In silico meta-analyses for putative biomarkers associated with Renal Cancer

Bhardwaj Anuradha[1,2*] and Ashra Himani[2]
1. School of Biotechnology, Gautam Buddha University, Greater Noida, Uttar Pradesh-201312, INDIA
2. International Institute of Health Management Research (IIHMR), New Delhi-110075, INDIA
*anumohitvats@gmail.com

Abstract

Renal cell carcinoma (RCC) is the most prevalent form of renal malignancy accounting for around 3% of all adult malignancies. Although new targeted medications are continually being developed, they are not able to treat all patients. Thus, a comprehensive investigation of the mode of progression and biomolecular mechanism of renal cancer is a need to identify its novel targets for better diagnostics and treatment strategies. We aim to identify the potential biomarkers of renal cancer, infer the cellular processes and pathways influenced by renal cancer, using in silico methods. We analysed three profiles of gene expression (GSE168845, GSE66270 and GSE781) from Gene Expression Omnibus (GEO) database to investigate possible treatment targets for RCC. Differentially expressed genes (DEGs) between RC and normal renal tissues were found using the GEO2R web program. Gene Ontology (GO) and KEGG pathway enrichment analysis were carried out using the Enrichr web-based tool. The DEGs were then organized into a protein-protein network using the STRING database tool. From an interaction network of multiple genes, we filtered the critical hub genes using the CytoHubba application of Cytoscape. To verify the predictive-value of the hub genes, we performed survival-analysis using a renal cancer database by plotting the Kaplan-Meier plots.

We identified a set of 30 DEGs (24 upregulated genes and 6 downregulated genes). Most of the DEGs were active in signaling and transportation mechanisms. The PPI network and Cytohubba results revealed ten critical hub genes including UMOD, SLC34A1, SLC22A6, SLC12A1, RHCG, NPHS2, KCNJ1, G6PC, FABP1 and ALB. The Kaplan-Meier plotter database confirmed that few genes enhanced the chances of survival, while others decreased them and some genes had no effect on RC patient survival. The identification of DEGs and the enrichment of their biological functions/key pathways offers more precise information about RC and allows identification of crucial biomarkers which will aid future research and help in efficient therapeutic strategies.

Keywords: Renal cancer, protein-protein interactions, differentially expressed genes, functional enrichment

Introduction

Renal cell cancer or kidney cancer is the most common cancer in the genitourinary system with highest mortality rate[1]. It has a high mortality rate and is responsible for almost 90% of all renal malignancies[2]. Furthermore, the prevalence of Renal Cell Carcinoma (RCC) has been continuously increasing during the past few decades[3]. Based on histopathological features, clinical phenotype and molecular biology, renal cell carcinoma may be categorized into four broad categories: clear cell RCC, papillary RCC, chromophobe RCC and collecting duct RCC. Clear cell RCC (CCRCC) contributes for 85 percent of them[4]. Diagnoses based on clinical signs, imaging, renal biopsy and other factors are commonly used to identify this disease.

However, few patients with tiny renal masses (RMs) have no clinical abnormalities until late in the disease and 30% of them have distant metastases when they are identified with clear cell RCC with the worst prognosis in the urinary system[5,6]. As a result, a sensitive and accurate CCRCC diagnostic approach is urgently required.

Renal cell carcinoma is resistant to all chemotherapeutic drugs and intense radio treatments[7] due to which surgically removing the cancerous tissues remains the only effective treatment. RCC has also shown limited sensitivity towards targeted treatments or immunotherapy[8,9]. This disease has a complicated etiology, caused by the complex interactions among genes. Studies over the past decades have identified VHL (von Hippel-Lindau), p53, p16, p21 and p27 as the primary tumor suppressor genes in RCC. It is evident that loss of function of VHL and p53 is critically associated with cancer[10].

VHL dysfunction causes constitutively abnormal hypoxia response activation such as overexpression of vascular endothelial growth factor (VEGF) enhancing tumor formation and angiogenesis[11,12]. In RCC, p53 has been demonstrated to inhibit tumor development and promote cell death[13,14].

Despite this significant knowledge in RCC disease, the molecular pathways behind this illness remain unknown. Due to the limited knowledge about the disease's etiology and pathophysiology, there are currently no effective treatments for renal cancer. As a result, investigation into the molecular processes is needed to identify the causative factors and critical molecular markers of renal cancer to investigate novel therapeutic options for the treatment of this disease. Biological interaction networks provide an exciting

opportunity to understand the molecular mechanisms behind complex diseases, interactions between different proteins and identification of potential drug targets. In recent studies, human RCC profiles are also generated using network analysis[15,16].

In this study, we critically analyzed top 30 Differentially Expressed Genes (DEGs) identified by performing a comparative investigation of the datasets of renal cancer patients with the control sample from the GEO database. We then performed the enrichment analysis (Gene Ontology) of the gene sets. To group genes into their active pathways, we used Kyoto Encyclopedia of Genes and Genomes (KEGG) database. We filtered 10 hub genes as potential biomarkers of renal cancer by identifying the protein-protein associations of the DEGs using STRING database and Cytoscape software. Finally, we fit the Kaplan-Meier plots to perform the overall survival analyses of the hub genes using the ‘‘Renal cancer’’ database.

Material and Methods

Microarray data collection: For this study, three microarray datasets were obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo) with the query “Human Renal Cancer”. After a critical review, three gene expression profiles (GSE168845, GSE66270 and GSE781) were selected for gene expression analysis. Among them, GSE168845 is based on GPL21185 platform (Agilent-072363 SurePrint G3 Human GE3 860K Microarray 039494), GSE66270 is based on GPL570 ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array) and GSE781 is of platform GPL96 ([HG-U133A] Affymetrix Human Genome U133A Array).

Screening for Differentially Expressed Genes (DEGs): We used GEO2R web-based tool of NCBI (https://www.ncbi.nlm.nih.gov/geo/geo2r/) to screen the differentially expressed genes among 3 datasets analyzing the samples as renal cancer and normal tissues. We employed the Benjamini and Hochberg (False discovery rate) to calculate the adjusted p-values. Genes that satisfied the threshold |logFC| >1.0 and P-value <0.05 were shortlisted for further analysis. As the number of DEGs discovered is so large, so another cut of |logFC| value was set to find the most critical DEGs for upregulated genes |logFC| ≥ 4 and for downregulated |logFC| ≤ -4 was taken. The Venn diagram was created to find out common DEGs among the 3 expression profiles using an online tool (https://bioinformatics.psb.ugent.be/webtools/Venn/).

Functional Enrichment and Pathway Analysis: The Enrichr classification system (https://maayanlab.cloud/Enrichr) was used to do an enrichment analysis of the key differentially expressed genes that were significantly up and down regulated. Using the Enrichr classification system, the DEGs were categorized based on their molecular function (MF), biological process (BP) and cellular component (CC). Further, the pathway analysis was carried out using the KEGG database in Enricher. KEGG database is enriched with wide information about gene function, biological pathways, genomes and diseases.

Protein-Protein Interaction (PPI) network and identification of critical hub genes: To generate the PPI networks, the DEGs were processed via a web-based PPI network generation tool, the Search Tool for the Retrieval of Interacting Genes (STRING) database (https://string-db.org/). The topology of the PPI network was then examined using the Cytoscape program (http://www.cytoscape.org/). CytoHubba, a Cytoscape plugin program, was used to compute the degree of each protein node. Hub genes were defined as those with scores of 10 or more gene degrees in the PPI network.

Survival analysis of hub genes: The Kaplan–Meier plotter (http://kmplot.com/analysis/), a web-based tool was used to perform the survival analyses of each hub gene with 95% confidence of the hazard ratio (HR).

Results

Identification of DEGs: Among the three selected renal cancer microarray datasets (GSE168845, GSE66270, GSE781), GSE168845 comprised of 8 samples (4 control and 4 renal cancer), GSE66270 had 28 samples (14 control and 14 renal cancer) and GSE781 comprised of 17 samples (8 control and 9 renal cancer). GEO2R was used to compare and to screen the DEGs between control and renal cancer with the set threshold criteria of |logFC| >1.0 and P-value <0.05. The results obtained with these thresholds ≥ 4 for upregulated and of |logFC| ≤ -4 were downregulated. We screened 569 DEGs from GSE168845 which included 410 genes upregulated and 159 genes downregulated.

Similarly, for dataset GSE66270, 394 DEGs were identified which included 272 upregulated and 122 downregulated. And for dataset GSE781, a set of 69 DEGs containing 55 upregulated and 6 downregulated was identified. The volcano plots showing the DEGs from our three gene expression profiles are shown in figure 1 (A, B and C) respectively.

We performed Venn analysis to screen the common up regulated and down regulated DEGs among all three datasets shown in figure 2 (A and B). A set of 30 critical genes were found common amongst all three groups, of which 24 were significantly upregulated genes and 6 were downregulated (Table 1).

Functional enrichment analyses of DEGs: The enrichment analysis for GO function and KEGG pathway for DEGs was performed using the Enrichr. The GO terms with enrichment scores included molecular function (MF), biological process (BP) and cellular component (CC) ontologies. The results of gene ontology enrichment analyses revealed that the DEGs are mainly enriched in BP including sodium-independent organic anion transport, negative regulation of lipase

activity, transmembrane transport, monovalent inorganic anion homeostasis, organic substance transport and triglyceride catabolic process (Figure 3A).

In terms of molecular functions, DEGs are enriched in monovalent inorganic cation transmembrane transporter activity, ion antiporter activity, anion: anion antiporter activity, inorganic anion exchanger activity, secondary active transmembrane transporter activity and sodium ion transmembrane transporter activity (Figure 3B). In CC, the DEGs are enriched in platelet alpha granule, platelet alpha granule lumen, secretory granule lumen, actin-based cell projections, very low density lipoprotein particle and vacuolar proton- transporting V- type ATPase complex (Figure 3C). Our results highlight the active involvement of most of the DEGs in transport mechanism, cell signaling and binding.

The results of KEGG analysis reveal the active participation of our DEGs in the PPAR signalling pathway, compliments and coagulation cascade, collecting duct caid secretion, cholesterol metabolism and tyrosine metabolism (Figure 3D).

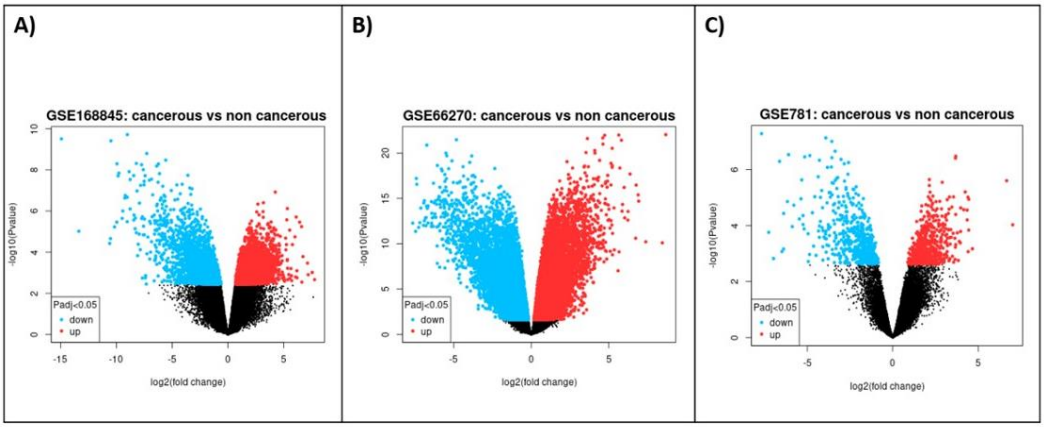

Figure 1: DEGs between RC samples and normal samples.
(A) The volcano plot for DEGs in GSE168845 data. (B) The volcano plot for DEGs in GSE66270 data.
(C) The volcano plot for DEGs in GSE781 data. X-axes index the log fold change and Y-axes index the –log (P-value).
The red dots represent upregulated genes screened based on fold change > 1.0 and adjusted P value of < 0.05.
The blue dots represent downregulated genes screened based on fold changes > 1.0 and adjusted P value of < 0.05.
The black dots represent genes with no significance

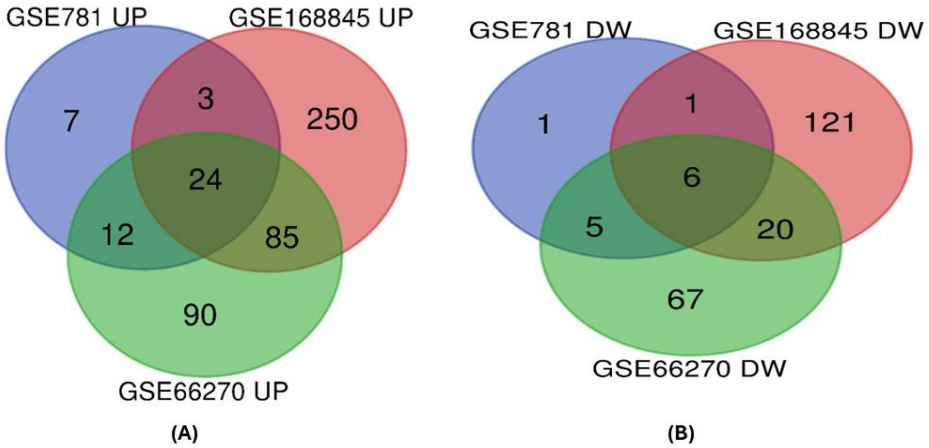

Figure 2: Venn diagram representing the overlaps between three GEO datasets.
(A) Venn diagram illustrating overlapping upregulated genes in GSE168845, GSE66270 and GSE781 dataset.
(B) Venn diagram illustrating overlapping downregulated genes in GSE168845, GSE66270 and GSE781 dataset.

Table 1
Screened DEGs in renal cancer by integrated microarray

Upregulated	RALYL PLG SLC22A8 HPD UMOD CALB1 DIO1 KCNJ1 CLDN8
	OLFM4 FABP1 G6PC NPHS2 ALB SLC12A1 ALDOB SLC22A6
	RHCG ATP6V1B1 SERPINA5 CLCNKB KNG1 SLC34A1
Downregulated	TNFAIP6 NDUFA4L2 NPTX2 FABP7 ANGPTL4 APOC1

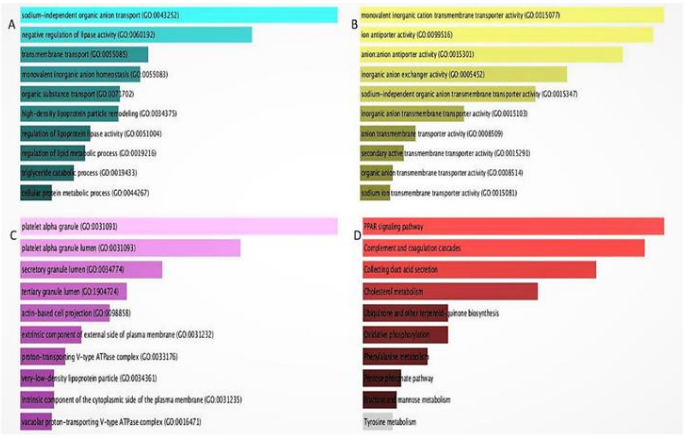

Figure 3: GO term and KEGG pathway enrichment analyses performed using Enrichr on DEGs identified from GC samples and normal samples. (A) The top 10 enriched biological process for DEGs. The horizontal axis represents the number of genes and Y-axis represents biological process. (B) The top 10 enriched molecular function for DEGs. The horizontal axis represents the number of genes and Y-axis represents molecular function. (C) The top 10 enriched cellular component for DEGs. The horizontal axis represents the number of genes and Y-axis represents cellular component. (D) The top 10 enriched KEGG pathway for DEGs. The horizontal axis represents the number of genes and Y-axis represents KEGG pathway.

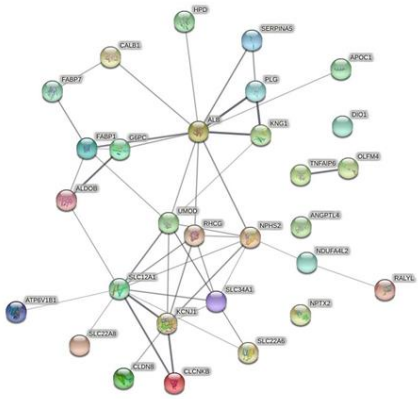

Figure 4: STRING protein-protein interaction network of 24 upregulated and 6 downregulated genes. The network includes 29 nodes and 46 edges. Circles represent genes, lines represent the interaction of proteins between genes and the results within the circle represent the structure of protein. Line colour represent evidence of the interaction between the proteins.

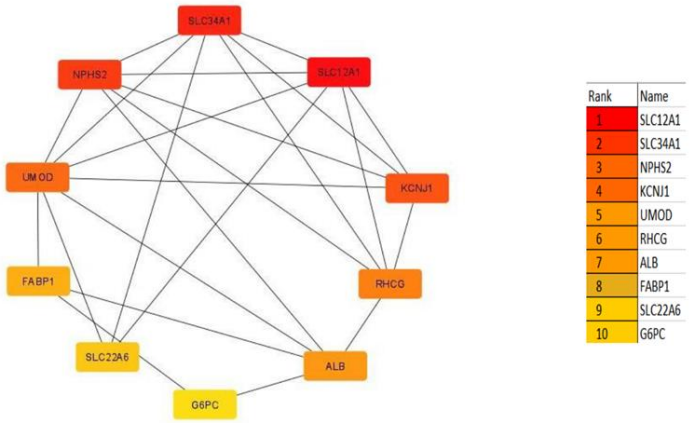

Figure 5: Subnetwork of top ten hub genes from protein-protein interaction network using Cytoscape software. Node colour reflects degree of connectivity. The pseudocolour scale from red to yellow represents and top nine hub rank from 1-10. Red colour represents highest degree and orange colour represents intermedia degree and yellow colour represents lowest degree

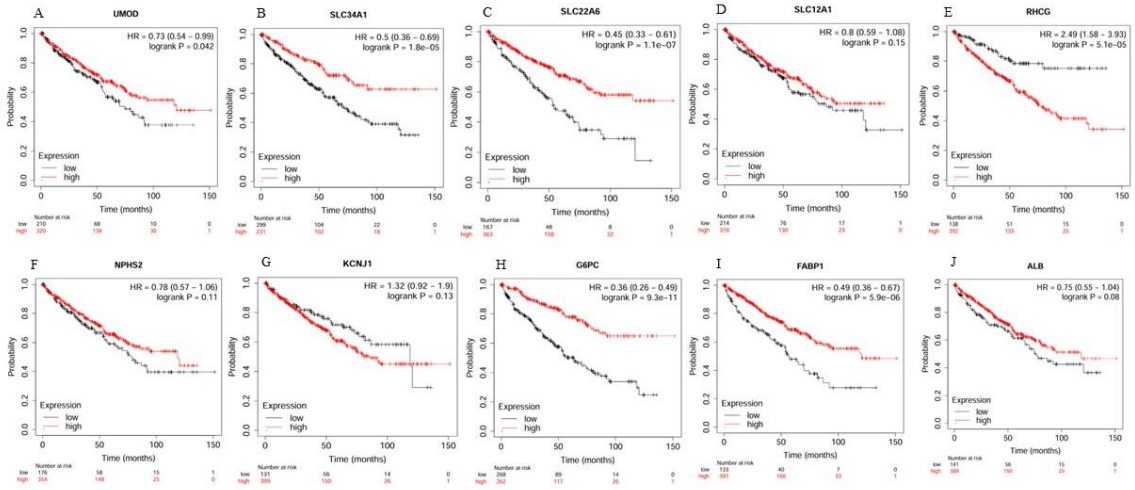

Figure 6: Kaplan-Meier overall survival analysis for the top nine hub genes expressed in GC patient's samples, Kaplan-Meier plot of overall survival in subjects with low versus high. (A) UMOD, (B) SLC34A1, (C) SLC22A6, (D) SLC12A1, (E) RHCG, (F) NPHS2, (G) KCNJ1, (H) G6PC, (I) FABP1 and (J) ALB mRNA expression

Protein-protein interaction network and critical hub gene identification: Protein interactions and associations among the different differentially expressed genes were identified using the web-based STRING-DB tool. A complex interaction network of 29 nodes representing proteins and 46 edges representing interactions was constructed as represented in figure 4. We then evaluated the top ten genes ranking them by their connectivity degree in the protein-protein interaction network using Cytohubba, as presented in figure 5.

The results showed that Solute Carrier Family 12 Member 1 (SLC12A1) gene had the highest connectivity degree = 59, followed by sodium-dependent phosphate transport protein 2A(SLC34A1 degree = 54), Stomatin Family Member, Podocin (NPHS2 degree = 53), potassium 58 Inwardly Rectifying Channel Subfamily J Member 1 (KCNJ1 degree= 51), Uromodulin (UMOD degree = 36), Rhesus Blood Group Family Type C Glycoprotein (RHCG degree = 26), albumin (ALB degree = 17), Fatty Acid Binding Protein 1 (FABP1 degree = 7), Solute Carrier Family 22 Member 6 (SLC22A6 degree = 6) and Glucose-6-Phosphatase Catalytic Subunit 1 (G6PC degree = 4). All these genes were observed to be upregulated in case of our renal cancer gene expression

Survival analysis of hub genes: The prognostic values of the 10 hub genes were further evaluated by performing the survival analysis (Figure 6) on the available dataset of 530 RC patients on the Kaplan–Meier plotter platform. The survival curves of two critical hub genes KCNJ1 and RHCG, were showing adverse effects on overall survival in GC patients. Also, four hub genes namely, ALB, NPHS2, SLC12A1 and UMOD were not associated with over survival in renal cancer patients. Furthermore, four significant up-regulated genes namely, FABP1, G6PC, SLC22A6 and SLC34A1, were assessed to be strongly linked with favourable survival in renal cancer patients.

Discussion

Investigation of critical biomarkers of complex disease such as cancer is important for diagnosis and treatment[17]. In this study, we performed analysis of 3 high-throughput gene expression datasets of RC versus normal human renal tissue and screened a set of 30 DEGs (24 upregulated DEGs and 6 downregulated DEGs). We enriched the DEGs among different ontologies based on their molecular function, biological process and cellular component. It was evident that most of the DEGs were actively participating in sodium-independent organic anion transport, negative regulation of lipase activity, transmembrane transport, monovalent inorganic anion homeostasis, organic substance transport and triglyceride catabolic process.

The distribution of DEGs across KEGG pathway majorly includes the signaling pathways such as PPAR signaling pathway, compliments and coagulation cascade, collecting duct caid secretion, cholesterol metabolism and tyrosine metabolism. A PPI network of the DEGs was constructed to study the correlations among them. The ten most interconnected genes were screened including UMOD, SLC34A1, SLC22A6, SLC12A1, RHCG, NPHS2, KCNJ1, G6PC, FABP1 and ALB which were found to be up regulated in RC patients.

Also, we assessed the effect of these hub genes on survival of RC patients KCNJ1 and RHCG which was associated to a poor prognosis in RC patients. In clear cell renal cell[18].

By inhibiting NF-B Signaling and MMP1 Expression, RHCG is reported in reducing Tumorigenicity and Metastasis in Esophageal Squamous Cell Carcinoma[19].

Upregulation of FABP1, G6PC, SLC22A6 and SLC34A1 was found to be favourable prognostic indicators in RC[22,23].

The genes SLC22A6 and SLC34A1 were among the top ten genes in terms of connectivity. The soluble carrier (SLC) family encodes passive transporters, ion coupled transporters and exchanger genes that play a significant role in cell metabolism. Elevated expression of nutrient transporter proteins is associated with glucose transport for the aggressive and highly proliferating malignant cancers[24]. The SLC34A1 of this family is a sodium-dependent Pi Cotransporter Involved in multiple diseases[21]. Our results identified several associations of SLC gene expression with prognosis of RC patients, indicating that these genes may represent possible oncogenes that could serve as therapeutic targets of RC.

Conclusion

Using an integrated bioinformatics method, we discovered that 10 hub genes are important in renal cancer. We built a complex PPI network of 29 shortlisted DEGs with 46 key connecting linkages. The most highly enriched sub-networks and hub genes were discovered to be involved in biological process including sodium-independent organic anion transport, negative regulation of lipase activity, transmembrane transport, monovalent inorganic anion homeostasis, organic substance transport and triglyceride catabolic process.

These findings aid our understanding of the cause and molecular events that contribute to renal cancer. These potential gene targets could be leveraged to develop efficient diagnostics and therapies.

References

1. Angulo J.C. et al, The role of epigenetics in the progression of clear cell renal cell carcinoma and the basis for future epigenetic treatments, *Cancers*, https://doi.org/10.3390/cancers13092071, **13(9)**, 2071 **(2021)**

2. Barata P.C. and Rini B.I., Treatment of renal cell carcinoma: Current status and future directions, *CA: A Cancer Journal for Clinicians*, **67(6)**, 507–524, https://doi.org/10.3322/caac.21411 **(2017)**

3. Cairns P., Renal cell carcinoma, *Cancer Biomarkers*, **9(1-6)**, 461–473, https://doi.org/10.3233/CBM-2011-0176 **(2011)**

4. Cohen H.T. and McGovern F.J., Renal-cell carcinoma, *The New England Journal of Medicine*, **353(23)**, 2477–2490, https://doi.org/10.1056/NEJMra043172 **(2009)**

5. Costa L.J. and Drabkin H.A., Renal cell carcinoma: New developments in molecular biology and potential for targeted therapies, *The Oncologist*, **12(12)**, 1404–1415, https://doi.org/10.1634/theoncologist.12-12-1404 **(2007)**

6. Goossens N., Nakagawa S., Sun X. and Hoshida Y., Cancer biomarker discovery and validation, *Translational Cancer Research*, **4(3)**, 256–269, https://doi.org/10.3978/j.issn.2218-676X.2015.06.04 **(2015)**

7. Guo Z. et al, KCNJ1 inhibits tumor proliferation and metastasis and is a prognostic factor in clear cell renal cell carcinoma, *Tumor Biology*, **36(2)**, 1251–1259, https://doi.org/10.1007/s13277-014-2746-7 **(2015)**

8. Hsieh J.J. et al, Renal cell carcinoma, *Nature Reviews Disease Primers*, **3**, 17009, https://doi.org/10.1038/nrdp.2017.9 **(2017)**

9. Jemal A. et al, Cancer statistics, *CA: A Cancer Journal for Clinicians*, **57(1)**, 43–66, https://doi.org/10.3322/canjclin.57.1.43 **(2007)**

10. Kang W. et al, The SLC family are candidate diagnostic and prognostic biomarkers in clear cell renal cell carcinoma, *BioMed Research International*, https://doi.org/10.1155/2020/1932948, 1932948 **(2020)**

11. Ku B.M. et al, Transglutaminase 2 inhibition found to induce p53-mediated apoptosis in renal cell carcinoma, *The FASEB Journal*, **27(9)**, 3487–3495, https://doi.org/10.1096/fj.12-22422 **(2013)**

12. Kuo M.H. et al, Glucose transporter 3 is essential for the survival of breast cancer cells in the brain, *Cells*, **8(12)**, 1568, https://doi.org/10.3390/cells8121568 **(2019)**

13. Li W. et al, Microarray profiling of human renal cell carcinoma: Identification of potential biomarkers and critical pathways, *Kidney and Blood Pressure Research*, **37(6)**, 506–513, https://doi.org/10.1159/000355726 **(2013)**

14. Lin L., Yee S.W., Kim R.B. and Giacomini K.M., SLC transporters as therapeutic targets: Emerging opportunities, *Nature Reviews Drug Discovery*, https://doi.org/10.1038/nrd4626, **14(8)**, 543–560 **(2015)**

15. Low G. et al, Review of renal cell carcinoma and its common subtypes in radiology, *World Journal of Radiology*, **8(5)**, 484–500, https://doi.org/10.4329/wjr.v8.i5.484 **(2016)**

16. Ming X.Y. et al, RHCG suppresses tumorigenicity and metastasis in esophageal squamous cell carcinoma via inhibiting NF-κB signaling and MMP1 expression, *Theranostics*, **8(1)**, 185–198, https://doi.org/10.7150/thno.21383 **(2018)**

17. Pelletier J. et al, The asparaginyl hydroxylase factor-inhibiting HIF is essential for tumor growth through suppression of the p53-p21 axis, *Oncogene*, https://doi.org/10.1038/onc.2011.471, **31(24)**, 2989–3001 **(2012)**

18. Razafinjatovo C.F. et al, VHL missense mutations in the p53 binding domain show different effects on p53 signaling and HIFα degradation in clear cell renal cell carcinoma, *Oncotarget*, **8(6)**, 10199–10210, https://doi.org/10.18632/oncotarget.14372 **(2017)**

19. Russo P., Contemporary understanding and management of renal cortical tumors, *Urologic Clinics of North America*, **35(4)**, 593–604, https://doi.org/10.1016/j.ucl.2008.07.016 **(2008)**

20. Tomar Nitika, Sharma Vivek and Tomar Sapna, Chemical reduction method-based copper nanoparticle synthesis and it's characterization, *Res. J. Chem. Environ*., **28(6)**, 37-44 **(2024)**

21. Wagner C.A., Hernando N., Forster I.C. and Biber J., The SLC34 family of sodium-dependent phosphate transporters,

Pflugers Archiv: European Journal of Physiology, **466**, 139–153, https://doi.org/10.1007/S00424-013-1418-6 **(2014)**

22. Wu J. et al, High expression of CD39 is associated with poor prognosis and immune infiltrates in clear cell renal cell carcinoma, *OncoTargets and Therapy*, https://doi.org/10.2147/OTT.S272553, **13**, 10453–10464 **(2020)**

23. Yang D.C. and Chen C.H., Potential new therapeutic approaches for renal cell carcinoma, *Seminars in Nephrology*, **40(1)**, 86–97, https://doi.org/10.1016/j.semnephrol.2019.12.010 **(2020)**

24. Zou X., Guo B., Ling Q. and Mo Z., Toll-like receptors serve as biomarkers for early diagnosis and prognosis assessment of kidney renal clear cell carcinoma by influencing the immune microenvironment: Comprehensive bioinformatics analysis combined with experimental validation, *Frontiers in Molecular Biosciences*, **9**, 24, https://doi.org/10.3389/fmolb.2022.832238 **(2022)**.

(Received 15th February 2025, accepted 18th April 2025)