Research Paper

Identification of neoantigens and immunological subtypes in clear cell renal cell carcinoma for mRNA vaccine development and patient selection

Daoqi Zhu[1,*], Jiabin Yang[1,*], Minyi Zhang[2,*], Zhongxiao Han[1], Meng Shao[3], Qin Fan[1], Yun Ma[4], Dandan Xie[5], Wei Xiao[1]

[1]School of Traditional Chinese Medicine, Southern Medical University, Guangzhou 510515, Guangdong, China
[2]Department of Epidemiology, School of Public Health, Southern Medical University, Guangzhou 510515, Guangdong, China
[3]Guangdong Provincial Key Laboratory of Chinese Medicine Pharmaceutics, School of Traditional Chinese Medicine, Southern Medical University, Guangzhou 510515, Guangdong, China
[4]Department of pharmacy, Nanfang Hospital, Southern Medical University, Guangzhou 510515, Guangdong, China
[5]The Affiliated TCM Hospital of Guangzhou Medical University, Guangzhou 510130, Guangdong, China
[*]Equal contribution

Correspondence to: Yun Ma, Dandan Xie, Wei Xiao; **email:** myxw1977@smu.edu.cn, 2020621986@gzhmu.edu.cn, xw7688@smu.edu.cn

ABSTRACT

Clear cell renal cell carcinoma (ccRCC) is a common urological malignancy with diverse histological types. This study aimed to detect neoantigens in ccRCC to develop mRNA vaccines and distinguish between ccRCC immunological subtypes for construction of an immune landscape to select patients suitable for vaccination. Using The Cancer Genome Atlas SpliceSeq database, The Cancer Genome Atlas, and the International Cancer Genome Consortium cohorts, we comprehensively analysed potential tumour antigens of ccRCC associated with aberrant alternative splicing, somatic mutation, nonsense-mediated mRNA decay factors, antigen-presenting cells, and overall survival. Immune subtypes (C1/C2) and nine immune gene modules of ccRCC were identified by consistency clustering and weighted correlation network analysis. The immune landscape as well as molecular and cellular characteristics of immunotypes were assessed. Rho-guanine nucleotide exchange factor 3 (ARHGEF3) was identified as a new ccRCC antigen for development of an mRNA vaccine. A higher tumour mutation burden, differential expression of immune checkpoints, and immunogenic cell death were observed in cases with the C2 immunotype. Cellular characteristics increased the complexity of the immune environment, and worse outcomes were observed in ccRCC cases with the C2 immunotype. We constructed the immune landscape for selecting patients with the C2 immunotype suitable for vaccination.

INTRODUCTION

Renal cell carcinoma (RCC) is among the top 10 cancers worldwide. It is a commonly diagnosed and histo-logically diverse urologic malignancy accounting for >90% of all renal neoplasms, with clear cell RCC (ccRCC) representing almost 75% of cases [1]. Localized RCC is treatable surgically by either radical or partial nephrectomy; nevertheless, ~30% of localized ccRCC metastasize at some point [2]. RCC major risk factors include obesity, hypertension, and smoking, and the incidence increases markedly with age. The

pathogenesis of ccCCR is closely related to several genetic mutations, including *VHL*, [3] *PBRM1*, [4] *SETD2*, [4] *BAP1*, [5] *KDM5C*, [4] and *mTOR* mutations [6]. However, ccRCC development requires supplementary epigenetic and genetic events [7]. As RCCs are highly vascular, it is not surprising that tyrosine kinase inhibitors, [8] anti-VEGF monoclonal antibodies, [9] and mTOR [10] inhibitors have been used to explore their features. Unfortunately, there are no available clinical markers to classify patients for therapies, despite intensive efforts. With respect to immunotherapy, nivolumab response rate is only 25%, and there was no significant decrease in tumor size in most patients treated with nivolumab [11]. The relevance of the International Metastatic Renal Cell Carcinoma Database Consortium (IMDC) prognostic criteria remains to be established in the era of frontline combination immunotherapy [12]. In the absence of prognostic standards based on alternative immuno-therapy, the IMDC criteria continue to be used in clinical trials to stratify patients at risk under the guide of clinical guidelines. Immune checkpoint (ICP) blockade (ICB) therapy increases the overall survival (OS) rates of advanced RCC cases treated with nivolumab [11].

Tumor mutation burden (TMB) is a clinically related parameter that highlights the molecular characteristics associated with immunotherapy responses [13]. Splicing of pre-mRNA is significant to the pathology of various diseases, especially cancer. Aberrant splicing isoforms are highlighted as tumor markers and cancer therapy targets [14]. Aberrantly expressed transcripts in cancer cells are degraded by nonsense-mediated mRNA decay (NMD), a process involved in the mRNA quality control system [15]. NMD disruption factors have significant correlation with the splicing isoform count in TCGA-Lung adenocarcinoma datasets; [16] the precise identification of aberrant transcripts in cancer cells is essential for identifying potential neoantigens. Particularly, the initial hypothesis based on TMB as an immunotherapy-relevant parameter is associated with the fact that somatic variants can generate tumor-specific neoantigens. Several clinical trials have used TMB as an important stratification factor or a landmark endpoint to decipher the role of TMB in cancer-type treatment decision-making [17]. A minority of somatic mutations in DNA can give rise to neoantigens, but not all neopeptides present on the cell surface are immunogenic [18]. Consequently, TMB analysis to detect the underlying mutations subsets responsible for immunogenicity may enable substantial optimization of biomarker accuracy and improved therapeutic targeting of neoantigens [19]. Hence, profiling abnormal transcripts is equally essential as a potential biomarker of an ICP inhibitor (ICI).

Although mRNA-based cancer vaccines have been extensively reviewed, with the recent approval of two mRNA lipid nanoparticle vaccines for coronavirus disease 2019 (COVID-19), the focus has again shifted to mRNA vaccines as a promising platform for cancer immunotherapy. Recently, the Massachusetts Institute of Technology (MIT) Technology Review released the 2021 list of "Top 10 Breakthrough Technologies in the World," and mRNA vaccines that have caused significant changes in the field of medicine topped the list. MIT commented that mRNA technology has broad application prospects for treating various infectious diseases (including COVID-19 and malaria) and cancer, among other diseases. mRNA vaccines are rapidly advancing in preclinical and clinical studies of cancer and infectious diseases [20]. We can design cancer vaccines to target tumor-associated unique antigens that have preferential expression in cancerous cells, and to achieve this goal, multiple preclinical and clinical trials of mRNA vaccines have been conducted [21]. Compared to an infectious disease vaccine, the challenge in developing a cancer vaccine lies in the limitation of clinical translation because of the difficulty in predicting antigens and their poor immunogenicity. Tumor antigens vary significantly among cancer patients; hence, identifying immunogenic tumor-associated antigens and tumor-specific antigens and overcoming the inhibitory tumor microenvironment remain significant obstacles for the development of mRNA-based cancer vaccines. Here, we aimed to identify novel ccRCC antigens and different immune subtypes for developing mRNA-based cancer vaccines and mapping the immune landscape of ccRCC to select patients fitting the criteria for vaccination.

RESULTS

Identification of aberrant alternative splicing (AS) events and potential ccRCC antigens

To identify the features of aberrant AS in ccRCC patients, seven AS events compared with normal samples were screened from TCGA cohort (Supplementary Figure 1A); exon skip (ES) had the highest incidence (15481), appearing in 6506 genes, followed by the alternative promoter (AP) event. In contrast, the lowest mutually exclusive exons (MEs) occurred 178 times in 172 genes (Figure 1A; Supplementary Figure 1B). A PSI value can be used to evaluate the variation of AS events; the PSI values for 11289 and 12049 AS events increased and decreased, respectively, in ccRCC patients (Figure 1B). The heatmap of PSI for the top 100 AS events is shown in Figure 1C. Gene mutations, especially splice site mutations, may lead to aberrant splicing isoforms; [22] consequently, we analyzed and visualized somatic mutations (MuTect2 variant aggregation and masking)
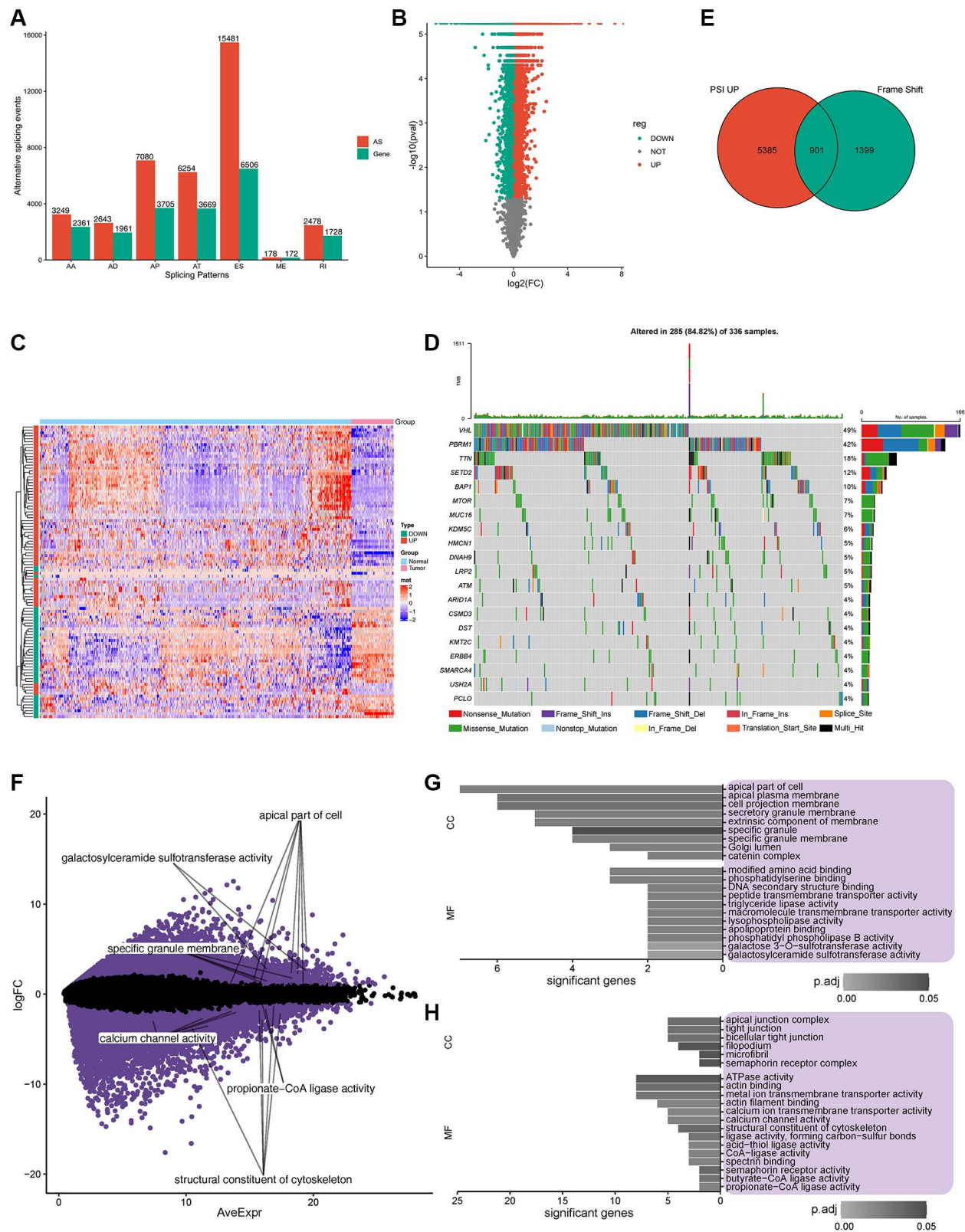
Figure 1. Potential antigen identification and enrichment of gene ontology (GO). (**A**) Overview of alternate splicing (AS) in ccRCC patients. (**B**) PSI of all the AS events, Green represents downregulated gene expression, while red represents upregulated gene expression. (**C**) Heatmap of PSI for the top 100 AS events. (**D**) Twenty highly mutated genes in ccRCC patients from TCGA cohort. (**E**) Total of 901 potential antigens were found in PSI, which were significantly upregulated after frame shift mutations, Green represents frame shift genes, while red represents PSI up genes. (**F**) Expression of 156 potential antigens was significantly regulated. (**G**, **H**). Frequency of somatic mutations and genes/patients involved. (**G**, **H**). GO enrichment of significantly upregulated/downregulated potential antigens.

in ccRCC patients from TCGA cohort (Figure 1D). Among the entire cohort of patients included in the database, it was observed that 49% of the patients presented with VHL gene mutations, while 42% had mutations in the PBRM1 gene [4, 23]. Importantly, within the subset of patients exhibiting mutations in either the VHL or PBRM1 gene, approximately half of them concurrently possessed mutations in both genes. Missense mutations occurred most frequently in 5689 genes, followed by frameshift deletion in 578 genes (Supplementary Figure 1C). Multiple mutations commonly occurred in one sample, and the top three mutation types were missense mutations, frameshift deletions, and nonsense mutations (Supplementary Figure 1D). The distributions of the top 10 mutated genes in all individuals and ccRCC patients are shown in Supplementary Figure 1E, 1F, respectively. Our data revealed that missense mutations and frameshift mutations occurred more frequently in patients with ccRCC.

In total, 901 genes were identified as encoding for potential antigens in the PSI significantly upregulated group (including frameshift deletion and frameshift insertion) (Figure 1E). GO enrichment analysis of the 901 potential antigens in terms of biological processes (BP), molecular function (MF), cellular component (CC) is shown in Supplementary Figure 1G–1I. Further analysis of the expression profiles of TCGA cohort revealed 10393 and 8134 genes with up-regulated and downregulated expression, respectively. Among these significantly regulated genes (Figure 1F), GO enrichment analysis indicated that potential upregulated antigens were related to the apical part of the cell, apical plasma membrane, and cell projection membrane in CC and modified amino acid binding and phosphatidylserine binding in MF (Figure 1G). Potential downregulated antigens were related to apical junction complex, tight junction, and bicellular tight junction in CC, and ATPase activity, actin binding, and metal ion transmembrane transporter activity in MF (Figure 1H). Furthermore, 156 candidate antigens showed upregulated aberrant AS events with frameshift mutations and abnormal expression.
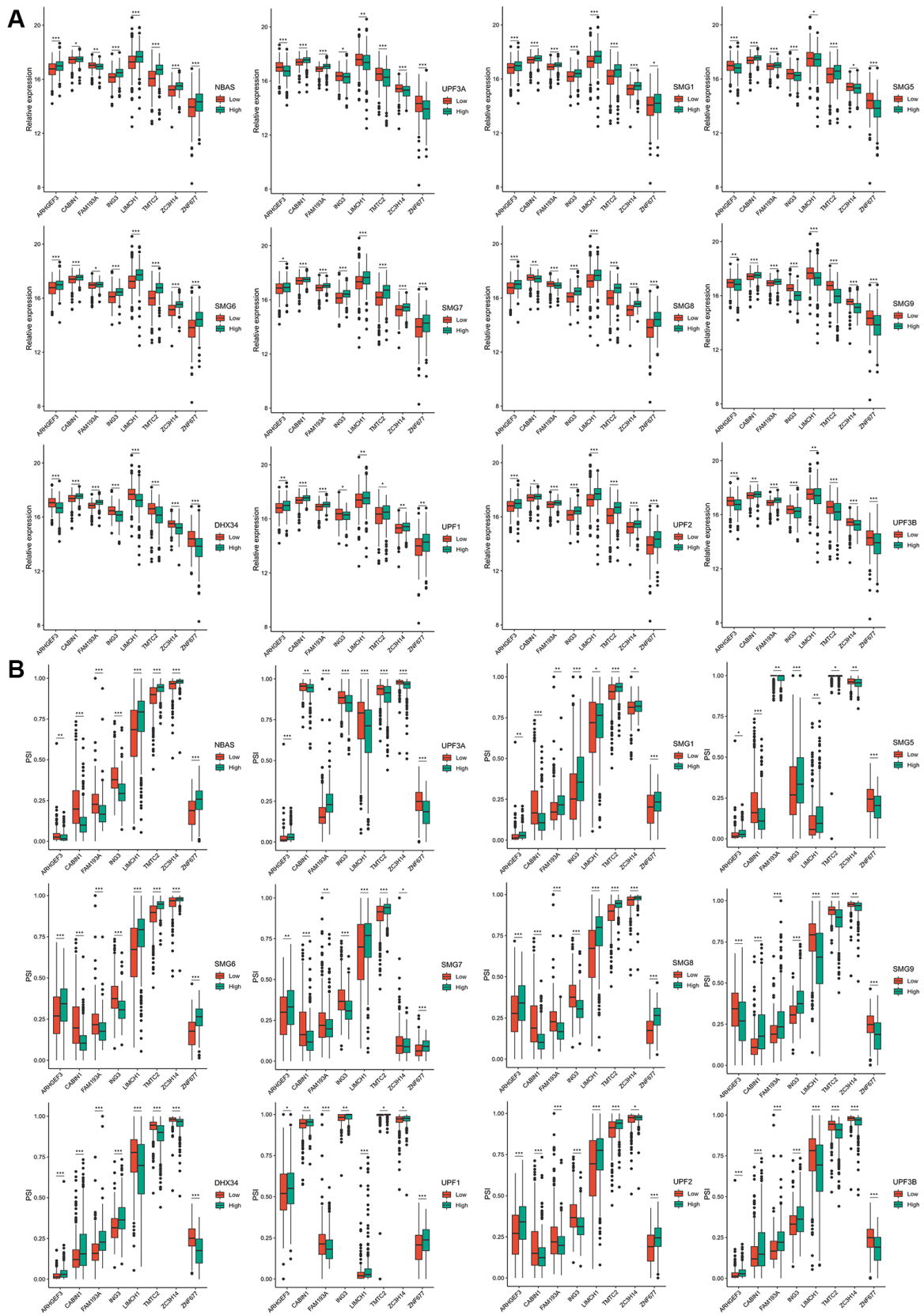
Potential antigens related to NMD

Twelve NMD factors were used to identify the potential antigens in ccRCC. The expression profiles of TCGA datasets were grouped according to the median expression level of each NMD factor. Forty-seven differentially expressed genes were identified with respect to the expression of 12 NMD factors. Notably, the expression of *ARHGEF3*, *CABIN1*, *FAM193A*, *ING3*, *LIMCH1*, *TMTC2*, *ZC3H14*, and *ZNF677* (the

low- and high-expression groups) was significantly different among the 12 NMD factors (Figure 2A). Furthermore, 157 differentially expressed PSI genes were observed among the 12 NMD factors (Supplementary Table 1). Similarly, the PSI of *ARHGEF3*, *CABIN1*, *FAM193A*, *ING3*, *LIMCH1*, *TMTC2*, *ZC3H14*, and *ZNF677* was significantly different among the 12 NMD factors in the low- and high-expression groups (Figure 2B).

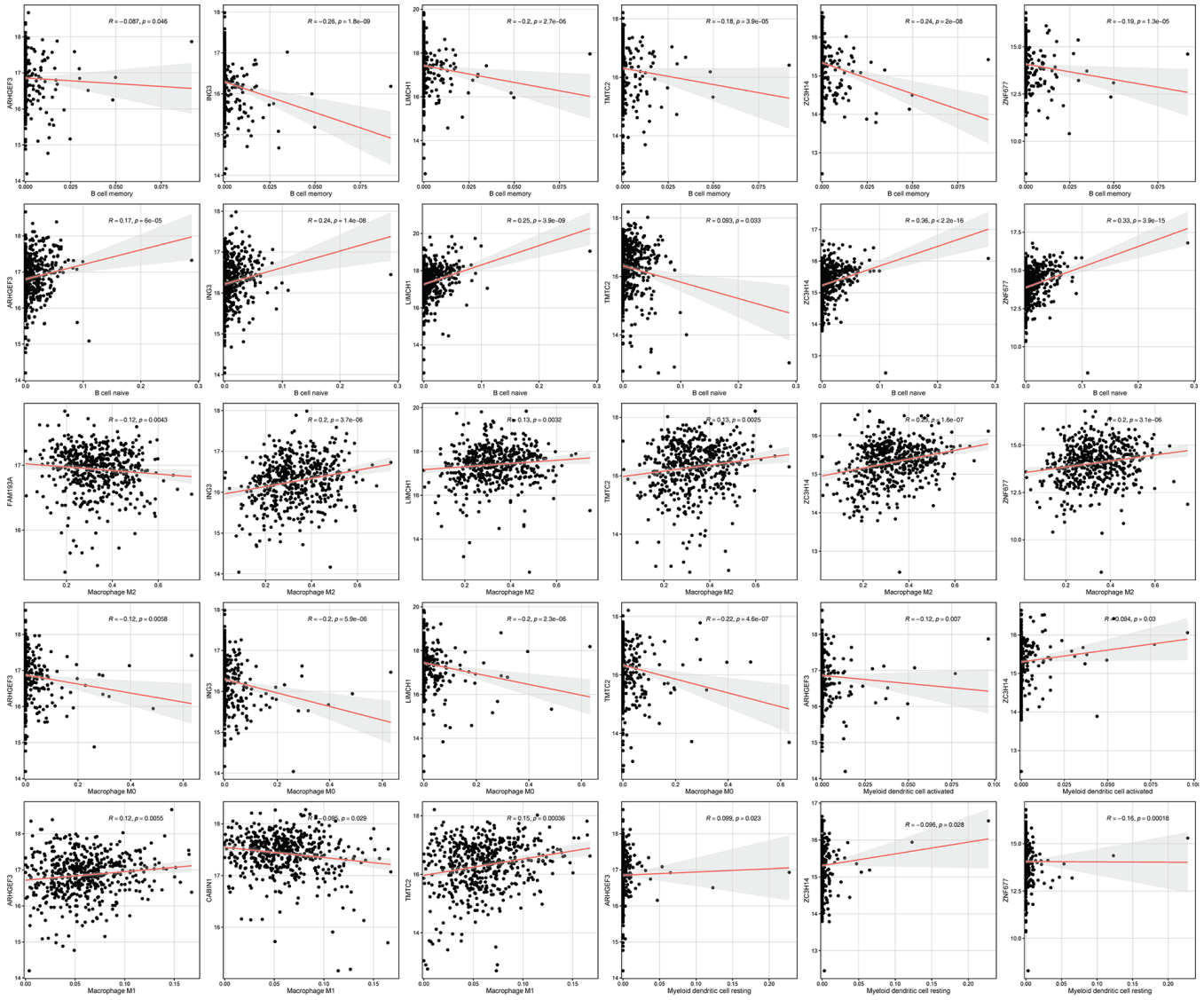**Potential antigens related to prognosis and antigen presenting cells (APCs) in ccRCC patients**

Cox regression modeling was performed on the potential antigens and OS or DFS data. A total of 510 genes were closely related to OS, whereas 40 genes were closely related to DFS. Twenty-one of these genes were common in both, and 15 potential genes were notably relevant to the survival rate. *PRPF39*, *SPG7*, *PISD*, *TUBGCP6*, *RBM6*, *SORBS2*, *PAM*, *ZFAT*, *DOCK7*, *ZNF266*, *RBM39*, *DFNA5*, *ERMAP*, *BTF3*, *GUSB* were significantly associated with OS (Supplementary Figure 2). Patients overexpressing *PRPF39, SPG7, PISD, TUBGCP6, RBM6, ZNF266, RBM39, DFNA5,* and *GUSB* in the tumor tissues had significantly shorter survival duration than patients in the low-expression group. In contrast, patients with insufficient expression of *SORBS2, PAM, ZFAT, DOCK7, ERMAP, BTF3*, and *GUSB* in the tumor tissues had shorter survival duration than patients in the high-expression group.

Furthermore, six candidate potential antigens related to NMD factors, including, *ARHGEF3*, *ING3*, *LIMCH1*, *TMTC2*, *ZC3H14*, and *ZNF677*, were closely related to APCs (Figure 3A). For example, *ARHGEF3* overexpression was notably associated with decreased infiltration of memory B cells, M0 macrophages, and activated myeloid dendritic cells in tumors; in contrast, infiltration of naïve B cells, macrophage M1 cells, and resting myeloid dendritic cells was increased. *ZNF677* overexpression was associated with decreased infiltration of memory B cells and resting myeloid dendritic cells, but increased infiltration of naïve B cells and M2 macrophages in tumors. These results suggest that we can directly process tumor antigens and present them by APCs to T cells and for recognition by B cells to elicit an immune response. These antigens hold promise as candidates for developing mRNA-based vaccines against ccRCC. Combining the aforementioned results, *ARHGEF3* and *ZNF677* were closely related to differences in the expression of NMD factors, OS, and APC function (Figure 3B–3D). According to TCIA database, ARHGEF3 is the best candidate antigen for mRNA-based vaccine for ccRCC.

Figure 2. The expression and PSI of eight potential antigens related to 12 NMD factors. (**A**) The expression of eight potential antigens between NBAS, UPF3A, SMG1, SMG5, SMG6, SMG7, SMG8, SMG9, DHX34, UPF1, UPF2, UPF3B high-expression and low-expression groups. (**B**) The PSI of eight potential antigens between NBAS, UPF3A, SMG1, SMG5, SMG6, SMG7, SMG8, SMG9, DHX34, UPF1, UPF2, UPF3B high-expression and low-expression groups. *padj < 0.05; **padj < 0.01; ***padj < 0.001.
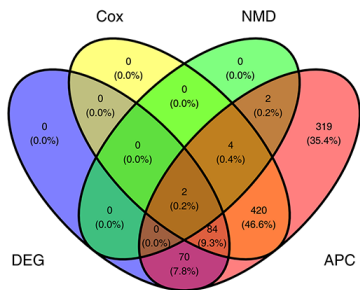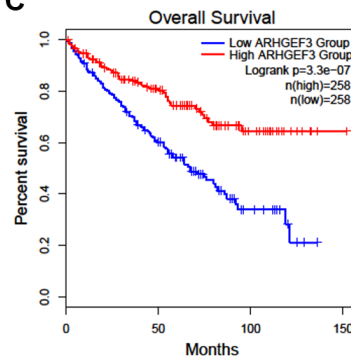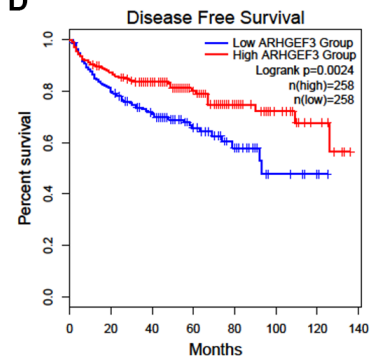
Figure 3. Potential antigens related to prognosis and antigen presenting cells (APC) in clear ccRCC patients. (**A**) ARHGEF3, ING3, LIMCH1, TMTC2, ZC3H14, and ZNF677 were closely related to APCs. (**B**) ARHGEF3 and ZNF677 were closely related to the expression difference, NMD factor difference, OS difference, and APC difference. (**C**) Effect of ARHGEF3 on overall survival rate in ccRCC patients. (**D**) Effect of ARHGEF3 on disease free survival rate in ccRCC patients.

Potential immune subtype identification in ccRCC patients

Immunotyping has been proven to be of clinical relevance in multiple tumors; it can present tumor-infiltrating immune cells and reveal the ccRCC microenvironment, helping in the identification of suitable patients to receive vaccines. The expression profiles of 577 immune-related genes were screened using the ImmPort database for consensus clustering. ccRCC patients from TCGA were clustered into two immune subtypes (C1 and C2) based on the clustering of immune-related genes (stabilized when K = 2) and their cumulative distribution function (CDF) and function delta area (Figure 4A; Supplementary Figure 3A–3C). Immune-related genes of C1/C2 could be clustered into five types as shown in the heatmap (Figure 4B). A better prognosis had a significant association with C1 (Figure 4C) and differed from that in the ICGC cohort (Figure 4D). Consistently, subtype was distributed along variable tumor grades and stages, which indicates that the C1 subtype showed a high proportion of stage I or grade 2/3 patients, whereas C2 showed a higher proportion of stage III/IV or grade 3/4 patients. ccRCC patients were divided into four molecular subtypes (KIRC1-4) according to TCGA cohort molecular classification. Interestingly, C1 accounted for the majority of KIRC1/2 subtypes, while C2 accounted for most KIRC3/4 subtypes. Subtype distribution across patients diagnosed with differential TNM pathologies was irregularly clustered; 11% of N1 patients belonged to C2, which was nearly four times the percent of patients belonging to C1 (3%) (Supplementary Figure 3D). In summary, immuno-typing is a possible prognostic marker for ccRCC patients and has a superior accuracy to the usual grading and staging, which was consistent across different cohorts.

Overexpression of *AHNAK2* and loss of *VHL* and *CDKN2A* are correlated with poor prognosis in patients with ccRCC [24, 25], and are generally used as prognostic markers for ccRCC. We analyzed the differences in the expression of *AHNAK2*, *CDKN2A*, and *VHL* between the C1 and C2 immunotypes in TCGA and ICGC cohorts. The results of TCGA cohort showed that AHNAK2 was significantly overexpressed in the C2 subtype, but the expression of *VHL* and *CDKN2A* was not significantly different (Supplementary Figure 3E). However, the results of the ICGC cohort did not agree with those of TCGA cohort, indicating that these prognostic markers may not be suitable as potential antigens due to their inconsistent expression profiles (Supplementary Figure 3F). Thus, immunotyping is superior to *VHL*, *AHNAK2*, and *CDKN2A* expression profile in prognosis prediction of ccRCC patients.

Molecular characteristics of immune subtypes

TMB is a clinically related parameter associated with immunotherapy response. We detected somatic mutations and CNV in TCGA cohort between the two immune subtypes. TMB of the C2 subtype was higher than that of C1, and a higher number of mutant genes was present in C2 (Figure 4E; Supplementary Figure 3G). Among all tested genes, including *VHL* and *PBRM1*, the frequency of missense mutations was the highest in both immune subtypes. Mutation in *PBRM1* plays a suppressing role in ccRCC [26] and occurs at a lower frequency in the C2 subtype (37%) than in the C1 subtype (44%), indicating that C2 may be more sensitive to immunotherapy (Supplementary Figure 3H, 3I). However, no significant dissimilarity in CNV between C1 and C2 was observed (Supplementary Figure 3J). A recent study reported that ccRCC tumors show a high infiltration of CD8+ T cells, favorable *PBRM1* mutation depletion, and enrichment for unfavorable chromosomal losses of 9p21.3 [27].

Similarly, the results of CNV analysis showed that CNV loss occurred frequently in chromosome 9 of the patients in the C2 subtype (Supplementary Figure 4A). Neoantigen load (NAL) is a predictive biomarker for ICI therapy. The higher NAL in C2 subtype patients suggests a higher immunogenicity of the mRNA vaccine in this subtype (Figure 4F). Chromosomal instability is the main cause of tumor evolution, and tumor cells that show chromosomal instability opt for chronically active innate immunity pathways to metastasize to remote organs [28]. Hence, ccRCC patients with the C2 subtype have a higher risk of metastasis (Figure 4G). Homologous recombination deficiency (HRD) can lead to sensitivity to poly (ADP-ribose) polymerase inhibitors and is used as a biomarker for monitoring treatment efficacy [29]. MRNAsi is an index that can be used for quantification of stemlike indices and has a significant correlation with patient outcomes, and high mRNAsi is associated with upregulation of immunosuppressive checkpoints [30]. Unfortunately, we could not determine the difference in HRD and mRNAsi between the two subtypes (Supplementary Figure 4B, 4C).

We detected the AS of 901 potential neoantigens between the two subtypes. PSI difference was noted in 337 potential neoantigens (37.4%) between the two subtypes. The immune system induces diverse repertoire of antigen receptors, which are expressed by B and T cells and are capable of recognizing a variety of protein antigens [31]. Downregulated expression ICP-related genes can sensitize cells to immunotherapy, and immunogenic cell death (ICD) can stimulate the dysfunctional anti-tumor immune system. In TCGA
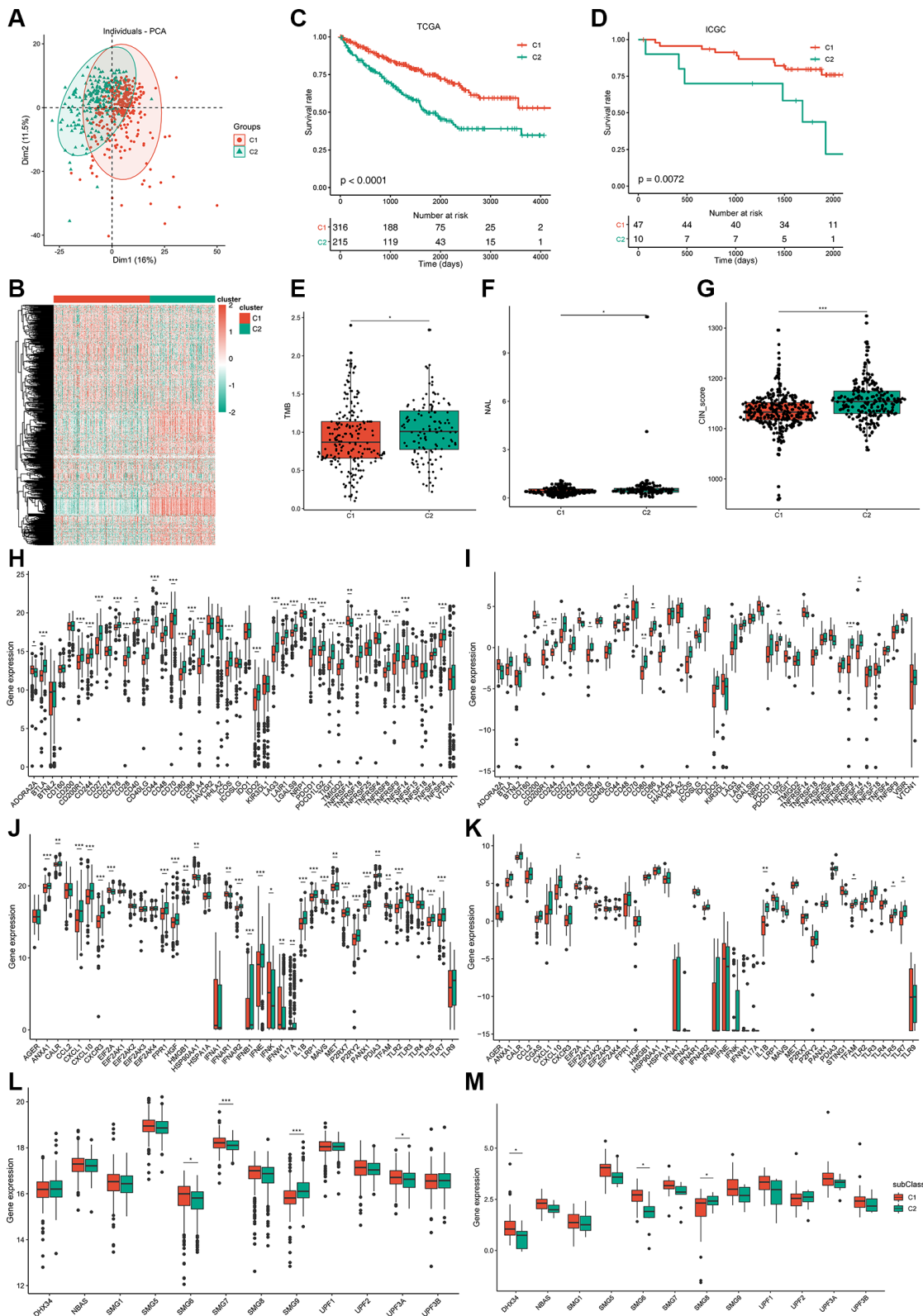
Figure 4. Identification of potential immune subtypes of ccRCC patients and molecular characteristics. (**A**) PCA of ccRCC patients in TCGA cohort. (**B**) Heatmap of immune-related genes of C1/C2 ccRCC patients. (**C**, **D**) Kaplan-Meier curves showing overall survival (OS) of ccRCC immune subtypes in TCGA and ICGC cohorts. (**E**) TMB and number of mutated genes in C1/C2 ccRCC patients. (**F**, **G**) NAL and CIN score of C1/C2 ccRCC patients. (**H**, **I**) Expression of ICP genes in C1/C2 ccRCC patients from TCGA cohort and ICGC cohort separately. (**J**, **K**) Expression of ICD factors in C1/C2 ccRCC patients from TCGA cohort and ICGC cohort separately. (**L**, **M**) Expression of NMD factors in C1/C2 ccRCC patients from TCGA cohort and ICGC cohort separately. *$p$adj < 0.05; **$p$adj < 0.01; ***$p$adj < 0.001.

cohort, significant and differential expression of 32 ICP genes were found between the two subtypes (Figure 4H), and most of them had a higher expression in the C2 subtype.

We also verified 10 ICP genes that had differential expression between the two subtypes in the ICGC cohort (Figure 4I). A total of 29 ICD factors and four NMD factors were significantly and differentially expressed between the two subtypes from TCGA cohort (Figure 4J–4L), and five ICD factors and four NMD factors were verified in the ICGC cohort (Figure 4K–4M). Therefore, reflection of the expression levels of ICP and ICD modulators can be achieved through immunotyping, which is potentially a therapeutic biomarker for mRNA vaccines.
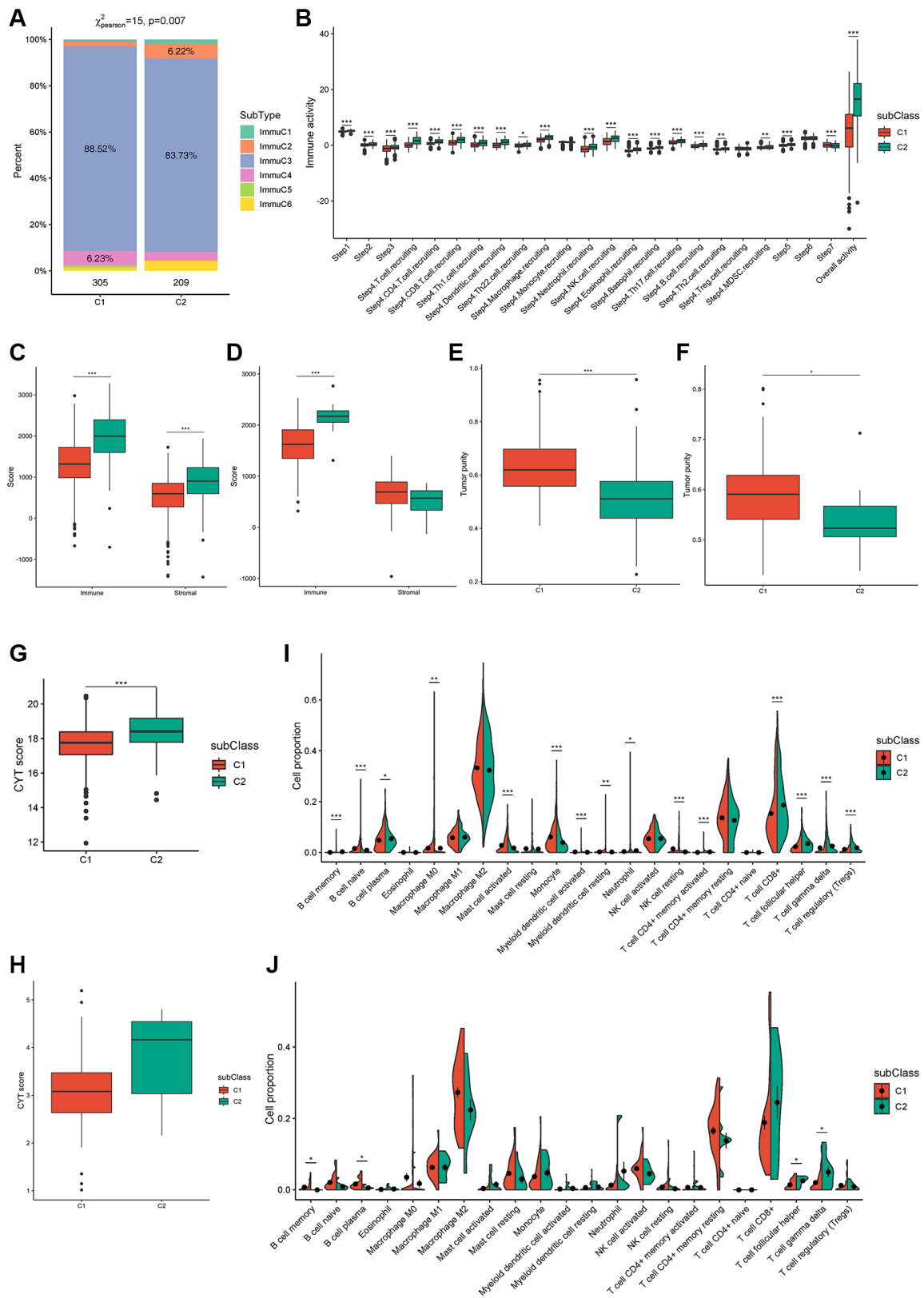
Cellular characteristics of immune subtypes

To demonstrate immunotyping reliability, we analyzed the percentage of six earlier demonstrated pan-cancer immune subtypes (ImmuC1-ImmuC6) among the C1 and C2 immunotypes. The proportions of pan-cancer immuC3 were 88.52% and 83.73% in the C1 and C2 immunotypes, respectively, and immuC4 was the secondary subtype (6.23%) in the C1 immunotype, whereas immuC2 was the secondary subtype (6.22%) in the C2 immunotype. Compared to the C1 immunotype, the lack of immuC5 and abundance of immuC6 in the C2 immunotype may induce some differences between them (Figure 5A). These outcomes correspond to the higher survival observed in the C1 immunotype tumors than in the C2 immunotype tumors. Tumor immuno-phenotype systematic tracking is required for under-standing cancer immunity fundamental mechanisms and improvement of cancer immunotherapy clinical benefits [32]. As shown in Figure 5B, the overall activity of immune cells in the C2 group was significantly higher than that in the C1 group, and most anti-cancer immune response steps correlated with this observation. In particular, the difference in Th2 cell-recruiting activity during the trafficking of immune cells to tumors was consistent with other reports, which showed that the increase in the expression of type-2 T helper cell signature was associated with low survival in ccRCC, papillary RCC, and chromophobe RCC [24].

In patients, immune and stromal scores can help determine tumor purity and immune cell infiltration in the tumor microenvironment, with high immune scores indicating cytolytic good immune responses and better prognosis [33, 34]. Consistently, immune and stromal scores were remarkably higher in the C2 subtype in TCGA cohort (Figure 5C) and the immune scores of the two subtypes were verified in the ICGC cohort (Figure

5D). In addition, the tumor purity of C2 was remarkably lower than that of C1 in both TCGA and ICGC cohorts (Figure 5E, 5F). CYT reflects the cell-killing function and can be used to assess immune-mediated attacks against cancer cells; moreover, it is associated with the mutational burden [35]. The CYT score from TCGA and ICGC cohorts indicated a higher mutational burden and more complex immune microenvironment in the C2 subtype (Figure 5G, 5H). The proportion of 15 types of immune cells, comprising memory B cells, naïve B cells, plasma B cells, M0 macrophages, activated mast cells, monocytes, activated myeloid dendritic cells, resting myeloid dendritic cells, neutrophils, resting natural killer calls (NKCs), activated CD4+ memory T cells, CD8+ T cells, follicular helper T cells, gamma delta T cells, and regulatory T cells (Tregs) was remarkably different between the two subtypes in TCGA cohort (Figure 5I). Although memory B cells, plasma B cells, follicular helper T cells, and gamma delta T cells were verified in the ICGC cohort, some other immune cell types, such as CD8+ T cells, resting NKCs, and activated CD4+ memory T cells showed the same trend (Figure 5J). These results indicated that patients with the C2 subtype had a significant tumor infiltration of CD8+ T cells, [27] suggesting that these patients were more suitable for our mRNA vaccine. Taken together, these results suggest that the immune subtypes are promising candidates for mRNA vaccines, and patients with C2 tumors with/without an immunosuppressive microenvironment are potentially more suitable for mRNA vaccination.

ccRCC immune landscape

The immune landscape of ccRCC was created from 28 immune cell types of pan-cancer gene expression profiles using ssGSEA and monocle (Figure 6A). C1 integral distribution was opposite to that of C2. The correlation between PCA1/2 and 28 types of pan-cancer immune cells is shown in Figure 6B. The pan-cancer immune cells did not show any positive correlation with PCA1 expression. Nevertheless, activated and immature B cells, myeloid-derived suppressor cells, and T follicular helper cells were the most negatively correlated immune cells. Immature dendritic cells, mast cells, and memory B cells showed the highest positive correlation with PCA2. Moreover, differential distribution was displayed in the same subtype, indicating significant intra-cluster heterogeneity within the subtypes. Based on the location of immune cell populations, C1 was further divided into four subsets defined as C1a, C1b, C1c, and C1d, and C2 was further divided into three subsets separately defined as C2a, C2b, and C2c (Figure 6C). The enrichment score of several immune cells showed significant difference between every subset (Figure 6D). C1c and C2c showed

Figure 5. Cellular characteristics of immune subtypes. (**A**) Distribution of immune subtypes (ImmuC1-C6) across C1/C2 ccRCC patients. (**B**) Overall activity of immune cells in C1/C2 ccRCC patients. (**C**, **D**) Immune score and stromal score of C1/C2 ccRCC patients from TCGA and ICGC cohorts. (**E**–**H**) Tumor purity and CYT score of C1/C2 ccRCC patients from TCGA (**E**, **G**) and ICGC (**F**, **H**) cohorts. (**I**, **J**) Immune cell proportions of C1/C2 ccRCC patients from TCGA and ICGC cohorts separately. *$p$adj < 0.05; **$p$adj < 0.01; ***$p$adj < 0.001.
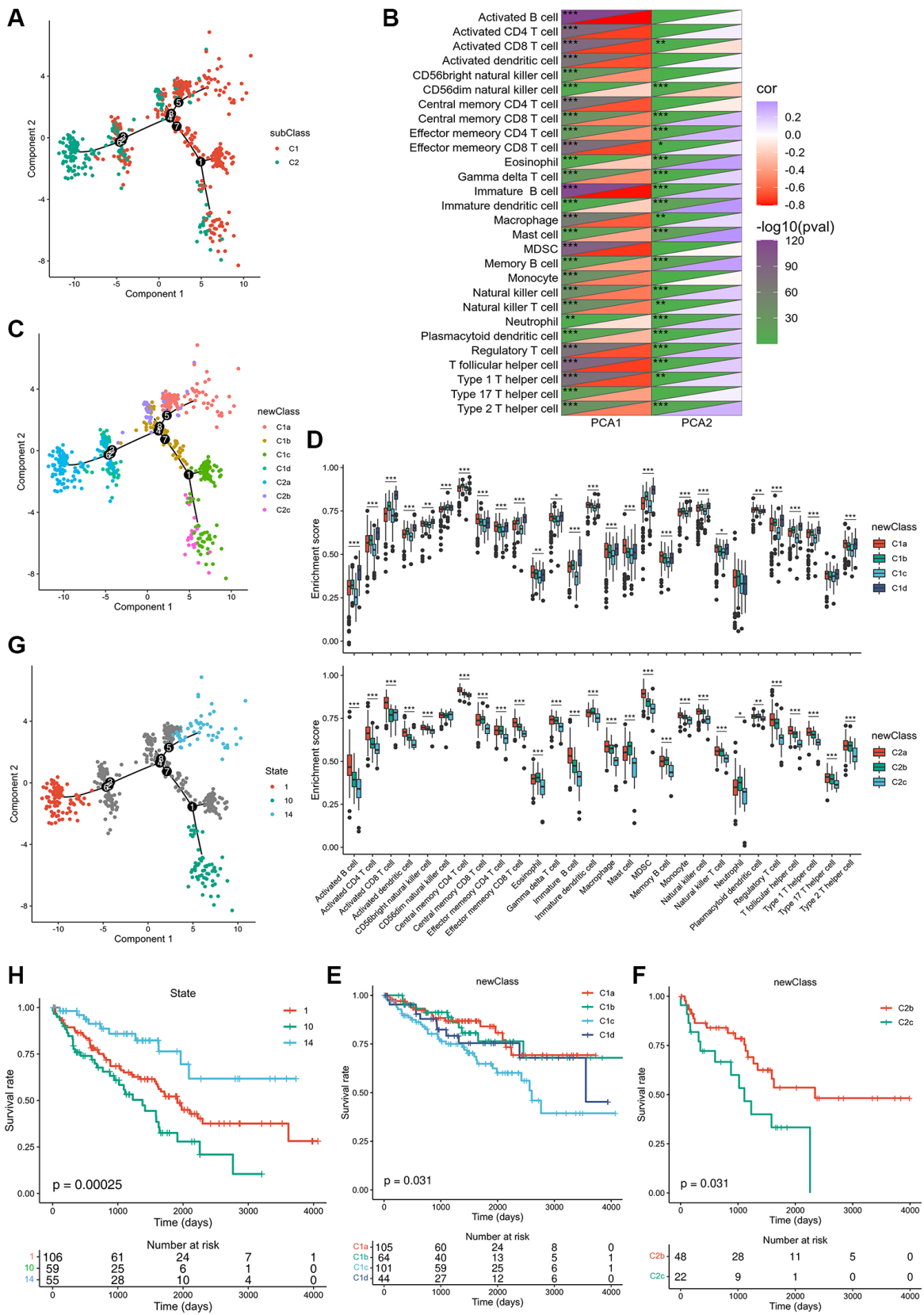
Figure 6. Immune landscape of ccRCC. (**A**) Immune landscape of ccRCC. (**B**) Heat map of two principal components with 28 immune cell signatures. (**C**) Immune landscape of the subsets of ccRCC immune subtypes. (**D**) Differential enrichment scores of 28 immune cell signatures in the above subsets. (**E**, **F**) The prognostic status of each subset of C1 and C2 separately. (**G**) Immune landscape of samples from three extreme locations and (**H**) their prognostic status. *$p$adj < 0.05; **$p$adj < 0.01; ***$p$adj < 0.001.

lower enrichment scores for most immune cells, such as activated B cells, activated CD4 T cells, activated CD8 T cells, central memory CD8 T cells, effector memory CD4 T cells, effector memory CD8 T cells, and memory B cells, suggesting that patients with these subsets may be more suitable for mRNA vaccination. C1a subtype patients showed the highest survival rates, while subtypes of C1c and C2c showed the worst survival rates (Figure 6E, 6F). In addition, samples that showed utmost immune landscape distribution positions were compared in terms of prognosis, and the probability of survival was the best in a group 14 of patients, consistent with the aforementioned results (Figure 6G, 6H). Taken together, the immune landscape based on immune subtypes can be used for identification of immune components of each ccRCC patient and prediction of their prognoses and can preferably be used to provide a personalized mRNA vaccine therapy.

ccRCC hub genes and gene co-expression module identification

TCGA co-expression module identification was performed by sample clustering based on WGCNA (Supplementary Figure 5A) with a soft threshold of 4 for a scale-free network (Supplementary Figure 5B, 5C). Nine co-expression modules obtained from TCGA are shown in black, blue, green, pink, red, turquoise, yellow, gray, and brown (Figure 7A; Supplementary Figure 5D). The expression of the black, blue, green, pink, red, turquoise, and yellow module eigengenes was significantly different between the two subtypes. C2 showed higher eigengenes in the black, blue, pink, and turquoise modules and lower eigengenes in the green, red, and yellow modules (Figure 7B). Furthermore, the black, gray, red, and yellow modules showed significant positive correlations with prognosis, whereas the green and turquoise modules were notably negatively correlated with prognosis (Figure 7C). GO enrichment of the green, black, gray, and turquoise modules indicated that the prognostic modules were related to small molecule catabolic processes, nuclear division, organelle fission, and epidermis development. Regulation of membrane potential, kidney epithelium development, and renal tubule development were enriched in the KEGG pathway, indicating that the prognostic modules were closely linked to kidney health (Supplementary Figure 5E).
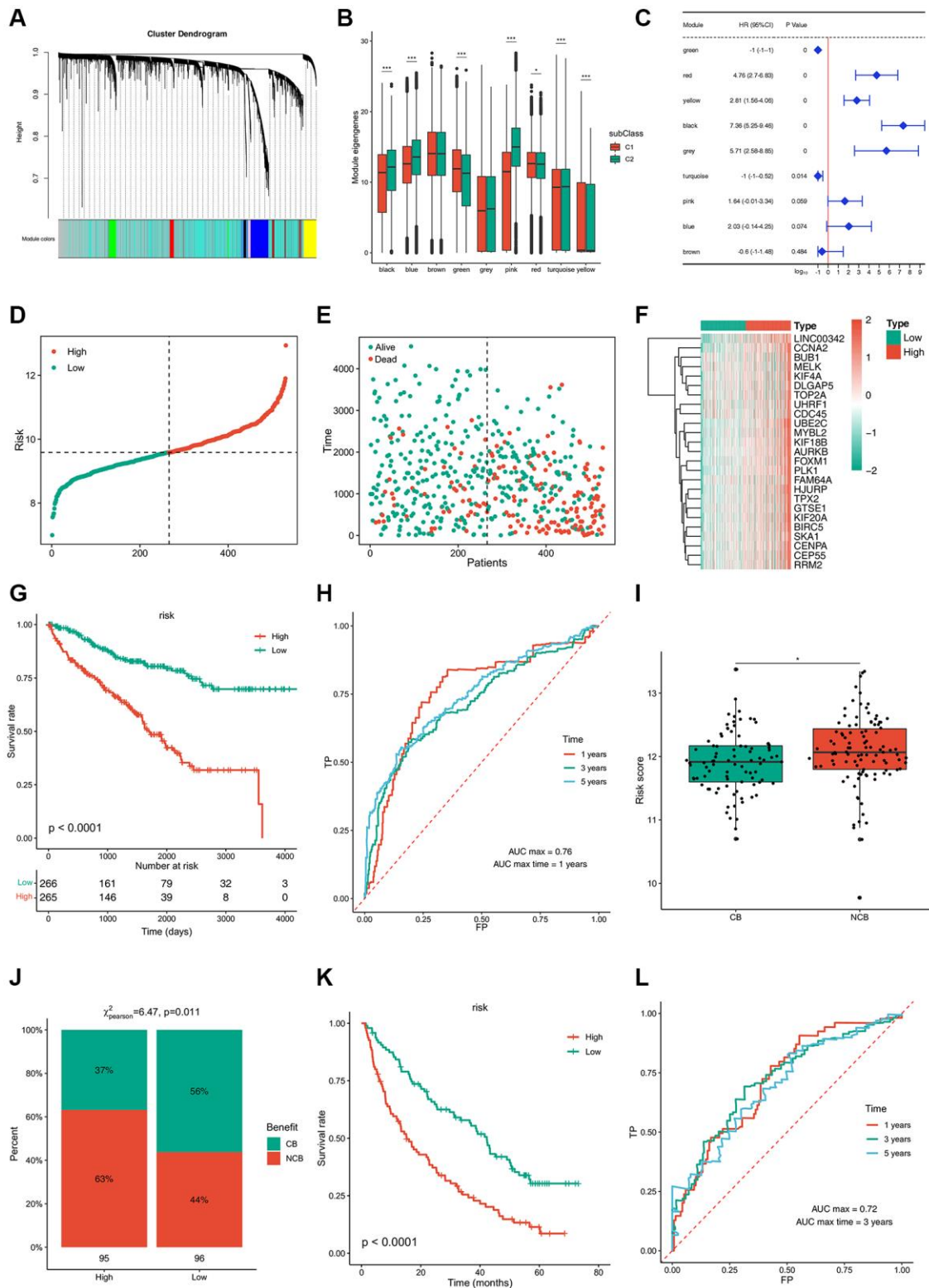
A risk estimation model was established to predict tumor risk according to the expression of 25 hub genes that were obtained from prognostic modules (Figure 7D–7F). Low-risk patients showed a better survival rate (Figure 7G). The time-dependent receiver operating characteristic (ROC) curve suggested that this model

can accurately predict the 1- to 5-year survival rate of ccRCC patients (Figure 7H). Thus, this model can be used to identify patients who can benefit from mRNA vaccines through risk scoring and provide a reference for immunotherapy. Next, 592 patients treated with programmed cell death protein 1 (PD-1) blockade were divided into clinical benefit (CB) and no clinical benefit (NCB) groups; NCB patients showed a higher risk score (Figure 7I). In contrast, 56% of patients with CB were concentrated in the low-risk group, while 63% of patients in the high-risk group had a poor survival rate (Figure 7J, 7K). The time-dependent ROC curve suggested that this model can accurately predict 1- to 5-year survival rate of ccRCC patients treated with PD-1 blockade (Figure 7L). Collectively, these results provide evidence for the accuracy of the risk prediction model, based on which, more patients can benefit from mRNA vaccines.

DISCUSSION

The study aimed to identify potential tumor antigens and immune subtypes of ccRCC for mRNA vaccine development. We constructed the aberrant AS and mutational landscape of ccRCC and identified *ARHGEF3* as a promising mRNA vaccine candidate. *RhoGEFs* play critical roles in various signaling cascades and cellular processes and are involved in multiple cancers. *ARHGEF3* downregulation may be associated with invasion, metastasis, and proliferation in osteosarcoma because this gene specifically activates *RHOA* and *RHOB*, which play a role in bone cell biology [36, 37]. According to the Human Protein Atlas database, high *ARHGEF3* expression is favorable for patient survival and potentially a prognostic marker for renal cancer, endometrial cancer, and head and neck cancer, and others. In contrast, elevated *ARHGEF3* expression is a crucial contributor to the pathogenesis of nasopharyngeal carcinoma (NPC) by inhibiting cell apoptosis and potentially used as a novel marker for prognosis and an effective target for treatment [38]. Alternatively, spliced transcript variants encoding different isoforms have been found for *ARHGEF3*, suggesting that aberrant splicing isoforms are responsible for the diverse effects of *ARHGEF3* in various tumors. According to TCGA SpliceSeq database, the PSI of the alternate promoter from exon 7 in ccRCC patients was twice as high as normal. Hence, the mRNA vaccine of *ARHGEF3* for ccRCC patients was designed to target this specific splicing isoform.

The aberrant splicing transcript of *ARHGEF3* not only had an association with bad prognosis but also with higher APCs and infiltration of B cell. Given that the benefit of mRNA vaccines is applicable for some cancer patients only, we divided ccRCC patients into two

Figure 7. Immune gene co-expression modules and immune hub genes for ccRCC identification and the established risk-estimating model. (**A**) Dendrogram of all differentially expressed genes clustered based on a dissimilarity measure (1-TOM). Clustering of samples. (**B**) Differential distribution of feature vectors of each module in ccRCC subtypes. (**C**) Forest maps of single-factor survival analysis of nine modules of ccRCC. (**D–F**) Risk-estimating model was established from prognostic modules and (**G**) their prognostic status from TCGA cohort. (**H**) ROC curve of risk of ccRCC patients from TCGA cohort. (**I, J**) Risk score of ccRCC patients treated with PD-1 and (**K**) their prognostic status from TCGA cohort. (**L**) ROC curve of risk of ccRCC patients treated with PD-1 from TCGA cohort. *$p$adj < 0.05; **$p$adj < 0.01; ***$p$adj < 0.001.

groups (C1 and C2) according to 577 immune-related genes. The two immune groups showed distinct molecular, cellular, and clinical characteristics. The worse prognosis of patients with the C2 subtype in both ICGC and TCGA cohorts indicates that we can use immunotyping as prognostic marker in patients with ccRCC as it is superior to traditional grading and staging in terms of accuracy. Further, the prognostic performance of immune subtype is superior to that of traditional tumor markers (*VHL*, *AHNAK2*, and *CDKN2A*) in ccRCC patients. As a promising predictor of immune checkpoint inhibitor efficacy, the tumor's ability to produce new antigens is possibly reflected indirectly by TMB. A high TMB is an indicator supporting the use of monotherapy in children and adults with solid tumors. In addition to its prognostic potential, immunotyping can predict mRNA vaccine therapeutic response. The molecular characteristics of immune subtypes indicated that the TMB, number of mutated genes, NAL, and chromosomal instability in patients with C2 tumors were higher than those in patients with C1 tumors, suggesting that ccRCC patients with the C2 subtype may respond better to mRNA vaccines.

ccRCC poses a challenge to known standards in cancer immunology because of the modest mutation burden; however, it responds to immunotherapies, and a higher CD8+ T cell infiltration has traditionally been linked to poor prognosis [39, 40]. Aberrant splicing transcripts of ARHGEF3 play a critical role in ccRCC progression and development and can be processed directly and presented to CD8+ T cells to elicit an immune response. However, the high expression of partial ICPs, such as *CD200R1*, *CD244*, *CD28*, *CD48*, *CD80*, *CD86*, *ICOS*, *PDCD1LG2*, *TNFRSF9*, and *TNFSF14*, in C2 tumors suggests that the tumor microenvironment is immuno-suppressive; thus, mRNA vaccine inhibition may be caused by eliciting an effective immune response. Further, the response to ICIs is related to the depletion of *PBRM1* mutations and enrichment of chromosomal losses of 9p21.3, [27] while the elevated expression of partial ICD modulators, such as *EIF2A*, *IL1B*, *TFAM*, *TLR5*, and *TLR7*, in C2 tumors suggests that mRNA vaccines are effective in these immune subtypes. Moreover, the immunological complexity of the ccRCC landscape considerably suggests heterogeneity between individual cases and within the same immune subtype, thus decreasing the immune components needed to develop personalized mRNA vaccine-based therapeutics. Patients in group 10 showed the worst prognosis, consistent with the results for the C1c and C2c subsets. C1c and C2c showed significantly lower scores for activated CD8 T cells, eosinophils, activated B cells, effector memory CD4 T cells, effector memory CD8 T cells, and monocytes than for other subsets, indicating that C1c and C2c are immunologically "hot," and

therefore, should respond better to mRNA vaccines. Furthermore, the risk estimation model established based on 25 hub genes of prognostic modules can be used to accurately determine CB ccRCC patients and enable more patients to benefit from mRNA vaccines.

In conclusion, ARHGEF3 is a potential ccRCC antigen for mRNA vaccine development. Candidates with immune subtype 2, especially C2c, were the most suitable for vaccination. This study provides a theoretical basis for developing anti-ccRCC mRNA vaccines, predicting patient outcomes, and selecting vaccination candidates. We recommend further study and validation of the identified vaccine antigens and other prognostic markers. Moreover, we are attempting to verify the potential application of ARGEF3 as a promising vaccine antigen. By constructing an aberrant AS *AHRGEF3* cell line, cell-line-derived xenograft, and ccRCC patient-derived tumor xenograft, we intent to understand how aberrant AS *AHRGEF3* induces changes in the tumor immune environment in ccRCC. Moreover, the intrinsic connection between the novel ccRCC antigen *ARHGEF3* and immune subtype warrants further studies. Similar investigations should be conducted for other cancer types as well.

MATERIALS AND METHODS

Data acquisition and preprocessing

Percent spliced in (PSI) values for the splice events of all genes and samples of kidney renal clear carcinoma (KIRC) were downloaded from TCGA SpliceSeq database. MuTect2 Variant Aggregation and Masking ($n = 336$) GDC Hub, Phenotype ($n = 985$) GDC Hub, GISTIC - focal score by gene ($n = 536$) GDC Hub, and gene expression RNAseq ($n = 607$) were obtained from the clinical and follow-up information from the XENA TCGA-KIRC database. The clinical characteristics of the samples from TCGA are shown in Supplementary Table 2. The RECA-EU project data were obtained from the ICGC cohort.

AS and enrichment analysis

The difference in PSI between ccRCC and normal control was determined using the t-test, and significance was set at $p < 0.05$. Gene ontology (GO) enrichment analysis of the target gene set was performed by the ClusterProfiler package [41] in R (version "4.1").

Gene expression analysis

Gene expression data (FPKM values) from RNAseq ($n = 607$) of ccRCC and normal control from TCGA-

KIRC database were compared using the limma package in R (version "4.1") [42] and filtered using |LogFC| > 1 and $p < 0.01$ as threshold criteria.

Cox proportional-hazards model and survival analysis

The Cox proportional-hazards model and survival analysis were used for regression modeling of the characteristic gene and OS or disease-free survival (DFS) data using the coxph function in the survival package [43] in R (version "4.1") and filters with $p < 0.05$ as the threshold. The effect sizes were displayed in a forest plot using corresponding modeling parameters. For multifactor Cox analysis, multigene pairs were built using the coxph function of the survival package for regression modeling. The genes identified in the survival analysis were grouped after screening for the median expression level, and the OS was fitted using the survfit function of the survival package. Analysis and visualization were performed using the ggsurvplot function of the Survminer package in R (version "4.1").

TIMER2 analysis

Tumor Immune Estimation Resource [42] was employed for analysis and visualization of the association between tumor-infiltrating cells (TICs) and ccRCC-related genes using analytical modules for somatic mutations, gene expression, clinical outcomes, and somatic copy number alterations. Adjustment of purity was performed by Spearman's correlation analysis. Statistical significance was set at $p < 0.05$.

Consensus clustering analysis

Screening of the immune gene set downloaded from ImPort using single-factor Cox yielded 577 immune-related genes associated with prognosis. The prognosis-related immunogenetic expression data of TCGA-KIRC patients were screened and normalized with the median, and then analysed using the R package Consensus ClusterPlus for Consensus clustering. The clustering algorithm was k-means algorithm, and sampling was performed 500 times.

Analysis of somatic mutation and copy number variation (CNV) between immune subtypes

We downloaded TCGA-KIRC somatic mutation maf file from the TCGA, calculated the TMB using the maftools package in R, and visualized the mutation type. The frequency of CNV gain or loss was calculated through GISTIC file from the Xena database, visualized using maftools.

Analysis of intertype immune activity

Data for the immune activity of TCGA-KIRC were downloaded from the TIP database, and the t-test was conducted for testing differences in the immune activity scores of different subtypes.

Homologous recombination defects, new antigen loads, chromosomal instability, and dryness index analysis

The DNA damage repair (DDR) score was downloaded from https://gdc.cancer.gov/about-data/publications/PanCan-DDR-2018 and The Cancer Immunome Atlas (TCIA) database. Maftools was used to calculate the new antigen load TMB from TCGA-KIRC. A list of unstable chromosomal genes was obtained from a published paper [44], and the sum of the expression values of these genes was used as the chromosomal instability score. mRNAsi was used to calculate the dryness index of tumor subtypes according to a previously published paper [PMID:29625051], and t-test was used to calculate the differences between the four indicators.

Immune cell analysis

The immune, matrix, and tumor purity evaluations of different TCGA-KIRC subtypes were performed using R-package estimate, and the expression means of the genes granzyme A and perforin 1 were used as immune cytolytic activity (CYT) scores. The immuno-immersion score file from TCGA was downloaded from the TIMER2 database, the data related to TCGA-KIRC sample was included, the CIBERSORT score data were analyzed, the immune immersion score of test data set was calculated using CIBERSORT, and a Wilcox-test/t-test was used to calculate the differences in the immunization score index.

Immune landscape of ccRCC

Data of 28 pancellular immune cell types were downloaded from a published paper, [45] using the sSGSEA algorithm to analyze the relative abundance of each immune cell type in TCGA-KIRC. The DDRTree dimension reduction of TCGA-KIRC was then carried out using the monocle package, and the correlation between the first two main components and 28 pancellular immune cell types was calculated. The subtype was subdivided according to the pseudotime analysis of TCGA-KIRC samples. The relative abundance difference of 28 pancellular immune cell types in each subcategory was determined using single-factor analysis of variance (ANOVA).

WGCNA co-expression network construction

The expression level of differentially expressed genes in TCGA-KIRC was screened, and the co-expression network of TCGA-KIRC expression matrix was constructed using the WGCNA package. We set 0.8 as the threshold for the average connectivity of the scale-free fit index and network with the change of the soft threshold parameter, when soft threshold (softpower) changed from 1 to 10. The softpower was set to 4, such that it met the scale-less network, a co-expression network was built, and the prognosis module was identified, wherein the minimum number of genes of the module was 30 and the combined correlation was greater than 0.7. The Pearson correlation coefficient between the genes in the module and the module feature genes was calculated. $|cor| > 0.9$ and $p < 0.001$ was used to identify hub genes, and a risk score was determined according to a multivariate Cox model of the hub genes.

Data availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

AUTHOR CONTRIBUTIONS

D.Z. wrote the manuscript. D.Z. and D.X. and W.X. and Y.M. designed experiments. J.Y. and M.Z. and M.S. acquired, analyzed, and interpreted data. D.Z. and J.Y. and M.Z. edited the manuscript.

ACKNOWLEDGMENTS

CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

FUNDING

REFERENCES

1. Hsieh JJ, Purdue MP, Signoretti S, Swanton C, Albiges L, Schmidinger M, Heng DY, Larkin J, Ficarra V. Renal cell carcinoma. Nat Rev Dis Primers. 2017; 3:17009. https://doi.org/10.1038/nrdp.2017.9 PMID:28276433

2. Meskawi M, Sun M, Trinh QD, Bianchi M, Hansen J, Tian Z, Rink M, Ismail S, Shariat SF, Montorsi F, Perrotte P, Karakiewicz PI. A review of integrated staging systems for renal cell carcinoma. Eur Urol. 2012; 62:303–14. https://doi.org/10.1016/j.eururo.2012.04.049 PMID:22575911

3. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature. 2013; 499:43–9. https://doi.org/10.1038/nature12222 PMID:23792563

4. Hakimi AA, Chen YB, Wren J, Gonen M, Abdel-Wahab O, Heguy A, Liu H, Takeda S, Tickoo SK, Reuter VE, Voss MH, Motzer RJ, Coleman JA, et al. Clinical and pathologic impact of select chromatin-modulating tumor suppressors in clear cell renal cell carcinoma. Eur Urol. 2013; 63:848–54. https://doi.org/10.1016/j.eururo.2012.09.005 PMID:23036577

5. Peña-Llopis S, Vega-Rubín-de-Celis S, Liao A, Leng N, Pavía-Jiménez A, Wang S, Yamasaki T, Zhrebker L, Sivanand S, Spence P, Kinch L, Hambuch T, Jain S, et al. BAP1 loss defines a new class of renal cell carcinoma. Nat Genet. 2012; 44:751–9. https://doi.org/10.1038/ng.2323 PMID:22683710

6. Xu J, Pham CG, Albanese SK, Dong Y, Oyama T, Lee CH, Rodrik-Outmezguine V, Yao Z, Han S, Chen D, Parton DL, Chodera JD, Rosen N, et al. Mechanistically distinct cancer-associated mTOR activation clusters predict sensitivity to rapamycin. J Clin Invest. 2016; 126:3526–40. https://doi.org/10.1172/JCI86120 PMID:27482884

7. Wei EY, Hsieh JJ. A river model to map convergent cancer evolution and guide therapy in RCC. Nat Rev Urol. 2015; 12:706–12. https://doi.org/10.1038/nrurol.2015.260 PMID:26526752

8. Motzer RJ, Hutson TE, Cella D, Reeves J, Hawkins R, Guo J, Nathan P, Staehler M, de Souza P, Merchan JR, Boleti E, Fife K, Jin J, et al. Pazopanib versus sunitinib in metastatic renal-cell carcinoma. N Engl J Med. 2013; 369:722–31. https://doi.org/10.1056/NEJMoa1303989 PMID:23964934

9. Escudier B, Pluzanska A, Koralewski P, Ravaud A, Bracarda S, Szczylik C, Chevreau C, Filipek M, Melichar B, Bajetta E, Gorbunova V, Bay JO, Bodrogi I, et al, and

AVOREN Trial investigators. Bevacizumab plus interferon alfa-2a for treatment of metastatic renal cell carcinoma: a randomised, double-blind phase III trial. Lancet. 2007; 370:2103–11.
https://doi.org/10.1016/S0140-6736(07)61904-7
PMID:18156031

10. Motzer RJ, Escudier B, Oudard S, Hutson TE, Porta C, Bracarda S, Grünwald V, Thompson JA, Figlin RA, Hollaender N, Urbanowitz G, Berg WJ, Kay A, et al, and RECORD-1 Study Group. Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomised, placebo-controlled phase III trial. Lancet. 2008; 372:449–56.
https://doi.org/10.1016/S0140-6736(08)61039-9
PMID:18653228

11. Motzer RJ, Escudier B, McDermott DF, George S, Hammers HJ, Srinivas S, Tykodi SS, Sosman JA, Procopio G, Plimack ER, Castellano D, Choueiri TK, Gurney H, et al, and CheckMate 025 Investigators. Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma. N Engl J Med. 2015; 373:1803–13.
https://doi.org/10.1056/NEJMoa1510665
PMID:26406148

12. Motzer RJ, Agarwal N, Beard C, Bolger GB, Boston B, Carducci MA, Choueiri TK, Figlin RA, Fishman M, Hancock SL, Hudes GR, Jonasch E, Kessinger A, et al. NCCN clinical practice guidelines in oncology: kidney cancer. J Natl Compr Canc Netw. 2009; 7:618–30.
https://doi.org/10.6004/jnccn.2009.0043
PMID:19555584

13. Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, Barron DA, Zehir A, Jordan EJ, Omuro A, Kaley TJ, Kendall SM, Motzer RJ, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. Nat Genet. 2019; 51:202–6.
https://doi.org/10.1038/s41588-018-0312-8
PMID:30643254

14. Di C, Syafrizayanti, Zhang Q, Chen Y, Wang Y, Zhang X, Liu Y, Sun C, Zhang H, Hoheisel JD. Function, clinical application, and strategies of Pre-mRNA splicing in cancer. Cell Death Differ. 2019; 26:1181–94.
https://doi.org/10.1038/s41418-018-0231-3
PMID:30464224

15. Baker KE, Parker R. Nonsense-mediated mRNA decay: terminating erroneous gene expression. Curr Opin Cell Biol. 2004; 16:293–9.
https://doi.org/10.1016/j.ceb.2004.03.003
PMID:15145354

16. Oka M, Xu L, Suzuki T, Yoshikawa T, Sakamoto H, Uemura H, Yoshizawa AC, Suzuki Y, Nakatsura T, Ishihama Y, Suzuki A, Seki M. Aberrant splicing isoforms detected by full-length transcriptome sequencing as transcripts of potential neoantigens in non-small cell lung cancer. Genome Biol. 2021; 22:9.
https://doi.org/10.1186/s13059-020-02240-8
PMID:33397462

17. Chan TA, Yarchoan M, Jaffee E, Swanton C, Quezada SA, Stenzinger A, Peters S. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. Ann Oncol. 2019; 30:44–56.
https://doi.org/10.1093/annonc/mdy495
PMID:30395155

18. Carreno BM, Magrini V, Becker-Hapak M, Kaabinejadian S, Hundal J, Petti AA, Ly A, Lie WR, Hildebrand WH, Mardis ER, Linette GP. Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. Science. 2015; 348:803–8.
https://doi.org/10.1126/science.aaa3828
PMID:25837513

19. Litchfield K, Reading JL, Lim EL, Xu H, Liu P, Al-Bakir M, Wong YNS, Rowan A, Funt SA, Merghoub T, Perkins D, Lauss M, Svane IM, et al. Escape from nonsense-mediated decay associates with anti-tumor immunogenicity. Nat Commun. 2020; 11:3800.
https://doi.org/10.1038/s41467-020-17526-5
PMID:32733040

20. Pardi N, Hogan MJ, Porter FW, Weissman D. mRNA vaccines - a new era in vaccinology. Nat Rev Drug Discov. 2018; 17:261–79.
https://doi.org/10.1038/nrd.2017.243
PMID:29326426

21. Türeci Ö, Vormehr M, Diken M, Kreiter S, Huber C, Sahin U. Targeting the Heterogeneity of Cancer with Individualized Neoepitope Vaccines. Clin Cancer Res. 2016; 22:1885–96.
https://doi.org/10.1158/1078-0432.CCR-15-1509
PMID:27084742

22. Berget SM. Exon recognition in vertebrate splicing. J Biol Chem. 1995; 270:2411–4.
https://doi.org/10.1074/jbc.270.6.2411
PMID:7852296

23. Gnarra JR, Tory K, Weng Y, Schmidt L, Wei MH, Li H, Latif F, Liu S, Chen F, Duh FM. Mutations of the VHL tumour suppressor gene in renal carcinoma. Nat Genet. 1994; 7:85–90.
https://doi.org/10.1038/ng0594-85
PMID:7915601

24. Linehan WM, Ricketts CJ. The Cancer Genome Atlas of renal cell carcinoma: findings and clinical implications. Nat Rev Urol. 2019; 16:539–52.
https://doi.org/10.1038/s41585-019-0211-5
PMID:31278395

25. Wang M, Li X, Zhang J, Yang Q, Chen W, Jin W, Huang YR, Yang R, Gao WQ. AHNAK2 is a Novel Prognostic Marker and Oncogenic Protein for Clear Cell Renal Cell Carcinoma. Theranostics. 2017; 7:1100–13.
https://doi.org/10.7150/thno.18198
PMID:28435451

26. Kim SH, Park WS, Park EY, Park B, Joo J, Joung JY, Seo HK, Lee KH, Chung J. The prognostic value of BAP1, PBRM1, pS6, PTEN, TGase2, PD-L1, CA9, PSMA, and Ki-67 tissue markers in localized renal cell carcinoma: A retrospective study of tissue microarrays using immunohistochemistry. PLoS One. 2017; 12:e0179610.
https://doi.org/10.1371/journal.pone.0179610
PMID:28654655

27. Braun DA, Hou Y, Bakouny Z, Ficial M, Sant' Angelo M, Forman J, Ross-Macdonald P, Berger AC, Jegede OA, Elagina L, Steinharter J, Sun M, Wind-Rotolo M, et al. Interplay of somatic alterations and immune infiltration modulates response to PD-1 blockade in advanced clear cell renal cell carcinoma. Nat Med. 2020; 26:909–18.
https://doi.org/10.1038/s41591-020-0839-y
PMID:32472114

28. Bakhoum SF, Ngo B, Laughney AM, Cavallo JA, Murphy CJ, Ly P, Shah P, Sriram RK, Watkins TBK, Taunk NK, Duran M, Pauli C, Shaw C, et al. Chromosomal instability drives metastasis through a cytosolic DNA response. Nature. 2018; 553:467–72.
https://doi.org/10.1038/nature25432
PMID:29342134

29. Frey MK, Pothuri B. Homologous recombination deficiency (HRD) testing in ovarian cancer clinical practice: a review of the literature. Gynecol Oncol Res Pract. 2017; 4:4.
https://doi.org/10.1186/s40661-017-0039-8
PMID:28250960

30. Zhang C, Chen T, Li Z, Liu A, Xu Y, Gao Y, Xu D. Depiction of tumor stemlike features and underlying relationships with hazard immune infiltrations based on large prostate cancer cohorts. Brief Bioinform. 2021; 22:bbaa211.
https://doi.org/10.1093/bib/bbaa211
PMID:32856039

31. Yabas M, Elliott H, Hoyne GF. The Role of Alternative Splicing in the Control of Immune Homeostasis and Cellular Differentiation. Int J Mol Sci. 2015; 17:3.
https://doi.org/10.3390/ijms17010003
PMID:26703587

32. Xu L, Deng C, Pang B, Zhang X, Liu W, Liao G, Yuan H, Cheng P, Li F, Long Z, Yan M, Zhao T, Xiao Y, Li X. TIP: A Web Server for Resolving Tumor Immunophenotype Profiling. Cancer Res. 2018; 78:6575–80.

33. Yuan W, Cai W, Huang X, Peng S. Prognostic value of immune scores in the microenvironment of colorectal cancer. Oncol Lett. 2020; 20:256.
https://doi.org/10.3892/ol.2020.12119
PMID:32994819

34. Ge PL, Li SF, Wang WW, Li CB, Fu YB, Feng ZK, Li L, Zhang G, Gao ZQ, Dang XW, Wu Y. Prognostic values of immune scores and immune microenvironment-related genes for hepatocellular carcinoma. Aging (Albany NY). 2020; 12:5479–99.
https://doi.org/10.18632/aging.102971
PMID:32213661

35. Oshi M, Kawaguchi T, Yan L, Peng X, Qi Q, Tian W, Schulze A, McDonald KA, Narayanan S, Young J, Liu S, Morris LG, Chan TA, et al. Immune cytolytic activity is associated with reduced intra-tumoral genetic heterogeneity and with better clinical outcomes in triple negative breast cancer. Am J Cancer Res. 2021; 11:3628–44.
PMID:34354864

36. Gong J, Tang W, Lv B, Zhang S, Fan T, Gao G, Chen D, Liu Y. ARHGEF3 Associated with Invasion, Metastasis, and Proliferation in Human Osteosarcoma. Biomed Res Int. 2021; 2021:3381957.
https://doi.org/10.1155/2021/3381957
PMID:34350290

37. Mullin BH, Prince RL, Dick IM, Hart DJ, Spector TD, Dudbridge F, Wilson SG. Identification of a role for the ARHGEF3 gene in postmenopausal osteoporosis. Am J Hum Genet. 2008; 82:1262–9.
https://doi.org/10.1016/j.ajhg.2008.04.016
PMID:18499081

38. Liu TH, Zheng F, Cai MY, Guo L, Lin HX, Chen JW, Liao YJ, Kung HF, Zeng YX, Xie D. The putative tumor activator ARHGEF3 promotes nasopharyngeal carcinoma cell pathogenesis by inhibiting cellular apoptosis. Oncotarget. 2016; 7:25836–48.
https://doi.org/10.18632/oncotarget.8283
PMID:27028992

39. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, Kiezun A, Hammerman PS, McKenna A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013; 499:214–8.
https://doi.org/10.1038/nature12213
PMID:23770567

40. Fridman WH, Zitvogel L, Sautès-Fridman C, Kroemer G. The immune contexture in cancer prognosis and treatment. Nat Rev Clin Oncol. 2017; 14:717–34.

https://doi.org/10.1038/nrclinonc.2017.101
PMID:28741618

41. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012; 16:284–7.
https://doi.org/10.1089/omi.2011.0118
PMID:22455463

42. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015; 43:e47.
https://doi.org/10.1093/nar/gkv007
PMID:25605792

43. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. Statistics for Biology and Health. Springer: New York, NY. 2000; XIV, 350.
https://doi.org/10.1007/978-1-4757-3294-8

44. Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. Nat Genet. 2006; 38:1043–8.
https://doi.org/10.1038/ng1861
PMID:16921376

45. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep. 2017; 18:248–62.
https://doi.org/10.1016/j.celrep.2016.12.019
PMID:28052254
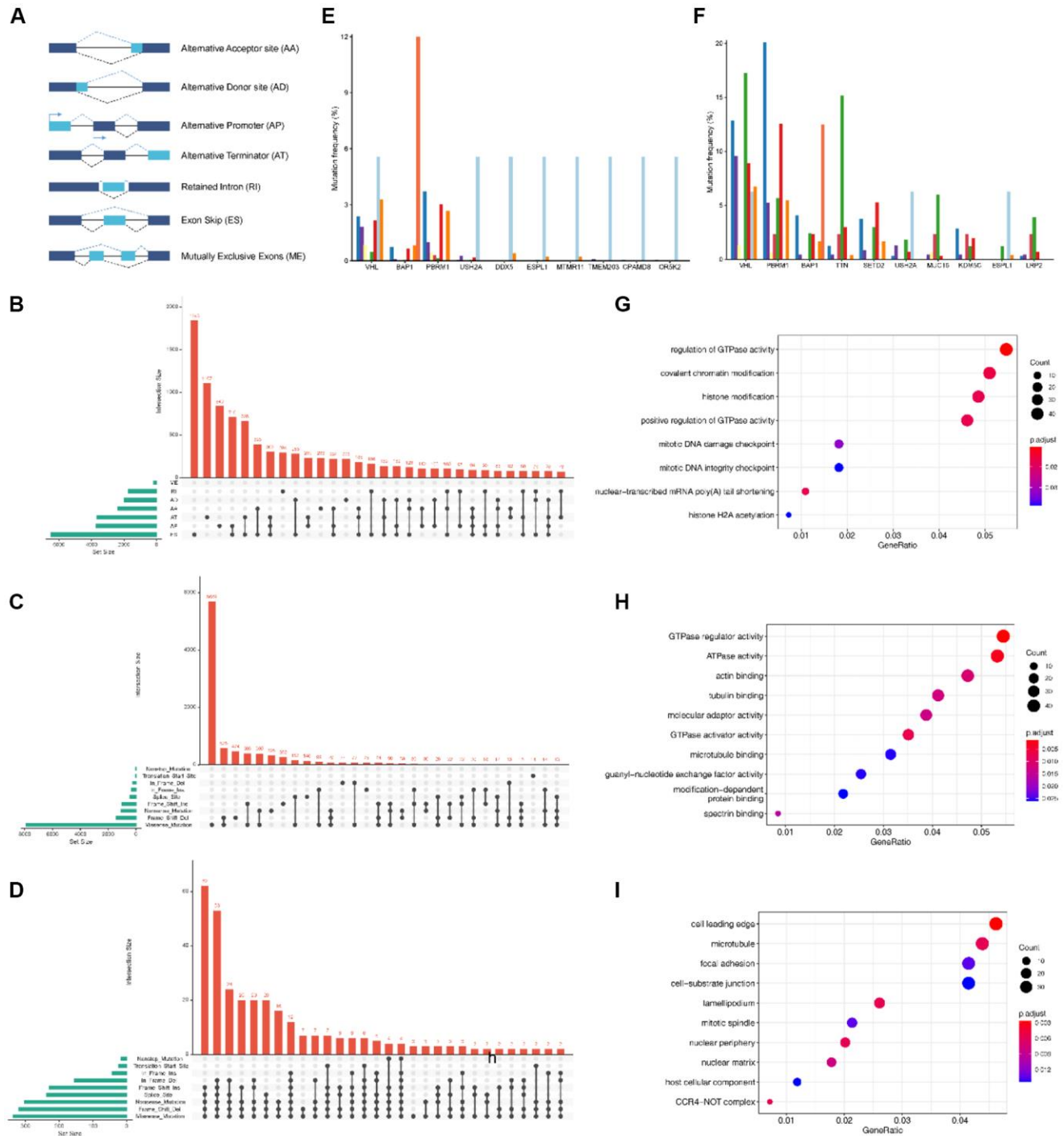
Supplementary Figure 1. Potential antigen identification and enrichment of gene ontology (GO). (**A**) Seven AS events in ccRCC patients. (**B**) Number of genes involved in AS events. (**C**, **D**) Frequency of somatic mutations and genes/patients involved. (**E**, **F**) The mutation frequency of top 10 genes in all samples/ccRCC patients. (**G**–**I**) biological process (BP), molecular function (MF), and cellular component (CC) enrichment of 901 potential antigens.
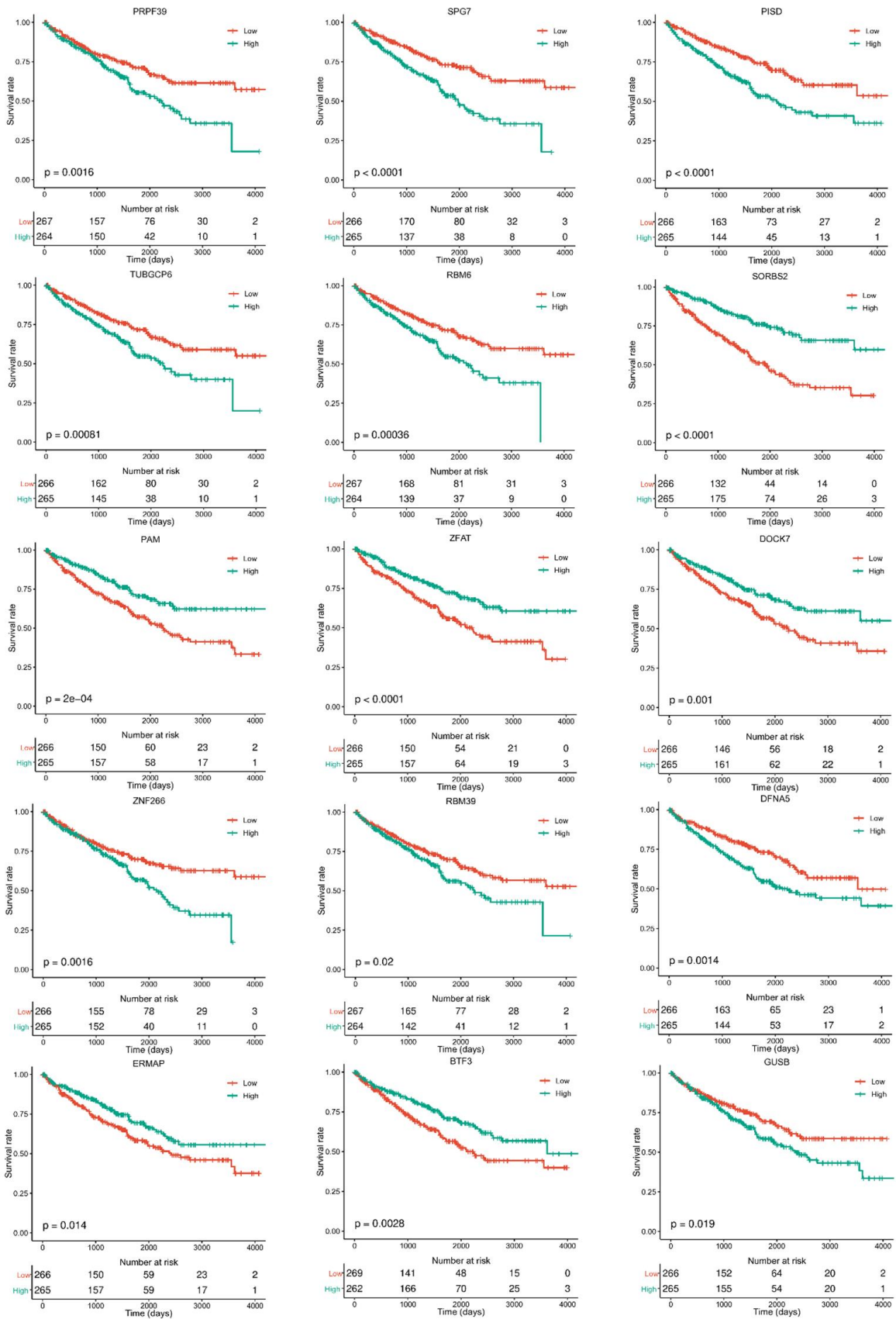
Supplementary Figure 2. Fifteen candidates significantly associated with the overall survival rate. Effect of PRPF39, SPG7, PISD, TUBGCP6, RBM6, SORBS2, PAM, ZFAT, DOCK7, ZNF266, RBM39, DFNA5, ERMAP, BTF3, and GUSB on overall survival rate.
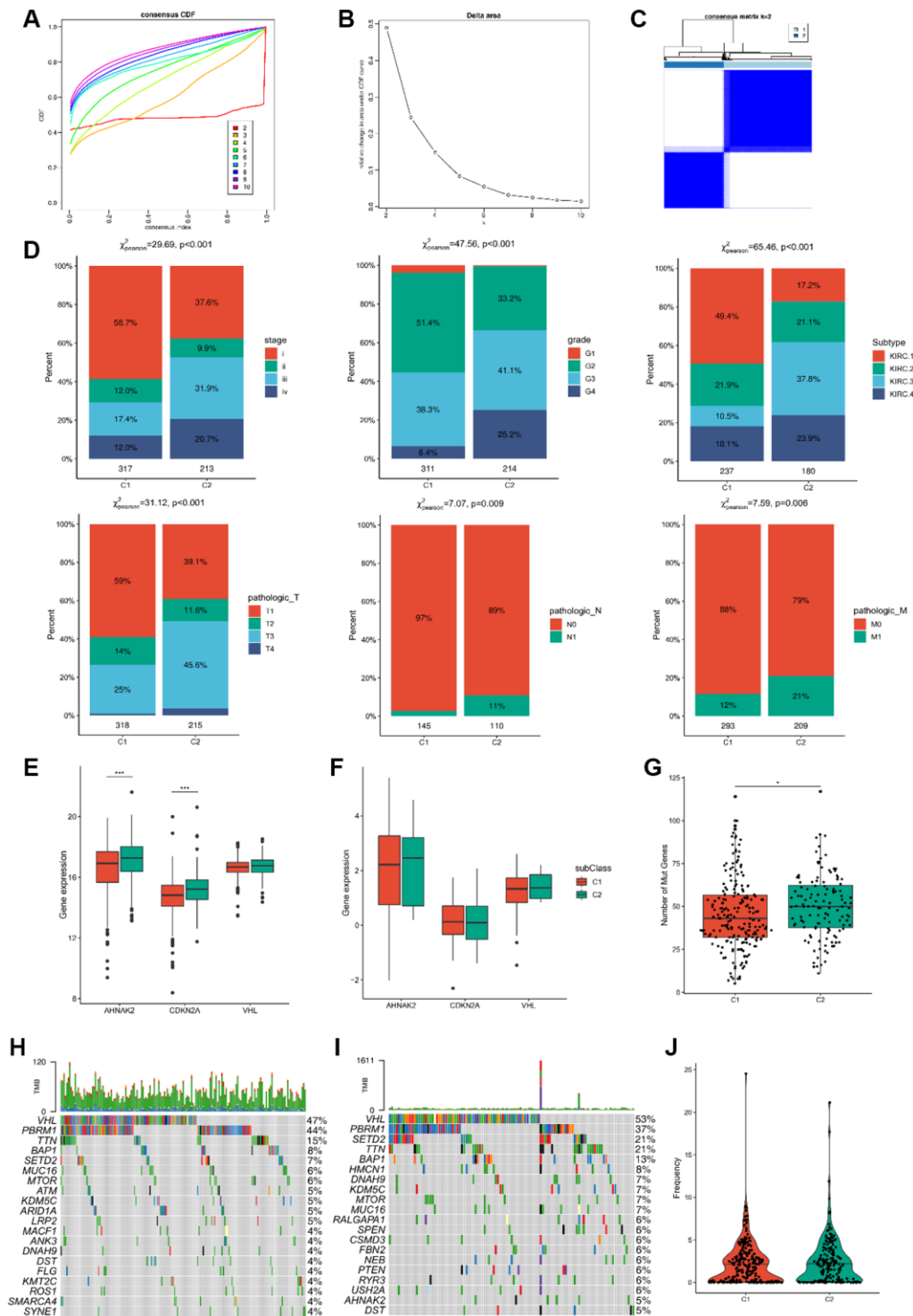
Supplementary Figure 3. Identification of potential immune subtypes of ccRCC patients. (**A**) Cumulative distribution function curve and (**B**) delta area of immune-related genes in TCGA cohort. (**C**) Sample clustering heat map. (**D**) Distribution of C1/C2 across ccRCC stages, grades, molecular subtypes, T, N, M staging of TCGA cohort. (**E**, **F**) The expression of ccRCC patient biomarkers in C1/C2 subtypes. (**G**) Number of mutated genes in C1/C2 ccRCC patients. (**H**, **I**). Twenty highly mutated genes in C1/C2 ccRCC patients. (**J**) CNV frequency of C1/C2 ccRCC patients.
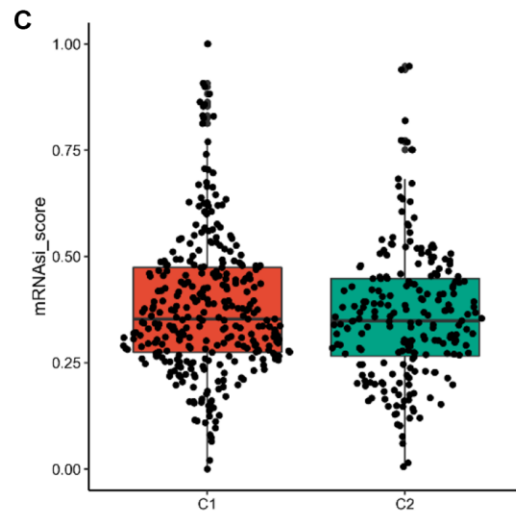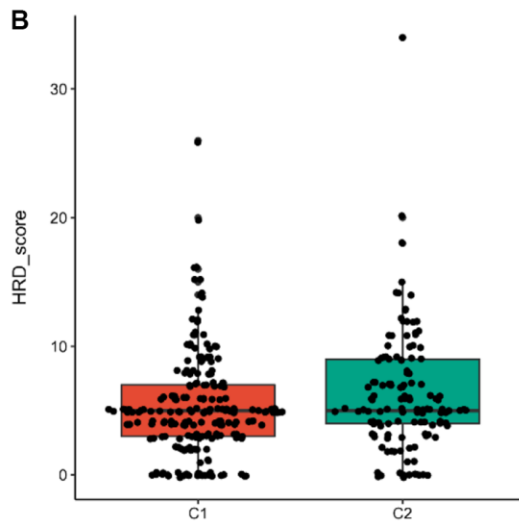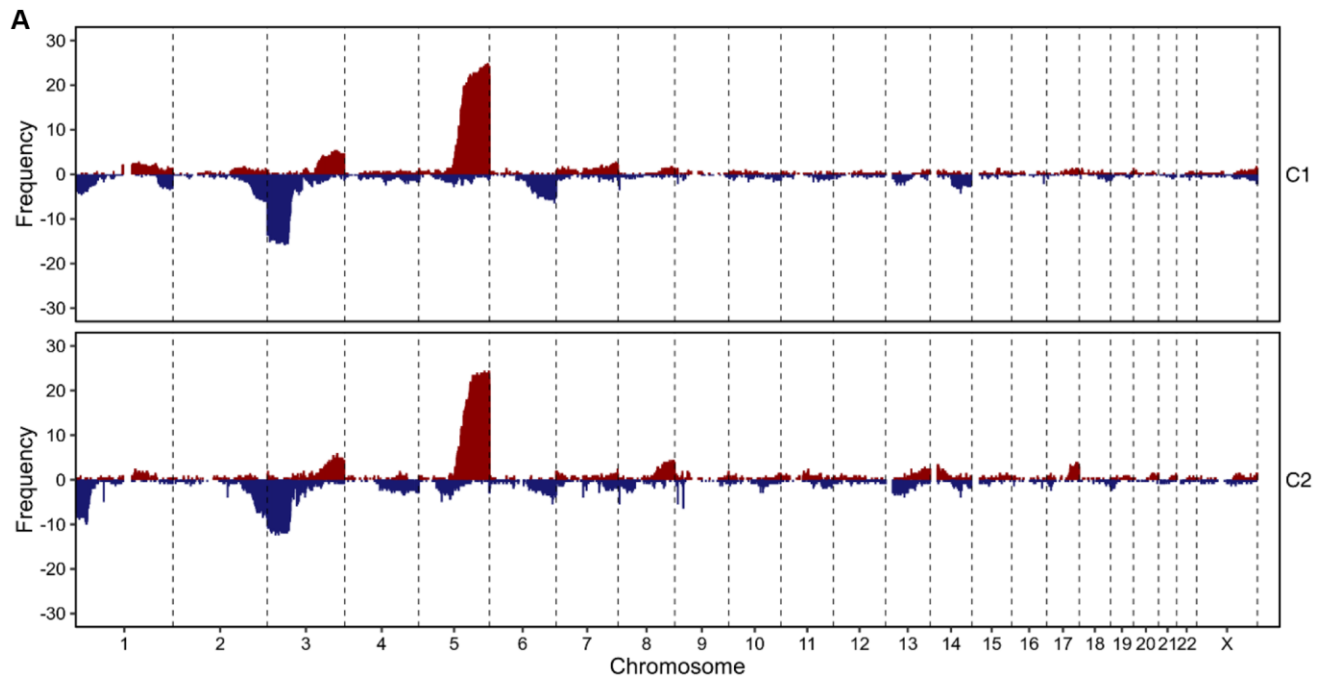
Supplementary Figure 4. Molecular characteristics of immune subtypes. (**A**) Overview of CNV in C1/C2 ccRCC patients. (**B**, **C**) HRD score and mRNAsi score of C1/C2 ccRCC patients. $^*p$adj < 0.05; $^{**}p$adj < 0.01; $^{***}p$adj < 0.001.
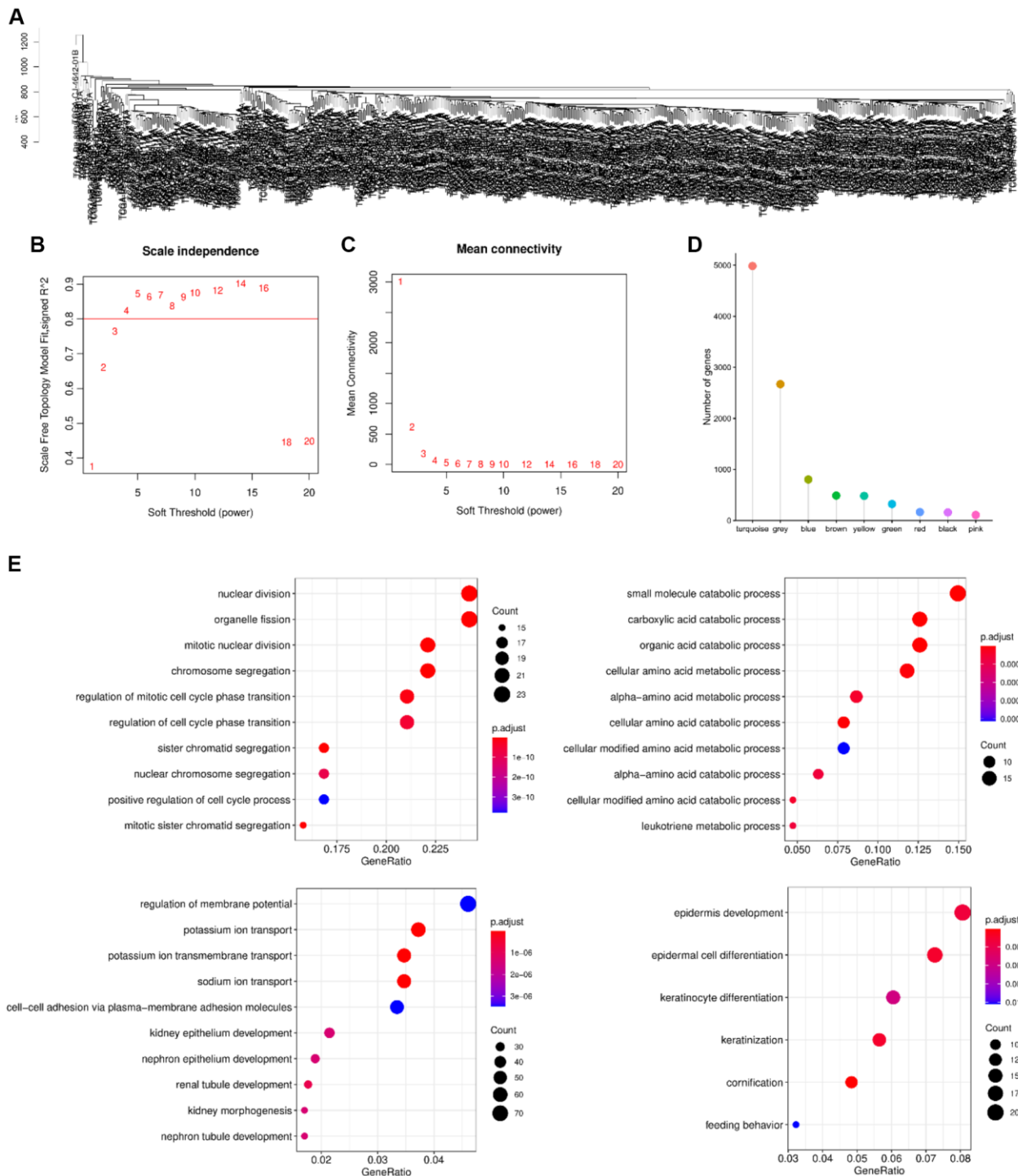
Supplementary Figure 5. Immune gene co-expression modules and immune hub genes for ccRCC identification. (**A**) Clustering of samples. (**B**) Scale-free fit index for various soft-thresholding powers (β). (**C**) Mean connectivity for various soft-thresholding powers. (**D**) Gene numbers in each module. (**E**) BP, MF, CC, and KEGG enrichment of green, black, gray, and turquoise modules. The dot size and color intensity represent the gene count and enrichment level, respectively.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Table 1.

Supplementary Table 1. The list of 157 differentially expressed PSI genes.

Supplementary Table 2. The clinical characteristics of samples from TCGA.

Characteristic	Tumor	Normal
Sample	531	71
Stage I	265	−
II	58	−
III	123	−
IV	82	−
Grade I	13	−
II	230	−
III	206	−
IV	74	−
KIRC subtype 1	147	−
2	90	−
3	93	−
4	86	−
Pathologic T1	271	−
T2	70	−
T3	179	−
T4	11	−
Pathologic N0	239	−
N1	16	−
Pathologic M0	422	−
M1	78	−