Identification and Verification of Immune-Related Genes Prognostic Signature Based on ssGSEA for Adrenocortical Carcinoma (ACC)

Kaisheng Yuan, Ruiqi Zeng, Pengteng Deng, Aiping Zhang, Huiqian Liu, Ning Wang, Yongxi Tang, Zhikang Yin, Hang Liu

Department of Urology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, 400016, People’s Republic of China

Correspondence: Hang Liu, Department of Urology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, 400016, People’s Republic of China, Tel +86-185-8030-9681, Email liuhang723@163.com

Purpose: Adrenocortical carcinoma (ACC) is an endocrine malignant tumor with poor prognosis. The study aimed to construct ACC immune-related gene prognostic signature and verify the efficacy of prognostic signature.

Methods: ACC RNA-seq data and clinical information are downloaded from TCGA databases and GEO databases. We used single sample gene set enrichment analysis (ssGSEA) to assess immune cell infiltration in ACC patients and ACC patients were divided into high- and low-immune cell infiltration clusters. The validity of ssGSEA grouping was verified using the ESTIMATE algorithm. A total of 275 differentially expressed immune-related genes (IRGs) were obtained from the intersection of IRGs and differentially expressed genes (DEGs) in high and low immune cell infiltration clusters. LASSO analysis was used to identify 13 IRGs that regulate the prognosis of ACC patients through immune infiltration. Kaplan–Meier analysis, ROC curve, univariate and multivariate Cox regression further confirmed that these 13 immune-related gene signatures were innovative and significant prognostic factors, which were independent of clinical features. Finally, ACC prognostic nomogram was constructed, ROC curve and calibration curve were drawn to evaluate the accuracy of the prognostic nomogram.

Results: LASSO regression analysis was used to screen out ACC survival-related genes. Univariate and multivariate Cox proportional risk regression models were used to analyze and construct the ACC prognosis nomogram. The AUC for predicting 1-, 3- and 5-year overall survival rate of ACC patients was 0.799, 0.966 and 0.969, suggesting good prediction accuracy. The calibration curve shows that the predicted results of the prognostic nomogram are in good agreement with the actual situation.

Conclusion: ssGSEA technique plays an important role in the construction of ACC prognostic model. Based on IRGs associated with survival independently predicted ACC prognosis, we identified thirteen immune-related genes as prognostic signature for ACC.

Keywords: immune, prognostic, adrenocortical carcinoma, ACC, single sample gene set enrichment analysis, ssGSEA, risk score

Introduction

Adrenocortical carcinoma (ACC) is an invasive endocrine malignant tumor originating in the adrenal cortex with poor prognosis. Statistically, the incidence worldwide of ACC is between 0.7 and 2 cases per million population. Clinically, patients with adrenal cortical carcinoma are more likely to be diagnosed at an advanced stage with a 5-year survival rate of less than 15%, the rate of metastasis and recurrence is very high even after radical resection. However, globally, there are relatively few cases of ACC available for study, and the potential molecular mechanism and prognostic factors of ACC are not particularly clear. Therefore, robust and trustworthy prognostic biomarkers associated with biological heterogeneity of ACC, except for the usual clinical characteristics, are necessary to improve prognosis.

Today, immunotherapy is a new treatment that has revolutionized cancer treatment. Based on observations of adrenalitis in patients receiving immune checkpoint blockade, the presence of programmed death-ligand 1 (PD-L1) expression in the tumor cell membrane and tumor-infiltrating mononuclear cells in surgically treated ACC, we think that evaluation of immunotherapy in ACC was convincing. Furthermore, Bagante et al constructed a multi-centre study and
found that upregulation of immunologic markers neutrophil–lymphocyte ratio (NLR) and platelet–lymphocyte ratio (PLR) was significantly associated with poor prognosis after resection of ACC patients. Although there is a lot of evidence that immunotherapy may be a breakthrough treatment for ACC, the related molecular mechanisms, especially the regulation of immune-related genes (IRGs), are still unclear. Therefore, systematic analysis of the relationship between immune-related genes and prognosis may provide motivation and target for further study on immunotherapy of ACC.

Therefore, in this study, ACC patients were divided into high- and low-immune cell infiltration clusters based on single sample gene enrichment analysis (ssGSEA), which was verified by ESTIMATE and CIBERSORT analysis. Gene expression and clinical data of ACC patients were downloaded from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO). The prognostic genes of ACC were identified by survival analysis. The prognostic model was constructed by using prognostic genes and the accuracy of the model would be verified. Finally, we constructed the nomogram to predict the survival rate of ACC patients, which will provide reference for the prognosis of ACC patients.

Materials and Methods

Collect and Cluster of ACC Data

ACC (n = 92) RNA sequencing data can be downloaded from TCGA (The Cancer Genome Atlas) (https://portal.gdc.cancer.gov/). Seventy-nine of the samples had relevant clinical information. Supplement Table S1A shows all the patients’ data from TCGA dataset. All IRGs can be obtained from IMMPORT (https://www.immport.org/shared/home). We performed a systematic search on the GEO database (https://www.ncbi.nlm.nih.gov/geo/) to find ACC gene expression datasets. GSE76201 dataset from the GEO database was used for external validation in this study. GSE76201 is a microarray dataset that includes 29 adrenal carcinoma tissues with available clinical information. The platform of the dataset is GPL96. Supplement Table S1B shows all the patients’ data from GEO dataset. We downloaded ACC mutation data from TCGA and used “perl” to calculate tumor mutation burden (TMB). SsGSEA was used to score IRGs and immune-related functions using the R package “GSVA”. Based on the results of ssGSEA score, immune cells in ACC expression profile in TCGA database were divided into high- and low-immune cell infiltration cluster using the R package “sparcl”. Supplement Figure S1 shows a flow diagram of the design of the work in this study.

Validation of the Effectuality of Immune Clustering

According to the expression levels of immune cells and stromal cells, the scores of stromal cells infiltration levels, the scores of immune cells infiltration levels and tumor purity in ACC tumor microenvironment were calculated using the R package “ESTIMATE”. First, we use ESTIMATE algorithm to calculate Stromal Score, Immune Score, ESTIMATE Score and Tumor Purity of 79 ACC samples from two clusters in TCGA database. Then, R Package “Pheatmap” was used to draw the heatmap. The R package “ggpubr” was used to draw violin plots (vioplot) of tumor purity, estimated score, immunity score and matrix score of two clusters. Then, we used the R package “CIBERSORT” to count the contents of 22 immune cells in all ACC samples, and plotted the boxplots of differently expressed immune cells in different cluster groups using the R package “GGplot2”. Furthermore, we reconfirmed differences in human leukocyte antigen (HLA)-related gene expression between the high- and low-immune cell infiltration cluster.

GSEA Enrichment Analysis

GSEA software was used for gene enrichment analysis of ACC data. R Package “GGplot2” was applied to transform the results of GSEA enrichment analysis into annotated visualized bar graph and bubble graph. P < 0.05 was considered statistically significant.
Identification of Differentially Expressed IRGs
To identify the DEGs between two clusters in the TCGA-ACC cohort, the R package “limma” was applied in the standard comparison mode. The significance criteria of DEGs were set as | logFC | >0.585, FDR < 0.05. The volcano map was also plotted to identify DEGs with “ggplot2” package in R software. Common genes of IRGs and DEGs were obtained by Venn analysis. Then, R package “pheatmap” was used to draw the heatmap of DEGs and differentially expressed IRGs. Finally, we used R package “sva” to calculate the differentially expressed IRGs and their expression levels in TCGA and GEO, respectively.

Identification of Immune-Related Prognostic Genes for ACC
Univariate Cox regression analysis was used to distinguish immune prognosis genes among ACC samples (n = 79) with complete clinical information from TCGA database and draw the corresponding forest map. The significance criteria for determining ACC immune prognostic genes was set as p < 0.05. Hazard radio >1 is the high-risk gene, <1 is the low-risk gene.

Establishment of PPI Network
Then, a PPI network of correlated nodes was established by the STRING (v11.5; https://string-db.org/). Nodes represent genes or proteins, and the lines represent protein–protein associations.

Construction of IRGs Prognostic Signature for TCGA-ACC
In the next study, the data extracted from the TCGA dataset were used as the experimental group, and the GEO dataset was set as the verification group. First, we screened survival-related genes from TCGA dataset using LASSO regression analysis to obtain significant results from univariate Cox regression by using R “glmnet” package. One thousand rounds of cross validation for tuning parameter selection are used to prevent overfitting and partial likelihood bias meets minimum criteria. Second, we constructed prognostic markers of ACC based on the expression of IRGs and correlation coefficients of LASSO regression analysis. These genes were selected to further calculate the risk score of each patient. The median value of the risk score was defined by using the R package “survminer”, where ACC patients were divided into high-risk and low-risk groups. Time-dependent receiver-operating characteristic (ROC) and Kaplan–Meier (K-M) curves were used to evaluate the clinical prognostic ability of risk scores using the R “timeROC”, “survival”, and “survminer” packages. Third, to evaluate whether risk score can be regarded as an independent predictor of overall survival in ACC patients, univariate Cox and multivariate Cox regression analysis were performed with risk score, gender and tumor grade as variables using the R “survival” package. Finally, we used R package “ggplot2” to analyze the co-expression of 22 immune cells and 13 ACC prognosis genes and draw relevant heat maps.

Verification of IRGs Prognostic Signature Used by GSE76201
In the verification group, ACC patients were also divided into high-risk and low-risk groups according to LASSO regression analysis. Then, ROC and K-M curves were used to verify the clinical prognostic ability of risk scores using the R “timeROC”, “survival”, and “survminer” packages.

The Establishment and Evaluation of the Nomogram
The nomogram was constructed to predict the survival status of ACC patients. Gender, N classification, risk, clinical stage, and T classification were used to construct the nomogram using R “rms” and “survival” packages. ROC curves and calibration curves were used to assess the accuracy of the nomogram using R “regplot”, “survminer” and “timeROC” packages.

Statistical Analysis
All statistical analyses were conducted using the R software version 4.1.0 and perl version 5.32.1. p < 0.05 was regarded as statistically significant.
Results
Validation of the Effectuality of Immune Clustering
We obtained 79 ACC samples with relevant clinical information from the GDC database. All IRGs were obtained from IMMPORT. We used ssGSEA to quantify ACC RNA sequencing data for immune cell infiltration. Finally, 29 infiltrating immune cells were obtained: aDCs, Mast cells, Type II IFN Response, Type I IFN Response, DCs, MHC class I, B cells, Neutrophils, pDCs, APC co-stimulation, HLA, T helper cells, iDCs, NK cells, CD8+ T cells, Inflammation promoting, Th1 cells, Check point, T cell co-inhibition, T cell co-stimulation, Cytolytic-activity, TIL, Th, APC co-inhibition, Parainflammation, CCR, Treg, Macrophages, Th2 cells. By applying an unsupervised hierarchical clustering algorithm (cutoff = 1.0), ACC samples were divided into two groups according to immune infiltration: high immune cell infiltration cluster (n = 17) and low immune cell infiltration cluster (n = 62) (Figure 1A). To verify the effectiveness of ssGSEA grouping, we used ESTIMATE algorithm to calculate the stromal score, immune score, ESTIMATE score and tumor purity of two clusters, respectively. The results show that stromal score, immune score, and ESTIMATE score in high immune cell infiltration cluster were higher than those in low immune cell infiltration cluster. However, tumor purity was the opposite (Figure 1B). The violin plot also shows that stromal score, immune score and ESTIMATE score in the two clusters have significant differences (P < 0.001) (Figure 1C). In addition, the box plot also showed that the expression of high immune cell infiltration cluster was significantly higher than that of low immune cell infiltration cluster in 24 HLA. (“****”, “***”, “**”, “*”, “ns” are “P<0.001” “P<0.01” “P<0.05” “no significance”) (Figure 1D).

GSEA Enrichment Analysis
KEGG analysis of genes in high and low immune cell infiltration clusters showed that these genes were related to KEGG HEMATOPOIETIC CELL LINEAGE, TYPE I DIABETES MELLITUS, AUTOIMMUNE THYROID DISEASE, ALLOGRAFT REJECTION, CYTOKINE CYTOKINE RECEPTOR INTERACTION, INTESTINAL IMMUNE NETWORK FOR IGA PRODUCTION, CELL ADHESION MOLECULES CAMS (Figure 2A and B).

Figure 1 Validation of the effectuality of immune clustering. (A) 79 ACC samples were divided into high immune cell infiltration group (Red) and low immune cell infiltration group (Blue). (B) The enrichment levels of 29 immune-related cells and types in the high immune cell infiltration group (Immunity_H) and the low immune cell infiltration group (Immunity_L). The tumor purity, ESTIMATE score, immune score and stromal score of every patient’s gene were showed combine with the clustering information. (C) The violin plot showed the difference in ESTIMATE score, immune score and stromal score between two clusters. (D) The expression of most HLAs was a significant difference in high- (red) and low- (blue) immune cell infiltration cluster. *p < 0.05, **p < 0.01, ***p < 0.001. Abbreviation: ns, no statistically significant.
Identification of Differentially Expressed IRGs Between High and Low Immune Cell Infiltration Clusters

By using the threshold value of |lgFC| > 0.585 and FDR < 0.05, we identified 2890 DEGs with high and low immune cell infiltration in the 79 ACC samples of TCGA. It included 769 upregulated genes and 2111 downregulated genes (Figure 3A). A total of 1793 IRGs were downloaded from IMMPORT and Venn analysis was used to yield 275 differentially expressed IRGs by making intersection with DEGs (Figure 3B). At the same time, we plot the heat map (Figure 3C and D) for DEGs and differentially expressed IRGs.

Identification of IRGs Prognostic Signature for ACC

We used univariate Cox regression analysis to analyze data from 79 ACC patients with complete clinical data. Finally, 57 differentially expressed IRGs were significantly associated with ACC overall survival rate (P < 0.05), with 38 high-risk genes and 19 low-risk genes (Figure 4). Supplement Table S2 shows differentially expressed IRGs.

Establishment of PPI Network

A protein interaction network analysis was performed on 109 co-expressed nodes and these nodes representing genes or proteins. If there was a connection between nodes, there was a protein interaction relationship between nodes (Figure 5). The lines of different colors represent different meanings. The lines represent protein–protein associations. The light blue line means that both proteins are from curated databases. The pink line means that both proteins are experimentally determined. The green line means that both proteins are in a gene neighborhood. The red line means that both proteins are gene fusions. The dark blue line means that both proteins are gene co-occurrence. The yellow line means that both proteins are text mining. The black line means that both proteins are co-expression and the purple line means that both proteins are protein homology.

Construction of IRGs Prognostic Signature

The data extracted from the TCGA dataset was used as the experimental group, and the GEO dataset was set as the verification group. We used the GSE76201 dataset to verify the validity of immune-related gene prognostic signature. LASSO Regression analysis was performed on 57 differentially expressed IRGs. Eventually, 13 genes including RAC2, KRAS, IL11RA, C3, GHR, VGF, TOR2A, PTGFR, TGFB2, TNFAIP3, PAK4, IRF7, SDC4 were validated (Figure 6A and B). Risk scores were calculated based on the expression levels and their coefficients of these 13 genes. Risk score = (−0.0722415*expression of RAC2) + (0.6064765*expression of KRAS) + (0.6911548*expression of IL11RA) + (−0.0375132 * expression of C3) + (−0.1608705*expression of GHR) + (0.0646402 *expression of VGF) + (0.4313901*expression of TOR2A) + (−0.0413595 * expression of PTGFR) + (−0.0223629*expression of TGFB2) + (0.3557257 *expression of TNFAIP3) + (0.1153306*expression of PAK4) + (0.5941446 *expression of IRF7) + (−0.0655002*expression of SDC4).
Patients were divided into high-risk group and low-risk group by using the median of the risk score as the critical value in both groups. The results of Kaplan–Meier survival analyses revealed that the survival rate of the low-risk group was significantly better than that of the high-risk group (P < 0.001), indicating that the risk score predicted the prognosis of ACC patients effectively (Figure 6C and D). Risk curves (Figure 6E and F) and scatter plots (Figure 6G and H) show the risk scores and survival status for all ACC patients.

Verification of IRGs Prognostic Signature
The ROC curve was used to assess the accuracy of the 13 IRGs in predicting the 1-, 3-, and 5-year overall survival of ACC patients. The AUC of 1, 3 and 5 years were 0.973, 0.970 and 0.936, respectively (Figure 7A) and the AUC of 1, 3 and 5 years were 0.722, 0.822 and 0.817, respectively (Figure 7B). It showed that the risk score has a strong predictive power. The larger the area under the curve, the better the prediction effect. These results suggest that 13 IRGs signature can be used as an important indicator for the assessment of ACC prognosis. The predictive power of the nomogram was evaluated and quantified by measuring the degree of fit between the C index in the standard curve and the baseline time predicted by the model. From the 1-, 3- and 5-year relapse calibration curves shown in Figure 7C and D, the nomogram has a good predictor on the outcome of ACC patients.

Figure 3 Identification of differentially expressed IRGs between high and low immune cell infiltration clusters. (A) The volcano plot showed that 769 up-regulated genes (red) and 2111 down-regulated genes (green) between high and low immune cell infiltration cluster in TCGA dataset. Black dots mean meaningless (P>0.05). (B) Using venn diagram to pick up the intersection, 275 differentially expressed IRGs were obtained. (C) The heat map showed the expression levels of 2890 DEGs with high and low immune cell infiltration in 79 ACC samples from TCGA. (D) The heat map showed the expression levels of 275 differentially expressed IRGs.
We applied univariate (Figure 7E) and multivariate regression analysis (Figure 7F) to verify whether the 13 IRGs signature is an independent predictor of other features, eg, gender, tumor, stage, and risk score. The results showed that the risk score is an independent predictor (P < 0.001). Finally, the heatmap was used to show the relationship between 22 immune cells and 13 IRGs signature (Figure 7G).

The Establishment and Evaluation of the Nomogram
The nomogram was constructed to predict survival status in ACC patients. Gender, N classification, risk, clinical stage, and T classification were treated as variables. Each factor in this nomogram has a specific score. Based on the reality of each sample, the score of each prognostic factor was added to get a total score (Figure 8A). We can obtain patients’ survival status of 1, 3 and 5 years and plot the corresponding ROC curves (Figure 8B). The ROC curve predicted that the
AUC of ACC patients 1-, 3- and 5-year overall survival rate was 0.799, 0.969 and 0.969, indicating good prediction accuracy. The calibration curve obtained compared well with the ideal model (Figure 8C).

**Discussion**

ACC is a rare heterogeneous malignancy with unclear mechanisms of tumorigenesis and poor prognosis. In recent years, the findings of the tumor microenvironment have increased a deep understanding of tumor progression and immunotherapy. Habra et al conducted a single-center Phase II clinical trial on 16 ACC patients with a previous history of treatment failure in the past 6 months. They found that single-agent pembrolizumab has modest efficacy as a salvage therapy in ACC regardless of the tumor’s hormonal function, MSI status, or PD-L1 status. There was a low rate of severe adverse events in most study participants. Many studies have shown that immunotherapy plays a critical role in many tumors and is likely to be a key part of the clinical management of tumors. Therefore, the identification of ACC IRGs will be critical to predict the prognosis of ACC and the development of immunotherapy. In the study, we identified 13 IRGs regulating ACC prognosis by using the TCGA ACC dataset combined with ssGSEA and LASSO analysis. Then, we used the GSE76201 dataset to verify the validity of 13 IRGs. And we...
Figure 6 Construction of IRGs prognostic signature (A and B). The LASSO Cox analysis identified 13 genes associated with prognosis. (C) Train group, (D) Test group. The results of Kaplan–Meier survival analysis revealed that the survival rate of the low-risk group was significantly higher than that of the high-risk group (P<0.001). (E) Train group, (F) Test group. (G) The risk curve of every sample arranged by risk score. (H) Train group, (I) Test group. The scatter plot of ACC samples survival overview. The red and green dots stand for death and survival, respectively.
Figure 7 Verification of IRGs prognostic signature. (A) Train group, (B) Test group. The ROC curve was used to evaluate the precision of risk score in predicting overall survival of ACC patients at 1-, 3-, and 5-year. (C) Train group, (D) Test group. Calibration curve for 1-, 3-, and 5-year recurrence rate of ACC in TCGA datasets. (E) Univariate Cox regression demonstrated that clinical stage and the risk score could predict Overall Survival (OS). (F) Multivariate Cox regression showed that the risk score was independent prognostic factors. (G) Heatmap of correlation matrix between the immune cell and 13 IRGs signature. *p < 0.05, **p < 0.01, ***p < 0.001.
screened independent risk factors influencing ACC prognosis and constructed an ACC prognostic nomogram for clinical reference.

In this study, we applied an unsupervised hierarchical clustering algorithm to classify 79 samples based on the enrichment of 29 immune cell types into high- (n = 17) and low immune cell infiltration (n = 62). The stromal score, immune score, ESTIMATE score and tumor purity varied significantly between the two clusters (P < 0.05). Then, the variability in the ACC immune microenvironment was verified by the expression levels of HLA. Finally, LASSO regression analysis of differentially expressed IRGs was performed to construct 13 IRGs signature to predict the prognosis of ACC patients. The samples were divided into a low-risk group and a high-risk group based on the median risk score, with a lower survival rate for ACC patients in the high-risk group. K-M curve and ROC curve showed that 13 IRGs had good predictive ability.

KRAS, IL11RA, C3, TOR2A, TNFAIP3, PAK4, IRF7 were highly expressed in ACC and the rest of the IRGs (RAC2 GHR, VGF PTGFR, TGFBR2, SDC4) were low. Most of the 13 IRGs have been found to be associated with prognosis in different types.
of cancer. KRAS G12D resulted in the transition of macrophages into M2-like tumor-promoting phenotype by STAT3-dependent fatty acid oxidation. Thus, disruption of KRAS G12D release and uptake eliminated macrophage-mediated pancreatic cancer stimulation in mouse models. Some studies showed that KRAS in macrophages expression levels were correlated with poor survival in pancreatic cancer patients. IL11RA may play a role in the progression of endometrial cancer, and IL11 can be used as a diagnostic marker for early endometrial cancer. Local C3 deposition in the tumor microenvironment is a relevant immune feature for predicting gastric cancer prognosis. It may abnormally activate the JAK2/STAT3 pathway, thereby allowing for progression of gastric cancer. The inhibition of enzyme topoisomerase 2 alpha (TOP2a) during replication and transcription is one of the main mechanisms of anthracyclines, it perturbs DNA structure and function, leading to cell death. Elevated TOP2a activity and elevated TOP2a gene expression levels caused by amplification of the TOP2a site (17q21.2) in tumor cells were correlated with anthracycline chemosensitivity. TNFAIP3 is a potent regulator of ubiquitin (UB)-dependent signaling. Deletion of the TNFAIP3 gene is associated with the NFκB pathway in lymphoma patients and can slow down recurrence or progression in diagnosed patients. It was shown that PAK4 as a promoter of breast cancer, can overcome the caducity of oncogene-induced and reveal a selective vulnerability of cancer to PAK4 inhibition. A large number of genes suppressed in bone metastasis in breast cancer are targets of the interferon regulator Irf7. Recovery of Irf7 or interferon administration in tumor cells resulted in reduced bone metastasis and prolonged survival.

We then verified whether gender, tumor, stage, and risk score were independent predictors of other features by applying both univariate and multivariate regression analyses, showing that the risk score was an independent predictor. Finally, we constructed a nomogram predicting the survival status of ACC patients based on gender, N classification, risk, clinical stage, T classification. From this nomogram, we were able to predict survival at 1, 3, and 5 years for ACC patients, with 79 samples having 100% survival for less than 5 years. The calibration curves indicate the nomogram could better predict prognosis in ACC patients. Nevertheless, there are still some limitations of our study. First, larger samples are needed to further strengthen our findings. Second, all data downloaded online come from public datasets and require further validation in prospective clinical trials, but ACC cases are clinically rare, so the feasibility of prospective studies needs to be further explored.

In conclusion, this study identified 13 IRGs regulating ACC prognosis and constructed a nomogram predicting survival status in ACC patients. This may provide a reference for immunotherapy in ACC patients.

Statement of Ethics
The study was submitted for review by the Ethics Committee of The First Affiliated Hospital of Chongqing Medical and a waiver for ethics was granted.

Consent for Publication
All authors have read and approved the content, and they agree to submit it for consideration for publication in the journal.

Funding
There is no funding to report.

Disclosure
The authors report no conflicts of interest in this work.

References