Upregulated VEGFA and DLL4 act as potential prognostic genes for clear cell renal cell carcinoma

Xilong Wang
Jun Zhang
Yangyun Wang
Minqi Tu
Ying Wang
Guowei Shi
Department of Urology, The Fifth People’s Hospital of Shanghai, Fudan University, Shanghai, China

Purpose: As a typical hypervascular tumor, clear cell renal cell carcinoma (ccRCC) is the most common type of RCC. This study was aimed to explore the prognostic genes for ccRCC, focusing on the roles of vascular endothelial growth factor A (VEGFA) and Delta-like ligand 4 (DLL4) in the disease.

Materials and methods: The mRNA-sequencing data of kidney renal clear cell carcinoma (KIRC) were obtained from The Cancer Genome Atlas (TCGA) database, including 469 tumor samples and 68 adjacent normal samples. Using limma package, differentially expressed genes (DEGs) were analyzed by differential expression and subgroup analyses and confirmed using validation dataset GSE53757. Followed by enrichment analysis, protein–protein interaction (PPI) network analysis and protein subcellular localization were performed using multifaceted analysis tool for human transcriptome tool, and Cytoscape software and InnateDB database, respectively. Moreover, survival analysis was conducted to identify key prognosis-associated genes. In addition, VEGFA and DLL4 levels were detected using real-time quantitative PCR (qRT-PCR).

Results: A total of 1,984 DEGs were screened in the KIRC tumor samples. VEGFA was located in extracellular space and could interact with placental growth factor (PGF) and angiopoietin 2 (ANGPT2) in the PPI network. Subgroup analysis suggested that VEGFA was significantly upregulated in stages I, II, and III ccRCC tumor samples. Survival analysis showed that TIMP1 was among the top four prognosis-associated genes. qRT-PCR analysis confirmed that the expression levels of DLL4 and VEGFA were significantly upregulated in tumor samples.

Conclusion: VEGFA and DLL4 might be prognostic genes for ccRCC. Besides, PGF, ANGPT2, and TIMP1 might also be related to the prognosis of ccRCC patients.

Keywords: clear cell renal cell carcinoma, differentially expressed genes, protein–protein interaction network, subgroup analysis, survival analysis

Introduction
Renal cell carcinoma (RCC) is the most common type of kidney cancer and accounts for about 90%–95% of kidney cancer cases.1 It usually starts with weight loss, flank pain, blood in the urine, high blood pressure, fever, feeling unwell, and night sweats.2 RCC is responsible for 2%–3% of all malignancies in adult, ranking 7th and 9th in common cancer cases among men and women, respectively.3 Globally, RCC induces ~2,09,000 new cases and results in 1,02,000 deaths per year.4 When RCC is diagnosed, in ~30% of cases, the cells spread into the ipsilateral renal vein, and in 5%–10% of cases, it spreads to the inferior vena cava.5 RCC commonly metastasizes to the adrenal glands, lymph nodes, liver, lungs, brain, or bones, and targeted therapy can be used to improve the outcome of metastatic RCC.6 Clear cell renal cell carcinoma (ccRCC) is the most
common type of RCC, which is responsible for 75% of the cases. Therefore, investigating the mechanisms of ccRCC is important for improving therapies.

In patients with low-risk ccRCC, the expression of BRCA1-associated protein-1 (BAP1) can serve as a prognostic marker independently. Galectin-1 (GALI) regulates migration and invasion of ccRCC via the hypoxia-inducible factor-1α (HIF-1α)-mammalian target of rapamycin (mTOR) signaling axis; thus, GALI is a promising prognostic indicator and therapeutic target for the disease. Previous studies report that both the mRNA and protein levels of AT-rich interactive domain 1A (ARID1A) have statistical significance in predicting the prognosis of ccRCC. Vascular endothelial growth factor A (VEGFA), which is a critical angiogenic cytokine, plays an important role in tumor angiogenesis and may be a key target for cancer therapy. As an endothelial Notch ligand, delta-like ligand 4 (DLL4) plays a critical role in regulating tumor angiogenesis and thus is a potential antiangiogenesis target in clinical applications. RCC is a typical hypervascular tumor, and its tumor cells can promote tumor growth and progression through producing pro-angiogenesis factors. However, the prognostic genes involved in ccRCC have not been fully reported. Thus, this study was designed to investigate the genes associated with the prognosis of ccRCC, focusing on the roles of VEGFA and DLL4 in ccRCC.

In this study, the mRNA-sequencing (mRNA-seq) data of kidney renal clear cell carcinoma (KIRC) were obtained from The Cancer Genome Atlas (TCGA) database. Differentially expressed genes (DEGs) between tumor samples and adjacent normal samples were identified, and then enrichment analysis was performed. Followed by protein–protein interaction (PPI) network analysis, protein subcellular localization, subgroup analysis, and survival analysis were successively performed to explore the key genes involved in the prognosis of ccRCC. Finally, VEGFA and DLL4 expression levels were detected by real-time quantitative PCR (qRT-PCR).

Materials and methods

Date source

The level 3 mRNA-seq data (downloaded in November 2016) of KIRC were downloaded from TCGA (https://cancergenome.nih.gov) database, including 469 tumor samples and 68 adjacent normal samples. Meanwhile, clinical information including age, gender, race, tumor stage, survival time, and outcome were also obtained.

Data preprocessing and DEGs screening

Using the edgeR package (version 3.4, http://www.bioconductor.org/packages/release/bioc/html/edgeR.html) in R, the raw data were successively normalized into log continuous phase modulation (CPM) values, performed with linear modeling, and the relationship between average variances were mediated. Based on the empirical Bayes method in limma package (version 3.10.3, http://www.bioconductor.org/packages/2.9/bioc/html/limma.html), differential expression analysis was performed for the tumor samples and adjacent normal samples. The p-values obtained from t-test was adjusted into false discovery rates (FDRs; that were adjusted p-values) using Benjamini–Hochberg method. The genes with FDR <0.05 and |log₂ fold change (FC)| >2 were identified as DEGs.

Functional and pathway enrichment analysis

Gene Ontology (GO, http://www.geneontology.org) database can be applied for performing functional enrichment for genes and gene products. The Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/) database, which includes known genes and corresponding biochemical functionalities, can be used for pathway enrichment analysis. BioCloud is an online platform (http://www.biocloudservice.com) that was developed for computing high-throughput data. Using the multifaceted analysis tool for human transcriptome tool in the BioCloud platform, the DEGs were performed with functional and pathway enrichment analyses. The p-value <0.01 was set as the cutoff criterion.

PPI network analysis and protein subcellular localization

The intersection of the PPI pairs included in Mentha (http://mentha.uniroma2.it/about.php), BioGRID (https://wiki.thebiogrid.org), and HPRD (http://www.hprd.org) databases was taken as background, and then the DEGs were mapped into the background to obtain their PPI pairs. Subsequently, the PPI network for the DEGs was visualized by Cytoscape software (http://www.cytoscape.org). Using the CytoNCA plugin (version 2.1.6, http://apps.cytoscape.org/apps/cytonca) in Cytoscape software, betweenness centrality (BC), degree centrality (DC), and closeness centrality (CC) of the nodes in PPI network were analyzed to further identify hub nodes. In addition, the information of human protein subcellular localization were obtained from InnateDB database (http://www.innatedb.com), including extracellular space, cell surface, cytoplasm, plasma membrane, and nucleus. Afterward, protein subcellular localizations of the nodes in the PPI network were identified.

Subgroup analysis based on tumor stage

To further explore DEGs, KIRC tumor samples were divided into four groups (stages I, II, III, and IV) based on their tumor
stage. Then differential expression analysis between the tumor samples in each group and the adjacent normal samples was conducted by limma package,\(^18\) with FDR <0.05 and \(|\log_2 FC| >2\) as the thresholds.

**Differential expression analysis and subgroup analysis of validation dataset**

The raw CEL files and annotation files under GSE53757, which were sequenced on the platform of GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array, were obtained from Gene Expression Omnibus (GEO, [http://www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/)) database (downloaded in November 2016). GSE53757 included 72 ccRCC tumor samples and 72 adjacent normal samples. After the raw data of GSE53757 were read by the affy package (version 1.50.0, [http://www.bioconductor.org/packages/release/bioc/html/affy.html](http://www.bioconductor.org/packages/release/bioc/html/affy.html))\(^9\) in R, they were performed with background correction, normalization, and expression calculation using robust multi-array average (RMA) method.\(^10\) Based on the annotation files, the probes having no matched gene symbol were eliminated. For probes having the same gene symbol, their average value was taken as the final expression value of the gene. Finally, differential expression analysis and subgroup analysis were performed separately for the ccRCC tumor samples and adjacent normal samples using limma package\(^16\) with FDR <0.05 and \(|\log_2 FC| >2\) as the thresholds.

**Survival analysis**

To obtain prognosis-associated genes, the intersection DEGs between TCGA dataset and GSE53757 were identified. Then the intersection DEGs were divided into groups with high expression and low expression based on their medians. Using Cox model,\(^13\) variables such as age, gender, and tumor stage were adjusted. The genes with \(p\)-value <0.05 were taken as prognosis-associated genes. The hazard ratios (HRs) for survival were predicted for the prognosis-associated genes. In addition, high-expressed genes with HR >1 and low-expressed genes with HR <1 were identified, and then Kaplan–Meier (KM) survival curve was drawn.

**qRT-PCR analysis**

A total of 11 pairs of ccRCC tumor samples and adjacent normal samples were obtained from our hospital. This study was approved by the ethic committee of The Fifth People’s Hospital of Shanghai, Fudan University, and written informed consent was obtained from relevant patients. Total RNA was isolated from the samples with RNAiso Plus (TaKaRa, Tokyo, Japan) following the manufacturer’s manual. Followed by the concentration and quality of total RNA were detected using microplate reader (Tecan, Mannedorf, Switzerland). Subsequently, first-strand cDNA was synthesized using a reverse transcription kit (TaKaRa) and then stored at \(-20^\circ C\).

For qRT-PCR experiments, primers were designed and synthesized separately by Primer Premier 6.0 software (Premier Software Inc., Cherry Hill, NJ, USA) and Sangon Biotech Co., Ltd (Shanghai, China) (Table 1), respectively. The expression levels of VEGFA and DLL4 in ccRCC tumor samples and adjacent normal samples were measured using SYBR green kit (Thermo Fisher Scientific, Waltham, MA, USA). The 10 \(µL\) amplification system included 5 \(µL\) SYBR Premix EX Taq (2x), 3.4 \(µL\) cDNA template (100 ng/\(µL\)), 0.3 \(µL\) forward primer (10 \(µM\)), 0.3 \(µL\) reverse primer (10 \(µM\)), and 1 \(µL\) ddH\(_2\)O. The reaction program was as follows: 50°C for 3 min, 95°C for 3 min, and 95°C for 10 s, and 60°C for 30 s for 40 cycles. In addition, a melting curve was created in the end. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was taken as the reference gene, and all samples had three repeats.

**Statistical analysis**

Based on the 2-ΔΔCt method,\(^32\) the expression levels of VEGFA and DLL4 were analyzed. All results were shown as mean ± standard error of mean (SEM). GraphPad Prism (GraphPad Software, Inc., La Jolla, CA, USA) was used for statistical analysis and drawing pictures. The \(p<0.05\) was taken as the threshold for significant difference.

**Results**

**DEG screening**

A total of 1,984 DEGs were identified in the KIRC tumor samples in relative to adjacent normal samples, including 806 upregulated genes (including VEGFA and DLL4) and 1,178 downregulated genes. There were more downregulated genes than upregulated genes.

**Functional and pathway enrichment analysis**

Both the upregulated genes and downregulated genes were performed with enrichment analysis. The top 10 GO terms and pathways are shown in Figure 1A and B, respectively.

---

**Table 1** The primers used for quantitative real-time PCR analysis

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLL4 forward</td>
<td>GGGGCCAATCTATGCTTGTGA</td>
</tr>
<tr>
<td>DLL4 reverse</td>
<td>CACAGTAGGTCGCCGTGAAAT</td>
</tr>
<tr>
<td>VEGFA forward</td>
<td>CTGCTAATGGCCTGGAGGCC</td>
</tr>
<tr>
<td>VEGFA reverse</td>
<td>AGCCGGATCTGTTTTCGGC</td>
</tr>
<tr>
<td>GAPDH forward</td>
<td>TGACAACTTGGATCTGCGAGAGG</td>
</tr>
<tr>
<td>GAPDH reverse</td>
<td>AGGCAGGATGATTTCTGGAGAGG</td>
</tr>
</tbody>
</table>
Figure 1: The top 10 gene ontology (GO) terms (A) and pathways (B) enriched for the upregulated genes and the downregulated genes, respectively.

For the upregulated genes, the enriched GO terms mainly included immune response, regulation of immune response, and inflammatory response. Meanwhile, the pathways enriched for the upregulated genes mainly included *Staphylococcus aureus* infection, natural killer cell-mediated cytotoxicity, and cytokine–cytokine receptor interaction. Besides, the GO terms enriched for the downregulated genes mainly included excretion, ion transmembrane transport,
and chloride transmembrane transport. Moreover, the downregulated genes were enriched in pathways such as neuroactive ligand–receptor interaction, protein digestion and absorption, and tyrosine metabolism.

**PPI network analysis and protein subcellular localization**

A total of 729 nodes and 1,105 interactions were obtained from the intersection of the PPI pairs included in Mentha, BioGRID, and HPRD databases. Then the PPI network involving 337 upregulated genes and 392 downregulated genes was constructed (Figure 2). Combined with BC, CC, and DC scores, the top 10 nodes are listed in Table 2. According to the results of protein subcellular localization, 72, 33, 28, 68, and 91 nodes in the PPI network were located in extracellular space, cell surface, cytoplasm, plasma membrane, and nucleus, respectively. In particular, VEGFA could interact with placental growth factor (PGF) and angiopoietin 2 (ANGPT2) and was located in extracellular space.

**Subgroup analysis based on the tumor stage**

Differential expression analysis were performed for stage I versus normal, stage II versus normal, stage III versus normal, and stage IV versus normal comparison groups. There were 1,849 (700 upregulated genes and 1,149 downregulated genes), 2,006 (717 upregulated genes and 1,289 downregulated genes), 2,174 (974 upregulated genes and 1,227 downregulated genes), and 2,288 (1,007 upregulated genes and 1,281 downregulated genes) DEGs in stage I versus normal, stage II versus normal, stage III versus normal, and stage IV versus normal comparison groups, respectively. The Venn diagram of subgroup analysis (Figure 3) showed that the common DEGs among the four comparison groups were in majority. Importantly, VEGFA were differentially expressed in all the four comparison groups, and DLL4 had significantly differential expression in stage I versus normal, stage II versus normal, and stage III versus normal comparison groups (Figure 4).

**Figure 2** The protein–protein interaction network constructed for the differentially expressed genes. Circles and diamonds represent upregulated genes and downregulated genes, respectively. Green, blue, purple, brown, red, and gray represent nodes located in cell surface, cytoplasm, extracellular space, nucleus, plasma membrane, and unknown, respectively. The size of a node indicates its degree.
Table 2 The top 10 nodes in the protein–protein interaction network

<table>
<thead>
<tr>
<th>Gene</th>
<th>DC</th>
<th>Gene</th>
<th>BC</th>
<th>Gene</th>
<th>CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALB</td>
<td>27</td>
<td>KRT40</td>
<td>24</td>
<td>KRT40</td>
<td>24</td>
</tr>
<tr>
<td>KRT40</td>
<td>24</td>
<td>ALB</td>
<td>27</td>
<td>FASLG</td>
<td>14</td>
</tr>
<tr>
<td>CD247</td>
<td>22</td>
<td>TFAP2A</td>
<td>17</td>
<td>HCK</td>
<td>17</td>
</tr>
<tr>
<td>ZAP70</td>
<td>21</td>
<td>FASLG</td>
<td>14</td>
<td>ALB</td>
<td>27</td>
</tr>
<tr>
<td>HCK</td>
<td>17</td>
<td>ZAP70</td>
<td>21</td>
<td>TFAP2A</td>
<td>17</td>
</tr>
<tr>
<td>TFAP2A</td>
<td>17</td>
<td>PLG</td>
<td>16</td>
<td>BTK</td>
<td>13</td>
</tr>
<tr>
<td>VIM</td>
<td>16</td>
<td>WT1</td>
<td>8</td>
<td>LCP2</td>
<td>12</td>
</tr>
<tr>
<td>PLG</td>
<td>16</td>
<td>HCK</td>
<td>17</td>
<td>LAT</td>
<td>12</td>
</tr>
<tr>
<td>PLK1</td>
<td>15</td>
<td>CFTR</td>
<td>14</td>
<td>VAV1</td>
<td>12</td>
</tr>
<tr>
<td>CFTR</td>
<td>14</td>
<td>DLG2</td>
<td>11</td>
<td>ITK</td>
<td>8</td>
</tr>
</tbody>
</table>

Abbreviations: DC, degree centrality; BC, betweenness centrality; CC, closeness centrality.

Differential expression analysis and subgroup analysis of validation dataset

After the raw data of GSE53757 were preprocessed, a total of 398 DEGs were identified in ccRCC tumor samples compared with adjacent normal samples, including 147 upregulated genes (such as VEGFA) and 251 downregulated genes. Based on the tumor stage, ccRCC tumor samples were also divided into four groups (stages I, II, III, and IV). Then subgroup analysis was performed for the ccRCC tumor samples and adjacent normal samples, finding that VEGFA was significantly upregulated in stages I, II, and III ccRCC tumor samples (Figure 5).

Survival analysis

There were 270 intersection DEGs between TCGA dataset and GSE53757, including 103 upregulated genes and 167 downregulated genes. Then a total of 40 prognosis-associated genes were identified from the intersection DEGs. The KM survival curves for the top four prognosis-associated genes (according to p-values) (including aldehyde dehydrogenase 6 family, member A1, ALDH6A1; WD repeat domain 72, WDR72; phospholipase C-like 1, PLCL1; and TIMP metalloproteinase inhibitor 1, TIMP1) are shown in Figure 6.

qRT-PCR analysis

Furthermore, the expression levels of DLL4 and VEGFA in ccRCC tumor samples and adjacent normal samples were detected by using qRT-PCR. In tumor samples, the expression levels of DLL4 (p<0.001, Figure 7A) and VEGFA (p<0.01, Figure 7B) were significantly upregulated in relative to adjacent normal samples.

Discussion

In this study, a total of 1,984 DEGs (including upregulated VEGFA and DLL4) were identified in the KIRC tumor samples in relative to adjacent normal samples. Subgroup analysis for the KIRC tumor samples showed that VEGFA were differentially expressed in all the four comparison groups, and DLL4 had significantly differential expression in stage I versus normal, stage II versus normal, and stage III versus normal comparison groups. Meanwhile, subgroup analysis for the validation dataset showed that VEGFA was significantly upregulated in stages I, II, and III ccRCC tumor samples. In addition, qRT-PCR analysis confirmed that the expression levels of DLL4 and VEGFA were significantly upregulated in tumor samples.

Figure 3 The Venn diagram for upregulated (A) and downregulated genes (B) in stage I versus normal, stage II versus normal, stage III versus normal, and stage IV versus normal comparison groups.
In RCC, increased serum level of VEGF is correlated with adverse survival and can serve as a prognostic factor.\(^3\text{3}\) VEGF expression is associated with tumor stage and prognosis, suggesting that VEGF functions in the growth and progression of RCC.\(^3\text{4}\text{,3\text{5}}\) Huang et al reported that the activation of DLL4/Notch signaling and the interaction of endothelium and cells contribute to the progression of RCC.\(^3\text{6}\) DLL4 blockade has a promising antitumor activity in RCC patient-derived tumors, and the simultaneous targeting of the VEGF and DLL4 signaling pathways has a combination benefit.\(^3\text{7}\) Therefore, VEGFA and DLL4 might be prognostic genes for ccRCC.

Besides, VEGFA was located in extracellular space and had interactions with PGF and ANGPT2 in the PPI network. Angiogenic protein PGF belongs to the VEGF family, and anti-PGF antibody may be used as antiangiogenic agent that is minor toxicity when combined with anti-VEGF strategies.\(^3\text{8}\) Plasma levels of PGF have significant correlation with the clinical features and VEGF levels and thus can be used as an independent prognostic factor for RCC.\(^3\text{9}\) Plasma ANGPT2 concentration is increased in RCC patients, and ANGPT2 can serve as a promising target for the treatment of tyrosine kinase-resistant RCC.\(^4\text{0}\) Through targeting oncogenes ANGPT2 and neural precursor cell expressed, developmentally downregulated 9 (NEDD9), miR-145 acts as tumor suppressor and therapeutic target in patients with RCC.\(^4\text{1}\) By activating Tie2 receptor tyrosine kinase, ANGPT2 protects tumor endothelial cells and suppresses the antivascular effects of VEGF inhibition.\(^4\text{2}\) This indicates that PGF and ANGPT2 might be involved in the prognosis of ccRCC through interacting with VEGFA.

Moreover, 40 prognosis-associated genes were identified from the intersection DEGs between TCGA dataset and GSE5375, and TIMP1 was among the top four prognosis-associated genes. Through upregulating matrix metalloproteinase-2 (MMP-2) and MMP-9 and downregulating TIMP1, S-phase kinase-associated protein-2 (SKP2) signaling pathway contributes to the invasion and metastasis of RCC cells.\(^4\text{3}\text{,4\text{4}}\) Elevated protein levels of MMP2, MMP9, TIMP1, and TIMP2 are related to shortened patient survival and thus predict poor prognosis in RCC.\(^4\text{5}\) MMPs and TIMPs function in maintaining extracellular matrix homeostasis, and TIMP1 and TIMP2 levels are relevant in RCC.\(^4\text{6}\) By upregulating TIMP1 and TIMP2, Rac signaling inhibits the invasion of epithelial tumor cells in RCC patients.\(^4\text{7}\) Thus, TIMP1 might also be associated with the prognosis of ccRCC.

**Conclusion**

A series of bioinformatics analyses were carried out, finding a total of 1,984 DEGs in the KIRC tumor samples. Besides, VEGFA and DLL4 might be prognostic genes for ccRCC. Furthermore, PGF, ANGPT2, and TIMP1 might also be related to the prognosis of ccRCC patients. However, in-depth experimental research studies should be performed to confirm our findings.
Figure 6 The Kaplan–Meier (KM) survival curves for ALDH6A1 (A), WDR72 (B), PLCL1 (C), and TIMP1 (D).

Figure 7 The expression levels of DLL4 (A) and VEGFA (B) in clear cell renal cell carcinoma (ccRCC) tumor samples and adjacent normal samples. ***p<0.001; **p<0.01; *p<0.05.
Acknowledgments
This work was supported by Minhang District natural science research project (project number: 2014MHZ007) and the Shanghai Key Medical Specialty Program (project number: ZK2015B04).

Disclosure
The authors report no conflicts of interest in this work.

References


