

RESEARCH ARTICLE

Identification of biomarkers and construction of a microRNA-mRNA regulatory network for clear cell renal cell carcinoma using integrated bioinformatics analysis

Miaoru Han, Haifeng Yan, Kang Yang, Boya Fan, Panying Liu, Hongtao Yang *

Department of Nephrology, First Teaching Hospital of Tianjin University of Traditional Chinese Medicine, Tianjin, P. R. China

* tjcmht@126.com OPEN ACCESS

Citation: Han M, Yan H, Yang K, Fan B, Liu P, Yang H (2021) Identification of biomarkers and construction of a microRNA-mRNA regulatory network for clear cell renal cell carcinoma using integrated bioinformatics analysis. PLoS ONE 16(1): e0244394. <https://doi.org/10.1371/journal.pone.0244394>

Editor: Edwin Wang, University of Calgary, CANADA

Received: June 23, 2020

Accepted: December 7, 2020

Published: January 12, 2021

Copyright: © 2021 Han et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The chip data analyzed in this study are available in the [Supporting Information](#) files. These data are all from GeneExpression Omnibus (GEO) database <http://www.ncbi.nlm.nih.gov/geo/>. GSE16441 was downloaded from <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE16441>, and GSE66270 was downloaded from <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE66270>.

Abstract

With the recent research development, the importance of microRNAs (miRNAs) in renal clear cell carcinoma (CCRCC) has become widely known. The purpose of this study is to screen out the potential biomarkers of renal clear cell carcinoma (CCRCC) by microarray analysis. The miRNA chip (GSE16441) and mRNA chip (GSE66270) related to CCRCC were downloaded from the Gene Expression Omnibus (GEO) database. After data filtering and pretreating, R platform and a series of analysis tools (funrich3.1.3, string, Cytoscape_3.2.1, David, etc.) were used to analyze chip data and identify the specific and highly sensitive biomarkers. Finally, by constructing the miRNA-mRNA interaction network, it was determined that five miRNAs (*hsa-mir-199a-5p*, *hsa-mir-199b-5p*, *hsa-mir-532-3p* and *hsa-mir-429*) and two key genes (*ETS1* and *hapln1*) are significantly related to the overall survival rate of patients.

1 Introduction

Renal cell carcinoma is a common malignant tumor, which accounts for 90% of all renal tumors [1]. According to the characteristics of histopathology and molecular biology, renal cell carcinoma can be divided into four parts, including clear renal cell carcinoma, papillary renal cell carcinoma, chromophobe renal cell carcinoma and Bellini collecting duct carcinoma. Among them, 85% are clear renal cell carcinoma (CCRCC) [2]. At present, it is often diagnosed according to the clinical symptoms, imaging, renal biopsy, etc. However, many small renal mass (RMs) patients have no symptoms until the late stage of the disease, 30% of them have had distant metastasis when they are diagnosed as CCRCC [3]. Therefore, a sensitive and accurate diagnosis method of CCRCC is urgently needed. Recent studies have found that microRNAs (miRNAs) plays an important role in the pathogenesis of renal cancer [4,5]. Phosphatase and tensin homolog (*PTEN*) is one of the most extensively studied tumor suppressor genes related to tumor metastasis. Studies have shown that the expression of *miR-22* is down-regulated in the tumor tissue of CCRCC patients. In addition, in order to inhibit the growth,

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

migration and invasion of tumor cells [6]. *miR-22* may directly target *PTEN* in renal cell carcinoma. *MiR-21* is involved in the fine-tuning of von Hippel Lindau tumor suppressor (VHL) expression, and VHL tumor suppressor gene is involved in the regulation of hypoxia response through the regulation of hyperoxia inducible factors (HIFs) [7] VHL-HIF signaling pathways play an important role in the pathogenesis of CCRCC. In addition, *miR-155* [8], *miR-185* [9] play a key role in the prognosis of renal cell carcinoma. In this study, the common data sets of GSE16441 and GSE66270, the differential expression of miRNA chip and mRNA chip, and the functional enrichment and pathway enrichment of different genes were analyzed. Finally, the miRNA-mRNA regulatory network of CCRCC was constructed to identify and screen out the specific and highly sensitive prognosis biomarkers related to CCRCC. It include five miRNAs (*hsa-mir-199a-5p*, *hsa-mir-199b-5p*, *hsa-mir-532-3p* and *hsa-mir-429*) and two key genes (*ETS1* and *HAPLN1*). However, the regulatory network between miRNA and mRNA is very complex and needs to be studied further.

2 Materials and methods

2.1 MiRNA and mRNA expression profiles of CCRCC

The GeneExpression Omnibus (GEO) database <http://www.ncbi.nlm.nih.gov/geo/> served as an international public data archive, distributes high-throughput gene expression and other functional genomic datasets free of charge [10]. miRNA and mRNA micro-array datasets were downloaded from the GEO database under the accession number GSE16441 and GSE66270. The miRNA data set GSE16441 includes 18 normal control samples and 18 clear renal cell carcinoma samples (S1 Table). All samples were detected by the platform GPL8659 Agilent human miRNA microarray rel12.0. The gene expression profile GSE66270 includes 17 normal control samples and 17 clear renal cell carcinoma samples (S2 Table). The chip platform is GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array.

2.2 Data processing and differentially expressed RNAs

In <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE16441> download the series matrix file (s) and GPL8659 Agilent human miRNA microarray rel12.0 platform files. According to the corresponding platform annotation information, the samples were grouped by Perl software, and the probe ID was converted into gene symbols. Also in <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE66270> Download series matrix file (s) and GPL570 [HG-U133_Plus_2] Affymetrix human genome U133 Plus 2.0 array platform file, the samples were grouped in the same way, and the probe ID was converted into gene symbol. R/Bioconductor's "limma" package was used to identify the differentially expressed miRNAs (DEmiRs) in individual studies. Initially Benjamin Hochberg (BH) corrected p value was to be < 0.05 and $|\log_2 \text{fold change (FC)}|$ was set to be > 1 (S3 Table). Using the same method, set the adjusted value of Benjamin Hochberg (BH) was adjusted to be < 0.05 and $|\log_2 \text{fold change (FC)}|$ was adjusted to be > 2 to identify the differentially expressed mRNAs (DEGs) (S4 Table). Finally, the volcano map was drawn using R software.

2.3 Functional and pathway enrichment analysis of miRNA with differential expression

In order to further understand the biological function of DEmiRs, the gene ontology (GO) and the Kyoto Encyclopedia of genes and genomes (KEGG) pathway enrichment of DEmiRs were studied through `funrich3.1.3` [11]. GO enrichment analysis included "biological process" (BP),

"cellular component" (CC) and "molecular function" (MF). There was statistical significance in setting $P < 0.05$.

2.4 PPI network construction and module selection

Upload DEGs to search tool for retrieval of interacting genes (string) database [11] (<https://string-db.org/>), build PPI network, and set the confidence score ≥ 0.95 (S5 Table). Then import the network into Cytoscape_3.2.1 [12] software for visualization, using the Molecular Complex Detection (MCODE) plug-in to filter the DEGs module (degree cutoff = 2, k-core = 2, max. depth = 100 and node score cutoff = 0.2). Then according to the score value, the first four modules are selected for subsequent analysis. In order to further understand the biological function of DEGs in the module, the R / cluster Profiler [13] software package was used to perform Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis on DEmiRs. $P < 0.05$ was the threshold. R / ggplot2 [14] software package was used for visualization of GO enrichment analysis, and the first ten item parameters are limited for display. The R / goplot [15] software package was also used to visualize the enrichment analysis of KEGG pathway. The first six pathway entries were defined for display.

2.5 miRNA-mRNA Interactions analysis and visualization

Similarly, funRich3.1.3 was used to predict the target genes of DEmiRs, and we obtained a total of 2070 target genes (S6 Table), then we used Practical Extraction and Report Language (Perl v5.30.0) [16] to integrate the target genes, DEmiRs and DEGs data to identify the commonly Genes in the predicted target genes and DEGs. Finally, we obtained the regulatory network of miRNA target genes and DEGs (S7 Table). Using Cytoscape_3.2.1 software to visualize the regulatory network. This regulatory network includes a total of 69 nodes and 66 edges (Fig 4).

2.6 Survival analyses

Gene Expression Profiling Interactive Analysis (GEPIA, <http://gepia.cancer-pku.cn/index.html>) based on TCGA database [17] was used to conduct online analysis of DEGs in the regulatory network. Kaplan Meier survival chart was used to compare the overall survival of the two groups of patients. Group cutoff was set as median, consistency interval was 95%, and log rank p value < 0.05 was statistically significant. The central genes were identified from the predicted target genes for subsequent topological analysis, and degree was used as the screening condition for nodes with high connectivity.

3 Results

3.1 Identification of differentially expressed miRNAs and mRNAs in CCRCC

After searching the appropriate datasets in GEO database according to the eligibility criteria, two datasets (GSE16441 and GSE66270) regarding CCRCC were chosen as discovery sets. By setting the threshold of adjust $P < 0.05$ and $|\log_2 \text{FC}| > 1.0$. A result of 90 DEmiRs was obtained from GSE16441 data set, including 45 were up-regulated and 45 were down-regulated. (Fig 1A). By adopting the same method and setting threshold of adjust $P < 0.05$ and $|\log_2 \text{FC}| > 2.0$, and a total of 1309 DEGs were obtained from GSE66270 data set, of which 718 were up-regulated and 591 were down-regulated. (Fig 1B).

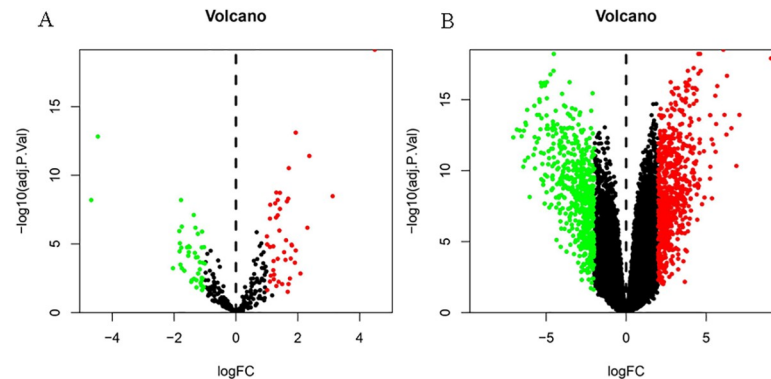


Fig 1. Volcano plot of DE miRNAs and DEGs. A, Volcano plot of differentially expressed miRNAs. The red dot represents upregulated miRNAs and the green dot represents downregulated miRNAs. B, Volcano plot of differentially expressed mRNAs. The red dot represents upregulated mRNAs and green dot represents downregulated mRNAs.

<https://doi.org/10.1371/journal.pone.0244394.g001>

3.2 Enrichment analysis of differentially expressed miRNAs

In order to study the biological function of DE miRNAs, GO and KEGG functional enrichment analysis were conducted. The results showed that these DE miRNAs were significantly enriched in various functional characteristics. GO analysis results showed that in terms of biological process, DE miRNAs are mainly enriched in cell communication, signal transduction and Regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolism (Fig 2A); that cellular

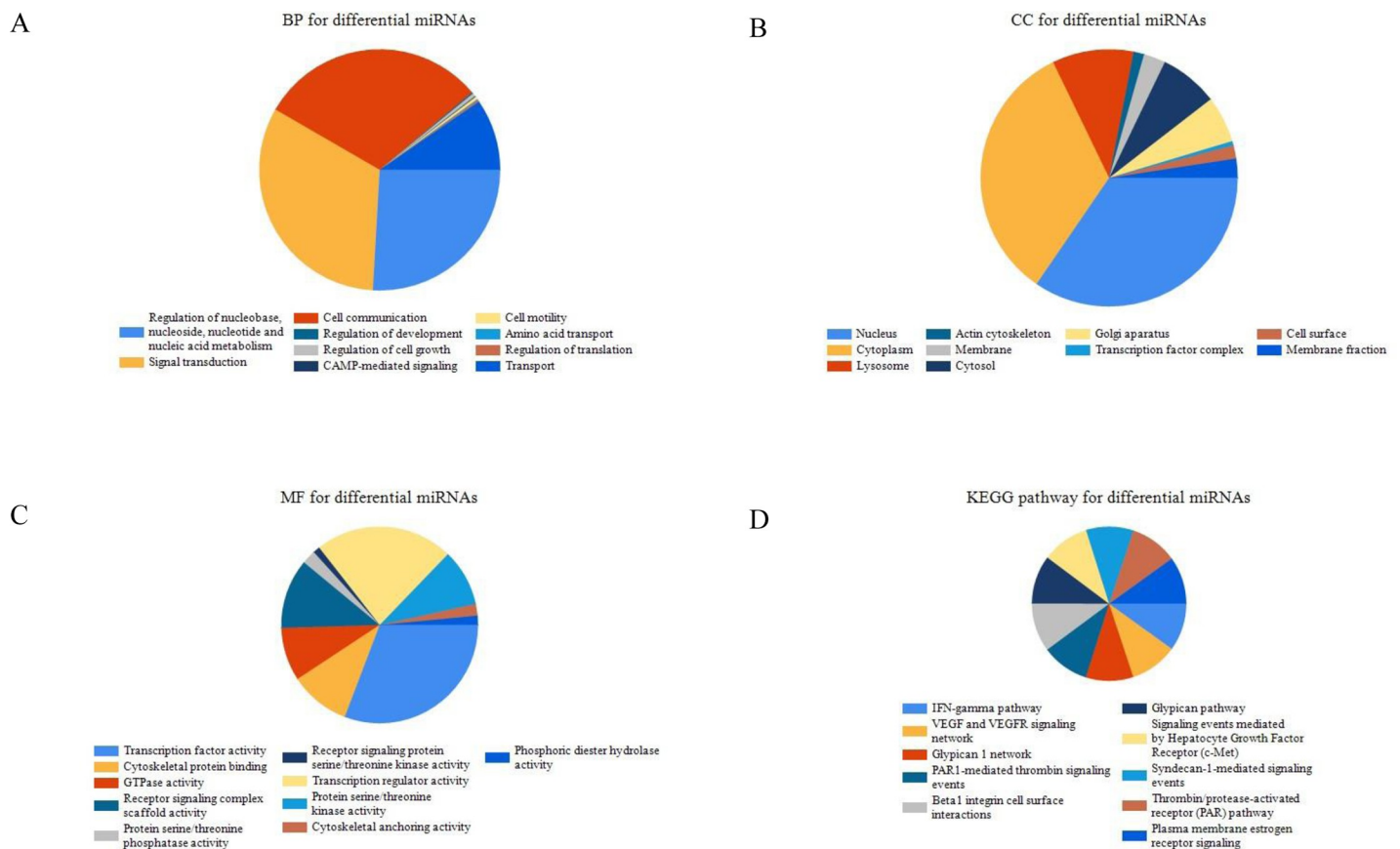


Fig 2. GO and KEGG functional enrichment analysis of DE miRNAs. A Biological Process for differential miRNA. B Cellular Component for differential miRNA. C Molecular Function for differential miRNA. KEGG pathway of differential miRNAs.

<https://doi.org/10.1371/journal.pone.0244394.g002>

components are mainly enriched in Cytoplasm, Nucleus Etc. (Fig 2B); that molecular functions are mainly enriched in Transcription regulator activity, Transcription factor activity, etc. (Fig 2C). KEGG functional enrichment analysis showed that DE miRNAs are more uniformly enriched in tumor-related pathways, such as Beta1 integrin cell surface interactions, Glypican pathway and VEGF and VEGFR signaling network (Fig 2D).

3.3 PPI network construction and module selection

Using string online tool to generate PPI network of DEGs, confidence Score ≥ 0.95 , the network was imported into the software of Cytoscape. The network included 336 nodes and 706 edges, including 247 up-regulated genes and 89 down-regulated genes (Fig 3). Then the cytochromeMCODE was used for further analysis. The MCODE analysis showed that there were 16 available modules in total, and the first four modules with the highest score were selected for analysis (Fig 4), including 47 genes in total. Then GO (Table 1 and Fig 5) and KEGG enrichment analysis (Table 2 and Fig 6) were carried out for the DEGs between these modules. GO enrichment analysis results showed that in the aspect of biological process, the DEGs in the module were mainly enriched in the aspects of nuclear division, sister chromosome segregation, mitotic nuclear division, etc; the cellular component was mainly enriched in the aspects of vacuole proton transporting V-type ATPase complex, proton transporting V-type ATPase complex, platelet alpha grain lumen, etc; molecular function was mainly enriched in the aspects of chemokine receptor binding, chemokine activity proton exporting ATPase activity, etc. KEGG pathway enrichment analysis results showed that DEGs is mainly enriched in those aspects, for example, collecting duct acid secret, chemokine signaling pathway, vital protein interaction with cytokine and cytokine receptor, etc.

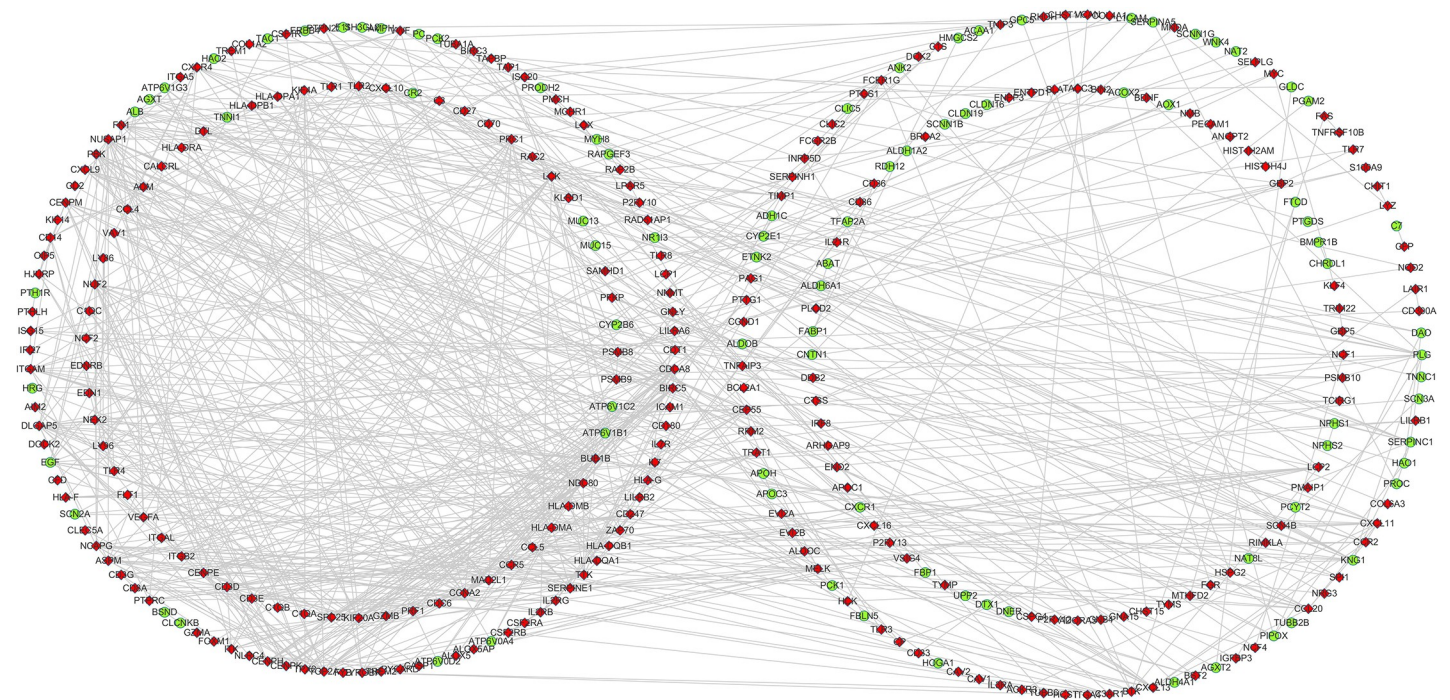


Fig 3. Protein-protein interaction regulatory network of DEGs in CCRCC. Nodes represent DEGs. Red nodes indicate upregulated DEGs and green nodes indicated downregulated DEGs. Edges stand for the regulatory association between any 2 nodes.

<https://doi.org/10.1371/journal.pone.0244394.g003>

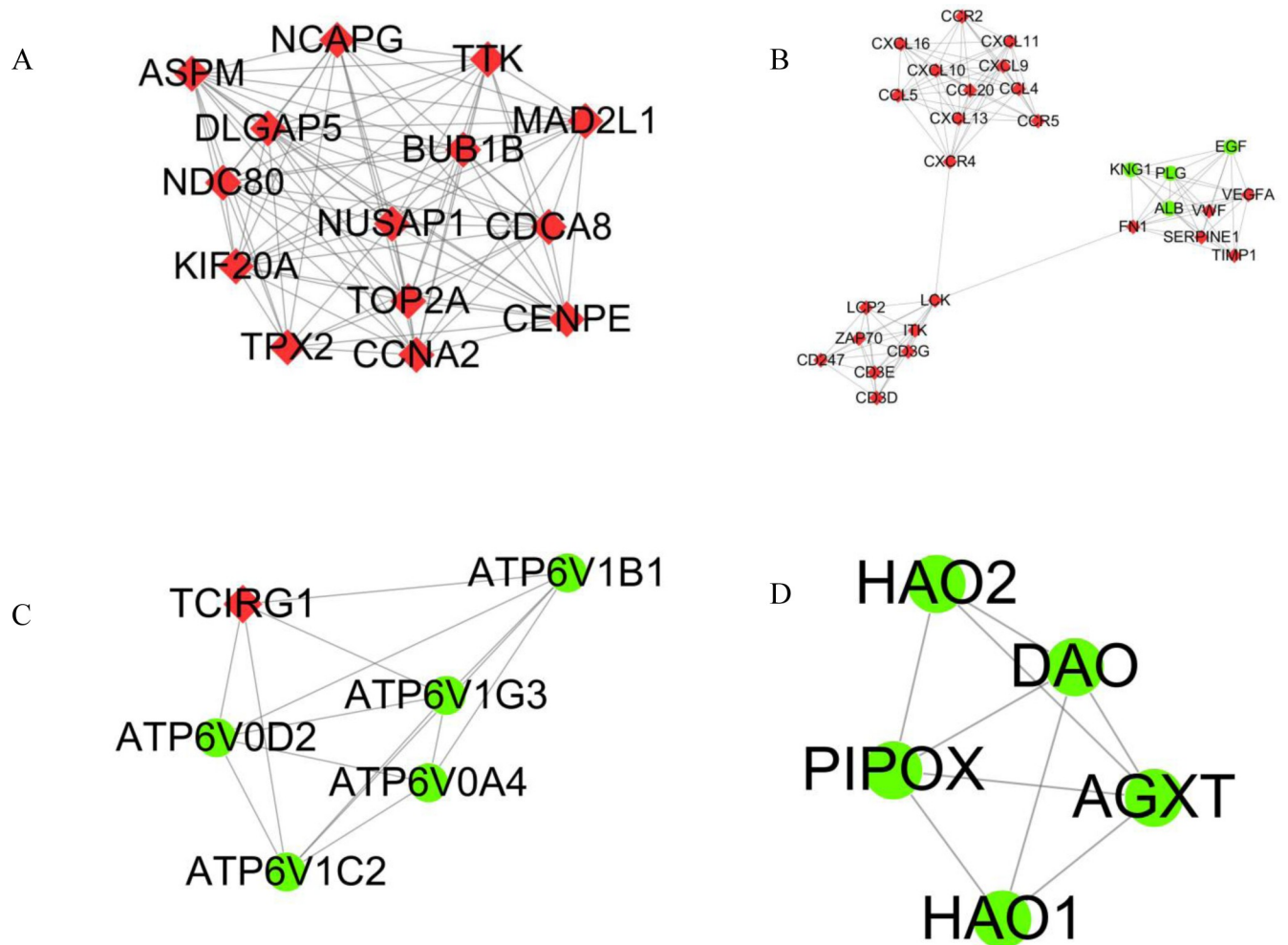


Fig 4. Top 4 modules in the protein-protein interaction network for DEGs. A Module 1. B Module 2. C Module 3. D Module 4. Nodes represent DEGs. Edges stand for the regulation association between any 2 nodes. Red and green nodes represent upregulated and downregulated genes.

<https://doi.org/10.1371/journal.pone.0244394.g004>

3.4 miRNA-mRNA interactions analysis and visualization

By using FunRich software, the target genes of DEmiRs was predicted and a total of 2070 target genes were obtained. The targeted regulatory relationship between DEmiRs and DEGS was obtained with Perl (V5.30.0). The subsequent steps involved uploading this control relationship to Cytoscape_3.2.1 software for visual processing, and constructing the control network of DEmiRs and DEGS (Fig 7). In this control network, a total of 69 nodes and 66 edges were included. The elliptical node represents DEmiRs, the triangular node represents DEGS, and the connection represents the targeted regulatory relationship between the two. Green represents miRNA or mRNA down-regulation in CCRCC samples, and red represents miRNA or mRNA up-regulation in CCRCC samples.

3.5 Survival analyses

In order to further verify the characteristics of DEGS in the network and its correlation strength with survival, Cox regression analysis, Kaplan-Meier curve and log-rank analysis were adopted to study more about the survival of DEGS in the network. The results showed

Table 1. Gene ontology analysis of DEGs in CCRCC.

ONTOLOGY	ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue
BP	GO:0000280	nuclear division	13/46	407/18670	1.10E-11	1.02E-08	6.65E-09
BP	GO:0000819	sister chromatid segregation	10/46	189/18670	2.65E-11	1.02E-08	6.65E-09
BP	GO:0140014	mitotic nuclear division	11/46	264/18670	3.16E-11	1.02E-08	6.65E-09
BP	GO:0048285	organelle fission	13/46	449/18670	3.75E-11	1.02E-08	6.65E-09
BP	GO:0070098	chemokine-mediated signaling pathway	8/46	88/18670	3.97E-11	1.02E-08	6.65E-09
BP	GO:0090383	phagosome acidification	6/46	28/18670	5.77E-11	1.23E-08	8.05E-09
BP	GO:1990868	response to chemokine	8/46	97/18670	8.78E-11	1.40E-08	9.19E-09
BP	GO:1990869	cellular response to chemokine	8/46	97/18670	8.78E-11	1.40E-08	9.19E-09
BP	GO:0000070	mitotic sister chromatid segregation	9/46	151/18670	9.94E-11	1.41E-08	9.26E-09
BP	GO:0050900	leukocyte migration	13/46	499/18670	1.38E-10	1.64E-08	1.08E-08
CC	GO:0016471	vacuolar proton-transporting V-type ATPase complex	6/46	17/19717	1.39E-12	1.84E-10	8.96E-11
CC	GO:0033176	proton-transporting V-type ATPase complex	6/46	26/19717	2.55E-11	1.69E-09	8.20E-10
CC	GO:0031093	platelet alpha granule lumen	7/46	67/19717	1.83E-10	8.04E-09	3.91E-09
CC	GO:0031091	platelet alpha granule	7/46	91/19717	1.63E-09	5.05E-08	2.46E-08
CC	GO:0016469	proton-transporting two-sector ATPase complex	6/46	51/19717	1.91E-09	5.05E-08	2.46E-08
CC	GO:0005819	spindle	10/46	347/19717	5.82E-09	1.28E-07	6.23E-08
CC	GO:0000775	chromosome, centromeric region	7/46	193/19717	2.99E-07	5.64E-06	2.75E-06
CC	GO:0033179	proton-transporting V-type ATPase, V0 domain	3/46	10/19717	1.41E-06	2.33E-05	1.13E-05
CC	GO:0072686	mitotic spindle	5/46	109/19717	5.39E-06	7.17E-05	3.49E-05
CC	GO:0005782	peroxisomal matrix	4/46	51/19717	5.98E-06	7.17E-05	3.49E-05
MF	GO:0042379	chemokine receptor binding	8/46	66/17697	5.63E-12	8.39E-10	4.86E-10
MF	GO:0008009	chemokine activity	7/46	49/17697	3.94E-11	2.93E-09	1.70E-09
MF	GO:0036442	proton-exporting ATPase activity	5/46	30/17697	1.29E-08	4.85E-07	2.81E-07
MF	GO:0045236	CXCR chemokine receptor binding	4/46	11/17697	1.30E-08	4.85E-07	2.81E-07
MF	GO:0046961	proton-transporting ATPase activity, rotational mechanism	4/46	22/17697	2.82E-07	8.41E-06	4.87E-06
MF	GO:0001664	G protein-coupled receptor binding	8/46	280/17697	5.51E-07	1.21E-05	7.01E-06
MF	GO:0005126	cytokine receptor binding	8/46	286/17697	6.46E-07	1.21E-05	7.01E-06
MF	GO:0019829	cation-transporting ATPase activity	5/46	66/17697	7.53E-07	1.21E-05	7.01E-06
MF	GO:0022853	active ion transmembrane transporter activity	5/46	67/17697	8.12E-07	1.21E-05	7.01E-06
MF	GO:0042625	ATPase coupled ion transmembrane transporter activity	5/46	67/17697	8.12E-07	1.21E-05	7.01E-06

<https://doi.org/10.1371/journal.pone.0244394.t001>

that a total of 19 DEGs were ultimately deemed to have significant differences in the survival status of CCRCC patients (Fig 8). These included 13 up-regulated DEGs (*ANLN*, *ARHGAP9*, *CDCA8*, *CENPK*, *ETS1*, *FAM49A*, *FAM57A*, *HAPLN1*, *KLF4*, *PMAIP1*, *PPP1R18*, *RASSF2*, *TGFBI*), and six down-regulated DEGs (*GPD1L*, *MAL*, *PRLR*, *PGS9BP*, *SLC39A5*, *TCF21*). Through network topology analysis, it was found that *ETS1* and *HAPLN1*'s degree are the highest, that *ETS1* is regulated by *hsa-miR-199a-5p*, *hsa-miR-199b-5p*, *hsa-miR-532-3p* and *hsa-miR-429*, and that *HAPLN1* is controlled by *hsa-miR-133b*, *hsa-miR-199a-5p* and *hsa-miR-199b-5p* regulation.

4 Results and discussion

Renal cell carcinoma is one of the most common malignant tumors in the urinary system, and renal clear cell carcinoma is the most common histological type. Early-stage renal cell carcinoma patients have a higher cure rate, with a 5-year survival rate of more than 90%. However, distant metastasis was found in many patients. Once it was found, and the 5-year survival rate has dropped to 12% [18]. Therefore, there is an urgent need to find out the sensitive early

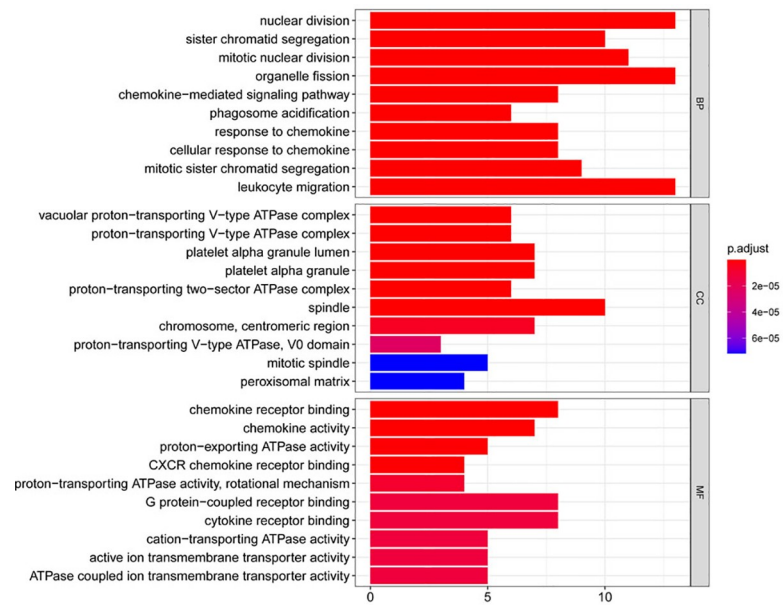


Fig 5. Gene ontology analysis of DEGs in CCRCC.

<https://doi.org/10.1371/journal.pone.0244394.g005>

biomarkers. Now RNA sequencing technology has been used to screen out general genetic changes in tumorigenesis and predict potential biomarkers for prognosis [19]. A single miRNA may be involved in the regulation of target mRNA in different carcinogenic or tumor suppressor pathways, hence, it is of great importance for excavating novel prognostic indicators of CCRCC to analyze the regulatory network of miRNA-mRNA interaction at the system level, which may also contribute to control malignant transformation of CCRCC.

In this study, the mRNA and miRNA microarrays of CCRCC patients were analyzed to assess the changes in genes and miRNAs of renal clear cell carcinoma, with a total of 90 DE miRs (45 up-regulated and 45 down-regulated) and 1309 DEGs (718 Up-regulation, 591 down-regulation). The study found that they have a clear difference in GO functional enrichment analysis after in-depth analysis of the two functional and pathway enrichment. In pathway enrichment analysis, DE miRs are mainly enriched in Beta1 integrin cell surface interactions, Glypican pathway and VEGF and VEGFR signaling network, DEGs are mainly enriched in collecting duct acid secretion, Chemokine signaling pathway, viral protein

Table 2. KEGG pathway analysis of DEGs in CCRCC.

ID	Description	GeneRatio	pvalue	p.adjust	qvalue	geneID	Count
hsa04966	Collecting duct acid secretion	6/37	1.72E-09	1.75E-07	1.25E-07	ATP6V1C2/ATP6V1B1/TCIRG1/ATP6V0A4/ATP6V0D2/ATP6V1G3	6
hsa04062	Chemokine signaling pathway	10/37	8.16E-09	4.16E-07	2.97E-07	CCL4/ITK/CXCL10/CXCL16/CXCL9/CXCR4/CXCL11/CCR2/CCL20/CXCL13	10
hsa04061	Viral protein interaction with cytokine and cytokine receptor	8/37	1.24E-08	4.21E-07	3.00E-07	CCL4/CXCL10/CXCL9/CXCR4/CXCL11/CCR2/CCL20/CXCL13	8
hsa05110	Vibrio cholerae infection	6/37	8.53E-08	2.18E-06	1.55E-06	ATP6V1C2/ATP6V1B1/TCIRG1/ATP6V0A4/ATP6V0D2/ATP6V1G3	6
hsa05323	Rheumatoid arthritis	7/37	1.71E-07	3.49E-06	2.49E-06	CCL20/ATP6V1C2/ATP6V1B1/TCIRG1/ATP6V0A4/ATP6V0D2/ATP6V1G3	7
hsa05120	Epithelial cell signaling in Helicobacter pylori infection	6/37	6.59E-07	1.12E-05	7.98E-06	ATP6V1C2/ATP6V1B1/TCIRG1/ATP6V0A4/ATP6V0D2/ATP6V1G3	6

<https://doi.org/10.1371/journal.pone.0244394.t002>

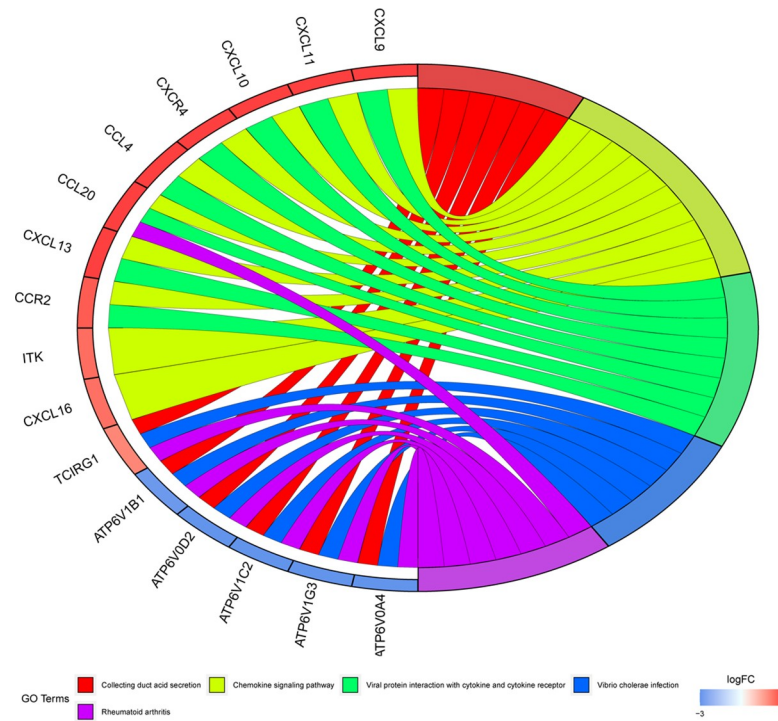


Fig 6. KEGG pathway analysis of DEGs in CCRCC.

<https://doi.org/10.1371/journal.pone.0244394.g006>

interaction with cytokine and cytokine receptor related pathways, integrin is a transmembrane heterodimer, mainly plays the role of cell adhesion, connecting extracellular matrix and intracellular actin skeleton [20]. Studies have shown that in tumor cells, vascular endothelial growth factor receptor (*VEGFR*) can interact with integrin, thereby activating integrin signal transduction [21]. Now, it is evident that the *VEGF-A* gene and angiogenic growth factors are involved in the development of CCRCC. Many researchers have also begun to focus on the biological research of *VEGFs* and *VEGFR1-3* and the intricate angiogenesis pathway [22]. Glypican and Glypican 1 network are also significantly enriched in pathways. Glypican 1 and other glypicans belong to the heparan sulfate proteoglycan family, which mainly regulates growth factors [23]. Studies have shown that in some cancers, they are highly expressed and are involved in the development of certain cancers. It is reported that Glypican is expected to become a new biomarker in the field of cancer [24]. Glypicans, as the core protein of glycocalyx, are involved in promoting integrin aggregation and growth factor signaling [25]. Therefore, it is guessed that the occurrence and development of CCRCC are closely related to the above pathways. In the pathway enrichment analysis of DEGs, the most significant pathway of enrichment is Chemokine signaling pathway, and chemokines belong to the large family of chemokines. So far, 48 chemokines have been identified. Chemokine CXC subfamily 13 (CXCL13), also known as B cell-attracting chemokine-1 (BCA-1), is a member of CXC-chemokine family. It has been shown that the tissue damage and hypoxic environment exacerbate the progression of prostate cancer, which is achieved by inducing the expression of CXC13 in tumor myofibroblasts [26]. Other studies also indicated that CXCL13 has a good diagnostic and prognostic value for renal clear cell carcinoma [27]. Therefore, chemokine signaling pathway also plays an important role in renal cell carcinoma, and its molecular mechanism needs further study.

In order to further screen the target gene, the target genes of differentially expressed miRNAs were predicted. We took the intersection between them the target regulatory relationship

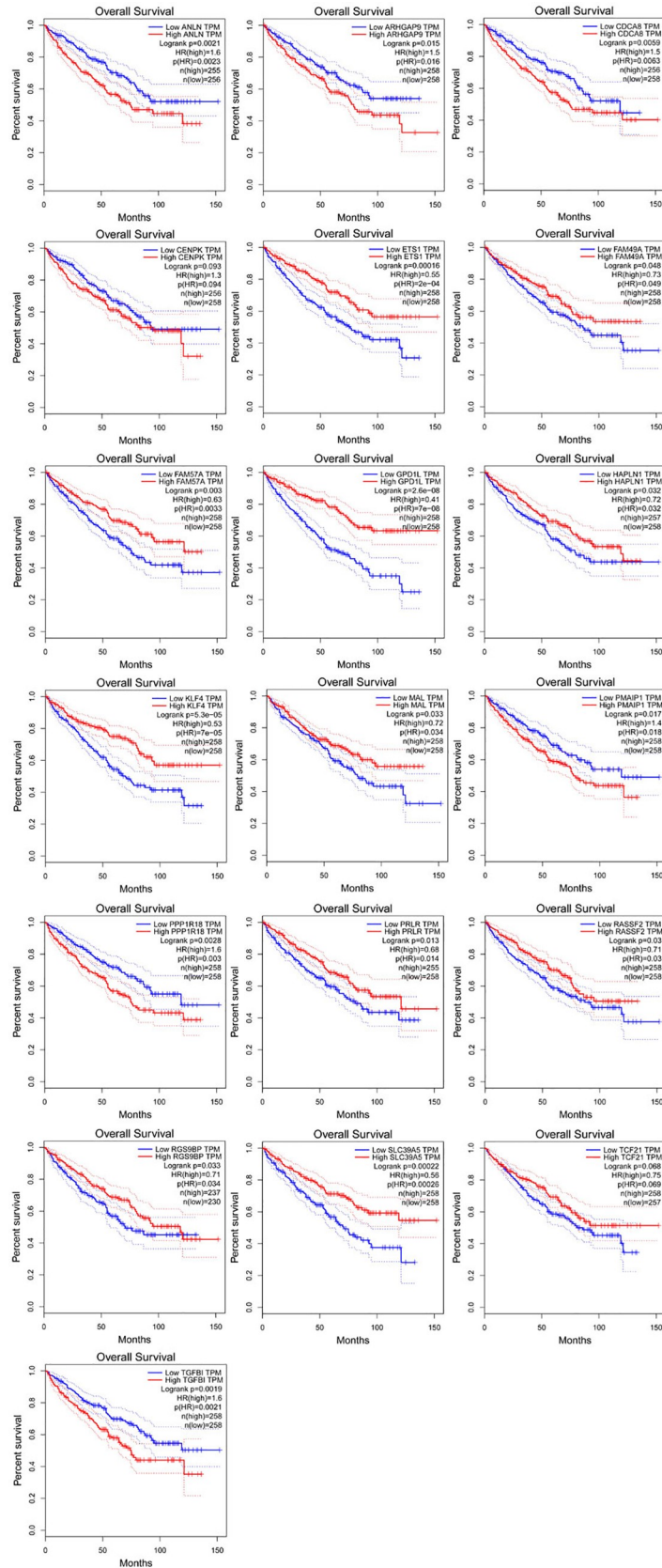


Fig 8. Nineteen differentially expressed mRNAs were associated with survival in CCRCC patients by using the Kaplan-Meier curve. A, ANLN; B, ARHGAP9; C, CDCA8; D, CENPK; E, ETS1; F, FAM49A; G, FAM57A; H, GPD1L; I, HAPLN1; J, KLF4; K, MAL; L, PMAIP1. M, PPP1R18; N, PRLR; O, RASSF2; P, RGS9BP; Q, SLC39A5; R, TCF21; S, TGFB.

<https://doi.org/10.1371/journal.pone.0244394.g008>

related to the overall survival rate of CCRCC. Through topology analysis, it was found that E26 transformation specificity-1 (*ETS1*) and Hyaluronan and proteoglycan link protein 1 (*HAPLN1*)'s degree are the highest, and these two core genes are highly expressed in CCRCC patients, besides, *ETS1* is affected by *hsa-miR-199a-5p*, *hsa-miR-199b-5p*, *hsa-miR-532-3p* and *hsa-miR-429* regulated, *HAPLN1* is regulated by *hsa-miR-133b*, *hsa-miR-199a-5p* and *hsa-miR-199b-5p* simultaneously. As we all know, miRNAs bind to the 3'-UTR of target genes to reduce the protein amount through inhibiting the translation process or promoting the degradation of corresponding mRNA [28]. *ETS1* is a member of ETS transcription factor family. Studies have shown that as an oncogene, *ETS1* is highly expressed in a variety of solid tumors, including lung cancer, colorectal and squamous cancer, breast cancer, ovarian cancer and cervical cancer, etc [29,30]. The high expression of *ETS1* is associated with a poor survival prognosis, which is consistent with our results. Previous studies have shown that *ETS1* can directly bind to a special region of the miR-532-5p promoter and inhibit its transcription. As an oncogene of renal cell carcinoma, *ETS1* is significantly associated with poor survival in large cohort specimens of renal cell carcinoma [31]. In this study, it was also found that *hsa-mir-199a-5p*, *hsa-mir-199b-5p*, *hsa-mir-532-3p* and *hsa-mir-429* regulate the expression of *ETS1* at the same time. Therefore, the effect of these four miRNA gene expressions on the expression of *ETS1* in renal cancer tissues and cell lines deserves further study. *HAPLN1* is another gene with the most significant difference, and its high expression is also significantly related to the survival prognosis of CCRCC patients. *HAPLN1* is negatively regulated by *hsa-miR-133b*, *hsa-miR-199a-5p* and *hsa-miR-199b-5p*. *HAPLN1*, known as cartilage connexin, is expressed in various tissues such as fetal cerebral cortex [32], fetal and adult heart tissue, etc [33]. Alla V et al. [34] found that overexpression of *HAPLN1* and its SP and IgV domains increased the tumorigenicity of mesothelioma. Sihem Mebarki et al. [35] found that knockout of the *HAPLN1* gene downregulated the expression of key markers of liver progenitor cells, such as *Snail*, *LGR5*, *type IV collagen*, and *α -SMA*. Through bioinformatics analysis, Zengzeng Wang et al. [36] also found that *hapln1*, *hsa-mir-204* and *hsa-mir-218* can be used as biomarkers of renal clear cell carcinoma, which further verified our prediction. In this study, the regulatory relationship and mechanism of *hapln1*, *hsa-mir-133b*, *hsa-mir-199a-5p* and *hsa-mir-199b-5p* in CCRCC need further study.

In recent years, more and more studies have shown that miRNA plays an important role in inducing complex human diseases [37,38]. Large scale experiments are expensive and time-consuming to explore miRNA disease association. Therefore, researchers try to construct complex networks on the basis of existing biological databases, establish the calculation model of potential miRNA disease association, screen out the most likely potential association, and then conduct small-scale verification through biological experiments, which plays an important role in the study of miRNA disease association. Shi Identified the potential association between miRNA and disease genes by using the association between miRNA targets and disease genes in the protein-protein interaction (PPI) network [39]. Chen [40] calculated the prediction score of miRNA disease pair by combining miRNA similarity and disease similarity, and proposed the induction matrix completion model (IMCMDA) for miRNA disease association prediction. By using Gaussian spectral kernel similarity method to deal with miRNA similarity and disease similarity, a miRNA disease association prediction (MDHGI) calculation model was proposed [41]. In 2019, Chen et al. [42] proposed a miRNA disease association prediction

(EDTMDA) calculation method based on decision tree integration. The three calculation models above can predict the potential relationship between miRNA and disease from different perspectives, and promote the understanding of the pathogenesis of complex human diseases, such as cancer. With the development of bioinformatics and sequencing technology, genomics, transcriptomics, proteomics, metabonomics and other omics data research continues to deepen, the integration of multi omics data to achieve efficient cancer biomarker recognition algorithm, find more valuable cancer biomarkers, and provide targeted experimental exploration for researchers will be one of the direction of future researches.

In conclusion, with method of bioinformatics, five miRNAs (*hsa-mir-199a-5p*, *hsa-mir-199b-5p*, *hsa-mir-532-3p* and *hsa-mir-429*) and two key genes (*ETS1* and *HAPLN1*) were found to be the potential biomarkers for the prognosis of patients with CCRCC. However, it is worth noting that the network of miRNA- mRNA interactions is complex, and more scientific exploration are needed to confirm our findings and verify their clinical application potential in improving the survival and prognosis of patients with CCRCC.

Supporting information

S1 Table. GSE16441-GPL8659_series_matrix.

(XLS)

S2 Table. GSE66270_series_matrix.

(XLS)

S3 Table. The differentially expressed miRNAs.

(XLS)

S4 Table. The differentially expressed mRNAs.

(XLS)

S5 Table. Protein-protein network (PPI) based on the DEmRNAs with a combined score was ≥ 0.95 .

(XLS)

S6 Table. The miRNA target genes predicted by Funrich 3.1.3.

(XLS)

S7 Table. The regulatory network of miRNA target genes and differentially expressed mRNAs (DEGS).

(XLS)

Acknowledgments

We appreciate all those who helped us in this research.

Author Contributions

Conceptualization: Kang Yang.

Data curation: Haifeng Yan.

Formal analysis: Kang Yang.

Investigation: Boya Fan, Panying Liu.

Methodology: Boya Fan.

Project administration: Hongtao Yang.

Resources: Hongtao Yang.

Visualization: Miaoru Han.

Writing – original draft: Miaoru Han.

Writing – review & editing: Miaoru Han.

References

1. Campbell S, Uzzo RG, Allaf ME, Bass EB, Cadeddu JA, Chang A, et al. Renal Mass and Localized Renal Cancer: AUA Guideline. *J Urol*. 2017; 198(3):520–9. <https://doi.org/10.1016/j.juro.2017.04.100> PMID: 28479239.
2. Hsieh JJ, Purdue MP, Signoretti S, Swanton C, Albiges L, Schmidinger M, et al. Renal cell carcinoma. *Nature reviews Disease primers*. 2017; 3:17009. Epub 2017/03/10. <https://doi.org/10.1038/nrdp.2017.9> PMID: 28276433; PubMed Central PMCID: PMC5936048.
3. Ljungberg B, Bensalah K, Canfield S, Dabestani S, Hofmann F, Hora M, et al. EAU guidelines on renal cell carcinoma: 2014 update. *Eur Urol*. 2015; 67(5):913–24. <https://doi.org/10.1016/j.eururo.2015.01.005> PMID: 25616710.
4. Chen X, Xie D, Zhao Q, You ZH. MicroRNAs and complex diseases: from experimental results to computational models. *Brief Bioinform*. 2019; 20(2):515–39. <https://doi.org/10.1093/bib/bbx130> PMID: 29045685.
5. He YH, Chen C, Shi Z. The biological roles and clinical implications of microRNAs in clear cell renal cell carcinoma. *Journal of cellular physiology*. 2018; 233(6):4458–65. Epub 2017/12/08. <https://doi.org/10.1002/jcp.26347> PMID: 29215721.
6. Fan W, Huang J, Xiao H, Liang Z. MicroRNA-22 is downregulated in clear cell renal cell carcinoma, and inhibits cell growth, migration and invasion by targeting PTEN. *Mol Med Rep*. 2016; 13(6):4800–6. <https://doi.org/10.3892/mmr.2016.5101> PMID: 27082730.
7. Fritz HKM, Lindgren D, Ljungberg B, Axelson H, Dahlback B. The miR(21/10b) ratio as a prognostic marker in clear cell renal cell carcinoma. *Eur J Cancer*. 2014; 50(10):1758–65. <https://doi.org/10.1016/j.ejca.2014.03.281> PMID: 24793999.
8. Shinmei S, Sakamoto N, Goto K, Sentani K, Anami K, Hayashi T, et al. MicroRNA-155 is a predictive marker for survival in patients with clear cell renal cell carcinoma. *Int J Urol*. 2013; 20(5):468–77. <https://doi.org/10.1111/j.1442-2042.2012.03182.x> PMID: 23050614.
9. Yuan HX, Zhang JP, Kong WT, Liu YJ, Lin ZM, Wang WP, et al. Elevated microRNA-185 is associated with high vascular endothelial growth factor receptor 2 expression levels and high microvessel density in clear cell renal cell carcinoma. *Tumour Biol*. 2014; 35(12):12757–63. <https://doi.org/10.1007/s13277-014-2602-9> PMID: 25217984.
10. Clough E, Barrett T. The Gene Expression Omnibus Database. *Methods Mol Biol*. 2016; 1418:93–110. https://doi.org/10.1007/978-1-4939-3578-9_5 PMID: 27008011; PubMed Central PMCID: PMC4944384.
11. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019; 47(D1):D607–D13. <https://doi.org/10.1093/nar/gky1131> PMID: 30476243; PubMed Central PMCID: PMC6323986.
12. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003; 13(11):2498–504. Epub 2003/11/05. <https://doi.org/10.1101/gr.1239303> PMID: 14597658; PubMed Central PMCID: PMC403769.
13. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012; 16(5):284–7. <https://doi.org/10.1089/omi.2011.0118> PMID: 22455463.
14. Villanueva Chen. ggplot2: Elegant Graphics for Data Analysis (2nd ed.). Measurement: Interdisciplinary Research and Perspectives. 2019; 17(3).
15. Wencke W, Fátima S-C, Mercedes R. GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics (Oxford, England)*. 2015; 31(17).
16. Yan XQ, Tu K, Xie L, Li YX, Yin B, Gong YH, et al. MiRE: a graphical R package for microRNA-related analysis. *Chinese medical sciences journal = Chung-kuo i hsueh k'o hsueh tsa chih*. 2008; 23(4):202–4. Epub 2009/02/03. [https://doi.org/10.1016/s1001-9294\(09\)60039-0](https://doi.org/10.1016/s1001-9294(09)60039-0) PMID: 19180879.

17. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 2017; 45(W1):W98–w102. Epub 2017/04/14. <https://doi.org/10.1093/nar/gkx247> PMID: 28407145; PubMed Central PMCID: PMC5570223.
18. Atkins MB, Tannir NM. Current and emerging therapies for first-line treatment of metastatic clear cell renal cell carcinoma. *Cancer Treat Rev.* 2018; 70:127–37. <https://doi.org/10.1016/j.ctrv.2018.07.009> PMID: 30173085.
19. Wang J, Dean DC, Hornicek FJ, Shi H, Duan Z. RNA sequencing (RNA-Seq) and its application in ovarian cancer. *Gynecologic oncology.* 2019; 152(1):194–201. Epub 2018/10/10. <https://doi.org/10.1016/j.ygyno.2018.10.002> PMID: 30297273.
20. Sun Z, Guo SS, Fassler R. Integrin-mediated mechanotransduction. *J Cell Biol.* 2016; 215(4):445–56. <https://doi.org/10.1083/jcb.201609037> PMID: 27872252; PubMed Central PMCID: PMC5119943.
21. Parveen A, Subedi L, Kim HW, Khan Z, Zahra Z, Farooqi MQ, et al. Phytochemicals Targeting VEGF and VEGF-Related Multifactors as Anticancer Therapy. *J Clin Med.* 2019; 8(3). <https://doi.org/10.3390/jcm8030350> PMID: 30871059; PubMed Central PMCID: PMC6462934.
22. Roskoski R Jr. Vascular endothelial growth factor (VEGF) and VEGF receptor inhibitors in the treatment of renal cell carcinomas. *Pharmacol Res.* 2017; 120:116–32. <https://doi.org/10.1016/j.phrs.2017.03.010> PMID: 28330784.
23. Wang S, Qiu Y, Bai B. The Expression, Regulation, and Biomarker Potential of Glypican-1 in Cancer. *Front Oncol.* 2019; 9:614. <https://doi.org/10.3389/fonc.2019.00614> PMID: 31355137; PubMed Central PMCID: PMC6640540.
24. Nagarajan A, Malvi P, Wajapeyee N. Heparan Sulfate and Heparan Sulfate Proteoglycans in Cancer Initiation and Progression. *Frontiers in endocrinology.* 2018; 9:483. Epub 2018/09/11. <https://doi.org/10.3389/fendo.2018.00483> PMID: 30197623; PubMed Central PMCID: PMC6118229.
25. Tarbell JM, Cancel LM. The glycocalyx and its significance in human medicine. *J Intern Med.* 2016; 280(1):97–113. <https://doi.org/10.1111/joim.12465> PMID: 26749537.
26. Ammirante M, Shalpour S, Kang Y, Jamieson CA, Karin M. Tissue injury and hypoxia promote malignant progression of prostate cancer by inducing CXCL13 expression in tumor myofibroblasts. *Proceedings of the National Academy of Sciences of the United States of America.* 2014; 111(41):14776–81. Epub 2014/10/01. <https://doi.org/10.1073/pnas.1416498111> PMID: 25267627; PubMed Central PMCID: PMC4205637.
27. Xu T, Ruan H, Song Z, Cao Q, Wang K, Bao L, et al. Identification of CXCL13 as a potential biomarker in clear cell renal cell carcinoma via comprehensive bioinformatics analysis. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie.* 2019; 118:109264. Epub 2019/08/08. <https://doi.org/10.1016/j.biopha.2019.109264> PMID: 31390578.
28. Fabian MR, Sonenberg N, Filipowicz W. Regulation of mRNA translation and stability by microRNAs. *Annual review of biochemistry.* 2010; 79:351–79. Epub 2010/06/11. <https://doi.org/10.1146/annurev-biochem-060308-103103> PMID: 20533884.
29. Dittmer J. The biology of the Ets1 proto-oncogene. *Molecular cancer.* 2003; 2:29. Epub 2003/09/16. <https://doi.org/10.1186/1476-4598-2-29> PMID: 12971829; PubMed Central PMCID: PMC194255.
30. Testoni M, Chung EY, Priebe V, Bertoni F. The transcription factor ETS1 in lymphomas: friend or foe? *Leuk Lymphoma.* 2015; 56(7):1975–80. <https://doi.org/10.3109/10428194.2014.981670> PMID: 25363344.
31. Zhai W, Ma J, Zhu R, Xu C, Zhang J, Chen Y, et al. MiR-532-5p suppresses renal cancer cell proliferation by disrupting the ETS1-mediated positive feedback loop with the KRAS-NAP1L1/P-ERK axis. *British journal of cancer.* 2018; 119(5):591–604. Epub 2018/08/08. <https://doi.org/10.1038/s41416-018-0196-5> PMID: 30082686; PubMed Central PMCID: PMC6162242.
32. Long KR, Newland B, Florio M, Kalebic N, Langen B, Kolterer A, et al. Extracellular Matrix Components HAPLN1, Lumican, and Collagen I Cause Hyaluronic Acid-Dependent Folding of the Developing Human Neocortex. *Neuron.* 2018; 99(4):702–19.e6. Epub 2018/08/07. <https://doi.org/10.1016/j.neuron.2018.07.013> PMID: 30078576.
33. Wirrig EE, Snarr BS, Chintalapudi MR, O'Neal JL, Phelps AL, Barth JL, et al. Cartilage link protein 1 (Crtl1), an extracellular matrix component playing an important role in heart development. *Dev Biol.* 2007; 310(2):291–303. <https://doi.org/10.1016/j.ydbio.2007.07.041> PMID: 17822691; PubMed Central PMCID: PMC2254939.
34. Ivanova AV, Goparaju CM, Ivanov SV, Nonaka D, Cruz C, Beck A, et al. Protumorigenic role of HAPLN1 and its IgV domain in malignant pleural mesothelioma. *Clin Cancer Res.* 2009; 15(8):2602–11. <https://doi.org/10.1158/1078-0432.CCR-08-2755> PMID: 19351750; PubMed Central PMCID: PMC3761224.
35. Mebarki S, Désert R, Sulpice L, Sicard M, Desille M, Canal F, et al. De novo HAPLN1 expression hallmarks Wnt-induced stem cell and fibrogenic networks leading to aggressive human hepatocellular

- carcinomas. *Oncotarget*. 2016; 7(26):39026–43. Epub 2016/10/26. <https://doi.org/10.18632/oncotarget.9346> PMID: 27191501; PubMed Central PMCID: PMC5129911.
36. Wang Z, Zhang Z, Zhang C, Xu Y. Identification of potential pathogenic biomarkers in clear cell renal cell carcinoma. *Oncology letters*. 2018; 15(6):8491–9. Epub 2018/05/29. <https://doi.org/10.3892/ol.2018.8398> PMID: 29805586; PubMed Central PMCID: PMC5950538.
 37. Liu Z, Sall A, Yang D. MicroRNA: An emerging therapeutic target and intervention tool. *International journal of molecular sciences*. 2008; 9(6):978–99. Epub 2009/03/28. <https://doi.org/10.3390/ijms9060978> PMID: 19325841; PubMed Central PMCID: PMC2658779.
 38. Rupaimoole R, Slack FJ. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. *Nature reviews Drug discovery*. 2017; 16(3):203–22. Epub 2017/02/18. <https://doi.org/10.1038/nrd.2016.246> PMID: 28209991.
 39. Shi H, Xu J, Zhang G, Xu L, Li C, Wang L, et al. Walking the interactome to identify human miRNA-disease associations through the functional link between miRNA targets and disease genes. *BMC systems biology*. 2013; 7:101. Epub 2013/10/10. <https://doi.org/10.1186/1752-0509-7-101> PMID: 24103777; PubMed Central PMCID: PMC4124764.
 40. Chen X, Wang L, Qu J, Guan NN, Li JQ. Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics (Oxford, England)*. 2018; 34(24):4256–65. Epub 2018/06/26. <https://doi.org/10.1093/bioinformatics/bty503> PMID: 29939227.
 41. Chen X, Yin J, Qu J, Huang L. MDHGI: Matrix Decomposition and Heterogeneous Graph Inference for miRNA-disease association prediction. *PLoS computational biology*. 2018; 14(8):e1006418. Epub 2018/08/25. <https://doi.org/10.1371/journal.pcbi.1006418> PMID: 30142158; PubMed Central PMCID: PMC6126877.
 42. Chen X, Zhu CC, Yin J. Ensemble of decision tree reveals potential miRNA-disease associations. *PLoS computational biology*. 2019; 15(7):e1007209. Epub 2019/07/23. <https://doi.org/10.1371/journal.pcbi.1007209> PMID: 31329575; PubMed Central PMCID: PMC6675125.