

Original Article

MUC20 as a novel prognostic biomarker in ccRCC correlating with tumor immune microenvironment modulation

Bo Xue¹, Wen-Min Guo², Jie-Dong Jia², Gaoher Kadeerhan², Hua-Ping Liu², Tao Bai³, Yuan Shao¹, Dong-Wen Wang^{1,2}

¹Shanxi Medical University, Taiyuan 030001, Shanxi, China; ²Department of Urology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shenzhen 518116, Guangdong, China; ³Department of Pathology, First Hospital of Shanxi Medical University, Taiyuan 030001, Shanxi, China

Received December 27, 2021; Accepted January 24, 2022; Epub February 15, 2022; Published February 28, 2022

Abstract: Tumor microenvironment (TME) broadly participates in genesis development of clear cell renal cell carcinoma (ccRCC). To recognize the immune and stromal modulation in TME, we screened the differentially expressed TME-related genes generated by the ESTIMATE algorithm in ccRCC specimens. Following the construction of protein-protein interaction (PPI) network and univariate COX regression, mucin 20 (MUC20) was judged to be a predictive factor. Further analysis, including immunohistochemistry (IHC) showed that MUC20 was positively correlated with survival and negatively correlated with the clinicopathologic characteristics (grade, clinical and TNM stages) in ccRCC patients. Gene Set Enrichment Analysis suggested that the low-expression MUC20 group was primarily enriched in immune-related activities, inflammation and epithelial-mesenchymal transition. Based on the CIBERSORT analysis for tumor-infiltrating immune cells (TICs), MUC20 was positively correlated with CD8⁺ T cells and resting mast cells and negatively correlated with activated CD4⁺ memory T cells, Treg cells, and plasma cells, implying that MUC20 may contribute to immune component in TME. Additionally, the patients with low MUC20 expression had better response to immune checkpoint blockades (ICBs) and 17 potential anticancer drugs were screened regarding calculating IC₅₀ value. Thus, MUC20 may contain a value of prognosis assessment for ccRCC patients and indicate the immune modulation status of TME, which provided a novel insight for comprehensive immunotherapy.

Keywords: Mucin 20, ccRCC, prognostic biomarker, TME, tumor-infiltrating immune cell

Introduction

Clear cell renal cell carcinoma (ccRCC), as the main pathologic type of renal carcinoma, is featured with rapid progression after diagnosis, early metastasis with poor survival and heterogeneity in response to immunotherapy [1]. Current research on tumor genomics reveals that the complex heterogeneity of genome profiles within or between the tumor may contribute to different clinical outcomes [1-3]. Therefore, identifying disease-related genes is crucial for exploring the molecular mechanism of ccRCC, promoting early diagnosis and therapy.

Over the last few years, the tumor microenvironment (TME) has gained increased attention

due to broadly involved in promoting tumor development, such as increasing proliferation, resistance to apoptosis and immune escape. These malignant cancer phenotypes resulted from a collaborative environment consisting of tumor cells and supporting cells; the recruited immune component and the structural stromal component contributed to this complex microenvironment. Among these, the resident stromal cells were involved in extracellular matrix remodeling and tumor angiogenesis [4, 5]. Meanwhile, growing studies concerning the recruited immune cells in TME have focused on tumor biological behavior. Tumor-infiltrating immune cells (TICs) have been identified as potential indicators for targets of cancer treatment [6]. Giraldo found that those ccRCC patients with more PD1⁺, CD8⁺ T cells and Treg⁺

immune cells infiltrated had poorer prognosis, but benefited from the treatment with immune checkpoint blockades (ICBs) and TME-modulating agents [7]. Fu et al. suggested that immune evasion was accompanied by glutamine metabolism through interleukin (IL)-23 in ccRCC tumor cells and the patients with higher IL-23 expression levels had poorer outcomes [8]. The findings above shown that the adaptive immune response has profound effects on the early stage of ccRCC. Accordingly, accurate genetic analysis can identify immune-related genes and disclose their moderating mode in TME.

Bioinformatics analysis has been used to investigate genetic alterations and identify potential biomarkers in oncogenesis [9], thus exploring the functions of various components in TME can be carried out by reviewing previous data collected. In this study, we evaluated the construction of immune and stromal cells and the proportion of TICs in ccRCC patients from The Cancer Genome Atlas (TCGA) database. After further analyzing, an immune-related biomarker mucin 20 (MUC20, cell surface associated) was identified. MUC20 is a transmembrane mucin secreted by various epithelial cells to form a protecting mucous barrier, which performs a variety of roles ranging from physical protection to the regulation of immunity and cell signaling [10, 11]. Chen et al. [12] reported that MUC20 overexpression promoted ovarian cancer cells migration and invasion via activating the integrin β 1 pathway. Previous article also suggested that the mucin families were associated with tumor biology, which deserved to explore their roles of tumor markers and therapeutic targets [13]. Here, we generated TME-related differential expressed genes (DEGs) of ccRCC samples and made a series analysis to screen immune-related genes, which revealed that MUC20 is worth identifying as a marker for tumor immune microenvironment (TIME) modulation status in ccRCC.

Materials and methods

Original datum and processing

RNA-seq transcriptome and clinical data of 607 human samples (535 ccRCC; 72 paracancerous normal samples) were gathered from TCGA dataset (UCSC browser, <http://xena.ucsc.edu/>). After removing the invalid data without

gene expression data and matched clinic information, 576 samples (505 tumor and 71 matched normal samples) were conducted further analysis (Table S1).

TME components assessment for survival analysis and difference analysis

Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data (ESTIMATE) is a method for scoring tumor purity. The ImmuneScore, StromalScore and ESTIMATEScore loaded with “estimate” package [14] in R software (Version 4.0.3; <https://www.r-project.org/>) respectively represent the immune component, stromal component, and the sum of both for each sample, which means the higher the scores acquired, the more the corresponding component contributed to TME. We plotted the Kaplan-Meier (KM) curve in R language to describe the correlation between the overall survival (OS) and three scores. And the difference analysis between estimated scores and clinicopathological characteristics of ccRCC samples were performed.

Screening DEGs regarding ImmuneScore and StromalScore

We divided 505 ccRCC patients into high/low score groups by the median ImmuneScore/StromalScore. The DEGs were detected by comparing groups in package edgeR [15]. After preprocessing raw biological data of 505 samples, low-count filtering, bias removal and normalization were performed. $|\log_2FC| > 1$ and adjust $P < 0.05$ were set as the cut-off criteria.

Volcano plots and Venn plots

Volcano plots of DEGs screened by immune/stromal score were plotted by R language with package of ggplot2. Venn plots were plotted on E-Venn website (<http://www.ehbio.com/test/venn>).

Function and pathway enrichment analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis for 341 DEGs were performed with the package of “clusterProfiler” [16] in R language. Only terms with both P -value and q -value less than 0.05 were set as significance.

Construction of protein-protein interaction (PPI) network and COX regression analysis

We constructed PPI network of 341 DEGs on STRING [17] database (<http://string-db.org>) and Cytoscape [18] software (version 3.7.2; <https://cytoscape.org/>). 78 nodes with interaction score larger than 0.9 were selected. The numbers of degree representing the connectivity of each node larger than three were shown in the plot with count from small to larger. DEGs were calculated with univariate COX regression in the package of “survival”.

Immunohistochemistry (IHC) analysis

We stained the tissue sections from ccRCC and paracancerous normal kidney samples by standard IHC protocols. In brief, after overnight deparaffinization and rehydration, the paraffin sections were put into boiling antigen retrieval buffer (ARB) at 120°C for 5 min. After adding 3% hydrogen peroxide, the sections were incubated with primary antibodies against MUC20 (D221008, Sangon Biotech, China) and secondary antibodies, and the aimed color was observed with diaminobenzidine (DAB) under the microscope. The staining intensity was described as: no staining, weak staining, moderate staining and strong staining.

Gene set enrichment analysis (GSEA)

Hallmark and C7 gene sets (immunologic signatures) were set as the target sets on GSEA_4.1.0 software (<http://software.broadinstitute.org/gsea/index.jsp>). All transcriptome data of tumor samples was loaded in GSEA [19] with $P < 0.05$ and $FDR < 0.25$ set as significance.

TICs profile estimation

The abundance profile of 22 types of TICs in 505 tumor samples was estimated with CIBERSORT [20] algorithm, which were plotted in R language with “barplot” and “pheatmap” packages. Only 399 samples with $P < 0.05$ were selected for analysis.

Potential response to ICBs and estimated IC_{50} of anticancer drugs

We analyzed the correlation between MUC20 and immune checkpoints in ccRCC samples based on the TISIDB [21] database (<http://cis.hku.hk/TISIDB>). The result of the response to

ICBs were downloaded from Immune Cell Abundant Identifier database [22] (<http://bioinfo.life.hust.edu.cn/ImmuCellAI>). To further assess the potential response to anticancer drugs for ccRCC patients with different MUC20 expression levels, we performed pRRophetic [23] package in R language to screen potential drugs included in Cancer Genome Project (CGP) by comparing estimated half maximal inhibitory concentration (IC_{50}).

Statistical analysis

KM method was used to calculate OS with the log-rank test for comparing statistically. The difference test for MUC20 with clinicopathologic characteristics, TICs and IC_{50} were compared by Kruskal-Wallis or Wilcoxon rank sum test. The correlation between target gene and TICs were assessed with Pearson coefficient test. All data analyses were performed with R software and $P < 0.05$ was considered as significance.

Results

Analysis process

Figure 1 depicts the analyzing procedure of this research. To evaluate the immune-related components in TME of ccRCC patients, we processed the RNA-seq data of 607 ccRCC samples from TCGA dataset and performed expression data with ESTIMATE algorithm. DEGs screened by ImmuneScore and StromalScore were analyzed using the PPI network and univariate COX regression. After filtering five genes in the intersection, we focused on MUC20 for further analysis, including difference test with IHC validation, correlation analysis for survival, clinical characteristics and GSEA. For further exploring the role of MUC20 in immune activities, we conducted on correlation analysis for 22 types of TICs evaluated with CIBERSORT algorithm, estimated response to ICBs and potential anticancer drugs from CGP database.

ImmuneScore was correlated with OS of ccRCC patients

To acquire the correlation between the immune-related components and OS of ccRCC patients, KM survival analysis was performed with estimated scores. Higher scores of ImmuneScore/StromalScore meant a greater proportion in TME and ESTIMATEScore was

TIME-related gene MUC20 predicts prognosis in ccRCC

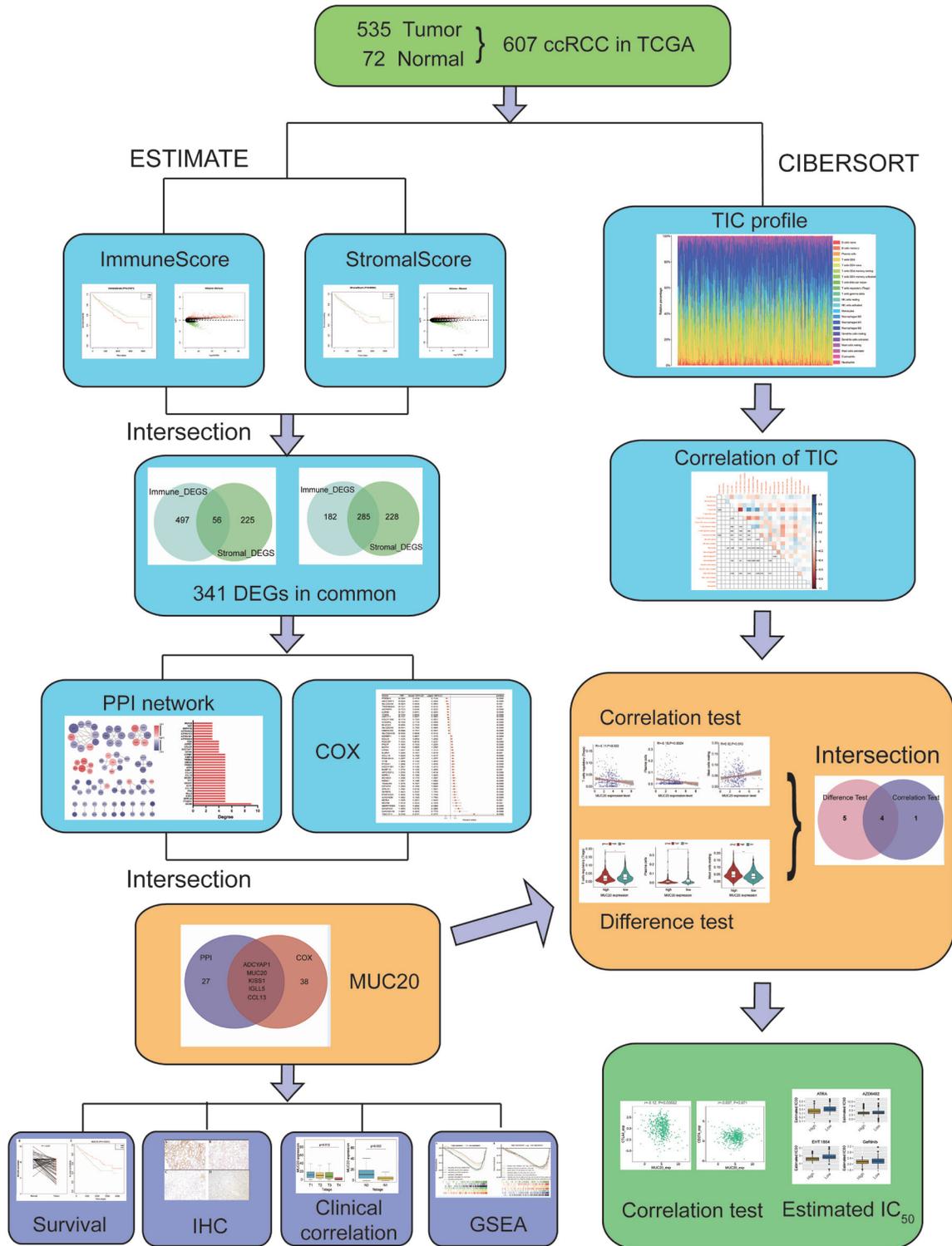


Figure 1. Analysis workflow of research.

comprehensively evaluated by these two components. The patients with lower Immune-Scores had better prognosis; however, there was no significant correlation with OS in the

ESTIMATEScore and StromalScore groups (**Figure 2**). These results supported the prognostic value of immune component in TME for ccRCC patients.

TIME-related gene MUC20 predicts prognosis in ccRCC

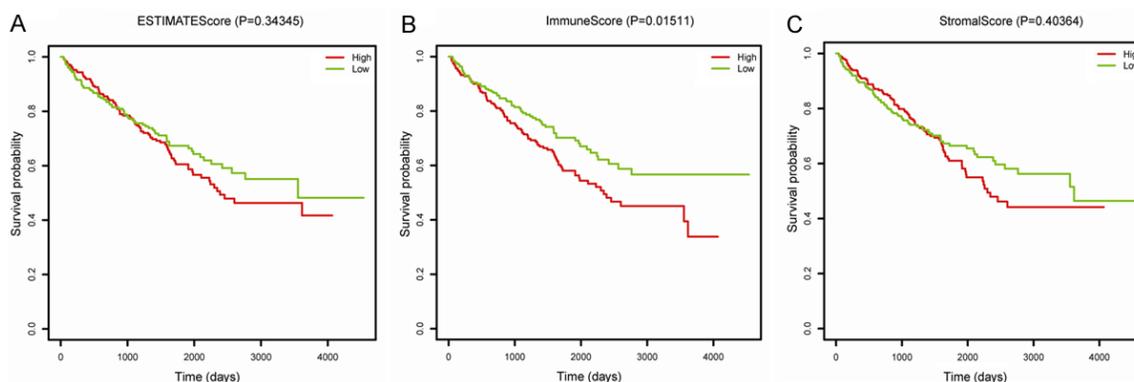


Figure 2. Survival analysis for estimated scores of ccRCC patients. A-C. KM survival analysis for ccRCC patients grouped into high or low score with ESTIMATEScore/ImmuneScore/StromalScore ($P=0.34345$, 0.01511 , 0.40364 by log-rank test, respectively).

Estimated scores were associated with the clinicopathological characteristics of ccRCC patients

To explain the role of immune-related components in ccRCC patients, the corresponding clinical information was analyzed. As presented in **Figure 3**, ESTIMATEScore and ImmuneScore significantly increased with the increase of clinic stages, pathologic Furman grade and TN stages ($P<0.05$) and the patients with higher M stages had higher ImmuneScores; however, StromalScore showed no significance in correlation with any clinicopathological characteristics. Thus, immune component in TME might promote the invasion and metastasis of ccRCC.

TME-related DEGs were mainly enriched in immune activity, metabolic process and ion transmembrane transport in TME

To evaluate the exact alterations and effects in TME, we screened TME-related DEGs for enrichment analysis. As described in **Figure 4A-D**, we obtained 1002 DEGs from the ImmuneScore group (high score vs. low score), consisting of 553 up-regulated genes and 467 down-regulated genes. Similarly, 794 DEGs were acquired from the StromalScore group. Following the intersection analysis, 56 co-upregulated genes and 285 co-downregulated genes in two groups were displayed in the Venn plot. These co-DEGs (total 341 genes) might be hub genes for TME moderation status. GO enrichment analysis suggested that these 341 DEGs were mainly mapped on the ion transmembrane transport, metabolic process and immune-related GO terms. Of these,

humoral immune response and monocyte chemotaxis were parts of immune activity (**Figure 4E**). KEGG pathway analysis also complicated the enrichment of primary immunodeficiency (**Figure 4F**). Thus, these co-DEGs seemed to be closely correlated with ion transmembrane transport, metabolic process and immune activities, which represented the significant features of TME in ccRCC.

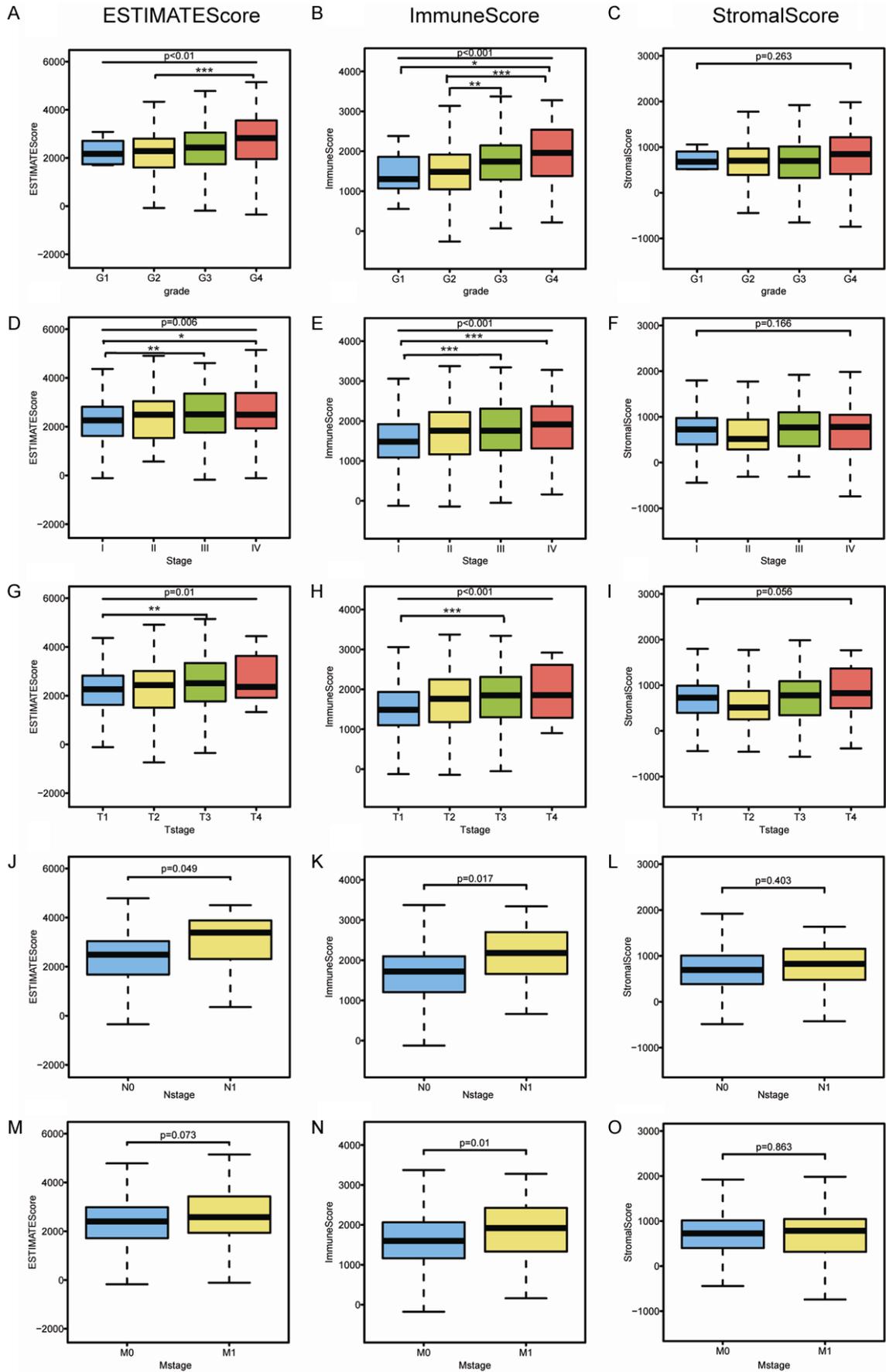
Intersection analysis of PPI network and univariate COX regression

A joint analysis of PPI network and univariate COX regression for molecular mechanism were conducted. The interactions between 341 genes were illustrated, and the top 32 genes with more nodes were displayed on the bar plot (**Figure 5A, 5B**). The top 43 factors from univariate COX regression analysis for OS were listed (**Figure 5C**). Following the intersection between 32 leading nodes and 43 factors, five hub genes (ADCYAP1, MUC20, KISS1, IGLL5 and CCL13) were identified (**Figure 5D**). Then, these hub genes were performed with survival analysis and correlation analysis between expression levels and clinical classifications; however, only MUC20 showed a statistical difference in all evaluation standards. Thus, we put MUC20 gene for further investigation.

The correlation of MUC20 expression with clinical features in ccRCC patients

MUC20 participated in many biological processes in different tissues, such as physical protection, modulation of immune and cell signaling. Recently, some mucin family members have been identified as potential biomarkers

TIME-related gene MUC20 predicts prognosis in ccRCC



TIME-related gene MUC20 predicts prognosis in ccRCC

Figure 3. Correlation of estimated scores with clinicopathological characteristics. A-C. Distribution of ESTIMATE-Score, ImmuneScore and StromalScore in pathologic Furman grade. The $P < 0.01$, $P < 0.001$, and $P = 0.263$, respectively, by Kruskal-Wallis rank test. D-F. Estimated scores in clinical stage ($P = 0.006$, $P < 0.001$, $P = 0.166$). G-O. Estimated scores in TNM classifications (by Kruskal-Wallis or Wilcoxon rank sum test). The comparison between two groups were performed with Tukey test ($*P < 0.05$, $**P < 0.01$, $***P < 0.001$).

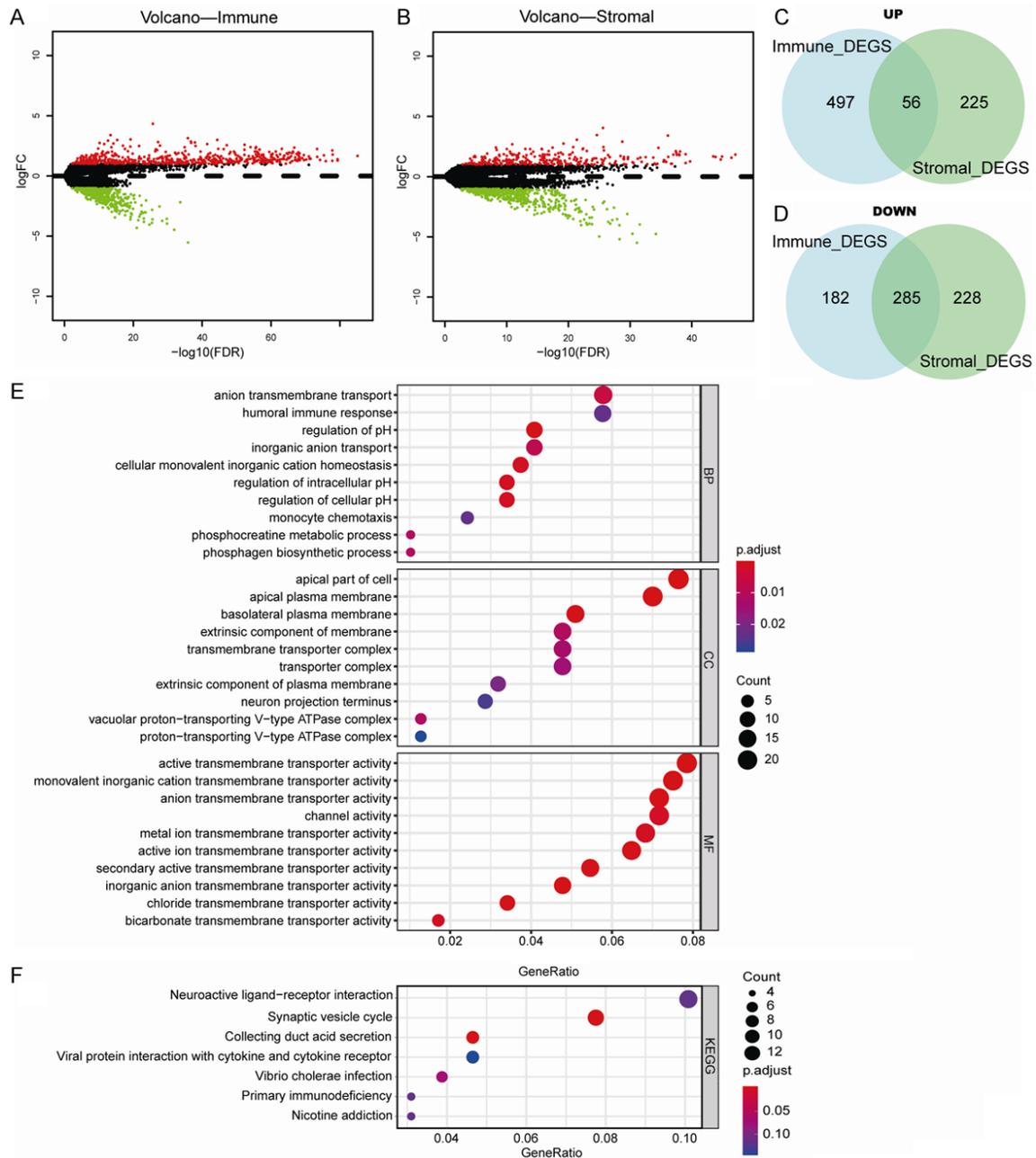


Figure 4. Volcano plots, Venn plots, and enrichment analysis of GO and KEGG for DEGs. A, B. Volcano plot for DEGs generated from ImmuneScore/StromalScore. $|\log_2FC| > 1$ and $P = 0.05$ as significance. C, D. Venn plots showed common up-regulated and down-regulated DEGs. E, F. GO and KEGG enrichment analysis for 341 DEGs, terms with $P < 0.05$ and $q < 0.05$ were set as enriched significantly.

and therapy targets in some malignant tumors. However, the function of MUC20 is unclear in ccRCC. In the presented study, the expression

of MUC20 in the tumor samples was significantly lower than that in the normal samples (Figure 6A), and the pairing analysis showed

TIME-related gene MUC20 predicts prognosis in ccRCC

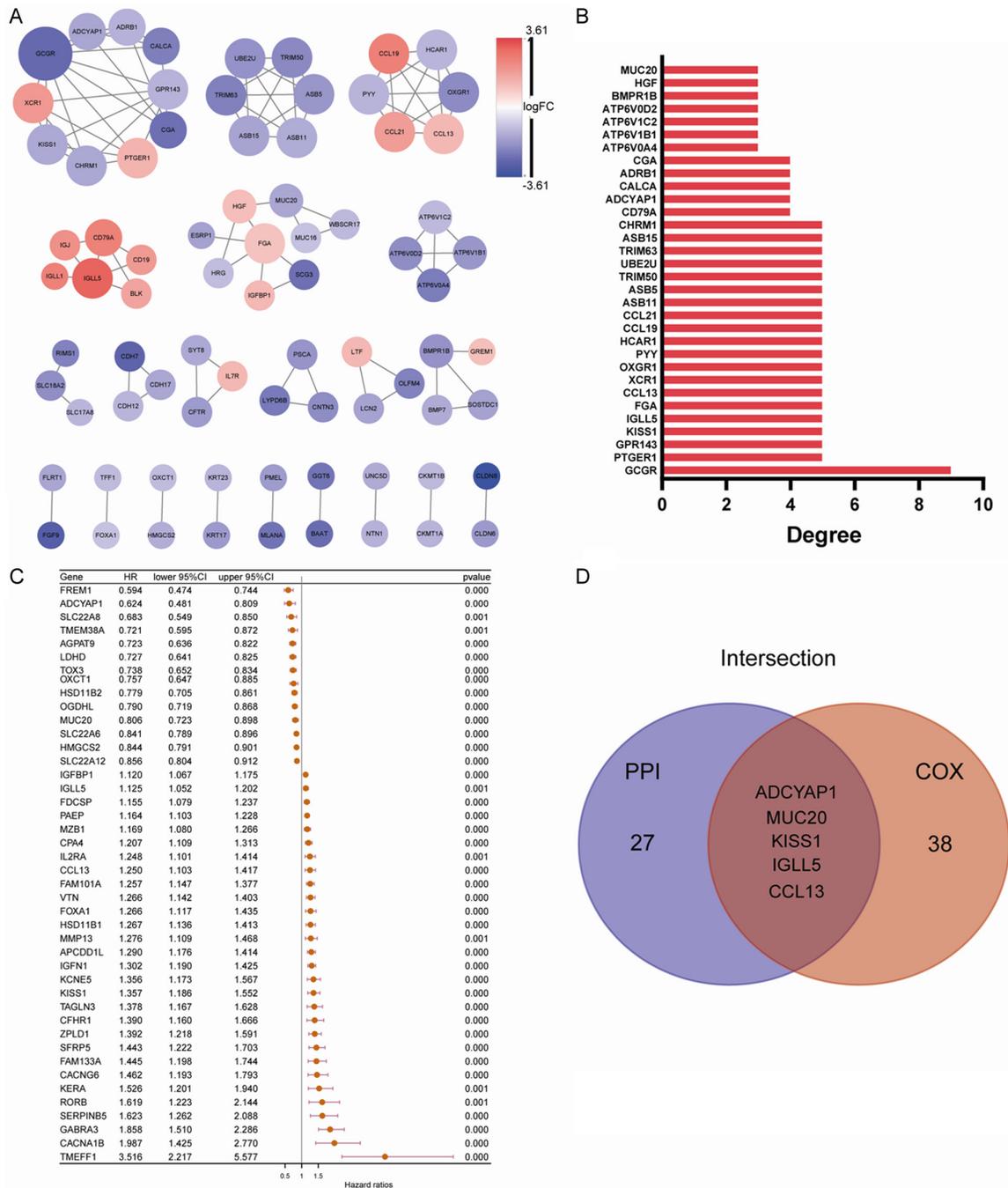
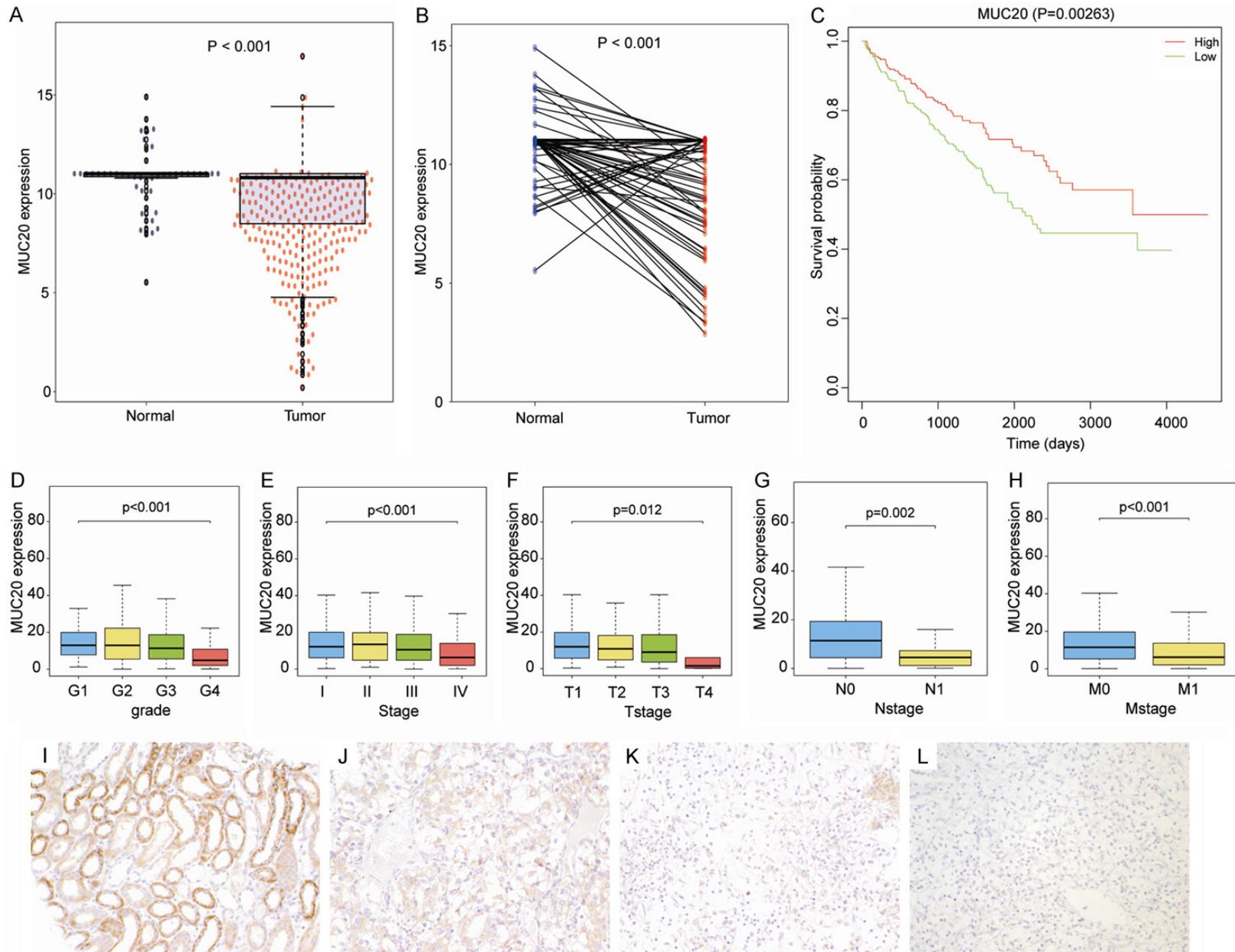


Figure 5. PPI network and univariate COX regression for DEGs. A. The interaction network constructed by the nodes with interaction confidence value >0.9. B. The 32 leading nodes ranked by degree. C. The top factors of 341 DEGs in univariate COX regression analysis ($P < 0.001$). D. The common genes intersected by 32 leading nodes and 43 top factors in Venn plot.

similar results (Figure 6B). All ccRCC samples were divided into MUC20 high-expression group and low-expression group, and the ccRCC patients with high-expression MUC20 had longer survival than low-expression group (Figure 6C). Additional IHC analysis performed on our in-house cases and verified the difference of

MUC20 expression level among normal kidney and different grade of ccRCC (Figure 6I-L). The above results indicated that down-regulated expression of MUC20 predicted poor prognosis, high pathologic grade and late stage of TNM clinical stages in ccRCC patients (Figure 6D-H).

TIME-related gene MUC20 predicts prognosis in ccRCC



TIME-related gene MUC20 predicts prognosis in ccRCC

Figure 6. The correlation of MUC20 expression with clinical features of 576 samples (505 ccRCC and 71 normal samples). A, B. Total and paired differential expression of MUC20 in tumor and normal samples. C. Survival analysis for ccRCC patients with different MUC20 expression. D-H. The correlation of MUC20 expression with clinicopathological characteristics. I-L. Representative IHC staining of MUC20 expression in ccRCC specimens at 400 magnification. I. High expression of MUC20 in normal kidney tissue as a contrast, appearing as brown yellow or brown particles. J. Moderate expression of MUC20 with Fuhrman 1-2. K, L. Low expression of MUC20 with Fuhrman 3-4.

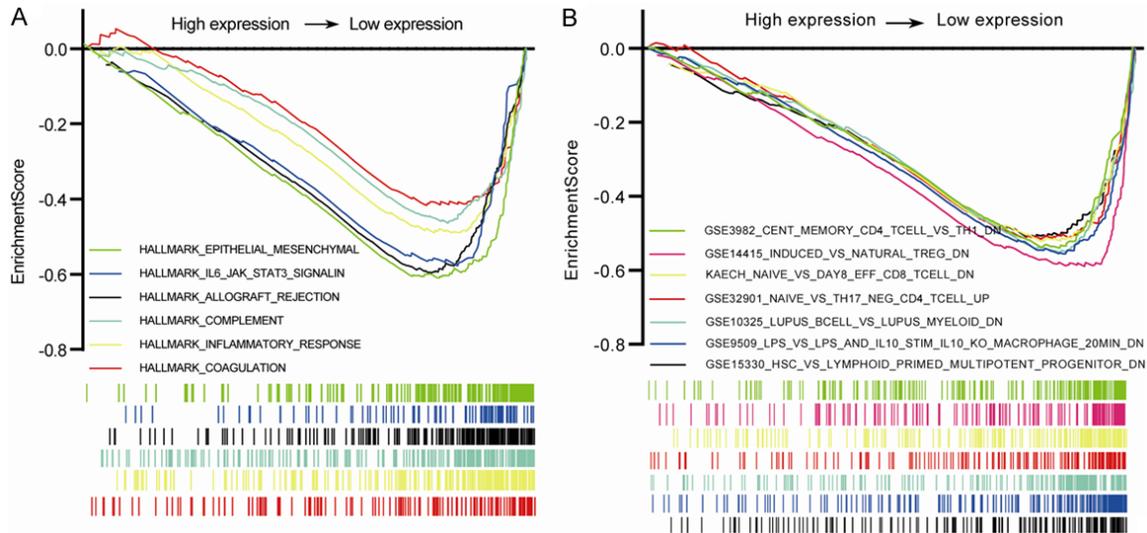


Figure 7. GSEA for ccRCC samples with low expression of MUC20. A. The enriched gene sets in HALLMARK. B. Several leading enriched signatures in C7 gene set. Only the gene sets with $P < 0.05$ and $FDR < 0.25$ were shown in the plot.

MUC20 might be a potential indicator of TIME modulation in ccRCC

Given that the expression of MUC20 was negatively correlated with clinicopathological parameters in ccRCC, GSEA was applied to investigate bio-behavior involved. As shown in **Figure 7A** and **Table S2**, the related-genes in the MUC20 low-expression group were mainly enriched in immune-related activities (allograft rejection, complement), inflammation (IL-6-JAK-STAT3 signaling, inflammatory response) and epithelial-mesenchymal transition (EMT). These immune-related gene sets were shown in **Figure 7B** and **Table S2**, such as memory CD4⁺ T cell, macrophage, natural Treg, B cell and lymphoid primed progenitor, etc. However, there was no obvious gene set enriched in the MUC20 high-expression group. The results described above implied that MUC20 might contribute to immune status modulation in ccRCC.

Correlation of MUC20 expression with TICs in ccRCC

To mining the precision MUC20-regulated immune component, we assessed 22 types of

TICs in ccRCC samples (**Figure 8**). The correlated TICs demonstrated a precise immunophenotypic profile in ccRCC. The difference analysis and correlation analysis for MUC20 and TICs showed that resting mast cells and CD8⁺ T cells were positively correlated with MUC20 expression; activated CD4⁺ memory T cells, T cells regulatory (Treg), and plasma cells were negatively correlated with MUC20 expression (**Figure 9**; **Table S3**). The result of intersection between difference test and correlation test provided four types of TICs deserved more attention. The above results confirmed that MUC20 affected the immune activity of TME in ccRCC and several TICs were mentioned particularly.

Correlation of MUC20 expression with immune checkpoints and estimated response to anti-cancer drugs

ccRCC patients with positive immune checkpoints usually have better prognosis. Thus, we considered whether the expression of MUC20 can be used as a guide for immunotherapy strategies for ccRCC patients. As shown in **Figure 10A**, there were several immune-check-

TIME-related gene MUC20 predicts prognosis in ccRCC

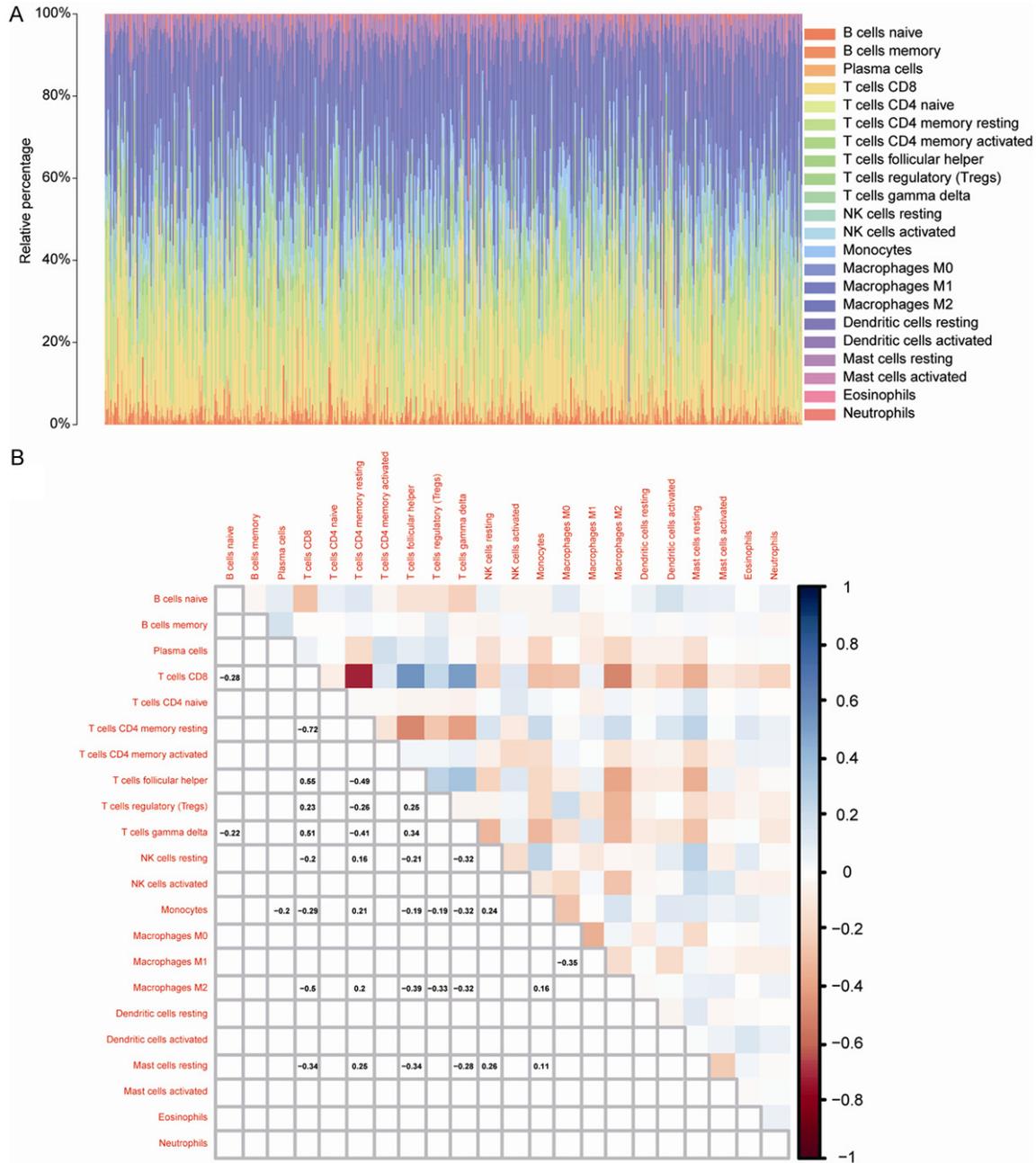


Figure 8. TICs profile in ccRCC samples and correlation analysis. A. Barplot showed the proportion of 22 types of TICs in ccRCC. B. The correlation between each TIC was shown in heatmap. The numerical value in box indicated the r value of correlation with Pearson coefficient test.

point genes (TGFB1, IL10/IL-10RB, CD96, BTLA) negatively correlated with the expression of MUC20 ($P < 0.05$). Additionally, the low MUC20 expression group (15.6%) had more patients responding to ICBs than high-expression group (9%) ($P = 0.0453$) (Table S4). The estimated IC_{50} for 138 kinds of anticancer drugs in CGP were measured, and we screened 17 potential drugs with significance from the database ($P < 0.05$) (Figure 10B). The esti-

mated response to different drugs for ccRCC patients with different MUC20 expression provided us with more opportunity to explore the potential direction of therapy.

Discussion

The remodeling of TME from tumor-suppressive to tumor-activated status draws attention on the potential therapeutic targets in recent

TIME-related gene MUC20 predicts prognosis in ccRCC

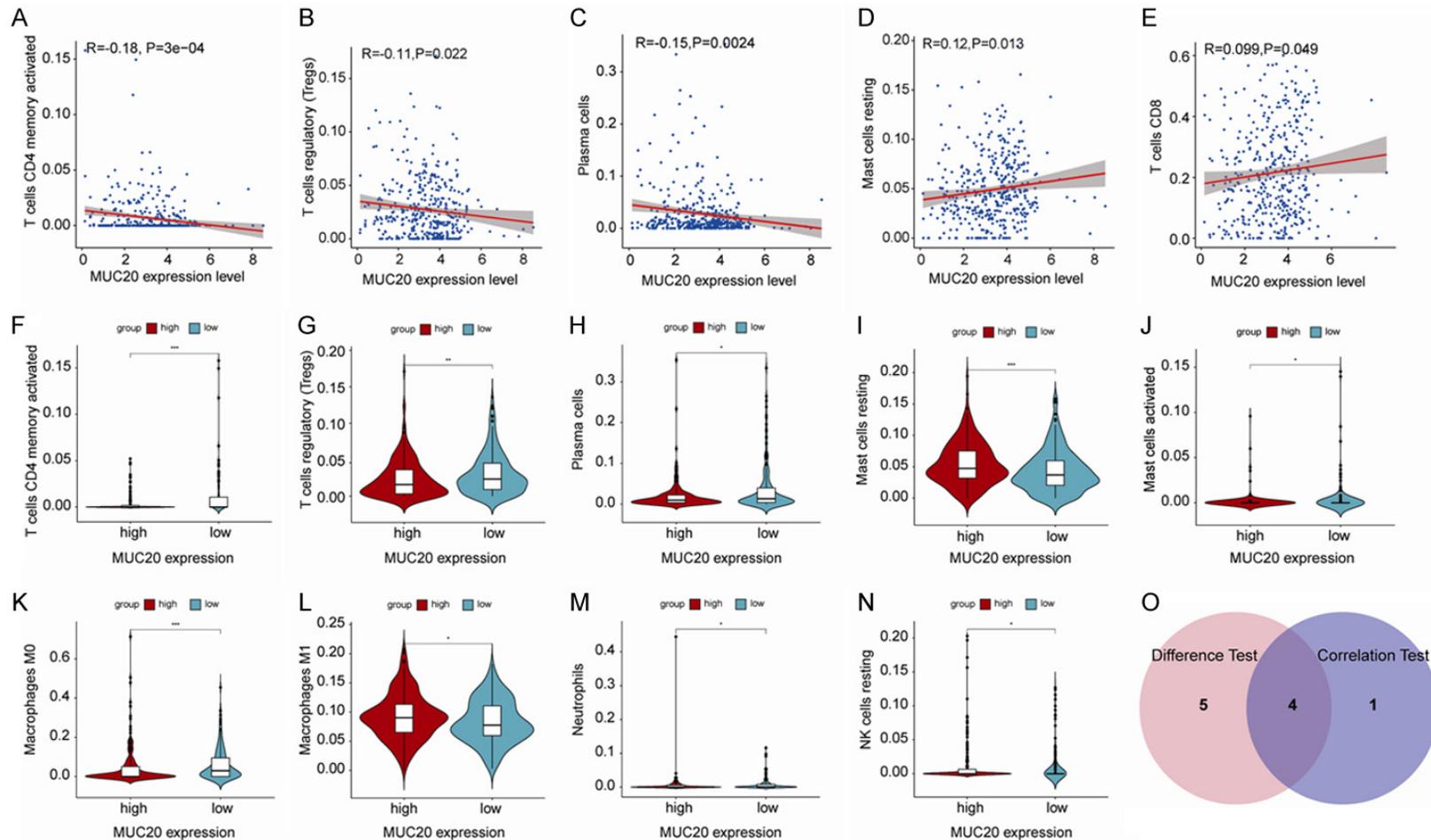
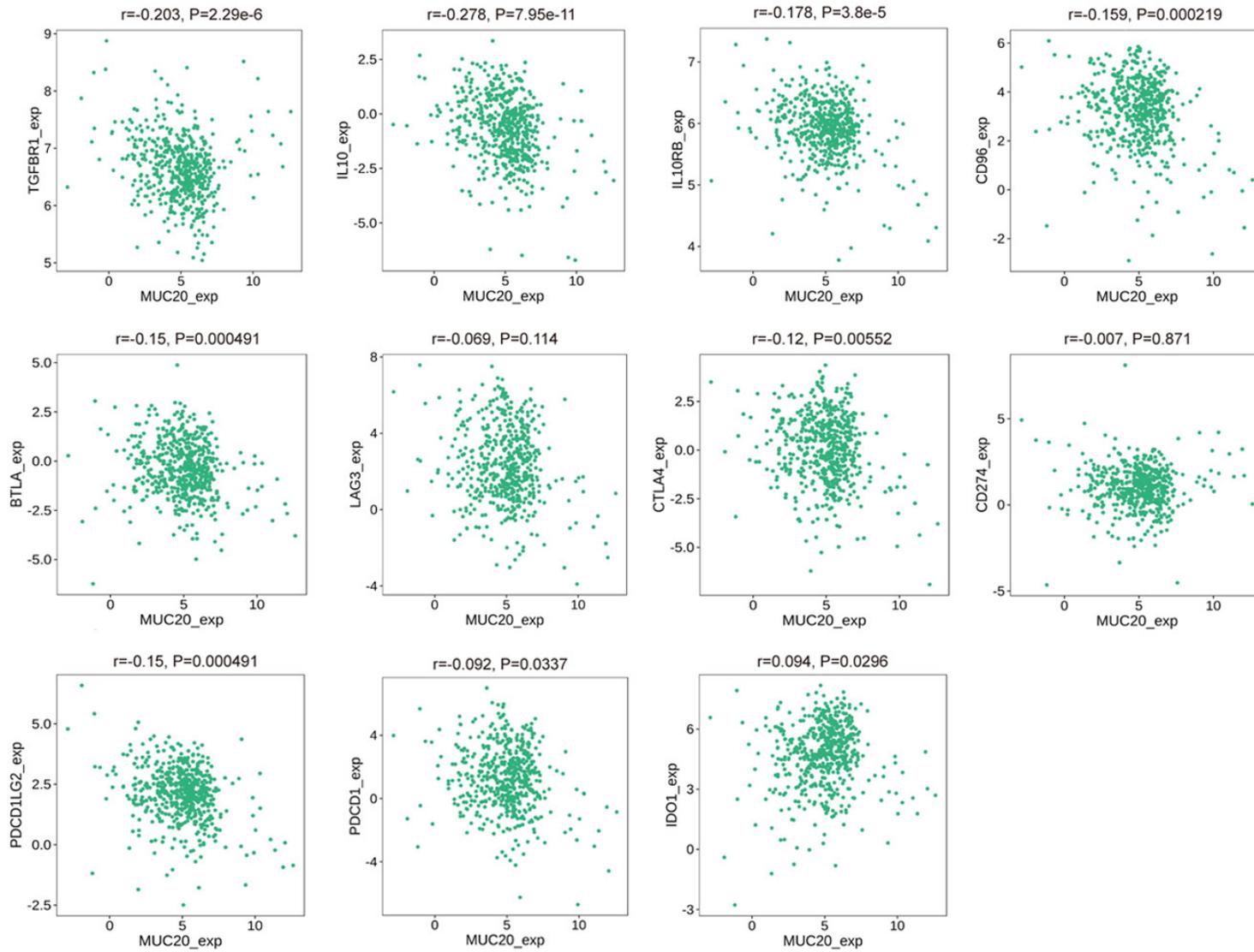


Figure 9. Correlation of TICs proportion with MUC20 expression. A-E. The correlation of five types of TICs with the MUC20 expression by Pearson coefficient test ($P<0.05$). The red shaded line represented the trend of TIC along with MUC20 expression. F-N. Violin plot displayed the difference of TICs between ccRCC samples and MUC20 expression. O. Venn plot presented common TICs from intersection of difference and correlation tests.

TIME-related gene MUC20 predicts prognosis in ccRCC

A



TIME-related gene MUC20 predicts prognosis in ccRCC

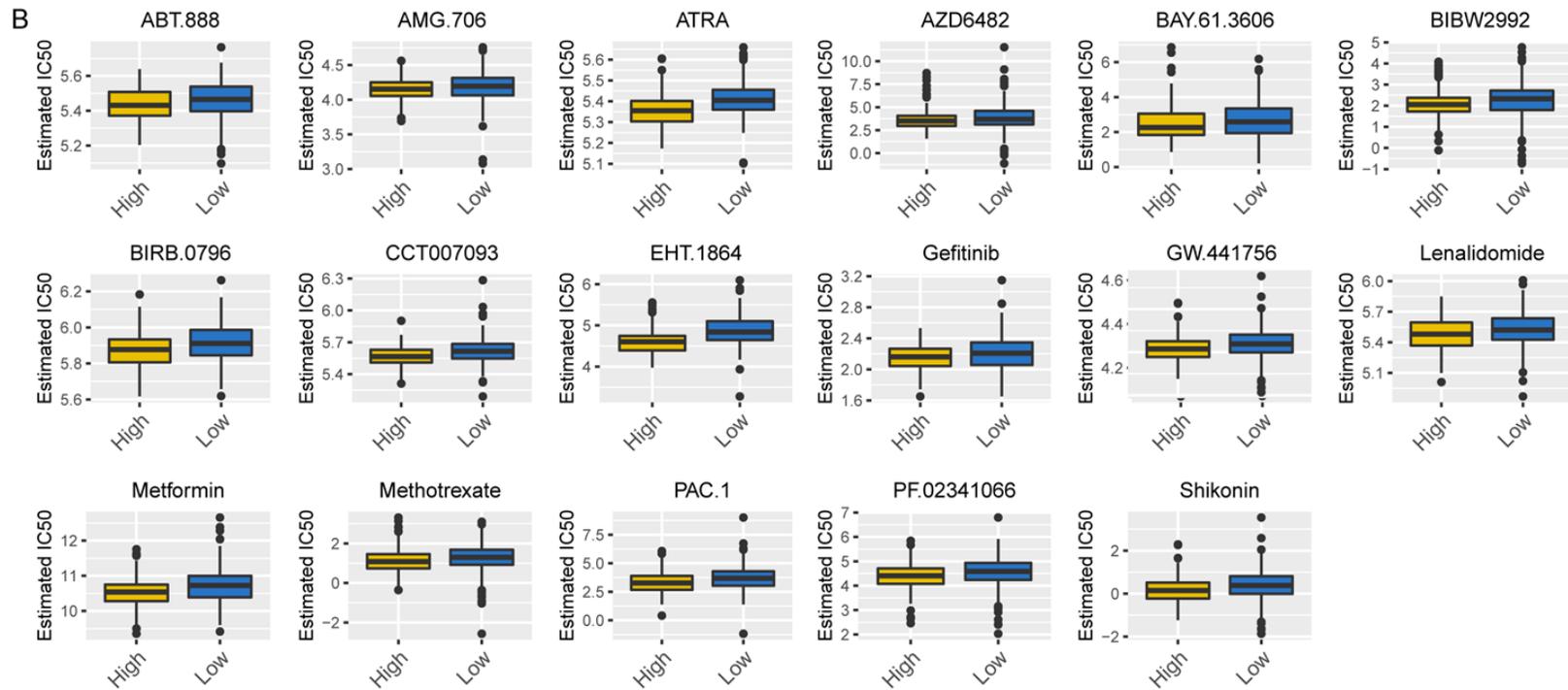


Figure 10. Correlation of immune checkpoints with MUC20 expression and putative differential response to anticancer drugs. A. Scatter plot showed the correlation of five kinds of ICBs with the MUC20 expression with Spearman coefficient test. B. The estimated IC₅₀ for anticancer drugs in MUC20 high-expression vs. low-expression group with significance ($P < 0.05$).

years. In this study, we aimed to screen TME-related DEGs correlated with clinic characteristics and patient survival in ccRCC. According to promoting tumor proliferation, invasion and metastasis, the immune component in TME influenced the prognosis of ccRCC patients. A series of analyses inspired us that MUC20 is an indicator for TIME modulation status in ccRCC.

Currently, the first-line recommendation for advanced ccRCC including immune combined anti-vascular targeted therapy has been published in the latest NCCN Kidney Cancer Guidelines (Version 1, 2021) [24]. Although advancements in immunotherapy have been made, only a small part of advanced renal cancer patients scored clinical benefit [25]. Moreover, the effect of immune checkpoint inhibitor (ICI) combined with other drugs has been better than single-agent therapy but with more immune-related adverse events [26]. Therefore, it is crucial to explore the predictive biomarkers reflecting patients who benefit from ICIs. Here we conducted enrichment analysis of immune-related DEGs screened from the transcriptome data, and the results suggested that these DEGs not only contribute to immune activity, but also participated in complex transmembrane transport, channel activity and regulation of metabolic process in ccRCC (**Figure 4**). Focused on exact mechanism, the decrease in expression of MUC20 was significantly associated with the advanced clinical characteristics (stage, grade and TNM stages) and poor prognosis in ccRCC (**Figure 6**). Additionally, those patients with low-expression MUC20 might benefit from treatment with some ICBs due to the expression of several immune checkpoints (TGFB1, IL10/IL10RB, CD96, BTLA) (**Figure 10**). Although we found that some immune checkpoints had low correlation coefficients with MUC20, such as PD-L1 ($r=0.092$), CTLA4 ($r=0.12$), the patients treated with the ICBs target for these checkpoints exhibit better clinical outcomes than no ICBs treatment in some clinical trials [27, 28]. A study of 155 renal cancer samples comparing the positive rates of PD-L1, TMB and MSI in various solid tumors found that the positive rate of PD-L1 in kidney cancer was 46/155 [29]. The low positivity rate of some immune checkpoints may therefore be responsible for these low correlation coefficients. The above

results may contribute to the reason those ccRCC patients with lower MUC20 expression had more response to ICBs ([Table S4](#)).

Mucins (MUCs) are divided into secreted and membrane-associated proteins depended on their function and structure [30]. MUCs, usually as a protective barrier, will induce morphogenetic signal transduction when epithelial cells are resisting outside injuries. Previous studies have reported that most alteration of MUCs expression are related to tumor progression and cellular properties moderation, including cellular motility, differentiation and growth [31]. MUC20, a transmembrane MUC, was overexpressed in several types of tumors and its aberrant expression was correlated with poor survival. Xiao et al. reported that the overexpression of MUC20 could enhance invasion and migration of colorectal carcinoma [32]. The overexpression of MUC20 enhanced epidermal growth factor (EGF)-induced development of cancer by activating the EGFR-STAT3 pathway in endometrial carcinoma [33]. Furthermore, increased intercellular communication and altered energy metabolism were mainly enriched bio-behavior associated with MUC20 in tumor progression [34]. Interestingly, our results revealed that MUC20 gene decreased in the advanced clinical stage of ccRCC patients, which seemed to be inconsistent with other solid tumors. Some studies suggested that the loss of certain mucins is a consequence of Wnt/ β -catenin pathway activation in cancer [35]. Moreover, the C-terminus of MUC20 protein is correlated with MET proto-oncogene, which suppresses the Grb2-Ras pathway leading to proliferation [36]. As a regulator of MET signaling cascade, MUC20 participates in the development and maintenance of kidney function and contributes to the repairment and regeneration of renal tubules [37]. The decreased expression of MUC20 due to multiple reasons may block the tissue repair and suppress the proliferation of renal cells. Therefore, despite varying expression levels in disparate cancers, the importance of MUC20 in disease progression is evident and MUC20 might play dual roles in either promoting survival or inducing poor outcome.

In addition, it has been reported that tumor-associated MUCs regulate the host-immune system via cellular crosstalk in TME [38]. We

found that several immune-related signatures, such as IL6-JAK-STAT3, allograft rejection and complement were enriched in MUC20 low-expression group. Meanwhile, CD4⁺ T cells, CD8⁺ T cells, B cells, macrophages and lymphoid primed multipotent progenitor cells, which were involved in the immune-related processes, were highly enriched (**Figure 7**). In order to further verify the direct site of action between MUC20 and immune activity, we analyzed the TICs of ccRCC patients. Tumor infiltrating lymphocytes are usually divided into two interacting groups. One group suppresses tumor growth and metastasis through specific immune responses, which are consisted of CD4⁺ Th1 cells, CD8⁺ toxic T lymphocytes (CTL), NK cells and macrophages M1. Another group mainly leads to tumor immune escape through immune resistance, such as regulatory T cells (Tregs), CD4⁺ Th2 cells, macrophages M2 and other related suppressor cells. The synergistic action among different components of immune cells contributes to the remodeling process from anti-tumor to tumor promotion. The majority of these immune cells were detected extensively infiltrating in ccRCC samples, suggesting active immune activity in TME (**Figure 8**). The TME-related TICs shown in **Figure 9** further demonstrate that MUC20 affects the immune modulation of TME in ccRCC. Though only 5 types of TICs were weakly correlated with MUC20 expression, 4 cells among these showed statistical differences between high and low expression levels of MUC20, which prompted us to further explore the role of MUC20 and these 4 TICs in TME modulation. Otherwise, ccRCC patients are usually not sensitive to common chemotherapy drugs; interestingly, the potential response to anticancer drugs indicated that ccRCC patients with high-expression MUC20 had better outcomes with some drugs, which could supply the direction of exploration in different MUC20 expression.

In conclusion, the down-regulation of MUC20 along with the advancing stage of ccRCC, the reduction of anti-tumor TICs and the increase of TICs related to immune escape in TME all supported that MUC20 might play an anti-tumor role in ccRCC. It is crucial to conduct further investigation for the combined analysis of MUC20 expression and accurate TIC components, contributing to potential MUC20-guided treatment for ccRCC patients.

Acknowledgements

This study was supported by Shanxi “1331 project” Key Innovative Research Team (No. 3c332019001) and Sanming Project of Medicine in Shenzhen (No. SZSM202111003).

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Dong-Wen Wang, Department of Urology, National Cancer Center/ National Clinical Research Center for Cancer/ Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, 113 Baohe Avenue, Longgang District, Shenzhen 518116, Guangdong, China. Tel: +86-13530336918; E-mail: wdw1urology@sxmu.edu.cn

References

- [1] Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* 2013; 499: 43-49.
- [2] Gerlinger M, Horswell S, Larkin J, Rowan AJ, Salm MP, Varela I, Fisher R, McGranahan N, Matthews N, Santos CR, Martinez P, Phillimore B, Begum S, Rabinowitz A, Spencer-Dene B, Gulati S, Bates PA, Stamp G, Pickering L, Gore M, Nicol DL, Hazell S, Futreal PA, Stewart A and Swanton C. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat Genet* 2014; 46: 225-233.
- [3] Hsieh JJ, Manley BJ, Khan N, Gao J, Carlo MI and Cheng EH. Overcome tumor heterogeneity-imposed therapeutic barriers through convergent genomic biomarker discovery: a braided cancer river model of kidney cancer. *Semin Cell Dev Biol* 2017; 64: 98-106.
- [4] Bussard KM, Mutkus L, Stumpf K, Gomez-Manzano C and Marini FC. Tumor-associated stromal cells as key contributors to the tumor microenvironment. *Breast Cancer Res* 2016; 18: 84.
- [5] Winkler J, Abisoye-Ogunniyan A, Metcalf KJ and Werb Z. Concepts of extracellular matrix remodelling in tumour progression and metastasis. *Nat Commun* 2020; 11: 5120.
- [6] Gajewski TF, Schreiber H and Fu YX. Innate and adaptive immune cells in the tumor microenvironment. *Nat Immunol* 2013; 14: 1014-1022.
- [7] Giraldo NA, Becht E, Vano Y, Petitprez F, Lacroix L, Validire P, Sanchez-Salas R, Ingels A, Oudard S, Moatti A, Buttard B, Bourass S, Ger-

- main C, Cathelineau X, Fridman WH and Sautès-Fridman C. Tumor-infiltrating and peripheral blood T-cell immunophenotypes predict early relapse in localized clear cell renal cell carcinoma. *Clin Cancer Res* 2017; 23: 4416-4428.
- [8] Fu Q, Xu L, Wang Y, Jiang Q, Liu Z, Zhang J, Zhou Q, Zeng H, Tong S, Wang T, Qi Y, Hu B, Fu H, Xie H, Zhou L, Chang Y, Zhu Y, Dai B, Zhang W and Xu J. Tumor-associated macrophage-derived interleukin-23 interlinks kidney cancer glutamine addiction with immune evasion. *Eur Urol* 2019; 75: 752-763.
- [9] Song E, Song W, Ren M, Xing L, Ni W, Li Y, Gong M, Zhao M, Ma X, Zhang X and An R. Identification of potential crucial genes associated with carcinogenesis of clear cell renal cell carcinoma. *J Cell Biochem* 2018; 119: 5163-5174.
- [10] Hattrup CL and Gendler SJ. Structure and function of the cell surface (tethered) mucins. *Annu Rev Physiol* 2008; 70: 431-457.
- [11] Corfield AP. Mucins: a biologically relevant glycan barrier in mucosal protection. *Biochim Biophys Acta* 2015; 1850: 236-252.
- [12] Chen CH, Shyu MK, Wang SW, Chou CH, Huang MJ, Lin TC, Chen ST, Lin HH and Huang MC. MUC20 promotes aggressive phenotypes of epithelial ovarian cancer cells via activation of the integrin beta1 pathway. *Gynecol Oncol* 2016; 140: 131-137.
- [13] Thomas A and Hassan R. Immunotherapies for non-small-cell lung cancer and mesothelioma. *Lancet Oncol* 2012; 13: e301-310.
- [14] Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Trevino V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, Mills GB and Verhaak RG. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 2013; 4: 2612.
- [15] Robinson MD, McCarthy DJ and Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010; 26: 139-140.
- [16] Yu G, Wang LG, Han Y and He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012; 16: 284-287.
- [17] Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ and Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 2019; 47: D607-D613.
- [18] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; 13: 2498-2504.
- [19] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES and Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005; 102: 15545-15550.
- [20] Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M and Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 2015; 12: 453-457.
- [21] Ru B, Wong CN, Tong Y, Zhong JY, Zhong SSW, Wu WC, Chu KC, Wong CY, Lau CY, Chen I, Chan NW and Zhang J. TISIDB: an integrated repository portal for tumor-immune system interactions. *Bioinformatics* 2019; 35: 4200-4202.
- [22] Miao YR, Zhang Q, Lei Q, Luo M, Xie GY, Wang H and Guo AY. ImmuCellAI: a unique method for comprehensive T-cell subsets abundance prediction and its application in cancer immunotherapy. *Adv Sci (Weinh)* 2020; 7: 1902880.
- [23] Geeleher P, Cox N and Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS One* 2014; 9: e107468.
- [24] Motzer RJ, Jonasch E, Boyle S, Carlo MI, Manley B, Agarwal N, Alva A, Beckermann K, Choueiri TK, Costello BA, Derweesh IH, Desai A, George S, Gore JL, Haas N, Hancock SL, Kyriakopoulos C, Lam ET, Lau C, Lewis B, Madoff DC, McCreery B, Michaelson MD, Mortazavi A, Nandagopal L, Pierorazio PM, Plimack ER, Ponsky L, Ramalingam S, Shuch B, Smith ZL, Somer B, Sosman J, Dwyer MA and Motter AD. NCCN guidelines insights: kidney cancer, version 1.2021. *J Natl Compr Canc Netw* 2020; 18: 1160-1170.
- [25] Flippot R, Escudier B and Albiges L. Immune checkpoint inhibitors: toward new paradigms in renal cell carcinoma. *Drugs* 2018; 78: 1443-1457.
- [26] Motzer RJ, Tannir NM, McDermott DF, Aren Frontera O, Melichar B, Choueiri TK, Plimack ER, Barthelemy P, Porta C, George S, Powles T, Donskov F, Neiman V, Kollmannsberger CK, Salman P, Gurney H, Hawkins R, Ravaud A, Grimm MO, Bracarda S, Barrios CH, Tomita Y, Castellano D, Rini BI, Chen AC, Mekan S, McHenry MB, Wind-Rotolo M, Doan J, Sharma P, Hammers HJ and Escudier B; CheckMate 214 Investigators. Nivolumab plus ipilimumab versus sunitinib in advanced renal-cell carcinoma. *N Engl J Med* 2018; 378: 1277-1290.
- [27] Barbee MS, Ogunniyi A, Horvat TZ and Dang TO. Current status and future directions of the

TIME-related gene MUC20 predicts prognosis in ccRCC

- immune checkpoint inhibitors ipilimumab, pembrolizumab, and nivolumab in oncology. *Ann Pharmacother* 2015; 49: 907-937.
- [28] Motzer RJ, Rini BI, McDermott DF, Redman BG, Kuzel TM, Harrison MR, Vaishampayan UN, Drabkin HA, George S, Logan TF, Margolin KA, Plimack ER, Lambert AM, Waxman IM and Hammers HJ. Nivolumab for metastatic renal cell carcinoma: results of a randomized phase II trial. *J Clin Oncol* 2015; 33: 1430-1437.
- [29] Vanderwalde A, Spetzler D, Xiao N, Gatalica Z and Marshall J. Microsatellite instability status determined by next-generation sequencing and compared with PD-L1 and tumor mutational burden in 11,348 patients. *Cancer Med* 2018; 7: 746-756.
- [30] Hollingsworth MA and Swanson BJ. Mucins in cancer: protection and control of the cell surface. *Nat Rev Cancer* 2004; 4: 45-60.
- [31] Bafna S, Kaur S and Batra SK. Membrane-bound mucins: the mechanistic basis for alterations in the growth and survival of cancer cells. *Oncogene* 2010; 29: 2893-2904.
- [32] Xiao X, Wang L, Wei P, Chi Y, Li D, Wang Q, Ni S, Tan C, Sheng W, Sun M, Zhou X and Du X. Role of MUC20 overexpression as a predictor of recurrence and poor outcome in colorectal cancer. *J Transl Med* 2013; 11: 151.
- [33] Chen CH, Wang SW, Chen CW, Huang MR, Hung JS, Huang HC, Lin HH, Chen RJ, Shyu MK and Huang MC. MUC20 overexpression predicts poor prognosis and enhances EGF-induced malignant phenotypes via activation of the EGFR-STAT3 pathway in endometrial cancer. *Gynecol Oncol* 2013; 128: 560-567.
- [34] Zheng F, Yu H and Lu J. High expression of MUC20 drives tumorigenesis and predicts poor survival in endometrial cancer. *J Cell Biochem* 2019; [Epub ahead of print].
- [35] Pai P, Rachagani S, Dhawan P and Batra SK. Mucins and Wnt/beta-catenin signaling in gastrointestinal cancers: an unholy nexus. *Carcinogenesis* 2016; 37: 223-232.
- [36] Higuchi T, Orita T, Katsuya K, Yamasaki Y, Akiyama K, Li H, Yamamoto T, Saito Y and Nakamura M. MUC20 suppresses the hepatocyte growth factor-induced Grb2-Ras pathway by binding to a multifunctional docking site of met. *Mol Cell Biol* 2004; 24: 7456-7468.
- [37] Higuchi T, Orita T, Nakanishi S, Katsuya K, Watanabe H, Yamasaki Y, Waga I, Nanayama T, Yamamoto Y, Munger W, Sun HW, Falk RJ, Jenette JC, Alcorta DA, Li H, Yamamoto T, Saito Y and Nakamura M. Molecular cloning, genomic structure, and expression analysis of MUC20, a novel mucin protein, up-regulated in injured kidney. *J Biol Chem* 2004; 279: 1968-1979.
- [38] Bhatia R, Gautam SK, Cannon A, Thompson C, Hall BR, Aithal A, Banerjee K, Jain M, Solheim JC, Kumar S and Batra SK. Cancer-associated mucins: role in immune modulation and metastasis. *Cancer Metastasis Rev* 2019; 38: 223-236.

TIME-related gene MUC20 predicts prognosis in ccRCC

Table S1. Clinicopathological staging characteristics statistics of ccRCC patients from TCGA

Clinical characteristics		Total (505)	%
Age at diagnosis (y)	young age (≤ 60)	258	51.1
	old age (> 60)	247	48.9
Gender	Male	332	65.7
	Female	173	34.3
Neoplasm Histologic Grade	I	12	2.4
	II	216	42.7
	III	200	39.6
	IV	72	14.3
Ajcc Pathologic Stage	I	251	49.7
	II	55	10.9
	III	117	23.2
	IV	82	16.2
T classification	T1	257	50.9
	T2	67	13.3
	T3	150	29.7
	T4	11	2.1
N classification	N0	228	45.1
	N1	15	3.0
	NX	262	51.9
M classification	M0	404	80.0
	M1	77	15.2
	MX	24	4.8

Table S2. Enriched gene sets

MSigDB collection	Gene set name	NES	NOM P-value	FDR q-value
h.all.v7.4.symbols.gmt (Hallmarks)				
MUC20 low expression	HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	-2.014	0.013	0.028
	HALLMARK_IL6_JAK_STAT3_SIGNALING	-1.947	0.002	0.041
	HALLMARK_ALLOGRAFT_REJECTION	-1.922	0.016	0.033
	HALLMARK_COMPLEMENT	-1.801	0.010	0.066
	HALLMARK_INFLAMMATORY_RESPONSE	-1.788	0.013	0.060
	HALLMARK_COAGULATION	-1.637	0.020	0.156
c7.all.v7.4.symbols.gmt (Immunologic Signatures)				
MUC20 low expression	GSE3982_CENT_MEMORY_CD4_TCELL_VS_TH1_DN	-1.982	0	0.037
	GSE9509_LPS_VS_LPS_AND_IL10_STIM_IL10_KO_MACROPHAGE_20MIN_DN	-1.981	0	0.035
	GSE14415_INDUCED_VS_NATURAL_TREG_DN	-1.976	0.010	0.035
	GSE10325_LUPUS_BCELL_VS_LUPUS_MYELOID_DN	-1.969	0.006	0.035
	KAECH_NAIVE_VS_DAY8_EFF_CD8_TCELL_DN	-1.969	0.006	0.033
	GSE32901_NAIVE_VS_TH17_NEG_CD4_TCELL_UP	-1.967	0.006	0.032
	GSE15330_HSC_VS_LYMPHOID_PRIMED_MULTIPOTENT_PROGENITOR_DN	-1.967	0.006	0.030

NES: normalized enrichment score; NOM: nominal P-value; FDR: false discovery rate. Gene sets with NOM P-value < 0.05 and FDR q-value < 0.25 were considered as statistical significance. Only several leading sets enriched in MUC20 low expression both in HALLMARK and C7 were listed here due to the large number of enriched gene sets.

TIME-related gene MUC20 predicts prognosis in ccRCC

Table S3. TICs co-determined by correlation test and difference test

TICs	Correlation test (<i>P</i> -value)	Difference test (<i>P</i> -value)
Plasma cells	-0.15 (0.002)	<0.05
Mast cells resting	-0.12 (0.013)	<0.001
T cells CD4 memory activated	-0.18 (<0.001)	<0.001
T cells regulatory (Tregs)	-0.11 (0.022)	<0.01

The Pearson coefficient was used for the correlation test and Wilcoxon rank sum was performed for the difference test. Only the *P* value <0.05 was set as significance.

Table S4. Immunophenotype of different MUC20 expression groups and response to ICBs

Classification		NO. of patients (%)	
		MUC20 High Expression (n=200)	MUC20 Low Expression (n=199)
Immunophenotype	C1	0 (0.0)	2 (1.0)
	C2	8 (4.0)	21 (10.6)
	C3	189 (94.5)	170 (85.4)
	C4	1 (0.5)	4 (2.0)
	C5	0 (0.0)	0 (0.0)
	C6	2 (1.0)	2 (1.0)
Response to ICBs	R	18 (9.0)	31 (15.6)*
	NR	182 (91.0)	168 (84.4)

The chi-square test is performed for the significance test. **P*=0.0453<0.05.