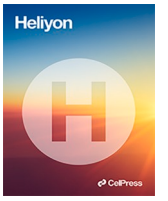
Heliyon 10 (2024) e31816

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Identification of clear cell renal cell carcinoma subtypes by integrating radiomics and transcriptomics

Ruizhi Gao [a,1], Jinshu Pang [a,1], Peng Lin [b], Rong Wen [a], Dongyue Wen [a], Yiqiong Liang [c], Zhen Ma [d], Li Liang [e], Yun He [a], Hong Yang [a,*]

[a] *Department of Medical Ultrasound, The First Affiliated Hospital of Guangxi Medical University, 6 Shuangyong Road, Nanning, Guangxi Zhuang Autonomous Region, PR China*
[b] *Department of Medical Ultrasound, Fujian Medical University Union Hospital, Fuzhou, Fujian Province, PR China*
[c] *Department of Radiology, The International Zhuang Medical Affiliated Hospital of Guangxi University of Chinese Medicine, Nanning, Guangxi Zhuang Autonomous Region, PR China*
[d] *Department of Medical Ultrasound, The International Zhuang Medical Affiliated Hospital of Guangxi University of Chinese Medicine, Nanning, Guangxi Zhuang Autonomous Region, PR China*
[e] *Department of Medical Ultrasound, Liuzhou People's Hospital, No. 8 Wenchang Road, Liuzhou, Guangxi Zhuang Autonomous Region, PR China*

A R T I C L E I N F O

Keywords:
Clear cell renal cell carcinoma
Transcriptomics
Radiomics
Computed tomography
Unsupervised clustering

A B S T R A C T

Objective: This study aimed to delineate the clear cell renal cell carcinoma (ccRCC) intrinsic subtypes through unsupervised clustering of radiomics and transcriptomics data and to evaluate their associations with clinicopathological features, prognosis, and molecular characteristics.
Methods: Using a retrospective dual-center approach, we gathered transcriptomic and clinical data from ccRCC patients registered in The Cancer Genome Atlas and contrast-enhanced computed tomography images from The Cancer Imaging Archive and local databases. Following the segmentation of images, radiomics feature extraction, and feature preprocessing, we performed unsupervised clustering based on the "CancerSubtypes" package to identify distinct radio-transcriptomic subtypes, which were then correlated with clinical-pathological, prognostic, immune, and molecular characteristics.
Results: Clustering identified three subtypes, C1, C2, and C3, each of which displayed unique clinicopathological, prognostic, immune, and molecular distinctions. Notably, subtypes C1 and C3 were associated with poorer survival outcomes than subtype C2. Pathway analysis highlighted immune pathway activation in C1 and metabolic pathway prominence in C2. Gene mutation analysis identified VHL and PBRM1 as the most commonly mutated genes, with more mutated genes observed in the C3 subtype. Despite similar tumor mutation burdens, microsatellite instability, and RNA interference across subtypes, C1 and C3 demonstrated greater tumor immune dysfunction and rejection. In the validation cohort, the various subtypes showed comparable results in terms of clinicopathological features and prognosis to those observed in the training cohort, thus confirming the efficacy of our algorithm.

* Corresponding author. Department of Medical Ultrasound, The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region, PR China.
E-mail address: yanghong@gxmu.edu.cn (H. Yang).
[1] Contributed equally as first-authors.

https://doi.org/10.1016/j.heliyon.2024.e31816
Received 9 April 2024; Received in revised form 21 May 2024; Accepted 22 May 2024
Available online 23 May 2024

Conclusion: Unsupervised clustering based on radiotranscriptomics can identify the intrinsic subtypes of ccRCC, and radiotranscriptomic subtypes can characterize the prognosis and molecular features of tumors, enabling noninvasive tumor risk stratification.

1. Introduction

Clear cell renal cell carcinoma (ccRCC) accounts for >80 % of all renal cancer cases, making it the most prevalent malignant kidney tumor [1]. In its early stages, ccRCC often exhibits minimal or no discernible symptoms, and there is currently no widely accepted serum biomarker for the early screening of ccRCC [2]. An estimated 25 % of patients are diagnosed when the tumor has already progressed to an advanced stage, necessitating comprehensive treatment approaches involving a combination of surgery and targeted therapies or immunotherapy [3,4]. At present, therapeutic outcome prediction predominantly relies on percutaneous biopsy, which carries the risk of complications, including bleeding, infection, and needle tract metastasis. However, there is substantial heterogeneity in the individual prognosis of ccRCC patients, and the small tissue samples obtained from biopsies may not accurately reflect the overall tumor heterogeneity [5,6]. Conventional noninvasive imaging techniques, like ultrasound, CT, and MRI, lack quantitative standards and face limitations in distinguishing between different pathologies, assessing preoperative characteristics, and identifying prognostic markers in ccRCC [7]. Meanwhile, most radiomics studies primarily focus on the quantitative analysis of image features within a computer semantics framework, often falling short in providing detailed molecular insights, thus limiting their clinical applicability [8]. Therefore, molecular markers lack noninvasive detection, whereas radiomics features lack molecular information and pathological confirmation. This is where the emerging field of radiotranscriptomics has come to play.

Radiotranscriptomics, an emerging discipline in recent years, integrates radiomics and genomics to provide a noninvasive and relatively efficient prediction of tumor molecules and the tumor microenvironment [9]. Radiotranscriptomics has demonstrated significant potential in various aspects of cancer research, including gene mutation, subtype differentiation, metastasis prediction, and prognosis assessment [10,11]. In our previous study, we identified three subtypes (proliferation, metabolism, and immune activation) by integrating radiomics and transcriptomics analyses, which confirmed that the radiotranscriptomics strategy helped stratify non-small cell lung cancer patients and provided radiomics features with molecular explanations and prognostic prediction [12]. Another study also suggested that SUVmax features could correlate with tumor mutational burden (TMB), a prognostic biomarker associated with improved outcomes following immunotherapy. However, assessing TMB typically requires a tissue biopsy [13]. Therefore, a systematic exploration of the correlation between radiomics features and genomic characteristics in ccRCC not only enhances our understanding of its biological mechanisms but also holds paramount importance in identifying biomarkers for targeted or immunotherapy prognosis assessment. This will pave the way for personalized and precise treatment strategies for ccRCC.

This study leveraged radiotranscriptomic data as its focal point and innovatively employed an unsupervised clustering approach to explore the intrinsic subtypes of ccRCC, which evaluated the clinicopathological, prognostic, immune, and molecular characteristics of different subtypes. The results were validated on an external validation cohort. Furthermore, a comprehensive investigation was conducted into the functional aspects of related genes at the multi-omics level revealing the potential clinical significance of these genes in ccRCC. Ultimately, our research provides a noninvasive and effective tumor risk stratification method that facilitates the advancement of personalized and precise treatment strategies for ccRCC.

2. Materials and methods

2.1. Study population

This retrospective study was approved by the Ethics Committee of the First Affiliated Hospital of Guangxi Medical University (Approval No. 2024-E141-01) and the requirement for written informed consent was waived. This was a dual-center study, with Queue 1 (the training cohort) consisting of data from a public database and Queue 2 (the validating cohort) comprising retrospective data from our hospital. In Queue 1, the transcriptomic data, clinicopathological parameters, and follow-up data of ccRCC patients were obtained from The Cancer Genome Atlas, whereas contrast-enhanced computed tomography (CECT) images were acquired from The Cancer Imaging Archive (TCIA) [14]. In Queue 2, the CECT images, clinicopathological characteristics, and follow-up data were obtained from our hospital. The inclusion criteria were as follows: (1) Patients with pathological diagnosis of ccRCC, (2) Patients who underwent preoperative CECT, and (3) Patients with solitary lesions. The exclusion criteria were as follows: (1) poor image quality and (2) missing clinical and follow-up data.

2.2. Region of interest delineation and radiomics feature extraction

We manually segmented 3D regions of interest from the CECT images of each patient using ITK-SNAP software (version 3.8) [15]. Regions of necrosis, hemorrhage, cystic changes, and calcifications within the lesions were delineated, as they represent the internal features of the tumors. Experienced radiologists performed the procedure to ensure accuracy and consistency.

Before feature extraction, we spatially resampled all images to ($1 \times 1 \times 1$ mm^3) voxel spacing. Quantitative radiomics features were extracted from the volumes of interest using PyRadiomics (https://github.com/Radiomics/pyradiomics) [16]. All radiomics features fully adhered to the standard Image Biomarker Standardization Initiative (IBSI) to ensure consistency and compatibility in data

analysis and interpretation. Through feature extraction, 107 radiomics features were obtained, including 18 first-order features, 14 shape features, 16 gray-level size zone matrix features, 14 gray-level dependence matrix features, 24 gray-level co-occurrence matrix features, five neighboring gray-tone difference matrix features, and 16 gray-level run-length matrix features.

2.3. Data preprocessing

First, we employed standardization techniques for both high-dimensional radiomic features and transcriptome data. Radiomics features underwent z-score normalization to ensure uniform scaling and enhance comparability between features. This normalization process entails proportional scaling of the original data. For the transcriptome data, we utilized the log2 (FPKM+1) transformation to standardize the dataset. This transformation facilitates data normalization and ensures comparability across samples. To enhance the reliability and relevance of our analysis, we selectively identified the top 300 highly variable genes from the transcriptome data for subsequent cluster analysis. These genes were chosen based on significant intersample variation, ensuring their capacity to capture the most informative and distinguishing aspects of gene expression patterns.

2.4. Tumor subtype identification

To determine the intrinsic radiotranscriptomics subtypes of ccRCC, we employed the "CancerSubtypes" package in the R language for unsupervised clustering, which is a tool specifically designed for cancer subtype identification [17]. Specifically, we utilized partitioning around the medoid clustering method, employing the Euclidean distance metric to generate subtypes. Throughout the process, 80 % of the patients were randomly sampled to obtain a random subsampling dataset. This process was repeated 1000 times to obtain a conservative matrix. We conducted analyses for two to six different subtype numbers. The optimal number of clusters was determined based on factors, including the relative change between k and k-1 under the cumulative distribution function curve, the degree of separation between consistent clusters in the heatmap, and the average silhouette distance of consistent clusters.

2.5. Comparative analysis of subtypes

To explore the prognostic characteristics, clinicopathological factors, molecules, and immune characteristics among the different subtypes, we conducted comparisons. This included: 1) Comparisons of overall survival (OS) and disease-free survival (DFS) between subtypes; 2) Analysis of clinical-pathological features among different subtypes, including age, sex, tumor stage, and grade; 3) Comparisons of transcriptomics and molecular features (e.g., upregulated differentially expressed genes, signaling pathways, immune cell infiltration, immune function, and certain immune checkpoints) among subtypes; 4) Analysis of gene mutation profiles among different subtypes to determine molecular genetic variations; 5) Comparisons of characteristics such as TMB, microsatellite instability (MSI), RNA interference (RNAi), and tumor immune dysfunction and rejection (TIDE) between subtypes.

2.6. Radiomics subtype validation

To validate the tumor subtyping results, we used local radiomic data from the First Affiliated Hospital of Guangxi Medical University. Firstly, differential analysis of radiomics features in the training cohort was conducted using the "Limma" package to identify significantly different radiomics features between subtypes (with conditions set at $P < 0.05$, Log2 FC > 2). The top ten upregulated radiomics features were selected as signature radiomics features for each subtype, thereby defining each subtype with characteristic radiomics features. Subsequently, the scores for each subtype were computed by averaging the values of the subtype-specific radiomic features for each sample in the validation cohort. Using these subtype scores, the samples in the validation cohort were categorized to determine their respective subtypes. Finally, the value of unsupervised cluster subtyping was validated by comparing the prognosis and clinicopathological characteristics of different subtypes in the validation cohort.

2.7. Statistical analysis

Statistical analyses were conducted using Python (version 3.6) and R (version 3.4.3). The extraction of radiomics features was performed using the "pyradiomics" package, and unsupervised clustering was conducted using the "CancerSubtypes" packages. Baseline characteristics were summarized using descriptive statistics, with frequencies and percentages for categorical variables, and means and standard deviations for continuous variables. Clinicopathological parameters were compared between the different subgroups using the chi-square test. For survival data, Kaplan–Meier analysis was performed using the log-rank test from the "survminer" package, and Univariate Cox regression analyses were performed using the "survival" package. The "Limma" package was used for analyzing differentially expressed genes and radiomics features between distinct subtype groups. Gene set enrichment analysis (GSEA) was also performed to detect pathway alterations in different subtypes. Immune cell infiltration analysis, including 29 immune-related factors, was performed using single-sample GSEA (ssGSEA). We also calculated the Tumor Immune Dysfunction and Exclusion (TIDE) score to evaluate the likelihood of ccRCC subtypes responding to immunotherapy. The TIDE algorithm uses a suite of checkpoint-related signatures to analyze two key immune escape mechanisms: cytotoxic T lymphocyte (CTL) dysfunction and CTL rejection through immunosuppressive factors. Elevated TIDE scores suggest a diminished response and limited benefits from immune checkpoint blockade (ICB) therapy. Lastly, the R package "maftools" served to read MAF files and visualize the mutation landscape between different subgroups.

3. Results

3.1. Clinical features of training and validating cohorts

Fig. 1 shows the technical workflow of this study. In this study, 191 patients with ccRCC were enrolled in the training cohort, whereas the validation cohort consisted of 136 ccRCC patients. The training cohort was used for model construction using unsupervised clustering, whereas the validation cohort was used for model validation. Detailed information about the participants is presented in Table 1.

3.2. Radiotranscriptomics revealed three clusters of ccRCC

A total of 107 radiomic features and 300 highly variable genes were included in cluster analysis. The analysis revealed significant heterogeneity in both radiomic features (Fig. 2A) and gene expression profiles (Fig. 2B) among these patients. Following appropriate data normalization, radiomics and transcriptomics data were merged for radiotranscriptomic analysis, which clustered training patients with ccRCC into distinct subgroups. The findings demonstrated an optimal classification for the consensus clusters with a group (k) = 3, offering the clearest separation among the different groups and the highest average silhouette width (0.82). Consequently, the training cohort of 191 patients with ccRCC was stratified into three distinct subtypes: 84 individuals in cluster 1 (C1), 77 individuals in cluster 2 (C2), and 30 individuals in cluster 3 (C3) (Fig. 3A–I).

3.3. Clinical characteristics of each cluster

The variations in the clinical characteristics and outcomes among the three subtypes were investigated. Kaplan–Meier survival curves revealed that individuals with subtypes C1 and C3 exhibited notably worse OS ($P < 0.05$) than those with the C2 subtype (Fig. 4A). Although there was no statistically significant difference in DFS, both C1 and subtype C3 exhibit survival trends similar to OS when compared with subtype C2 (Fig. 4B). However, no significant associations were observed between the subtypes in terms of

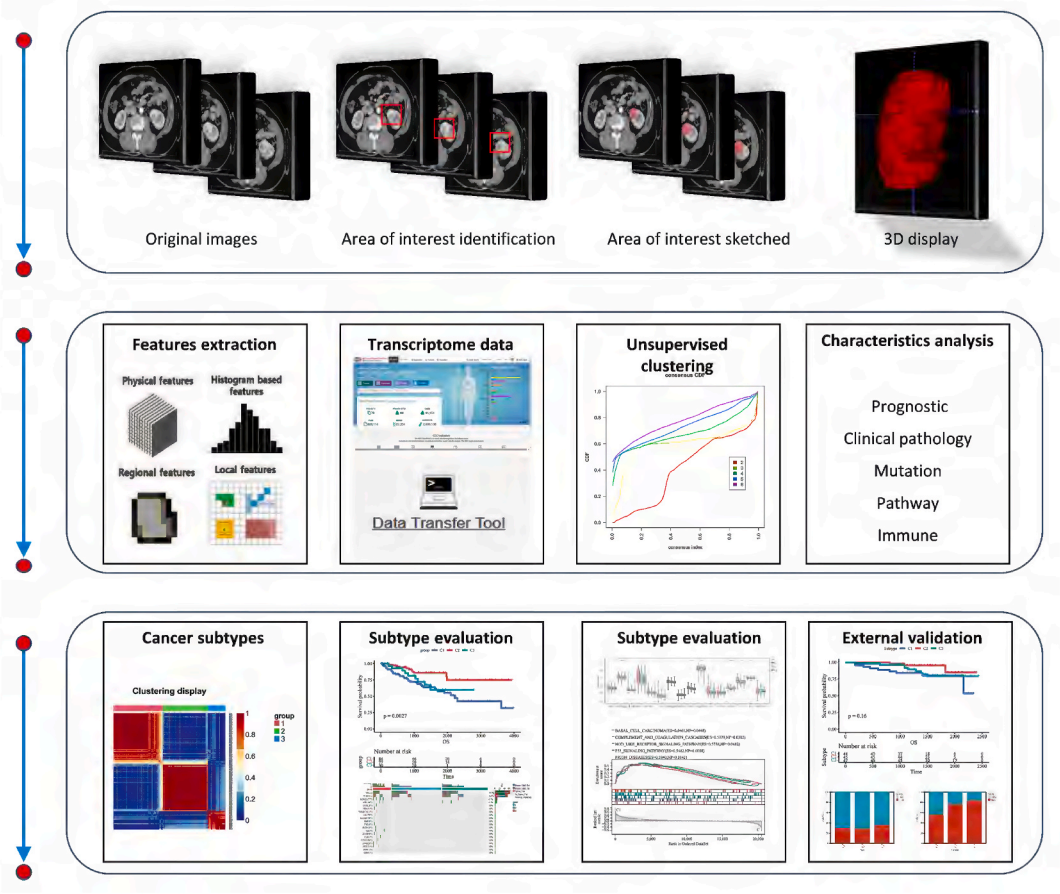


Fig. 1. Data processing steps and technology roadmap.

Table 1
Clinicopathologic baseline characteristics of training and validation cohorts.

Variables	Training cohort (N = 191)	Validation cohort (N = 136)
Sex		
Female	67.0 (35.1 %)	39.0 (28.7 %)
Male	124 (64.9 %)	97.0 (71.3 %)
Age		
Mean (SD)	59.5 (12.1)	52.5 (12.9)
T Stage		
T1	104 (54.5 %)	83.0 (61.0 %)
T2	19.0 (9.9 %)	35.0 (25.7 %)
T3	65.0 (34.0 %)	13.0 (9.6 %)
T4	3.00 (1.6 %)	5.00 (3.7 %)
N Stage		
0	77.0 (40.3 %)	129 (94.9 %)
1	5.00 (2.6 %)	7.00 (5.1 %)
NX	109 (57.1 %)	0 (0 %)
M Stage		
0	164 (85.9 %)	126 (92.6 %)
1	26.0 (13.6 %)	10.0 (7.4 %)
MX	1.00 (0.5 %)	0 (0 %)
Clinical stage		
I	101 (52.9 %)	82.0 (60.3 %)
II	16.0 (8.4 %)	31.0 (22.8 %)
III	47.0 (24.6 %)	15.0 (11.0 %)
IV	27.0 (14.1 %)	8.00 (5.9 %)
Grade		
I	1.00 (0.5 %)	27.0 (19.9 %)
II	75.0 (39.3 %)	56.0 (41.2 %)
III	85.0 (44.5 %)	47.0 (34.6 %)
IV	30.0 (15.7 %)	6.00 (4.4 %)

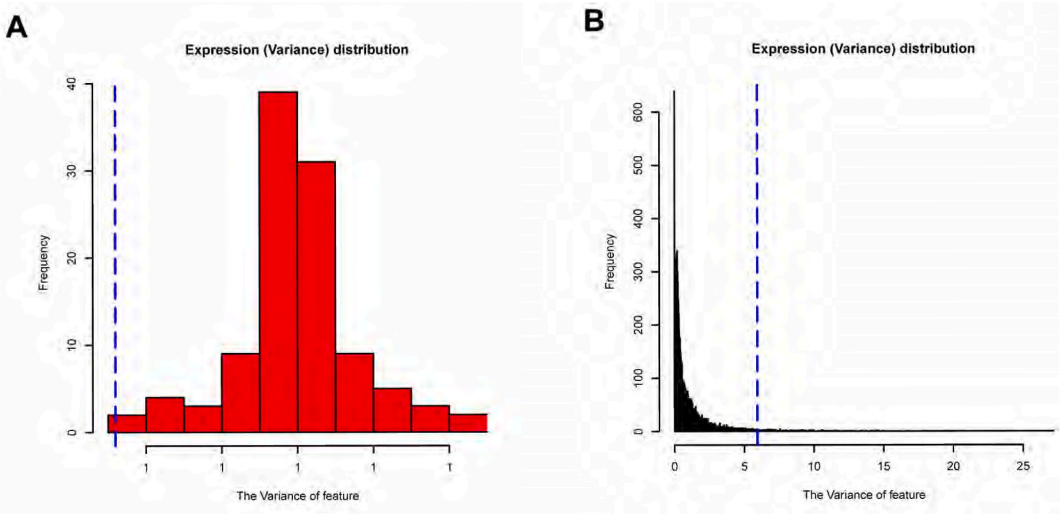


Fig. 2. Distribution of variance in radiomics and transcriptomics features. This figure illustrates the variability of radiomics features and transcriptomics. The horizontal axis represents the variance of features or genes, while the vertical axis represents frequency. Each point represents the variance distribution of expression or feature intensity of a feature or gene in the samples. A) Histogram depicting the frequency distribution of variance in radiomics features. B) Histogram displaying the frequency distribution of variance in transcriptomics.

histological grade (chi-square = 0.2316, P = 0.8907), clinical stage (chi-square = 6.3583, P = 0.3843), age (chi-square = 2.3721, P = 0.3054), or sex (chi-square = 3.4282, P = 0.1801) (Table 2) (Fig. 4C–F).

3.4. Molecular characteristics of each cluster

In addition to assessing the clinical disparities among the three subtypes, this study investigated the disparities in their molecular characteristics. Numerous molecular features varied across the clusters. Univariate Cox analysis of the top 10 upregulated genes suggested that most of the upregulated genes in the C1 subtype were associated with increased risk, whereas those in the C2 subtype

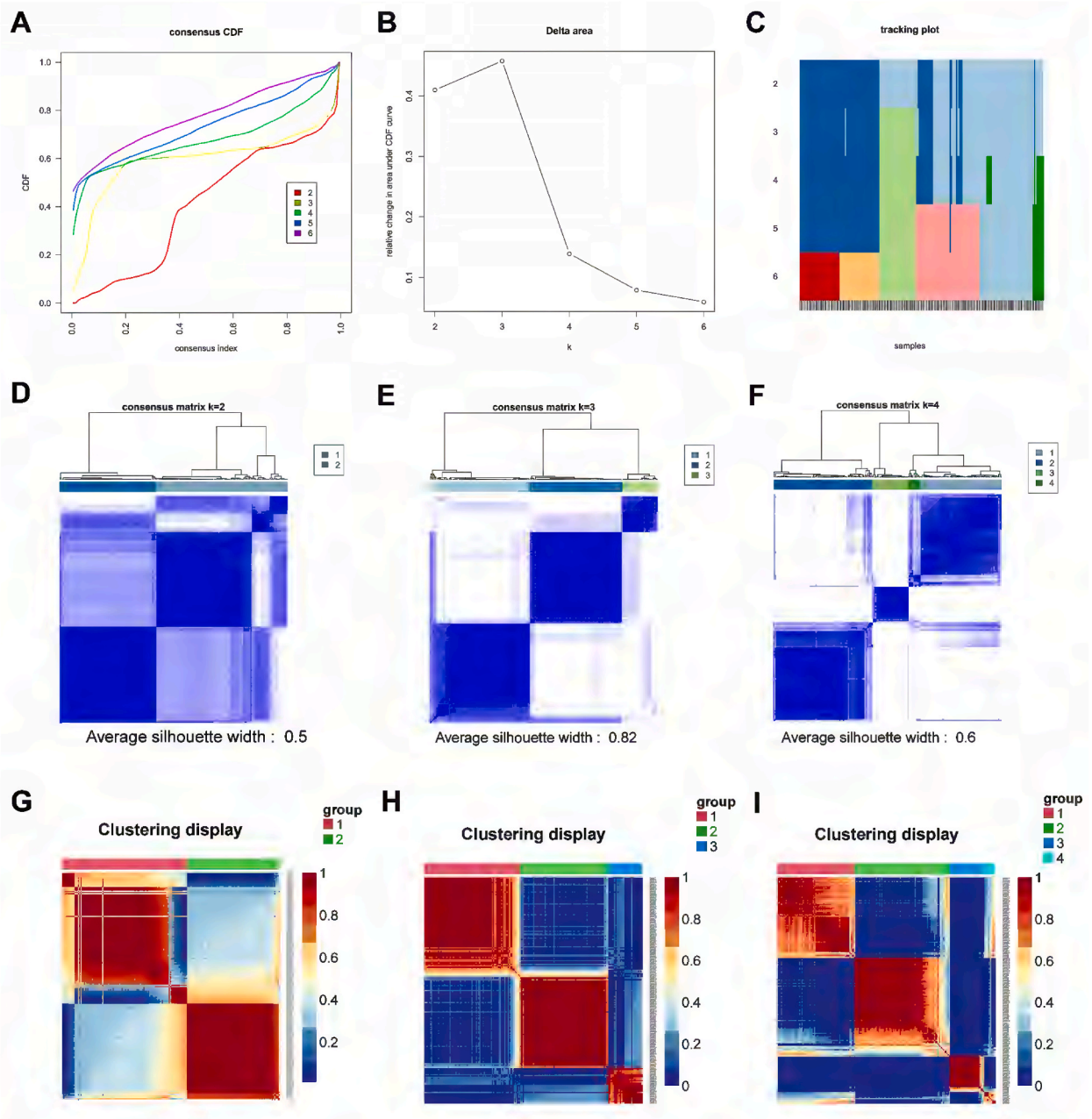


Fig. 3. Unsupervised clustering analysis for identifying distinct subtypes. A) Consensus cumulative distribution function for different numbers of clusters (k), showing the stability of clustering. B) Δ Area plot used to determine the optimal number of clusters based on the relative change in the area under the cumulative distribution function curve. C) Tracking plot demonstrating the stability of sample clustering assignments across different numbers of clusters. D-F) Consistency matrixes for k = 2, k = 3, and k = 4, displaying the average silhouette width, which measures the similarity of objects within their clusters compared to other clusters. G-I) Heatmaps for k = 2, k = 3, and k = 4, depicting cluster displays, where higher consistency indices indicate stronger consistency in cluster assignments. Each matrix visualizes the consistency (1 for complete consistency, 0 for no consistency) of sample clustering, providing insights into clustering structure and stability.

were linked to protective effects. However, most of the top 10 upregulated genes in the C3 subtype were not significantly associated with survival (Fig. 5A–B). Using GSEA, we observed that tumors belonging to the C1 subtype exhibited significant activation of the p53, complement, and coagulation cascades, as well as the NOD-like receptor signaling pathways. The C2 subtype exhibited significant activation of metabolic pathways, including the PPAR signaling pathway, histidine metabolism, valine leucine and isoleucine degradation, propanoate metabolism, and limonene and pinene degradation (Fig. 5C–E). Additionally, cluster 1 exhibited elevated scores for most infiltrating immune cells, infiltrating immune functions, and higher levels of certain immune checkpoint molecules

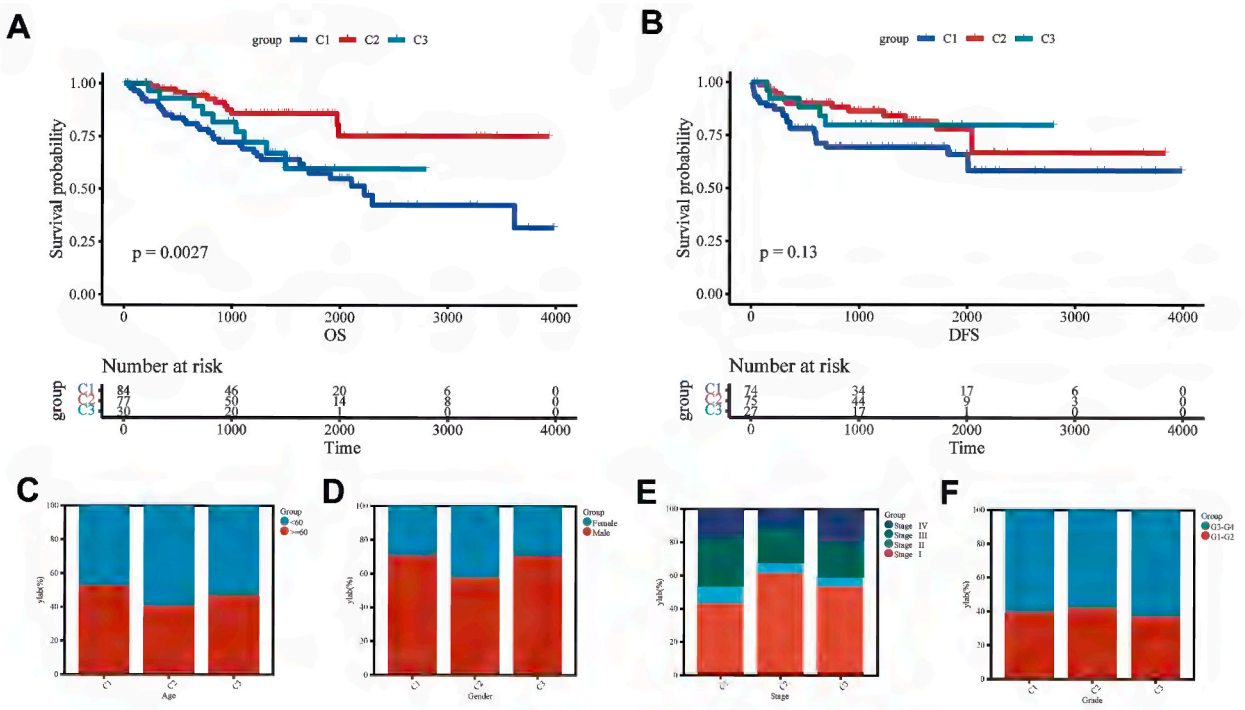


Fig. 4. Kaplan–Meier survival curves and distribution of clinical variables in unsupervised clustering subgroups.
A) Overall survival status of the three patient clusters (C1, C2, C3), with p-values indicating statistically significant differences in survival among groups.
B) Disease-free survival curves for the same clusters, with p-values indicating no significant differences.
C) Distribution of patients across age groups (<60, ≥60) across the three clusters.
D) Distribution of sex within each cluster.
E) Distribution of patients across clinical stages (I, II, III, and IV) across clusters.
F) Distribution of patients within each cluster according to tumor grade (G1-G2, G3-G4).
Each bar graph in panels C–F represents the percentage of patients in each clinical variable category in each cluster.

Table 2
The comparison of clinical characteristics among the different subgroups in the training cohort.

Variables	C1	C2	C3	χ2	p
Sex					
Male	59	44	21	3.4282	0.1801
Female	25	33	9		
Age (year)					
≥60	44	31	14	2.3721	0.3054
<60	40	46	16		
Grade					
G1-G2	33	32	11	0.2316	0.8907
G3-G4	51	45	19		
Clinical stage					
I	36	47	18	6.3583	0.3843
II	9	5	2		
III	24	16	7		
IV	15	9	3		

such as SIGLEC15, TIGIT, and LAG3 (Fig. 6A–C).

Associations between the clusters and genetic mutation data are shown in Fig. 7A. VHL and PBRM1 were the most frequently mutated genes, displaying a higher mutation rate across all subtypes, without significant variation. A comparison of the molecular features associated with gene mutations among the subtypes revealed that C3 exhibited a greater number of mutated genes, including KDM5C and LOXL2. Based on the training cohort, Kaplan–Meier analysis indicated that the mutation status of VHL, PBRM1, KDM5C, MUC16, LOXL2, and FREM1 did not have a significant impact on OS of patients (Fig. 7B–G). Additionally, we calculated the TMB, MSI, RNAi, and TIDE for each sample. TMB, MSI, and RNAi showed no differences among clusters. Fig. 7H demonstrates that the TIDE

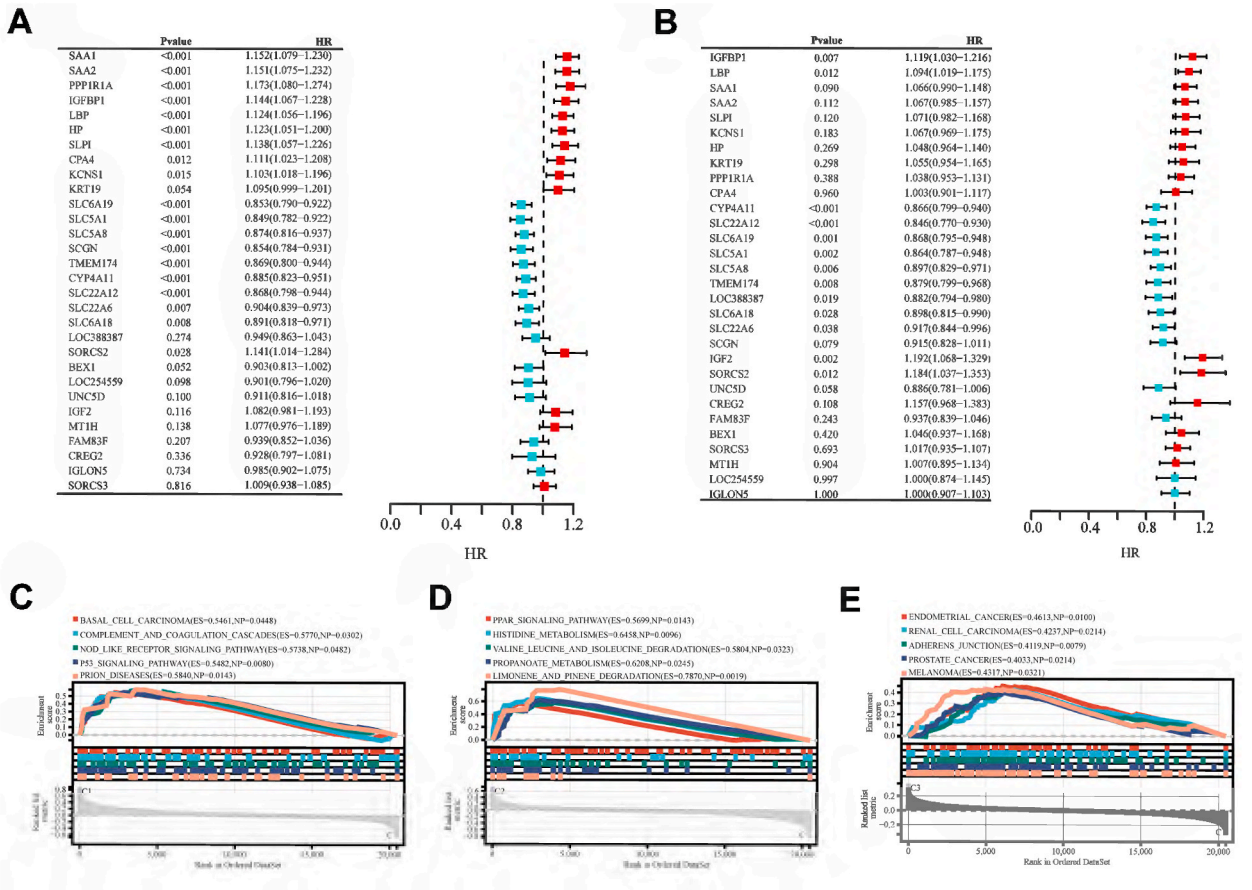


Fig. 5. Univariate Cox analysis and gene set enrichment analysis (GSEA).
A) Forest plot showing hazard ratios (HR) and p-values of the top 10 upregulated genes associated with overall survival for different subtypes. Genes with HR > 1 indicate a higher risk of adverse events, while HR < 1 suggests a protective effect.
B) Forest plot showing HR and p-values of the top 10 upregulated genes associated with disease-free survival for each subtype.
C-E) GSEA plots for subtypes C1, C2, and C3, illustrating the enrichment of various biological pathways. Enrichment scores and p-values for the most significantly enriched pathways are provided. Each plot displays the running enrichment score of gene sets across the genome, ranked by their expression, with peaks indicating the maximum enrichment score of the set.

distribution across all clusters was higher in subtypes C1 and C3, indicating a lower potential response to ICB therapy.

3.5. Algorithm validation with external dataset

To further validate the effectiveness of the radiomic tumor subtyping algorithm, we identified cluster-specific imaging biomarkers. We employed the top 10 radiomic features that exhibited the most significant upregulation based on differential expression analysis as subtype-specific markers (Fig. 8A–F). Subsequently, the scores for C1, C2, and C3 within each subgroup were computed by averaging the values of the subtype-specific radiomic features in the validation cohort. Patients were then allocated to their respective subtypes based on their highest subtype scores. The internal validation cohort included 136 patients who were categorized into subtypes C1 ($n = 44$), C2 ($n = 45$), and C3 ($n = 47$). The radiomic subtypes in the validation cohort displayed similar characteristics in terms of clinicopathological features (Table 3) and prognostic indications to those in the training cohort, suggesting that C1 exhibited a propensity for worse prognostic outcomes in both OS and DFS (Fig. 9A–F). Although not statistically significant, these results support the potential value of radiotranscriptomic analysis in identifying ccRCC subtypes.

4. Discussion

ccRCC is a highly heterogeneous tumor and the most common subtype of renal cell carcinoma (RCC) [18]. The identification of cancer subtypes contributes to a deeper understanding of the clinical, pathological, molecular, and prognostic characteristics of tumors, enabling risk stratification and optimization of individualized diagnosis and treatment [19]. Therefore, the integration of multi-omics data to define ccRCC subtypes is crucial [20]. In this study, we propose a framework to integrate radiotranscriptomic data

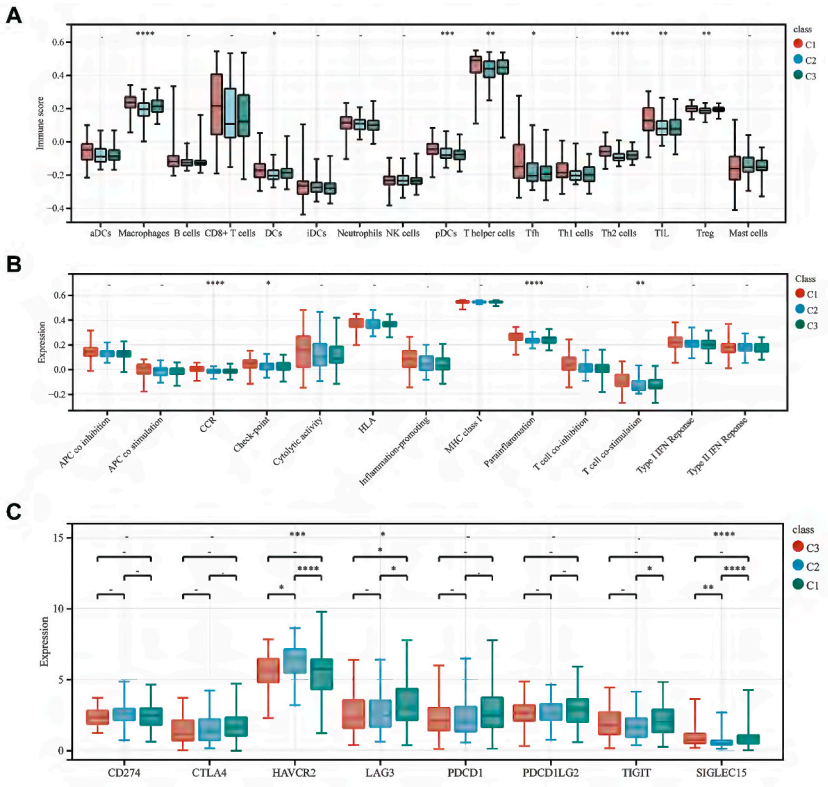


Fig. 6. Comparative analysis of immune infiltration, immune function, and immune checkpoint expression among different subtypes.
A) Boxplots depict the immune scores of infiltrating immune cell types into the tumor microenvironment across the identified subtypes (C1, C2, C3). The immune score quantitatively represents the relative abundance of immune cells.
B) Boxplots display the expression levels of immune functions within each subtype. These functions encompass various immune activities such as antigen presentation, cell chemotaxis, and cytolytic activity.
C) Boxplots illustrate the expression levels of selected immune checkpoint molecules within each subtype.

for ccRCC subtype identification, elucidating the clinical, pathological, molecular, and prognostic features of different ccRCC subtypes. Our results revealed three potential ccRCC subtypes identified through unsupervised clustering that exhibited distinct clinicopathological, prognostic, molecular, and pathway characteristics. Validation using an independent external dataset yielded similar results.

In an era of rapid multi-omics development, transcriptomics, and radiomics have demonstrated great potential for assessing and characterizing tumor heterogeneity. Both have shown promising efficacy in evaluating the clinical and prognostic characteristics of ccRCC. Researchers have constructed models based on transcriptomic information, such as mRNA and miRNAs, to predict the staging and OS of ccRCC patients [21–24]. Similarly, scholars have integrated CT radiomics and clinical information to predict the tumor grade and prognosis of RCC patients [25–28]. These methods fall within the category of supervised machine learning classification. Unlike supervised learning methods, consensus clustering analysis, as an unsupervised learning approach, explores and understands the inherent structure of data without the need for predefined labels, uncovering hidden patterns and relationships in the data, which makes it highly advantageous for exploratory analysis and handling of unknown label data [29,30]. In this study, we identified multiple ccRCC subtypes using unsupervised clustering methods and determined three distinct subtypes as the best fit based on the average silhouette width index.

Upon subtype classification, we conducted comparative analyses of clinicopathological, molecular, and prognostic characteristics among the different subgroups. The results revealed significant discrepancies in the OS rates among these subgroups. The C1 subtype exhibited inferior OS and DFS. These findings were corroborated, to some extent, by external data. Genetic mutation analysis identified the top six mutated genes as VHL, PBRM1, KDM5C, MUC16, LOXL2, and FREM1. The VHL mutation is the most prevalent in ccRCC and is present in approximately 70 % of cases [31]. VHL is known to play a critical role as a tumor suppressor in ccRCC, and its high mutation rate is considered an initiating event in the pathogenesis of ccRCC [32]. PBRM1 is the second most common tumor suppressor gene mutation in ccRCC (40%–50 %) and is often co-deleted with VHL. Previous studies have indicated that although VHL and PBRM1 mutations are widespread in ccRCC, their correlation with ccRCC prognosis remains inconclusive [33–37], which is consistent with our findings. Moreover, mutations in KDM5C, MUC16, and LOXL2 exhibited significant differences in their distribution among the three subtypes. KDM5C and LOXL2 mutations were predominantly found in the C1 subtype, while MUC16 mutations were mainly observed in the C3 subtype. KDM5C mutations occur in approximately 6%–7% of ccRCC cases [35]. Their prognostic significance remains controversial, with some studies suggesting an association between KDM5C and prolonged survival in metastatic ccRCC [38]. LOXL2 is

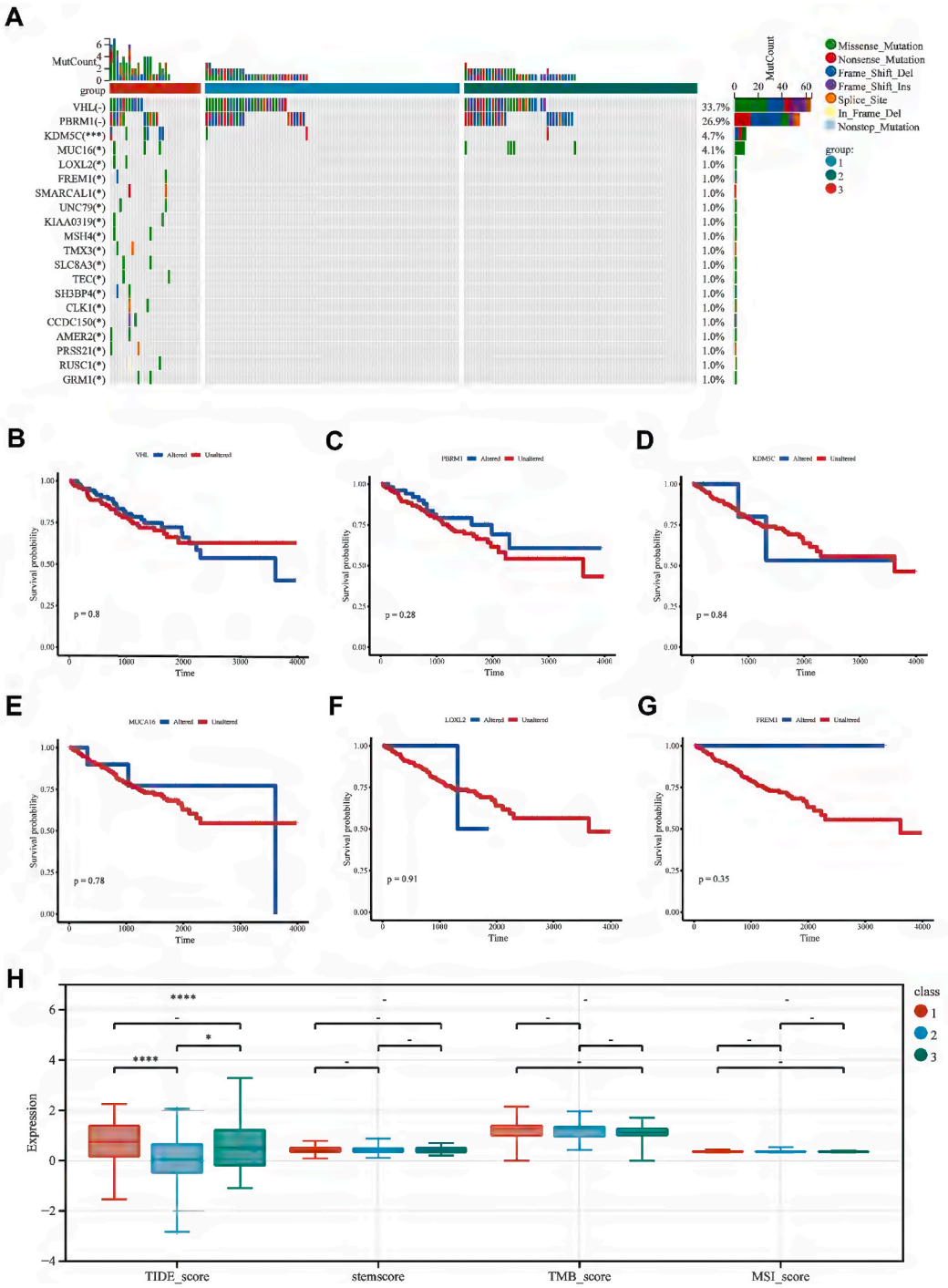


Fig. 7. Analysis of mutation landscape, survival analysis of gene mutations, and evaluation of tumor mutation burden (TMB), microsatellite instability (MSI), RNA interference (RNAi), and tumor immune dysfunction and exclusion (TIDE) among different subtypes.
A) Visualization of mutation landscape across different subtypes, displaying mutation frequencies and types of various genes. The bar plot above the heatmap represents the total mutation counts for each sample, color-coded according to mutation types.
B-G) Kaplan–Meier survival curves compare the survival outcomes between mutated and non-mutated groups for the top six genes with the highest mutation frequencies. p-values indicate the statistical significance of survival differences between groups.
H) Boxplots depict the distribution of TIDE scores, cytolytic scores, TMB scores, and MSI scores across different subtypes. These scores reflect the potential response to immune therapy and the overall immune landscape of the tumor. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

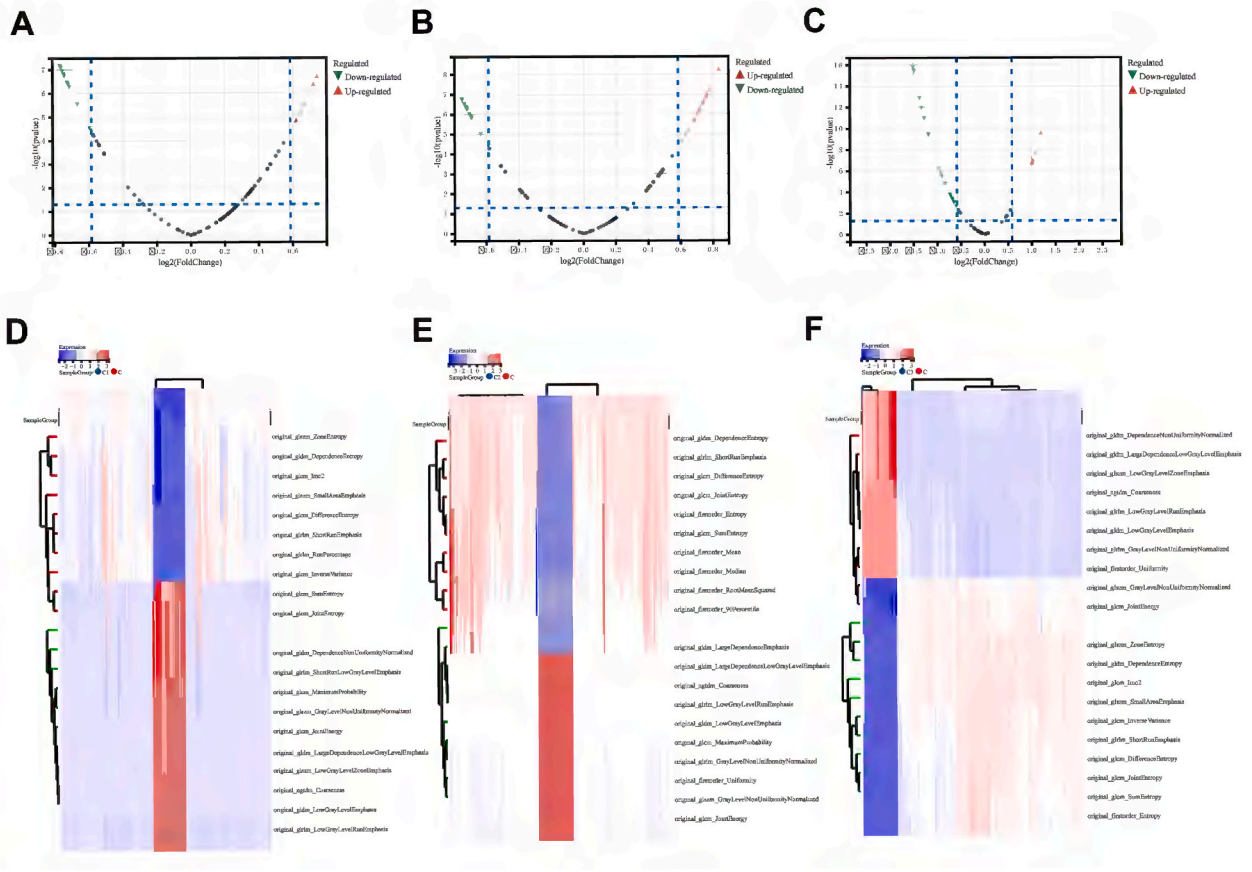


Fig. 8. Volcano plots and heatmaps visualizing radiomics features between subtypes C1, C2, and C3.
A-C) Volcano plots illustrate the differential expression of radiomics features between different subtypes. A) Comparing C1 with C2+C3, B) Contrasting C2 with C1+C3, C) Displaying C3 compared to C1+C2. These plots highlight significantly upregulated (red triangles) and downregulated (green triangles) features in the interested subtype compared to others, with the y-axis showing the negative logarithm of p-values and the x-axis indicating the log2-fold change.
D-F) Heatmaps display the top 10 differentially expressed radiomics features between upregulated and downregulated groups within each subtype. D) Shows the comparison between C1 and C2+C3, E) Describes the contrast between C2 and C1+ C3, F) Represents the comparison between C3 and C1+C2. The heatmaps provide a visual representation of expression patterns for these features, with red indicating higher expression within the respective subtype and blue indicating lower expression. Each column represents a sample, and each row represents a radiomics feature. The dendrogram on the left illustrates hierarchical clustering based on the expression patterns of features. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 3
The comparison of clinical characteristics among the different subgroups in the validation cohort.

Variables	C1	C2	C3	$\chi 2$	p
Sex					
Male	24	36	37	9.8798	0.0072
Female	20	12	7		
Age (year)					
≥60	12	14	14	0.2211	0.8953
<60	32	34	30		
Grade					
G1-G2	29	28	26	0.6567	0.7201
G3-G4	15	20	18		
Clinical stage					
I	24	31	27	8.6869	0.1920
II	13	6	12		
III	4	9	2		
IV	3	2	3		

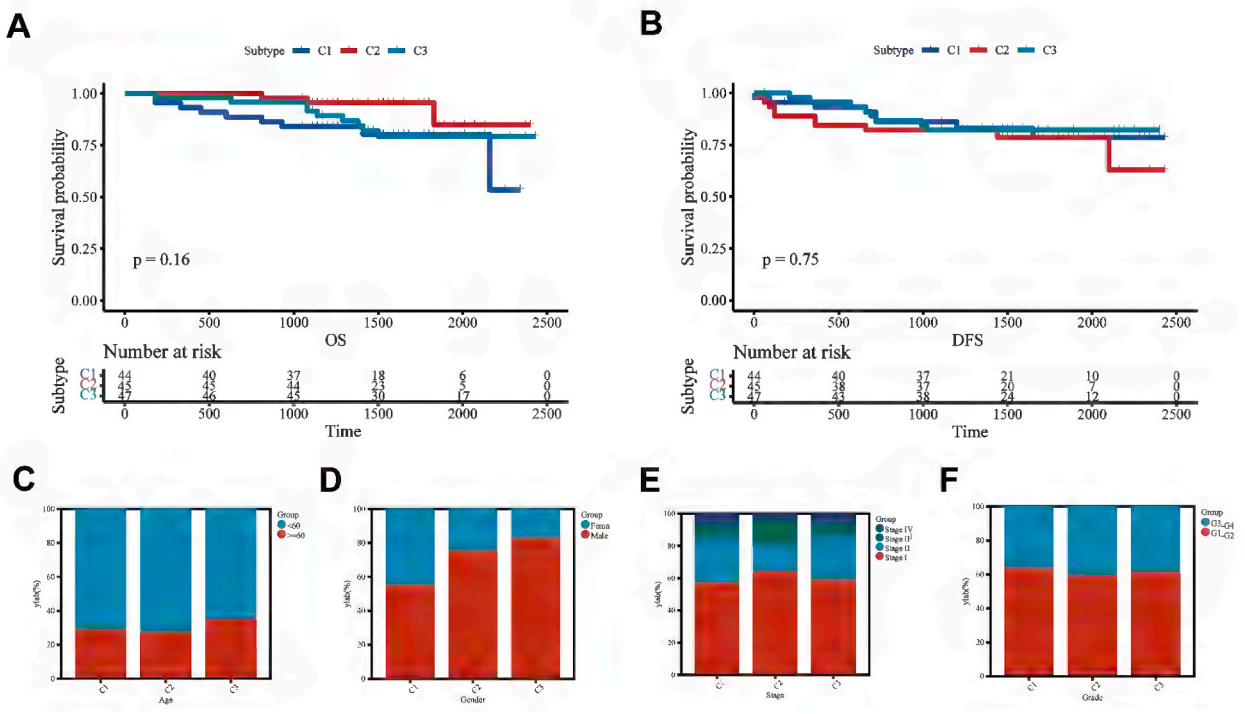


Fig. 9. Kaplan–Meier survival curves and distribution of clinical variables among different subtypes in the validation set.
A) Overall survival curves for subtypes C1, C2, and C3, with p-values indicating no significant difference in survival probability between subtypes.
B) Disease-free survival curves for subtypes C1, C2, and C3, with p-values indicating no significant difference in survival.
C) Bar graph depicting the distribution of age within each subtype, showing the percentage of patients below and above 60 years old.
D) Bar graph illustrating the distribution of sex across subtypes, displaying the percentage of female and male patients.
E) Bar graph demonstrating the distribution of cancer stages (I, II, III, IV) within each subtype.
F) Bar graph representing the distribution of tumor grades (G1-G2, G3-G4) across subtypes.
Each bar in the graph represents the proportion of patients in each clinical variable category within each subtype.

overexpressed in RCC, and its silencing suppresses the migration and invasion of cancer cells in ccRCC cell lines [39]. FREM1 has been identified as an independent prognostic factor in ccRCC and is negatively correlated with tumor staging [40,41]. Although MUC16 exhibits a high mutation frequency in ccRCC, studies on its prognostic implications are limited [42]. Further studies are required to elucidate the mechanisms by which these genes affect ccRCC development.

In-depth research on cancer signaling pathways and immune information will contribute to a comprehensive understanding of ccRCC cellular biology and disease mechanisms, thereby laying a crucial foundation for the development of more effective immunotherapy strategies. To this end, we conducted feature analyses including GSEA pathways, immune infiltration, and tumor stemness.

Interestingly, among the upregulated pathways enriched in the C1 subtype, P53_SIGNALING_PATHWAY emerged as a key pathway associated with cell cycle control and DNA repair. Its enrichment may reflect specific abnormalities in the cellular biology of the C1 subtype, potentially related to the development and progression of RCC. The COMPLEMENT_AND_COAGULATION_CASCADES pathway, involving the immune response and coagulation processes, may imply specific biological characteristics of the immune response and coagulation in C1 subtype ccRCC, possibly associated with tumor infiltration and metastasis. The NOD_LIKE_RECEPTOR_SIGNALING_PATHWAY, which is related to the immune response and inflammation, may indicate an important role of the C1 subtype ccRCC in regulating immune and inflammatory processes, potentially linked to tumor immune escape. Functional pathways involving cell cycle regulation, immune and inflammatory regulation, and tumor immune escape may be associated with the poor prognosis of the C1 subtype.

Within the C2 pathway, the PPAR SIGNALING PATHWAY, HISTIDINE METABOLISM, VALINE LEUCINE AND ISOLEUCINE DEGRADATION, PROPANOATE METABOLISM, and LIMONENE AND PINENE DEGRADATION are involved in various types of molecular metabolism and energy production. These pathways may be related to cancer cell metabolism, signal transduction, and cellular functions. However, the specific functional mechanisms require further research. Additionally, we analyzed the differential infiltration characteristics of TME cells into these three subtypes and evaluated 29 different immune characteristics using the ssGESA algorithm. Evidence suggests that compared to C2 and C3, C1 exhibits increased immune cell infiltration due to its TME. Furthermore, the distribution of TIDE across all clusters was higher in subtypes C1 and C3, indicating potentially lower responses to ICB therapy, which could also contribute to the poor prognosis observed in subtype C1.

This study has some limitations. First, it was a retrospective analysis, validated externally through a local dataset; however,

prospective studies with larger sample sizes are needed for validation in the future. Second, the tumor volumes of interest were extracted through manual segmentation. Nevertheless, with the advancement of deep-learning techniques, automated segmentation methods have become the mainstream approach in tumor research. This would save time in the segmentation process and facilitate better clinical application by reducing human error and improving efficiency. Finally, this study primarily focused on traditional radiomics features. However, we anticipate that future research will explore new domains, like deep learning features or habitat features. These novel methods and technologies have the potential to further enhance the accuracy and efficacy of tumor diagnosis and treatment. Future research should focus on addressing these limitations to contribute to the advancement of the field and further enhance the clinical applicability of our findings.

In summary, characterizing subtypes based on radiotranscriptomics provides a powerful means to evaluate ccRCC tumor heterogeneity. The findings of this study suggest that radiotranscriptomics can accurately stratify the risk for patients with ccRCC, particularly in prognosis assessment and evaluating molecular features. This comprehensive approach reveals the complex characteristics of tumors, providing clinicians with accurate patient information and laying a solid foundation for personalized treatment and precision medicine.

Ethics statement

The research, including clinical images, received ethical approval from the Ethics Committee of the First Affiliated Hospital of Guangxi Medical University (Approval No. 2024-E141-01). Written informed consent was not required for this retrospective study.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author.

CRediT authorship contribution statement

Ruizhi Gao: Writing – review & editing, Writing – original draft, Visualization, Validation. **Jinshu Pang:** Validation, Software, Investigation. **Peng Lin:** Software, Resources, Project administration. **Rong Wen:** Formal analysis, Data curation. **Dongyue Wen:** Visualization, Methodology. **Yiqiong Liang:** Data curation, Conceptualization. **Zhen Ma:** Formal analysis, Data curation, Conceptualization. **Li Liang:** Visualization, Software, Resources. **Yun He:** Supervision, Project administration. **Hong Yang:** Writing – review & editing, Supervision, Software, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Y.Z. Li, H.C. Zhu, Y. Du, H.C. Zhao, L. Wang, Silencing lncRNA SLC16A1-AS1 induced ferroptosis in renal cell carcinoma through miR-143-3p/SLC7A11 signaling, Technol. Cancer Res. Treat. 21 (2022) 15330338221077803.
[2] X. He, F. Tian, F. Guo, F. Zhang, H. Zhang, J. Ji, L. Zhao, J. He, Y. Xiao, L. Li, C. Wei, C. Huang, Y. Li, F. Zhang, B. Yang, H. Ye, F. Wang, Circulating exosomal mRNA signatures for the early diagnosis of clear cell renal cell carcinoma, BMC Med. 20 (1) (2022) 270.
[3] X. Zhang, H. Kang, J. Xiao, B. Shi, X. Li, G. Chen, LHPP inhibits the proliferation and metastasis of renal cell carcinoma, BioMed Res. Int. 2020 (2020) 7020924.
[4] R. Iaxx, F. Lefort, C. Domblides, A. Ravaud, J.C. Bernhard, M. Gross-Goupil, An evaluation of cabozantinib for the treatment of renal cell carcinoma: focus on patient selection and perspectives, Therapeut. Clin. Risk Manag. 18 (2022) 619–632.
[5] W.P. Huang, S.Y. Liu, Y.J. Han, L.M. Li, P. Liang, J.B. Gao, Development of CT-based imaging signature for preoperative prediction of invasive behavior in pancreatic solid pseudopapillary neoplasm, Front. Oncol. 11 (2021) 677814.
[6] H. Liao, Y. Li, Y. Yang, H. Liu, J. Zhang, H. Liang, G. Yan, Y. Liu, Comparison of multiple radiomics models for identifying histological grade of pancreatic ductal adenocarcinoma preoperatively based on multiphasic contrast-enhanced computed tomography: a two-center study in southwest China, Diagnostics 12 (8) (2022).
[7] X. Wang, B. Hao, C. Xu, X. Zhao, C. Liu, X. Chu, Y. Lv, Y. Zhao, S. Zhang, P. Wang, Y. Wang, Involvement of erbB4 and tumor marker genes in renal carcinoma regulatory network, Saudi J. Biol. Sci. 24 (8) (2017) 1787–1791.
[8] S.W. Ra, T. Lee, H.J. Cha, C.R. Park, J. Baek, Y. Chee, W.J. Kwon, Rotation aiding technique for endobronchial ultrasound-guided transbronchial needle aspiration biopsy of intrathoracic lymph nodes: a complementary approach to the conventional jabbing method, Thoracic cancer 13 (11) (2022) 1712–1718.
[9] Q. Liu, P. Hu, Radiogenomic association of deep MR imaging features with genomic profiles and clinical characteristics in breast cancer, Biomark. Res. 11 (1) (2023) 9.
[10] T. Wakabayashi, F. Ouhmich, C. Gonzalez-Cabrera, E. Felli, A. Saviano, V. Agnus, P. Savadjiev, T.F. Baumert, P. Pessaux, J. Marescaux, B. Gallix, Radiomics in hepatocellular carcinoma: a quantitative review, Hepatology international 13 (5) (2019) 546–559.
[11] X.M. He, J.X. Zhao, D.L. He, J.L. Ren, L.P. Zhao, G. Huang, Radiogenomics study to predict the nuclear grade of renal clear cell carcinoma, European journal of radiology open 10 (2023) 100476.
[12] P. Lin, Y.Q. Lin, R.Z. Gao, W.J. Wan, Y. He, H. Yang, Integrative radiomics and transcriptomics analyses reveal subtype characterization of non-small cell lung cancer, Eur. Radiol. 33 (9) (2023) 6414–6425.
[13] A. Haghighat Jahromi, D.A. Barkauskas, M. Zabel, A.M. Goodman, G. Frampton, M. Nikanjam, C.K. Hoh, R. Kurzrock, Relationship between tumor mutational burden and maximum standardized uptake value in 2-[(18)F]FDG PET (positron emission tomography) scan in cancer patients, EJNMMI Res. 10 (1) (2020) 150.
[14] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, F. Prior, The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository, J. Digit. Imag. 26 (6) (2013) 1045–1057.
[15] P.A. Yushkevich, J. Piven, H.C. Hazlett, R.G. Smith, S. Ho, J.C. Gee, G. Gerig, User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability, Neuroimage 31 (3) (2006) 1116–1128.

[16] J.J.M. van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R.G.H. Beets-Tan, J.C. Fillion-Robin, S. Pieper, H. Aerts, Computational radiomics system to decode the radiographic phenotype, Cancer Res. 77 (21) (2017) e104–e107.
[17] T. Xu, T.D. Le, L. Liu, N. Su, R. Wang, B. Sun, A. Colaprico, G. Bontempi, J. Li, CancerSubtypes: an R/Bioconductor package for molecular cancer subtype identification, validation and visualization, Bioinformatics 33 (19) (2017) 3131–3133 (Oxford, England).
[18] C. Qian, Y. Sun, S. Di, H. Wang, Y. Tian, C. Wang, X. Cui, Construction and validation of a prognostic model for predicting clear cell renal cell carcinoma based on complement-related genes, Transl. Androl. Urol. 12 (4) (2023) 659–672.
[19] C.D. Arvanitis, G.B. Ferraro, R.K. Jain, The blood-brain barrier and blood-tumour barrier in brain tumours and metastases, Nat. Rev. Cancer 20 (1) (2020) 26–41.
[20] Y. Guo, J. Zheng, X. Shang, Z. Li, A similarity regression fusion model for integrating multi-omics data to identify cancer subtypes, Genes 9 (7) (2018).
[21] F. Li, M. Yang, Y. Li, M. Zhang, W. Wang, D. Yuan, D. Tang, An improved clear cell renal cell carcinoma stage prediction model based on gene sets, BMC Bioinf. 21 (1) (2020) 232.
[22] Y. Wang, J. Liu, L. Zhang, Y. Li, A predictive model based on pyroptosis-related gene features can effectively predict clear cell renal cell carcinoma prognosis and may Be an underlying target for immunotherapy, Dis. Markers 2022 (2022) 6402599.
[23] A. Jiang, Y. Bao, A. Wang, W. Gong, X. Gan, J. Wang, Y. Bao, Z. Wu, B. Liu, J. Lu, L. Wang, Establishment of a prognostic prediction and drug selection model for patients with clear cell renal cell carcinoma by multiomics data analysis, Oxid. Med. Cell. Longev. 2022 (2022) 3617775.
[24] Z. Zhang, Y. Tang, Y. Liu, H. Zhuang, E. Lin, L. Xie, X. Feng, K. Tian, J. Zeng, J. Liu, Y. Yu, A novel immune-related lncRNA-based model for survival prediction in clear cell renal cell carcinoma, Journal of immunology research 2021 (2021) 9921466.
[25] C.G. Moldovanu, B. Boca, A. Lebovici, A. Tamas-Szora, D.S. Feier, N. Crisan, I. Andras, M.M. Buruian, Preoperative predicting the WHO/ISUP nuclear grade of clear cell renal cell carcinoma by computed tomography-based radiomics features, J. Personalized Med. 11 (1) (2020).
[26] G. Yang, P. Nie, L. Yan, M. Zhang, Y. Wang, L. Zhao, M. Li, F. Xie, H. Xie, X. Li, F. Xiang, N. Wang, N. Cheng, X. Zhao, N. Wang, Y. Wang, C. Chen, C. Yun, J. Cui, S. Duan, R. Zhang, D. Hao, X. Wang, Z. Wang, H. Niu, The radiomics-based tumor heterogeneity adds incremental value to the existing prognostic models for predicting outcome in localized clear cell renal cell carcinoma: a multicenter study, Eur. J. Nucl. Med. Mol. Imag. 49 (8) (2022) 2949–2959.
[27] Z. Khodabakhshi, M. Amini, S. Mostafaei, A. Haddadi Avval, M. Nazari, M. Oveisi, I. Shiri, H. Zaidi, Overall survival prediction in renal cell carcinoma patients using computed tomography radiomic and clinical information, J. Digit. Imag. 34 (5) (2021) 1086–1098.
[28] H. He, Z. Jin, J. Dai, H. Wang, J. Sun, D. Xu, Computed tomography-based radiomics prediction of CTLA4 expression and prognosis in clear cell renal cell carcinoma, Cancer Med. 12 (6) (2023) 7627–7638.
[29] H. Alashwal, M. El Halaby, J.J. Crouse, A. Abdalla, A.A. Moustafa, The application of unsupervised clustering methods to alzheimer's disease, Front. Comput. Neurosci. 13 (2019) 31.
[30] L. Manganaro, S. Bianco, P. Bironzo, F. Cipollini, D. Colombi, D. Corà, G. Corti, G. Doronzo, L. Errico, P. Falco, L. Gandolfi, F. Guerrera, V. Monica, S. Novello, M. Papotti, S. Parab, A. Pittaro, L. Primo, L. Righi, G. Sabbatini, A. Sandri, S. Vattakunnel, F. Bussolino, G.V. Scagliotti, Consensus clustering methodology to improve molecular stratification of non-small cell lung cancer, Sci. Rep. 13 (1) (2023) 7759.
[31] J. Zhou, J. Wang, W. Kong, J. Zhang, X. Wu, J. Huang, J. Zheng, Y. Chen, W. Zhai, W. Xue, VHL and DNA damage repair pathway alterations as potential clinical biomarkers for first-line TKIs in metastatic clear cell renal cell carcinomas, Cell. Oncol. 45 (4) (2022) 677–687.
[32] F. Massari, C. Ciccarese, M. Santoni, M. Brunelli, F. Piva, A. Modena, D. Bimbatti, E. Fantinel, D. Santini, L. Cheng, S. Cascinu, R. Montironi, G. Tortora, Metabolic alterations in renal cell carcinoma, Cancer Treat Rev. 41 (9) (2015) 767–776.
[33] Y. Sato, T. Yoshizato, Y. Shiraishi, S. Maekawa, Y. Okuno, T. Kamura, T. Shimamura, A. Sato-Otsubo, G. Nagae, H. Suzuki, Y. Nagata, K. Yoshida, A. Kon, Y. Suzuki, K. Chiba, H. Tanaka, A. Niida, A. Fujimoto, T. Tsunoda, T. Morikawa, D. Maeda, H. Kume, S. Sugano, M. Fukayama, H. Aburatani, M. Sanada, S. Miyano, Y. Homma, S. Ogawa, Integrated molecular analysis of clear-cell renal cell carcinoma, Nat. Genet. 45 (8) (2013) 860–867.
[34] A.A. Hakimi, M.H. Voss, F. Kuo, A. Sanchez, M. Liu, B.G. Nixon, L. Vuong, I. Ostrovnaya, Y.B. Chen, V. Reuter, N. Riaz, Y. Cheng, P. Patel, M. Marker, A. Reising, M.O. Li, T.A. Chan, R.J. Motzer, Transcriptomic profiling of the tumor microenvironment reveals distinct subgroups of clear cell renal cell cancer: data from a randomized phase III trial, Cancer Discov. 9 (4) (2019) 510–525.
[35] Comprehensive molecular characterization of clear cell renal cell carcinoma, Nature 499 (7456) (2013) 43–49.
[36] M. Gerlinger, S. Horswell, J. Larkin, A.J. Rowan, M.P. Salm, I. Varela, R. Fisher, N. McGranahan, N. Matthews, C.R. Santos, P. Martinez, B. Phillimore, S. Begum, A. Rabinowitz, B. Spencer-Dene, S. Gulati, P.A. Bates, G. Stamp, L. Pickering, M. Gore, D.L. Nicol, S. Hazell, P.A. Futreal, A. Stewart, C. Swanton, Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing, Nat. Genet. 46 (3) (2014) 225–233.
[37] S.J. Nam, C. Lee, J.H. Park, K.C. Moon, Decreased PBRM1 expression predicts unfavorable prognosis in patients with clear cell renal cell carcinoma, Urol. Oncol. 33 (8) (2015), 340.e9-16.
[38] D.M. Tennenbaum, B.J. Manley, E. Zabor, M.F. Becerra, M.I. Carlo, J. Casuscelli, A. Redzematovic, N. Khan, M.E. Arcila, M.H. Voss, D.R. Feldman, R.J. Motzer, N.E. Benfante, J.A. Coleman, P. Russo, J.J. Hsieh, A.A. Hakimi, Genomic alterations as predictors of survival among patients within a combined cohort with clear cell renal cell carcinoma undergoing cytoreductive nephrectomy, Urol. Oncol. 35 (8) (2017) 532.e7, 532.e13.
[39] R. Nishikawa, T. Chiyomaru, H. Enokida, S. Inoguchi, T. Ishihara, R. Matsushita, Y. Goto, I. Fukumoto, M. Nakagawa, N. Seki, Tumour-suppressive microRNA-29s directly regulate LOXL2 expression and inhibit cancer cell migration and invasion in renal cell carcinoma, FEBS Lett. 589 (16) (2015) 2136–2145.
[40] T. Zhou, W. Chen, Z. Wu, J. Cai, C. Zhou, A newly defined basement membrane-related gene signature for the prognosis of clear-cell renal cell carcinoma, Front. Genet. 13 (2022) 994208.
[41] J. Luo, Y. Xie, Y. Zheng, C. Wang, F. Qi, J. Hu, Y. Xu, Comprehensive insights on pivotal prognostic signature involved in clear cell renal cell carcinoma microenvironment using the ESTIMATE algorithm, Cancer Med. 9 (12) (2020) 4310–4323.
[42] Z. Yang, J. Zhao, L. Zhou, Y. Wang, P. Hao, Z. Fang, De novo mutations contribute to the development of clear-cell renal-cell carcinoma in a 5-year-old girl, Clin. Genitourin. Cancer 16 (3) (2018) e553–e556.