Biomarkers associated with cell-in-cell structure in kidney renal clear cell carcinoma based on transcriptome sequencing

Zehua Wang[1] and Zhongxiao Zhang[2]

[1] Department of Urology, Qilu Hospital, Shandong University, Jinan, China
[2] Department of Urology, Qilu Hospital (Qingdao), Shandong University, Qingdao, China

ABSTRACT

Background. Kidney renal clear cell carcinoma (KIRC), the main histological subtype of renal cell carcinoma, has a high incidence globally. Cell-in-cell structures (CICs), as a cellular biological phenomenon, play pivotal roles in cell competition, immune evasion and tumor progression in the context of KIRC.

Methods. Data for this study were sourced from The Cancer Genome Atlas (TCGA), International Cancer Genome Consortium (ICGC), and Gene Expression Omnibus (GEO) databases. Differentially expressed genes (DEGs) were identified using the limma package. Enrichment analyses were performed using the clusterProfiler package. Support vector machine-recursive feature elimination (SVM-RFE) and Least Absolute Shrinkage and Selection Operator (LASSO) regression, implemented *via* the caret and glmnet packages in R, were used to select biomarkers. The accuracy of these biomarkers was verified by using the receiver operating characteristic (ROC) curve as well as *in vitro* experiments (CCK-8 assay, wound healing assay, Transwell assay, and quantitative real-time PCR). The CIBERSORT algorithm was applied to explore the association between immune infiltration and the biomarkers. Further analysis explored the association between these biomarkers and clinicopathological characteristics of KIRC. For single-cell data, the Seurat package is used to read the sample data, and the SCTransform function is employed for normalization.

Results. This study identified 1,256 DEGs which enriched in T-cell immune system regulation processes. Five hub genes (*CDKN2A*, *VIM*, *TGFB1*, *CTSS*, and *CDC20*) were biomarkers with area under the curve (AUC) values > 0.8, indicating high predictive performance. *In vitro* validation experiments demonstrated that the expressions of all five biomarkers in KIRC cells were elevated, and the knockdown of *CTSS* could inhibit the migration and invasion of KIRC cells. Immune infiltration analysis showed higher proportions of T-cells and macrophages in tumor tissues. *CDKN2A* and *CDC20* expressions correlated significantly with stage and grade, while *TGFB1*, *CDKN2A*, and *CDC20* were highly expressed in proliferative tumor cells.

Conclusion. This study provides new biomarkers for KIRC, offering valuable insights into its developmental mechanisms for the research of CIC in this disease.

INTRODUCTION

Renal cell carcinoma (RCC) is the most frequent malignancy of the kidney (*Makhov et al., 2018a*). Globally, there are approximately 434,419 new cases each year, accounting for about 2.2% of all cancers and ranking 14th (*Bray et al., 2024*; *Tian et al., 2024*). Its mortality rate is approximately 1.6%, placing it 16th among all cancers (*Bray et al., 2024*). Smoking, overweight, and obesity are established risk factors for kidney cancer (*Frew & Moch, 2015*). Kidney renal clear cell carcinoma (KIRC) is currently the main histological subtype of kidney cancer, accounting for about 80–90% of kidney cancer patients, but it has a poor prognosis (*Wang et al., 2019*). Studies have reported that KIRC exhibits heterogeneity in clinical pathology, molecular, and cellular aspects (*Xie et al., 2020*). However, due to the limited availability of biomarkers for early detection and prognosis prediction, the prognosis for KIRC patients is generally poor (*Cui et al., 2020*). Hence, it is urgent to investigate the pathogenesis of KIRC and explore new molecular biomarkers for diagnosis and prognosis (*Seyfinejad & Jouyban, 2022*; *Liu et al., 2024*; *Zhang et al., 2023*).

Cell-in-cell structures (CICs) represent a cellular biological phenomenon where one living cell is enclosed within another living cell (*Wang et al., 2020*). CICs were initially discovered in the context of tumor biology and are considered to play pivotal roles in cell competition, immune evasion, and tumor progression (*Fais & Overholtzer, 2018*; *Huang, Chen & Sun, 2015*). CICs exert various effects on cellular behavior and the functions of both external and internal cells, encompassing cell death, cell proliferation, and immune regulation (*Wang, 2015*). Increasing evidence suggests that CICs may hold prognostic and diagnostic value for cancer patients (*Su et al., 2022*; *Chen et al., 2013*). It has been proposed that tumor cells may utilize CIC-mediated internal cell death as a means of immune evasion (*Sun & Chen, 2022*). Consequently, research on CICs in renal cancer is of great significance for understanding tumor biology and developing novel therapeutic strategies.

Based on the above background, this study aimed to mine molecular biomarkers for the diagnosis and prognosis of KIRC based on CICs-related genes, and establish a corresponding diagnostic model. Additionally, we conducted in-depth analyses of the correlation between these biomarkers and immune infiltration, their relationship with the clinicopathological characteristics of KIRC, and their distribution and expression in the kidney. The significance of this study lies not only in providing new scientific evidence for early diagnosis, prognosis assessment, and personalized treatment of KIRC, but also in offering new possibilities for exploring the pathogenesis of renal cancer and developing novel therapeutic strategies.

MATERIALS AND METHODS

Data acquisition

The current research encompasses datasets from four aspects. Firstly, RNA-Seq data for The Cancer Genome Atlas (TCGA)-KIRC were downloaded using the TCGA Genomic Data Commons (GDC) application programming interface (API). The FPKM values were converted to TPM and then log2-transformed, with a total of 513 primary tumor samples and 72 adjacent normal control samples retained. Secondly, expression profiles for The Renal Cell Cancer-European Union (RECA-EU)/Renal cell carcinoma were obtained from the International Cancer Genome Consortium (ICGC) database, encompassing 91 primary tumor samples and 45 adjacent normal control samples. Furthermore, the KIRC single-cell dataset GSE224630 was obtained from the Gene Expression Omnibus (GEO) database. The GSE224630 dataset includes tumor samples from 6 patients with untreated clear cell renal cell carcinoma. Finally, 101 CIC-correlated genes were sourced from previous literature (*Song et al., 2022*; *Ren et al., 2024*).

Identification and enrichment analysis of DEGs

The limma package (*Ritchie et al., 2015*) was utilized in the TCGA-KIRC and ICGC datasets to screen for DEGs between KIRC patients and normal controls, and to identify common upregulated genes across both datasets (*Song et al., 2023a*). The criteria for selecting DEGs were a $p < 0.05$ and $|\log2FC| > 1$. Additionally, the clusterProfiler package in R (*Yu et al., 2012*) was employed to enrich the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) of these DEGs. Adjusted $p < 0.05$ indicated significantly enriched pathways (*Song et al., 2023b*).

Machine learning for key gene selection, diagnostic model development, and validation

The overlapping genes obtained from the intersection of DEGs and CIC genes were further screened using machine learning techniques. The rfe function from the R package caret (*Kuhn, 2008*) was employed, utilizing the svmlinear method. During the selection of the number of features through recursive feature elimination (RFE), the cross-validation (CV) accuracy of the support vector machine (SVM) model was used to screen for key disease genes. Feature selection was also conducted using Least Absolute Shrinkage and Selection Operator (LASSO) regression from the glmnet package (*Friedman et al., 2021*), with key parameters set to nfolds = 10 and family = 'binomial'. Finally, the feature genes selected by both the SVM-RFE and LASSO methods were intersected to obtain the biomarkers for this study. Subsequently, the R package e1071 (*Meyer et al., 2019*) was used to construct a diagnostic model using the SVM method. The accuracy of the biomarkers selected by machine learning was tested by plotting receiver operating characteristic (ROC) curves for KIRC samples and normal controls in the TCGA-KIRC training set. A larger the area under the curve (AUC) indicated a higher accuracy of considering the genes as hub genes. The validity of these genes was further verified using the ICGC validation set using the same method.

Correlation analysis of biomarkers with immune infiltration

The CIBERSORT package (*Newman et al., 2015*) was employed to quantify different immune cells in KIRC samples and control samples. Spearman correlation coefficients were used to perform correlation analyses between biomarkers and immune cells.

Processing and analysis of scRNA-seq data from KIRC

The Read10X function from the Seurat package (*Hao et al., 2024*) was utilized to read the scRNA-seq data for each sample in GSE224630, retaining cells with a gene count between 200 and 5000 and a mitochondrial gene proportion of less than 10%. Subsequently, the SCTransform function (*Hao et al., 2024*) was applied for normalization. After Principal Component Analysis (PCA) dimensionality reduction, the harmony package (*Korsunsky et al., 2019*) was used to remove batch effects among different samples. The RunTSNE function (*Hao et al., 2024*) was then employed for t-Distributed Stochastic Neighbor Embedding (TSNE) dimensionality reduction. Finally, the FindNeighbors and FindClusters functions (*Hao et al., 2024*) were used for clustering, with parameters set to dims = 1:25 and resolution = 0.1. Cell subpopulations were annotated based on marker genes provided by the CellMarker2.0 database (*Hu et al., 2022*). The expression patterns of the biomarkers obtained in this study across different cell types were subsequently investigated.

Cell culture and transfection

Two cell lines, human embryonic kidney 293T (CRL-3216) and human renal clear cell adenocarcinoma cell 786-O (CRL-1932), were all obtained from the American Type Culture Collection (Manassas, MD, USA). These cells were cultured in DMEM (11965092, Gibco, Waltham, MA, USA) or RPMI 1640 medium (11875093, Gibco, Waltham, MA, USA) with the supplementation of 10% fetal bovine serum (FBS) (S9020, Solarbio Lifesciences, Beijing, China) and 1% penicillin-streptomycin (15140148, Gibco, Waltham, MA, USA). All cells were tested *via* short tandem repeat profiling and incubated in the incubator at 37 °C with 5% CO_2.

For the liposome transfection, the small interfering RNA against *CTSS* (si-*CTSS*) and the control small interfering RNA (si-NC) were all purchased from GenePharma (Shanghai, China) and transfected into 786-O cells with the use of lipofectamine 2000 transfection reagent (11668027, Invitrogen, Carlsbad, CA, USA) as per the manuals. The sequences applied for the transfection were 5′-TCACATATAAGTCAAACCCTA-3′.

CCK8 assay

During the logarithmic phase, 786-O cells were plated in a 96-well plate with a density of 1×10^4 cells per well and incubated at 37 °C in an atmosphere containing 5% CO_2 for 0, 24, or 48 h. Subsequently, 10 μL of CCK-8 reagent was added to the medium, and the samples were incubated at 37 °C for 2 h. To generate the CCK-8 curve, the absorbance was measured at 450 nm, which served as the ordinate, while time was represented on the abscissa. The results were averaged from three independent experimental repeats.

Cell migration assay

The transfected 786-O cells (5×10^5 cells/well) were cultivated in a six-well plate with serum-free media. When they achieved complete confluence, a 200-μL sterile pipette tip

was used to create an artificial scratch on the monolayer. Forty-eight h later, the cells were photographed by an inverted optical microscope (DP27, Olympus, Tokyo, Japan), and the wound closure (%) was quantified correspondingly to assess the migration of KIRC cells (*Zhang et al., 2024*). Wound closure (%) = (Initial scratch width–scratch width at measurement time point)/Initial scratch width ×100%.

Cell invasion assay

For the invasion assay, 786-O cells (1×10^5/100 µL) were suspended in 200 µL serum-free medium and plated in the upper Transwell chamber (3422, Corning, Inc., Corning, NY, USA) coated with matrix gel (C0372, Beyotime, China), while the lower chamber was filled with 700 µL culture media containing 10% bovine calf serum. After 48 h, the invaded cells were fixed by 4% paraformaldehyde (P0099, Beyotime, Shanghai, China) and stained with 0.1% crystal violet (C0121, Beyotime, Shanghai, China) for 30 min. Then, three random fields were observed under an inverted optical microscope (DP27, Olympus, Japan), and the number of invaded cells was quantified (*Wang et al., 2023*).

QRT-PCR experiment

Following the instructions, total RNA was isolated from 293T and 786-O cells using the TriZol total RNA extraction kit (15596026, Invitrogen, Carlsbad, CA, USA). Subsequently, the concentration of the isolated RNA was determined. Then, complementary DNA was synthesized by reverse transcription with a relevant assay kit (D7178S, Beyotime, Shanghai, China). After that, SYBR Green qPCR Mix (D7260, Beyotime, Shanghai, China) was used for the PCR assay according to the protocols. Finally, the relative level was calculated by the $2^{-\Delta\Delta CT}$ method with *GAPDH* as the reference gene. The qRT-PCR primers used in this study were designed according to National Center for Biotechnology Information (NCBI) sequences using Primer Premier 6 software. The sequences of the primers used were presented in Table S1.

Statistical analysis

All statistical analysis was conducted using R (version 3.6.0). The *t*-test was performed on continuous variables between two groups. Data normality was assessed using the Shapiro–Wilk test. If normality assumptions were violated, the Wilcoxon rank-sum test was performed as a supplementary analysis. To investigate the correlation between gene expression and immune cell fractions, Spearman's rank correlation test was employed. A $p < 0.05$ signified a significant level. Sangerbox (http://sangerbox.com/) provided assistance with this study (*Shen et al., 2022*).

RESULTS

Differential gene selection and enrichment analysis

Comprehensive differential gene expression analysis was conducted on tumor and control samples within the TCGA-KIRC and ICGC datasets (Figs. 1A–1B). Subsequently, upregulated genes common to both the TCGA and ICGC datasets were identified, totaling 1,256 (Fig. 1C). Functional enrichment analysis, including KEGG pathway analysis and GO

analysis, was performed on these commonly upregulated genes. The top 10 enriched KEGG pathways included Human T-cell leukemia virus 1 infection, PI3K-Akt signaling pathway, Cytokine-cytokine receptor interaction, Phagosome, Epstein-Barr virus infection, and cell adhesion molecules (CAMs) among others (Fig. 1D). The results of the GO functional enrichment analysis covered biological process (BP), cellular component (CC), and molecular function (MF) aspects. The top five GO functional enrichments indicated that these genes were significantly involved in critical processes related to immune system regulation, such as regulation of lymphocyte activation, MHC protein complex, regulation of T cell activation, peptide antigen binding, T cell activation, regulation of leukocyte activation, and leukocyte cell–cell adhesion (Fig. 1E).

Machine learning for biomarkers screening

Twelve overlapping genes were obtained by intersecting the DEGs with CIC genes. When selecting the number of features using the RFE method, the trend of the CV accuracy of the SVM model with the number of features can be observed. It can be seen that when the number of selected features is six, the model's CV accuracy reaches its maximum (Fig. 2A). To further screen the genes, LASSO regression analysis was employed. Figure 2B displayed the changes in regression coefficients for different gene features in LASSO regression as the penalty parameter (λ) varies, showing the deviance at the optimal λ value determined by 10-fold CV. By intersecting the genes obtained from both the RFE-SVM and LASSO methods, five hub genes were ultimately identified as biomarkers in this study, namely *CDKN2A*, *VIM*, *TGFB1*, *CTSS*, and *CDC20* (Fig. 2C). Figures 2D–2E demonstrated the differential expression of these hub genes between tumor samples and normal controls, showing that all biomarkers are expressed higher in the tumor group than the control group in both the TCGA training set and the ICGC validation set ($p < 0.0001$).

Construction and verification of the diagnostic model

The five biomarkers were integrated to establish a diagnostic model, and the predictive ability of this model in the TCGA training set was assessed using ROC curves. Each curve represents a hub gene, and the AUC value was used to explore the gene's ability to distinguish between diseases. The results showed that the AUC > 0.8 (Fig. 3A), indicating that these genes had high predictive power. Figure 3B displayed the ROC curve for the overall predictive ability of the constructed SVM model, with an AUC of 0.962, suggesting that the model exhibits very high predictive performance in distinguishing between KIRC and normal samples. The confusion matrix of the classification model demonstrated the classification results for the tumor group and the control group (Fig. 3C), further validating the robustness of the model in distinguishing between the two groups of samples. Subsequently, the same method was used for validation in the ICGC validation set. Both the AUC values for individual genes and the overall AUC value for the SVM model were above 0.8, indicating that the diagnostic model in this study has high predictive performance (Figs. 3D–3F).
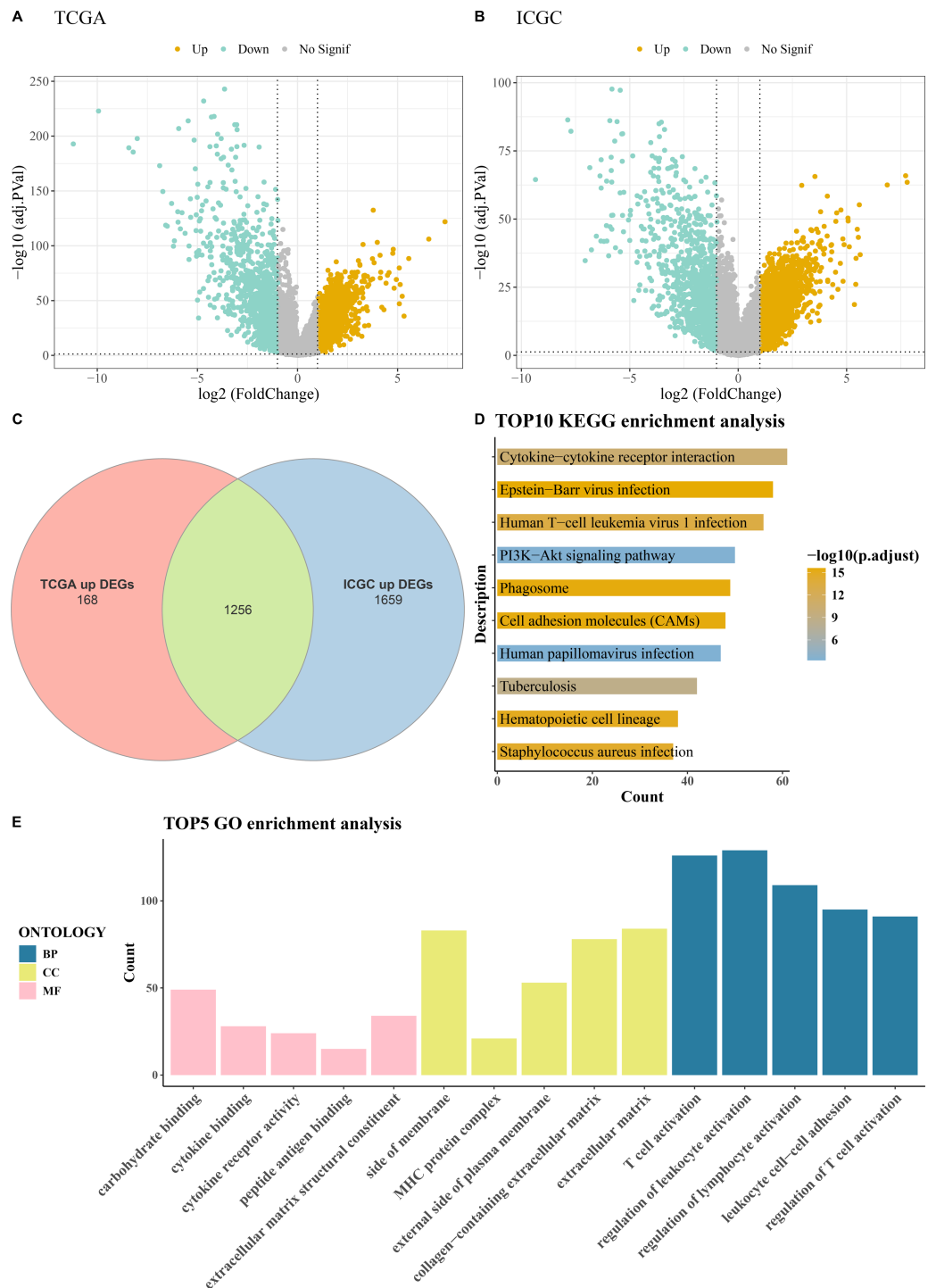
Figure 1 Acquisition and enrichment analysis of differential genes. (A) Volcano plot of DEGs in the TCGA cohort; (B) Volcano plot of DEGs in the ICGC cohort; (C) Venn diagram of upregulated genes common to both the TCGA and ICGC cohorts; (D) KEGG pathway enrichment analysis of DEGs; (E) GO enrichment analysis of DEGs in terms of BP, CC, and MF.
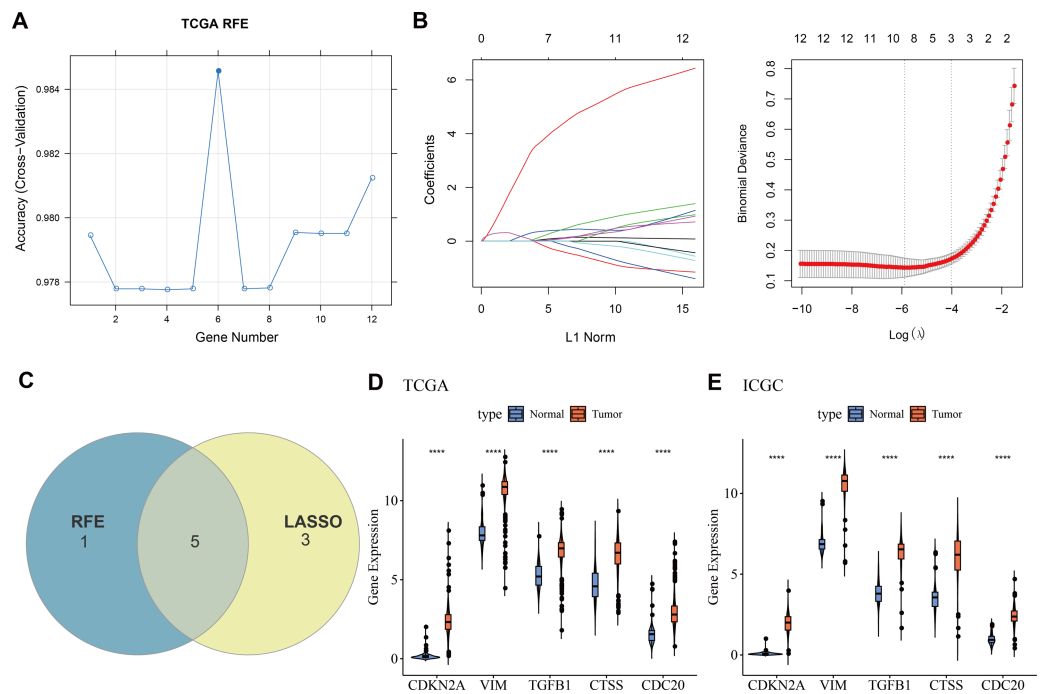
Full-size 🖾 DOI: 10.7717/peerj.19246/fig-1

Figure 2 Machine learning for biomarkers screening. (A) The curve of CV accuracy varying with the number of features selected by the RFE method; (B) The changes in regression coefficients of gene features in the LASSO regression model and the optimal penalty parameter (λ) determined through CV. The red dashed line indicates the selected optimal λ value, which corresponds to a relatively small number of features while ensuring good predictive performance of the model; (C) A Venn diagram showing the intersection of feature genes selected by both the RFE-SVM and LASSO methods; (D) The expression levels of feature genes in the tumor group *versus* the control group within the TCGA training set, with **** indicating $p < 0.0001$; (E) The expression levels of feature genes in the tumor group *versus* the control group within the ICGC validation set, with **** indicating $p < 0.0001$.

Full-size 🖼 DOI: 10.7717/peerj.19246/fig-2

Correlation between biomarkers and immune infiltration

To investigate the relationship between immune cells and key biomarkers, Figs. 4A–4B presented the distribution of different types of immune cells in normal and KIRC tissues. Overall, immune cells such as macrophages (M0, M1, M2) and T-cells (including CD8$^+$ T-cells, regulatory T-cells), and others were significantly more abundant in tumor tissues compared to normal tissues ($p < 0.05$), while B-cells (B cells naive) were more prevalent in normal tissues ($p < 0.05$). These differences showed slight variations between the TCGA training set and the ICGC validation set, for example, there was a significant difference in M2 macrophages between the tumor group and the control group in the TCGA training set, but not in the ICGC validation set. However, the overall results were consistent. Finally, the correlation between hub genes and immune cell scores in tumor samples was calculated and presented using correlation coefficients (ranging from −1 to 1) and significance levels (indicated by asterisks). The results indicated that these genes have significant correlations with multiple types of immune cells, and these correlations were particularly concentrated in T-cells and macrophages (Figs. 4C–4D). Among them, T cells CD4 memory activated
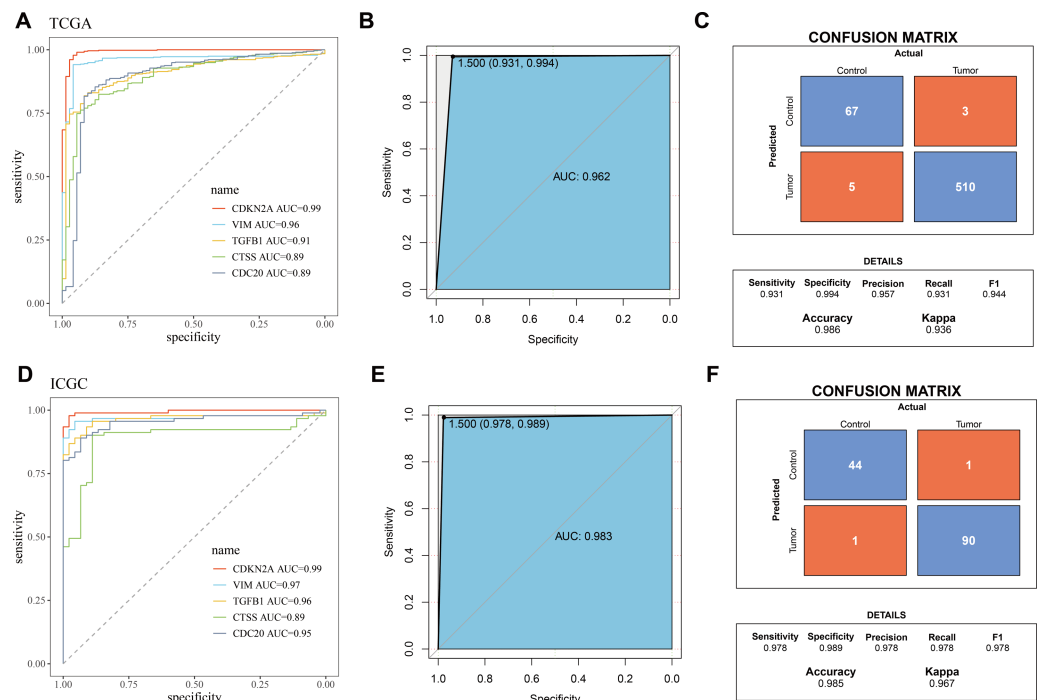
Figure 3 **Validation of biomarkers and diagnostic model.** (A) ROC curves for hub genes' expression in TCGA; (B) ROC curve for overall predictive performance of the model in TCGA; (C) Confusion matrix showing classification results for tumor group and control group in TCGA; (D) ROC curves for hub genes' expression in ICGC; (E) ROC curve for overall predictive performance of the model in ICGC; (F) Confusion matrix showing classification results for tumor group and control group in ICGC.

Full-size 🖼 DOI: 10.7717/peerj.19246/fig-3

and T cells CD4 memory resting showed significant differences across 5 biomarkers in the TCGA training set ($p < 0.05$).

Relationship between hub genes and clinicopathological characteristics of KIRC

Next, we observed at the expression levels of the five hub genes in relation to clinical stage and grade. As shown in Fig. 5A, the expression levels of *CDKN2A* were significantly different in different stages and grades ($p < 0.01$). The expression of *VIM* and *CTSS* did not change significantly across different stages or grades (Figs. 5B, 5D). The trend in the expression level of the *TGFB1* gene across different stages and grades was not significant, although there was a slight increase in Stage III and IV, it did not reach statistical significance (Fig. 5C). The expression level of *CDC20* gene produced significant changes with the progression of clinical stage and grading, and its expression level was significantly up-regulated in Stage IV and G4 (Fig. 5E, $p < 0.001$). These results imply that these hub genes may be closely associated with tumor progression in KIRC.

Distribution and expression of hub genes in the kidney

Six major cell clusters were identified from public scRNA-seq datasets. These clusters included endothelial cells, Natural killer T (NKT) cells, fibroblasts, B cells, tumor cells,
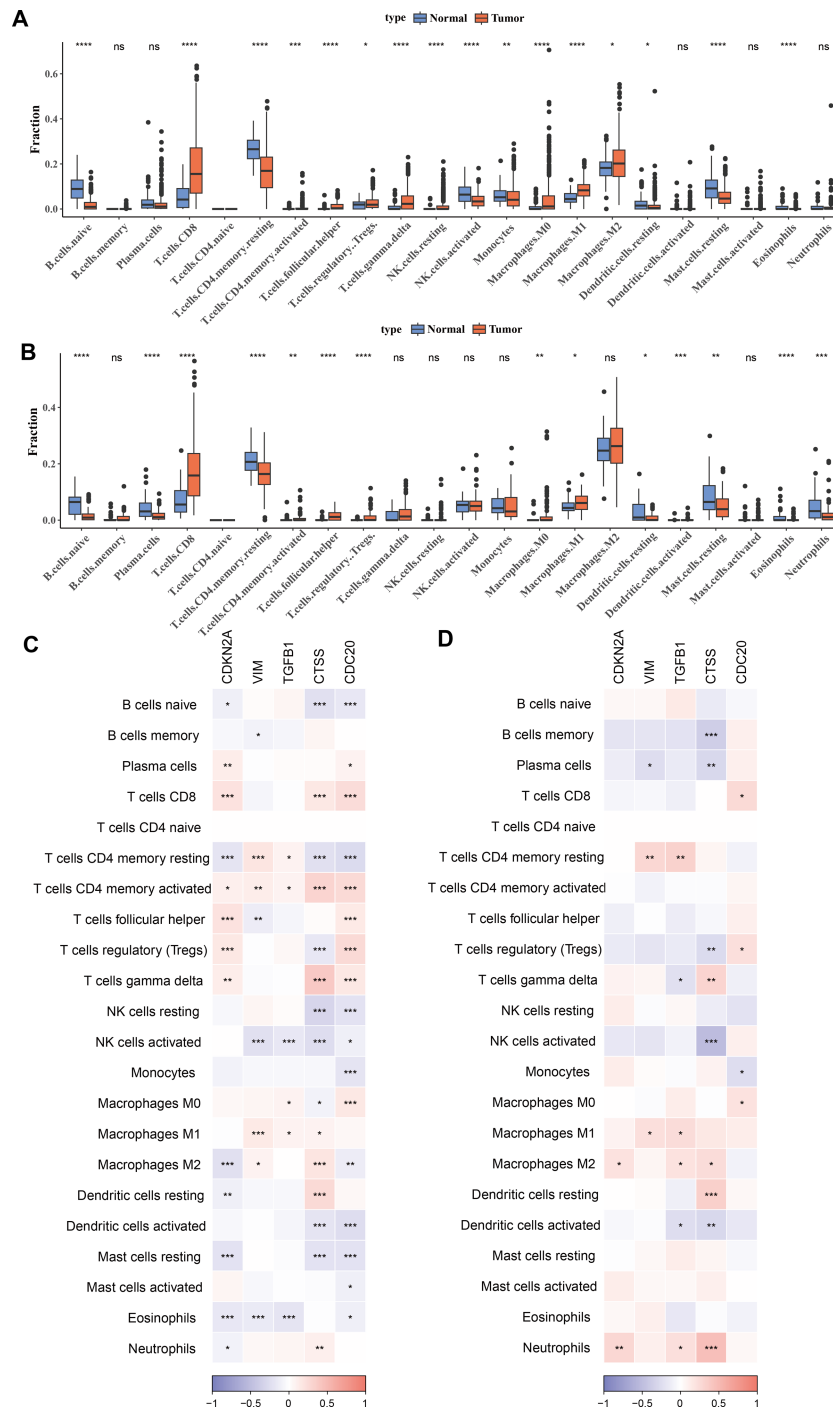
Figure 4 Correlation between biomarkers and immune infiltration. (A) Distribution of different types of immune cells in normal and tumor tissues in the TCGA training set. (B) Distribution of different types of immune cells in normal and tumor tissues in the ICGC validation set. (C) Heatmap of correlations between biomarkers (CDKN2A, VIM, TGFB1, CTSS, CDC20) and different types of immune cells in the TCGA training set. (D) Heatmap of correlations between biomarkers (CDKN2A, VIM, TGFB1, CTSS, CDC20) and different types of immune cells in the ICGC validation set. $* p < 0.05$, $** p < 0.01$, $*** p < 0.001$, $**** p < 0.0001$.
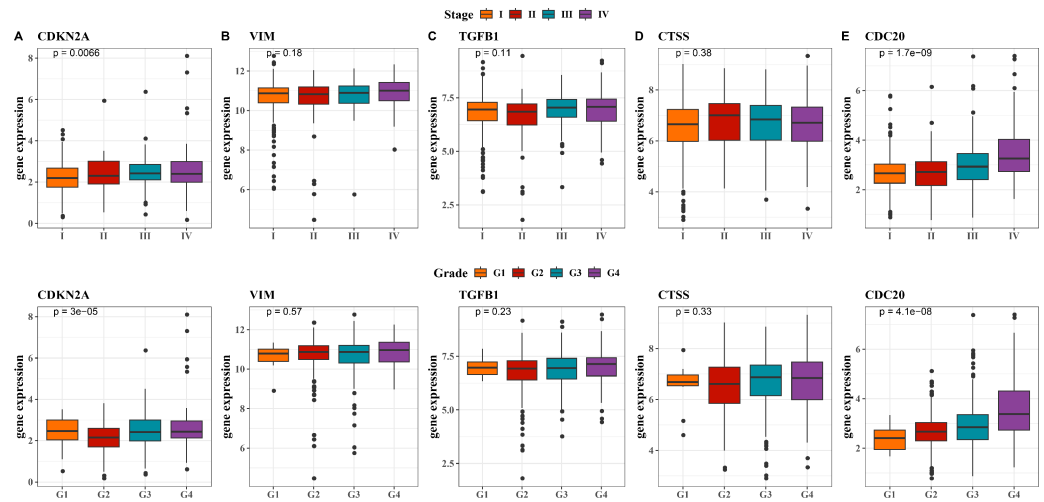
Full-size ☑ DOI: 10.7717/peerj.19246/fig-4

Figure 5 **Expression levels of 5 biomarkers (CDKN2A, VIM, TGFB1, CTSS, CDC20) in patients at different stages and grades.** (A) The expression levels of CDKN2A in patients with different stages and grades in the TCGA training set and the ICGC validation set; (B) The expression levels of VIM in patients with different stages and grades in the TCGA training set and the ICGC validation set; (C) The expression levels of TGFB1 in patients with different stages and grades in the TCGA training set and the ICGC validation set; (D) The expression levels of CTSS in patients with different stages and grades in the TCGA training set and the ICGC validation set; (E) The expression levels of CDC20 in patients with different stages and grades in the TCGA training set and the ICGC validation set.

Full-size ⬜ DOI: 10.7717/peerj.19246/fig-5

and proliferative tumor cells (Fig. 6A). Further gene expression characteristics were used to demonstrate the marker genes of various cell types. For instance, genes highly expressed in B cells included *CD79A* and *MS4A1*, endothelial cells specifically expressed *PLVAP*, *VWF*, and *EMCN*, fibroblasts specifically expressed *COL3A1* and *COL1A2*, NKT cells specifically expressed *GZMA*, *NKG7*, and *GNLY*, while proliferative tumor cells and tumor cells expressed specific tumor marker genes such as *MKI67* and *EPCAM*, respectively (Fig. 6B). Comparison of the proportions of various cell types in the samples revealed that tumor cells accounted for the highest proportion, reaching 42%, followed by endothelial cells and B cells, each accounting for 19%, fibroblasts accounting for 17%, and NKT cells and proliferative tumor cells accounting for the smallest proportions (Fig. 6C). Finally, the expressions of key hub genes in different cell types was further analyzed. These results demonstrated that *TGFB1*, *CDKN2A*, and *CDC20* were significantly expressed in proliferative tumor cells, *CTSS* was most prominently expressed in B cells, while *VIM* was expressed in multiple cell types (Fig. 6D).

Cellular validation based on *in vitro* experiment

The possibilities of *CDKN2A*, *VIM*, *TGFB1*, *CTSS* and *CDC20* as biomarkers of KIRC were further verified through *in vitro* experiments. The mRNA expression levels of these five genes were calculated, and it was found that the expressions of all these five genes in 786-O KIRC cells were significantly higher than those in 293T cells (Fig. 7A, $p < 0.05$). Since we observed a significant up-regulation of *CTSS* expression in KIRC cell lines relative to the other key genes screened, and its less studied in *CTSS* in KIRC. For this reason,
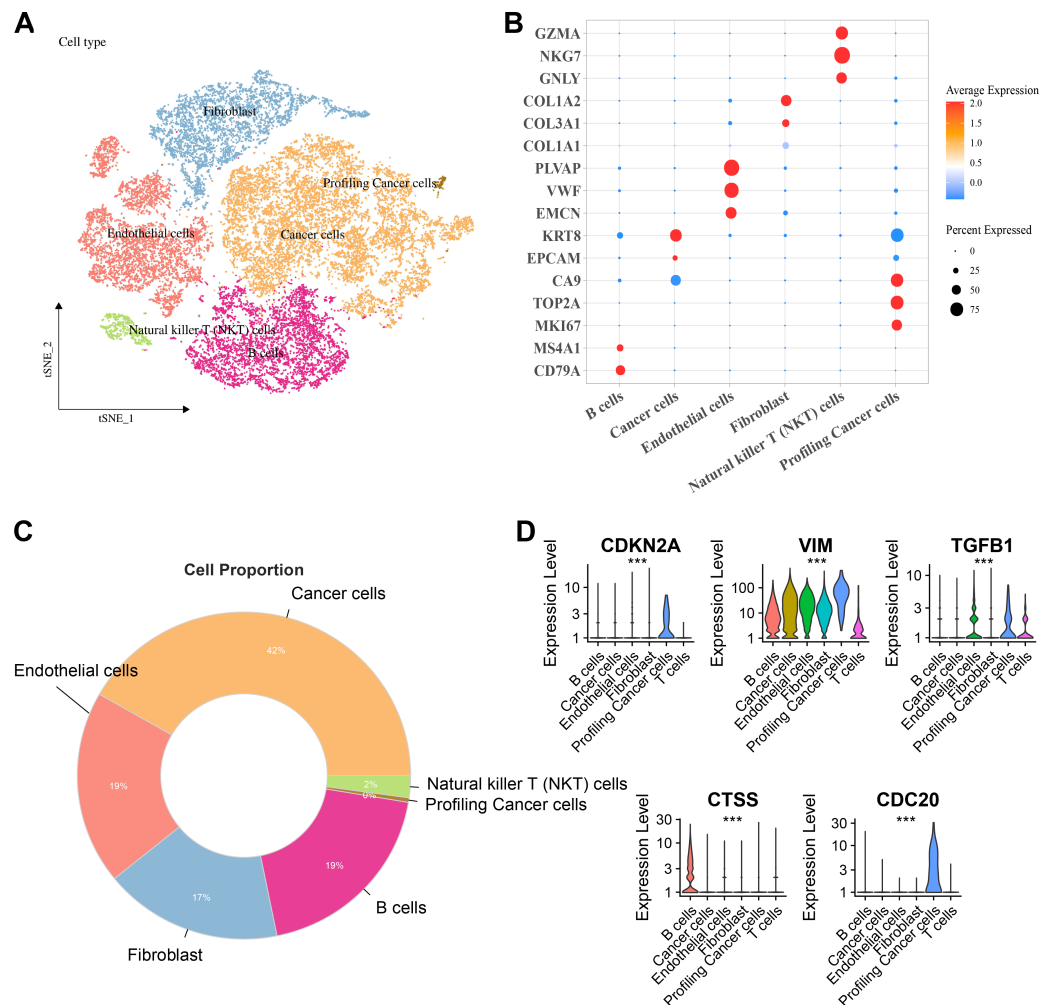
Figure 6 **Distribution and expression of hub genes across six major cell clusters.** (A) Cluster analysis of KIRC cells using the TSNE dimensionality reduction method; (B) Bubble plot showing the expression levels of marker genes for each cell type; (C) Proportion distribution of various cell types across the entire KIRC sample; (D) Expression profile of hub genes in six major cell types. ***$p < 0.001$.

Full-size 🖼 DOI: 10.7717/peerj.19246/fig-6

the impact of *CTSS* silencing on the biological function of KIRC cells was investigated (Fig. 7B). CCK-8 results showed that silencing the expression of *CTSS* significantly reduced the proliferative capacity of 786-O cells (Fig. 7C, $p < 0.001$). Furthermore, the relevant results demonstrated that the silencing of *CTSS* led to a reduction in the migration and invasion capabilities of KIRC cells (Figs. 7D–7E, $p < 0.001$).

DISCUSSION

RCC accounts for about 2.2% of all newly diagnosed cancer cases, and over the past three decades, its incidence has been steadily rising across all stages (*Bray et al., 2024*). While early diagnosis of RCC is associated with a relatively favorable prognosis, KIRC is characterized by the absence of early warning signs (*Cuadros et al., 2013*). Notably, CICs
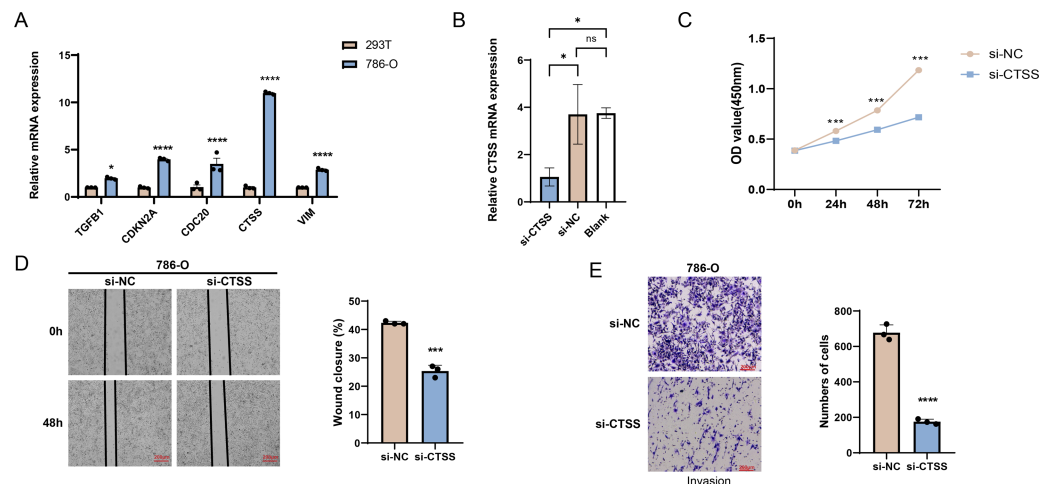
Figure 7 ***In vitro*** **validation results.** (A) Quantified mRNA expression levels of *CDKN2A*, *VIM*, *TGFB1*, *CTSS*, and *CDC20* in 786-O cells and 293T cells. (B) qRT-PCR to verify *CTSS* knockdown efficiency in 786-O cells. (C) The effect of CTSS knockdown on 786-O cells proliferation was determined based on CCK-8. (D) The wound healing assay was used to assess the effect of *CTSS* knockdown on 786-O cell migration. (E) Transwell assay was used to assess the effect of *CTSS* knockdown on the invasive capacity of 786-O cells. All data of three independent trials were expressed as mean ± standard deviation. * $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$, and ns stands for no significant difference.

Full-size 🖼 DOI: 10.7717/peerj.19246/fig-7

have been identified to occur between homotypic tumor cells or between immune cells and tumor (or other tissue cells) (*Mackay & Muller, 2019*), presenting a novel direction for KIRC research. Based on this, the current study employed machine learning methods to screen and identify five biomarkers (*CDKN2A*, *VIM*, *TGFB1*, *CTSS*, and *CDC20*). The reliability of these biomarkers has been verified by *in vitro* experiments. A diagnostic model with good predictive performance was established using these five biomarkers. In additional, the results of immune infiltration showed a higher proportion of T-cells and macrophages in tumor tissues. These findings open up new potential avenues for exploring and developing novel therapeutic approaches for KIRC, offering potential possibilities for improving patient treatment outcomes and prognosis.

Numerous studies have demonstrated that the PI3K-Akt signaling pathway exhibits aberrant activation during the genesis and development of various tumor types, KIRC included (*Makhov et al., 2018b*). Activation of this pathway typically results in dysregulated cell—cycle control, augmented anti-apoptotic capabilities, and immune evasion within the tumor microenvironment, all of which drive tumor progression and metastasis (*Xie et al., 2020*). Specifically in KIRC, the abnormal activation of the PI3K-Akt pathway is intricately associated with tumor cell proliferation, angiogenesis, and drug resistance (*Chen et al., 2023*). In our research, we observed a remarkable enrichment of upregulated genes from both the TCGA and ICGC datasets in this very pathway. Consequently, the PI3K—Akt pathway holds great potential as a viable therapeutic target for KIRC. Inhibiting its activity could be an effective strategy to curtail tumor growth and enhance the effectiveness of treatment regimens.

Importantly, this study identified five biomarkers, namely *CDKN2A*, *VIM*, *TGFB1*, *CTSS*, and *CDC20*. Among them, the *CDKN2A* gene is located in the frequently deleted p21 region on chromosome 9 and is widely recognized as a tumor suppressor (*Zhao et al., 2016*; *Gil & Peters, 2006*). The accumulation of various genetic alterations, including the *CDKN2A* gene, underlies the development of RCC (*Dulaimi et al., 2004*). *CDKN2A* is considered a key target for 9p deletions in multiple tumors, particularly RCC, due to its frequent inactivation through homozygous deletions or hypermethylation in the promoter region (*Vidaurreta et al., 2008*; *Schraml et al., 2001*). In 9.5% of RCC samples, the *CDKN2A* gene is lost along with other genetic materials, and these samples exhibit sarcomatoid features, a highly aggressive form of RCC that may benefit from immunotherapy (*Kiatprungvech et al., 2024*), which also corroborates with the significant correlation between the *CDKN2A* gene and immune cell infiltration observed in our study. Additionally, research has proven that inhibiting *CDKN2A* effectively promotes the formation of homologous CICs, and the activation of CIC-mediated cell death can serve as a barrier against potential malignant transformation induced by the inactivation of tumor suppressor genes like *CDKN2A* (*Liang et al., 2018*), suggesting that *CDKN2A*, as a CIC-related gene, holds potential therapeutic promise for KIRC and warrants further investigation. In this study, we observed an upward trend in *CDKN2A* gene expression levels in patients with KIRC, particularly those at G3 and G4 stages. However, contrary to this, earlier studies in laryngeal squamous cell carcinoma found an increased frequency of *CDKN2A* gene hypermethylation in patients at the G3 stage (*Smigiel et al., 2004*). This implies that the expression regulatory mechanisms of the *CDKN2A* gene may differ among different types of cancers, and there may be a complex relationship between its expression changes and cancer progression stages.

The *VIM* gene is located on chromosome 10p13 and serves as a major component of the mesenchymal cytoskeleton (*Shi et al., 2015*). Research on malignant tumors has shown that *VIM* functions crucially in cell cycle regulation, migration, adhesion, and the epithelial-mesenchymal transition (EMT) process in cancer (*Yao et al., 2020*). Recent research reports indicate that VIM protein can influence immune cells infiltration in the tumor microenvironment (*Dutsch-Wicherek, Lazar & Tomaszewska, 2011*). Prior studies have also found that *VIM* is an independent factor for the prognosis of KIRC (*Xu et al., 2020*). *TGFB1*, a cytokine with regulatory functions, has been reported in the literature to exhibit both stimulatory and inhibitory properties in regulating tissue homeostasis, developmental processes, tissue remodeling, and disease states such as cancer (*Ingman & Robertson, 2009*; *Ciftci et al., 2014*). Previous experiments have confirmed that the expression level of *TGFB1* is significantly elevated in KIRC tissues than normal kidney tissues (*Takahara et al., 2022*). *CTSS*, one of the 11 members of the cysteine protease family, is closely associated with various pathological conditions, including in cancers (*Wilkinson et al., 2019*). Recent research efforts have elucidated the key role of *CTSS* in influencing the pathogenesis of chronic kidney disease (*Steubl et al., 2017*). Experimental results show that the expression of *CTSS* is noticeably promoted in KIRC tissues compared to normal kidney tissues (*Zhou et al., 2024*). Based on this, in this study, *CTSS* was silenced for *in vitro* experimental verification, and the results also demonstrated that *CTSS* silencing could inhibit the migration and invasion of KIRC cells. Additionally, this study also found

that *CTSS* expression is particularly prominent in B cells. This finding echoed previous research, which also pointed out that *CTSS* exhibits high expression in antigen-presenting cells (APCs) and the lysosomes of malignant B cells (*Bararia et al., 2020*). In summary, these discoveries all confirm the correlation between *VIM*, *TGFB1*, and *CTSS* with KIRC, thus suggesting their potential as biomarkers for KIRC.

CDC20 functions as an oncogenic regulator at multiple critical nodes of the cell cycle and is negatively regulated by the tumor suppressor protein p53, thus being considered a highly promising therapeutic target (*Wang et al., 2015*; *Kidokoro et al., 2008*). This study reveals that *CDC20* expression levels exhibit a significant increase with the advancement of clinical stage and grade. Notably, *CDC20* can directly bind to and activate the anaphase-promoting complex in conjunction with another important regulatory molecule, E-cadherin, which plays a vital part in the precise regulation of cell entry into and exit from mitosis (*Schrock et al., 2020*). Given that previous studies have shown substances such as APC^{CDC20} to function in the transition from metaphase to anaphase by disrupting key cell cycle regulators (*Yu, 2007*), this discovery further implies that *CDC20* may occupy a pivotal position in disease progression (*Zeng et al., 2010*). Numerous studies have indicated that *CDC20* is not only a potential effective target for various cancer therapies but also a potential biomarker for prognosis (*Yuan et al., 2017*). Moreover, in previous research on KIRC, *CDC20* was also identified as a biomarker (*Gu et al., 2017*).

Our analysis of immune infiltration showed that the proportions of immune cells such as T cells and macrophages in tumor tissues were remarkably higher than those in normal tissues, and five biomarkers were identified to have significant correlations with T cells and macrophages. Previous studies have shown that in KIRC, the deletion or dysfunction of *CDKN2A* is closely associated with an inflammatory immune phenotype and the exhaustion state of $CD8^+$ T cells (*Sobottka et al., 2024*). Notably, the number of these exhausted $CD8^+$ T cells tend to increase relatively in metastatic sites, which may be linked to the immune escape mechanisms of tumors (*Sobottka et al., 2024*). On the other hand, *CTSS* within the cysteine cathepsin family is unique due to its limited tissue expression, primarily associated with antigen-presenting cells in lymph nodes and spleen, as well as other immune cells, especially macrophages (*Wilkinson et al., 2015*). Additionally, this study also found that the upregulated DEGs were enriched in immune regulation-related pathways such as T cell activation, Human T-cell leukemia virus 1 infection (HTLV-1), and regulation of T cell activation. Adult T-cell Leukemia/Lymphoma (ATL) is a $CD4^+$ T-cell malignancy caused by infection with HTLV-1 (*Liu et al., 2005*). *CDC20* plays an important part in the pathogenesis and development of ATL by mediating mitotic defects and the advancement of aneuploid cells (*Bruno et al., 2022*). This discovery also corroborates the correlation between the biomarkers obtained in this study and T-cell-related immune pathways. In summary, these findings could improve our understanding of the immune microenvironment in KIRC, but also offer potential targets for the development of novel immunotherapies targeting this disease.

However, this study has certain limitations. Firstly, although the data of this study underwent rigorous screening and analysis, it is constrained by the sample size and the homogeneity of data sources, which may, to a certain extent, undermine the broad

applicability and reliability of the research findings. To enhance the universality and persuasiveness of the conclusions, future research needs to expand the sample size and strive to encompass diverse patient populations and disease stages, while incorporating high-quality data from multiple sources. Furthermore, the specific molecular mechanism by which the biomarkers identified in this study affect immune infiltration remains unclear. Looking ahead, we plan to use advanced molecular biology techniques to deeply investigate the interactions between the identified biomarkers and various intracellular signaling pathways, with a particular focus on the specific pathways and mechanisms by which they act on immune infiltration, in the hope of revealing the underlying biological mysteries and providing a more solid theoretical basis for the diagnosis of KIRC and treatment.

CONCLUSION

This study identified five core biomarkers associated with CICs in KIRC through transcriptome analysis and machine learning methods: *CDKN2A*, *VIM*, *TGFB1*, *CTSS*, and *CDC20*. The diagnostic model constructed based on these biomarkers demonstrated good predictive performance. Importantly, these biomarkers were significantly correlated with the infiltration of specific immune cells in the tumor microenvironment, suggesting that these genes may be involved in regulating the immune evasion mechanism of KIRC. Additionally, the expression levels of CDKN2A and CDC20 showed significant differences in clinical stage and pathological grade. Finally, *in vitro* experiments verified that silencing CTSS inhibited KIRC cell proliferation, migration, and invasion, further supporting its potential as a therapeutic target. In conclusion, this study provides new molecular targets for the early diagnosis, prognosis assessment, and personalized treatment of KIRC, offering important insights for future therapeutic strategies.

Abbreviations

RCC	renal cell carcinoma
CIC	cell-in-cell structure
KIRC	kidney renal clear cell carcinoma
TCGA	The Cancer Genome Atlas
GDC	Genomic Data Commons
API	Application Programming Interface
FPKM	Fragments Per Kilobase of exon model per Million mapped fragments
TPM	Transcripts Per Million
RECA-EU	The Renal Cell Cancer-European Union
ICGC	International Cancer Genome Consortium
GEO	Gene Expression Omnibus
DEG	differentially expressed gene
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
BP	biological processes
CC	cellular components
MF	molecular functions

RFE	Recursive Feature Elimination
CV	cross-validation
SVM	Support Vector Machine
LASSO	Least Absolute Shrinkage and Selection Operator
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
scRNA-seq	single-cell RNA sequencing
PCA	Principal Component Analysis
TSNE	t-Distributed Stochastic Neighbor Embedding
DMEM	Dulbecco's Modified Eagle Medium
RPMI	Roswell Park Memorial Institute
FBS	fetal bovine serum
qRT-PCR	quantitative real-time PCR
NCBI	National Center for Biotechnology Information
CAM	Cell adhesion molecule
NKT cell	Natural killer T-cell
EMT	epithelial-mesenchymal transition
APC	antigen-presenting cell
ATL	Adult T-cell Leukemia/Lymphoma
HTLV-1	Human T-cell Leukemia Virus Type 1

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests
The authors declare there are no competing interests.

Author Contributions
- Zehua Wang conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Zhongxiao Zhang conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

Data Availability
The following information was supplied regarding data availability:

The sequences are available at GEO: GSE224630.

The raw data is available in GitHub and Zenodo:

- https://github.com/zhongxiaoZhang/Raw-data.git

- zhongxiaoZhang. (2025). zhongxiaoZhang/Raw-data: raw data (v.1.1.1). Zenodo. https://doi.org/10.5281/zenodo.14922228.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.19246#supplemental-information.

REFERENCES

Bararia D, Hildebrand JA, Stolz S, Haebe S, Alig S, Trevisani CP, Osorio-Barrios F, Bartoschek MD, Mentz M, Pastore A, Gaitzsch E, Heide M, Jurinovic V, Rautter K, Gunawardana J, Sabdia MB, Szczepanowski M, Richter J, Klapper W, Louissaint Jr A, Ludwig C, Bultmann S, Leonhardt H, Eustermann S, Hopfner K-P, Hiddemann W, Von Bergwelt-Baildon M, Steidl C, Kridel R, Tobin JWD, Gandhi MK, Weinstock DM, Schmidt-Supprian M, Sárosi MB, Rudelius M, Passerini V, Mautner J, Weigert O. 2020. Cathepsin S alterations induce a tumor-promoting immune microenvironment in follicular lymphoma. *Cell Reports* **31(5)**:107522 DOI 10.1016/j.celrep.2020.107522.

Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. 2024. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **74(3)**:229–263 DOI 10.3322/caac.21834.

Bruno S, Ghelli Luserna di Rorà A, Napolitano R, Soverini S, Martinelli G, Simonetti G. 2022. CDC20 in and out of mitosis: a prognostic factor and therapeutic target in hematological malignancies. *Journal of Experimental & Clinical Cancer Research* **41(1)**:159 DOI 10.1186/s13046-022-02363-9.

Chen J, Lin Y, Zheng S, Chen Q, Tang S, Zhong X. 2023. CBX3 promotes clear cell renal carcinoma through PI3K/AKT activation and aberrant immunity. *Journal of Translational Medicine* **21(1)**:600 DOI 10.1186/s12967-023-04478-9.

Chen Y-H, Wang S, He M-F, Wang Y, Zhao H, Zhu H-Y, Yu X-M, Ma J, Che X-J, Wang J-F, Wang Y, Wang X-N. 2013. Prevalence of heterotypic tumor/immune cell-in-cell structure *in vitro* and *in vivo* leading to formation of aneuploidy. *PLOS ONE* **8(3)**:e59418 DOI 10.1371/journal.pone.0059418.

Ciftci R, Tas F, Yasasever CT, Aksit E, Karabulut S, Sen F, Keskin S, Kilic L, Yildiz İ, Bozbey HU, Duranyildiz D, Vatansever S. 2014. High serum transforming growth factor beta 1 (TGFB1) level predicts better survival in breast cancer. *Tumor Biology* **35(7)**:6941–6948 DOI 10.1007/s13277-014-1932-y.

Cuadros T, Trilla E, Vilà MR, De Torres I, Vilardell J, Messaoud NB, Salcedo M, Sarró E, López-Hellin J, Blanco A, Mir C, Ramóny Cajal S, Itarte E, Morote J, Meseguer A. 2013. Hepatitis A virus cellular receptor 1/kidney injury molecule-1 is a susceptibility gene for clear cell renal cell carcinoma and hepatitis A virus cellular receptor/kidney injury molecule-1 ectodomain shedding a predictive biomarker of tumour progression. *European Journal of Cancer* **49(8)**:2034–2047 DOI 10.1016/j.ejca.2012.12.020.

Cui H, Shan H, Miao MZ, Jiang Z, Meng Y, Chen R, Zhang L, Liu Y. 2020. Identification of the key genes and pathways involved in the tumorigenesis and

prognosis of kidney renal clear cell carcinoma. *Scientific Reports* **10(1)**:4271 DOI 10.1038/s41598-020-61162-4.

Dulaimi E, De Caceres II, Uzzo RG, Al-Saleem T, Greenberg RE, Polascik TJ, Babb JS, Grizzle WE, Cairns P. 2004. Promoter hypermethylation profile of kidney cancer. *Clinical Cancer Research* **10(12)**:3972–3979 DOI 10.1158/1078-0432.CCR-04-0175.

Dutsch-Wicherek M, Lazar A, Tomaszewska R. 2011. The potential role of MT and Vimentin immunoreactivity in the remodeling of the microenvironment of parotid adenocarcinoma. *Cancer Microenvironment* **4(1)**:105–113 DOI 10.1007/s12307-010-0058-z.

Fais S, Overholtzer M. 2018. Cell-in-cell phenomena in cancer. *Nature Reviews Cancer* **18(12)**:758–766 DOI 10.1038/s41568-018-0073-9.

Frew IJ, Moch H. 2015. A clearer view of the molecular complexity of clear cell renal cell carcinoma. *Annual Review of Pathology: Mechanisms of Disease* **10(Volume 10 2015)**:263–289 DOI 10.1146/annurev-pathol-012414-040306.

Friedman J, Hastie T, Tibshirani R, Narasimhan B, Tay K, Simon N, Qian J. 2021. Package 'glmnet'. CRAN R Repositary. 595. *Available at* https://cran.r-project.org/.

Gil J, Peters G. 2006. Regulation of the INK4b–ARF–INK4a tumour suppressor locus: all for one or one for all. *Nature Reviews Molecular Cell Biology* **7(9)**:667–677 DOI 10.1038/nrm1987.

Gu Y, Lu L, Wu L, Chen H, Zhu W, He Y. 2017. Identification of prognostic genes in kidney renal clear cell carcinoma by RNA-seq data analysis. *Molecular Medicine Reports* **15(4)**:1661–1667 DOI 10.3892/mmr.2017.6194.

Hao YH, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, Srivastava A, Molla G, Madad S, Fernandez-Granda C, Satija R. 2024. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature Biotechnology* **42(2)**:293–304 DOI 10.1038/s41587-023-01767-y.

Hu C, Li T, Xu Y, Zhang X, Li F, Bai J, Chen J, Jiang W, Yang K, Ou Q, Li X, Wang P, Zhang Y. 2022. CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. *Nucleic Acids Research* **51(D1)**:D870–D876 DOI 10.1093/nar/gkac947.

Huang H, Chen Z, Sun Q. 2015. Mammalian cell competitions, cell-in-cell phenomena and their biomedical implications. *Current Molecular Medicine* **15(9)**:852–860 DOI 10.2174/1566524015666151026101101.

Ingman WV, Robertson SA. 2009. The essential roles of TGFB1 in reproduction. *Cytokine & Growth Factor Reviews* **20(3)**:233–239 DOI 10.1016/j.cytogfr.2009.05.003.

Kiatprungvech N, Sangkum P, Malinee R, Sommaluan S, Korkiatsakul V, Worawicha-wong S, Rerkamnuaychoke B, Kongruang A, Aeesoa S, Lertsithichai P, Kijvikai K, Kongchareonsombat W, Siriboonpiputtana T. 2024. Genetic study of the CDKN2A and CDKN2B genes in renal cell carcinoma patients. *Practical Laboratory Medicine* **40**:e00410 DOI 10.1016/j.plabm.2024.e00410.

Kidokoro T, Tanikawa C, Furukawa Y, Katagiri T, Nakamura Y, Matsuda K. 2008. CDC20, a potential cancer therapeutic target, is negatively regulated by p53. *Oncogene* **27(11)**:1562–1571 DOI 10.1038/sj.onc.1210799.

Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh P-r, Raychaudhuri S. 2019. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods* **16(12)**:1289–1296 DOI 10.1038/s41592-019-0619-0.

Kuhn M. 2008. Building predictive models in R using the caret package. *Journal of Statistical Software* **28(5)**:1–26.

Liang J, Fan J, Wang M, Niu Z, Zhang Z, Yuan L, Tai Y, Chen Z, Song S, Wang X, Liu X, Huang H, Sun Q. 2018. CDKN2A inhibits formation of homotypic cell-in-cell structures. *Oncogenesis* **7(6)**:50 DOI 10.1038/s41389-018-0056-4.

Liu B, Hong S, Tang Z, Yu H, Giam C-Z. 2005. HTLV-I tax directly binds the Cdc20-associated anaphase-promoting complex and activates it ahead of schedule. *Proceedings of the National Academy of Sciences of the United States of America* **102(1)**:63–68 DOI 10.1073/pnas.0406424101.

Liu J, Zhang Z, Xiao L, Guo Y, Luo S, Zhou B. 2024. Establishment of a prognostic signature of disulfidptosis-related lncRNAs for predicting survival and immune landscape in clear cell renal cell carcinoma. *Oncologie* **26(4)**:603–618 DOI 10.1515/oncologie-2024-0049.

Mackay HL, Muller PAJ. 2019. Biological relevance of cell-in-cell in cancers. *Biochemical Society Transactions* **47(2)**:725–732 DOI 10.1042/BST20180618.

Makhov P, Joshi S, Ghatalia P, Kutikov A, Uzzo RG, Kolenko VM. 2018a. Resistance to systemic therapies in clear cell renal cell carcinoma: mechanisms and management strategies. *Molecular Cancer Therapeutics* **17(7)**:1355–1364 DOI 10.1158/1535-7163.MCT-17-1299.

Makhov P, Joshi S, Ghatalia P, Kutikov A, Uzzo RG, Kolenko VM. 2018b. Resistance to systemic therapies in clear cell renal cell carcinoma: mechanisms and management strategies. *Molecular Cancer Therapeutics* **17(7)**:1355–1364 DOI 10.1158/1535-7163.MCT-17-1299.

Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F, Chang C-C, Lin C-CJRpv. 2019. e1071: misc functions of the department of statistics, probability theory group (formerly: E1071), TU Wien. R package version, 1(2). *Available at https://cloud.r-project.org/package=e1071*.

Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. 2015. Robust enumeration of cell subsets from tissue expression profiles. *Nature Methods* **12(5)**:453–457 DOI 10.1038/nmeth.3337.

Ren W, Li J, Zeng R, Zhu L. 2024. RNF26 up-regulates PD-L1 to regulate the cancer immune response in ccRCC. *Biocell* **48(9)**:1323–1330 DOI 10.32604/biocell.2024.051747.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43(7)**:e47 DOI 10.1093/nar/gkv007.

Schraml P, Struckmann K, Bednar R, Fu W, Gasser T, Wilber K, Kononen J, Sauter G, Mihatsch MJ, Moch H. 2001. CDKN2A mutation analysis, protein expression, and deletion mapping of chromosome 9p in conventional clear-cell renal carcinomas:

evidence for a second tumor suppressor gene proximal to CDKN2A. *The American Journal of Pathology* **158**(**2**):593–601 DOI 10.1016/S0002-9440(10)64001-1.

Schrock MS, Stromberg BR, Scarberry L, Summers MK. 2020. APC/C ubiquitin ligase: functions and mechanisms in tumorigenesis. *Seminars in Cancer Biology* **67**:80–91 DOI 10.1016/j.semcancer.2020.03.001.

Seyfinejad B, Jouyban A. 2022. Importance of method validation in the analysis of biomarker. *Current Pharmaceutical Analysis* **18**(**6**):567–569 DOI 10.2174/1573412918666211213142638.

Shen W, Song Z, Zhong X, Huang M, Shen D, Gao P, Qian X, Wang M, He X, Wang T. 2022. Sangerbox: a comprehensive, interaction-friendly clinical bioinformatics analysis platform. *Imeta* **1**(**3**):e36 DOI 10.1002/imt2.36.

Shi ZG, Li SQ, Li ZJ, Zhu XJ, Xu P, Liu G. 2015. Expression of vimentin and survivin in clear cell renal cell carcinoma and correlation with p53. *Clinical and Translational Oncology* **17**(**1**):65–73 DOI 10.1007/s12094-014-1199-1.

Smigiel R, Sasiadek M, Krecicki T, Ramsey D, Jagielski J, Blin N. 2004. Inactivation of the cyclin-dependent kinase inhibitor 2A (CDKN2A) gene in squamous cell carcinoma of the larynx. *Molecular Carcinogenesis: Published in Cooperation with the University of Texas MD Anderson Cancer Center* **39**(**3**):147–154 DOI 10.1002/mc.20007.

Sobottka B, Vetter V, Banaei-Esfahani A, Nowak M, Lorch A, Sirek A, Mertz KD, Brunelli M, Berthold D, De Leval L, Kahraman A, Koelzer VH, Moch H. 2024. Immune phenotype-genotype associations in primary clear cell renal cell carcinoma and matched metastatic tissue. *Modern Pathology* **37**(**10**):100558 DOI 10.1016/j.modpat.2024.100558.

Song J, Ruze R, Chen Y, Xu R, Yin X, Wang C, Xu Q, Zhao Y. 2022. Construction of a novel model based on cell-in-cell-related genes and validation of KRT7 as a biomarker for predicting survival and immune microenvironment in pancreatic cancer. *BMC Cancer* **22**(**1**):894 DOI 10.1186/s12885-022-09983-6.

Song ZG, Yu JB, Wang MM, Shen WT, Wang CC, Lu TY, Shan GJ, Dong G, Wang YR, Zhao JY. 2023a. CHDTEPDB: transcriptome expression profile database and interactive analysis platform for congenital heart disease. *Congenital Heart Disease* **18**(**6**):693–701 DOI 10.32604/chd.2024.048081.

Song Z, Yu J, Wang M, Shen W, Wang C, Lu T, Shan G, Dong G, Wang Y, Zhao J. 2023b. CHDTEPDB: transcriptome expression profile database and interactive analysis platform for congenital heart disease. *Congenital Heart Disease* **18**(**6**):693–701 DOI 10.32604/chd.2024.048081.

Steubl D, Kumar SV, Tato M, Mulay SR, Larsson A, Lind L, Risérus U, Renders L, Heemann U, Carlsson AC, Ärnlöv J, Anders H-J. 2017. Circulating cathepsin-S levels correlate with GFR decline and sTNFR1 and sTNFR2 levels in mice and humans. *Scientific Reports* **7**(**1**):43538 DOI 10.1038/srep43538.

Su Y, Huang H, Luo T, Zheng Y, Fan J, Ren H, Tang M, Niu Z, Wang C, Wang Y, Zhang Z, Liang J, Ruan B, Gao L, Chen Z, Melino G, Wang X, Sun Q. 2022. Cell-in-cell structure mediates in-cell killing suppressed by CD44. *Cell Discovery* **8**(**1**):35 DOI 10.1038/s41421-022-00387-1.

Sun Q, Chen W. 2022. Cell-in-cell: an emerging player in COVID-19 and immune disorders. *National Science Open* **1(1)**:2021004.

Takahara T, Tsuyuki T, Satou A, Wada E, Sakurai K, Ueda R, Tsuzuki T. 2022. TGFB1 mRNA expression is associated with poor prognosis and specific features of inflammation in ccRCC. *Virchows Archiv* **480(3)**:635–643 DOI 10.1007/s00428-021-03256-6.

Tian K, Chen H, Wang Q, Jiang F, Feng C, Li T, Pu X, Tang Y, Liu J. 2024. CARD11 serves as a therapeutic biomarker for the drug therapies of ccRCC. *Biocell* **48(5)**:817–834 DOI 10.32604/biocell.2024.048737.

Vidaurreta M, Maestro ML, Sanz-Casla MT, Maestro C, Rafael S, Veganzones S, Moreno J, Blanco J, Silmi A, Arroyo M. 2008. Inactivation of p16 by CpG hypermethylation in renal cell carcinoma. *Urologic Oncology: Seminars and Original Investigations* **26(3)**:239–245 DOI 10.1016/j.urolonc.2007.01.018.

Wang X. 2015. Cell-in-cell phenomenon: a new paradigm in life sciences. *Current Molecular Medicine* **15(9)**:810–818 DOI 10.2174/1566524015666151026095730.

Wang J, Cai Y, Sheng Z, Dong Z. 2023. EGFR inhibitor CL-387785 suppresses the progression of lung adenocarcinoma. *Current Molecular Pharmacology* **16(2)**:211–216 DOI 10.2174/1874467215666220329212300.

Wang C, Chen A, Ruan B, Niu Z, Su Y, Qin H, Zheng Y, Zhang B, Gao L, Chen Z, Huang H, Wang X, Sun Q. 2020. PCDH7 inhibits the formation of homotypic cell-in-cell structure. *Frontiers in Cell and Developmental Biology* **8(8)**:329 DOI 10.3389/fcell.2020.00329.

Wang Q, Zhang H, Chen Q, Wan Z, Gao X, Qian W. 2019. Identification of METTL14 in kidney renal clear cell carcinoma using bioinformatics analysis. *Disease Markers* **2019(1)**:5648783 DOI 10.1155/2019/5648783.

Wang L, Zhang J, Wan L, Zhou X, Wang Z, Wei W. 2015. Targeting Cdc20 as a novel cancer therapeutic strategy. *Pharmacology & Therapeutics* **151**:141–151 DOI 10.1016/j.pharmthera.2015.04.002.

Wilkinson RDA, Burden RE, McDowell SH, McArt DG, McQuaid S, Bingham V, Williams R, Cox ÓT, O'Connor R, McCabe N, Kennedy RD, Buckley NE, Scott CJ. 2019. A novel role for cathepsin s as a potential biomarker in triple negative breast cancer. *Journal of Oncology* **2019(1)**:3980273 DOI 10.1155/2019/3980273.

Wilkinson RDA, Williams R, Scott CJ, Burden RE. 2015. Cathepsin S: therapeutic, diagnostic, and prognostic potential. *Biological Chemistry* **396(8)**:867–882 DOI 10.1515/hsz-2015-0114.

Xie L, Li H, Zhang L, Ma X, Dang Y, Guo J, Liu J, Ge L, Nan F, Dong H, Yan Z, Guo X. 2020. Autophagy-related gene P4HB: a novel diagnosis and prognosis marker for kidney renal clear cell carcinoma. *Aging* **12(2)**:1828–1842 DOI 10.18632/aging.102715.

Xu H, Xu WH, Ren F, Wang J, Wang HK, Cao DL, Shi GH, Qu YY, Zhang HL, Ye DW. 2020. Prognostic value of epithelial-mesenchymal transition markers in clear cell renal cell carcinoma. *Aging* **12(1)**:866–883 DOI 10.18632/aging.102660.

Yao JX, Chen X, Zhu YJ, Wang H, Hu XY, Guo JM. 2020. Prognostic value of vimentin is associated with immunosuppression in metastatic renal cell carcinoma. *Frontiers in Oncology* **10**:1181 DOI 10.3389/fonc.2020.01181.

Yu H. 2007. Cdc20: a WD40 activator for a cell cycle degradation machine. *Molecular Cell* **27(1)**:3–16 DOI 10.1016/j.molcel.2007.06.009.

Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16(5)**:284–287 DOI 10.1089/omi.2011.0118.

Yuan L, Chen L, Qian K, Qian G, Wu C-L, Wang X, Xiao Y. 2017. Co-expression network analysis identified six hub genes in association with progression and prognosis in human clear cell renal cell carcinoma (ccRCC). *Genomics Data* **14**:132–140 DOI 10.1016/j.gdata.2017.10.006.

Zeng X, Sigoillot F, Gaur S, Choi S, Pfaff KL, Oh D-C, Hathaway N, Dimova N, Cuny GD, King RW. 2010. Pharmacologic inhibition of the anaphase-promoting complex induces a spindle checkpoint-dependent mitotic arrest in the absence of spindle damage. *Cancer Cell* **18(4)**:382–395 DOI 10.1016/j.ccr.2010.08.010.

Zhang X, Jin M, Yao X, Liu J, Yang Y, Huang J, Jin G, Liu S, Zhang B. 2024. Upregulation of LncRNA WT1-AS inhibits tumor growth and promotes autophagy in gastric cancer *via* suppression of PI3K/Akt/mTOR pathway. *Current Molecular Pharmacology* **17**:e18761429318398 DOI 10.2174/0118761429318398240918063450.

Zhang L, Yang H, Liu J, Wang K, Cai X, Xiao W, Wang L, Wang M, Zhang C, Zhang J. 2023. Metabolomics-based approach to analyze the therapeutic targets and metabolites of a synovitis ointment for knee osteoarthritis. *Current Pharmaceutical Analysis* **19(3)**:222–234 DOI 10.2174/1573412919666221223152915.

Zhao R, Choi BY, Lee M-H, Bode AM, Dong Z. 2016. Implications of genetic and epigenetic alterations of CDKN2A (p16INK4a) in cancer. *EBioMedicine* **8**:30–39 DOI 10.1016/j.ebiom.2016.04.017.

Zhou X, Chen H, Huang D, Guan G, Ma X, Cai W, Liao J, Guan T. 2024. Reduced expression of cathepsin F predicts poor prognosis in patients with clear cell renal cell carcinoma. *Scientific Reports* **14(1)**:13556 DOI 10.1038/s41598-024-64542-2.