Cellular Oncology (2025) 48:591–615
https://doi.org/10.1007/s13402-024-01030-9

RESEARCH

Construction and validation of a regulatory T cells-based classification of renal cell carcinoma: an integrated bioinformatic analysis and clinical cohort study

Yuntao Yao[1] · Yifan Liu[1] · Bingnan Lu[1] · Guo Ji[2] · Lei Wang[1] · Keqin Dong[1] · Zihui Zhao[1] · Donghao Lyu[1] · Maodong Wei[1] · Siqi Tu[1] · Xukun Lyu[1] · Yuanan Li[1] · Runzhi Huang[3] · Wang Zhou[1] · Guofeng Xu[1] · Xiuwu Pan[1] · Xingang Cui[1]

Accepted: 18 December 2024 / Published online: 23 December 2024

Abstract

Purpose Renal cell carcinoma (RCC), exhibiting remarkable heterogeneity, can be highly infiltrated by regulatory T cells (Tregs). However, the relationship between Treg and the heterogeneity of RCC remains to be explored.

Methods We acquired single-cell RNA-seq profiles and 537 bulk RNA-seq profiles of TCGA-KIRC cohort. Through clustering, monocle2 pseudotime and prognostic analyses, we identified Treg states-related prognostic genes (TSRPGs), then constructing the RCC Treg states-related prognostic classification (RCC-TSC). We also explored its prognostic significance and multi-omics landmarks. Additionally, we utilized correlation analysis to establish regulatory networks, and predicted candidate inhibitors. More importantly, in Xinhua cohort of 370 patients with kidney neoplasm, we used immunohistochemical (IHC) staining for classification, then employing statistical analyses including Chi-square tests and multivariate Cox proportional hazards regression analysis to explore its clinical relevance.

Results We defined 44 TSRPGs in four different monocle states, and identified high immune infiltration RCC (HIRC, LAG3+, Mki67+) as the highly exhausted subtype with the worst prognosis in RCC-TSC ($p < 0.001$). BATF-LAG3-immune cells axis might be its underlying metastasis-related mechanism. Immunotherapy and inhibitors including sunitinib potentially conferred best therapeutic effects for HIRC. Furthermore, we successfully validated HIRC subtype as an independent prognostic factor within the Xinhua cohort (OS, HR = 16.68, 95% CI = 1.88–148.1, $p = 0.011$; PFS, HR = 4.43, 95% CI = 1.55–12.6, $p = 0.005$).

Conclusion Through integrated bioinformatics analysis and a large-sample retrospective clinical study, we successfully established RCC-TSC and a diagnostic kit, which could stratify RCC patients with different prognosis and to guide personalized treatment.

Synopsis

Utilizing integrated bioinformatics analyses and a large-sample retrospective clinical study, a Treg states-related RCC prognostic classification is successfully established, which could stratify RCC patients with different prognosis and to guide personalized treatment.

Keywords Renal cell carcinoma (RCC) · Molecular classification system · Regulatory T cell (Treg) · Single-cell sequencing · Clinical study · Immunohistochemical staining

Yuntao Yao, Yifan Liu, Bingnan Lu, Guo Ji and Lei Wang contributed equally to this work and shared first authorship.

Xingang Cui: First correspondence author.

Extended author information available on the last page of the article

Abbreviations

RCC	Renal cell carcinoma
KIRC	Kidney renal clear cell carcinoma
TIME	Tumor immune microenvironment
NK	Natural killer
scRNA	Single-cell RNA
TCR	T cell receptor
Treg	Regulatory T cells
FOXP3	Forkhead box protein P3
ICB	Immune checkpoint blockade
TSRPGs	Treg states-related prognostic genes
GEO	Gene expression omnibus
OS	Overall survival
TCGA	The Cancer Genome Atlas
TF	Transcription factor
CCLE	Cancer Cell Line Encyclopedia
GDSC	Genomics of Drug Sensitivity in Cancer
ORA	Over-representation analysis
MSigDB	Molecular Signatures Database
QC	Quality control
DEGs	Differentially expressed genes
PCA	Principal component analysis
UMAP	Uniform Manifold Approximation and Projection
FC	Fold Change
GSVA	Gene set variation analysis
TSRGs	Treg states-related genes
TSRPGs	Treg states-related prognostic genes
K-M	Kaplan-Meier
RCC-TSC	RCC Treg states-related prognostic classification
H-PCA	High PCA score group
L-PCA	Low PCA score group
TIDE	Tumor immune dysfunction and exclusion
IPS	Landmarks and immunophenoscore
ssGSEA	Single-sample gene set enrichment analysis
MDSC	Myeloid-derived suppressor cell
TAM	Tumor-associated macrophage, CAF, cancer-associated fibroblast
MSI	Microsatellite instability
TMB	Tumor mutation burden
CNV	Copy number variation
DETFs	Differentially expressed transcription factor
ATAC-seq	Assay for transposase-accessible chromatin sequencing
KEGG	Kyoto encyclopedia of genes and genomes
LASSO	Least absolute shrinkage and selection operator
ROC	Receiver operating characteristic
AUC	Area under curve
GO	Gene oncology
IHC	Immunohistochemical
PFS	Progression-free survival
TMA	Tissue microarrays
DC	Dendritic cell
MIRC	Moderate immune infiltration RCC
HIRC	High immune infiltration RCC
LIRC	Low immune infiltration RCC
CAF	Cancer-associated fibroblast
ATRA	All-trans retinoic acid
TIL	Tumor infiltrating lymphocyte
MKi67	Marker of proliferation ki-67
LAG3	Lymphocyte activation gene 3
CD7	Cluster of differentiation 7
SOCS1	Suppressor of cytokine signaling 1

1 Introduction

Renal cell carcinoma (RCC) is one of the most prevalent cancers worldwide, ranking sixth in men and ninth in women, and it accounts for approximately 3% of all adult malignancies [1–3]. RCC is highly heterogenous, and kidney renal clear cell carcinoma (KIRC) comprises 70–80% neoplasms that originate from the renal tubular epithelium [4].

Tumor immune microenvironment (TIME) is a complicated ecosystem that is paramount in tumor progression and clinical outcomes of patients [5–7]. RCC is highly immunogenic and typically exhibits substantial leukocyte infiltrates, with CD8 + T cells and CD4+ T cells dominating in the TIME [8]. Recently, a study combing single-cell RNA (scRNA) and T cell receptor (TCR) sequencing found terminally exhausted CD8 + T cells and M2-phenotype macrophages were more abundant in the TIME of advanced and metastatic RCC, which were also significantly associated with the dysfunctional state and poor prognosis of patients [9]. More importantly, the hypofunctional state of intra-tumoral CD8 + and CD4+ T cells in RCC was mediated by multiple factors, including the profound effects of regulatory T cells (Treg) in the surrounding TIME, which significantly highlighted its critical role in the pathogenesis of RCC [8, 10–13].

Forkhead box protein P3 (FOXP3) + Treg, a highly immunosuppressive CD4+ T cell subtype, could maintain immune homeostasis [14, 15]. Nevertheless, major immunosuppressive functions of Treg also could suppress anticancer immunity to influence tumor pathogenesis [13, 16]. Notably, the dynamics and heterogeneity of T cells were revealed by a single-cell atlas in pan-cancer level, which characterized the phenotypes and compositions across different TIME, including Tregs with different marker genes (TNFRSF9+, OAS1+, RTKN2+, S1PR1+) [10]. Moreover, it was stated that proliferative Tregs exhibited highly

activated and immune-suppressive phenotype [17, 18], and abundant proliferative Tregs were a positive indicator of tumor progression [19–21]. However, relationship between different phenotypes of Tregs and the heterogeneity of RCC remains largely unexplored. Consequently, there is an urgent need to develop a molecular classification system based on Tregs for precise risk stratification and personalized therapy.

Herein, we performed an integrated study encompassing scRNA-seq analysis, bulk RNA-seq analysis, multi-omics analysis, and a retrospective clinical study, to construct a novel molecular RCC classification system. Utilizing monocle 2 pseudotime analysis to delineate distinct Tregs subtypes and performing prognostic analysis, we characterized Treg states-related prognostic genes (TSRPGs). Then, ConsensusClusterPlus clustering analysis was employed to compare the expression levels of TSRPGs among KIRC patients, identifying three subtypes with different prognostic, immunologic, biological, and genomic characteristics. Finally, using descriptive statistics, univariate analysis, multivariate analysis, model diagnostics, we validated the feasibility of the classification system in the Xinhua retrospective cohort, effectively stratifying patients with various kidney neoplasms into groups with distinct prognoses.

2 Materials and methods

2.1 Data sources of our research

The scRNA-seq profiles that used in Fig. 1 were retrieved from the Gene Expression Omnibus (GEO) database (accession No. GSE156728) (https://www.ncbi.nlm.nih.gov/geo) [22]. The scRNA-seq profiles used for validation were available at the following website (https://github.com/Aleksobrad/single-cell-rcc-pipeline). We also downloaded the microarray profiles and bulk RNA-sequencing profiles and the corresponding clinical data of patients with renal cancers from the GEO database (accession No. GSE22541).

We acquired bulk RNA-sequencing profiles and clinical data encompassing overall survival (OS), demographic characteristics (age, gender, and race), and disease conditions (TNM stage, pathological stage, grade), for 537 KIRC patients in The Cancer Genome Atlas (TCGA) database (https://cancergenome.nih.gov/) [23]. The transcription fa ctor (TF) profiles were sourced from the Cistrome cancer database (http://cistrome.org/) [24]. Meanwhile, the data pertaining to metastatic samples were from both MET500 (https://met500.path.med.umich.edu/) [25] and TCGA databases.

Additionally, genomic and transcriptomic characteristics, along with data regarding the sensitivity to cytotoxic drugs for cancer cell lines, were sourced from the Cancer Cell Line Encyclopedia (CCLE) database (https://sites.broadinstitute.org/ccle) [26], and the Genomics of Drug Sensitivity in Cancer (GDSC) database (https://www.cancerrxgene.org/) [27]. The 3D conformations of the LAG3 protein and ligands were downloaded from the alphafold website (https://alphafold.ebi.ac.uk/) and PubChem (https://pubchem.ncbi.nlm.nih.gov/), respectively. The pocket information for molecular docking was retrieved from the Protein Plus (https://proteins.plus/). Lastly, fifty molecular signaling pathways and nine pathway categories used for over-representation analysis (ORA) were retrieved from the Molecular Signatures Database (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp) (MSigDB, version 7.4) [28]. The code and information of the patients we utilized for analysis were uploaded in the supplementary material.

2.2 ScRNA-seq analysis processes

Cellranger (v3.0, http://10xgenomics.com/) was used to demultiplex the raw scRNA-seq data bcl files, generating two FASTQ files. The construction of the gene expression matrix was subsequently performed based on the two FASTQ files. Next, the Seurat R toolkit (v4.0.1; was employed for data standardization, quality control (QC), dimensionality reduction, unsupervised clustering, and analysis of differentially expressed genes (DEGs) [29, 30]. Furthermore, during the QC processes, only cells expressing more than 100,000 transcripts and exhibiting mitochondrial genes comprising less than 10% of all cellular genes were retained. After QC, "vst" algorithm was utilized to identify the top 2000 highly variable genes. Afterwards, these genes were scaled and standardized, and employed for dimensionality reduction and clustering. Finally, features of different cell types were extracted through principal component analysis (PCA), and shown in the Uniform Manifold Approximation and Projection (UMAP) plots.

2.3 Annotation of CD4+T cells in RCC

Genes expressed in more than 25% of cells, with $|\log_2$Fold Change (FC)$| > 0.25$ and an adjusted p value < 0.05 were considered eligible for distinguishing cell populations. Afterwards, we annotated the CD4+T cells with Seurat R toolkit, Single R package, the CellMarker database and the published literature (https://bio-bigdata.hrbmu.edu.cn/CellMarker/) [10, 31, 32]. Additionally, we calculated and illustrated the cell number and cellular composition of each cluster, as well as the sample proportions for the cell types through bar plots. A heatmap displaying the top five highly expressed genes per cluster and a Cleveland dot plot featuring four key marker genes were also displayed.

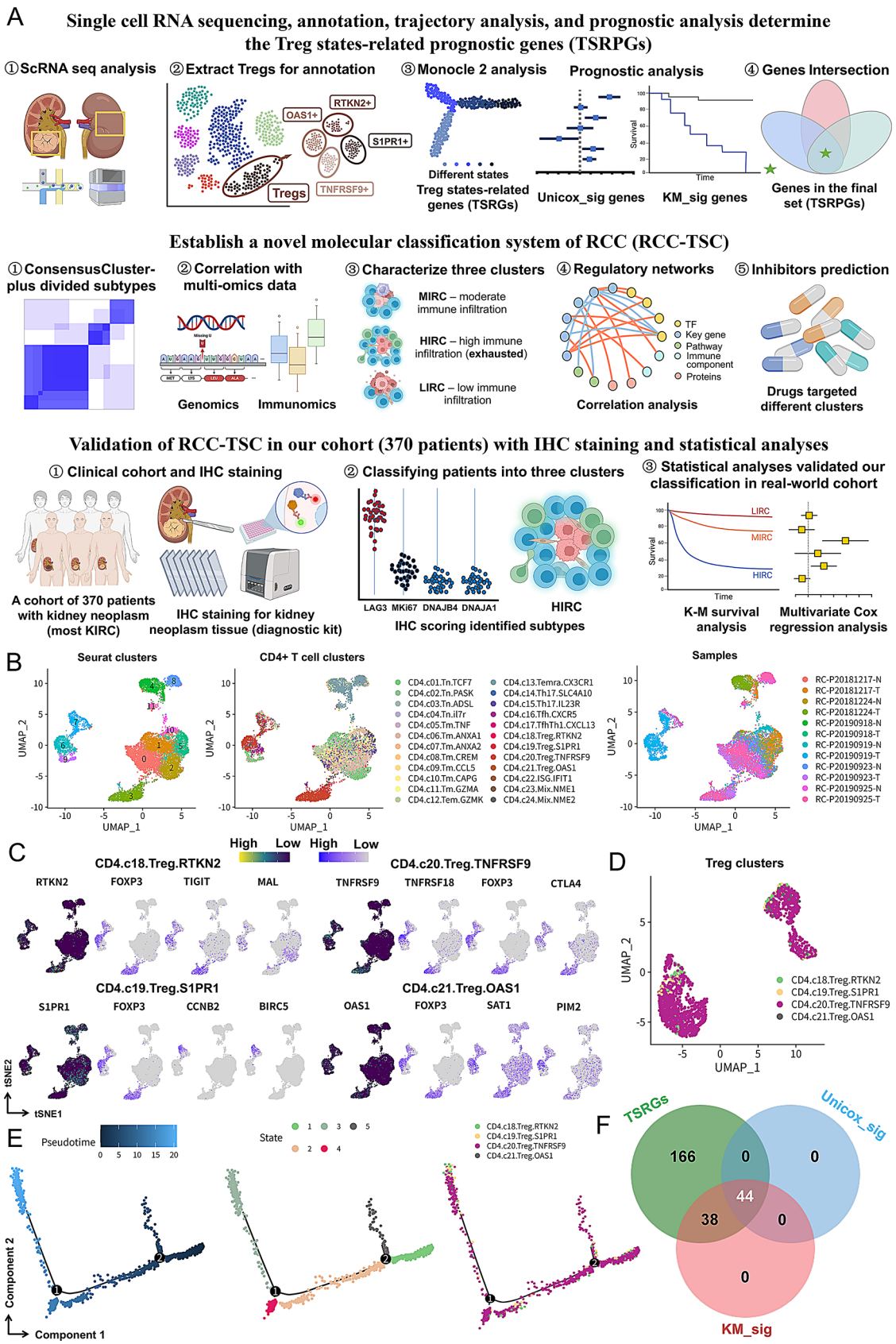

A
Single cell RNA sequencing, annotation, trajectory analysis, and prognostic analysis determine the Treg states-related prognostic genes (TSRPGs)
①ScRNA seq analysis
②Extract Tregs for annotation
③Monocle 2 analysis
Prognostic analysis
④Genes Intersection
RTKN2+
OAS1+
S1PR1+
Tregs
TNFRSF9+
Different states
Treg states-related genes (TSRGs)
Survival
Time
Unicox_sig genes
KM_sig genes
Genes in the final set (TSRPGs)
Establish a novel molecular classification system of RCC (RCC-TSC)
①ConsensusCluster-plus divided subtypes
②Correlation with multi-omics data
③Characterize three clusters
④Regulatory networks
⑤Inhibitors prediction
MIRC – moderate immune infiltration
HIRC – high immune infiltration (exhausted)
LIRC – low immune infiltration
TF
Key gene
Pathway
Immune component
Proteins
Genomics
Immunomics
Correlation analysis
Drugs targeted different clusters
Validation of RCC-TSC in our cohort (370 patients) with IHC staining and statistical analyses
① Clinical cohort and IHC staining
② Classifying patients into three clusters
③ Statistical analyses validated our classification in real-world cohort
A cohort of 370 patients with kidney neoplasm (most KIRC)
IHC staining for kidney neoplasm tissue (diagnostic kit)
LAG3 MKi67 DNAJB4 DNAJA1
HIRC
IHC scoring identified subtypes
LIRC
MIRC
HIRC
K-M survival analysis
Multivariate Cox regression analysis
B
Seurat clusters
CD4+ T cell clusters
Samples
UMAP_1
UMAP_2
CD4.c01.Tn.TCF7
CD4.c02.Tn.PASK
CD4.c03.Tn.ADSL
CD4.c04.Tn.il7r
CD4.c05.Tm.TNF
CD4.c06.Tm.ANXA1
CD4.c07.Tm.ANXA2
CD4.c08.Tm.CREM
CD4.c09.Tm.CCL5
CD4.c10.Tm.CAPG
CD4.c11.Tm.GZMA
CD4.c12.Tem.GZMK
CD4.c13.Temra.CX3CR1
CD4.c14.Th17.SLC4A10
CD4.c15.Th17.IL23R
CD4.c16.Tfh.CXCR5
CD4.c17.TfhTh1.CXCL13
CD4.c18.Treg.RTKN2
CD4.c19.Treg.S1PR1
CD4.c20.Treg.TNFRSF9
CD4.c21.Treg.OAS1
CD4.c22.ISG.IFIT1
CD4.c23.Mix.NME1
CD4.c24.Mix.NME2
RC-P20181217-N
RC-P20181217-T
RC-P20181224-N
RC-P20181224-T
RC-P20190918-N
RC-P20190918-T
RC-P20190919-N
RC-P20190919-T
RC-P20190923-N
RC-P20190923-T
RC-P20190925-N
RC-P20190925-T
C
High Low High Low
CD4.c18.Treg.RTKN2
CD4.c20.Treg.TNFRSF9
RTKN2 FOXP3 TIGIT MAL
TNFRSF9 TNFRSF18 FOXP3 CTLA4
CD4.c19.Treg.S1PR1
CD4.c21.Treg.OAS1
S1PR1 FOXP3 CCNB2 BIRC5
OAS1 FOXP3 SAT1 PIM2
tSNE1
tSNE2
D
Treg clusters
CD4.c18.Treg.RTKN2
CD4.c19.Treg.S1PR1
CD4.c20.Treg.TNFRSF9
CD4.c21.Treg.OAS1
E
Pseudotime
State
Component 1
Component 2
F
TSRGs
Unicox_sig
KM_sig
166
0
0
44
38
0
0

Fig. 1 ScRNA-seq analysis, monocle 2 trajectory analysis, and the determination of TSRPGs for CD4 T cells in RCC. **A** The schematic diagram of our study, primarily three parts. Firstly, we performed scRNA-seq analysis and specifically extracted Tregs for annotation (S1PR1+, TNFRSF9+, RTKN2+, OAS1+). Using Monocle 2 pseudotime analysis and prognostic analysis, we determined the TSRPGs. Secondly, ConsensusClusterPlus analysis established a novel classification (RCC-TSC, three subtypes), with significant prognostic and multi-omics meanings. Thirdly, after IHC staining for four key genes (DNAJA1, DNAJB4, LAG3, MKi67) and statistical analyses, we robustly validated RCC-TSC in Xinhua cohort. **B** The UMAP plot shown that, for scRNA-seq analysis, CD4+ T cells were specifically extracted, and 24 different cell clusters were annotated among the 12 samples. **C** The relative expression level and spatial distribution of the significant marker genes were demonstrated for Tregs. **D** The UMAP plots demonstrated the four Treg clusters, with the marker genes specifically pointed out. **E** The distribution of Tregs at different pseudotime and spatial locations was illustrated. With two bifurcation points, five Treg trajectory states were determined. Trajectory states 1 and 5 were positioned at the beginning of pseudotime, whereas trajectory states 3 and 4 were situated at the terminal branch. **F** The Venn plot demonstrated that with intersection of the TSRGs with genes significantly in Kaplan-Meier survival analysis and univariate Cox proportional hazards regression analysis ($p < 0.001$, K-M_sig, Unicox_sig), a gene set called the TSRPGs with 44 genes was built. ScRNA-seq, single-cell RNA sequencing; TSRPGs, Treg states-related prognostic genes; RCC, renal cell carcinoma; Treg, regulatory T cell; RCC-TSC, RCC-Treg states-related prognostic classification; IHC, immunohistochemical; UMAP, Uniform Manifold Approximation and Projection; TSRG, Treg states-related genes; OS, overall survival

2.4 Extract Tregs for scRNA-seq analysis

We extracted the four Treg clusters separately to specifically interrogate their biological role in RCC. Highly variable genes in each Treg cluster were selected based on a filtration standard of $|\log2\ FC| > 0.5$ and $p < 0.05$. Following the PCA and UMAP dimension reduction, marker genes of each Treg cluster were visualized through a heatmap and a cleveland dot plot. Moreover, gene set variation analysis (GSVA) was conducted to quantify the enrichment of cancer hallmarks, which were subsequently visualized in a heatmap [33].

2.5 Cellular communication analysis and cell cycle analysis

To elucidate the cellular communication and identify ligand-receptor pairs among all CD4+ T cells we performed cellular communication analysis by the "iTALK" R package (Version: 0.1.0) [34]. We also performed similar analysis with a specific focus on Tregs. Next, we selected the top 500 expressed genes to categorize cellular communication types into cytokines, growth factors, checkpoints and others. Then, the "FindLR" function helped identify the ligands and receptors pairs with the strongest connections, which were visualized through communication networks and LRplots. Moreover, to gain deeper insights into the cellular proliferation states, we performed cell cycle analysis using the "CellCycleScoring" function.

2.6 Trajectory analysis and identification of Treg states-related genes (TSRGs)

To identify the distinct cell states of Tregs at different pseudotime, trajectory analysis was conducted using the "Monocle" R package (Version: 2.18.0) [35, 36]. We filtrated out the low-quality cells and genes with low dispersion, and employed the "DifferentialGeneTest" function to select the top 1000 genes for ordering. Afterwards, the "DDRtree" method helped accomplish dimensionality reduction and identify the branch points in cell fates. After the pseudotime of the Tregs was determined, the DEGs for ordering were utilized to distinguish Tregs in different states, illustrating how differential gene expression prompted the divergence of cell fates toward distinct directions. Lastly, genes significantly associated with Treg states ($p < 0.05$) were named as TSRGs.

2.7 Determination of Treg States-related prognostic genes (TSRPGs)

To select the genes most statistically related to the OS of KIRC patients, we performed Kaplan-Meier (K-M) survival and univariate Cox proportional hazards regression analyses with strict criteria ($p < 0.001$) [37]. Afterwards, we intersected the genes with TSRPs, to build a gene set called TSRPGs. Specifically, the extracted genes needed to be expressed in both scRNA-seq and bulk-RNA-seq datasets, and a final set comprising 44 genes was established. The prognostic information and interrelationships between TSRPGs were visualized.

2.8 ConsensusClusterPlus analysis identified three novel molecular subtypes of RCC

To identify novel molecular subtypes based on the expression levels of TSRPGs among KIRC patients, we employed the "ConsensusClusterPlus" package for unsupervised clustering [38]. Taking the cumulative density function, the number of patients in each cluster (avoiding clusters with too few patients), as well as the high intra-cluster and low inter-cluster correlation into consideration, we determined that the optimal number of RCC subtypes was three. Next, we showed the heatmap that demonstrated the expression levels of TSRPGs in the three subtypes together with their clinical information, and we named our classification as RCC Treg states-related prognostic classification (RCC-TSC). Furthermore, a statistical comparison of TSRPGs'

expression among the three subtypes was performed and presented in boxplots.

2.9 Using PCA for patients of three subtypes to further validate RCC-TSC

After PCA, the principal components (PCs) of each subtype were visualized, and PCA scores for patients in each subtype were also calculated. The PCA score, a continuous variable, could provide additional meaning to verify RCC-TSC. By calculating the optimal cut-off value using "surv_cutpoint" function, patients were stratified into high PCA score group (H-PCA) and low PCA score group (L-PCA). Subsequent K-M survival analysis was performed to explore the predictive capability of PCA scores.

2.10 Correlation between multi-omics landscapes and RCC-TSC

The original multi-omics data, the Tumor Immune Dysfunction and Exclusion (TIDE) landmarks and immunophenoscore (IPS) files of KIRC patients were downloaded from TCGA and TIDE database respectively [39]. CIBERSORT was employed to comprehensively calculate the infiltration levels of 22 different immune cell subtypes in each patient. We assessed the immune infiltration landscapes and evaluate the immune function enrichment score using single-sample gene set enrichment analysis (ssGSEA) [40]. We also calculated the immune enrichment score of TIDE, dysfunction, exclusion, Merck18, myeloid-derived suppressor cell (MDSC), tumor-associated macrophage (TAM) M2, cancer-associated fibroblast (CAF), microsatellite instability (MSI), PD-L1, and IPS file (CTLA-4 and PD-1 expression), then comparing them among three RCC subtypes and exploring its relationship with PCA scores.

The mutation profiles were analyzed using the "Maftools" R package, which identified different mutation types. Next, we calculated the tumor mutation burden (TMB) and MSI [41]. Then, waterfall plots depicting the genes with the most frequent mutations were displayed for both L-PCA and H-PCA. Also, we compared the TMB/MSI between L-PCA and H-PCA, assessing their correlation. Subsequently, we employed K-M survival analysis to examine the combined predictive value of PCA scores and TMB in predicting prognosis. Stepwise, we determined the copy number variation (CNV) frequencies of the 44 genes via read-depth analysis, and the mutated genes were then mapped onto the chromosome and visualized using the "RCircos" R package.

2.11 Differential analysis of primary and metastatic KIRC in TCGA and MET500 databases

DEGs were identified between primary and metastatic KIRC patients in TCGA and MET500 databases using the "edge" R package [42] with thresholds set at $|\log2\ FC| > 1.0$ and adjusted p-value < 0.05. Next, expression levels of TSRPGs were compared between the primary and metastatic KIRC. Additionally, differential expression analysis was conducted for transcription factors and cancer hallmarks. The results were visualized by heatmaps, volcano plots respectively.

2.12 Construction of the regulatory networks of each RCC subtype

For each subtype, we performed correlation analysis between TSRPGs (with significant differences between primary and metastatic KIRC) and various components including upstream differentially expressed transcription factors (DETFs), and downstream factors including immune cells, immune function, cancer hallmarks, as well as proteomics. The thresholds of the correlation analysis were set at $|R$ (correlation coefficient)$| > n$ and $p < 0.05$, and n varied for different components and three subtypes, so that we could find the most significant pairs (with highest correlation coefficient). The results were depicted in three heatmaps, and the most significant components were chosen to construct the regulatory networks.

2.13 Anti-RCC drug response prediction for three different subtypes

Firstly, we used "pRRophetic" R package to predict potential drugs for each RCC subtype [43, 44]. Through analyzing the expression profiles of each subtype, the significant drugs were discovered and their drug sensitivity (IC50) was predicted. We also employed the CCLE and GDSC databases to predict inhibitors. Initially, we extracted the transcriptome and drug sensitivity data of the cancer cell lines. The drugs that were cytotoxic for more than 10% of the cell lines were excluded from our subsequent research. Next, we matched the CCLE cell lines to the three RCC subtypes. Secondly, we downloaded the sensitivity of tumors cell lines to cytotoxic drugs from GDSC database and explored their similarity to the CCLE cell lines. Finally, we could use the drug responses of the GDSC cell lines to evaluate the drug responses of the similar RCC subtypes. Besides, molecular docking analysis was performed using PyMOL (Version 1.3, Schrödinger, LLC) to prepare the 3D structure of LAG3 by removing water molecules. Adding hydrogen atoms and the generation of PDBQT files for both receptors and ligands were performed utilizing AutoDock Tools [45],

which also enabled the setup of a 3D grid box around the receptor's binding site. AutoDock Vina was then employed to predict the optimal binding conformations between ligands and receptors, with the docking results subsequently analyzed in PyMOL [46].

2.14 Assay for transposase-accessible chromatin sequencing

Assay from transposase-accessible chromatin sequencing (ATAC-seq) data of KIRC samples were obtained from the TCGA cohort of chromatin accessibility landscape of primary human cancers (https://gdc.cancer.gov/about-data/publications/ATACseq-AWG) [47]. ATAC-seq could validate the chromatin accessibility of the key genes in the three regulatory networks, and the analysis was performed with the UCSC genome browser (www.genome.ucsc.edu), "org.Hs.eg.db" R package and "TxDb.Hsapiens.UCSC.hg38.knownGene" R package. Each chromosome was partitioned into 92 segments, and the number of binding peaks was normalized by peak density. Subsequently, the total number of peaks in these segments was counted to characterize chromatin accessibility for the corresponding region.

2.15 Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analysis

KEGG pathway enrichment analysis was conducted for each subtype to identify the top 10 significant pathways. In the visualized bubble plots, the size of each bubble represented the number of genes enriched in the corresponding pathway, while the color indicated the significance of the enrichment.

2.16 Construction and validation of a prognostic risk-related prediction with 15 TSRPGs

Patients in the TCGA database with intact data including the information of metastasis were included. Firstly, univariate Cox proportional hazards regression analysis was conducted for TSRPGs. Secondly, least absolute shrinkage and selection operator (LASSO) regression analysis was performed to optimize the number of genes used in constructing the prognostic model [48]. Thirdly, 522 patients were randomly divided into a training set and a test set in a 6:4 ratio for model construction.

The formula of risk score was: risk score $= \beta_1 \times gene_1 + \beta_2 \times gene_2 + \ldots + \beta_n \times gene_n$ [49].

In this equation, "β" represents the coefficient of each gene, "gene" denotes the expression level of each gene, and "n" signifis the sequential number of the gene. Afterwards, patients in the train, test, all sets were respectively divided into the low-risk group and the high-risk group, and K-M survival analysis was performed. Additionally, receiver operating characteristic (ROC) curves were used to evaluate the sensitivity and specificity of the prediction model, and area under the curve (AUC) was calculated to quantify its effectiveness. Moreover, univariate and multivariate Cox proportional hazards regression analyses were conducted for age, TNM stage, and the risk score, to verify risk score as a significant and independent prognostic factor. Then, we explored the correlation between risk score and clinical variables using Chi-square tests. Moreover, we performed gene oncology (GO) and KEGG analysis and GSEA for the low-risk and high-risk groups. Finally, the fractions of various immune cell types and the TIDE landscapes were compared between low-risk and high-risk patients.

2.17 Using immunohistochemical (IHC) staining to classify patients of a large real-world cohort into different subtypes of RCC-TSC

A retrospective cohort study included 370 patients diagnosed with kidney neoplasms, primarily KIRC, who underwent curative surgery at Xinhua Hospital between 2016 and 2018. The paraffin samples from these patients were available and the patient information including demographic factors, tumor characteristics, follow-up data were collected. The follow-up period was ended in March 2021, and the survival status of each patient including OS and progression-free survival (PFS) were retrieved. PFS censor was defined as progression or death, and we classified patients into high or low PFS/OS groups with the corresponding median.

We selected four TSRPGs as the marker genes for subtype 2 (LAG3, Mki67), and subtype 3 (DNAJA1, DNAJB4) respectively. The four marker genes were selected based on strict criteria. Firstly, gene expression could be detected at the proteomic level in the tissue of renal tumor. Second, the gene expression level should have significant differences among the three subtypes and had the highest expression in the subtype with best or worst prognosis. Thirdly, the regulatory networks and previous literature were taken into account. We then performed IHC staining of the four markers on the tissue microarrays (TMA) of the patients. The processes of IHC staining were as follows. Above all, the kidney neoplasm tissues were fixed, embedded, and sliced into paraffin sections. Next, dewaxing and rehydration were performed, followed by antigen retrieval to expose specific antigens (LAG3, Mki67, DNAJA1, DNAJB4) via EDTA (pH = 9.0) or citric acid (pH = 6.0) antigen-retrieving buffers. Subsequently, to block the endogenous peroxidase, the samples were incubated in 3% hydrogen peroxide solution, and then washed in phosphate buffer saline solution. Also, the samples were treated with 3% bovine serum albumin

for prevention of the non-specific antibody binding. The primary antibodies, encompassing LAG3 (Proteintech, 29548-1-AP, 1:200), Mki67 (Proteintech, 27309-1-AP, 1:500), DNAJA1 (Affinity, DF8940, 1:100), and DNAJB4 (Proteintech, 13064-1-AP. 1:50) with different concentration, were then introduced to interact with the exposed antigens. Subsequently, secondary antibodies marked with HRP (an enzyme) were utilized to bind to the primary antigen-antibody complex. The specific substrate DAB for HRP was applied for the following chromogenic reaction, and counterstain with hematoxylin that could provide contrast and highlight the nuclei was also employed for inducing an evident color change. Following this, the samples underwent dehydration with alcohols of different concentration gradients. Eventually, pictures of all TMAs were obtained utilizing a microarray scanner.

Two experienced pathologists were enrolled to independently review and evaluate these TMAs. The conflicted results would be carefully checked by a third pathologist, and we finally obtained a comprehensive and dependable result of the acquired data. Specifically, the IHC scoring for each sample was determined based on both the area and intensity of staining. The area was defined as the proportion of stained tumor cells in the TMA, with 0%, 1–10%, 11–40%, 41–70%, 71–100% corresponding to 0–4 scores respectively. The intensity of the renal tumor cells was classified into negative (0), weak (1), moderate (2), and strong (3). Subsequently, the area and intensity scores were multiplied to get the IHC scores (ranging from 0 to 12), which were stratified into low (0–4), medium (5–8), and high (9–12). Finally, with the differential expression of the four key markers, we classified the patients into three subtypes of RCC-TSC.

2.18 Validation of RCC-TSC utilizing statistical analyses

At the beginning, we performed detailed descriptive statistics for the scores of IHC staining, together with all clinical variables, including age, gender, histopathological classification, Fuhrman nuclear stage, TNM stage, stage, OS censor, OS, PFS censor, PFS, and number of patients in three RCC-TSC subtypes. Notably, patients who with missing information were excluded when we conducted the following analyses. Firstly, we conducted K-M survival analysis to explore the relationship between OS, PFS and RCC-TSC in KIRC patients, all patients with kidney neoplasm, and patients excluding those with KIRC. Next, we performed Chi-square tests between other clinical variables and RCC-TSC. Then, we included age, gender, histopathological classifications, Fuhrman nuclear stage, stage and RCC-TSC into the multivariate Cox proportional hazards regression analysis, to explore whether RCC-TSC was an independent prognostic factor for OS and PFS. Based on the model, nomograms were constructed for OS and PFS prediction. Also, we generated risk score and explored its relationship with OS and PFS. Finally, we performed schoenfeld tests, and presented residual plots, ROC curves of the sensitivity and specificity tests, and calibration curves to comprehensively validate our model.

2.19 Quantitative statistical analysis

In our study, results meeting both a two-tailed p-value < 0.05 and a false discovery rate < 0.05 were considered statistically significant, unless explicitly otherwise. Analytical processes were performed using R (Institute for Statistics and Mathematics, version 4.2.3; www.r-project.org; Vienna, Austria) and Python version 3.6 software (https://www.python.org/). Descriptive statistics utilized the mean ± standard deviation for normally distributed continuous variables and the median with range for non-normally distributed continuous variables. Furthermore, parametric tests were employed to compare means of different groups when variables accorded with Gaussian distribution and homogeneity of variance. Otherwise, non-parametric test was utilized. Parametric and non-parametric tests respectively comprised the t-test, analysis of variance, Pearson correlation, and Wilcoxon test, Mann-Whitney U test, Kruskal-Wallis test, Spearman correlation.

3 Results

3.1 ScRNA-seq analysis revealed molecular characteristics of CD4+T cells in RCC

The schematic diagram and detailed flowchart were respectively presented in Fig. 1A and Figure S1, illustrating the primary processes of our study. Generally, we firstly performed scRNA sequencing, annotation, trajectory analysis, and prognostic analysis (TCGA cohort) to determine the TSRPGs. Secondly, we established a novel molecular classification system with three RCC subtypes (RCC-TSC), which demonstrated significant prognostic and immunological implications for RCC patients. Finally, we validated RCC-TSC in Xinhua cohort (370 patients) with IHC staining and statistical analyses, and a diagnostic kit was designed, which could be applied for patient diagnosis in the near future.

For scRNA-seq analysis, CD4+T cells were specifically extracted, and 24 different cell clusters were annotated among the 12 samples (Fig. 1B). We firstly demonstrated all DEGs in a scatter diagram (Figure S2A). The relative

expression level and spatial distribution of the significant marker genes were demonstrated for Tregs (Fig. 1C) and other CD4+ T cells (Figure S2B). Following that, the cleveland dot plot depicted the expression levels of the four key marker genes (CD3D, FOXP3, IL17A, IL21) across all 24 cell clusters (Figure S3A). CD3D was highly expressed in all clusters, while FOXP3 had higher expression in four Treg clusters and two clusters with mixed cells. Additionally, the percentage of each sample for the cell types were shown in bar plots, and Figure S3B visualized the cell numbers and proportions of the 12 sample sources. Also, the top 5 key genes of each cell type were illustrated in a heatmap (Figure S3C). In cellular communication analysis, Figure S3D showed that Tregs had many interactions with other CD4+ T cells. The cell cycle score plots revealed that a high proportion of S1PR1 + Treg and NME1 + mix T cells were at G2M phase (Figure S3E).

3.2 Integrated analyses of DEGs, cellular communication, and functional analysis in Tregs

Tregs exert significant inhibitory effects on the TIME, and to better elucidate the biological functions of Tregs in RCC, we extracted the Treg clusters for further analysis (Fig. 1D). Next, Figure S4A shows the expression levels of four key marker genes (S1PR1, TNFRSF9, RTKN2, OAS1) that characterized the four clusters. In Figure S4B, heatmap of the top five key genes for each Treg cluster displayed great heterogeneity among them. Cellular communication analysis revealed strong and complex interactions among the four clusters (Figure S4C). Considering the biological functions, GSVA identified the activated cancer hallmarks (Figure S4D). Notably, Hedgehog signaling, Kras signaling dn, and epithelial mesenchymal transition signaling pathways were most activated in all the four Treg clusters.

3.3 Trajectory analysis unveiled the temporal dynamics of gene expression and determined different Treg states

To figure out the heterogeneity and temporal dynamics of the four Treg clusters, we performed the trajectory analysis. Figure 1E illustrates the distribution of Tregs across different pseudotime points and spatial locations. With two bifurcation points, five Treg trajectory states were determined. Trajectory states 1 and 5 were positioned at the beginning of pseudotime, whereas trajectory states 3 and 4 were situated at the terminal branch. The four Tregs clusters were also plotted at the trajectory (Fig. 1E, Figure S5A). Comparatively, RTKN2-Treg and OSA1-Treg were respectively located at the beginning and ending of the pseudotime. Furthermore, the heatmap depicted that differential gene expression resulted in the divergence of distinct cell states (Figure S5B), validating their pivotal role in determining Treg states. As a result, these genes were termed Treg states-related genes (TSRGs), which were the candidate biomarkers in determining the Treg states.

3.4 Determination of the Treg States-related prognostic genes (TSRPGs)

Though the TSRGs were significantly associated with Treg temporal dynamics and states, their prognostic implications required further exploration. Consequently, we performed prognostic analysis on the 537 patients in the TCGA-KIRC cohort, whose baseline information is demonstrated in Table S1. As a result, 44 significant genes in univariate Cox proportional hazards regression analysis (Unicox_sig, $p < 0.001$) and 82 significant genes in K-M survival analysis (KM_sig, $p < 0.001$) for OS were obtained. Finally, we intersected the TSRGs, Unicox_sig, and KM_sig to build a gene set called the TSRPGs, which contained 44 genes (Fig. 1F). The prognostic information of the TSRPGs were demonstrated in Figure S6A and Table S2. We also built the co-expression network to visualize the relationship between the TSRPGs (Figure S6B). Moreover, TSRPGs in state 1 and state 3 were all identified as risk factors, while KLF6 in state 4 was a favorable prognostic factor.

3.5 Prognostic and immunological implications for RCC-TSC

To improve the integration and operability of the TSRPGs for prognostic prediction, we used clustering analysis to construct a classification system based on the expression levels of TSRPGs from TCGA database. As a result, we named the classification system of three RCC molecular subtypes as RCC Treg states-related prognostic classification (RCC-TSC). The CDF curve, delta area of the CDF, and consensus matrix were demonstrated in Figure S7A-C respectively. In Fig. 2A integrates clinical information, disease conditions, expression levels of TSRPGs, trajectory states, and three clusters in the heatmap. More importantly, in Fig. 2B, the survival probability of cluster 2 was the lowest among the three clusters ($p < 0.001$), manifesting the significant prognostic relation of RCC-TSC. It should be noted that cluster 2 and cluster 3 respectively had the highest and the lowest expression of TSRPGs in state 1 and 3 (all were risk factors), which could partially explain their differential prognosis (Fig. 2C). The immune infiltration of most immune cells was highest in cluster 2, particularly activated CD8 + T cells, activated dendritic cells (DCs), MDSCs, macrophages, NK cells and various CD4+ T cells (Fig. 2D). Consequently, we renamed cluster 1, 2, 3 as moderate immune infiltration

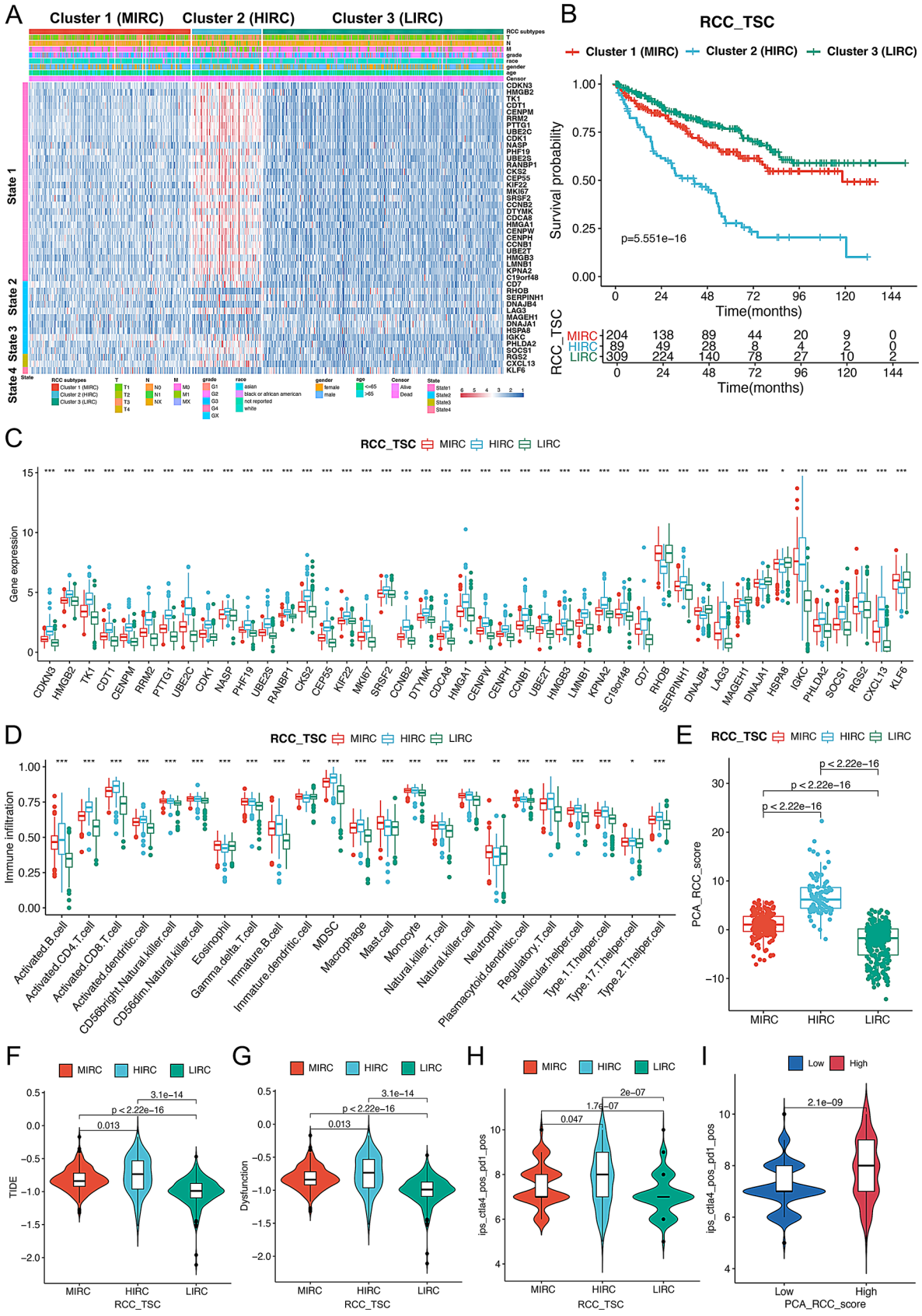

A
Cluster 1 (MIRC)
Cluster 2 (HIRC)
Cluster 3 (LIRC)
State 1
State 2
State 3
State 4
B
RCC_TSC
Cluster 1 (MIRC)
Cluster 2 (HIRC)
Cluster 3 (LIRC)
Survival probability
p=5.551e−16
Time(months)
MIRC 204 138 89 44 20 9 0
HIRC 89 49 28 8 4 2 0
LIRC 309 224 140 78 27 10 2
C
RCC_TSC MIRC HIRC LIRC
Gene expression
D
RCC_TSC MIRC HIRC LIRC
Immune infiltration
E
RCC_TSC MIRC HIRC LIRC
PCA_RCC_score
p < 2.22e−16
F
TIDE
RCC_TSC
0.013
3.1e−14
p < 2.22e−16
G
Dysfunction
RCC_TSC
H
ips_ctla4_pos_pd1_pos
RCC_TSC
0.047
1.7e−07
2e−07
I
Low High
ips_ctla4_pos_pd1_pos
PCA_RCC_score
2.1e−09

Fig. 2 Construction of a novel molecular classification (RCC-TSC) with significant prognostic and immunological implications. **A** The heatmap integrated the clinical information, disease conditions, expression levels of TSRPGs, trajectory states, and three RCC subtypes. **B** In the K-M survival plot, the survival probability of LIRC and HIRC were the highest and the lowest respectively ($p < 0.001$), manifesting the significant prognostic relation of RCC-TSC. **C** The expression level of TSRPGs was compared among the three clusters, which was all significant. Notably, cluster 2 (HIRC) and cluster 3 (LIRC) respectively had the highest and the lowest expression of genes in state 1 and 3, which were all risk factors. **D** The immune infiltration of most immune cells was highest in cluster 2, especially activated CD8 + T cells, activated DCs, MDSC, macrophage, NK cell and various CD4+ T cells. **E** The PCA scores of patients were calculated, while PCA scores in HIRC and LIRC were correspondingly the highest and the lowest. **F** The TIDE score was highest in HIRC, followed by MIRC and LIRC. **G** The dysfunction score was highest in HIRC, indicating HIRC had the highest number of T cells (especially CD8 cells) at late stage of dysfunction. **H** More patients in HIRC had higher expression for both CTLA4 and PD1. **I** Patients in H-PCA group had higher expression for both CTLA4 and PD1. RCC-TSC, renal cell carcinoma-Treg states-related prognostic classification; TSRPGs, Treg states-related prognostic genes; LIRC, low immune infiltration RCC, HIRC, high immune infiltration RCC. DC, dendritic cell; MDSC, myeloid-derived suppressor cell; PCA, principal component analysis; TIDE, tumor immune dysfunction and exclusion; MIRC, moderate immune infiltration RCC; H-PCA, high-PCA group

RCC (MIRC), high immune infiltration RCC (HIRC), low immune infiltration RCC (LIRC) respectively.

To further explore the prognostic relevance of RCC-TSC, and to reduce the potential bias introduced by categorical variables, we conducted PCA for the three RCC subtypes. The results were illustrated in Figure S7D, indicating the HIRC was greatly separated from LIRC. The PCA scores of patients in HIRC and LIRC were correspondingly the highest and the lowest (Fig. 2E). Then, Figure S7E suggests that high PCA scores of patients signified significantly shorter OS ($p < 0.001$). Finally, Figure S7F shows that PCA scores were positively correlated to the infiltration of most immune cells. In summary, using TSRPGs, we proposed an effective molecular classification with distinct OS and immune infiltration features for RCC patients.

3.6 Unveiling the characteristics of TIME for three RCC subtypes

Given the significant association between RCC-TSC and prognosis, as well as immune infiltration in RCC, we further investigated its relationship with immune dysfunction phenotypes. CD8+, IFNG, and Merck 18 (T-cell-inflamed signature) scores were highest in HIRC (Figure S8A-C). However, the TIDE score (Fig. 2F), dysfunction score (Fig. 2G), were also highest in HIRC, followed by MIRC and LIRC. It indicated HIRC had the highest number of pro-inflammatory cells (especially CD8 + cells), but with a high proportion at late stage of dysfunction [50, 51], which were associated with resistance of ICB reprogramming [51, 52]. No significant difference was found for exclusion score (Figure S8D), while MIRC and LIRC respectively had highest infiltration of CAF, MDSC and TAM M2 (Figure S8E-G). Besides, MSI was highest in LIRC (Figure S8H). PD-L1 did not differ significantly among three subtypes (Figure S8I), but had higher expression in H-PCA group (Figure S8J). Furthermore, patients in HIRC or H-PCA had higher expression of CTLA4 and PD1 (Fig. 2H-I, Figure S8K-L). These analyses revealed that the TIME of HIRC patients with H-PCA was the most dysfunctional, potentially correlating with a higher response rate to immune checkpoint blockade (ICB) therapies.

3.7 Multi-omics analysis of genomics and epigenomics

We delved into the genomic and epigenomic landscapes. Figure S9A and Figure S9B respectively illustrate the top 20 significantly mutated genes in the H-PCA and L-PCA groups. The mutation rate in KIRC patients was approximately 80%, suggesting a high level of genomic instability. VHL mutation dominated in both groups, followed by PBRM1, TTN, BAP1, and SETD2. Next, TMB and MSI showed no significant differences between H-PCA and L-PCA groups, and exhibited no linear relationship with PCA score (Figure S9C). However, patients with high TMB had significantly better OS (Figure S9D). Moreover, patients with H-PCA and H-TMB had the longest OS among all KIRC patients ($p < 0.001$) (Figure S9E). Almost all the TSRPGs exhibited CNV, with frequent changes observed in PTTG1, LMNB1 and DTYMK (Figure S9F), located in chromosomes 5, 5 and 2 respectively.

3.8 Performing differential analysis for genes, TFs, cancer hallmarks between primary and metastatic KIRC samples

Given that TIME is closely associated with tumorigenesis, progression, metastasis, while metastasis usually indicated a poor outcome, we conducted the differential analysis between primary and metastatic KIRC samples. Figure S10A-B presents the DEGs between primary and metastatic KIRC. All TSRPGs were validated as DEGs, and were specifically pointed out in the volcano plot. Subsequently, DETFs were identified and visualized in Figure S10C-D. Then, GSVA was employed to compare the differentially activated cancer hallmarks (Figure S10E). Afterwards, the GSVA scores were quantified, with activated and inhibited pathways presented at the top and bottom respectively (Figure S10F). Unfolded protein response, DNA repair, E2F targets were most activated, while KRAS signaling dn, pancreas beta cells, spermatogenesis were most downregulated.

3.9 Construction of the regulatory networks for three RCC subtypes

To figure out the metastasis-related molecular mechanisms, we built the regulatory networks by correlation analysis. These networks included the midstream TSRPGs and critical biological components (upstream DETFs and downstream factors including immune cells, immune functions, cancer hallmarks, as well as proteomics) (Fig. 3A). Components in the regulatory networks may play critical roles in the metastasis of three RCC subtypes correspondingly. The correlation coefficients between all the key biological components are shown in Fig. 3B. In MIRC, the dominant regulatory TSRPGs were LMNB1 and PTTG1, both of which exhibited significant co-expression patterns with DETFs (MYBL2, NCAPG, CENPA, LMNB1), cancer hallmarks (E2F targets), protein (CyclinB1), and immune components (T cells CD4+ memory activated and Tfh). In HIRC, the most aggressive subtype, the regulatory TSRPGs were LAG3 and HSPA8. LAG3 exhibited strong co-expression with upstream BATF ($R=0.77$), and downstream T cells CD8+ ($R=0.69$), CD8 + T cells ($R=0.65$), T cell co-inhibition ($R=0.74$), and Tfh ($R=0.69$). In LIRC, KLF6 and CXCL13 were the key TSRPGs in the regulatory network. CXCL13 showed a high correlation with IRF4 ($R=0.69$) and Tfh ($R=0.7$). Additionally, the expression levels of DETFs and TSRPGs in the three regulatory networks are illustrated in the heatmaps (Figure S11A). Furthermore, the 10 most significant pathways for the three RCC subtypes are displayed in the KEGG bubble plots (Figure S11B). Finally, ATAC-seq was conducted, validating the chromatin accessibility of six key regulatory TSRPGs (Figure S11C).

3.10 Prediction of drugs for three RCC subtypes

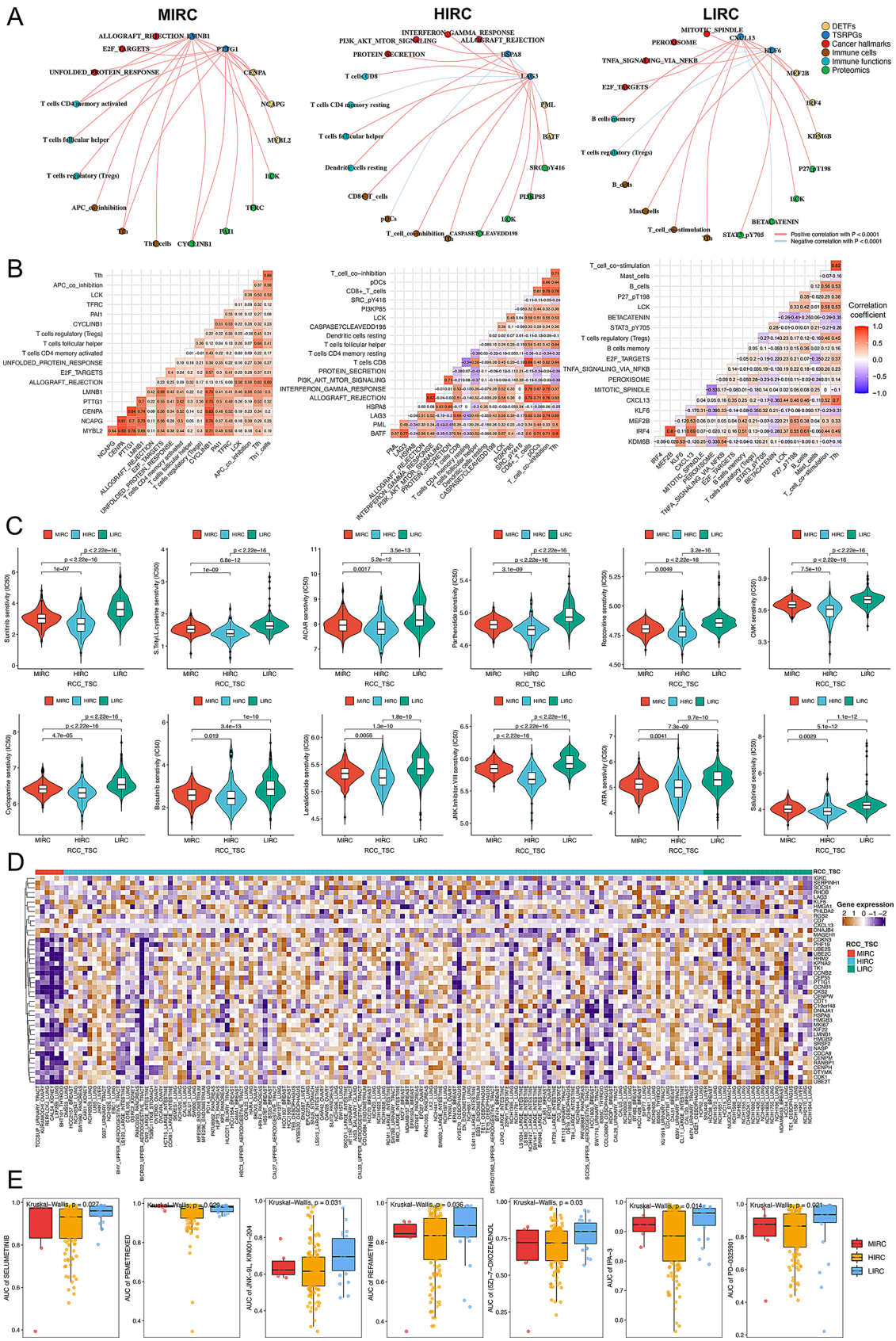
To individualize therapeutic strategies for the three RCC subtypes, we selected the drugs that could confer best therapeutic efficacy for each RCC subtype. We then utilized "pRRophetic" R package to predict the potential candidates of drugs (Fig. 3C). Drugs including sunitinib (a well-developed TKI for RCC patients) [2], cyclopamine (targets the Hedgehog pathway) [53, 54], bosutinib, lenalidomide, all-trans retinoic acid (ATRA) were potential drugs for HIRC. Drugs for treatment of MIRC and LIRC are respectively shown in Figure S12A-B. We subsequently mapped the expression matrix of three RCC subtypes with the cell lines in the CCLE database and the results are displayed with TSRPGs in Fig. 3D, and significant results among the predicted drugs for MIRC, HIRC, LIRC were respectively depicted in Fig. 3E and Figure S12C-D. Among them, selumetinib, pemetrexed, refametinib emerged as potential effective drugs for HIRC.

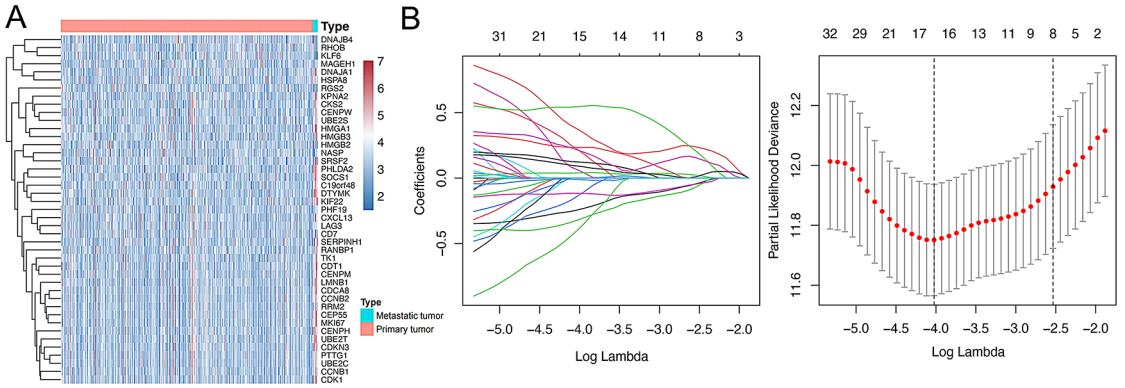
Fig. 3 The regulatory networks and potential inhibitors for three RCC subtypes. **A** The regulatory networks were built by correlation analysis, which contained the midstream TSRPGs and critical biological components (upstream DETFs and downstream factors including immune components, cancer hallmarks, as well as proteomics). **B** The correlation coefficients between all the key biological components were shown. In MIRC, the dominant regulatory TSRPGs were LMNB1 and PTTG1, which both exhibited significant co-expression pattern with DETFs (MYBL2, NCAPG, CENPA, LMNB1), cancer hallmarks (E2F targets), proteomics (CyclinB1), and immune components (T cells CD4 memory activated and Tfh). In HIRC, the most aggressive subtype, regulatory TSRPGs were LAG3 and HSPA8. LAG3 shown strong co-expression with upstream BATF ($R=0.77$), and downstream T cells CD8 ($R=0.69$), CD8 + T cells ($R=0.65$), T cell co-inhibition ($R=0.74$), and Tfh ($R=0.69$). In LIRC, KLF6 and CXCL13 dominated. And CXCL13 had high correlation with IRF4 ($R=0.69$) and Tfh ($R=0.7$). **C** Utilizing the "pRRophetic" R package, the potential candidates of drugs were predicted. Drugs including sunitinib, parthenolide, roscovitine, cyclopamine, bosutinib, lenalidomide, ATRA, salubrinal were potentially feasible drugs specifically for HIRC. **D** After mapping the resemblances of the expression matrix of three RCC subtypes with the cell lines in the CCLE database, the results with TSRPGs were displayed. **E** A comparative analysis of the therapeutic efficacy of drugs was conducted across the three clusters, and significant results of HIRC were illustrated in the boxplots. Selumetinib, pemetrexed, refametinib, could be potential drugs for HIRC. RCC, renal cell carcinoma; TSRPGs, Treg states-related prognostic genes; DETFs, differentially expressed transcription factors; MIRC, moderate immune infiltration RCC; HIRC, high immune infiltration RCC; LIRC, low immune infiltration RCC; ATRA, all-trans retinoic acid; CCLE, Cancer Cell Line Encyclopedia

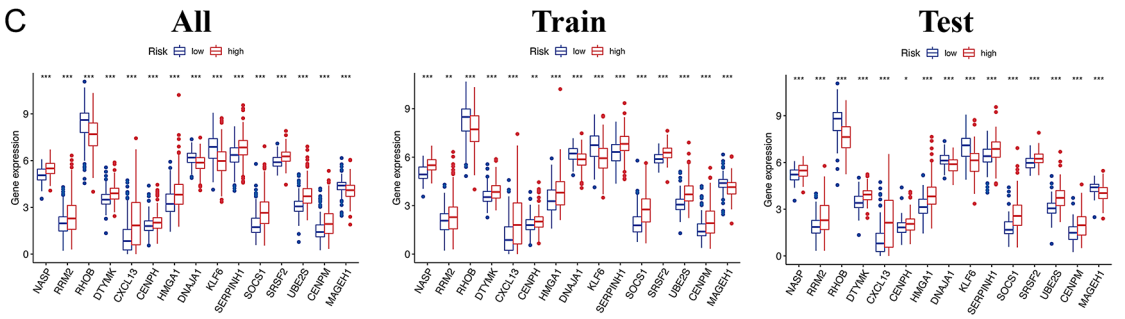
Potential binding patterns between LAG3 and sunitinib, cyclopamine were analyzed with molecular docking analysis (Figure S12E). When the sunitinib bonded with VAL-231, SER-182, and THR-266, it showed lowest affinity of -7.2 kcal/mol. While for the cyclopamine, it showed lowest affinity of -7.9 kcal/mol in the conformation bonded with ARG-210.

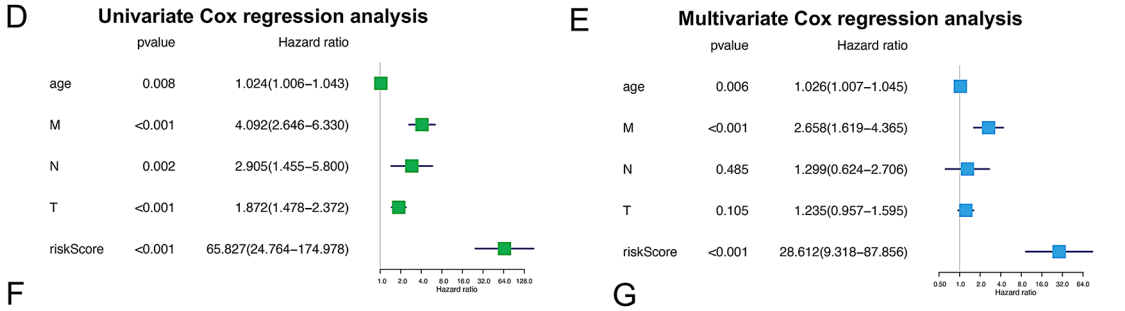
3.11 Construction and validation of a prognostic prediction model with TSRPGs

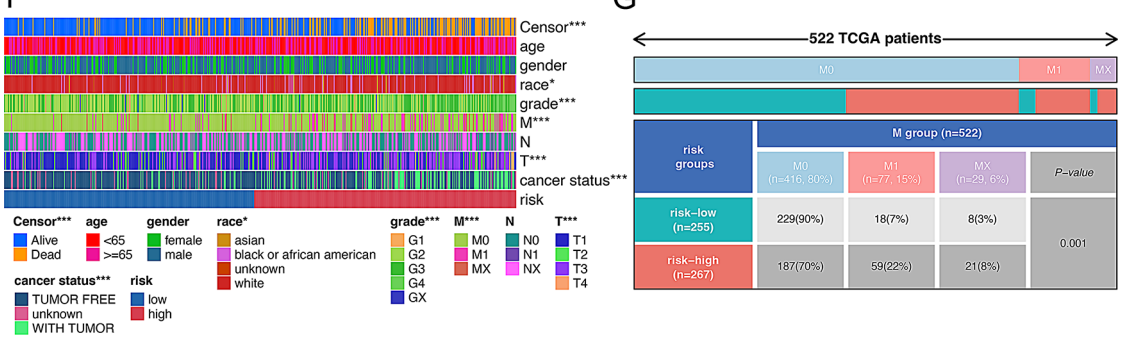
Using the TSRPGs to construct a prognostic prediction model, and demonstrated the expression level of TSRPGs between primary and metastatic KIRC (Fig. 4A). ORA unveiled TSRPGs were enriched in immune functions, cell cycle, cell division, and cancer-associated pathways (Figure S13A). The hazard ratio (HR) of TSRPGs in univariate Cox proportional hazards regression analysis are shown in Figure S13B. To mitigate overfitting of TSPRGs in our model, LASSO regression analysis was employed (Fig. 4B). A total of 15 genes, along with their corresponding coefficients were determined for the model. Subsequently, KIRC patients were randomly separated into a training set and a test set. The expression levels of the 15 genes were visualized in heatmaps (Figure S13C) and compared between the low-risk and high-risk patients in


A
MIRC
HIRC
LIRC
DETFs
TSRPGs
Cancer hallmarks
Immune cells
Immune functions
Proteomics
Positive correlation with P < 0.0001
Negative correlation with P < 0.0001
B
Correlation coefficient
C
MIRC HIRC LIRC
RCC_TSC
Sunitinib sensitivity (IC50)
S.Trityl.Lcysteine sensitivity (IC50)
AICAR sensitivity (IC50)
Parthenolide sensitivity (IC50)
Roscovitine sensitivity (IC50)
CMK sensitivity (IC50)
Cytarabine sensitivity (IC50)
Bosutinib sensitivity (IC50)
Lenalidomide sensitivity (IC50)
JNK.Inhibitor.VIII sensitivity (IC50)
ATRA sensitivity (IC50)
Salubrinal sensitivity (IC50)
D
Gene expression
RCC_TSC
MIRC
HIRC
LIRC
E
Kruskal–Wallis, p = 0.027
Kruskal–Wallis, p = 0.029
Kruskal–Wallis, p = 0.031
Kruskal–Wallis, p = 0.036
Kruskal–Wallis, p = 0.03
Kruskal–Wallis, p = 0.014
Kruskal–Wallis, p = 0.021
AUC of SELUMETINIB
AUC of PEMETREXED
AUC of JNK−IN−8, KINOM1−204
AUC of REFAMETINIB
AUC of (5Z)−7−OXOZEAENOL
AUC of ISPA−3
AUC of PD−0325901
MIRC
HIRC
LIRC


A
Type
DNAJB4
RHOB
KLF6
MAGEH1
DNAJA1
HSPA8
RGS2
KPNA2
CKS2
CENPW
UBE2S
HMGA1
HMGB3
HMGB2
NASP
SRSF2
PHLDA2
SOCS1
C19orf48
DTYMK
KIF22
PHF19
CXCL13
LAG3
CD7
SERPINH1
RANBP1
TK1
CDT1
CENPM
LMNB1
CDCA8
CCNB2
RRM2
CEP55
MKI67
CENPH
UBE2T
CDKN3
PTTG1
UBE2C
CCNB1
CDK1
7
6
5
4
3
2
Type
Metastatic tumor
Primary tumor
B
31 21 15 14 11 8 3
Coefficients
0.5
0.0
−0.5
−5.0 −4.5 −4.0 −3.5 −3.0 −2.5 −2.0
Log Lambda
32 29 21 17 16 13 11 9 8 5 2
Partial Likelihood Deviance
12.2
12.0
11.8
11.6
−5.0 −4.5 −4.0 −3.5 −3.0 −2.5 −2.0
Log Lambda


C
All
Train
Test
Risk low high
Gene expression
NASP RRM2 RHOB DTYMK CXCL13 CENPH HMGA1 DNAJA1 KLF6 SERPINH1 SOCS1 SRSF2 UBE2S CENPM MAGEH1


D
Univariate Cox regression analysis
pvalue Hazard ratio
age 0.008 1.024(1.006−1.043)
M <0.001 4.092(2.646−6.330)
N 0.002 2.905(1.455−5.800)
T <0.001 1.872(1.478−2.372)
riskScore <0.001 65.827(24.764−174.978)
1.0 2.0 4.0 8.0 16.0 32.0 64.0 128.0
Hazard ratio
E
Multivariate Cox regression analysis
pvalue Hazard ratio
age 0.006 1.026(1.007−1.045)
M <0.001 2.658(1.619−4.365)
N 0.485 1.299(0.624−2.706)
T 0.105 1.235(0.957−1.595)
riskScore <0.001 28.612(9.318−87.856)
0.50 1.0 2.0 4.0 8.0 16.0 32.0 64.0
Hazard ratio


F
Censor***
age
gender
race*
grade***
M***
N
T***
cancer status***
risk
Censor*** Alive Dead
age <65 >=65
gender female male
race* asian black or african american unknown white
grade*** G1 G2 G3 G4 GX
M*** M0 M1 MX
N N0 N1 NX
T*** T1 T2 T3 T4
cancer status*** TUMOR FREE unknown WITH TUMOR
risk low high
G
522 TCGA patients
M0 M1 MX
risk groups
M group (n=522)
M0 (n=416, 80%)
M1 (n=77, 15%)
MX (n=29, 6%)
P-value
risk−low (n=255)
229(90%) 18(7%) 8(3%)
risk−high (n=267)
187(70%) 59(22%) 21(8%)
0.001

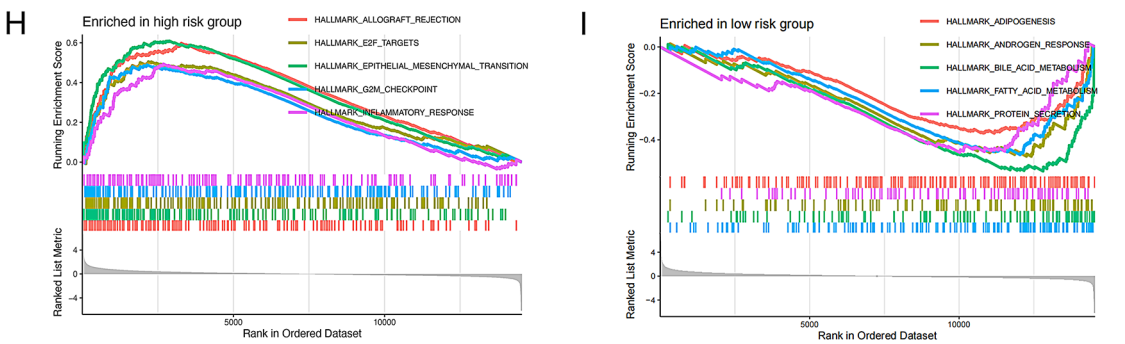

H
Enriched in high risk group
HALLMARK_ALLOGRAFT_REJECTION
HALLMARK_E2F_TARGETS
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION
HALLMARK_G2M_CHECKPOINT
HALLMARK_INFLAMMATORY_RESPONSE
Running Enrichment Score
Ranked List Metric
Rank in Ordered Dataset
I
Enriched in low risk group
HALLMARK_ADIPOGENESIS
HALLMARK_ANDROGEN_RESPONSE
HALLMARK_BILE_ACID_METABOLISM
HALLMARK_FATTY_ACID_METABOLISM
HALLMARK_PROTEIN_SECRETION
Running Enrichment Score
Ranked List Metric
Rank in Ordered Dataset

Fig. 4 Construction and validation of a prognostic prediction model with TSRPGs. **A** The heatmap demonstrated the expression level of TSRPGs between primary and metastatic RCC. **B** To mitigate the overfitting impact of TSPRGs in our model, we employed LASSO regression analysis, and 15 genes together with their corresponding coefficients were determined for the model. **C** The expression level of the selected genes was compared between the low-risk and high-risk patients in all, train, test sets respectively, which all demonstrated statistical significances. **D** Through univariate Cox regression analysis, risk score was verified as a significant factor (HR = 65.827, 95% CI = 24.764-174.978, $p < 0.001$). **E** Through multivariate Cox regression analysis, risk score was validated as an independent factor (HR = 28.612, 95% CI = 9.318–87.856, $p < 0.001$). **F** The Chi-square test demonstrated patients in high-risk group were significantly associated higher probability of death, higher grade, T stage, M stage, and worse cancer status. **G** High-risk patients had significantly lower proportion of M0 stage ($p = 0.001$). **H** The five most significant hallmarks for high-risk patients were presented, prominently featuring cell cycle-related pathways, EMT and inflammatory response pathways. **I** Adipogenesis, androgen responses signaling hallmarks were highly activated in the low-risk patients. TSRPGs, Treg states-related prognostic genes; RCC, renal cell carcinoma; LASSO, least absolute shrinkage and selection operator; HR, hazard ratio; CI, confidential interval

three sets respectively (Fig. 4C), all of which demonstrated statistical significances. Next, the risk score of each patient was calculated, and combined the risk score together with OS in scatter plots (Figure S13D). A high risk score was associated with shorter OS, as validated in K-M survival plots (Figure S13E). ROC curves and AUC values demonstrated the robust performance of the model (Figure S13F). Then, univariate and multivariate Cox proportional hazards regression analyses were used, validating risk score as a significant risk factor (Fig. 4D, HR = 65.827, 95% CI = 24.764-174.978, $p < 0.001$) and an independent risk factor (Fig. 4E, HR = 28.612, 95% CI = 9.318–87.856, $p < 0.001$). Moreover, Fig. 4F shows patients in high-risk group were significantly associated a higher probability of death, higher grade, T stage and M stage (Fig. 4G), and worse cancer status. In Fig. 4H presents the five most significant hallmarks for high-risk patients, prominently featuring cell cycle-related pathways, EMT, and inflammatory response pathways. In Fig. 5I demonstrates that adipogenesis and androgen responses signaling hallmarks dominate in the low-risk patients. GO and KEGG pathways for the two groups are respectively demonstrated in Figure S13G-H.

3.12 Exploring the association between immune components and KIRC patients with different risks

In the immune subtype analysis, low-risk KIRC patients were more frequently associated with C3-subtype (inflammatory) (Figure S14A, $p = 0.002$), characterized by elevated expression of Th17 and Th1 genes. The C3-subtype exhibited low to moderate cellular proliferation, lower levels of chromosomal aneuploidy, higher somatic CNVs, and the most favorable prognosis [55]. Next, the proportion of 22 immune cells were extracted (Figure S14B), and the immune cells and scores of immune functions were then compared between the low-risk and high-risk groups (Figure S14C-D). Significant differences were observed between high-risk group and low-risk group, which was further significantly associated with distinct survival probabilities (Figure S14E-F). Additionally, TIDE, dysfunction, CD8+, Merck18, and exclusion scores were higher in high-risk KIRC patients (Figure S15A-E). Low-risk KIRC patients exhibited higher MSI, CD274, TAM M2 and MDSC infiltration, but lower CAF infiltration (Figure S15F-K). These findings validated the significant association between immune components and KIRC patients with different risks.

3.13 RCC-TSC was validated as an independent prognostic factor in Xinhua cohort

The flowchart of our cohort study is shown in Fig. 5A, and the analytic processes primarily composed of three parts, including IHC staining for the retrospective cohort, molecular classification, and statistical validation. The relative expression levels of the markers used for IHC staining and classification are shown (Fig. 5B). MKi67 and LAG3 expressed highest in HIRC, while DNAJA1 and DNAJB4 had highest expression in LIRC. Our retrospective cohort contained 370 patients, and the detailed information was presented in Table 1; Fig. 5C. 6.22% of the patients were deceased and 11.08% patients suffered PFS censor. Considering the age category, patients in 21–40, 41–60, 61–88 groups were 8.92%, 52.43%, 38.65% respectively. 69.19% patients were male and most patients (88.65%) were diagnosed with KIRC. Besides, 48.11% and 28.11% patients were in T stage 1a and 1b, while only 0.54% patients had node metastasis and no patient had distant metastasis, manifesting that most patients in this cohort were in early stages. The K-M survival plots with significant results regarding OS and PFS for the clinical variables of our cohort are shown in Figure S16A-B respectively. We illustrate part of the results for IHC staining in Fig. 5D, and we classified the 370 patients as HIRC (122, 32.97%), MIRC (139, 37.57%), and LIRC (108, 29.19%) respectively. The IHC scores of the four markers were compared among the three subtypes (Figure S16C), demonstrating significant differences.

The K-M survival plots showed that patients in the HIRC group had the lowest OS and PFS in KIRC patients (Fig. 6A, OS, $p = 0.003$; PFS, $p < 0.001$). More importantly, the K-M survival plots in patients with kidney neoplasm (OS, PFS, $p < 0.001$), and patients excluding those with KIRC (OS, $p = 0.050$; PFS, $p = 0.052$) were demonstrated respectively (Figure S17). These results validated

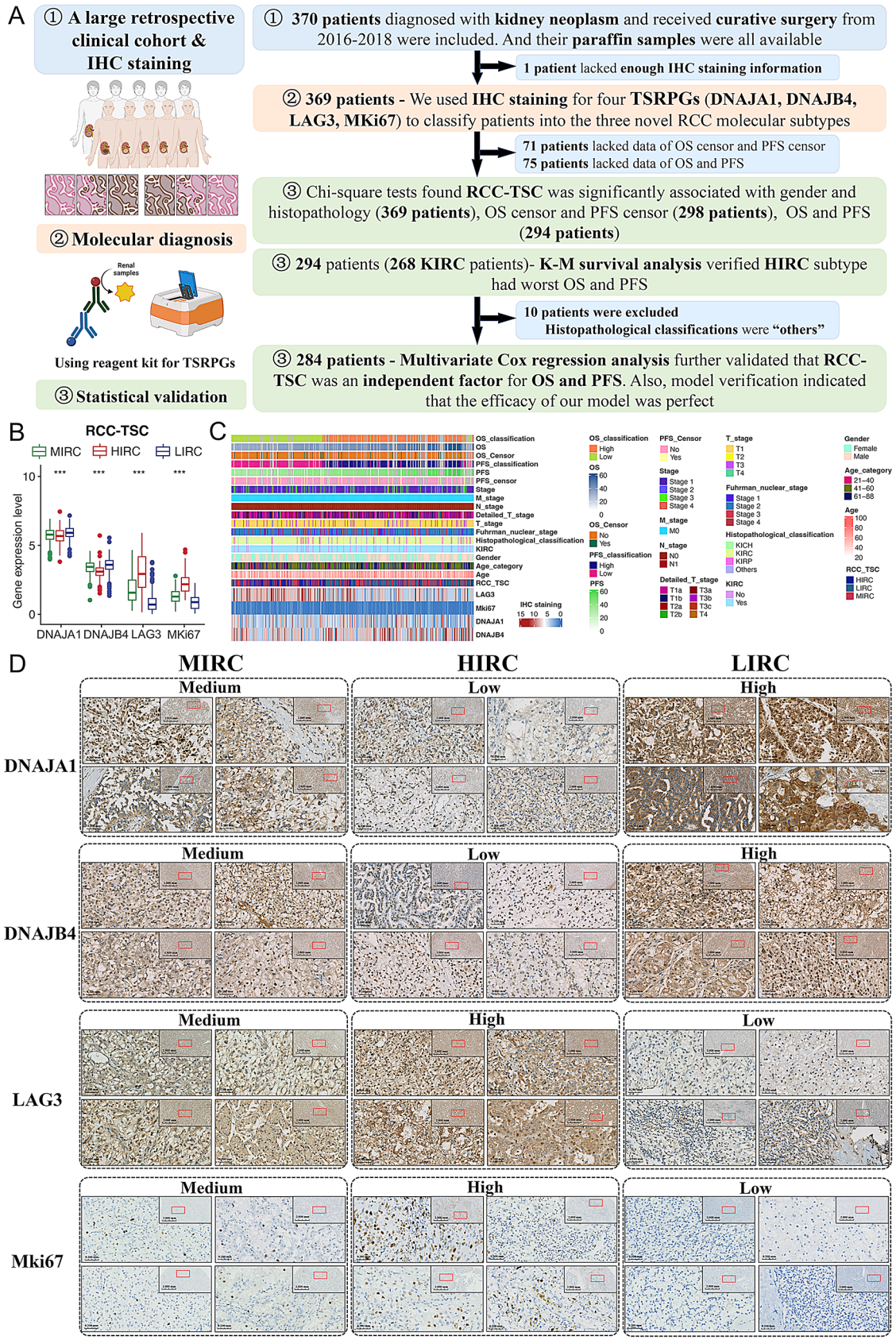

A
① A large retrospective clinical cohort & IHC staining
② Molecular diagnosis
Renal samples
Using reagent kit for TSRPGs
③ Statistical validation
① 370 patients diagnosed with kidney neoplasm and received curative surgery from 2016-2018 were included. And their paraffin samples were all available
1 patient lacked enough IHC staining information
② 369 patients - We used IHC staining for four TSRPGs (DNAJA1, DNAJB4, LAG3, MKi67) to classify patients into the three novel RCC molecular subtypes
71 patients lacked data of OS censor and PFS censor
75 patients lacked data of OS and PFS
③ Chi-square tests found RCC-TSC was significantly associated with gender and histopathology (369 patients), OS censor and PFS censor (298 patients), OS and PFS (294 patients)
③ 294 patients (268 KIRC patients)- K-M survival analysis verified HIRC subtype had worst OS and PFS
10 patients were excluded
Histopathological classifications were "others"
③ 284 patients - Multivariate Cox regression analysis further validated that RCC-TSC was an independent factor for OS and PFS. Also, model verification indicated that the efficacy of our model was perfect
B
RCC-TSC
MIRC HIRC LIRC
Gene expression level
DNAJA1 DNAJB4 LAG3 MKi67
C
OS_classification
OS
OS_Censor
PFS_classification
PFS
PFS_censor
Stage
M_stage
N_stage
Detailed_T_stage
T_stage
Fuhrman_nuclear_stage
Histopathological_classification
KIRC
Gender
Age_category
Age
RCC_TSC
LAG3
Mki67
DNAJA1
DNAJB4
IHC staining
D
MIRC
HIRC
LIRC
DNAJA1
DNAJB4
LAG3
Mki67
Medium
Low
High

Fig. 5 IHC staining and scoring for classifying patients in Xinhua cohort based on RCC-TSC. **A** The flowchart of our study. Our clinical study was mainly composed of three analytic processes, including IHC staining for the tumor tissue of the 370 patients with kidney neoplasm, molecular diagnosis based on the RCC-TSC, and statistical validation. The inclusion and exclusion criteria were also shown in detail. Firstly, 370 patients with kidney neoplasm and received curative surgery from 2016–2018 were included. Secondly, due to the incompleteness of the IHC staining information (DNAJA1, DNAJB4, LAG3, Mki67), one patient was excluded. Thirdly, when performing Chi-square tests and Kaplan-Meier survival analysis, 71 (lacking OS censor and PFS censor) or 75 patients (lacking OS and PFS) were correspondingly excluded. Finally, when conducting the multivariate proportional hazards Cox regression analysis, 10 patients whose histopathological classifications were not KICH, KIRC, or KIRP were excluded. **B** The boxplots showed the relative expression in transcriptional level for the four markers we used for IHC staining and classification. MKi67 and LAG3, DNAJA1 and DNAJB3 respectively expressed highest in HIRC and LIRC. **C** The heatmap illustrated the information of the 370 patients in our retrospective cohort. **D** IHC slides demonstrated the differential staining results of DNAJA1, DNAJB4, LAG3, Mki67 in three RCC-TSC subtypes. Specifically, DNAJA1 and DNAJB4 both had relatively high, medium, and low staining score in LIRC, MIRC, and HIRC. While LAG3, Mki67 both had relatively high, medium, and low staining score in HIRC, MIRC, and LIRC. IHC, immunohistochemical; RCC-TSC, renal cell carcinoma-Treg states-related prognostic classification; OS, overall survival; PFS, progression-free survival; KICH, kidney chromophobe; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LIRC, low immune infiltration RCC; MIRC, medium immune infiltration RCC; HIRC, high immune infiltration RCC

its effectiveness in KIRC, and its potential feasibility across various kidney neoplasms. Moreover, Fig. 6B shows that RCC-TSC was also significantly associated with gender, histopathological classification, OS censor, PFS classification, PFS censor and OS classification. The proportion of male was highest in LIRC (78%), while lowest in MIRC (63%) (Fig. 6C). Percentage of KIRC was highest in MIRC (93%), and all the classification labeled as others were in LIRC (Fig. 6D). Moreover, HIRC conferred higher probability of death (13%, Fig. 6E), rates of PFS censor (23%, Fig. 6F), as well as lower OS (Fig. 6G) and PFS classification (Fig. 6H). More importantly, Fig. 6I-J validated HIRC as an independent risk factor for both OS (HR = 16.68, 95% CI = 1.88–148.1, $p = 0.0115$) and PFS (HR = 4.43, 95% CI = 1.55–12.6, $p = 0.0054$). We also constructed nomograms for OS and PFS prediction (Figure S18A). Then, we used residual plots, schoenfeld tests (Figure S18B), calibration curves, ROC curves (Fig. 6K-L) to comprehensively validate our model's accuracy, robustness and consistency. Finally, we also built a risk-related model based on the multivariate Cox proportional hazards regression analysis, proving its significant association with patients' survival (Figure S18C) and tumor stages (Figure S18D).

Table 1 Demographic information and clinical characteristics of the 370 patients

Variables	Number (%)	Mean ± SD; Median (range)	*P* value
LAG3		6.60 ± 3.16; 8 (0–12)	
DNAJB4		6.78 ± 3.11; 8 (0–12)	
DNAJA1		4.89 ± 2.72; 4 (0–12)	
MKi67		0.27 ± 0.50; 0 (0–3)	
OS censor			0.008*
Alive	276 (74.59)		
Dead	23 (6.22)		
Unknown	71 (19.19)		
OS (months)		42.39 ± 9.18; 41 (7–62)	
OS classification			< 0.001*
High (≥ 41)	163 (44.05)		
Low (< 41)	132 (35.68)		
Unknown	75 (20.27)		
PFS censor			0.002*
Yes	41 (11.08)		
No death/progression	258 (69.73)		
Unknown	71 (19.19)		
PFS (months)		40.90 ± 11.07; 41 (0–62)	
PFS classification			< 0.001*
High (≥ 41)	154 (41.62)		
Low (< 41)	141 (38.11)		
Unknown	75 (20.27)		
Age		56.61 ± 12.47; 55 (21–88)	
Age (category)			0.502
21–40	33 (8.92)		
41–60	194 (52.43)		
61–88	143 (38.65)		
Gender			0.037*
Female	114 (30.81)		
Male	256 (69.19)		
KIRC			0.062
Yes	329 (88.92)		
No	41 (11.08)		
Histopathological classification			0.002*
KIRC	328 (88.65)		
KIRP	18 (4.86)		
KICH	8 (2.16)		
Others	16 (4.32)		
Fuhrman nuclear stage			0.227
Stage 1	69 (18.65)		
Stage 2	234 (63.24)		
Stage 3	51 (13.78)		
Stage 4	16 (4.32)		
T stage			0.801
T1	282 (76.22)		

Table 1 (continued)

Variables	Number (%)	Mean ± SD; Median (range)	*P* value
T2	29 (7.84)		
T3	53 (14.32)		
T4	3 (0.81)		
Unknown	3 (0.81)		
Detailed T stage			0.276
T1a	178 (48.11)		
T1b	104 (28.11)		
T2a	17 (4.59)		
T2b	12 (3.24)		
T3a	49 (13.24)		
T3b	2 (0.54)		
T3c	2 (0.54)		
T4	3 (0.81)		
Unknown	3 (0.81)		
N stage			
N1	2 (0.54)		
N0	368 (99.46)		
M stage			
M1	0 (0.00)		
M0	370 (100.00)		
Stage			0.781
Stage 1	282 (76.22)		
Stage 2	28 (7.84)		
Stage 3	54 (14.59)		
Stage 4	3 (0.81)		
Unknown	3 (0.81)		
RCC-TSC			
Cluster 1 (MIRC)	139 (37.57)		
Cluster 2 (HIRC)	122 (32.97)		
Cluster 3 (LIRC)	108 (29.19)		
Unknown	1 (0.27)		

Abbreviations SD, standard deviation; OS, overall survival; PFS, progression-free survival; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; KICH, kidney chromophobe renal cell carcinoma; RCC-TSC, renal cell carcinoma-Treg states-related prognostic classification; MIRC, moderate immune infiltration renal cell carcinoma; HIRC, high immune infiltration renal cell carcinoma; LIRC, low immune infiltration renal cell carcinoma. Variables whose p value of the Chi-square tests were less than 0.05 were signed with "*"

3.14 Using external databases to validate the results of our annotation in the scRNA-seq analysis and the risk-related prediction model

In Figure S19A, 7 bulk labels and 17 sub labels are defined in the scRNA-seq analysis. Feature plots of the marker genes, such as CD4, CD3E, CD3D, TGFβ1, and IL10 are displayed (Figure S19B). The cell number and cell proportion of different cell types are shown in the bar plots (Figure S19C). Subsequently, we compared the distribution of cell types between adjacent normal tissue and tumor tissue (Figure S20A). As illustrated in Figure S20B, the number and proportion of cells co-expressing double-positive markers were higher in the tumor tissue, supporting the potential promoting role of Tregs in the tumor microenvironment. Furthermore, in Figure S21A, we demonstrate strong co-expression of FOXP3, CD4, and the four previously mentioned markers (RTKN2, S1PR1, TNFRSF9, OAS1). Lastly, Figure S21B illustrates the four markers used for IHC staining.

In the GSE22541 dataset, 12 of the 15 genes in the risk-related prediction model were found to be significant for patients' OS, and eight genes were significant for patients' disease free survival (Figure S22A-B). Moreover, up to 32 TSRPGs had significantly different expression between samples of the primary KIRC and the pulmonary metastasis of KIRC (Figure S22C). These results collectively indicated the significance of TSRPGs and especially the genes in the risk-related prediction model.

4 Discussion

RCC, as a highly heterogenous malignancy, poses challenges in interpreting its biological characteristics and patients' differential prognosis [2, 4, 56]. Recent research has unveiled the highly immune infiltrated but exhausted characteristics of RCC in the TIME. The mechanisms underlying the phenomenon include the strong immunosuppressive functions of Tregs, which significantly inhibit the anti-tumor effect of CD8 + T cells [8, 57, 58]. Moreover, a pan-cancer atlas has depicted the dynamics and heterogeneity of four Treg subtypes (TNFRSF9+, OAS1+, RTKN2+, S1PR1+) [10]. To bridge the gap that relationship between different states of Tregs and the heterogeneity of RCC plus the prognosis of patients hasn't been determined, we conducted a comprehensive bioinformatic, pathological and clinical study.

Using scRNA-seq and Monocle 2 pseudotime analyses, we identified the genes that significantly determined Treg states (TSRG). Then, utilizing ConsensusClusterPlus algorithm, RCC-TSC was proposed with HIRC identified as the subtype exhibiting the worst prognosis and highest immune infiltration. Additionally, high scores of TIDE and dysfunction indicated a highly exhausted state for T cells [8]. Afterwards, we predicted the possible drugs for each subtype. It was found that drugs including sunitinib, cyclopamine, bosutinib, lenalidomide, all-trans retinoic acid (ATRA), selumetinib, pemetrexed, refametinib could be the potential inhibitors for HIRC. Besides, we used TSRPGs for construction of an effective prognostic model on the basis of risk score, and intercorrelated the model, and RCC-TSC with multi-omics data. Finally, utilizing the four TSRPGs (DNAJA1, DNAJB4, LAG3, Mki67), we successfully classified patients with kidney neoplasm into three RCC-TSC


A
OS
KIRC patients
PFS
MIRC
HIRC
LIRC
Overall survival
p=0.003
Progression−free survival
p<0.001
Time(months)
RCC_TSC
MIRC 104 103 103 83 21 0
HIRC 86 84 82 59 4 1
LIRC 78 78 78 75 35 18
MIRC 104 102 101 78 21 0
HIRC 86 80 76 53 4 1
LIRC 78 78 78 73 34 18
B
Age_category
Gender*
KIRC
Histopathological_classification**
Fuhrman_nuclear_stage
T_stage
Detailed_T_stage
Stage
PFS_censor**
PFS_classification***
OS_censor**
OS_classification***
RCC_TSC
21–40
41–60
61–88
Female
Male
No
Yes
KICH
KIRC
KIRP
Others
Stage 1
Stage 2
Stage 3
Stage 4
T1
T2
T3
T4
T1a
T1b
T2a
T2b
T3a
T3b
T3c
T4
High
Low
C
Gender
MIRC
HIRC
LIRC
63%
37%
69%
31%
78%
22%
Male
Female
D
Histopathological classification
93%
3%
4%
1%
89%
7%
3%
82%
10%
5%
3%
Others
KIRP
KIRC
KICH
E
OS Censor
8%
92%
13%
87%
1%
99%
Yes
No
F
PFS_Censor
11%
89%
23%
77%
7%
93%
G
OS classification
47%
53%
26%
74%
9%
91%
Low
High
H
PFS classification
51%
49%
23%
77%
10%
90%
I
OS
Hazard ratio
Age_category
21–60 (N=178) reference
61–88 (N=106) 1.12 (0.42 − 3.0) 0.8221
Gender
Female (N=87) reference
Male (N=197) 1.38 (0.48 − 4.0) 0.5475
Histopathological_classification
KIRC (N=268) reference
KIRP (N=9) 3.17 (0.79 − 12.6) 0.1024
KICH (N=7) 11.55 (1.23 − 108.6) 0.0324 *
Fuhrman_nuclear_stage
Stage 1 (N=57) reference
Stage 2 (N=178) 0.95 (0.19 − 4.8) 0.9501
Stage 3 & 4 (N=51) 1.84 (0.33 − 10.2) 0.4834
Stage
Stage 1 (N=282) reference
Stage 2 (N=19) 5.91 (1.33 − 26.2) 0.0195 *
Stage 3 & 4 (N=33) 7.62 (2.50 − 23.2) < 0.001 ***
RCC_TSC
LIRC (N=81) reference
HIRC (N=93) 16.68 (1.88 − 148.1) 0.0115 *
MIRC (N=110) 7.12 (0.82 − 62.2) 0.0758
Events: 19; Global p−value (Log−Rank): 2.7082e−05
AIC: 187.57; Concordance Index: 0.86
J
PFS
Hazard ratio
Age_category
21–60 (N=178) reference
61–88 (N=106) 0.93 (0.45 − 1.9) 0.8339
Gender
Female (N=87) reference
Male (N=197) 1.19 (0.54 − 2.6) 0.6741
Histopathological_classification
KIRC (N=268) reference
KIRP (N=9) 2.03 (0.58 − 7.0) 0.2655
KICH (N=7) 3.83 (0.47 − 31.1) 0.2087
Fuhrman_nuclear_stage
Stage 1 (N=57) reference
Stage 2 (N=178) 0.97 (0.31 − 3.0) 0.9619
Stage 3 & 4 (N=51) 1.66 (0.49 − 5.7) 0.4161
Stage
Stage 1 (N=282) reference
Stage 2 (N=19) 4.57 (1.50 − 14.0) 0.0076 **
Stage 3 & 4 (N=33) 8.03 (3.63 − 17.8) < 0.001 ***
RCC_TSC
LIRC (N=81) reference
HIRC (N=93) 4.45 (1.55 − 12.6) 0.0054 **
MIRC (N=110) 1.84 (0.61 − 5.6) 0.2802
Events: 34; Global p−value (Log−Rank): 4.0885e−07
AIC: 346.28; Concordance Index: 0.81
K
1−year−OS
2−year−OS
3−year−OS
4−year−OS
5−year−OS
Observed OS (%)
Nomogram−OS (%)
Sensitivity
1−Specificity
AUC at 1 year: 0.962
AUC at 2 years: 0.891
AUC at 3 years: 0.886
AUC at 4 years: 0.979
AUC at 5 years: 0.968
L
1−year−PFS
2−year−PFS
3−year−PFS
4−year−PFS
5−year−PFS
Observed PFS (%)
Nomogram−PFS (%)
Sensitivity
1−Specificity
AUC at 1 year: 0.915
AUC at 2 years: 0.887
AUC at 3 years: 0.838
AUC at 4 years: 0.834
AUC at 5 years: 0.805

Fig. 6 Statistical analyses validated the prognostic relevance of RCC-TSC in Xinhua cohort. **A** The K-M survival plots validated patients in the HIRC cluster had the lowest OS ($p=0.003$) and PFS ($p<0.001$) in KIRC patients. **B** The Chi-square test revealed that RCC-TSC were significantly associated with gender, histopathological classifications, PFS censor, PFS classification, OS censor and OS classification. **C** The pie chart displayed that the proportion of male was highest in LIRC (78%), while lowest in MIRC (63%). **D** The pie chart showed that percentage of KIRC was highest in MIRC (93%), while lowest in LIRC (82%). All the classification labeled as others were in LIRC. **E** The pie chart demonstrated that HIRC conferred higher probability of death (13%), while MIRC and LIRC were 8% and 1% respectively. **F** The pie chart suggested that HIRC conferred higher probability of PFS censor (23%), while MIRC and LIRC were 11% and 7% respectively. **G** The pie chart displayed that the patients of HIRC had higher proportion of low OS (74%), greatly surpassing MIRC (47%) and LIRC (9%). **H** The pie chart showed that patients of HIRC had higher proportion of low PFS (77%), higher than MIRC (51%) and LIRC (10%). **I** Multivariate proportional hazards Cox regression analysis validated RCC-TSC as an independent factor and HIRC was a risk factor for OS (HR = 16.68, 95% CI = 1.88–148.1, $p=0.0115$). **J** Multivariate proportional hazards Cox regression analysis validated HIRC as an independent factor for PFS (HR = 4.43, 95% CI = 1.55–12.6, $p=0.0054$). **K** The calibration curve demonstrated high concordance and high AUC values in 1 (0.962), 2 (0.891), 3 (0.886), 4 (0.979), 5 (0.968) years for OS. **L** The calibration curve also demonstrated high concordance and high AUC values in 1 (0.915), 2 (0.887), 3 (0.838), 4 (0.834), 5 (0.805) years for PFS. RCC-TSC, renal cell carcinoma-Treg states-related prognostic classification; K-M, Kaplan-Meier; HIRC, high immune infiltration RCC; OS, overall survival; PFS, progression-free survival; KIRC, kidney renal clear cell carcinoma; LIRC, low immune infiltration RCC; MIRC, moderate immune infiltration RCC; HR, hazard ratio; CI, confidential interval; AUC, area under curve

subtypes with different prognosis, validating the clinical significance of the model in real-world patients.

4.1 TSRPGs determined states of Tregs and heterogeneity of RCC-TSC

The TSRPGs were classified into four states, with differential average expression in three subtypes of RCC-TSC. Except RGS2, CXCL13, and KLF6, most of the TSRPGs belonged to state 1 and state 2.

All the TSRPGs of state 1 were in the most beginning of pseudotime trajectory. Moreover, they all served as risk factors and had highest expression in HIRC. The majority of TSRPGs in state 1 were reported to play a critical role in cancers including modulating the cell cycle, cell proliferation and chromosome stability, regulating cell death programs such as apoptosis and ubiquitination [59–67]. Among them, high expression of the proliferation marker ki-67 (Mki67) of HIRC in our IHC staining also significantly indicated its strong proliferative ability. Proliferative Tregs have been reported to exhibit highly activated phenotype, enabling clonal expansion and accumulation in tumors, where they exert strong suppressive effects [17, 18]. Further, abundant proliferative Tregs were a positive indicator of tumor progression, and could partially explain the poor prognosis of HIRC [19–21]. Besides the proliferation-related functions, HGMB2 was also upregulated in exhausted CD8 + T cells and was critical for their differentiation, which could be a potential marker and target for T cell-based immunotherapies [68]. RRM2 facilitates immune escape, and higher expression of RRM2 has also been associated with better ICB therapeutic effects in RCC [69].

In state 2, the intermediate temporal state, immune-related genes such as LAG3, CD7, and SOCS1 were more immune related and also had highest expression in HIRC. Lymphocyte activation gene 3 (LAG3) could induce T cell exhaustion through various mechanisms. LAG3 on the CD4+ and CD8 + T cells could directly bind ligand receptors [70]. More importantly, LAG3 is also expressed on Tregs, and has been shown to mediate inhibitory cytokines released by Tregs, thereby accelerating T cell dysfunction in TIME [71–73]. High expression of LAG3 is not only correlated with better ICB therapeutic effects and co-expressed with immunosuppressive markers such as CTLA-4, PD-1 and TIGIT [74–77], but also serves as a potential target for RCC itself [73]. Relatlimab, an antibody that could reverse LAG3-mediated inhibition, has been reported to have potent anti-tumor effect for melanoma in combination with nivolumab [78, 79]. Besides, for advanced melanoma patients who are not responsive to PD-1/PD-L1 blockade, therapy that combined anti-LAG-3 with anti-PD-1 has also showed significant benefits for them [80]. In our IHC staining, the high expression of LAG3 in HIRC further indicated the significant immune exhausted states in HIRC. Consequently, how to take advantage of LAG3 as a feasible target for immunotherapy in RCC patients remains to be studied, which might benefit HIRC patients a lot. Cluster of differentiation 7 (CD7), a transmembrane glycoprotein accountable for co-stimulation of T cells, has higher expression levels associated with aggressive biological functions, drug resistance and poor prognosis [81, 82]. Moreover, CD7 is an attractive therapeutic target for T cell malignancies (like T cell lymphoma and T-cell acute lymphoblastic leukemia). It might also be a potential target for the T cells of RCC, by using CAR-T cells targeting the molecules negatively regulate immune functions. Suppressor of cytokine signaling 1 (SOCS1), though acting as a tumor suppressor across various malignancies, was found to be a potent inhibitor of inflammation through controlling cytokines associated with JAK-STAT1/3 pathways in a negative-feedback approach [83–85]. SOCS1 could thus attenuate the magnitude of antitumor immunity in TIME of RCC, which also had therapeutic potentials.

To summarize, the biological functions of the TSRPGs, and their differential expression levels, together with the interplays between them, collectively shaped the unique molecular characteristics of three subtypes of RCC-TSC. HIRC, MIRC, LIRC was respectively characterized by high, medium, low infiltration of proliferative and suppressive Tregs and immune cells in exhausted states.

4.2 Prognostic, therapeutic and mechanistic implications of RCC-TSC

RCC-TSC, established by high-throughput bioinformatic analysis and validated through a real-world cohort (Xinhua cohort), has several significant clinical implications.

Firstly, RCC-TSC successfully stratified the RCC patients into distinct prognostic groups. It validated that through RNA sequencing or IHC staining, we could classify RCC patients according to the expression level of TSRPGs, and employ personalized therapeutic interventions. For the four key markers (DNAJA1, DNAJB4, LAG3, Mki67) that we used in our study, they have been discovered to be viable markers across several malignancies. DNAJA1 (known as HDJ-2), would promote the progression of pancreatic and colorectal cancer [86, 87], and has been reported to be a reliable marker in acute lymphocytic leukemia and pancreatic cancer [88–90]. Talking of DNAJB4 (known as HLJ1), it has also been revealed to be a tumor suppressor gene in lung, colon, and gastric cancers, a feasible biomarker, and a potential therapeutic target for lung cancer [91–95]. Considering LAG3, higher expression of LAG3 was linked to worse prognosis and more suppressive TIME in various solid tumors [96]. Mki67 was not only critical in evaluating the aggressiveness of cancers, but also validated to play a critical role in classifying cancers into different subtypes [97], including breast cancer [98]. In our research, we used IHC staining and successfully distinguished patients with differential expression of the four key genes. It validated that we could design a diagnostic kit in future, to handily apply RCC-TSC to classify RCC patients with different prognosis and to tailor personalized therapy. In the future, its efficacy in classifying patients with kidney neoplasms, besides KIRC, remains to be further validated.

Secondly, HIRC, MIRC, LIRC were characterized by different levels of immune infiltration and different states of immune cells, which had great significance for guidance of treatment. To be specific, HIRC had the highest immune infiltration but with proliferative and suppressive Tregs. Moreover, the immune exhausted scores were highest for HIRC, with higher expression of CTLA4, PD-L1, and PD-1. These results suggest that patients of HIRC would be more likely to benefit from ICB therapies, since ICIs target molecules that negatively regulate immune function (CTLA4 and PD1), and the ability of RCC TIME to be infiltrated by suppressive T cells was also a critical contributor to the immunotherapy clinical response [8]. Additionally, we have predicted the possible inhibitors for each subtype, and sunitinib benefited the HIRC most. Consequently, with the unique properties of HIRC, ICB therapy could be prioritized and its indication might even be expanded to patients who are not in the advanced or metastatic stage. This hypothesis warrants to be verified with future clinical studies.

Thirdly, the metastasis-related mechanisms of the three subtypes were diverse, which could be the most significant underlying mechanisms that promote their progression. In HIRC, the BATF had high correlation with LAG3, and downstream CD8 + T cells, pDCs, T-cell co-inhibition, and Tfh. It indicated that BATF and LAG3 might regulate the immune components [99], while T cell and immune exhaustion was possibly the most critical mechanism in progression and metastasis of HIRC. Considering MIRC, elevated CENPA [100], NCAPG [101], MYBL2 [102] and LMNB1 [103], PTTG1 [59] have been reported to influence cell cycle and apoptosis to promote the proliferation of cancer cells, which was associated with progression and metastasis. Moreover, cyclin B1 was the most correlated downstream molecule, suggesting dysregulated cell cycle might be the most critical mechanism of MIRC. For LIRC, IRF4 showed the strongest correlation with upstream MEF2B, as well as downstream CXCL13, B cells memory, and B cells. Furthermore, IRF4 and MEF2B were reported to have close association with B cell lymphoma [104–107], suggesting B cell-related mechanisms might play crucial roles in LIRC.

Nonetheless, this study has certain limitations. Firstly, due to the inherent biases in bioinformatics analyses, our investigation can't completely simulate the intricacies of the TIME and the developmental trajectories of RCC Tregs. Secondly, though our model captures the broader role of TSRPGs in the tumor microenvironment, we acknowledge that relying on overall tissue expression rather than Treg-specific expression may introduce a degree of bias. Moreover, the use of only DNAJA1 and DNAJB4 as markers for LIRC and LAG3 and MKi67 for HIRC may introduce potential bias in the classification of subtypes. Also, though we have completed a comprehensive study encompassing bioinformatic analysis, retrospective clinical and pathological study with a large cohort, the effectiveness of RCC-TSC warrants further prospective and multi-center clinical trials. Finally, it is imperative to conduct more cell and animal mechanistic experiments to validate the regulatory networks and the potential inhibitors for each RCC subtype.

5 Conclusion

Through integrated bioinformatics analysis and a large-sample retrospective study, we established RCC-TSC, which classified RCC (especially KIRC) patients into HIRC (high risk, LAG3+, Mki67+), MIRC (intermediate risk), and LIRC (low risk, DNAJA1+, DNAJB4+). Moreover, HIRC was validated as an independent prognostic factor for OS (HR = 16.68, 95% CI = 1.88–148.1, $p = 0.011$) and PFS (HR = 4.43, 95% CI = 1.55–12.6, $p = 0.005$), and the AUC of our model demonstrated perfect performances in 1, 2, 3, 4, 5 years for both OS (0.962, 0.891, 0.886, 0.979, 0.96) and PFS

(0.915, 0.887, 0.838, 0.834, 0.805). Besides, we identified that HIRC was highly infiltrated by exhausted immune cells, which possibly promoted its progression and metastasis, also resulting its high sensitivity to ICB therapies. Hopefully, with our diagnostic kit, RCC-TSC could effectively help clinicians to stratify patients with different prognosis, and could provide insights into guiding personalized treatment.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13402-024-01030-9.

Acknowledgements We thank GEO database, TCGA database, Cistrome cancer database, MET500 database, CCLE database, GDSC database, MSigDB database for allowing us to use their data.

Author contributions Conception/design: Y. Y, Y. L, B. L, R. H, W. Z, G. X, X. P, X. C Collection and/or assembly of data: K. D, X. P Experiment: G. J Data analysis: Y. Y, Y. L, B. L, G. J, Z. Z, R. H Manuscript writing: Y. Y, Y. L, B. L, G. J, L. W, K. D, Z. Z, D. L, M. W, S. T, X. L, Y. L, R. H, W. Z, G. X, X. P, X. C Manuscript revision: Y. Y, Y. L, B. L, R. H, W. Z, G. X, X. P, X. C Final approval of manuscript: Y. Y, Y. L, B. L, G. J, L. W, K. D, Z. Z, D. L, M. W, S. T, X. L, Y. L, R. H, W. Z, G.X, X. P, X. C.

Funding This work was sponsored by the National Natural Science Foundation of China (No. 82330094, 82173265, 82072806); Leading health talents of Shanghai Municipal Health Commission (2022LJ002); Shanghai Rising-Star Program (23QC1401400); Shanghai Rising-Star Program (Sailing Special Program) (23YF1458400); the Natural Science Foundation of Shanghai (23ZR1441300); Hospital Funded Clinical Research, Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine (21XHDB06); Bole Project of Shanghai Jiaotong University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability Please contact the corresponding author when necessary.

Declarations

Ethical approval The study was approved by the Ethics Committee of Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine (XHEC-C-2021-145-1). All authors confirmed that all methods were carried out in accordance with relevant guidelines and regulations.

Data access statement The datasets generated and/or analyzed during the current study are available on the Gene Expression Omnibus (GEO) database, The Cancer Genome Atlas (TCGA) database, Cistrome cancer database, MET500 database, Cancer Cell Line Encyclopedia (CCLE) database, Genomics of Drug Sensitivity in Cancer (GDSC) database and the Molecular Signatures Database (MSigDB, version 7.4). Due to the privacy of the patients in our center, please contact the corresponding authors to obtain the information of patients.

Patient and public involvement We included a retrospective cohort, which contained 370 patients that were diagnosed with kidney neoplasm and received treatment in our hospital from between 2016 and 2018. Before we used their clinical data for research, we have explained our study procedure and obtained informed consent from the participants, or from a family member of them when the participants had difficulty communicating or understanding. The study was approved by the Ethics Committee of Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine (XHEC-C-2021-145-1) in Dec 24th 2021.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Y.F. Wang et al., Evolution of cell therapy for renal cell carcinoma. Mol. Cancer. **23**(1), (2024)
2. J.J. Hsieh et al., Renal cell carcinoma. Nat. Reviews Disease Primers. **3**, (2017)
3. U. Capitanio et al., Epidemiology of renal cell carcinoma. Eur. Urol. **75**(1), 74–84 (2019)
4. H.T. Cohen, F.J. McGovern, Renal-cell carcinoma. N. Engl. J. Med. **353**(23), 2477–2490 (2005)
5. M. Binnewies et al., Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat. Med. **24**(5), 541–550 (2018)
6. T. Fu et al., Spatial architecture of the immune microenvironment orchestrates tumor immunity and therapeutic response. J. Hematol. Oncol. **14**(1), (2021)
7. S. Mariathasan et al., TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature. **554**(7693), 544– (2018)
8. C.M. Díaz-Montero, B.I. Rini, J.H. Finke, The immunology of renal cell carcinoma. Nat. Rev. Nephrol. **16**(12), 721–735 (2020)
9. D.A. Braun et al., Progressive immune dysfunction with advancing disease stage in renal cell carcinoma. Cancer Cell. **39**(5), 632– (2021)
10. L.T. Zheng et al., Pan-cancer single cell landscape of tumor-infiltrating T cells. Science. **374**(6574), 1462– (2021)
11. A. Chow et al., Clinical implications of T cell exhaustion for cancer immunotherapy. Nat. Reviews Clin. Oncol. **19**(12), 775–790 (2022)
12. Y. Togashi, K. Shitara, H. Nishikawa, Regulatory T cells in cancer immunosuppression - implications for anticancer therapy. Nat. Reviews Clin. Oncol. **16**(6), 356–371 (2019)
13. C. Tay, A. Tanaka, S. Sakaguchi, Tumor-infiltrating regulatory T cells as targets of cancer immunotherapy. Cancer Cell. **41**(3), 450–465 (2023)

14. J.D. Fontenot, M.A. Gavin, A.Y. Rudensky, Foxp3 programs the development and function of $CD4^{+}CD25^{+}$ regulatory T cells. J. Immunol. **198**(3), 986–992 (2017)
15. S. Hori, T. Nomura, S. Sakaguchi, Control of regulatory T cell development by the transcription factor Foxp3. Science. **299**(5609), 1057–1061 (2003)
16. A. Ribas, J.D. Wolchok, Cancer immunotherapy using checkpoint blockade. Science. **359**(6382), 1350– (2018)
17. L. Wyss et al., Affinity for self antigen selects T subreg cells with distinct functional properties. Nat. Immunol. **17**(9), 1093–1101 (2016)
18. M. Miyara et al., Functional delineation and differentiation dynamics of Human CD4+T cells expressing the FoxP3 transcription factor. Immunity. **30**(6), 899–911 (2009)
19. D. Ha et al., Differential control of human Treg and effector T cells in tumor immunity by Fc-engineered anti-CTLA-4 antibody. Proc. Natl. Acad. Sci. U.S.A. **116**(2), 609–618 (2019)
20. T. Kamada et al., PD-1+regulatory T cells amplified by PD-1 blockade promote hyperprogression of cancer. Proc. Natl. Acad. Sci. U.S.A. **116**(20), 9999–10008 (2019)
21. Y. Ohue, H. Nishikawa, Regulatory T (Treg) cells in cancer: can Treg cells be a new therapeutic target? Cancer Sci. **110**(7), 2080–2089 (2019)
22. T. Barrett et al., NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. **41**(D1), D991–D995 (2012)
23. K. Tomczak, P. Czerwińska, M. Wiznerowicz, Review the Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp. Oncology/Współczesna Onkologia. **2015**(1), 68–77 (2015)
24. R.B. Zheng et al., Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. Nucleic Acids Res. **47**(D1), D729–D735 (2019)
25. D.R. Robinson et al., Integrative clinical genomics of metastatic cancer. Nature. **548**(7667), 297– (2017)
26. D.P. Nusinow et al., Quant. Proteom. Cancer Cell. Line Encyclopedia Cell. **180**(2), 387– (2020)
27. W. Yang et al., Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. **41**(D1), D955–D961 (2012)
28. A. Liberzon, A description of the molecular signatures database (MSigDB) Web Site, in *Stem Cell Transcriptional Networks: Methods and Protocols*, ed. by B.L. Kidder (2014), pp. 153–160
29. A. Butler et al., Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. **36**(5), 411– (2018)
30. T. Stuart et al., Comprehensive Integration of Single-Cell Data. Cell. **177**(7), 1888– (2019)
31. X.X. Zhang et al., CellMarker: a manually curated resource of cell markers in human and mouse. Nucleic Acids Res. **47**(D1), D721–D728 (2019)
32. S.J. Cheng et al., A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. Cell. **184**(3), 792– (2021)
33. A. Liberzon et al., The Molecular signatures Database Hallmark Gene Set Collection. Cell. Syst. **1**(6), 417–425 (2015)
34. Y. Wang et al., *iTALK: an R Package to Characterize and Illustrate Intercellular Communication.* bioRxiv (2019), p. 507871
35. X.J. Qiu et al., Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods. **14**(10), 979– (2017)
36. S. Slovin et al., Single-Cell RNA Sequencing Analysis: A Step-by-Step Overview, in RNA Bioinformatics, 2 Edition, ed. E. Picardi, pp. 343–365 (2021)
37. J.T. Rich et al., A practical guide to understanding Kaplan-Meier curves. Otolaryngol. Head Neck Surg. **143**(3), 331–336 (2010)
38. M.D. Wilkerson, D.N. Hayes, ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. **26**(12), 1572–1573 (2010)
39. P. Jiang et al., Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat. Med. **24**(10), 1550 (2018)
40. M. Yi et al., ssGSEA score-based Ras dependency indexes derived from gene expression data reveal potential Ras addiction mechanisms with possible clinical implications. Sci. Rep. **10**(1), (2020)
41. M.E. Ritchie et al., Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. **43**(7), (2015)
42. H. Varet et al., SARTools: a DESeq2-and EdgeR-Based R Pipeline for Comprehensive Differential Analysis of RNA-Seq Data. Plos One. **11**(6) (2016)
43. P. Geeleher, N.J. Cox, R.S. Huang, Clinical drug response can be predicted using baseline gene expression levels and < i > in vitro drug sensitivity in cell lines. Genome Biol. **15**(3) (2014)
44. P. Geeleher, N. Cox, R.S. Huang, pRRophetic: an R Package for Prediction of Clinical Chemotherapeutic Response from Tumor Gene expression levels. Plos One. **9**(9) (2014)
45. G.M. Morris et al., AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J. Comput. Chem. **30**(16), 2785–2791 (2009)
46. O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. **31**(2), 455–461 (2010)
47. M.R. Corces et al., The chromatin accessibility landscape of primary human cancers. Science. **362**(6413), 420– (2018)
48. J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized Linear models via Coordinate Descent. J. Stat. Softw. **33**(1), 1–22 (2010)
49. M. Uhlen et al., Tissue-based map of the human proteome. Science. **347**(6220) (2015)
50. S.R. Woo, L. Corrales, T.F. Gajewski, The STING pathway and the T cell-inflamed tumor microenvironment. Trends Immunol. **36**(4), 250–256 (2015)
51. M. Philip et al., Chromatin states define tumour-specific T cell dysfunction and reprogramming. Nature. **545**(7655), 452– (2017)
52. A. Schietinger et al., Tumor-specific T cell dysfunction is a dynamic Antigen-Driven differentiation program initiated early during Tumorigenesis. Immunity. **45**(2), 389–401 (2016)
53. H. Shao et al., Asymmetric synthesis of Cyclopamine, a hedgehog (hh) signaling pathway inhibitor. J. Am. Chem. Soc. **145**(46), 25086–25092 (2023)
54. P. Heretsch, L. Tzagkaroulaki, A. Giannis, Cyclopamine and hedgehog signaling: Chemistry, Biology, Medical perspectives. Angewandte Chemie-International Ed. **49**(20), 3418–3427 (2010)
55. V. Thorsson et al., The Immune Landscape of Cancer. Immunity. **48**(4), 812– (2018)
56. E. Jonasch, J.J. Gao, W.K. Rathmell, Renal cell carcinoma. Bmj-British Med. J. **349** (2014)
57. T.F. Gajewski, H. Schreiber, Y.X. Fu, Innate and adaptive immune cells in the tumor microenvironment. Nat. Immunol. **14**(10), 1014–1022 (2013)
58. J.A. Joyce, D.T. Fearon, T cell exclusion, immune privilege, and the tumor microenvironment. Science. **348**(6230), 74–80 (2015)
59. X.W. Liu et al., PCLAF Promotes Neuroblastoma G1/S cell Cycle Progression via the E2F1/PTTG1 axis. Cell Death & Dis. **132**(2) (2002)
60. Z.W. Zuo et al., Ribonucleotide reductase M2 (RRM2): regulation, function and targeting strategy in human cancer. Genes Dis. **11**(1), 218–233 (2024)
61. S.Z. Zhang et al., The UBE2C/CDH1/DEPTOR axis is an oncogene and tumor suppressor cascade in lung cancer cells. J. Clin. Invest. **133**(4) (2023)

62. M. Malumbres, M. Barbacid, Cell cycle, CDKs and cancer: a changing paradigm. Nat. Rev. Cancer. **9**(3), 153–166 (2009)
63. P. Jain et al., PHF19 mediated regulation of proliferation and invasiveness in prostate cancer cells. Elife. **9** (2020)
64. L.K. Esser et al., Overexpression of Parkin in clear cell renal cell carcinoma decreases tumor aggressiveness by regulating CKS2 levels. Int. J. Oncol. **60**(2) (2022)
65. S. Hindriksen, A. Meppelink, S.M.A. Lens, Functionality of the chromosomal passenger complex in cancer. Biochem. Soc. Trans. **43**, 23–32 (2015)
66. L. Wang et al., High mobility Group A1 (HMGA1): structure, biological function, and therapeutic potential. Int. J. Biol. Sci. **18**(11), 4414–4431 (2022)
67. B.W. Xie et al., Cyclin B1/CDK1-regulated mitochondrial bioenergetics in cell cycle progression and tumor resistance. Cancer Lett. **443**, 56–66 (2019)
68. E.N. Neubert et al., HMGB2 regulates the differentiation and stemness of exhausted $CD8^+$ T cells during chronic viral infection and cancer. Nat. Commun. **14**(1) (2023)
69. W. Xiong et al., RRM2 regulates sensitivity to Sunitinib and PD-1 blockade in renal cancer by stabilizing ANXA1 and activating the AKT pathway. Adv. Sci. **8**(18) (2021)
70. C. Guy et al., LAG3 associates with TCR-CD3 complexes and suppresses signaling by driving co-receptor-lck dissociation. Nat. Immunol. **23**(5), 757– (2022)
71. T. Maruhashi et al., Binding of LAG-3 to stable peptide-MHC class II limits T cell function and suppresses autoimmunity and anti-cancer immunity. Immunity. **55**(5), 912– (2022)
72. C.T. Huang et al., Role of LAG-3 in regulatory T cells. Immunity. **21**(4), 503–513 (2004)
73. Y. Liu et al., Lymphocyte activation gene 3 served as a potential prognostic and immunological biomarker across various cancer types: a clinical and pan-cancer analysis. Clin. Transl Immunol. **13**(10), e70009 (2024)
74. Z. Gong, J. Zhang, W. Guo, Tumor purity as a prognosis and immunotherapy relevant feature in gastric cancer. Cancer Med. **9**(23), 9052–9063 (2020)
75. Y. Mao et al., Low tumor purity is associated with poor prognosis, heavy mutation burden, and intense immune phenotype in colon cancer. Cancer Manag Res. **10**, 3569–3577 (2018)
76. X. Qi et al., Immune-Stromal score signature: Novel Prognostic Tool of the Tumor Microenvironment in Lung Adenocarcinoma. Front. Oncol. **10**, 541330 (2020)
77. Y. Deng et al., Tumor purity as a prognosis and immunotherapy relevant feature in cervical cancer. Aging (Albany NY). **13**(22), 24768–24785 (2021)
78. H.A. Tawbi et al., Relatlimab and Nivolumab versus Nivolumab in Untreated Advanced Melanoma. N Engl. J. Med. **386**(1), 24–34 (2022)
79. K. Thudium et al., Preclinical characterization of Relatlimab, a human LAG-3-Blocking antibody, alone or in combination with Nivolumab. Cancer Immunol. Res. **10**(10), 1175–1189 (2022)
80. P.A. Ascierto et al., Nivolumab and Relatlimab in patients with Advanced Melanoma that had progressed on Anti-programmed Death-1/Programmed death Ligand 1 therapy: results from the phase I/IIa RELATIVITY-020 trial. J. Clin. Oncol. **41**(15), 2724–2735 (2023)
81. C.E. Brown, C.L. Mackall, CAR T cell therapy: inroads to response and resistance. Nat. Rev. Immunol. **19**(2), 73–74 (2019)
82. M. MacKay et al., The therapeutic landscape for cells engineered with chimeric antigen receptors. Nat. Biotechnol. **38**(2), 233– (2020)
83. J. Zhang et al., Role of SOCS1 in tumor progression and therapeutic application. Int. J. Cancer. **130**(9), 1971–1980 (2012)
84. S. Ilangumaran, S. Ramanathan, R. Rottapel, Regulation of the immune system by SOCS family adaptor proteins. Semin. Immunol. **16**(6), 351–365 (2004)
85. C. Beaurivage et al., SOCS1 in cancer: an oncogene and a tumor suppressor. Cytokine. **82**, 87–94 (2016)
86. D.D. Xu et al., Inhibition of mutant Kras and p53-driven pancreatic carcinogenesis by atorvastatin: mainly via targeting of the farnesylated DNAJA1 in chaperoning mutant p53. Mol. Carcinog. **58**(11), 2052–2064 (2019)
87. S.S. Yang et al., KNK437 restricts the growth and metastasis of colorectal cancer via targeting DNAJA1/CDC45 axis. Oncogene. **39**(2), 249–261 (2020)
88. J.S. Chen et al., Identification of novel markers for monitoring minimal residual disease in acute lymphoblastic leukemia. Blood. **97**(7), 2115–2120 (2001)
89. A. Patnaik et al., A phase I, pharmacokinetic, and biological study of the farnesyltransferase inhibitor tipifarnib in combination with gemcitabine in patients with advanced malignancies. Clin. Cancer Res. **9**(13), 4761–4771 (2003)
90. L.Q.M. Chow et al., A phase I safety, pharmacological, and biological study of the farnesyl protein transferase inhibitor, lonafarnib (SCH 663366), in combination with cisplatin and gemcitabine in patients with advanced solid tumors. Cancer Chemother. Pharmacol. **62**(4), 631–646 (2008)
91. Y. Liu et al., HLJ1 is a novel biomarker for colorectal carcinoma progression and overall patient survival. Int. J. Clin. Exp. Pathol. **7**(3), 969–977 (2014)
92. J. Simoes-Correia et al., DNAJB4 molecular chaperone distinguishes WT from mutant E-cadherin, determining their fate in vitro and in vivo. Human Mol. Genet. **23**(8), 2094–2105 (2014)
93. M.F. Tsai et al., A new tumor suppressor DnaJ-like heat shock protein, HLJ1, and survival of patients with non-small-cell lung carcinoma. Jnci-Journal Natl. Cancer Inst. **98**(12), 825–838 (2006)
94. C.C. Wang et al., The transcriptional factor YY1 upregulates the novel invasion suppressor HLJ1 expression and inhibits cancer cell invasion. Oncogene. **24**(25), 4081–4093 (2005)
95. H.W. Chen et al., Curcumin inhibits lung cancer cell invasion and metastasis through the tumor suppressor HLJ1. Cancer Res. **68**(18), 7428–7438 (2008)
96. R.Y. Li et al., Prognostic significance of lymphocyte-activation gene 3 (LAG3) in patients with solid tumors: a systematic review, meta-analysis and pan-cancer analysis. Cancer Cell Int. **23**(1) (2023)
97. C.B. Li et al., Single-cell transcriptomics reveals cellular heterogeneity and molecular stratification of cervical cancer. Commun. Biol. **5**(1) (2022)
98. S. Loibl et al., Breast cancer. Lancet. **397**(10286), 1750–1769 (2021)
99. T.L. Murphy, R. Tussiwand, K.M. Murphy, Specificity through cooperation: BATF-IRF interactions control immune-regulatory networks. Nat. Rev. Immunol. **13**(7), 499–509 (2013)
100. S. Giunta et al., CENP-A chromatin prevents replication stress at centromeres to avoid structural aneuploidy. Proc. Natl. Acad. Sci. U.S.A. **118**(10) (2021)
101. H.B. Li et al., NCAPG promotes the proliferation of renal clear cell carcinoma via mediating with CDK1. Dis. Markers. **2022** (2022)
102. R.J. Yu et al., Clinicopathologic features and prognostic implications of MYBL2 protein expression in pancreatic ductal adenocarcinoma. Pathol. Res. Pract. **213**(8), 964–968 (2017)
103. D. Tang et al., Silencing LMNB1 contributes to the suppression of lung Adenocarcinoma Development. Cancer Manage. Res. **13**, 2633–2642 (2021)

104. J.E. Ramis-Zaldivar et al., Distinct molecular profile of IRF4-rearranged large B-cell lymphoma. Blood. **135**(4), 274–286 (2020)
105. N. Schleussner et al., Transcriptional reprogramming by mutated IRF4 in lymphoma. Nat. Commun. **14**(1) (2023)
106. R.D. Morin et al., Frequent mutation of histone-modifying genes in non-hodgkin lymphoma. Nature. **476**(7360), 298–303 (2011)
107. P. Brescia et al., MEF2B instructs Germinal Center Development and acts as an Oncogene in B Cell Lymphomagenesis. Cancer Cell. **34**(3), 453– (2018)

Authors and Affiliations

Yuntao Yao[1] · Yifan Liu[1] · Bingnan Lu[1] · Guo Ji[2] · Lei Wang[1] · Keqin Dong[1] · Zihui Zhao[1] · Donghao Lyu[1] · Maodong Wei[1] · Siqi Tu[1] · Xukun Lyu[1] · Yuanan Li[1] · Runzhi Huang[3] · Wang Zhou[1] · Guofeng Xu[1] · Xiuwu Pan[1] · Xingang Cui[1]

✉ Runzhi Huang
runzhihuang2022@163.com

✉ Wang Zhou
brilliant212@163.com

✉ Guofeng Xu
xuguofeng@xinhuamed.com.cn

✉ Xiuwu Pan
panxiuwu@126.com

✉ Xingang Cui
cuixingang@xinhuamed.com.cn

[1] Department of Urology, Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai 200092, China

[2] Department of Pathology, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai, China

[3] Department of Burn Surgery, The First Affiliated Hospital of Naval Medical University, Shanghai 200433, China