Network spatio-temporal analysis predicts disease stage-related genes and pathways in renal cell carcinoma

X.H. Li¹, C.Z. Yang² and J. Wang³

¹Department of Internal Medicine, Ji’nan Central Hospital, Ji’nan, China
²Department of Clinical Laboratory,
Second Affiliated Hospital of Mudanjiang Medical College, Mudanjiang, China
³Department of Nephrology,
Second Affiliated Hospital of Mudanjiang Medical College, Mudanjiang, China

Corresponding author: J. Wang
E-mail: wangjingwenzhang@163.com

Received November 17, 2015
Accepted January 15, 2016
Published May 6, 2016
DOI http://dx.doi.org/10.4238/gmr.15028061

ABSTRACT. The purpose of this study was to screen the key genes and pathways of renal cell carcinoma (RCC) and lay the foundation for its diagnosis and therapy. Microarray data of normal subjects and RCC patients at different stages of disease were used to screen differentially expressed genes (DEGs). Based on the DEGs in the four disease stages, four co-expression networks were constructed using the Empirical Bayes method and hub genes were obtained by centrality analysis. The enriched pathways of the DEGs and the mutual hub genes in the cluster of each disease stage were investigated. The mutual hub genes of the four disease stages in RCC tissue were validated using reverse transcription-polymerase chain reaction (RT-PCR) and western blot analysis. A total of 432 DEGs were screened, including 233 upregulated and 199 downregulated genes, by statistical analysis. Centrality analysis of co-expression networks in different disease stages suggested that PLXDC1, IKZF1, RUNX2, and RNF125 were mutual
hub genes. Pathway analysis showed that the DEGs were significantly enriched in seven terms. The hub modules in stage I disease were significantly enriched in the complement coagulation cascade pathway and the hub modules of the other three disease stages were enriched in natural killer cell-mediated cytotoxicity. The expression levels of PLXDC1, IKZF1, RUNX2, and RNF125 were significantly different between normal subjects and RCC patients by RT-PCR and western blot. Our study revealed four hub genes (PLXDC1, IKZF1, RUNX2, and RNF125) and two biological pathways that might be underlying biomarkers involved in RCC.

Key words: Renal cell carcinoma; Hub genes; Western blot; RT-PCR; Pathway-enrichment analysis

INTRODUCTION

Renal cell carcinoma (RCC) is a kidney cancer that originates in the renal parenchyma (Zhai et al., 2012). It is the second most common genitourinary tumor, accounting for approximately 3% of all adult malignancies (Siegel et al., 2012); global incidence and mortality rates are rising at a rate of approximately 2-3% every decade (Wei et al., 2013). Owing to a lack of early diagnostic biochemical indicators and the cancer’s resistance to radiotherapy and chemotherapy (Zhu et al., 2012), early diagnosis and postoperative adjuvant therapy of RCC are hindered. Approximately, 30% of RCC patients are diagnosed with metastatic disease at first diagnosis, and postoperative recurrence rates of localized RCC is as high as 30% (Motzer et al., 1996; Cohen and McGovern, 2005). Although the new treatment programs have been improved, the prognosis of RCC is still poor, especially in later tumor stages (Jiang et al., 2006; Suh et al., 2009).

The use of high-throughput experimental techniques, such as microarray, has increased in recent years. Microarray technology has been used to discover the diagnostic gene signatures of diseases (Liu et al., 2013). Complex diseases are usually characterized by diverse etiology, activation of multiple-signal transduction pathways, and various gene mutations. Thus, a robust genetic marker will be beneficial to the diagnosis and targeted treatment of complex diseases in clinical practice. Although human tumors have been profiled extensively by genomics-based studies (Dalgliesh et al., 2010; Morris et al., 2011), little is known about how they form a network that contributes to aggressive disease and poor outcome. A systematic analysis of the network and pathways in which these genes interact may enable us to observe a set of alterations in a more accurate perspective.

Therefore, in order to better understand the complex pathology and identify the molecular networks involved in RCC, we adopted a systems biology approach to acquire and analyze changes in gene expression between samples from RCC patients and normal controls. Herein, we examined the differentially expressed genes (DEGs) between RCC and normal controls and the DEGs in four stages of the disease group. Then, the co-expression networks of the four disease stages were constructed and the hub genes were identified by centrality analysis. In addition, the related metabolic pathways were identified. Finally, several hub genes were validated by reverse transcription-polymerase chain reaction (RT-PCR) and western blot analysis. These results may provide significant information for the understanding of RCC.
MATERIAL AND METHODS

Data recruitment and preprocessing

A search on the ArrayExpress database (http://www.ebi.ac.uk/arrayexpress/) to research gene expression microarray data between normal subjects and RCC patients was conducted. Three microarray datasets were extracted under accession numbers E-GEOD-53757 (von Roemeling et al., 2014), E-GEOD-36895 (Peña-Llopis et al., 2012), and E-GEOD-6344 (Gumz et al., 2007). Each dataset contained expression data for all four disease stages of RCC.

Before the analysis, the original expression information from all conditions was preprocessed. For each dataset, background correction and normalization were performed to eliminate the influence of nonspecific hybridization via the robust multichip average (RMA) method (Ma et al., 2006) and quantile-based algorithm (Rifai and Ridker, 2001), respectively. Perfect match and mismatch values were revised by Micro Array Suite 5.0 (MAS 5.0) algorithm (Pepper et al., 2007), and the expression value was selected using the median polish. Then, the data were screened by the feature filter function of gene filter package (http://bioconductor.org/packages/genefilter/). Each probe was mapped to one gene by getSYMBOL, and the probe was discarded if it did not match any.

The three filtered expression datasets were merged and calculated using the inSilicoMerging package (InSilico Genomics S.A., Brussels, Belgium) (Taminau et al., 2012). The data distribution was visually estimated using Batch Mean-centering (BMC), a merged data method after adaptation according to Support Vector Machines (SVM).

Detection of DEGs

DEGs refer to genes with different expression profiles between the cases and controls. If one gene has a high (or low) expression in the case and it is the opposite in the normal group, the gene may be related to the occurrence of the disease, and merits further research and analysis.

In this study, DEGs between the four disease stages and normal subjects were screened by the LIMMA package (Smyth et al., 2005). t-test and F-test were performed on the matrix, and then the P values were transformed to -log_{10}P. Empirical Bayes (Datta et al., 2004) (eBayes or EB) statistics and a false-discovery rate (FDR) (Reiner et al., 2003) calibration of P values (£0.05) for the data were conducted by lmFit function. The genes were extracted from the linear model after inspection, which needed to meet the following conditions: |logFC| ≥ 2, P values <0.05 (P < 0.05).

Construction of co-expression networks of the four disease stages by using EB

It is critical to construct a co-expression network for identifying modules and the intra-modular connectivity. The EB approach (Dawson and Kendziorski, 2012) has proven to be a useful complement to the existing differential expression methods by simulations and case studies. In this section, we used this method to achieve the co-expression relationships between co-expressed gene pairs of the four disease stages.

Differently co-expressed (DC) gene pairs were processed using the following method. The expression values in an m-by-n matrix of X (where m indicated the number of genes/probes
under consideration, and $n$ indicated the total number of microarrays over all conditions) were generally represented on the log2 scale. The members of the conditional array with length $n$ took values in $1, \ldots, K$ ($K$ indicated the total number of conditions). They were then used to define the equal DC classes based on the unique values in the conditions array. Intra-group correlations for all $p = m^* (m - 1) / 2$ gene pairs from $X$ and the condition array were calculated using bi-weight mid-correlation. The $p$-by-$K$ of D matrix with correlations was obtained. MakeMyD function (Fraley and Raftery, 2002) was used to initialize the hyper-parameters to find the component normal mixture model, which could best fit the empirical distribution of correlations. The values of the component in a normal mixture model with component means, standard deviations, and weights were used to initialize the expectation-maximization (EM) algorithm. Finally, DC genes were distinguished from gene pairs having invariant expression by controlling the posterior expected FDR at 0.05 and the co-expression network of the four disease stages was constructed to represent the correlation between each pair of genes.

**Centricity analysis of the co-expression network of each disease stage**

To further obtain the hub genes in each disease stage of RCC, we had a centricity analysis based on the degree of nodes in different networks (Scardoni and Laudanna, 2012). The degree is the simplest topological index and is the equivalent of the number of nodes directly adjacent to a given node $v$ (indicates the degree of the vertex). Calculation of the degree allows determining the “degree distribution”, $P(k)$, which gives the probability that a selected node has exactly $k$ links. $P(k)$ is obtained by counting the number of nodes $N(k)$ with $k = 1, 2, 3, \ldots$ links and dividing by the total number of nodes $N$. Nodes with a high degree (highly connected) are called “hubs”. At last, we selected the mutual hub genes in all disease stages for subsequent research.

**Module analysis**

In order to explore the interconnected areas and the function of the network, we had a module analysis using the MCODE method (Xu et al., 2002) for the co-expression network. The module can more clearly reflect the genes involved in biological processes. In order to identify the network module, we made a series of important basic parameters including node density cutoff = 1, node score cutoff = 0.2, K-core = 2, and max. depth = 100. Meanwhile, the modules which comprised the mutual hub genes were considered as hub modules and were selected for further pathway analysis.

**Pathway-enrichment analysis**

To further investigate the enriched pathways of the total DEGs and hub modules in the four disease stages, a pathway analysis was performed based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (www.genome.jp/kegg/), a widely used and comprehensive resource for pathway mapping of genes. We imported the DEGs and the genes of hub modules in the four disease stages to the online tool of Database for Annotation, Visualization and Integrated Discovery (Huang et al., 2009) (DAVID, http://david.abcc.ncifcrf.gov/tools.jsp), and obtained all pathways in which these genes were enriched. We obtained the enrichment pathways according to the threshold of P value <0.05.
Validation by RT-PCR

RT-PCR was used to validate the selected mutual hub genes, which had the highest degree in the network, at the four disease stages. RT-PCR was performed as follows:

RNA extraction and cDNA synthesis

Total RNA was extracted from the tumor tissue of 10 kidney cancer patients and the normal tissues 2 cm away from the tumor by using Trizol RNA kit (Invitrogen, Carlsbad, CA, USA). cDNA was synthesized using the following reaction: 4.0 µL RNA and 3 µL Oligo(dT) 18 were mixed and placed in a 70°C water bath for 5 min degeneration, then immediately put on ice, mixed in 2 µL RNasin (40 U/µL), 5X reverse transcription buffer (8 µL), 8 µL dNTPs, and 2 µL AMV reverse transcriptase (5 U/µL) were added and the mixture was placed in a 37°C water bath for 60 min. AMV reverse transcriptase was inactivated (95°C for 5 min) and the mixture was stored at -20°C.

PCR amplification

The data were normalized to the cDNA template and the β-actin reference gene. The upstream and downstream primers are shown in Table 1. PCR reagents were mixed as follows: 10X PCR Buffer (10 µL), TaqDNA polymerase (1 µL of 5 U/µL), upstream and downstream primers (3 µL each), and dNTPs (8 µL). PCR conditions are shown in Table 2. We computed the mean and standard deviation of individual gene (PLXDC1, IKZF1, RUNX2, and RNF125) expression values in patient and normal samples.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer sequences (5'-3')</th>
<th>Length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward</td>
<td>Reverse</td>
<td></td>
</tr>
<tr>
<td>PLXDC1</td>
<td>CTCAGAATTCGGGAGCAGGTCACG</td>
<td>AGCAGCGCTTCAGGTTGTTCTGAA</td>
</tr>
<tr>
<td>IKZF1</td>
<td>CTTGTGCCTGCCGAGATCAGTCTT</td>
<td>GAAGGGCGGACGTCTTGTGCTTTC</td>
</tr>
<tr>
<td>RUNX2</td>
<td>TGCGAGAGGTACGAGATGGG</td>
<td>TGTAAACTCCGTTCACTCAGTCCTCA</td>
</tr>
<tr>
<td>RNF125</td>
<td>GTGACACCGGGTGTGGCTTC</td>
<td>TGGGCTACCGCTGTTGCTCGAG</td>
</tr>
<tr>
<td>β-actin</td>
<td>AAGTACTCCTGGTGATCAGGG</td>
<td>TCAAGTTGGGGGACAAAAAG</td>
</tr>
</tbody>
</table>

Validation by western blot

The samples were ground to a powder in liquid nitrogen. Then, the samples were treated with lysis buffer and centrifuged at 10,000 g for 20 min at 4°C. Protein concentration was determined by the Bradford protein assay (Bio-Rad, Hercules, CA, USA). We used 10 mg
protein for SDS-PAGE (12%) and transferred the protein (4°C, constant current 300 mA, 2 h) to a nitrocellulose membrane (NC). The NC was blocked in TBS-T containing 5% skimmed milk powder at 37°C for 2 h, following which a 1:10,000 dilution of rabbit anti-human IgG antibody (Sigma-Aldrich, St. Louis, MO, USA) was added and incubated at 37°C for 2 h. Unbound antibody was washed away with TBS-T (3 times) and the NC was incubated with horseradish peroxidase-labeled sheep anti-rabbit IgG secondary antibody (1:5000) at 37°C for 2 h. After washing with TBS-T, the NC was treated with chemiluminescence (Genview, Houston, TX, USA) for 3 min and exposed to film in a darkroom.

Data analysis

PCR products were analyzed by 1.5% agarose gel electrophoresis and images were obtained using the Bio-Rad gel imaging analyzer. Result analysis was performed using the gel imaging system Quantity One (Bio-Rad) and values are reported as the relative levels of the target gene and β-actin. Gray value of target bands by western blot analysis was done by Image J (NIH, USA) and values are reported as the relative levels of the target protein and GAPDH. Data were analyzed using the Student t-test comparison between groups (SPSS19.0, USA).

RESULTS

Detection of DEGs

In the present study, 4214 genes were screened using the intersection of the microarray datasets. In the conditions of the threshold for |logFC| > 1.5 and FDR < 0.05, we identified 432 genes that were consistently differentially expressed in RCC. Of these, 233 were upregulated and 199 were downregulated. The numbers of DEGs in stages I, II, III, and IV of disease were 417, 431, 429, and 408, respectively.

Construction of gene co-expression networks using EB

The EB approach was used to identify DC gene pairs based on the 417, 431, 429, and 408 identified DEGs in stages I, II, III, and IV of disease, respectively. A total of 93,096 protein pairs were produced and the relational values of all pairs were yielded after analyzing gene expression relationship using meta-analysis. The gene interaction network containing nodes and edges was constructed using the 93,096 protein pairs in our analysis.

Centrality analysis of the co-expression network

After performing centricity analysis on the co-expression network, we got the number of hub genes in stages I, II, III, and IV of disease, which were 42, 42, 41, and 41, respectively. Meanwhile, four mutual hub genes were found across all four disease stages: PLXDC1, IKZF1, RUNX2, and RNF125. These genes are likely crucial to maintain function and coherence of metabolic mechanisms.
Module analysis

After performing a cluster analysis for co-expression networks, we identified many modules. Among them, we found that 13 modules comprised the common hub genes in all four disease stages. They were 3 hub modules at stage I, 4 hub modules at stage II, 3 at stage III, and 3 at stage IV.

Pathway-enrichment analysis

Pathway analysis showed that the DEGs were significantly enriched in seven terms. Then, the pathways in which the hub modules of the four disease stages were enriched were mapped on to the total pathway. In stage I, the hub modules were significantly enriched in the complement coagulation cascade pathway. In stages II, III, and IV, the hub modules were enriched in the same pathway, which was natural killer cell-mediated cytotoxicity. The pathway results are shown in Table 3.

Table 3. Enriched pathways of total differentially expressed genes and hub modules of four periods.

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Total DEG-enriched pathways</th>
<th>Stage I</th>
<th>Stage II</th>
<th>Stage III</th>
<th>Stage IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nicotinate and nicotinamide metabolism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aldosterone-regulated sodium reabsorption</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complement and coagulation cascades</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Natural killer cell-mediated cytotoxicity</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Focal adhesion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toll-like receptor signaling pathway</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECM receptor interaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Validation by RT-PCR and western blot

We performed RT-PCR and western blot to validate the microarray data and examine the changes of each hub gene at the mRNA and protein levels. In the present study, the expression levels of four mutual hub genes (PLXDC1, IKZF1, RUNX2, and RNF125) were analyzed by RT-PCR and western blot in RCC patients. The results showed that the four mutual hub genes were significantly differentially expressed in RCC patients compared to those in normal subjects (Figures 1 and 2).

DISCUSSION

RCC is the sixth leading cause of cancer death and approximately one-third of patients present disease that is already metastatic and for which there is currently no adequate treatment (Perroud et al., 2006). Investigating biomarkers in complex diseases such as RCC will contribute to the understanding of the pathogenesis and diagnosis of the disease, potentially reducing its mortality and morbidity substantially. In this study, we identified 432 DEGs from RCC tissue and normal renal cortex, and obtained DEGs in different stages of RCC, then analyzed them by co-expression network, module analysis, and pathway-enrichment analysis. We found 4 mutual hub genes and several significant pathways associated with RCC. In the present study, four common hub genes (PLXDC1, IKZF1, RUNX2, and RNF125) associated with RCC were screened by network-based centricity analysis.
Figure 1. RT-PCR analysis of 4 hub genes from co-expression network. Lane M represents molecular marker, lane 1 is RCC, and lane 2 is normal control. A, B, C, and D, PLXDC1, IKZF1, RUNX2, and RNF125, respectively. All gels are shown on the left with their quantitation on the right. Significance of relative expression of one gene in RCC compared to normal controls is indicated by its P value (*P < 0.05).
Figure 2. Western blot analysis of 4 hub genes. G represents internal control, lane 1 is RCC, and lane 2 is normal control. A, B, C, and D, PLXDC1, IKZF1, RUNX2, and RNF125, respectively. All blots are shown on the left with their quantitation on the right. Significance of relative expression of one gene in RCC compared to normal controls is indicated by its P value (*P < 0.05).
Their functional annotations have been widely studied by several previous studies. *PLXDC1*, also known as tumor endothelial marker 7, is the most abundantly expressed cell surface marker in the vascular endothelium of human tumors (St. Croix et al., 2000). *PLXDC1* has been discovered as one of the genes involved in many types of human tumors, including prostate cancer (Schwarze et al., 2005) and invasive ovarian carcinoma (Lu et al., 2007). Bagley et al. (2011) suggested that it should be taken as a novel and attractive target for tumor antiangiogenic therapy. *IKZF1* (Ikaros family zinc finger protein 1) is a DNA-binding protein that displays crucial functions in the hematopoietic system, and has been proven a major tumor suppressor in several diseases, such as pediatric acute lymphoblastic leukemia (Kuiper et al., 2010), multiple myeloma (Krönke et al., 2014), and type 1 diabetes (Swafford et al., 2011). *RUNX2* and *RNF125* were also reported to play important roles in the progression and development of several tumors (Sugawara et al., 2011; Ardura et al., 2013; Yang et al., 2015). However, few studies have reported on the relationship between these hub genes and RCC. Therefore, this study focused on the RCC-related genes potentially contributing to its diagnosis and treatment.

Pathway analysis showed several significantly enriched pathways of total DEGs and hub modules of the four disease stages. When comparing these enrichment pathways, we found that the hub modules in stage I were significantly enriched in the complement coagulation cascade pathway and the hub modules in the other stages (II, III, and IV) were enriched in natural killer cell-mediated cytotoxicity. It has been proven that the metastatic potential of tumor cells correlated with natural killer cell-mediated cytotoxicity (Hanna, 1980). For example, it has been reported that natural killer cell cytotoxicity is related to multiple myeloma. Recent research highlights the fact that natural killer cells can affect lymphocytes by limiting or exacerbating immune responses (Vivier et al., 2008). Natural killer cells respond to the signals from the environment by producing effect molecules that can directly suppress tumor growth (Smyth et al., 2002). Only recently have we begun to appreciate the potential of natural killer cell-mediated cytotoxicity in the treatment of human RCC.

In conclusion, the present study screened key genes and pathways associated with RCC based on network spatio-temporal analysis and could be consider as a useful complement to the identification of early diagnostic biochemical indicators. Based on the research, the hub disease genes (*PLXDC1*, *IKZF1*, *RUNX2*, and *RNF125*) might be target genes for diagnosing RCC.

**Conflicts of interest**

The authors declare no conflict of interest.

**ACKNOWLEDGMENTS**

We would like to thank Ji’nan Evidence Based Medicine Science-Technology Center for their translation and polishing of our paper.

**REFERENCES**

Biomarkers in renal cell carcinoma


