve critical roles in the regulation of immune response via targeting PPRARGCA1 and IRF7, respectively. Considering these findings, miR-199b and miR-412 might be used as the prognostic target for ccRCC gene therapy.

Acknowledgements

There is no financial support and conflict of interest in this study.

Authors’ Contributions

Y.W., H.W.; Contributed to the design of research, acquisition of data, analysis and interpretation of data, statistical analysis and drafting the manuscript. L.S., L.X., J.B.; Contributed to the analysis and interpretation of data and statistical analysis. J.L.; Contributed to the design of research. All authors read and approved the final manuscript.

References


