Identification of Key Genes Involved in the Progression of Renal Fibrosis and Associated with Kidney Renal Clear Cell Carcinoma

Qiming Xu

Seventh People's Hospital of Shanghai University of Traditional Chinese Medicine

Jianrao Lu (✉ jianraolu@aliyun.com)

Seventh People's Hospital of Shanghai University of Traditional Chinese Medicine

Research Article

Keywords: Chronic Kidney Disease, KIRC, Immune response, LYZ.

Posted Date: December 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1119578/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Purpose: Renal fibrosis (RF) is the necessary way for Chronic kidney disease (CKD) to develop to End Stage Renal Disease (ESRD). Patients with chronic kidney disease suffer from high morbidity and premature death due to various complications and even cancer. Therefore, this study aims to identify key genes in the pathogenesis of RF and Kidney Renal Clear Cell Carcinoma (KIRC).

Method: We analyzed the gene expression characteristics of two databases (GSE6344 and GSE22459) and used geo2R tools to obtain the differentially expressed genes (DEG). Then, use Database for Annotation, Visualization and Integrated Discovery (DAVID) for Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) path analysis. Subsequently, we used the STRING database and built the protein-protein-interactions (PPI) network, the cytoHubba plug-ins of Cytoscape were used to select the hub. Then, we used The Cancer Genome Atlas(TCGA) database to verify hub genes and further screen out core genes. Then, TargetScanHuman, miRTarbase and miRWalk databases were used to reverse-predict targeted miRNA regulated by core genes and screen out core miRNA. mRNA and miRNA mutual aid network were established. At the same time, Gene Expression Profiling Interactive Analysis (GEPIA) database was used for survival analysis of screened core genes to find genes related to prognosis. Tumor Immune Estimation Resource (TIMER) database was used to evaluate the correlation between the expression of core genes and immune cell penetration. Then use the Gene Set Enrichment Analysis (GSEA) tool to analyze the LYZ gene, and finally use the Human Protein Atlas (HPA) online database to verify the expression level of the identified central gene.

Result: We filtered 2755 DEGs from the GSE6344 database, including 1292 upregulated DEGs and 1463 downregulated DEGs; 2552 DEGs were filtered from the GSE22459 database, including 2022 downregulated DEGs and 530 upregulated DEGs. We did functional enrichment analysis of down-regulated and up-regulated differential genes, Functional enrichment analysis of up-regulated genes shows that DEGs involves many functions and expression pathways. such as immune response, plasma membrane, membrane, integral component of plasma membrane, signal transduction, extracellular region and extracellular space. It is demonstrated in the PPI network constructed by 67 nodes (proteins) and 546 PPI edges (interactions); Functional enrichment analysis of down-regulated genes also shows that DEGs involves many functions and expression pathways. such as integral component of plasma membrane, plasma membrane, extracellular space and extracellular region. It is demonstrated in the PPI network constructed by 141 nodes (proteins) and 624 PPI edges (interactions). Then a gene LYZ was selected step by step in three rounds of validation through TCGA data set, GTEx data set, Timer database and HPA database. LYZ expression was significantly correlated with the immune infiltration levels of CD4+ T cells, CD8+ T cells, Macrophage, Myeloid dendritic, Neutrophil and B cell. The upstream hub miRNA that regulate this gene were identified: has-miR-4649-3p and has-miR-873-3p. Based on these findings, it is proposed that LYZ may be a potential novel diagnostic and prognostic biomarker of KIRC at the mRNA and protein levels, and has-miR-4649-3p and has-miR-873-3p at the molecular level, and can help us better manage the progression of renal fibrosis.

Conclusion: Our findings suggest that immune response, inflammation and other pathways play an extremely important role in RF and KIRC. LYZ, has-miR-4649-3p and has-miR-873-3p may become potential prognostic biomarkers of KIRC and contribute to the prevention and treatment of renal fibrosis, which also shows us a new therapeutic idea that provides the possibility to treat renal fibrosis from the perspective of immunity.

1. Introduction

CKD brings heavy economic burden to the society and the public, and has become a public health problem seriously harmful to human health worldwide. It is chronic renal structure and dysfunction caused by various reasons. It is divided into five stages according to the glomerular filtration rate, and about 10% of CKD patients will eventually develop ESRD(1). RF is the only way and main pathological basis for the development of CKD to ESRD. Prevention and treatment of RF is the key to curb the progress of CKD. Therefore, in-depth research on the molecular mechanism of the occurrence and development of RF, search for and clarify the specific molecular targets of RF and the mechanism of action of RF therapeutic drugs are the focus of current RESEARCH on CKD, which has important social significance and economic value for strengthening and promoting the prevention and treatment of CKD.

Previous research has shown that result in renal fibrosis is the basic mechanism and the mechanism of complex, chronic kidney disease related metabolic factors, such as tissue inflammation, oxidative stress, renal interstitial fibroblasts, cytokines and signal of activation and proliferation of the cascade, the function of many factors such as will result in a change to the basis for the development of renal fibrosis, Ultimately, ECM synthesis - degradation imbalance and deposition in renal tissue, clinical manifestations of glomerular, renal interstitium and renal vascular fibrosis (2). However, due to the complex disease of renal fibrosis and many etiological factors, there has been a controversy.

Previous studies have also shown that there is a membrane-bound protein Klotho in the human kidney, which is mainly anti-fibrotic, and can inhibit tumors by inhibiting FGF2, TGF+1, IGF1, FGF2 and Wnt signaling pathways (3); In recent years, tumor immunosuppressive therapy has been regarded as one of the most successful methods in the field of cancer treatment, such as rituximab in non-Hodgkin's lymphoma (NHL), diffuse large B-cell lymphoma (DLBCL) and follicular lymphoma (FL) and other tumors have made great breakthroughs in the treatment of tumors. The drug is now used as membranous nephropathy (4), type 5 lupus nephropathy (membranous lupus nephropathy) (5), Anti-Neutrophil Cytoplasmic Antibodies (ANCA)-related vasculitis and renal damage (6). The first-line drugs for kidney diseases also suggest that the two are related in treatment. Therefore, it can be inferred that renal fibrosis is closely related to the development of kidney tumors. However, the basic mechanisms and potential interrelationships between renal fibrosis and renal clear cell carcinoma are not fully known.

Therefore, we used GEO, TCGA and other databases for bioinformatics analysis of the data. We first identified the key genes and pathways in renal fibrosis disease and renal clear cell carcinoma.

Fortunately, we found that LYZ is the common hub gene of RF and KIRC, and the analysis results will provide meaningful clues for the treatment of renal fibrosis.

2. Result

2.1 DEG identification in different GEO data sets

In this study, gene expression characteristics of two datasets GSE6344 and GSE22459 were searched by NCBI-Geo database. In these Data sets, GSE6344 Microarray Data contained 20 samples, and further analysis of the specific composition revealed that it contained 10 renal cell carcinoma groups and 10 normal tissue samples. At the same time, GSE22459 Microarray Data contains 65 samples, which are specifically grouped into 24 renal fibrosis samples group and 25 normal controls group. We used GEO2R online tool to analyze the data obtained from GEO database. Finally, we found the genetic data of 2755 differentially expressed genes from GSE6344 dataset (Figure 1D). Its the cut-off criteria were $|\log FC| \geq 0.5$ and $P < 0.05$; We found from GSE22459 data set 2552 differentially expressed genes of genetic data (figure 1E), its the cut - off criteria were $|\log FC| \geq 1$ and $P < 0.05$. A total of 347 DEGs were detected (Figure 1A), of which 152 were down-regulated genes ($\log FC \leq -1$) (Figure 1B) and 72 DEGs were up-regulated genes ($\log FC \geq 1$) (Table 1 and Figure 1C).

Table 1

347 DEGS WERE IDENTIFIED FROM TWO PROFILE DATASETS	
DEGs	Gene Name
Up	CRTAM▯ANXA1▯DOCK2▯FPR3▯PLEK▯RUNX3▯SLC02B1▯RAC2▯CCL19▯LAMP3▯FYB▯RHOH▯ VASH1▯HLA-DQB1▯GZMK▯TRAT1▯NKG7▯XCL2▯XCL1▯PVRIG▯LYST▯SLA▯ARHGAP25▯ SEZ6L2▯LST1▯MELK▯SLC4A7▯GPR18▯RGS1▯PTPRC▯GNLY▯ITGA4▯C1S▯GZMH▯NLGN4X▯ PML▯ADGRE5▯CCL18▯ASPM▯GRIN2D▯HLA-DPB1▯CORO1A▯CD52▯TOP2A▯LOC389906▯PLEC▯ ISG20▯HRH1▯SLC17A2▯REG1A▯NR1H4▯SDC3▯DLGAP5▯IL7R▯SERPINE1▯VCAN▯SLC1A3▯ CCL5▯DPEP2▯SLC38A1▯LYZ▯CSF2RA▯PON2▯KIF20A▯IGSF6▯LY9▯TRBC1▯ LOC101928916▯NNMT▯NOD2▯APBB1IP▯IGK▯IGKC▯APOC1▯NUSAP1
Down	DMP1▯EPB41L4B▯SLC26A1▯COL4A5▯TCL6▯TRPM6▯GNAS▯ RPL21P28▯SNORA27▯SNORD102▯RPL21▯PDZRN3▯FKRP▯TF▯ZNHIT2▯SPTBN1▯TACR1▯ GRIK2▯CCDC181▯HABP4▯MFAP5▯CHI3L1▯PLA2R1▯CEL▯CBARP▯TRPV5▯IL5RA▯NDNF▯ PRDM12▯PROX1▯PPM1H▯PACSIN3▯EGF▯MYOZ2▯TNNT2▯APOA2▯BAIAP3▯GRM1▯CLIC5▯ KMT2B▯MFAP3L▯NR4A3▯MAGI1▯CR1▯SLC22A3▯SYNPO2L▯MAGEA8▯ASB4▯LINC00652▯ U2AF2▯ACR▯PPFIA2▯CELA3B▯TSC22D2▯POU6F2▯EYA3▯TACSTD2▯PHF2▯STRA6▯WISP1▯ HHLA3▯MINA▯PDE4DIP▯GAS1▯TNNI1▯TRPM3▯CA8▯KCNN2▯MBP▯GHRHR▯CD22▯RPARP-AS1▯ SUCLG2▯HSPB7▯TSHR▯CTTN▯REEP2▯GNRH2▯PTGER3▯TGFB2▯CHRD1▯ATXN7L1▯ASCL3▯ MIR124-1▯LINC00599▯LOC55338▯BCAN▯ARPP21▯ATOH1▯ADAM5▯CFAP46▯SCIN▯PCDHB8▯ HAO1▯TNFSF15▯SOS2▯CHGA▯PAPPA▯LOC100289518▯TNNC1▯CTAG2▯F11▯CYP2A6▯ LINC01361▯GABRB1▯MPO▯NEU3▯SLC26A4▯USP2▯NDOR1▯PBOV1▯CELF3▯TYRO3▯PEX6▯ KCNA10▯VIPR2▯DUOX1▯PCDHA10▯PTPRS▯PTPN1▯PLCG2▯ZNF428▯NUP188▯SSX2B▯SSX2▯ NLRP2▯EXPH5▯MAG▯ATP6V1H▯RPL3L▯PMP2▯EDDM3A▯TRPM1▯CCKAR▯DDN▯AVPR1A▯ ANKRD2▯HOXB6▯MC5R▯METTL7A▯CADM1▯LMO3▯CR2▯MYH8▯ADH1B▯RASAL2▯MLANA▯ NPY2R▯PAX3▯EYA4▯RFX4▯ANKRD1▯EFCAB1▯TPSD1▯FTCD▯SMG7-AS1▯GPLD1

2.2 GO and KEGG Enrichment Analyses

Then, we used the online tool DAVID software to conduct GO analysis and KEGG analysis for the 72 up-regulated DEGs and 152 down-regulated DEGs mentioned above. These DEGs can be divided into the following two basic types: Cellular Components (CC) and Biological Processes (BP). Functional enrichment analysis of up-regulated genes showed that (Figure 2A), Genes in Cell Component (CC) mainly involve plasma membrane, membrane, integral component of plasma membrane, extracellular region and extracellular space. In Biological Processes (BP), the main function of these different genes is immune response and signal transduction (Table 2-1); Functional enrichment analysis of down-regulated genes showed that (Figure 2B), Genes in Cell Component (CC) mainly involve integral component of plasma membrane, plasma membrane, extracellular region and extracellular space (Table 2-2). The cut-off criteria were Count ≥ 10 and $P < 0.05$. KEGG pathway enrichment analysis showed that a total of 12 important signaling pathways were identified in up-regulated genes (Figure 2C). Including Cell adhesion Molecules (CAMs), Staphylococcus aureus Infection, Chemokine Signaling Pathway, Intestinal immune network for IgA production, Cytokine-cytokine receptor interaction, Viral myocarditis, Inflammatory bowel disease (IBD), Leishmaniasis, Influenza A, Tuberculosis, Herpes simplex infection and Fc gamma r-

mediated phagocytosis (Table 2-3); A total of 6 important signaling pathways of down-regulated genes were identified (Figure 2D). Including Neuroactive ligand-receptor interaction, Calcium signaling Pathway, B Cell receptor signaling Pathway, and Dilated Cardiomyopathy, Hematopoietic cell lineage and Gap junction (Table 2-4). The filter criteria are PValue<0.05.

Table 2-1

GO Enrichment analysis of up-regulated DEGs

Category	Term	Count	PValue
GOTERM_BP_DIRECT	GO:0006955~immune response	13	6.39E-08
GOTERM_CC_DIRECT	GO:0005886~plasma membrane	35	4.37E-07
GOTERM_CC_DIRECT	GO:0016020~membrane	22	2.43E-05
GOTERM_CC_DIRECT	GO:0005887~integral component of plasma membrane	16	1.53E-04
GOTERM_BP_DIRECT	GO:0007165~signal transduction	12	0.004665
GOTERM_CC_DIRECT	GO:0005576~extracellular region	13	0.013557
GOTERM_CC_DIRECT	GO:0005615~extracellular space	11	0.024663

Table 2-2

GO enrichment analysis of down-regulated DEGs

Category	Term	Count	PValue
GOTERM_CC_DIRECT	GO:0005887~integral component of plasma membrane	28	4.31E-06
GOTERM_CC_DIRECT	GO:0005886~plasma membrane	48	8.05E-04
GOTERM_CC_DIRECT	GO:0005576~extracellular region	20	0.02846
GOTERM_CC_DIRECT	GO:0005615~extracellular space	17	0.04093

Table 2-3

KEGG pathway enrichment analysis of up-regulated DEGs

Term	Count	PValue
hsa04514: Cell adhesion molecules (CAMs)	7	2.24E-05
hsa05150: Staphylococcus aureus infection	4	0.001446
hsa04062: Chemokine signaling pathway	5	0.007241
hsa04672: Intestinal immune network for IgA production	3	0.016498
hsa04060: Cytokine-cytokine receptor interaction	5	0.018014
hsa05416: Viral myocarditis	3	0.023734
hsa05321: Inflammatory bowel disease (IBD)	3	0.029443
hsa05140: Leishmaniasis	3	0.035642
hsa05164: Influenza A	4	0.035903
hsa05152: Tuberculosis	4	0.037487
hsa05168: Herpes simplex infection	4	0.040764
hsa04666: Fc gamma R-mediated phagocytosis	3	0.048353

Table 2-4

KEGG pathway enrichment analysis of down-regulated DEGs

Term	Count	PValue
hsa04080: Neuroactive ligand-receptor interaction	12	3.01E-05
hsa04020: Calcium signaling pathway	8	0.001056
hsa04662: B cell receptor signaling pathway	4	0.023973
hsa05414: Dilated cardiomyopathy	4	0.039671
hsa04640: Hematopoietic cell lineage	4	0.043308
hsa04540: Gap junction	4	0.044557

2.3 PPI Network Analysis

We further introduced the above-mentioned up-regulated DEGs and down-regulated genes into the online analysis tool STRING, and obtained two different PPI networks to explore the interaction of up-regulated and down-regulated DEGs. In the PPI network of up-regulated differential genes, in addition to the disconnected nodes, it was also verified that 67 nodes (proteins) and 546 PPI edges (interactions) were in the PPI network constructed by up-regulated DEG (Figure 3A). The PPI network is imported into Cytoscape software, and the cytoHubba plug-in is used to determine hub genes (Figure 3C and Table 3), including

PTPRC, PLEK, CD52, FYB, NKG7, CORO1A, LYZ, RAC2, DOCK2, NOD2. Using the same method, we obtained the PPI network (Figure 3B) and hub genes (Figure 3D and Table 3) that down-regulated differential genes, including ANKRD1, MYH8, TNNI1, SYNPO2L, TNNC1, MYOZ2, TNNT2, HSPB7, GRM1, CELF3. These hub genes may serve as promising biomarkers for renal fibrosis and renal clear cell carcinoma, and may be involved in the treatment of RF and KIRC.

Table 3

Top ten hub genes with higher Maximal Clique Centrality (MCC) of connectivity

DEGs	Rank	Name	Score
Up	1	PTPRC	7.64E+09
	2	PLEK	7.64E+09
	3	CD52	7.62E+09
	4	FYB	7.53E+09
	5	NKG7	7.34E+09
	6	CORO1A	7.31E+09
	7	LYZ	7.26E+09
	8	RAC2	7.09E+09
	9	DOCK2	6.94E+09
	10	NOD2	6.92E+09
Down	1	ANKRD1	5936
	2	MYH8	5869
	3	TNNI1	5827
	4	SYNPO2L	5800
	5	TNNC1	5774
	6	MYOZ2	5762
	7	TNNT2	5162
	8	HSPB7	5047
	9	GRM1	3516
	10	CELF3	3277

2.4 Verification of TCGA Data Set

First, we downloaded the KIRC and normal sample group data sets in TCGA and GTEx, and then used the GEPIA database to verify the expression levels and predicted values of hub genes. The results showed that the mRNA expression levels of a total of 13 central genes were significantly higher than that of normal kidneys organization ($P < 0.05$) (Figure 4). These findings are consistent with the microarray data obtained.

2.5 mRNA-miRNA regulatory network

We reversely predicted the targeted miRNA regulated by 13 hub genes, and screened out 6 hub miRNAs through cytoHubba plug-in in the Cytoscape software (Figure 5B), including has-miR-3183, has-miR-4646-5p, has-miR-520a-5p, has-miR-4723-3p, has-miR-4649-3p and has-miR-873-3p, which plays an important role in regulating gene expression, cell cycle and developmental sequence of organisms. Then establish a mutual assistance network between mRNA and miRNA (Figure 5A). At the same time, four genes related to hub miRNAs were screened out, namely TNNT1, LYZ, DOCK2 and PLEK.

2.6 Hub genes associated with prognosis

In the previous step, we screened out four hub genes, and then we used GEPIA online database to analyze whether these central genes were associated with prognosis and delineate the survival curve. Analysis showed that only LYZ gene was associated with the overall survival rate of patients with KIRC ($P < 0.05$) (Figure 6). Meanwhile, we also noticed that the upstream miRNA directly associated with LYZ gene were has-miR-4649-3p and has-miR-873-3p, so we further screened out our core miRNA: has-miR-4649-3p and has-miR-873-3p.

2.7 The expression of LYZ is related to immune cell infiltration in KIRC

We obtained the final core gene LYZ through the previous step. Through prognostic analysis, we know that this gene is a protective gene (patients with high LYZ expression will survive for a long time), so we need to pick out immune cells with anti-tumor potential, including B cells and CD4+ T cells, CD8+ T cells, neutrophils, macrophages and dendritic cells. Then, the online analysis tool TIMER was used to analyze the correlation between multi-sample phasing and somatic cell copy number changes in KIRC samples and the above-mentioned immune cell infiltration. Our data showed that somatic copy number changes were significantly associated with the infiltration of CD8 + T cells ($p < 0.05$) and macrophages ($p < 0.05$) (Figure 7A). As shown in Figure 7B, the analysis showed that LYZ expression in KIRC was significantly correlated with tumor purity ($\text{cor} = -0.241$; $p = 1.67e - 07$), B cells ($\text{cor} = 0.429$; $p = 5.57e - 22$), CD8 + T cells ($\text{cor} = 0.383$; $p = 9.88e - 17$), CD4+ T cells ($\text{cor} = 0.283$; $p = 6.33e - 10$), macrophages ($\text{cor} = 0.592$; $p = 1.18e - 43$), neutrophils ($\text{cor} = 0.606$; $p = 2.83e - 47$) and Dendritic Cell ($\text{cor} = 0.635$; $p = 1.11e - 52$). According to the above results, the expression of LYZ was significantly positively correlated with these immune Cell infiltrates, and LYZ had a great influence on the Dendritic Cell infiltration of KIRC patients. Therefore, it was speculated that LYZ affected the survival of KIRC patients by affecting the degree of Dendritic Cell infiltration.

2.8 Rich analysis of LYZ functional networks in KIRC

The phenotype we need has been found above. Next, we need to explain what mechanism LYZ may influence the Dendritic Cell infiltration level of KIRC patients. We divided the high and low expression of LYZ in KIRC patients into two groups, and then conducted GSEA analysis to find out the pathway of immune cell infiltration regulation. The combined results showed that LYZ was found to be involved in both RF progression and KIRC. We used the functional module of LinkedOmics to analyze the LYZ mRNA sequencing data of 533 KIRC patients in TCGA. As shown in volcano diagram (FIG. 8A), LYZ was positively correlated with 4386 genes (dark red dots) and negatively correlated with 3166 genes (dark green dots) (FDR < 0.01). 50 significant genes positively correlated with LYZ are shown in the heat map (Figure 8B), and 50 significant gene sets negatively correlated with LYZ are shown in the heat map (Figure 8C). They are associated with carcinogenicity of cancer, progression of fibrosis, development of kidney disease and inflammation. KEGG pathway analysis of GSEA is shown in Figure 8D. LYZ-related genes are detected in Leishmaniasis (FDR=0, P value=0), Th1 and Th2 cell differentiation (FDR=0, Immune network for IgA production (FDR=0, P value=0), Influenza A (FDR=0, P value=0), NF-Kappa B Signaling Pathway (FDR=0, P value=0), Ribosome (FDR=0, P value=0), Vibrio Cholerae infection (FDR=0, P value=0), Ubiquinone and other terpenoid-Quinone biosynthesis (FDR=0, P value=0) and Alzheimer disease (FDR=0, P value=0) (Figure 8E-P). Therefore, LYZ may be used as a diagnostic and therapeutic target for RF and KIRC by regulating immune cell infiltration through the above-mentioned pathways.

2.9 Verifying the HPA database

Finally, the protein expression level of LYZ was analyzed in HPA online database. Forty-four partial histological images of KIRC and normal renal tissues were analyzed. The results showed that LYZ gene protein was mainly expressed in renal tubules, and the expression of LYZ gene in KIRC was significantly increased compared with normal tissues (Figure 9), which was consistent with the above verification of mRNA levels.

3. Discussion

In this study, in order to find the key genes between RF and KIRC and find out the connection between the two, we determined the key genes and biological pathways through bioinformatics analysis, found the differential genes through GSE6344 and GSE22459, and showed through functional enrichment analysis, Most DEGs are enriched in signal transduction, inflammatory response and immune response. KEGG pathway analysis revealed that DEGs were also significantly enriched in cellular adhesion molecules (CAM), chemokine signaling pathways, IgA producing intestinal immune network, and cytokine - cytokine receptor interactions. Interestingly, most of these pathways are involved in inflammation and immune-related functions. In the subsequent analysis, it was found that the core genes identified by us were also closely related to immune infiltration, carcinogenicity of cancer and fibrosis progression. Therefore, inflammatory and immune-related pathways may play an important role in RF and KIRC, and provide the possibility that immunoagents previously used to treat tumor-related renal fibrosis can be used.

Importantly, we found 13 central genes between RF and KIRC by analyzing GSE6344 and GSE22459 databases, among which 4 genes (TNNI1, LYZ, DOCK2 and PLEK) are related to the molecular mechanism network of RF and KIRC pathogenesis. Through further research, we found that LYZ gene is not only involved in the molecular mechanism network of disease, but also related to prognosis and is a protective gene for the development of disease. These results suggest that it can be identified as a novel biomarker and therapeutic target for RF and KIRC. Then we focused on the study of LYZ and conducted functional enrichment analysis of LYZ gene in KIRC by using the online database LinkedOmics. KEGG pathway analysis showed that the genes associated with LYZ were found in Leishmaniasis, Th1 and Th2 cell differentiation, immune network for IgA production, and Influenza A, NF-Kappa B Signaling Pathway, Ribosome, Vibrio Cholerae Infection, Ubiquinone and other Terpenoid-Quinone biosynthesis And Alzheimer disease were significantly enriched. Interestingly, for example, the NF-kappa B signaling pathway has been shown by most studies to play a key role in the activation of this pathway and the progression of chronic kidney inflammation (7), and in the treatment of CKD mainly shows kidney damage and inflammatory detoxicity (8). In addition, in this study, LYZ expression in tumor tissues was significantly higher than that in normal tissues, and somatic copy number change of LYZ was significantly correlated with infiltration of CD8 + T cells and macrophages. Meanwhile, LYZ expression was significantly correlated with tumor purity and Dendritic Cell infiltration. Therefore, LYZ expression may be related to tumor immune microenvironment. These data suggest that LYZ may be used as a diagnostic and therapeutic target for RF and KIRC.

Although we identified candidate hub genes between RF and KIRC using bioinformatics techniques, there are still some limitations in this study. First of all, the differential genes obtained in our study were from GEO and TCGA databases, and were not verified by RT-qPCR. In the subsequent study, we will collect clinical sample data of RF and KIRC patients respectively, and verify the expression level of core genes through RT-qPCR experiments. Secondly, we just normal and tumor tissues were analyzed between the infiltration of immune cells, and no attention to the differences between these subtypes, also did not notice the normal group and RF tissue between the infiltration of immune cells, therefore based on the immunity infiltration between both diseases need further exploration, in the end, we plan to explore in animal models of RF, KIRC mechanism of hub genes.

4. Conclusion

In conclusion, LYZ, has-Mir-4649-3p and has-Mir-873-3p identified by us may become potential prognostic markers of KIRC. Considering the relationship between immune infiltration and inflammatory response and KIRC and RF, immunotherapy methods may have broad prospects in the treatment of RF.

5. Materials And Methods

5.1 Extraction of Microarray Data

GSE6344 and GSE22459 Gene expression profiles from NCBI Gene expression Integration (GEO) (<http://www.ncbi.nlm.nih.gov/gds/>) database for download (9). A total of 20 samples were studied in the GSE6344 dataset, including 10 tumor renal cancer groups and 10 normal groups. In addition, a total of 49 samples were studied in the GSE22459 dataset, of which 25 samples were from the normal group and the other 24 samples were from the renal fibrosis group.

5.2 Identifying the DEGs

GEO2R is an online interactive networking tool that allows users to compare data sets in the GEO family to identify DEGs under different experimental conditions. We use GEO2R online tools to detect GSE6344 DEG renal tumor group and normal group, as a standard for $|\log FC| \geq 1$ and $P < 0.05$; Detection of renal fibrosis group and normal group in GSE22459 DEG, the standard for $|\log FC| \geq 0.05$ and $P < 0.05$. Subsequently, the original data recorded in TXT format were verified by online tools of Venn diagrams, in order to find important DEG in two data sets: when the test results showed $\log FC \leq -1$, the expression of this gene was considered to be down-regulated; If $\log FC \geq 1$, gene expression is considered up-regulated.

5.3 GO and pathway enrichment analysis

Ontology of genes (GO) is a database established by the Gene Ontology Consortium, which aims to define and describe the functions of genes and proteins for various species and keep updating as further research progresses (10). At the same time, The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a comprehensive database integrating genomic, chemical and systematic functional information designed to reveal the genetic material and chemical blueprint of life phenomena (11). The KEGG database can easily classify various genes of the system path management, KEGG comprehensively analyzed the functions and biochemical pathways of the selected DEGs, If $P < 0.05$, the result is considered to be statistically significant. The Database for Annotation, The Visualization and Integrated Discovery (DAVID) (<https://david.ncifcrf.gov/tools.jsp>) (12) is a biological information database, the integration of biological data and analysis tool for large-scale gene or protein list ID (hundreds of genes or proteins ID list) to provide systematic and comprehensive biological function annotation information, help the user to extract information. We use DAVID to analyze Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) on RF and KIRC enrichment analysis of related genes, including biological process (BP), molecular function (MF), cell composition (CC), and other DEGs to obtain more research information.

5.4 PPI network analysis and hub gene identification

STRING (<http://string-db.org>) is a database that searches for known and predicted interactions between proteins (13) and is used to retrieve interacting genes to construct PPI networks. In order to explore the central protein and hub genes related to RF and KIRC, we screened the differential genes whose total score of interaction between differential genes was ≥ 0.1 , and then constructed a PPI network using STRING database. Finally, we visualized the network using Cytoscape software and revealed the central genes with high (junction nodes) in the PPI network. These genes are considered to be central genes and may play an important role in the relationship between renal fibrosis and KIRC.

5.5 Validation of the TCGA dataset

We validated the selected central genes in TCGA and GTEx KIRC datasets using GEPIA online tools (<http://gepia.cancer-pku.cn>)(14), a bioinformatics tool that addresses important questions in cancer biology, reveal cancer subtypes, driver genes, alleles, differential expression or carcinogenic factors, thus digging into a database of novel cancer targets and markers for validation. Finally, we got nine up-regulated core genes and down-regulated core genes.

5.6 Prediction of candidate mRNA-miRNA regulatory networks

We used TargetScanHuman(15), miRTarbase(16) and miRWalk(17) databases to reversely predict the targeted miRNA regulated by 9 up-regulated genes and 4 down-regulated genes, then established a mutual network of mRNA and miRNA, and finally screened out the hub miRNA.

5.7 Screening hub genes associated with prognosis

Previously, we further screened out 3 up-regulated hub regulatory genes and 1 down-regulated core regulatory gene. Then, we used GEPIA online database to analyze the expression level and predictive value of these central genes, and plotted the survival analysis curve.

5.8 Immune cell infiltration

The Tumor Immune Estimation Resource (TIMER) algorithm database (<https://cistrome.shinyapps.io/timer/>) provides a more reliable estimate of the immune infiltration level for TCGA or the tumor atlas provided by the user, and provides the tumor Comprehensive analysis and visualization function of infiltrating immune cells (18). We calculated the expression of LYZ and 6 types of immune infiltrating cells (B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages and dendritic cells) in KIRC samples through this database. Correlation between the abundance of phages and dendritic cells.

5.9 Gene Set Enrichment Analysis (GSEA)

LinkedOmics database (<https://www.linkedomics.org/>) from all 32 kinds of TCGA cancer types and 10 from Clinical Proteomics Tumor Analysis Consortium (CPTAC) multibody data for cancer clusters (19). In addition, we can download a large amount of well-organized data from this database and have very powerful online analysis capabilities. In this study, GSEA was proposed to study the difference between high and low expression of LYZ in KIRC.

5.10 HPA Database verification

Finally, the protein expression level of LYZ was analyzed in HPA database. Partial histological images of KIRC and normal renal tissue were analyzed. The results showed that, compared with normal tissues, the expression of LYZ gene protein in KIRC tissues increased significantly.

6.abbreviations

RF: Renal fibrosis

CKD: Chronic kidney disease

ESRD: End Stage Renal Disease

KIRC: Kidney Renal Clear Cell Carcinoma

DEG: differentially expressed genes

DAVID: Database for Annotation, Visualization and Integrated Discovery

GO: Gene Ontology

KEGG: Kyoto Encyclopedia of Genes and Genomes

PPI: protein-protein-interactions

TCGA: The Cancer Genome Atlas

GEPIA: Gene Expression Profiling Interactive Analysis

TIMER: Tumor Immune Estimation Resource

GSEA: Gene Set Enrichment Analysis

HPA: Human Protein Atlas

NHL: non-Hodgkin's lymphoma

DLBCL: diffuse large B-cell lymphoma

FL: follicular lymphoma

ANCA: Anti-Neutrophil Cytoplasmic Antibodies

CC: Cellular Components

BP: Biological Processes

CAMs: Cell adhesion Molecules

IBD: Inflammatory bowel disease

MCC: Maximal Clique Centrality

Declarations

Consent for publication

Not applicable

Availability of data and materials

In this study, the analysis of the data set can be obtained from the corresponding author upon request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (DOI:82074261)

Authors' contributions

Data analysis and interpretation, manuscript writing by Qiming Xu. The development of methodology, concept design, review and revision work were completed by Jianrao Lu. Qiming Xu and Jianrao Lu contributed equally to this work.

References

1. Zhang L, Wang F, Wang L, Wang W, Liu B, Liu J, Chen M, He Q, Liao Y, Yu X, Chen N, Zhang JE, Hu Z, Liu F, Hong D, Ma L, Liu H, Zhou X, Chen J, Pan L, Chen W, Wang W, Li X, Wang H. Prevalence of chronic kidney disease in China: a cross-sectional survey. *Lancet*. 2012 Mar 3;379(9818):815-22. doi: 10.1016/S0140-6736(12)60033-6. Erratum in: *Lancet*. 2012 Aug 18;380(9842):650. PMID: 22386035.
2. Nogueira A, Pires MJ, Oliveira PA. Pathophysiological Mechanisms of Renal Fibrosis: A Review of Animal Models and Therapeutic Strategies. *In Vivo*. 2017 Jan 2;31(1):1-22. doi: 10.21873/in vivo.11019. PMID: 28064215; PMCID: PMC5354133.
3. Mencke R, Olauson H, Hillebrands JL. Effects of Klotho on fibrosis and cancer: A renal focus on mechanisms and therapeutic strategies. *Adv Drug Deliv Rev*. 2017 Nov 1;121:85-100. doi: 10.1016/j.addr.2017.07.009. Epub 2017 Jul 12. PMID: 28709936.
4. Zhang J, Bian L, Ma FZ, Jia Y, Lin P. Efficacy and safety of rituximab therapy for membranous nephropathy: a meta-analysis. *Eur Rev Med Pharmacol Sci*. 2018 Nov;22(22):8021-8029. doi: 10.26355/eurev_201811_16431. PMID: 30536351.
5. Schioppo T, Ingegnoli F. Current perspective on rituximab in rheumatic diseases. *Drug Des Devel Ther*. 2017 Oct 3;11:2891-2904. doi: 10.2147/DDDT.S139248. PMID: 29042750; PMCID: PMC5633295.

6. Hassan RI, Gaffo AL. Rituximab in ANCA-Associated Vasculitis. *Curr Rheumatol Rep*. 2017 Feb;19(2):6. doi: 10.1007/s11926-017-0632-1. PMID: 28155022.
7. Gómez-Garre D, Largo R, Tejera N, Fortes J, Manzarbeitia F, Egido J. Activation of NF-kappaB in tubular epithelial cells of rats with intense proteinuria: role of angiotensin II and endothelin-1. *Hypertension*. 2001 Apr;37(4):1171-8. doi: 10.1161/01.hyp.37.4.1171. PMID: 11304520.
8. Fujihara CK, Antunes GR, Mattar AL, Malheiros DM, Vieira JM Jr, Zatz R. Chronic inhibition of nuclear factor-kappaB attenuates renal injury in the 5/6 renal ablation model. *Am J Physiol Renal Physiol*. 2007 Jan;292(1):F92-9. doi: 10.1152/ajprenal.00184.2006. Epub 2006 Aug 8. PMID: 16896182.
9. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res*. 2013 Jan;41(Database issue):D991-5. doi: 10.1093/nar/gks1193. Epub 2012 Nov 27. PMID: 23193258; PMCID: PMC3531084.
10. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000 May;25(1):25-9. doi: 10.1038/75556. PMID: 10802651; PMCID: PMC3037419.
11. Altermann E, Klaenhammer TR. PathwayVoyager: pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. *BMC Genomics*. 2005 May 3;6:60. doi: 10.1186/1471-2164-6-60. PMID: 15869710; PMCID: PMC1112592.
12. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44-57. doi: 10.1038/nprot.2008.211. PMID: 19131956.
13. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, Jensen LJ, von Mering C. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*. 2021 Jan 8;49(D1):D605-D612. doi: 10.1093/nar/gkaa1074. Erratum in: *Nucleic Acids Res*. 2021 Oct 11;49(18):10800. PMID: 33237311; PMCID: PMC7779004.
14. 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 Jul 3;45(W1):W98-W102. doi: 10.1093/nar/gkx247. PMID: 28407145; PMCID: PMC5570223.
15. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *Elife*. 2015 Aug 12;4:e05005. doi: 10.7554/eLife.05005. PMID: 26267216; PMCID: PMC4532895.
16. Huang HY, Lin YC, Li J, Huang KY, Shrestha S, Hong HC, Tang Y, Chen YG, Jin CN, Yu Y, Xu JT, Li YM, Cai XX, Zhou ZY, Chen XH, Pei YY, Hu L, Su JJ, Cui SD, Wang F, Xie YY, Ding SY, Luo MF, Chou CH, Chang NW, Chen KW, Cheng YH, Wan XH, Hsu WL, Lee TY, Wei FX, Huang HD. miRTarBase 2020:

- updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res.* 2020 Jan 8;48(D1):D148-D154. doi: 10.1093/nar/gkz896. PMID: 31647101; PMCID: PMC7145596.
17. Sticht C, De La Torre C, Parveen A, Gretz N. miRWalk: An online resource for prediction of microRNA binding sites. *PLoS One.* 2018 Oct 18;13(10):e0206239. doi: 10.1371/journal.pone.0206239. PMID: 30335862; PMCID: PMC6193719.
 18. Li T, Fu J, Zeng Z, Cohen D, Li J, Chen Q, Li B, Liu XS. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* 2020 Jul 2;48(W1):W509-W514. doi: 10.1093/nar/gkaa407. PMID: 32442275; PMCID: PMC7319575.
 19. Vasaikar SV, Straub P, Wang J, Zhang B. LinkedOmics: analyzing multi-omics data within and across 32 cancer types. *Nucleic Acids Res.* 2018 Jan 4;46(D1):D956-D963. doi: 10.1093/nar/gkx1090. PMID: 29136207; PMCID: PMC5753188.

Figures

Figure 1

Identify DEGs in different GEO data sets. Figure 1A identifies 99 different genes for two data sets (GSE6344 and GSE22459). Different color regions represent different data sets, and the crossover part is DEG; Figure 1B for two data sets with a downward adjustment of DEGs; Figure 1C shows DEGs with upward adjustment function in the two data sets; FIG. 1D and FIG. 1E show the comparison of volcano plots between the experimental group and the normal group in the GSE6344 and GSE22459 data sets

respectively. The heatmap was plotted by <http://www.bioinformatics.com.cn>, a free online platform for data analysis and visualization. $\log FC \geq 1$ and $P < 0.05$ were used to identify differentially expressed genes. Red color is indicative of up-regulated genes and green is indicative of down-regulated genes. Non-changed genes were shown in grey color.

Figure 2

GO and KEGG Enrichment Analyses. Figure 2A and Figure 2B show the functional enrichment of a rich GO term BP, CC and MF three-in-one bar chart of up-regulated and down-regulated differential genes, respectively. Figure 2C shows the enrichment dot bubble of up-regulated DEGs. Figure 2D shows the enrichment dot bubble of down-regulated DEGs.

Figure 3

The establishment of protein-protein-interactions network and the screening of core genes. The PPI network shown in Figure 3A is established by increasing DEG and consists of 67 nodes and 546 PPI edges. The network shown in Figure 3B is a down-regulated gene establishment and consists of 141 nodes and 624 PPI edges. Nodes represent proteins, and edges represent protein interactions. Figure 3C and Figure 3D indicate that the up-regulated and down-regulated HUB genes were identified using the cytoHubba plug-in in the Cytoscape software.

Figure 4

Use GEPIA database KIRC tissue and normal kidney tissue to verify (A) CD52, (B) CORO1A, (C) DOCK2, (D) FYB, (E) HspB7, (F) LYZ, (G) NKG7, (H) PLEK, (I) PTPRC, (J) RAC2, (K) TNNC1, (L) TNNI1 and (M) TNNT2 mRNA expression levels. The thirteen boxes are based on 523 KIRC samples (marked in red) and 100 normal samples (marked in grey). * $P < 0.05$ is considered statistically significant.

Figure 5

The establishment of mRNA-miRNA regulatory network (Figure 5A), where red represents up-regulated genes and green represents down-regulated genes. Figure 5B shows the core genes and miRNAs screened by cytoHubba plug-in in the Cytoscape software, where the rectangle represents the hub miRNA and the triangle represents the hub gene.

Figure 6

Prognostic analysis of hub genes in patients with KIRC. (A) Overall survival analysis of DOCK2 by GEPIA. (B) GEPIA's overall survival analysis of LYZ. (C) GEPIA's overall survival analysis of PLEK. (D) Overall survival analysis of TNNI1 by GEPIA. $P < 0.05$ was considered statistically significant.

Figure 7

LYZ expression was related to immune cell infiltration in KIRC. (Figure 7A) Correlation between somatic copy number change and 6 immune cell infiltration. (Figure 7B) Correlation between LYZ expression and 6 kinds of immune cell infiltration in KIRC samples.

Figure 8

Function-rich analysis of LYZ in KIRC by LinkedOmics online tool. (A) Differentially expressed genes from LYZ in volcano map. (B, C) Heat maps of TOP 50 genes positively and negatively correlated with LYZ. (D) KEGG pathway analysis of LYZ by GSEA. (E-P) Leishmaniasis (FDR=0, P value=0), Th1 and Th2 cell differentiation (FDR=0, P value=0), Intestinal immune network for IgA production (FDR=0, P value=0), Influenza A (FDR=0, P value=0), NF-Kappa B Signaling Pathway (FDR=0, P value=0), Ribosome (FDR=0, P value=0), Vibrio

Cholerae infection (FDR=0, P value=0), Ubiquinone and other terpenoid-Quinone biosynthesis (FDR=0, P value=0) and Alzheimer disease (FDR=0, P value=0). The standard is false Discovery Rate <0.01 (FDR<0.01).

Figure 9

The results showed that LYZ gene protein was mainly expressed in renal tubules, and the expression of LYZ gene in KIRC was significantly increased compared with normal tissues (Figure 9), which was consistent with the above verification of mRNA levels.