

Research Article

Establishment of a Prognostic Prediction and Drug Selection Model for Patients with Clear Cell Renal Cell Carcinoma by Multiomics Data Analysis

Aimin Jiang ¹, Yewei Bao,¹ Anbang Wang ², Wenliang Gong,¹ Xinxin Gan,¹ Jie Wang,¹ Yi Bao,² Zhenjie Wu,¹ Bing Liu ³, Juan Lu ⁴, and Linhui Wang ¹

¹Department of Urology, Changhai Hospital, Naval Medical University (Second Military Medical University), Shanghai, China

²Department of Urology, Changzheng Hospital, Naval Medical University, (Second Military Medical University), Shanghai, China

³Department of Urology, The Third Affiliated Hospital, Naval Medical University (Second Military Medical University), Shanghai, China

⁴Vocational Education Center, Naval Medical University (Second Military Medical University), Shanghai, China

Correspondence should be addressed to Bing Liu; 13501616398@163.com, Juan Lu; lujuan506@sina.com, and Linhui Wang; wanglinhuicz@163.com

Received 19 July 2021; Revised 11 October 2021; Accepted 18 November 2021; Published 4 January 2022

Academic Editor: Qiangqiang Li

Copyright © 2022 Aimin Jiang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rationale. Patients with clear cell renal cell cancer (ccRCC) may have completely different treatment choices and prognoses due to the wide range of heterogeneity of the disease. However, there is a lack of effective models for risk stratification, treatment decision-making, and prognostic prediction of renal cancer patients. The aim of the present study was to establish a model to stratify ccRCC patients in terms of prognostic prediction and drug selection based on multiomics data analysis. **Methods.** This study was based on the multiomics data (including mRNA, lncRNA, miRNA, methylation, and WES) of 258 ccRCC patients from TCGA database. Firstly, we screened the feature values that had impact on the prognosis and obtained two subtypes. Then, we used 10 algorithms to achieve multiomics clustering and conducted pseudotiming analysis to further validate the robustness of our clustering method, based on which the two subtypes of ccRCC patients were further subtyped. Meanwhile, the immune infiltration was compared between the two subtypes, and drug sensitivity and potential drugs were analyzed. Furthermore, to analyze the heterogeneity of patients at the multiomics level, biological functions between two subtypes were compared. Finally, Boruta and PCA methods were used for dimensionality reduction and cluster analysis to construct a renal cancer risk model based on mRNA expression. **Results.** A prognosis predicting model of ccRCC was established by dividing patients into the high- and low-risk groups. It was found that overall survival (OS) and progression-free interval (PFI) were significantly different between the two groups ($p < 0.01$). The area under the OS time-dependent ROC curve for 1, 3, 5, and 10 years in the training set was 0.75, 0.72, 0.71, and 0.68, respectively. **Conclusion.** The model could precisely predict the prognosis of ccRCC patients and may have implications for drug selection for ccRCC patients.

1. Introduction

Renal cancer is a malignant tumor derived from the proximal tubule epithelium in the renal parenchyma [1]. It is one of the most common tumors of the urinary system and one of the ten most common tumors in the world [2]. The incidence and mortality of renal cancer are increasing yearly. The latest world cancer statistics show that the num-

ber of new cases of renal cancer is about 431,288 (2.2%) and the number of deaths is about 179,368 (1.8%) worldwide each year. Clear cell renal cell carcinoma (ccRCC) or kidney renal clear cell carcinoma (KIRC) is the most common pathological type of renal cancer, accounting for 70-85% of total RCC [3]. It is noteworthy that most renal cancer cases are insidious and lack typical clinical symptoms. Therefore, about 20-30% renal cancer patients have metastasized at

the initial diagnosis, and 30% patients may have recurrence or metastasis. Renal cancer is not sensitive to traditional radiotherapy and chemotherapy. The current main treatments include surgery, chemotherapy, and immunotherapy, but about 30% patients will develop primary or secondary drug resistance. The degree of CD8⁺ T cell infiltration is positively correlated with a better prognosis in most solid tumors, whereas it is often associated with a worse prognosis in ccRCC [4]. It is therefore urgent to gain a better understanding about the heterogeneity of renal cancer patients and develop an accurate and comprehensive risk model to stratify renal cancer patients for the sake of designing individualized treatment plans in terms of prognostic prediction and drug selection.

With the advancement of sequencing technology and a variety of machine learning algorithms, bioinformatics has exploded, which has greatly promoted people's understanding of tumors at the molecular level [5]. At the same time, with the development of data science, a variety of robust cluster algorithms have pushed the progression of multiomics bioinformatics science. Omics data reflect the various biological processes of cancer and provide a detailed description of the molecular mechanism of cells. Different levels of omics data reflect different relationships between the distribution of the whole genome and the occurrence, development and prognosis of cancer [6]. Unfortunately, there is a lack of typing studies based on multiple omics prognostic indicators in renal cancer omics research. In this study, we performed a stratification analysis of ccRCC patients by integrating multi omics and advanced cluster algorithms and found that different subgroups presented different landscapes in each single omics dataset [7–16]. Finally, we used Borta algorithm to reduce the dimension of subtype-specific signatures and constructed a robust risk stratification model.

2. Materials and Methods

2.1. Data Collection and Preprocessing. The workflow of this study is shown in Figure S1. TCGA database is a comprehensive database of the American Cancer and Tumor Project Gene Atlas Project consisting of multiple omics data of different tumor types. Japan ccRCC database was included to validate the consistency of cluster analysis [17]. We downloaded ccRCC-related datasets of 530 ccRCC samples and 71 normal control samples from TCGA, which contains RNA-seq profiles, miRNA-seq profiles, the Illumina 450K DNA methylation array, and WES (Whole Exome Sequencing) data. Patients were excluded if they (1) missed any single omics data above, (2) did not have prognostic information, and (3) died within 30 days. After data filtration, we finally included 258 cases of ccRCC with complete multiomics information and clinical information into the subsequent study.

2.2. Cluster Analysis. We divided 258 ccRCC patients into two groups by clustering their multiomics database. Ten clustering algorithms were included in the clustering process: SNF, PINSPlus, NEMO, COCA, LRAcluster, Con-

sensusClustering, IntNMF, CLMLR, MoCluster, and iClusterBayes. Based on consensus ensembles, the results of different clustering algorithms were integrated, and the unified samples were clustered into the same category in different algorithms, namely, $CM = \sum_{t=1}^t M^{(t)}$ and $cm_{ij} \in [0, 10]$ [18]. Finally, we calculated the consensus matrix to represent robust pairwise similarities for samples because it considers different multiomics integrative clustering algorithms. Then, dimensionality reduction and visualization of subtype data at multiomics level were performed based on the idea of t-SNE graph dimensionality reduction using R “Rtsne” package.

2.3. Comparison between Subgroups. Prognostic analysis of overall survival (OS) and progression-free interval (PFI) was performed based on the log-rank test between CS1 (Cancer Subtype 1) and CS2 (Cancer Subtype 2) subtypes obtained in the cluster analysis above. Differentially expressed genes (DEGs) between different subgroups were determined using the R “DEseq2” package and visualized by R “EnhancedVolcano” package. The screening criterion for DEGs was set at adjusted p value < 0.01 and abstract $\log_{2}FC > 1.8$. Gene ontology (GO), Kyoto encyclopedia of genes and genomes (KEGG), and gene set enrichment analysis (GSEA) were performed to identify enriched GO terms, pathways, and hallmarks by using the R package with a screening cutoff value at p value < 0.05 and an adjusted p value < 0.2 [19]. Gene set variation analysis (GSVA) was performed to identify cancer-related biological processes using R “GSVA” package [20]. The cancer-related hallmarks were downloaded from the MSigDB database (<https://www.gsea-msigdb.org/gsea/msigdb/>).

2.4. Immunity Analysis and Immunity-Related Gene Expression. To quantify the proportion of immune cells in the two subtypes, several immune-related algorithms were used to calculate the cellular components or immune cell enrichment scores in the ccRCC tissue, including TIMER, CIBERSORT, QUANTISEQ, MCPOUNTER, XCELL, and EPIC [21–24]. Single sample gene set enrichment analysis (ssGSEA) was employed to quantify the relative abundance of 28 immune cells in the tumor microenvironment (TME) of ccRCC [20]. Differences in immune cell infiltration in TME were visualized by Heatmap and boxplot. R “Estimate” package was used to identify the stromal component and immune component between the two subgroups. The Tumor Immune Dysfunction and Exclusion (TIDE) (<http://tide.dfci.harvard.edu>) algorithms were applied to predict the immunotherapy response of each ccRCC patient.

2.5. Copy Number Variation (CNV) Analysis. As the main analysis tool for somatic mutation data analysis and visualization, the R “maftools” package provides the possibility to compare the differences between different subtypes at the WES level [25]. Through the correlation function of this package, we analyzed the tumor mutation panorama, base conversion and transversion, amino acid mutation hotspot, mutation frequency of mutation alleles, copy number mutation, mutual exclusion or coexistence mutation, and gene

mutation survival in different subtypes of ccRCC patients. At the same time, the differences in drug-gene interactions and carcinogenic signaling pathways were also analyzed in this function module. Recurrent broad and focal somatic copy-number alteration (SCNA) analysis was performed by the GISTIC 2.0 algorithm (https://www.genepattern.org/modules/docs/GISTIC_2.0) on the basis of Euclidean distance of threshold copy number at recurring peaks using Ward's method [26].

2.6. Drug Sensitivity Analysis. The response of ccRCC patients to chemotherapy drugs and small-molecule prodrugs was predicted based on the public pharmacogenomics database-Genomics of Drug Sensitivity in Cancer (GDSC; <https://www.cancerrxgene.org>) [27]. The half maximal inhibitory concentration (IC50) was estimated by R “pRRophetic” package [28]. The prediction process was implemented by R “pRRophetic” package, where IC50 of the samples was estimated by ridge regression and the prediction accuracy was evaluated by 10-fold cross-validation based on the GDSC training set. Finally, the molecular structure tomographs of these candidate drugs were obtained from PubChem (<https://pubchem.ncbi>, <http://nlm.nih.gov/>).

2.7. Construction of the Subtype Landscape. To facilitate visualization and uncover the underlying structural distribution of individual ccRCC patients, we applied the graph learning-based dimensionality reduction technique to the risk-related gene expression profiles in an attempt to visualize the distribution of ccRCC subtypes across individual patients using reduceDimension function of R “Monocle” package with a Gaussian distribution [29]. The maximum number of components was set to 4, and the discriminative dimensionality reduction with trees was used. p value < 0.05 was considered statistically significant. The subtype landscape was visualized with the function plot cell trajectory with color-coded ccRCC subtypes.

2.8. External Data Validation. To test the repeatability of cluster analysis, we utilized two model-free means from R “MOVICS” package to validate the robustness of our cluster results [18]. First, MOVICS switches to the nearest template prediction (NTP) which can be flexibly applied to cross-platform, cross-species, and multiclass predictions without any optimization of analysis parameters. Then, we compared the survival outcome of the predicted cancer subtypes in external cohort and further checked the agreement between the predicted subtype and AJCC classification. In addition to NTP, MOVICS provides another model-free approach to predict subtypes, which first trains a partition around medoids (PAM) classifier in the discovery (training) cohort to predict the subtype for patients in the external validation (testing) cohort, and each sample in the validation cohort was assigned to a subtype label whose centroid had the highest Pearson correlation with the sample. Finally, the in-group proportion (IGP) statistic was performed to evaluate the similarity and reproducibility of the acquired subtypes between discovery and validation cohorts. In the end, we

used Kappa statistics to check the consistency between different prediction results.

2.9. Dimensionality Reduction and Risk Model Construction. Based on the multiomics classification, we obtained two subtypes, CS1 and CS2. In the previous analysis, the DEGs between the subtypes were obtained and the DEGs that were positively correlated with the signals of the CS1 and CS2 subtypes were divided into risk score signatures A and B, respectively [30]. Furthermore, the dimension reduction of the risk gene signatures A and B was conducted using the Boruta algorithm, and the Principle Component Analysis (PCA) algorithm was employed to draw principal component 1 as the score [31]. Finally, we ran a previously reported method to get the risk score of each patient: $\text{score} = \sum \text{PC1 A} - \sum \text{PC1 B}$ [32]. According to the median value of the patient's risk score, the patients were divided into a high-risk group and a low-risk group, and survival differences between the two groups was analyzed using R “survival” package.

2.10. Statistical Analysis. Spearman' or Pearson' correlation analysis was used to calculate the correlation coefficient between two variables. Wilcoxon rank-sum or t -test was applied to evaluate the difference when comparing between two continuous variables. Kruskal-Wallis or ANOVA test was introduced to compare difference among three or more groups. Also, a chi-squared or Fisher's test was applied to compare the difference in categorical variables. The nomogram was developed to assess individual outcome of ccRCC patients by using rms package. Calibration curves were calculated by calibrate function implemented in R “rms” package. Univariate Cox regression and multiple Cox regression analyses were performed to calculate the hazard ratios (HRs). The receiver operating characteristic (ROC) curves were plotted by R “timeROC” package. Area under the ROC curve (AUC) and Harrell's concordance index (C-index) were employed to evaluate the performance of the risk score in predicting OS and PFI. All operations are in R 4.0.3 (<https://www.r-project.org>). p values were two-sided, and p value < 0.05 was considered statistically significant. If multiple comparisons were involved, p values were adjusted using the Benjamini-Hochberg method. Visualization was achieved by R “ggplot2,” “ggpubr,” and “ComplexHeatmap” packages. In addition, this work has been preprinted in Biorxiv [33].

3. Results

3.1. Multiomics Landscapes of the Two ccRCC Subgroups. The most important parameter in any clustering study is the optimal number for cluster, where k needs to be small enough to reduce noise but on the other hand large enough to retain important information. Herein, we referred to Clustering Prediction Index (CPI) and Gaps-statistics to estimate the optimal number of clusters in this study, which turned out to be 2 [12, 34]. The number of data clusters was estimated using CPI and Gap statistics in combination with the aforementioned 10 algorithms. Both the correlation

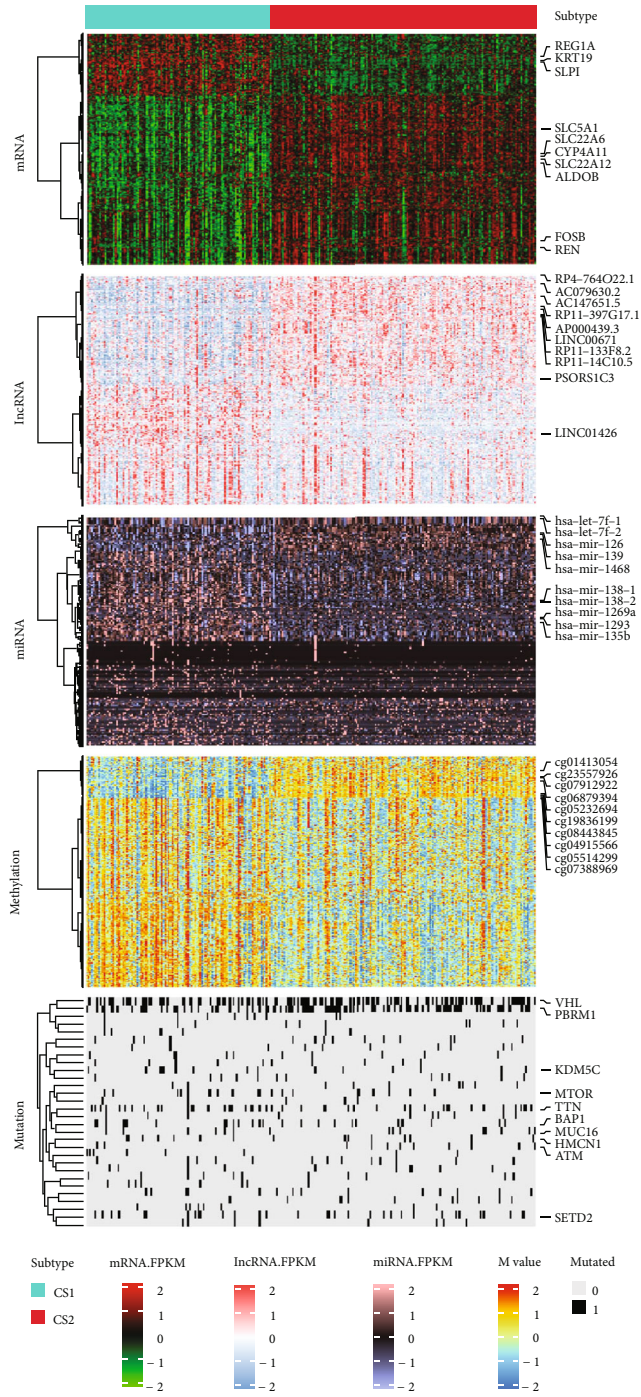
matrix heatmap and the silhouette map showed good intergroup heterogeneity and intragroup consistency between the two subgroups, indicating that our classification not only preserved the heterogeneity of different types of ccRCC but reduced unnecessary redundancy (Figure S2). The two types of ccRCC subtypes presented different landscapes in different omics including mRNA, lncRNA, miRNA, methylation, and somatic mutation levels. *REG1A*, *KRT19*, *SLPI*, *SLC5A1*, *SLC22A6*, *CYP4A11*, *SLC22A12*, *ALDOB*, *FOSB*, and *REN* were mRNAs that showed significant differences between the two subgroups; *RP4-764022.1*, *AC079630.2*, *AC147651.5*, *RP11-397G17.1*, *AP000439.3*, *LINC00671*, *RP11-133F8.2*, *RP11-14C10.5*, and *LINC01426* were differential lncRNAs; *has-let-7f-1*, *has-let-7f-2*, *has-mir-126*, *has-mir-139*, *has-mir-1468*, *has-mir-138-1*, *has-mir-138-2*, *has-mir-1269a*, *has-mir-1293*, and *has-mir-135b* were differential miRNAs; *cg01413054*, *cg23557926*, *cg07912922*, *cg06879394*, *cg05232694*, *cg19836199*, *cg08443845*, *cg04915566*, *cg05514299*, and *cg07388969* were differential methylation sites; *VHL*, *PBRM1*, *KDM5C*, *MTOR*, *TTN*, *BAP1*, *MUC16*, *HMCN1*, *ATM*, and *SETD2* were differential genes related to CNV (Figure 1(a)). The differences between the two tumor types were observed in the *t*-distributed stochastic neighbor embedding (*t*-SNE) dimensionality reduction distribution maps of each omics (Figure 1(b)). Prognostic analysis showed that the CS2 subtype was better than the CS1 subtype in terms of OS and PFI (Figure 1(c)). Finally, the classification of this study maintained a certain degree of consistency with the traditional AJCC classification and TNM stage classification was able to provide more accurate classification information for ccRCC patients (Figure 1(d) and Table 1).

3.2. Pseudotiming Analysis. This analysis cast individual patients into a manifold with sparse tree structures and defined the risk landscape of ccRCC. Consistent with our previously defined subtypes, many patients were segregated into two distinct clusters and presented different states (Figure 2(a)). The location of individual patients in the risk landscape represents the overall characteristics of the tumor immune microenvironment in the corresponding subtype. For instance, the high-risk subtype CS1 and low-risk subtype CS2 were distributed at the opposite end of the horizontal axis in the risk landscape (Figure 2(b)). Therefore, we hypothesized that the horizontal axis in the risk landscape represented the overall risk score. However, the vertical coordinate of the developmental trajectory appeared to be more complex and may reflect several characteristics. The risk landscape further revealed significant intracluster heterogeneity within each subtype. We observed that certain subtypes appeared to be more diverse and heterogeneous than others. For instance, CS1 could be further divided into three subgroups based on their location in the risk landscape, which showed different risk gene expression profiles in specific modules (Figure 2(c)). Interestingly, the three subgroups of patients in CS2 as stratified by the risk landscape were associated with distinct prognoses (Figures 2(e) and 2(f)). Similar results were obtained for CS2 (Figure 2(d)). The result of prognosis analysis showed two

distinct subtypes in the risk landscape within CS2, in which 2B showed poorer prognosis than other types in terms of OS and PFI (Figure 2(g) and 2(h)). These results indicate that our risk landscape analysis can provide a complementary value to previously identified risk subtypes.

3.3. Functional Enrichment Analysis. Through differential analysis, we identified a total of 950 DEGs, which are shown in volcano plot (Figure 3(a), Figure S3), including 907 upregulated genes and 43 downregulated genes, and biomarker signatures of each subtype are provided in Supplement Table 1. GO analysis showed that the differential genes between CS1 and CS2 subgroups were enriched in cornification, epidermis development, keratinocyte differentiation, skin development, and negative regulation of endopeptidase activity in Biological Process (BP) module. In Cellular Compartment (CC) module, DEGs were enriched in the collagen-containing extracellular matrix, intermediate filament, intermediate filament cytoskeleton, high-density lipoprotein particle, and plasma lipoprotein particle, as well as peptidase inhibitor, endopeptidase regulator, endopeptidase inhibitor, serine-type endopeptidase inhibitor, and peptidase regulator activities in Molecular Function (MF) module (Figure 3(b)). KEGG pathway analysis revealed differential pathways between CS1 and CS2 were mainly enriched in complement and coagulation cascades, ECM receptor interaction, protein digestion, and absorption (Figure 3(c)). Through GSEA analysis, we found that extracellular matrix organization, formation of the cornified envelope, keratinization, SLC-mediated transmembrane transport, and transport of small molecules were significantly enriched in the CS1 subgroup (Figure 3(d)). GSEA analysis showed that hypoxia, peroxisome, KRAS signaling pathway, myogenesis, and coagulation were significantly enriched in CS1, while G2M checkpoint, fatty acid metabolism, heme metabolism, and mitotic spindle were significantly enriched in CS2 (Figure 3(e)).

3.4. Comparison of Somatic Mutations. We had previously analyzed the differences between patients with different subtypes at the transcriptome level, and in the present study, we tried to analyze the differences between the subtypes at the genome level and further determine whether there was evidence of the disparity between CS1 and CS2 subgroups of patients at the genomic level. It was found that CS1 and CS2 presented different mutation landscapes. The top 20 most common mutation genes are shown in Figure 4(a). Among them, the mutation rate of *VHL*, *PBRM1*, and *ARID1A* in the CS1 subgroup was significantly lower than that in the CS2 group (41% vs. 54%; 26% vs. 54%; and 1% vs. 7%), while the mutation rate of *TTN*, *BAP1*, and *SETD2* in the CS1 subgroup was significantly higher than that in the CS2 subgroup (15% vs. 21%; 4% vs. 17%; and 10% vs. 16%) (Figure 4(b)). To be noted, *PBRM1* and *BAP1* were the most significantly different mutant genes between the two subgroups (Figure S4A). Some previous studies [35, 36] identified *VHL*, *PBRM1*, and *TTN* as the top three mutations in the two subtypes of patients, and they were involved in the occurrence and progression of ccRCC.



(a)

FIGURE 1: Continued.

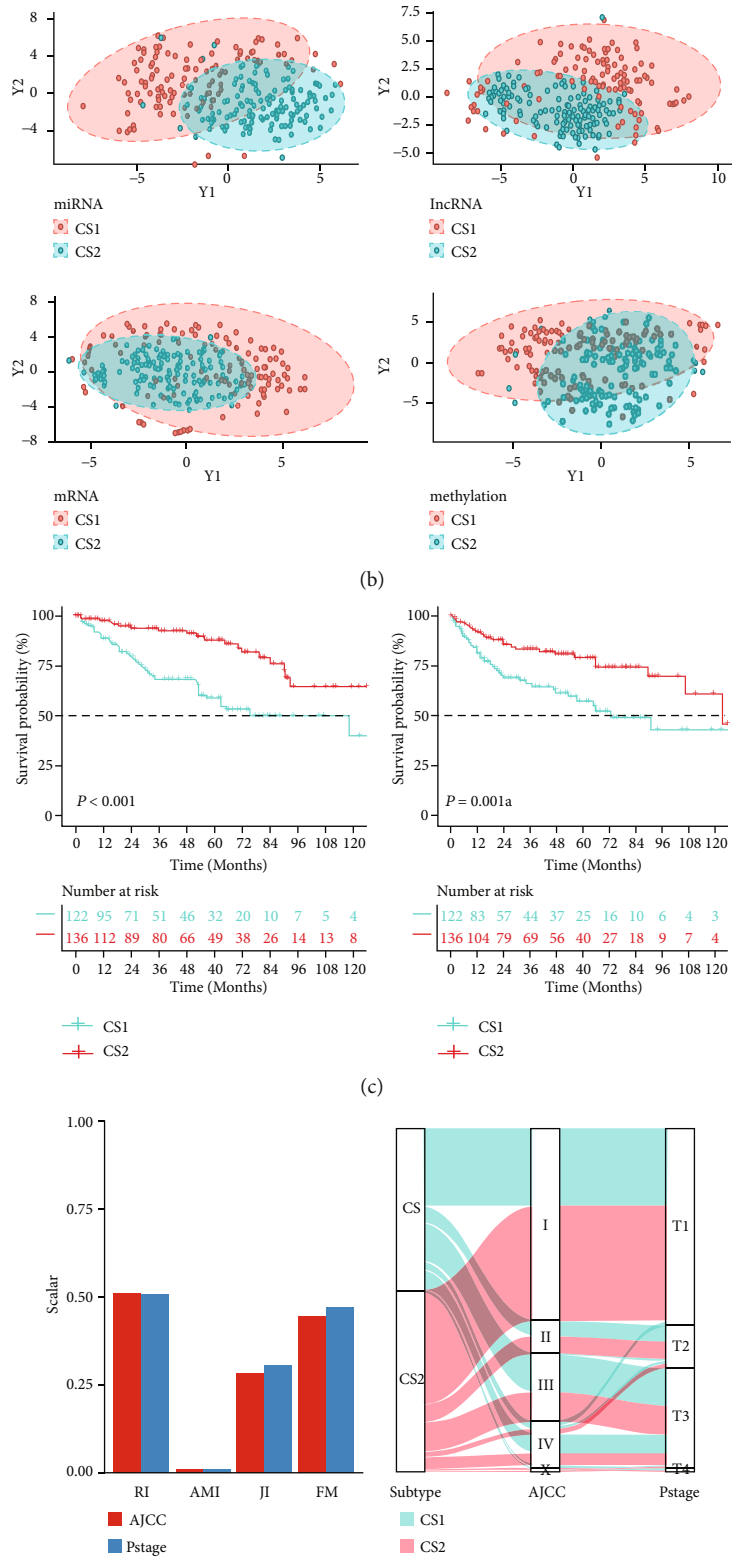


FIGURE 1: The landscape of multiomics differences in ccRCC. (a) Comprehensive heatmap of multiomics integrative clustering with dendrogram for samples. (b) Dot plot for two distinct clusters identified by the t-SNE algorithm based on mRNA, lncRNA, miRNA, and methylation profiles. (c) Kaplan-Meier survival curve of OS and PFI of the two identified subtypes of renal cancer in TCGA-KIRC cohort. (d) Agreement of the two identified subtypes of renal cancer with AJCC classification and pathological stage in TCGA-KIRC cohort. OS: overall survival; PFI: progression-free interval.

TABLE 1: Summary descriptive table by groups of “cluster.”.

	CS1 N = 122	CS2 N = 136	<i>p</i> .overall
fustat:			<0.001*
0	80 (65.6%)	117 (86.0%)	
1	42 (34.4%)	19 (14.0%)	
futime	1212 (984)	1506 (1131)	0.026
AJCC:			0.083
I	58 (47.5%)	86 (63.2%)	
II	12 (9.84%)	13 (9.56%)	
III	29 (23.8%)	22 (16.2%)	
IV	22 (18.0%)	14 (10.3%)	
X	1 (0.82%)	1 (0.74%)	
Age	60.7 (11.7)	60.8 (12.2)	0.970
PFI:			0.005
0	79 (64.8%)	110 (80.9%)	
1	43 (35.2%)	26 (19.1%)	
PFI.time	1038 (970)	1290 (1041)	0.045
Gender:			0.214
FEMALE	41 (33.6%)	57 (41.9%)	
MALE	81 (66.4%)	79 (58.1%)	
Grade:			<0.001*
G1	2 (1.64%)	7 (5.22%)	
G2	50 (41.0%)	68 (50.7%)	
G3	41 (33.6%)	51 (38.1%)	
G4	28 (23.0%)	8 (5.97%)	
GX	1 (0.82%)	0 (0.00%)	
Laterality:			0.658
Bilateral	1 (0.82%)	0 (0.00%)	
Left	55 (45.1%)	59 (43.4%)	
Right	66 (54.1%)	77 (56.6%)	

**p* value < 0.05 is considered statistically significant.

However, it was found in the present study that the mutation frequencies of the above genes were not consistent between the two subgroups of patients, suggesting that the classic ccRCC mutations play totally different roles in the CS1 and CS2 subtypes of ccRCC. For instance, PBRM1 is the second largest mutation gene in renal cancer, while the mutation event of PBRM1 in the two renal cancer subtypes described herein presented inconsistent survival benefits for the patients. PBRM1 mutation was identified as a protective factor in CS2 subgroup (Figure 4(c)). Similarly, the mutation sites were not consistent between the two subgroups of ccRCC (Figure 4(b)). Next, we investigated the cooccurring and exclusive mutations of the top 25 most frequently mutated genes using the CoMet algorithm and found that there existed some specific mutations in CS1 subtype such as PBRM1-VHL, PBRM1-NOTCH2, and PBRM1-KMT2C, and also there were specific mutations such as ARID1A-ATRX5, SETD2-THSD7B, and ARID1A-DNAH9 in the CS2 group, suggesting that they probably played a redundant role in the same pathway and harbored the selective advantages between them to keep more than

one copy of the mutations (Figure 4(d)). However, the ratio between transversion (Tv) and transition (Ti) in all SNVs was approximately 1:1 and remained stable in both cohorts (Figure S4B). The distribution of variant allele frequency (VAF) in different genes was found to be inconsistent between the two subtypes. Among them, the genes with the highest VAF in the CS1 group were KDM5C, VHL, and PBRM1, while the first three VAF genes in the CS2 group were VHL, PBRM1, and ATM (Figure S4C). Finally, there were differences in the sudden change landscape between the two subgroups. The different gene mutations showed distinct clinical outcomes for patients in the CS1 and CS2 subgroups (Figure S3D). For example, PBRM1 was a typical example to demonstrate the different mutation spots between the two cohorts and the plausible chain reaction of the differences took place in prognostic impact (Figure S4E). At the same time, based on the differences in CNV data between the two subgroups, we explored possible treatments based on the mutation data of the two subtypes using the drug-Interactions function in the maftools package and the DGIdb database (<https://www.dgiddb.org/>). Among them, the possible therapeutic targets in CS1 group mainly included BAP1, COL24A1, DST, HMCN1, and KDM5C. The potential therapeutic targets in the CS2 group mainly included ATM, COL6A6, DST, ERBB4, and HMCN1 (Figure S4F).

3.5. CNV Differences between CS1 and CS2 Subgroups of ccRCC. CNV differences were compared between patients with different subtypes of KIRC. The incidence of the rate of copy number variation in CS2 was higher than that in CS1 (Figure S5A). GISTIC2.0 was used to define recurrently amplified and deleted regions in the two subgroups (Supplement Table 2). The results showed that the two subtypes had frequent CNVs in the region containing oncogenes, tumor suppressor genes (e.g., VHL and PBRM1) and metabolism regulator (e.g., COL4A3 and COL4A4), indicating that CNVs might play a significant role in the tumorigenesis and progression of ccRCC. We found recurrent focal CNVs in CS1 including amplifications containing 5q35.1 (ADRA1B) and 3q26.33 (PIK3CA) and deletion of 9p21.3 (CDKN2A) and 3p21.31 (KIF9). Recurring focal CNVs in CS2 included amplifications of 5q33.2 (KIF4B) and deletion of 3p11.1 (HTR1F). These specific CNVs might contribute to the formation of the two subtypes (Figure S5B).

3.6. Carcinogenic Pathway Differences between CS1 and CS2 Subtypes of ccRCC. Through CNV data, we found that oncopathways were not completely consistent between the two subgroups. Among them, the most critical carcinogenic pathways in CS1 subgroup were RTK_Ras, Notch, PI3K, Hippo, WNT, TP53, MYC, NRF2, and Cell Cycle vs. RTK_Ras, Notch, Hippo, WNT, PI3K, TP53, MYC, Cell Cycle, NRF2, TGF_β, and other pathways in the CS2 subgroup (Figures 5(a) and 5(b)). At the same time, GSVA analysis was exploited to verify the difference in oncopathways between the two subgroups (Figure S6).

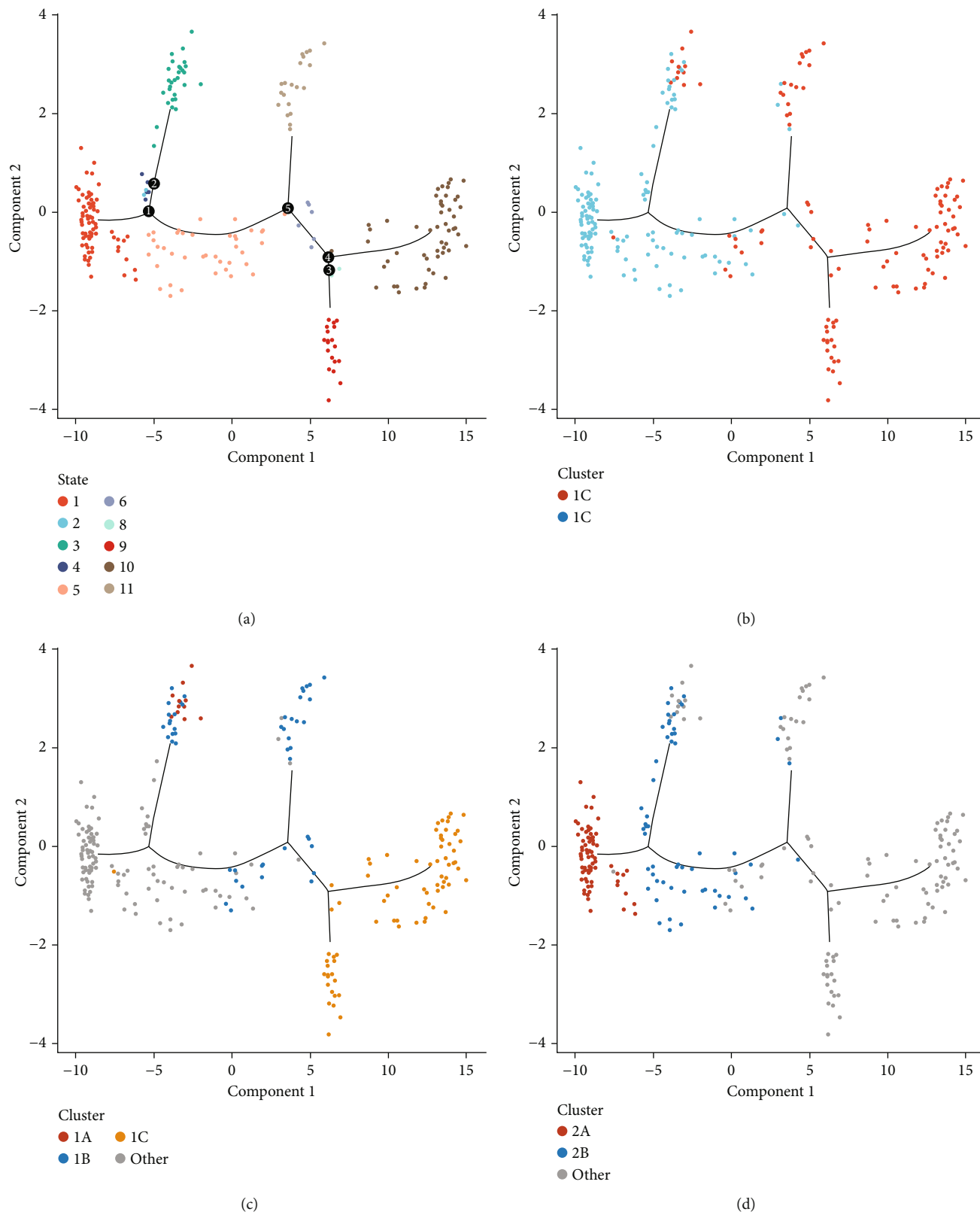


FIGURE 2: Continued.

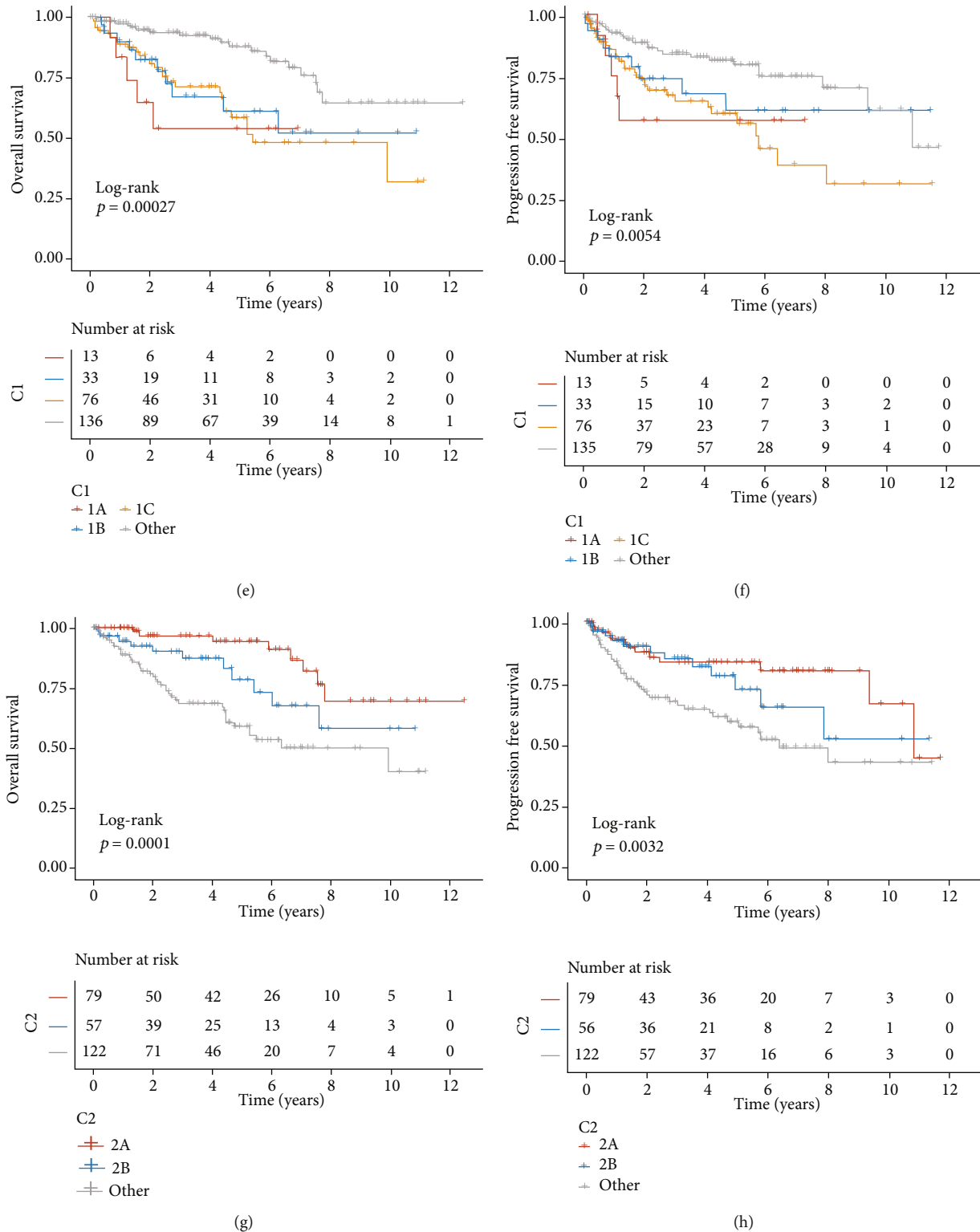
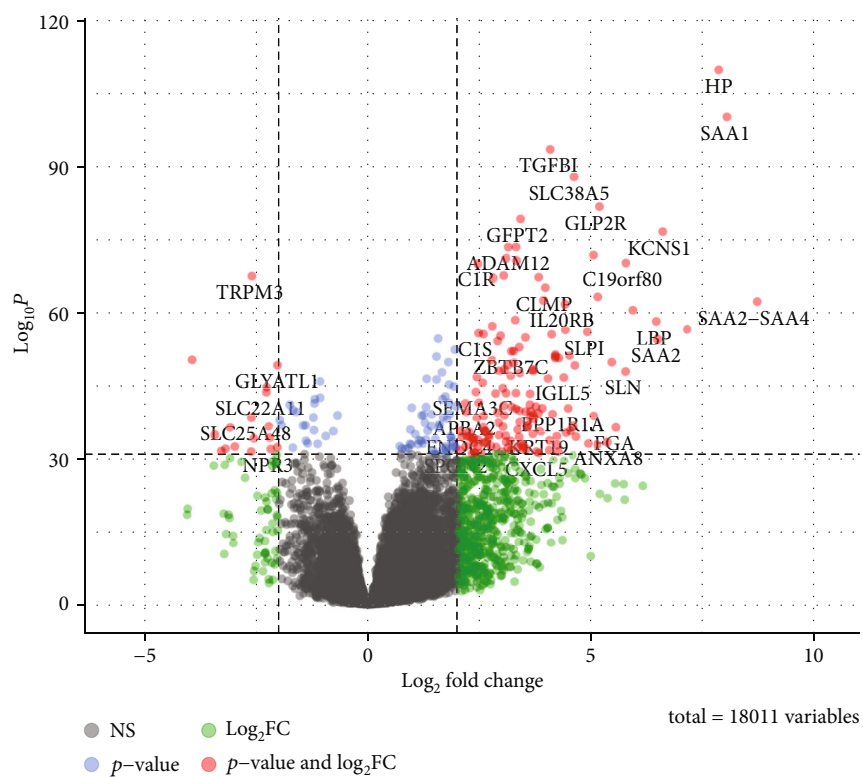
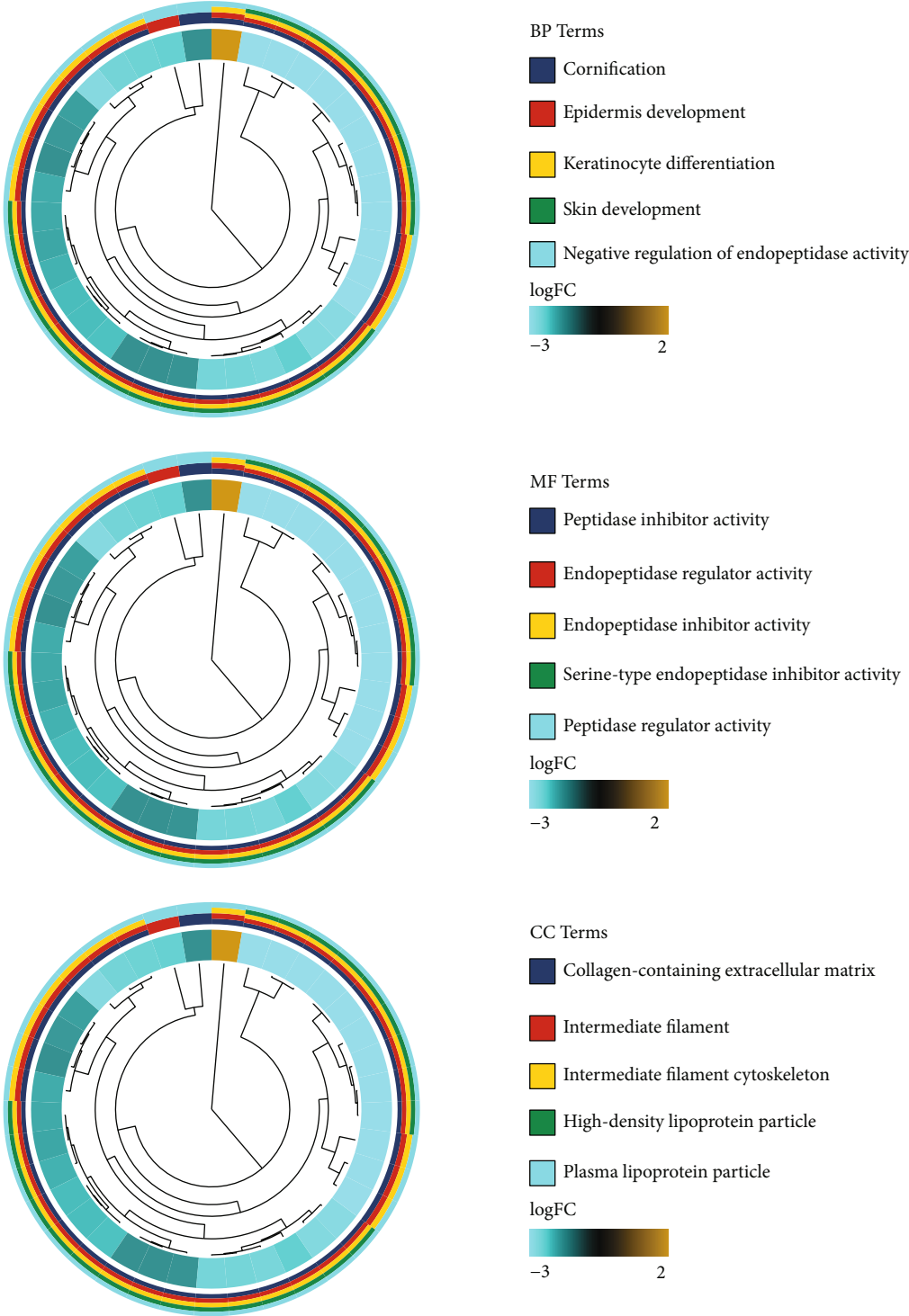


FIGURE 2: The risk landscape of ccRCC and the intracluster heterogeneity within subtypes. (a) The subtype of ccRCC clustered by state. (b) The risk-related mRNA landscape of ccRCC: each point represents a patient with colors corresponding to the subtype defined previously. (c) Patients of CS1 subtype could be further stratified into three subgroups based on their location in the risk-related mRNA landscape. (d) Patients of CS2 subtype could be further stratified into two subgroups based on their location in the risk related mRNA landscape. (e, f) The three subgroups of patients in CS1 as stratified by the risk-related mRNA landscape were associated with distinct prognoses of OS and PFI. Log-rank p value was calculated among subgroup stratification. (g, h) The two subgroups of patients in CS2 as stratified by the risk-related mRNA landscape were associated with distinct prognoses of OS and PFI. Log-rank p value was calculated among subgroup stratification.



(a)

FIGURE 3: Continued.



(b)

FIGURE 3: Continued.

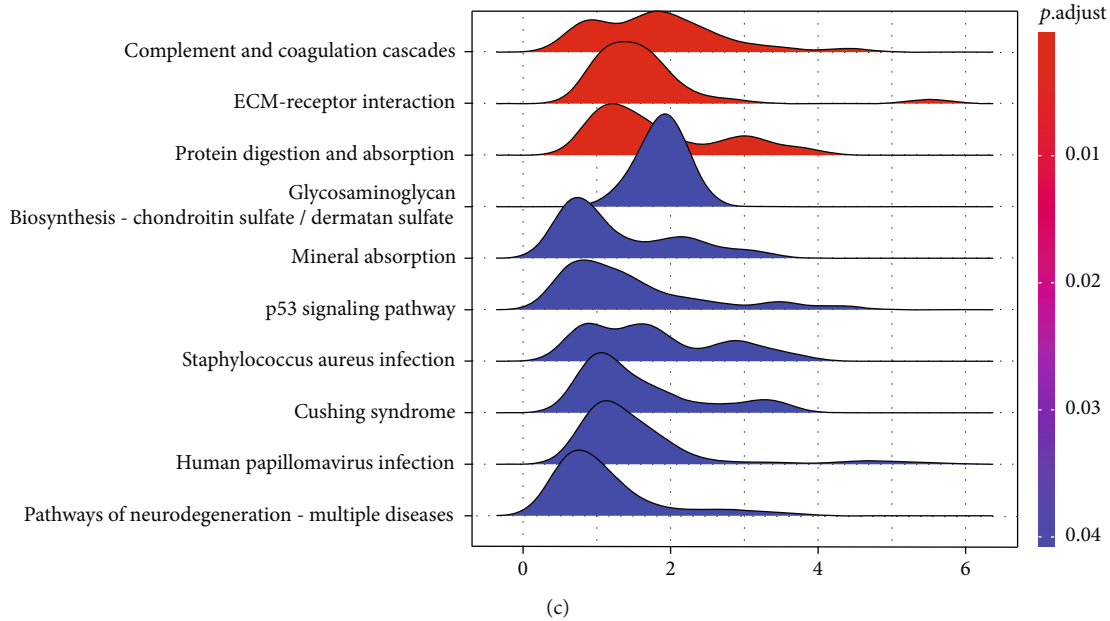


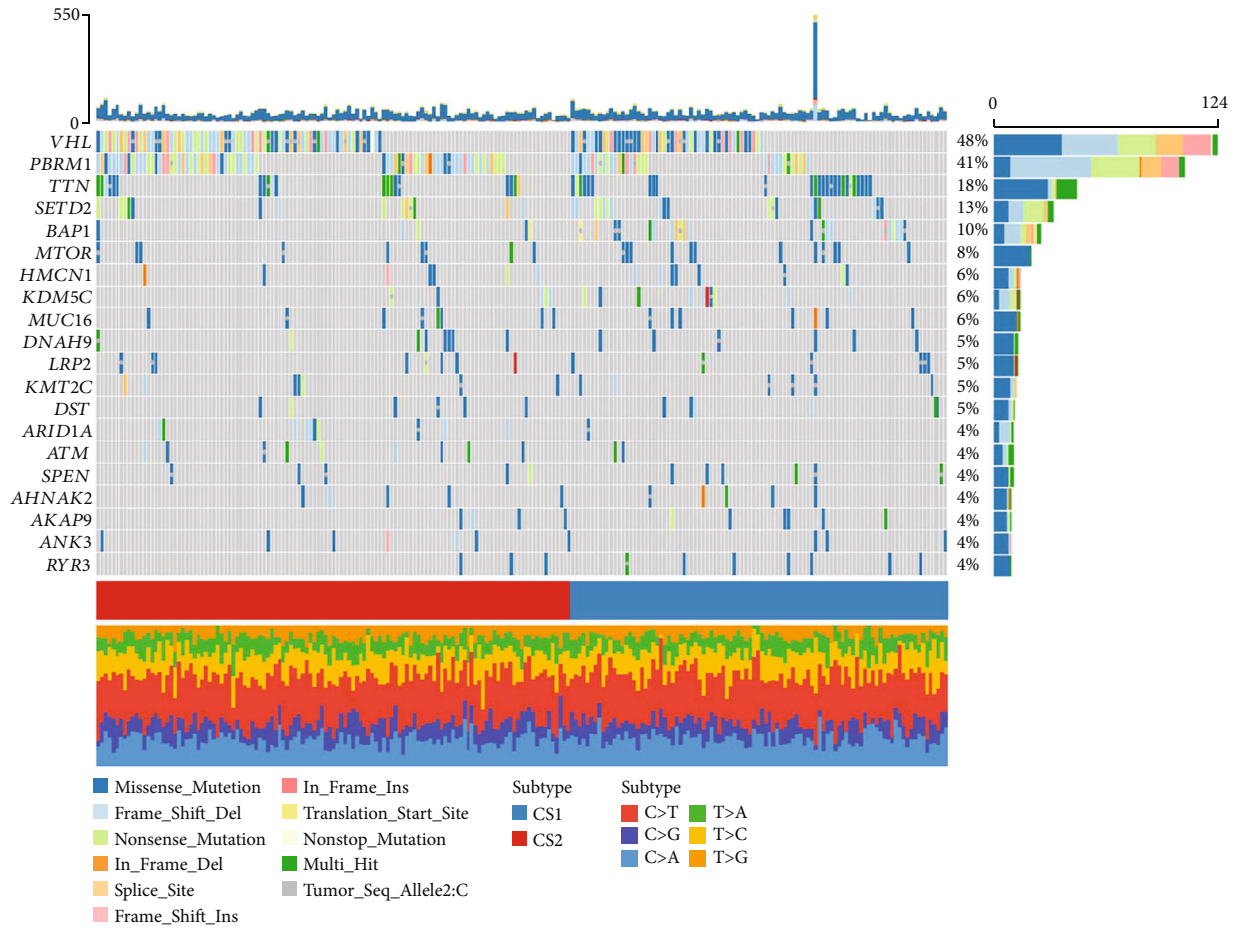
FIGURE 3: Functional enrichment analysis of DEGs between CS1 and CS2 subtypes. (a) Volcano map of different expression genes. (b) GO enrichment analysis. (c) KEGG pathway analysis. (d) Gene set enrichment analysis (GSEA): hallmarks in the subgroups. (e) GSVA: differences in enrichment between the subgroups.

3.7. Immune Infiltration Analysis. The result of our immune infiltration analysis showed a significant difference in the immune landscape between the CS1 and CS2 subgroups. The algorithms showed fewer infiltrating cells in the CS2 subgroup as compared with the CS1 subgroup (Figure 6(a)), and the expression of marker genes related to the immune micro-environment was also significantly different between the two groups (Figure S5). Among them, the chemokine family, costimulatory factors, coinhibitory molecules, interferon, and MHC molecules were included in the CS1 subgroup. GSVA analysis showed that differences in specific immune cell infiltration between the two groups were calculated, and the result showed a higher degree of multiple immune cell infiltration in the CS1 group than in the CS2 group (Figure 6(b)). GSVA analysis showed that more significantly immunity-related pathways were activated in CS1 subtype as compared with CS2 subtype, including interleukins, cytokines, B cell functions, and T cell functions (Figure 6(c)). However, there was no significant difference in tumor mutation burden (TMB) between the two subgroups (Figure 6(d)). However, the TIDE algorithm showed a significantly higher immune check inhibitor response rate in CS2 patients compared with CS1 patients ($p < 0.01$) (Figure 6(e)). Finally, with respect to differentially expressed common immune reverse regulator makers between the two groups, the expression of PD1 in the CS2 subgroup was higher than that in the subgroup ($p = 0.083$), while PDL1 in the CS2 subgroup was significantly lower than that in the CS1 subgroup ($p < 0.01$). But other common immunosuppressive markers including TIGIT, LAG3, CTLA4, and KLRB1 were not significantly different between the two subgroups (Figure S7A). The ESTIMATE results showed that the immune scores and

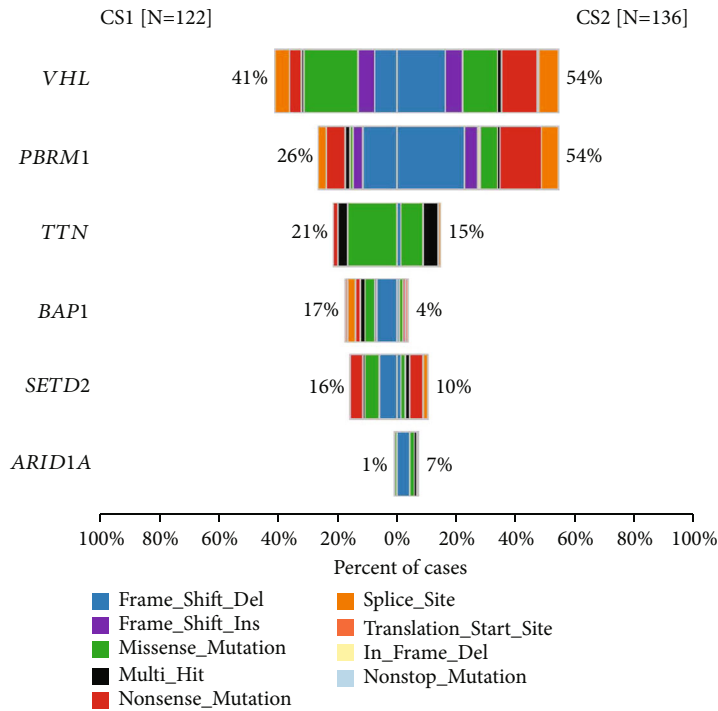
estimate scores in the CS1 group were lower than those in the CS2 group (Figure S7B).

3.8. Drug Sensitivity Analysis. In this section, we compared the sensitivity to Saracatinib, Lisitinib, Imatinib, Gefitinib, Erlotinib, Dasatinib, Crizobinib, and Sunitinib between the two subgroups of ccRCC and found no significant difference in sensitivity to Sunitinib and Erlotinib between the CS1 and CS2 subgroups, but patients in the CS2 subgroup were more sensitive to Sorafenib, Lisitinib, Gefitinib, and Dasatinib than those in the CS1 subgroup (Figure 7A). The previous prognosis analysis showed that the prognosis of the CS1 subgroup was poorer than that of the CS2 subgroup. Therefore, based on the GDSC database, we explored whether there were any other sensitive drugs for CS1 subgroup and finally discovered 53 kinds of small-molecule drugs that could be used as potential drugs for the treatment of CS1 patients (Supplement Table 3). The top 10 molecular drugs with the most significant differences were CGP.082996, CMK, JNJ.26854164, ZM.447439, GNF.2, LAM. A13, RO.3306, WO2009093972, CGP.60474, and VX.680 (Figure 7(b)). The detailed structures of the above molecules are provided in Figure S8.

3.9. External Data Verification for the Subtyping Method of ccRCC. In this part, we aimed to predict the possible subtypes of each sample in the external dataset. In most cases, this multiclassification issue is tricky for the simple reason that it is nearly impossible to entirely match recognized markers in the external cohort, so the reliability of model-based predicting methods might be low. To verify whether the subtype-specific signature could accurately classify renal cancer into CS1 and CS2 subtypes in an external dataset, we



(a)



(b)

FIGURE 4: Continued.

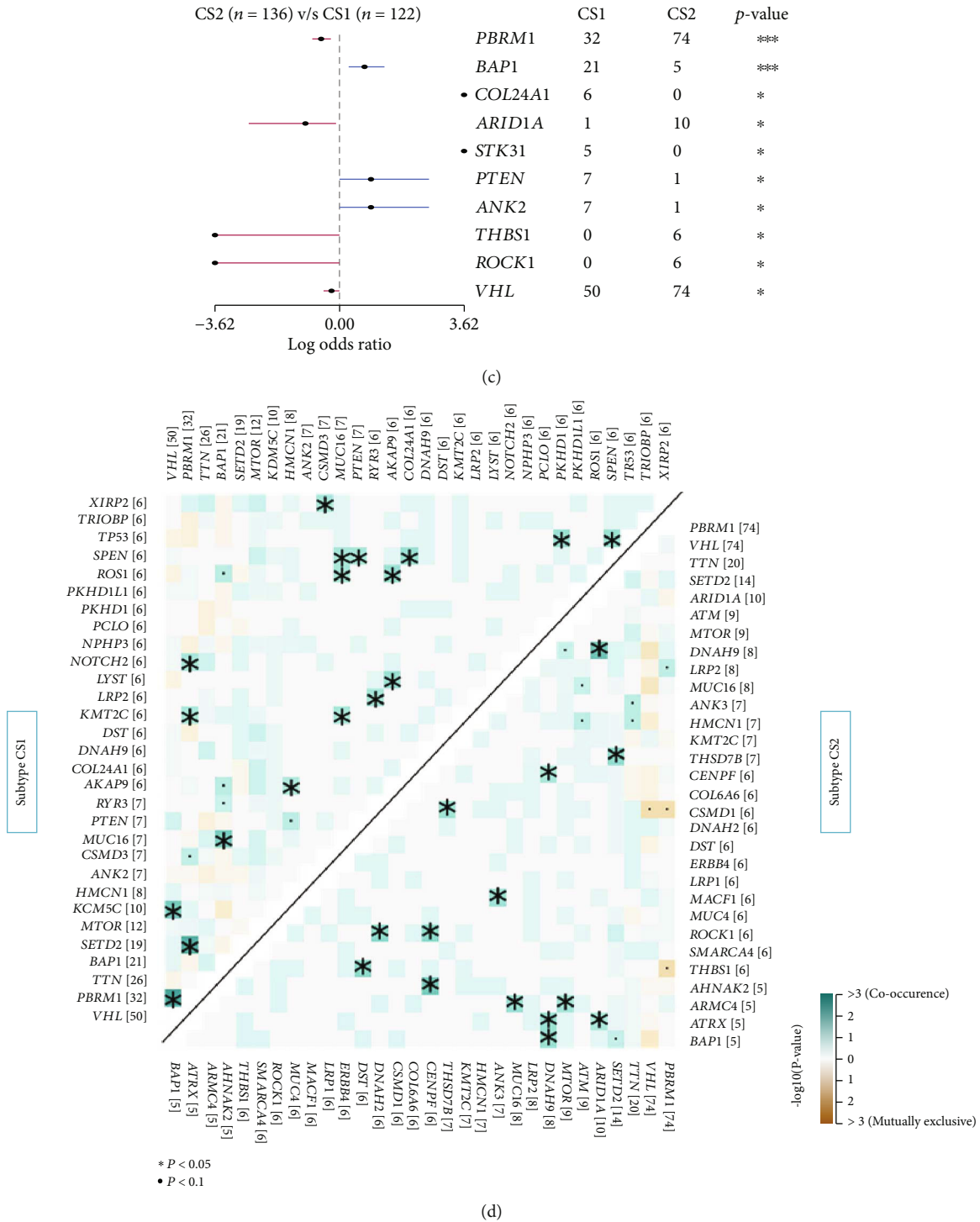
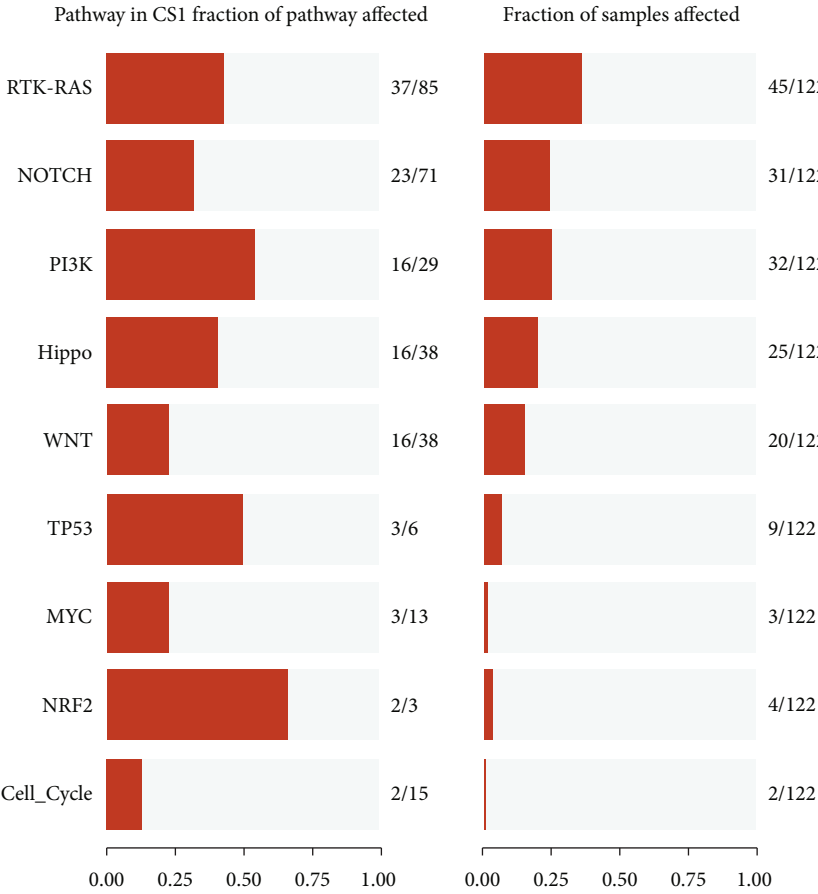


FIGURE 4: Landscape of somatic mutation in the CS1 and CS2 subgroups. (a, b) Waterfall plot shows the mutation distribution of the top 20 most frequently mutated genes. The central panel shows the type of mutation in each ccRCC sample. The upper panel shows the mutation frequency of each ccRCC sample. The bar plots on the left show the frequency and mutation type of genes mutated in 258 ccRCC samples. The lower part shows different subgroups including CS1 and CS2. (c) Forest plot displays the top 6 most significantly differentially mutated genes between the two subgroups. (d) Heatmap illustrates the mutually cooccurring and exclusive mutations of the top 25 frequently mutated genes. The color and symbol in each cell represent the statistical significance of the exclusivity or cooccurrence for each pair of genes.

used the remaining 273 renal cancer datasets in TCGA as a validation set to evaluate the robustness of the subtype signature classifier. It was found that the subtype signature could

accurately verify the outer data into two subgroups (Figure 8(a)). The classification of the outer dataset was consistent with the classic AJCC classification in some extent



(a)

FIGURE 5: Continued.

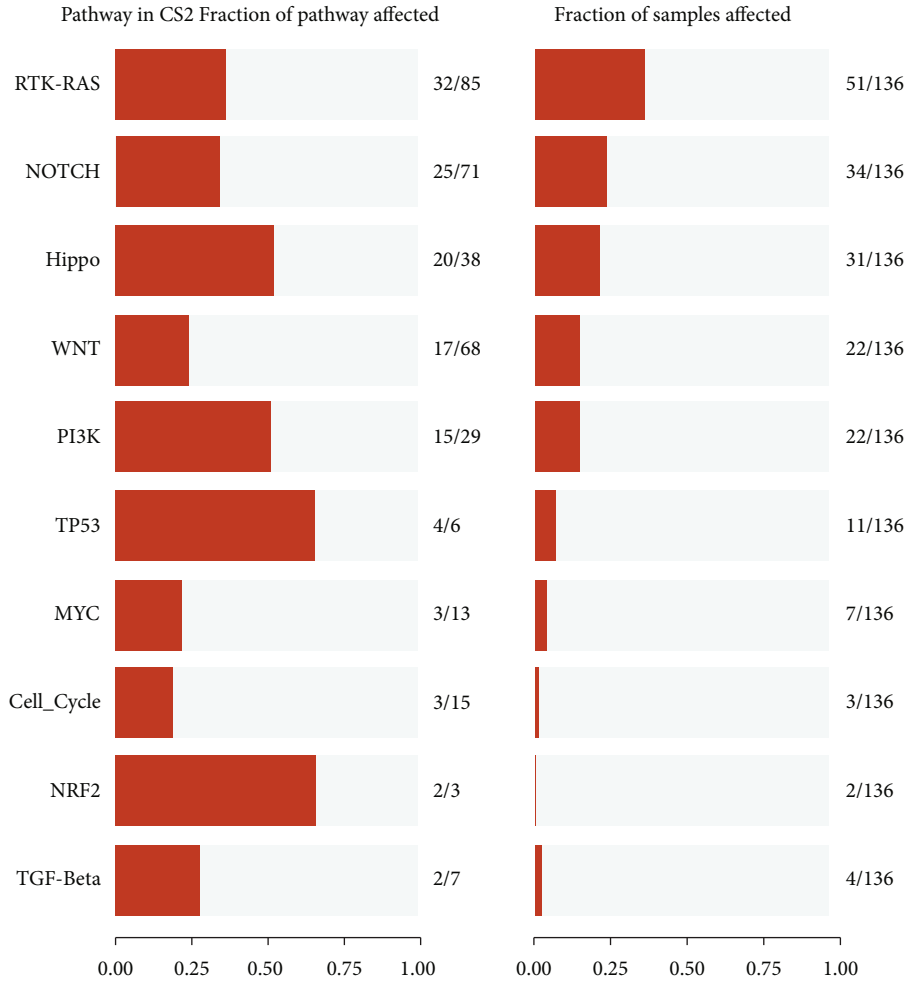
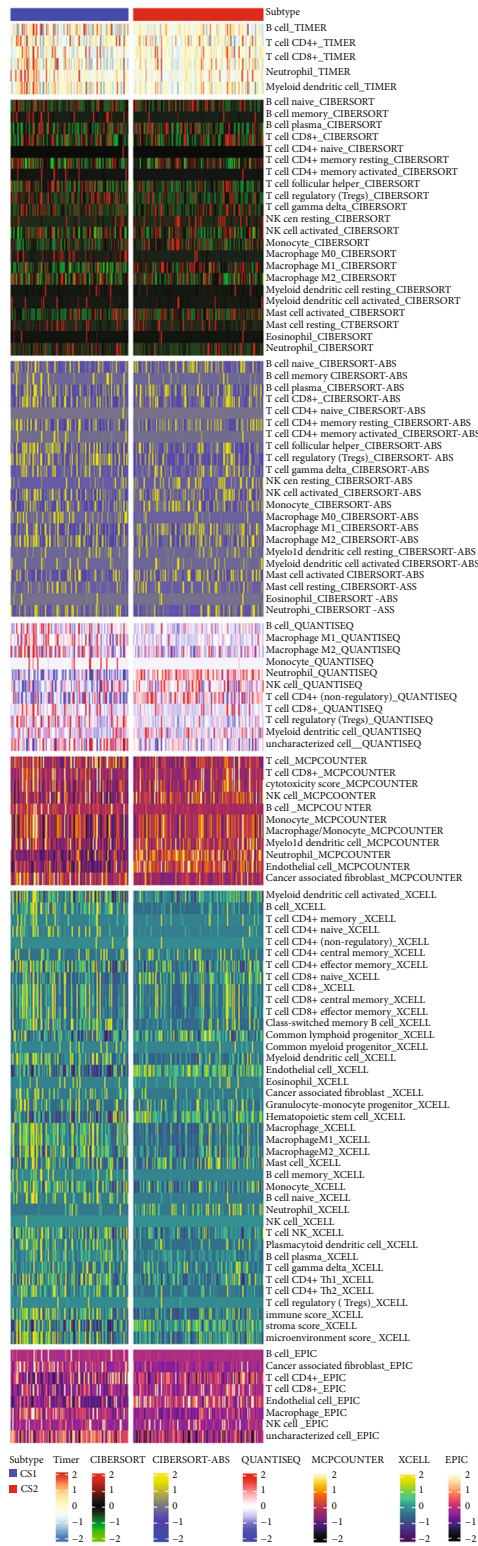


FIGURE 5: CNV differences between subgroups. (a, b) Bar plot of fraction genome altered in the two identified subtypes of renal cancer in TCGA-KIRC cohort.

(Figure 8(b)). Among the validation datasets, OS and PFI in the CS2 group were poorer than those in the CS1 group ($p < 0.01$) (Figure 8(c)). Finally, we used Kappa statistics for consistency evaluation, and we compared the consistency between NTP-predicted subtype and COMIC in discovery TCGA-KIRC (scores = 0.743, $p < 0.001$), between PAM-predicted subtype and COMIC in discovery TCGA-KIRC (scores = 0.899, $p < 0.001$), and between NTP and PAM-predicted subtype in validation ICGC-KIRC (scores = 0.75, $p < 0.001$) (Figure 8(d)). In addition, we obtained integrated data of Japanese ccRCC cohort and valid verification of our subtyping method was done in this cohort as well (Figure S9). The results demonstrated that our typing research was robust and reliable.

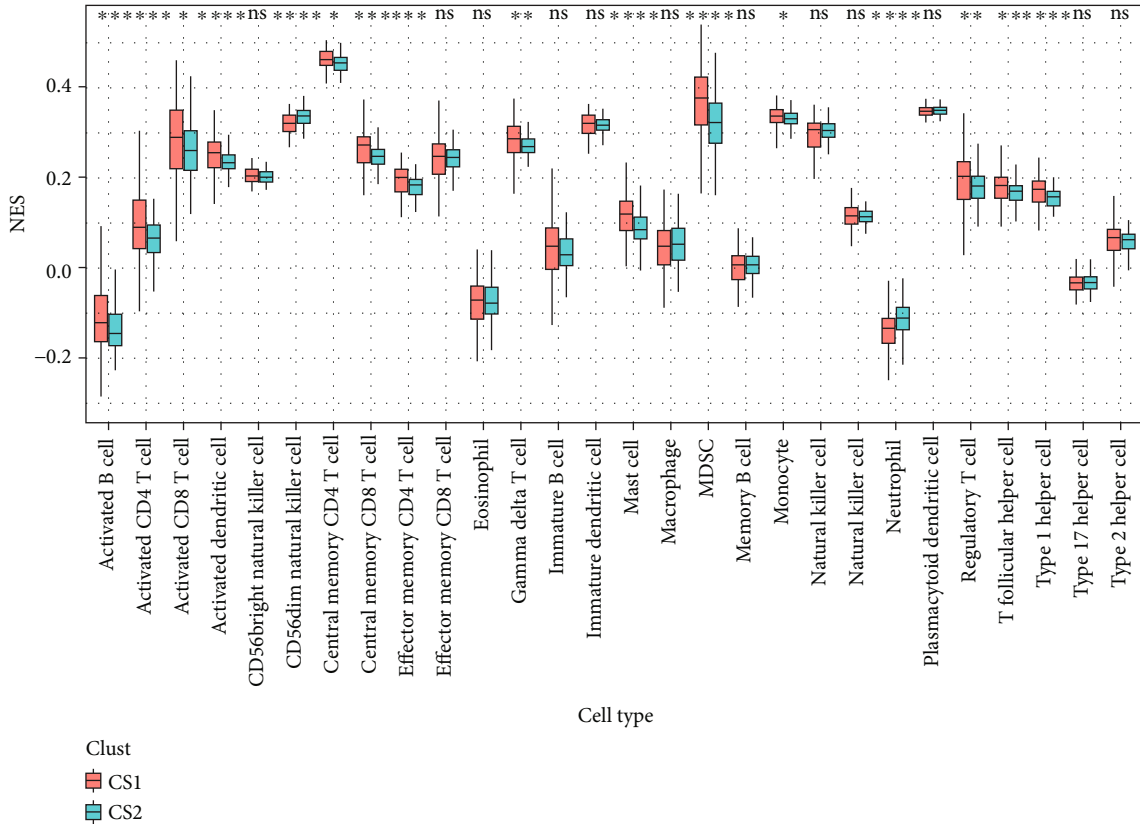
3.10. Dimensionality Reduction Clustering and Risk Model. To clarify the subtype signatures of the CS1 and CS2 subtypes, we used the DESeq2 package to perform differential analysis to determine the transcriptome differences between these subtypes, and then, we performed unsupervised clustering based on the differential genes and set the genes that

were positively related to CS1 as signature B, and the genes positively related to CS2 as signature A (Figure S10A). Next, to reduce noise or redundant genes, we used the Boruta algorithm to reduce the dimensionality of the signature genes of CS1 and CS2 subtypes (Figure S10B). At the same time, we chose the most critical signatures according to the degree of contribution (Figure S10C). Furthermore, a method similar to Gene expression grade index was applied to define the risk score of each patient [37]. Based on the risk scores obtained, the patients were divided into a high-risk group and a low-risk group (Figure S10D). Prognosis analysis showed that the prognosis of the patients in the high-risk group was significantly poorer than that of the patients in the low-risk group in terms of OS and PFI ($p < 0.01$), and the ROC curve showed that the risk score could more accurately assess the long-term survival rate of the patients (Figures 9(a)–9(h)). To develop a clinically relevant quantitative method for predicting the probability of patient mortality, we constructed a nomogram by integrating the risk score and other clinical prognostic factors (Figure S10E). The calibration plot indicated that the

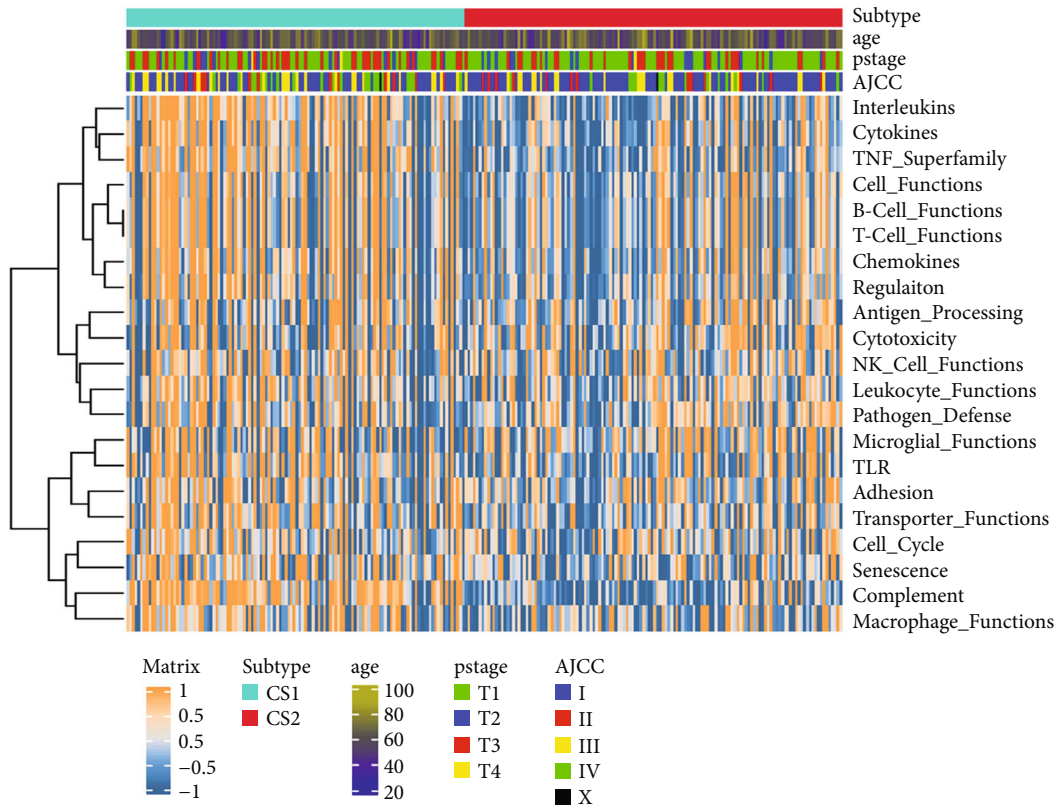


(a)

FIGURE 6: Continued.



(b)



(c)

FIGURE 6: Continued.

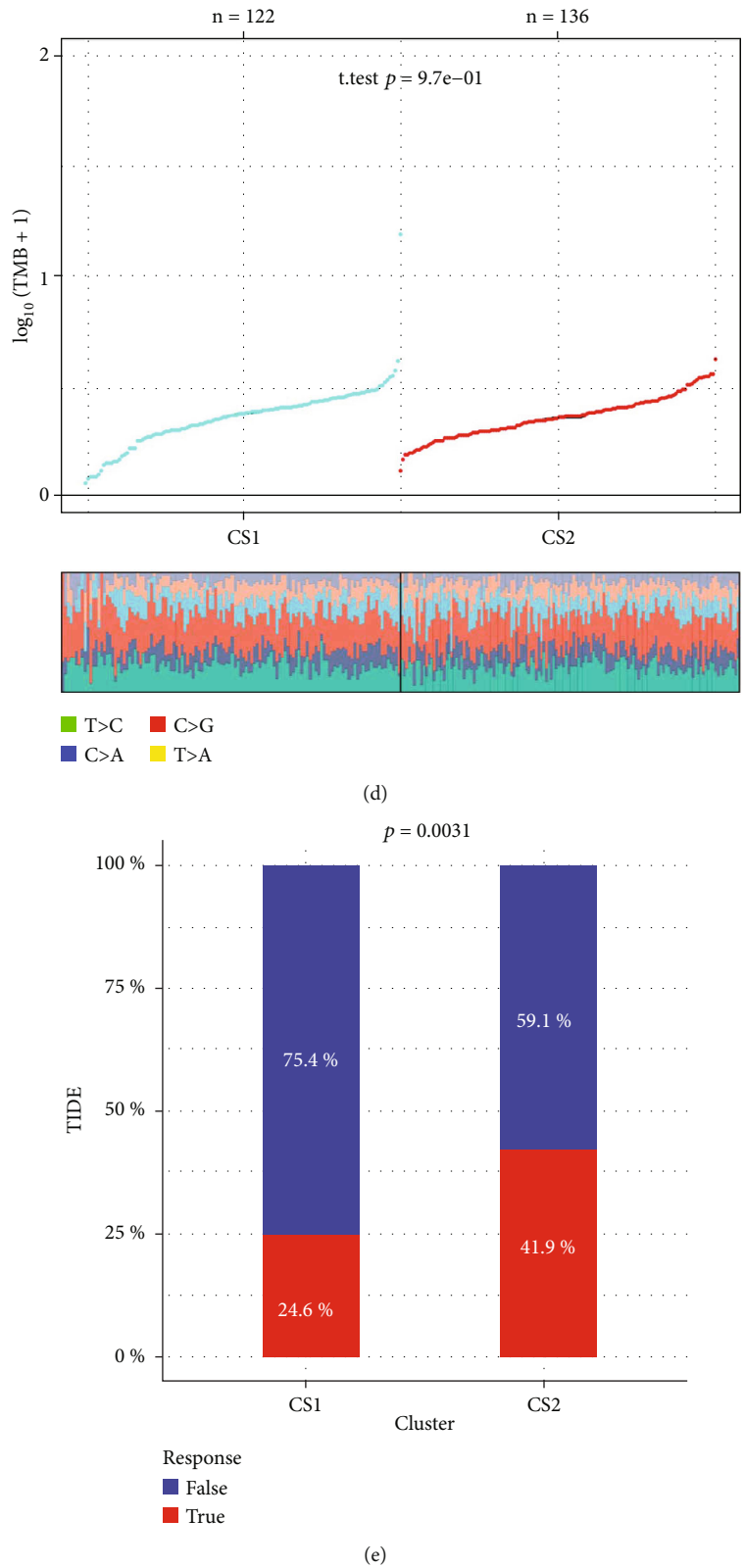
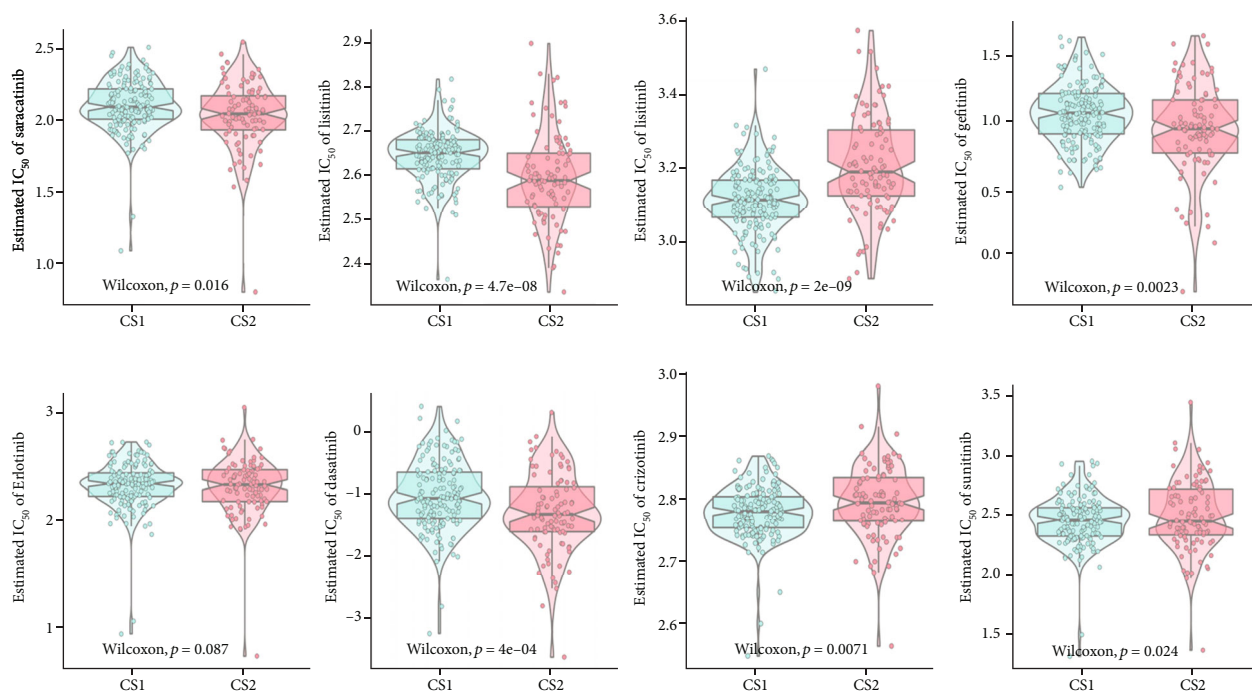
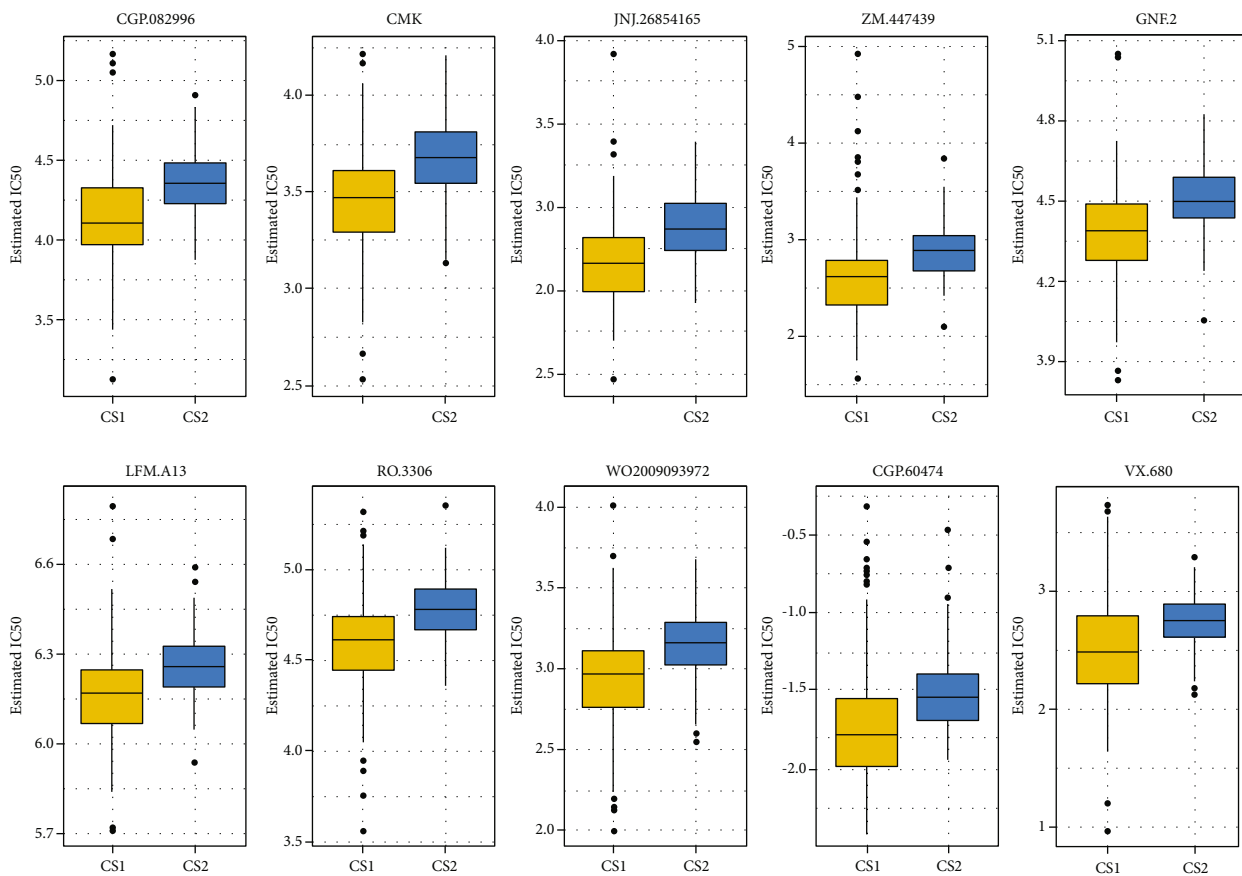


FIGURE 6: Identification of immune status. (a) Heatmap of tumor-related infiltrating immune cells based on TIMER, CIBERSORT, CIBERSORT-ABS, QUANTISEQ, MCPcounter, XCELL, and EPIC algorithms in the CS1 and CS2 subgroups. (b) Different normalized enrichment scores of immune cells between the subgroups. (c) Heatmap of different immune-related pathway enrichment scores between the subgroups. (d) Difference in TMB between the CS1 and CS2 subgroups. (e) Difference in response of immune check point inhibitor treatment based on the TIDE algorithm.

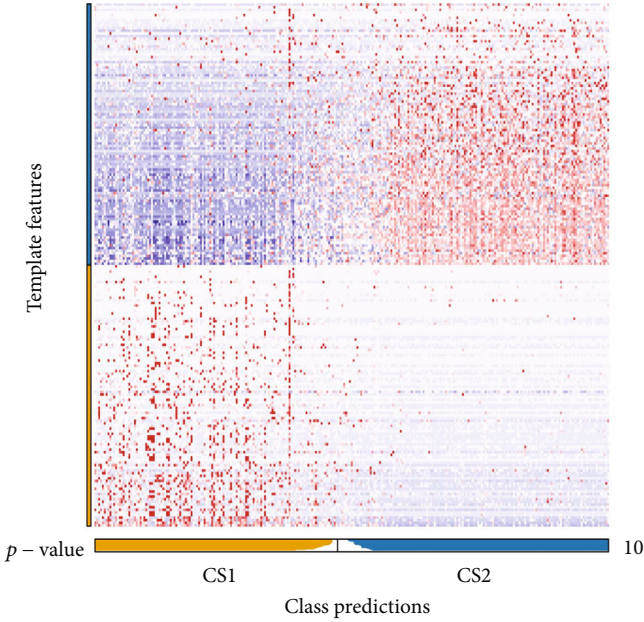


(a)

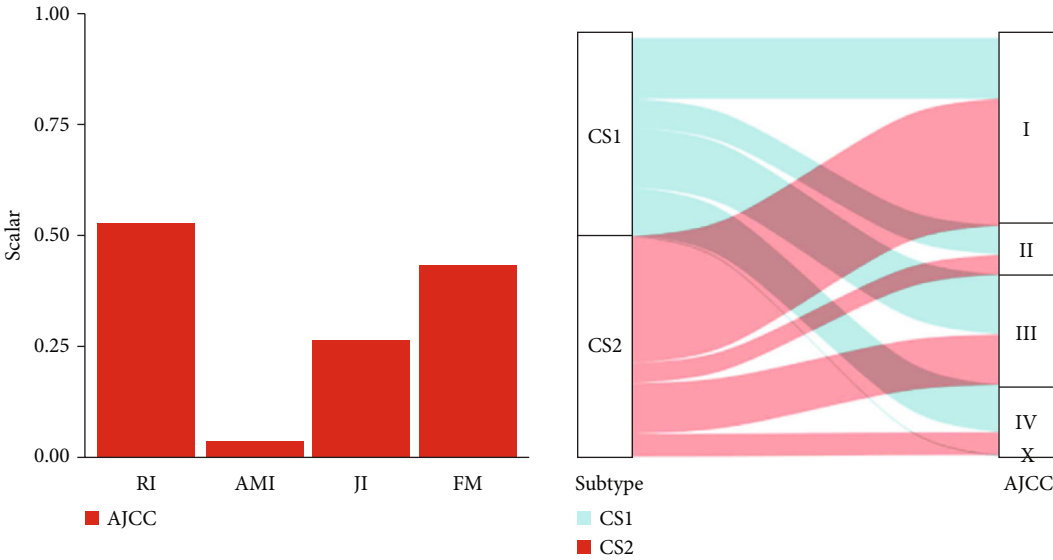


(b)

FIGURE 7: Difference of drug sensitivity. (a) Differences in estimated IC_{50} of the molecular targeted drugs (Saracatinib, Lisitinib, Imatinib, Gefitinib, Erlotinib, Dasatinib, Crizobininib, and Sunitinib) between the CS1 and CS2 subgroups. (b) Differences in response to 10 chemotherapy drugs between the two prognostic subtypes.



(a)



(b)

FIGURE 8: Continued.

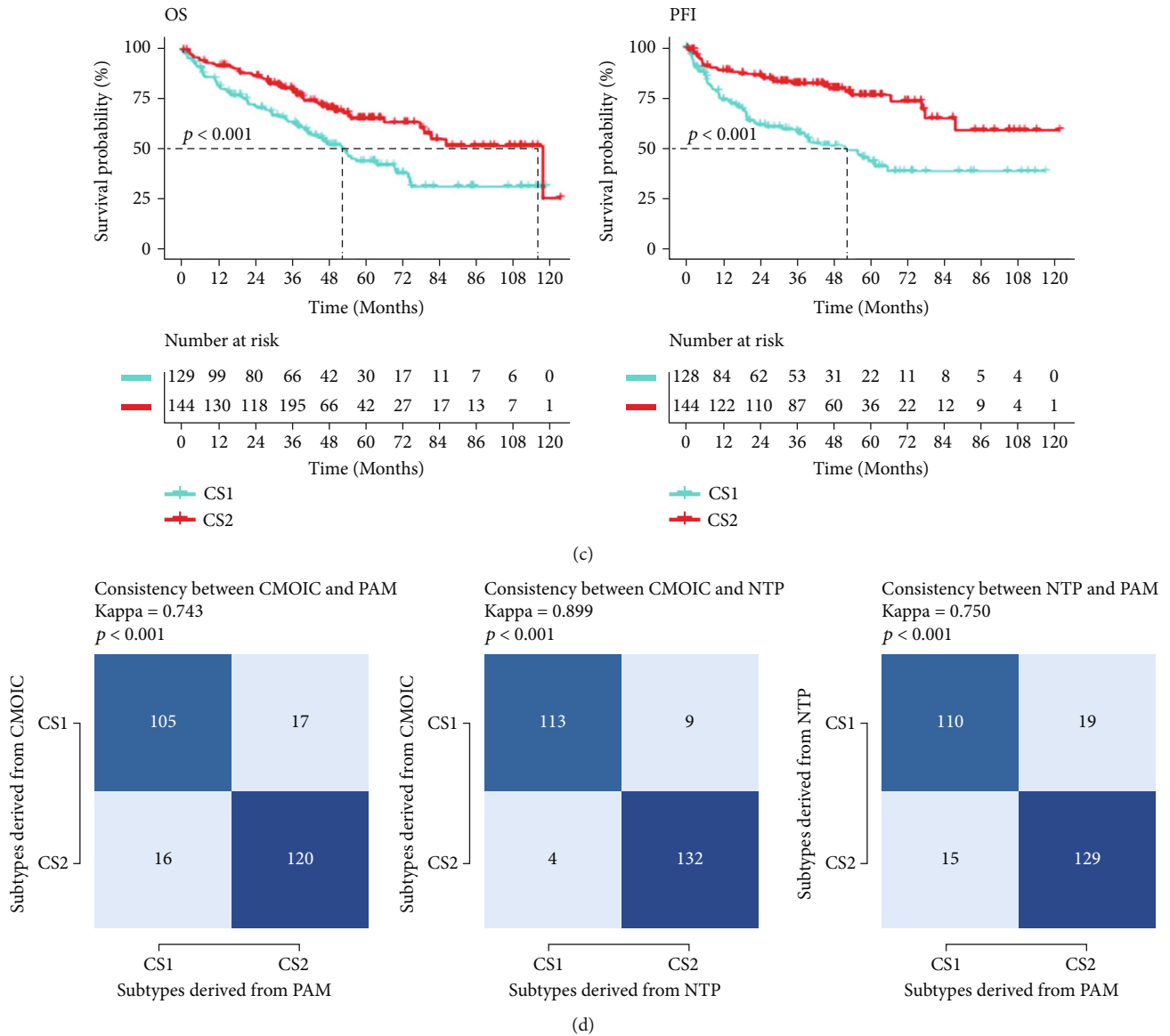


FIGURE 8: Verification in external dataset. (a) Heatmap of NTP in outer KIRC cohort using subtype-specific upregulated biomarkers identified from the TCGA-KIRC cohort. (b) Agreement of the two predicted subtypes of renal cancer with AJCC classification in outer KIRC cohort. (c) Kaplan-Meier survival curve of the two predicted subtypes of renal cancer in outer KIRC cohort. (d) Consistency heatmap using Kappa statistics.

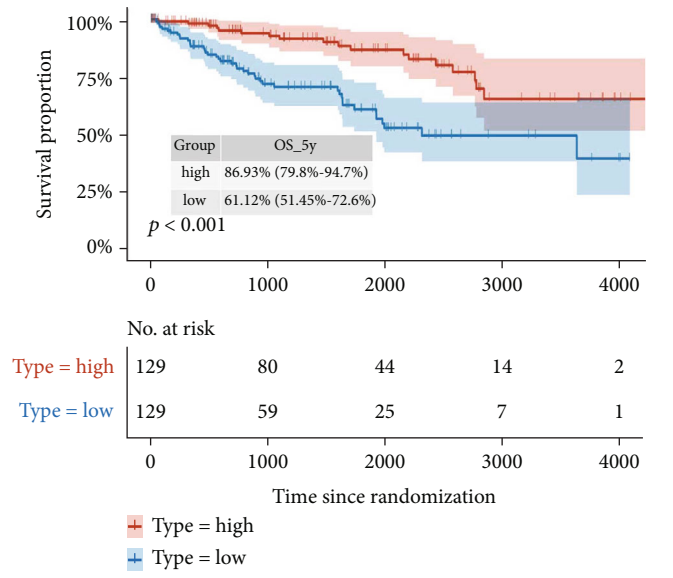
derived nomogram had better performance than that of an ideal model (Figure S10F).

4. Discussion

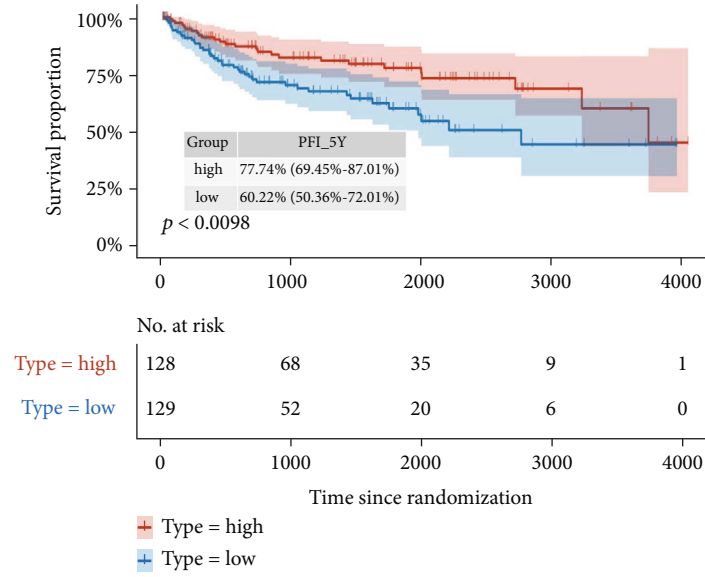
ccRCC is significantly different from many other solid tumors. Firstly, it has a special metabolic pattern and immune microenvironment as represented by its lipophilic, transparent, and metabolic pattern [38]. Although the TME of ccRCC is infiltrated with large numbers of immune cells, a typical feature is that patients with high-degree CD4⁺ and CD8⁺ T cell infiltration have a poorer prognosis [39]. Secondly, the heterogeneity of renal cancer is reflected by the fact that ccRCC patients with a similar stage and grade

may present completely different therapeutic response and OS after receiving the same treatment. They may present distinctive genetic and molecular alterations, experience different clinical courses, and exhibit different responses to the same therapy [40]. However, there are few studies trying to establish a model based on integrated data in the purpose of predicting both prognosis and drug selection in ccRCC patients. Hence, it is absolutely critical to develop a comprehensive and robust ccRCC prognosis and drug selection model based on integrated multiomics data and advanced strategies to help make prognosis prediction and specific personalized treatment plans.

With the rapid advance of high-throughput sequencing technology and fulfillment of The Cancer Genome Atlas,



(a)



(b)

FIGURE 9: Continued.

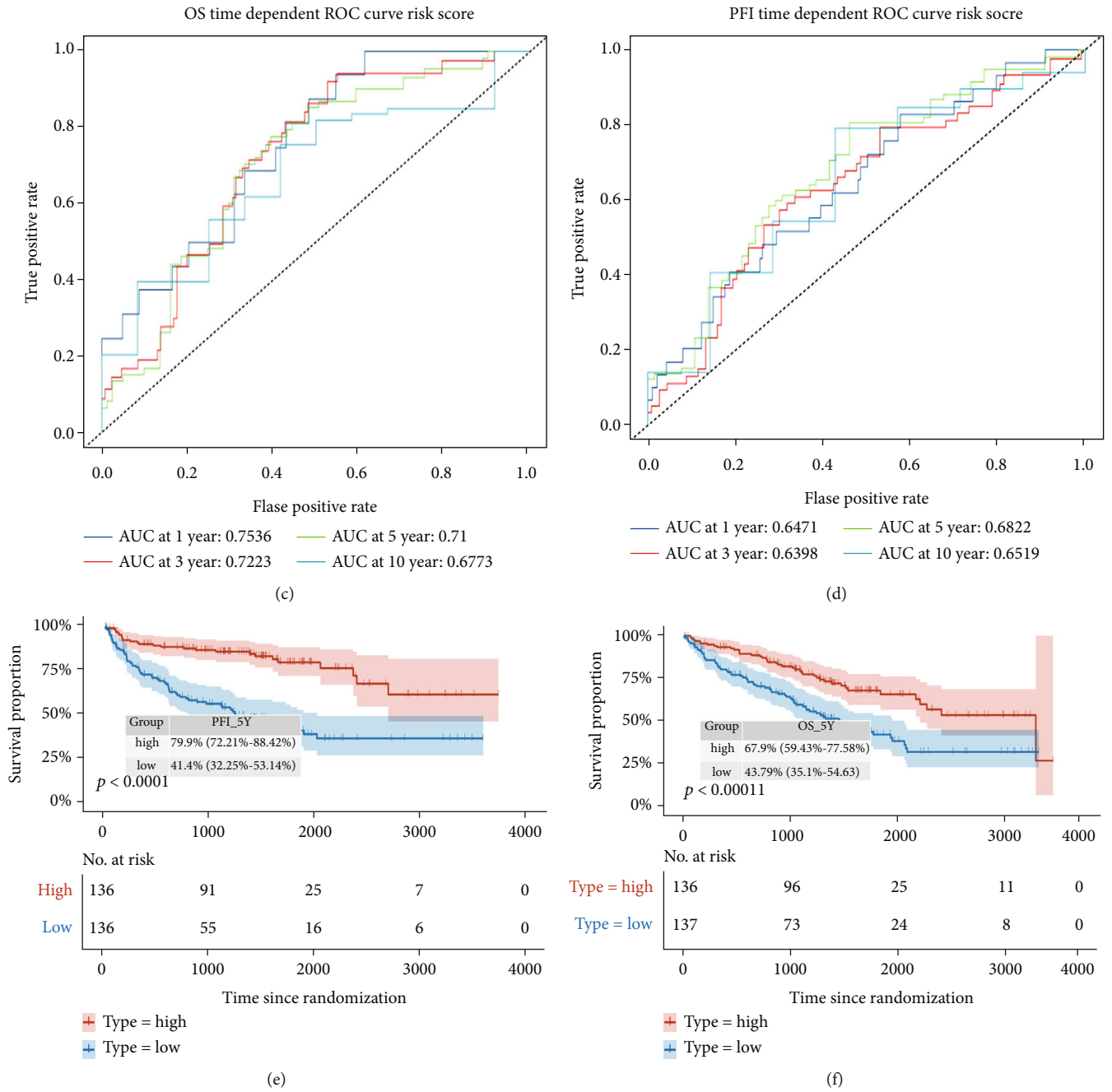


FIGURE 9: Continued.

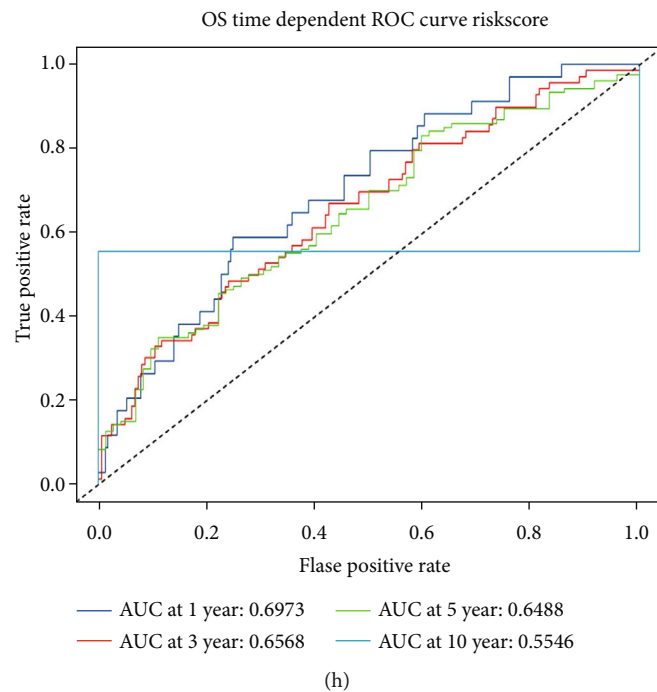
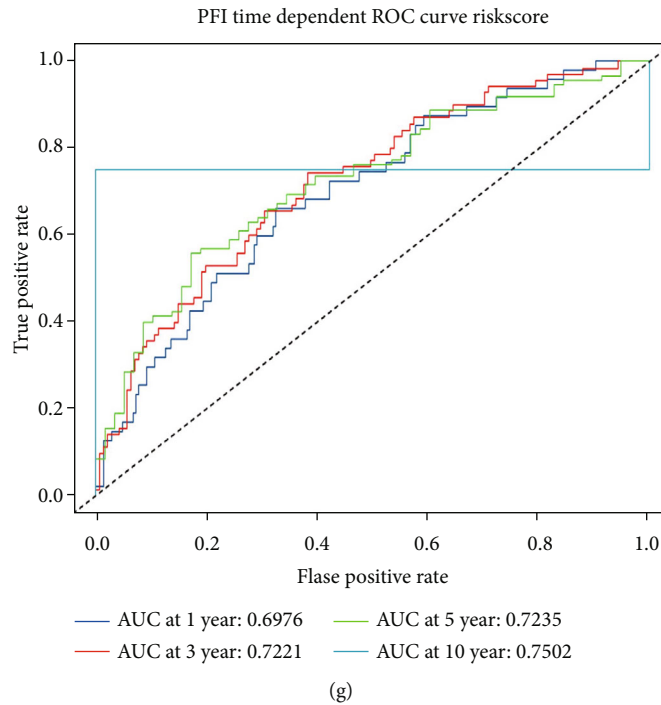


FIGURE 9: Construction of the risk scores. (a–d) ROC curve of OS and PFI, as well as Kaplan-Meier survival curve of OS and PFI of the two identified subtypes of renal cancer in train cohort. (e–h) ROC curve of OS and PFI, as well as Kaplan-Meier survival curve of OS and PFI of the two identified subtypes of renal cancer in train cohort.

the explosive growth of multiple omics has significantly contributed to researchers’ understanding of cancer at molecular level. The new challenges presented by massive amounts of omics data are data processing and analysis. In the present study, we made a comprehensive exploration of multiple omics including mRNA profiles, lncRNA profiles, miRNA profiles, DNA methylation profiles, and simple somatic mutation to construct a subtyping tool to help

address this important clinical issue. Omics data is complex, multilayered, and complementary so that a key goal of analyzing multiple omics data is to build effective models that predict phenotypic characteristics and elucidate the biological significance of important biomarkers. Another main goal of deciphering omics data is to reduce dimension and decrease inherent noise while at the same time to select relevant features and avoid overfit. As the purpose of this study

was to conduct a stratification model through multiomics data, we firstly used the Cox Proportional Typing model by R “Survival” package to screen the signatures in each set of omics data that are closely related to the prognosis of KIRC patients for subsequent analysis, then utilized ten robust machine learning algorithms to perform cross-validation and finally identified subtypes numbers. Through clustering the multiomics data of 258 ccRCC patients after filtration, we generated two subtypes of ccRCC and found significantly different landscapes of mRNA, lncRNA, miRNA, DNA methylation, somatic mutation, and CNVs between two subtypes of ccRCC patients. It was found that ccRCC patients in the CS2 subtype group showed better prognosis than those in the CS1 subtype group. In addition, pseudotiming analysis further validated the robustness of our subtyping method by two distinct clusters, and each of them was further stratified into four subtypes by the risk-related mRNA landscape. With the maturation of sequencing technology and the improvement of tumor atlas plans in recent years, more molecular typing studies on renal cancer have emerged. TCGA team remodeled ccRCC through a stratification remodeling study from the levels of genomic alterations, DNA methylation profiles, RNA, and proteomic signature and finally found that the PI3K/AKT pathway was recurrently mutated, suggesting that this pathway is a potential therapeutic target [41]. On this basis, numerous ccRCC typing studies have emerged by using a single omics profile or specific gene set. Although these classification studies have to some extent provided new directions for the classification and precise treatment of renal cancer to some extent, they have their respective shortcomings. For example, Chen et al. integrated multiple-omics data of all RCC cases only based on one algorithm (COCA) without incorporating the lncRNA profile into the typing study [42]. Besides, the classification methods used in these studies were relatively simple. Ricketts et al. integrated 843 RCC multiple-omics data and then categorized the data from each omics data alone, which did not achieve the true sense of multiple sets of data [43]. The methods used in these classification studies can hardly be applied to clinical practice because they lack accuracy. In contrast, the current study used a variety of mainstream clustering algorithms to integrate multiple profiles at the same time to perform a classification study on ccRCC patients by paying special attention to analysis at each omics level one by one and the heterogeneity of the two groups of patients in TME. Surprisingly, our ccRCC subtyping model constructed according to differential biomarkers showed good survival prediction in both training and validation sets.

Importantly, our study extended previous studies in patient subtyping based on sample clustering analysis. The discrete subtype information is unable to capture inter- and intracluster relationships and provide the overall structure of patient distribution. To remedy these shortcomings, we applied graph learning approaches to uncover the tree structures of expression profiles among patients, believing that they can provide complementary information for clustering analysis and offer new insight into the complex landscape of ccRCC. Based on the pseudochronological analysis, we found that there may be a certain evolutionary relation-

ship between CS1 and CS2. As the prognosis of the two subtypes is completely different, we performed a pseudotime analysis and found that the CS2 subtype was located at the beginning of the differentiation axis, while the CS1 subtype was in evolution. At the end of the axis, combined with the results of the prognostic analysis in previous studies, we speculate that the evolution of profiles between CS1 and CS2 can be used to explain the progressive condition of ccRCC patients. Among them, the up- and downregulated genes that participated in the progression from the CS2 state to the S1 state may be the hub genes and potential therapeutic targets for ccRCC patients.

Enrichment analysis showed that extracellular matrix organization, cornification, keratinocyte differentiation, and keratinization were distinctly enriched in the CS1 subgroup, which is in line with the following immune cell infiltration analysis. In GSVA analysis, we found that hypoxia, coagulation, and myogenesis pathways were upregulated in CS1 subtype, while G2_M checkpoint, fatty acid metabolism, and TGF_β signaling oncopathways were unregulated in CS2 subtype. All the GSVA results are consistent with the findings in the previously reported studies. Hypoxia is a typical character of RCC, which plays a crucial role in the development and progression of RCC together with angiogenesis and glucose metabolism, like a loop that is self-feeding [44]. In this study, we found significant differences in PBRM1 and BAP1 mutation between the two subgroups. PBRM1 is the second most common mutated gene in ccRCC after VHL, as well as a component of the SWI/SNF chromatin remodeling complex. Different studies investigated the biological consequences and the potential role of PBRM1 alteration in RCC prognosis and identified it as a drug response modulator, although some results are controversial [45]. The BAP1 gene has emerged as a major tumor suppressor mutated with various frequencies in numerous human malignancies, including uveal melanoma, malignant pleural mesothelioma, ccRCC, intrahepatic cholangiocarcinoma, hepatocellular carcinoma, and thymic epithelial tumor [46]. BAP1 orchestrates chromatin-associated processes including gene expression, DNA replication, and DNA repair [47]. BAP1 also exerts cytoplasmic functions, notably in regulating Ca²⁺ signaling at the endoplasmic reticulum [48]. Taken together, we speculate that the CS1 subtype of ccRCC tends to be infiltrated with more immune cells with apparent fibrosis or more mesenchyma as compared with the CS2 subtype, which seems to be a more “clear cell-like” classic renal carcinoma.

In recent years, immunotherapy has become one of the mainstays of treatment for renal cancer and been recommended by the American Food and Drug Administration (FDA) as the first-line treatment for advanced renal cancer [49, 50]. However, there is still a lack of targeted markers for predicting the benefit of immunotherapy in RCC patients. The TME contains a variety of components such as vascular interstitial components, chemokines, and inflammatory factors [51]. The composition and proportion of these components integrally determine the effectiveness of immunotherapy and directly determine the efficacy of anti-tumor immunity in renal cancer [52]. In the present study,

It was obvious that the proportion of different types of T cells in CS1 subtype was higher than that in CS2 subtype. They included activated CD4⁺ T cells, activated CD8⁺ T cells, regulatory T cells, T follicular helper cells, and type 1 T helper cells. Subsequent GSVA analysis also demonstrated that immune cell functions were mainly enriched in the CS1 subtype of ccRCC. However, although the infiltration scores of various immune cells in the CS1 group were relatively higher, the prognosis of patients in the CS1 group was poorer than that in the CS2 group. In addition, the TIDE algorithm revealed that the immunotherapy benefit rate in the CS2 group was significantly higher than that in CS1 group. Besides, the expression of immune checkpoints was significantly upregulated in CS1 subtype. All these results indicate that immune cell infiltration in CS1 subtype of ccRCC may play a coinhibitory role, while the immune microenvironment in CS2 subtype demonstrated a better antitumor effect. This hypothesis is in line with previous studies. ccRCC is generally recognized as an immunogenic tumor, but it is also known for its immune escaping through regulatory T cells by upregulating a variety of inhibitory functions [53]. Although patients with advanced renal cancer have more immune cell infiltration, these tumors often have immune escape. It is also proved that the degree of CD8⁺ T cell infiltration in renal cancer is inversely proportional to the patient's prognosis, meaning that a higher degree of CD8⁺ T cell infiltration predicts a worse prognosis of patients with renal cancer [54].

It was found in our study that ccRCC patients of the two subtypes exhibited a significantly different sensitivity to molecular targeted drugs: patients in the CS1 high-risk subgroup showed a higher sensitivity to Imatinib and Crizotinib as compared with those in the CS2 subgroup, while patients in the CS2 subgroup showed a higher sensitivity to Saracatinib, Licitinib, Gefitinib, and Dasatinib. To be noted, the response rate to immune checkpoint agents in ccRCC patients of CS2 subtype was higher than that in those of CS1 subtype. However, these immunotherapy markers are commonly used in other tumor types (such as PD1 and LAG3), and TMB was not identified as sensitive potential therapeutic target in ccRCC patients. Over the past 20 years, medical treatments for RCC have transitioned from nonspecific immune approaches to targeted therapy against vascular endothelial growth factor (VEGF) and today's novel immunotherapies. There are multiple targets, including platelet-derived growth factor and related receptors, and other inhibitors including mammalian target of rapamycin and the EMT and AXL tyrosine-protein kinase receptors. Nowadays, several immune-checkpoint inhibitors have been approved based on their significant curative effects in advanced RCC patients.

However, most other studies have focused on exploiting biomarkers for predicting the sensitivity of patients to immunotherapy and the optimal combination or selection of the existing agents. In the present study, we surprisingly noticed that CS1 subtype was insensitive to most molecular targeted drugs and immune inhibitory agents, which urged us to explore new potential drugs that could more effectively treat CS1 subtype. Also amazingly, we discovered a variety of

potential drugs specific to this high-risk subtype. Among them, compound CGP.082996 was found to be associated with the MYC locus, suggesting that it is a potent antiproliferative agent against retinoblastoma- (Rb-) positive tumor cells by exclusively arresting cells at G1 and reducing phospho-Ser 780/Ser 795 on the Rb protein [55]. CMK could inhibit the proliferation of lung adenocarcinoma cells both in vivo and vitro by targeting the RSK family members [56]. JNJ-26854165 could induce wild-type p53- and E2F1-mediated apoptosis in acute myeloid and lymphoid leukemia [57]. ZM 447439 is a novel promising aurora kinase inhibitor which can provoke antiproliferative and proapoptotic effects against gastroenteropancreatic neuroendocrine tumor diseases either alone or in combination with bio- and chemotherapeutic agents [58]. GNF-2 can affect structural dynamics of the ATP-binding site by binding to the myristate-binding site of Abl [59]. LFM-A13 exhibited a favorable pharmacokinetic behavior that was not adversely affected by the standard chemotherapy drugs vincristine, methylprednisolone, or l-asparaginase (when used as combination treatment, VPL) and significantly improved the chemotherapy response and survival outcome of mice challenged with BCL-1 leukemia cells [60]. RO.3306 could enhance p53-mediated Bax activation and mitochondrial apoptosis in AML by inhibiting cyclin-dependent kinase 1 [61]. AZD-6482 is a specific PI3K β inhibitor that exerts an antitumor effect by inhibiting proliferation and inducing apoptosis of human glioma cells [62]. Some of the potential drugs explored in the current study have not yet been reported before, such as AZD-6482, which is an intravenous platelet aggregation inhibitor which was reported to be used for the treatment of thrombosis but not cancer. It is worth noting that CGP-60474, an inhibitor of cyclin-dependent kinase, has proved to be the most potential drug in that it alleviated tumor necrosis factor- α (TNF- α) and interleukin-6 (IL-6) in activated macrophages by downregulating the NF- κ B activity and reducing the mortality rate in LPS-induced endotoxemia mice [63]. In addition, VX.680, a highly potent and selective small-molecule inhibitor of Aurora kinases, could block cell-cycle progression and induce apoptosis in a diverse range of human tumor types including leukemia and colon and pancreatic tumors [64]. Given that CS1 was more sensitive to the above-mentioned drugs and some drugs have achieved certain effects in clinical trials of some solid tumors, the above-mentioned drugs may provide new potential therapeutic targets for the high-risk subgroup of RCC.

As we discussed above, this study provides new insights into molecular subtyping of ccRCC and, at the same time, constructs a molecular prognostic model based on mRNA expression. Innovatively, this is the first study ever that established a robust prognosis and drug selection model by exploiting data at each omics level with several algorithms. Typically, we combined pseudochronological analysis into clustering analysis for better understanding of inter-/intracluster relationship as compared with previous studies. However, this study still has some limitations. Firstly, all the classification data used in this study include multiomics data. In real clinical work, obtaining so many omics data

from patients is expensive and time-consuming, so it is not easy to be applied in clinical practice. Secondly, although we compared drug sensitivities between different subgroups, specific experiments are still in need for further validation. As the current clinical multiomics panels are not yet popular, a risk model based on mRNA expression and the characteristics of subgroups needs to offer implications for the clinical work.

In summary, this is an innovative study on ccRCC molecular typing based on multiple omics, which describes differences between subgroups at multiple omics levels. As this prognostic model could efficiently predict the long-term survival and drug selection of ccRCC patients, it might bring new hope to the diagnosis and treatment of ccRCC.

Data Availability

The data used to support the findings of this study have been deposited in the TCGA repository (<https://portal.gdc.cancer.gov/>), ICGC repository (<https://dcc.icgc.org/>) and EAG repository (<https://ega-archive.org/>), respectively.

Disclosure

This article was also published as a preprint on BioRxiv (<https://www.biorxiv.org/content/10.1101/2021.08.06.455377v1.full.pdf>) doi:<https://doi.org/10.1101/2021.08.06.455377>.

Conflicts of Interest

The authors declare no conflicts.

Authors' Contributions

Aimin Jiang, Yewei Bao, and Anbang Wang have contributed equally to this work. Jie Wang, Bing Liu, Juan Lu, and Linhui Wang conceptualized and designed this study. Wenliang Gong, Xinxin Gan, Zhenjie Wu, and Yi Bao wrote the first draft of the manuscript. All authors contributed to the article and approved the submitted version. Aimin Jiang, Yewei Bao, and Anbang Wang have contributed equally to this work and share the first authorship.

Acknowledgments

We sincerely thank all participants involved in this study. This study was funded by the National Natural Science Foundation of China (82072812, 81730073, 81872074, and 81902560), Shanghai Sailing Program (19YF1448300), and Clinical science and technology innovation project of Shanghai Shenkang Hospital Development Center (SHDC12018108). We also thank Dr. Jianming Zeng (University of Macau) and all the members of his bioinformatics team, biotrainee, for generously sharing their experience and codes.

Supplementary Materials

Figure S1: the workflow of this study. Figure S2: (A) identification of the optimal cluster number by calculating CPI

(blue line) and Gaps-statistics (red line) in TCGA-KIRC cohort. (B) Consensus heatmap based on results from 10 multiomics integrative clustering algorithms with the cluster number of 2, 3, 4, and 5, and related quantification of sample similarity using the Silhouette score based on the consensus ensembles result. Figure S3: heatmap of subtype-specific upregulated and downregulated biomarkers using DEseq2 for the two identified subtypes in TCGA-KIRC cohort. Figure S4: (A) DNA substitution types including transition (Ti) and transversion (Tv) among CS1 and CS2. (B) The most significant difference of mutated genes between CS1 and CS2. (C) The lollipop plot illustrates the differential distribution of variants for PBRM1. (D) Kaplan-Meier curves show the independent relevance between overall survival and PBRM1 mutation in CS1 and CS2 subgroups. (E) The lollipop plot illustrates the differential distribution of variants for PBRM1. (F) Potential druggable gene categories from mutation dataset in CS1 and CS2. Figure S5: (A) differences in oncogenic signaling pathways between the CS1 and CS2 subgroups. (B) Heatmap of oncogenic pathways based on ssGSEA algorithm in the CS1 and CS2 subgroups. Figure S6: heatmap of oncogenic pathways based on the ssGSEA algorithm in the CS1 and CS2 subgroups. Figure S7: (A) heatmap of the immune-related gene family of chemokines and receptors, coinhibitors, costimulators, interferons, and receptors and MHC expression in the CS1 and CS2 subgroups. (B) Expression of immune checkpoints in CS1 and CS2 subgroups and difference in ESTIMATE scores between the subgroups. Figure S8: the structure tomographs of the candidate small-molecule drugs for the CS1 subgroup. Figure S9: (A) heatmap of NTP in outer KIRC cohort using subtype-specific upregulated biomarkers identified from Japan ccRCC cohort. (B) Kaplan-Meier survival curve of the two predicted subtypes of renal cancer in Japan ccRCC cohort. Figure S10: (A) common DEGs in CS1 and CS2 subtypes were classified into gene signature A and B subtypes by unsupervised clustering. (B) Use of Boruta algorithm to calculate variable importance score (green is important variable; blue is shadow variable). (C) The importance of the variables is sorted: from right to left, the importance is decreasing in order. (D) Estimated risk score between train and validation cohorts. (E) Nomograms for predicting the probability of patient mortality based on the risk score of 3- and 5-year OS. (F) Plots depict the calibration of nomograms based on the risk score of 3- and 5-year OS. Supplement Table 1: biomarker signatures of each subtype. Supplement Table 2: recurrently amplified and deleted regions in the two subgroups calculated by GISTIC2.0. Supplement Table 3: name of discovered 53 kinds of small-molecule drugs that could be used as potential drugs for the treatment of CS1. (*Supplementary Materials*)

References

- [1] U. Capitanio and F. Montorsi, "Renal cancer," *Lancet*, vol. 387, no. 10021, pp. 894–906, 2016.
- [2] P. C. Barata and B. I. Rini, "Treatment of renal cell carcinoma: current status and future directions," *CA: A Cancer Journal for Clinicians*, vol. 67, no. 6, pp. 507–524, 2017.

- [3] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2020," *CA: A Cancer Journal for Clinicians*, vol. 70, no. 1, pp. 7–30, 2020.
- [4] K. Geissler, P. Fornara, C. Lautenschläger, H.-J. Holzhausen, B. Seliger, and D. Riemann, "Immune signature of tumor infiltrating immune cells in renal cancer," *Oncoimmunology*, vol. 4, no. 1, article e985082, 2015.
- [5] S. J. Teran Hidalgo and S. Ma, "Clustering multilayer omics data using MuNCut," *BMC Genomics*, vol. 19, no. 1, p. 198, 2018.
- [6] J. M. Satagopan, E. S. Venkatraman, and C. B. Begg, "Two-stage designs for gene-disease association studies with sample size constraints," *Biometrics*, vol. 60, no. 3, pp. 589–597, 2004.
- [7] D. Ramazzotti, A. Lal, B. Wang, S. Batzoglou, and A. Sidow, "Multi-omic tumor data reveal diversity of molecular mechanisms that correlate with survival," *Nature Communications*, vol. 9, no. 1, p. 4453, 2018.
- [8] Q. Mo, R. Shen, C. Guo, M. Vannucci, K. S. Chan, and S. G. Hilsenbeck, "A fully Bayesian latent variable model for integrative clustering analysis of multi-type omics data," *Biostatistics*, vol. 19, no. 1, pp. 71–86, 2018.
- [9] C. Meng, D. Helm, M. Frejno, and B. Kuster, "moCluster: identifying joint patterns across multiple omics data sets," *Journal of Proteome Research*, vol. 15, no. 3, pp. 755–765, 2016.
- [10] K. A. Hoadley, C. Yau, D. M. Wolf et al., "Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin," *Cell*, vol. 158, no. 4, pp. 929–944, 2014.
- [11] S. Monti, P. Tamayo, J. Mesirov, and T. Golub, "Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data," *Machine Learning*, vol. 52, no. 1/2, pp. 91–118, 2003.
- [12] P. Chalise and B. L. Fridley, "Integrative clustering of multi-level 'omic' data based on non-negative matrix factorization algorithm," *PLoS One*, vol. 12, no. 5, article e0176278, 2017.
- [13] D. Wu, D. Wang, M. Q. Zhang, and J. Gu, "Fast dimension reduction and integrative clustering of multi-omics data using low-rank approximation: application to cancer molecular classification," *BMC Genomics*, vol. 16, no. 1, 2015.
- [14] N. Rappoport and R. Shamir, "NEMO: cancer subtyping by integration of partial multi-omic data," *Bioinformatics*, vol. 35, no. 18, pp. 3348–3356, 2019.
- [15] H. Nguyen, S. Shrestha, S. Draghici, and T. Nguyen, "PINS-Plus: a tool for tumor subtype discovery in integrated genomic data," *Bioinformatics*, vol. 35, no. 16, pp. 2843–2846, 2019.
- [16] B. Wang, A. M. Mezlini, F. Demir et al., "Similarity network fusion for aggregating data types on a genomic scale," *Nature Methods*, vol. 11, no. 3, pp. 333–337, 2014.
- [17] Y. Sato, T. Yoshizato, Y. Shiraishi et al., "Integrated molecular analysis of clear-cell renal cell carcinoma," *Nature Genetics*, vol. 45, no. 8, pp. 860–867, 2013.
- [18] X. Lu, J. Meng, Y. Zhou, L. Jiang, and F. Yan, "MOVICS: an R package for multi-omics integration and visualization in cancer subtyping," *Bioinformatics*, vol. 36, no. 22–23, pp. 5539–5541, 2021.
- [19] G. Yu, L.-G. Wang, Y. Han, and Q.-Y. He, "clusterProfiler: an R package for comparing biological themes among gene clusters," *Integrative Biology*, vol. 16, no. 5, pp. 284–287, 2012.
- [20] S. Hänzelmann, R. Castelo, and J. Guinney, "GSVA: gene set variation analysis for microarray and RNA-seq data," *BMC Bioinformatics*, vol. 14, no. 1, 2013.
- [21] T. Li, J. Fan, B. Wang et al., "TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells," *Cancer Research*, vol. 77, no. 21, pp. e108–e110, 2017.
- [22] E. Becht, N. A. Giraldo, L. Lacroix et al., "Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression," *Genome Biology*, vol. 17, no. 1, p. 218, 2016.
- [23] D. Aran, Z. Hu, and A. J. Butte, "xCell: digitally portraying the tissue cellular heterogeneity landscape," *Genome Biology*, vol. 18, no. 1, p. 220, 2017.
- [24] J. Racle and D. Gfeller, "EPIC: a tool to estimate the proportions of different cell types from bulk gene expression data," *Methods in Molecular Biology*, vol. 2120, pp. 233–248, 2020.
- [25] A. Mayakonda, D.-C. Lin, Y. Assenov, C. Plass, and H. P. Koefler, "Maftools: efficient and comprehensive analysis of somatic variants in cancer," *Genome Research*, vol. 28, no. 11, pp. 1747–1756, 2018.
- [26] "Comprehensive molecular characterization of gastric adenocarcinoma," *Nature*, vol. 513, no. 7517, pp. 202–209, 2014.
- [27] T. Cokelaer, E. Chen, F. Iorio et al., "GDSCTools for mining pharmacogenomic interactions in cancer," *Bioinformatics*, vol. 34, no. 7, pp. 1226–1228, 2018.
- [28] P. Geeleher, N. Cox, and R. S. Huang, "pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels," *PLoS One*, vol. 9, no. 9, article e107468, 2014.
- [29] C. Trapnell, D. Cacchiarelli, J. Grimsby et al., "The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells," *Nature Biotechnology*, vol. 32, no. 4, pp. 381–386, 2014.
- [30] M. B. Kursa and W. R. Rudnicki, "Feature selection with the Boruta package," *Journal of Statistical Software*, vol. 36, no. 1, pp. 1–13, 2010.
- [31] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1–3, pp. 37–52, 1987.
- [32] X. Zhang, M. Shi, T. Chen, and B. Zhang, "Characterization of the immune cell infiltration landscape in head and neck squamous cell carcinoma to aid immunotherapy," *Molecular Therapy-Nucleic Acids*, vol. 22, pp. 298–309, 2020.
- [33] A. Jiang, Y. Bao, A. Wang et al., "Establishment of a prognostic prediction and drug selection model for patients with clear cell renal cell carcinoma by multi-omics data analysis," *BioRxiv*, 2021.
- [34] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *Journal of the Royal Statistical Society: Series B*, vol. 63, no. 2, pp. 411–423, 2001.
- [35] J. R. Gnarr, K. Tory, Y. Weng et al., "Mutations of the *VHL* tumour suppressor gene in renal carcinoma," *Nature Genetics*, vol. 7, no. 1, pp. 85–90, 1994.
- [36] P. Kapur, S. Peña-Llopis, A. Christie et al., "Effects on survival of *BAP1* and *PBRM1* mutations in sporadic clear-cell renal-cell carcinoma: a retrospective analysis with independent validation," *The Lancet Oncology*, vol. 14, no. 2, pp. 159–167, 2013.
- [37] C. Sotiriou, P. Wirapati, S. Loi et al., "Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis," *Journal of the National Cancer Institute*, vol. 98, no. 4, pp. 262–272, 2006.
- [38] B. Owens, "Kidney cancer," *Nature*, vol. 537, no. 7620, p. S97, 2016.

- [39] S. Chevrier, J. H. Levine, V. R. T. Zanotelli et al., “An immune atlas of clear cell renal cell carcinoma,” *Cell*, vol. 169, no. 4, 2017.
- [40] N. Chowdhury and C. G. Drake, “Kidney Cancer: An overview of current therapeutic approaches,” *The Urologic Clinics of North America*, vol. 47, no. 4, pp. 419–431, 2020.
- [41] “Comprehensive molecular characterization of clear cell renal cell carcinoma,” *Nature*, vol. 499, no. 7456, pp. 43–49, 2013.
- [42] F. Chen, Y. Zhang, Y. Şenbabaoğlu et al., “Multilevel genomics-based taxonomy of renal cell carcinoma,” *Cell Reports*, vol. 14, no. 10, pp. 2476–2489, 2016.
- [43] C. J. Ricketts, A. A. de Cubas, H. Fan et al., “The cancer genome atlas comprehensive molecular characterization of renal cell carcinoma,” *Cell Reports*, vol. 23, no. 1, 2018.
- [44] G. L. Semenza, “HIF-1 mediates metabolic responses to intratumoral hypoxia and oncogenic mutations,” *The Journal of Clinical Investigation*, vol. 123, no. 9, pp. 3664–3671, 2013.
- [45] A. M. Nargund, C. G. Pham, Y. Dong et al., “The SWI/SNF protein PBRM1 restrains VHL-loss-driven clear cell renal cell carcinoma,” *Cell Reports*, vol. 18, no. 12, pp. 2893–2906, 2017.
- [46] S. Peña-Llopis, S. Vega-Rubín-de-Celis, A. Liao et al., “BAP1 loss defines a new class of renal cell carcinoma,” *Nature Genetics*, vol. 44, no. 7, 2012.
- [47] A. A. Hakimi, I. Ostrovskaya, B. Reva et al., “Adverse outcomes in clear cell renal cell carcinoma with mutations of 3p21 epigenetic regulators BAP1 and SETD2: a report by MSKCC and the KIRC TCGA research network,” *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, vol. 19, no. 12, pp. 3259–3267, 2013.
- [48] F. Dai, H. Lee, Y. Zhang et al., “BAP1 inhibits the ER stress gene regulatory network and modulates metabolic stress response,” *Proceedings of the National Academy of Sciences*, vol. 114, no. 12, pp. 3192–3197, 2017.
- [49] T. K. Choueiri and R. J. Motzer, “Systemic therapy for metastatic renal-cell carcinoma,” *The New England Journal of Medicine*, vol. 376, no. 4, pp. 354–366, 2017.
- [50] R. R. McKay, D. Bossé, and T. K. Choueiri, “Evolving systemic treatment landscape for patients with advanced renal cell carcinoma,” *Journal of Clinical Oncology*, vol. 36, no. 36, pp. 3615–3623, 2018.
- [51] S. I. Grivnenkov, F. R. Greten, and M. Karin, “Immunity, inflammation, and cancer,” *Cell*, vol. 140, no. 6, pp. 883–899, 2010.
- [52] N. Borcherding, A. Vishwakarma, A. P. Voigt et al., “Mapping the immune environment in clear cell renal carcinoma by single-cell genomics,” *Communications Biology*, vol. 4, no. 1, pp. 1–11, 2021.
- [53] C. M. Díaz-Montero, B. I. Rini, and J. H. Finke, “The immunology of renal cell carcinoma,” *Nature Reviews Nephrology*, vol. 16, no. 12, pp. 721–735, 2020.
- [54] W. H. Fridman, L. Zitvogel, C. Sautès-Fridman, and G. Kroemer, “The immune contexture in cancer prognosis and treatment,” *Nature Reviews Clinical Oncology*, vol. 14, no. 12, pp. 717–734, 2017.
- [55] D. W. Fry, P. J. Harvey, P. R. Keller et al., “Specific inhibition of cyclin-dependent kinase 4/6 by PD 0332991 and associated antitumor activity in human tumor xenografts,” *Molecular Cancer Therapeutics*, vol. 3, no. 11, pp. 1427–1438, 2004.
- [56] Q. He, R. He, W. Luo et al., “Expression of RSK4 in lung adenocarcinoma tissue and its clinicopathological value: a study based on RNA-seq data and immunohistochemistry,” *International Journal of Clinical and Experimental Pathology*, vol. 10, no. 12, pp. 11405–11414, 2017.
- [57] K. Kojima, J. K. Burks, J. Arts, and M. Andreeff, “The novel tryptamine derivative JNJ-26854165 induces wild-type p53- and E2F1-mediated apoptosis in acute myeloid and lymphoid leukemias,” *Molecular Cancer Therapeutics*, vol. 9, no. 9, pp. 2545–2557, 2010.
- [58] I. Georgieva, D. Koychev, Y. Wang, J. Holstein, M. Zeitz, and P. Grabowski, “ZM 447439, a novel promising aurora kinase inhibitor, provoking antiproliferative and pro-apoptotic effects alone and in combination with bio- and chemotherapeutic agents in gastroenteropancreatic neuroendocrine tumor disease,” *Molecular Cancer Therapeutics*, vol. 6, 11 Supplement, 2007.
- [59] J. Zhang, F. J. Adrián, W. Jahnke et al., “Targeting Bcr-Abl by combining allosteric with ATP-binding-site inhibitors,” *Nature*, vol. 463, no. 7280, pp. 501–506, 2010.
- [60] F. M. Uckun, Y. Zheng, M. Cetkovic-Cvrlje et al., “In vivo pharmacokinetic features, toxicity profile, and chemosensitizing activity of alpha-cyano-beta-hydroxy-beta-methyl-N-(2,5-dibromophenyl)propenamide (LFM-A13), a novel anti-leukemic agent targeting Bruton’s tyrosine kinase,” *Clinical Cancer Research*, vol. 8, no. 5, pp. 1224–1233, 2002.
- [61] K. Kojima, M. Shimanuki, M. Shikami, M. Andreeff, and H. Nakakuma, “Cyclin-dependent kinase 1 inhibitor RO-3306 enhances p53-mediated Bax activation and mitochondrial apoptosis in AML,” *Cancer Science*, vol. 100, no. 6, pp. 1128–1136, 2009.
- [62] P.-F. Xu, J.-A. Yang, J.-H. Liu et al., “PI3K β inhibitor AZD6482 exerts antiproliferative activity and induces apoptosis in human glioblastoma cells,” *Oncology Reports*, vol. 41, no. 1, pp. 125–132, 2019.
- [63] H.-W. Han, S. Hahn, H. Y. Jeong et al., “LINCS L1000 dataset-based repositioning of CGP-60474 as a highly potent anti-endotoxemic agent,” *Scientific Reports*, vol. 8, no. 1, article 14969, 2018.
- [64] E. A. Harrington, D. Bebbington, J. Moore et al., “VX-680, a potent and selective small-molecule inhibitor of the Aurora kinases, suppresses tumor growth *in vivo*,” *Nature Medicine*, vol. 10, no. 3, pp. 262–267, 2004.