MEDS Basic Medicine (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/medbm.2024.020113
ISSN 2616-2091 Vol. 2 Num. 1

Creating and Utilizing an Immune-Related Gene Prognosis Framework for Kidney Clear Cell Carcinoma

Jiahui Luo[1], Qingbo Zhou[2]

[1]*The Fourth School of Clinical Medicine, Zhejiang Chinese Medical University, Hangzhou, China*
[2]*Department of Internal Medicine, Shaoxing Yuecheng People's Hospital, Shaoxing, China*

Keywords: Immunomodulatory Genes, Renal clear cell carcinoma, Prognostic model, Consensus clustering, Prognostic model

Abstract: In this study, we initially classified renal cell carcinoma (ccRCC) samples into three distinct clusters based on immune expression patterns, utilizing an analysis of immunomodulatory genes. Through comprehensive examination of the differential genes across these clusters, a prognostic risk assessment framework was developed, aimed at determining risk levels to assist in prognostic evaluation and targeted therapy for ccRCC patients. Employing a consensus clustering method, 435 shared differential genes were pinpointed, predominantly associated with a range of immunomodulation pathways, through examining the immunomodulatory gene expressions in TCGA-KIRC patient samples. Furthermore, a univariate COX regression analysis allowed for the identification of 152 genes with significant prognostic associations, from which 11 pivotal genes were selected for the model through LASSO regression analysis. Utilizing these identified genes, a new prognostic risk assessment tool was crafted, and its predictive efficiency was confirmed via analysis of ROC curves. This newly developed prognostic tool, focusing on immunomodulatory genes, underwent thorough validation against external datasets, thereby enhancing the precision of clinical prognosis evaluations for ccRCC patients.

1. Introduction

Originating from the epithelial cells of renal tubules, renal cell carcinoma (RCC) stands as the predominant type of kidney cancer[1]. Specifically, clear cell RCC (ccRCC) accounts for a major part of urinary system cancers, characterized by its challenging prognosis. Each year, it is estimated that more than 225,000 individuals are diagnosed with this condition, resulting in over 100,000 deaths annually[1]. Despite progress in early detection and therapeutic strategies that have improved early-stage diagnosis and patient survival, ccRCC continues to exhibit high rates of occurrence and mortality[2]. In the absence of treatment, ccRCC is highly aggressive, leading to invasion and metastasis, and the survival rate drops to around 10% five years after metastasis has occurred[2]. While surgical intervention is pivotal in treating ccRCC, close to 40% of individuals undergo recurrence after surgery[3], posing a significant burden on both patients and their relatives[4].

Immunotherapy has emerged as a promising avenue in the management of renal cancer, underscoring the significance of immunomodulators (IMs) within the complexity of tumor behavior. Research into IM agonists and antagonists has illuminated their capability to directly affect the

tumor microenvironment and cellular interfaces, paving the way for innovative therapeutic strategies in treating clear cell renal cell carcinoma (ccRCC). Our study delves into the synergies between immunomodulatory genes and tumor dynamics, leading to the establishment of a gene expression-based prognostic evaluation tool. This tool, after rigorous validation with patient cohorts from ArrayExpress, has shown efficacy in forecasting the prognosis for ccRCC.

2. Materials and methods

2.1 Clear cell renal cell carcinomas datasets and preprocessing

RNA-seq data for Kidney renal clear cell carcinoma (KIRC) (HTSeq counts), along with clinicopathological information encompassing patient age, gender, grades and stages of pathology, TNM classification, life status, and duration of survival, were acquired from the UCSC Xena platform (https://xena.ucsc.edu/). Additionally, a collection of 75 immunomodulatory genes was sourced from the TISIDB database. Analysis excluded patients lacking survival data. To facilitate survival analysis, normalization of RNA-seq raw counts was performed by applying the TPM technique, followed by conversion to log2TPM.

The validation of our dataset utilized ArrayExpress (https://www.ebi.ac.uk/arrayexpress/), specifically focusing on the E-MTAB-1980 dataset, which includes data from 106 subjects. To assess our predictive model's efficacy regarding the response to anti-PD-1/L1 therapy, we sourced the IMvigor210 study data from (http://research-pub.gene.com/IMvigor210CoreBiologies/). Data normalization was achieved through Z-score transformation, employing the scale() function in R for standardization, followed by graphical representation to precede the phases of model construction and validation.

2.2 Consensus clustering

We implemented K-means clustering to categorize tumor samples into unique subtypes influenced by the expression of immunomodulatory (IM) genes related to cuproptosis. To identify the best cluster arrangement, we utilized the ConsensusClusterPlus package, following the guidelines of consensus clustering methods, and established the ideal cluster count. This process was rigorously validated through 1,000 iterations to ensure the robustness of our analysis. Furthermore, survival rates were evaluated across the identified tumor subtypes to ascertain associations with adverse survival outcomes.

2.3 Discovering Differential Gene Expression across Clusters

Sequencing data underwent analysis with the 'limma' toolkit to pinpoint genes exhibiting differential expression across the three clusters through pair-wise examination. Exploring molecular and biological disparities specific to KIRC species led to the discovery of differentially expressed genes (DEGs), selecting those with an adjusted p-value below 0.05 and an absolute log2 fold change greater than 0.5. By conducting comparative analyses between each pair of clusters, we delineated three sets of DEGs. We then proceeded to intersect these groups, extracting both their expression profiles within the specimens and pertinent clinical data for further investigative processes.

2.4 Crafting a Prognostic Scoring System Utilizing Differentially Expressed Genes

For the training set, the TCGA cohort of 531 primary KIRC samples with survival information

were selected as the training cohort. To be consistent with the rest of the testing data, we normalized both training and testing sets using the scale function. First, we performed an univariate COX analysis to identify significant prognostic genes. To avoid overfitting, Lasso regression was performed on the multifactor models. An analysis of stepwise regression was conducted to identify genes that were most important.

Patients within the training set were divided into subgroups of lower and higher risk, based on the median value of the prognostic scores as the division point. The effectiveness and precision of the prognostic model were assessed by utilizing the curves of receiver operating characteristic (ROC).

3. Results

3.1 Formulating a Predictive Scoring Model Leveraging Immunomodulatory Gene Signatures

Figue 1

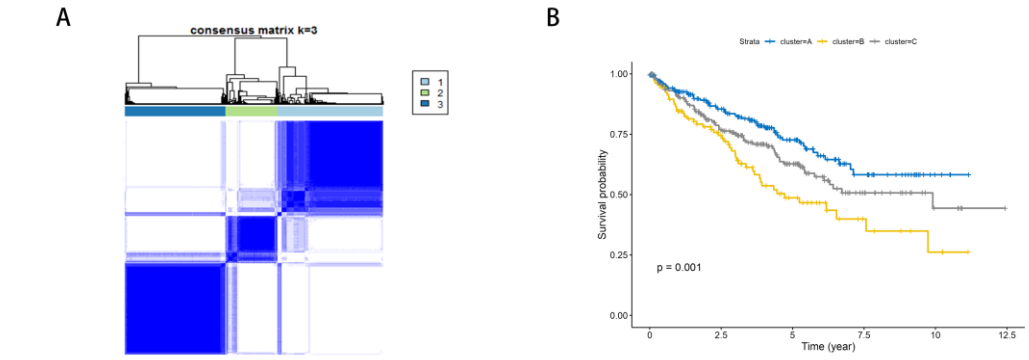

Figure 1A: The consensus matrix for all KIRC samples displays the clustering stability across 1,000 iterations, with all samples being categorized into an optimal number of subtypes (k equals 3).
Figure 1B: Kaplan-Meier plots illustrate the variance in overall survival among immunomodulatory subtypes A, B, and C, with a significant difference (p equals 0.001).

Utilizing IM genes, TCGA-KIRC patients were categorized into three distinct clusters via consensus clustering (Figure 1A). The decision on the optimal trio of clusters was informed by aiming for strong similarity within groups versus distinctiveness between them. The clusters, labeled A, B, and C, comprised 216, 107, and 208 individuals respectively. Survival outcomes, as illustrated by Kaplan-Meier plots, significantly varied across the clusters, with Cluster A displaying notably poorer survival outcomes when juxtaposed with Clusters B and C (log-rank p = 0.001) (Figure 1B).

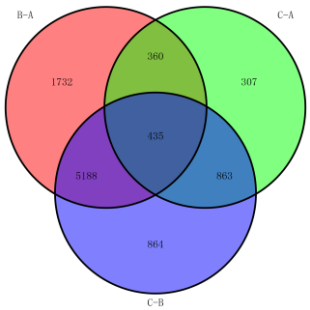

Figure 2: Wayne plot showing three differential gene sets obtained after two-by-two differential analysis of the three subtypes, with 435 overlapping differential genes.

We performed two-by-two analysis of variance (|Log FC > 0.5|, P<0.05) for clusters A, B and C, respectively, and obtained three sets of differential genes. Their intersection was taken and 435 overlapping differential genes were obtained (Figure 2).

Figure 3

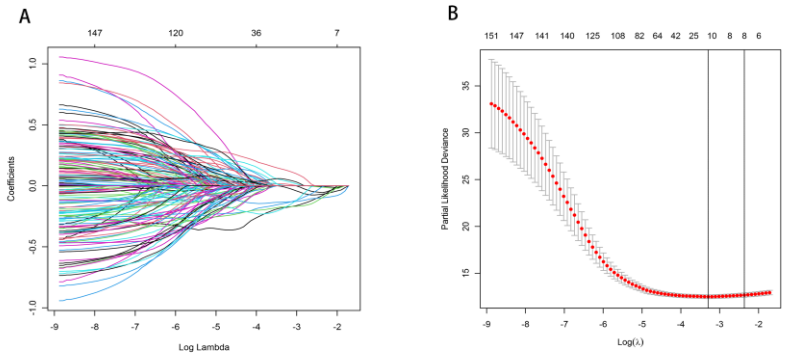

Figure 3A: Plot of LASSO regression coefficients. Figure 3B: Plot of LASSO regression parameters

Initially, the expression data for KIRC was standardized using the scale function. A univariate COX regression analysis then identified 152 genes with a significant connection to prognosis. To minimize the risk of overfitting in our study, the selection of λ was based on the smallest lambda value (lambda.min) in the subsequent LASSO regression analysis. This approach allowed us to precisely select the next 11 genes deemed most effective for inclusion in our model (Figure 3A-B).

Figure 4

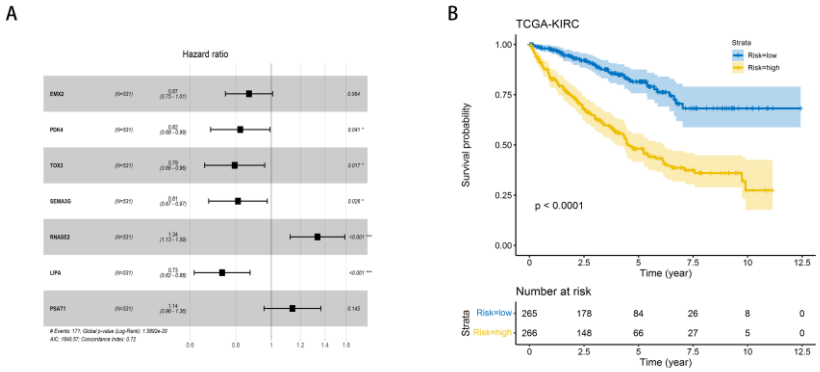

Figure 4A: Prognostic model of multifactorial COX regression. Figure 4B: Survival Outcomes Compared Between High- and Low-Risk Categories in TCGA via Kaplan-Meier Method

Subsequently, the optimal model was identified using a stepwise regression approach (Figure 4A). The risk scoring formula was derived by integrating the COX regression coefficients with the expression levels of the genes. For the 531 KIRC samples in our training dataset, risk scores ('riskScore') were computed according to the model's scoring formula. By setting the median 'riskScore' as a threshold, samples were categorized into groups of higher or lower risk. Analysis of survival outcomes revealed significantly reduced survival durations in the higher-risk category compared to the lower-risk one, with a noteworthy log-rank test result ($p < 0.001$) depicted in Figure 4B. An increase in 'riskScore' correlated with heightened mortality risks and shorter survival periods among KIRC patients, indicating the risk score's potential as a reliable prognostic indicator for KIRC patient outcomes.

3.2 Assessing the Predictive Value of the Prognostic Scoring System

To confirm the broad utility of our scoring framework, we conducted an evaluation using a validation cohort comprising 106 ccRCC samples from the E-MTAB-1980 dataset within the ArrayExpress archive. Initially, data normalization was carried out employing the 'scale' technique. Subsequently, 101 patients with ccRCC were categorized into groups of lower (n=55) and higher risk (n=46), utilizing the identical threshold as established in the training cohort. Kaplan-Meier survival analysis revealed a notable disparity in survival outcomes between these groups ($p < 0.001$) illustrated in Figure 5, with individuals in the higher risk category experiencing less favorable clinical prognoses. Furthermore, the effectiveness of our model was underscored through ROC curve evaluation, demonstrating impressive AUC metrics of 0.87, 0.85, and 0.84 at 1, 3, and 5-year intervals, respectively.

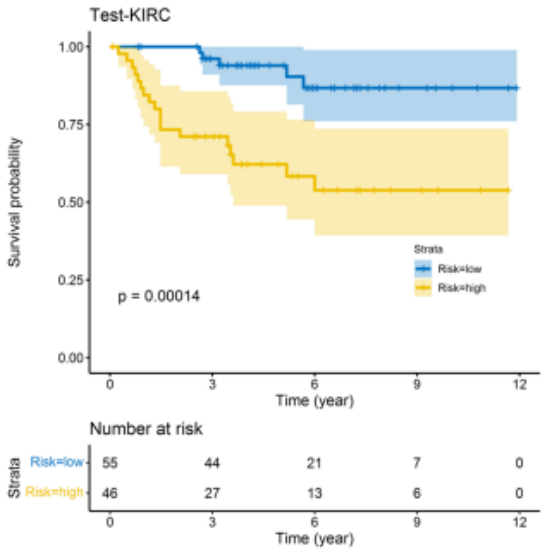

Figure 5: Comparative Survival Analysis Between High- and Low-Risk Categories Within the Array Express Validation Set via Kaplan-Meier Estimation.

In analyzing various gene clusters, notable variances in the expression of immunoregulatory genes were observed, aligning with anticipated outcomes. Furthermore, an examination of how risk scores correlate with the three distinct clusters revealed significant variations. Specifically, cluster B exhibited the highest median risk score, with cluster C following, and cluster A showing the lowest score, mirroring their respective prognostic impacts. This underlines the predictive accuracy of the risk scores across different clusters.

4. Discussion

In our research, we have pioneeringly utilized immunomodulatory genes for the classification of ccRCC specimens into three unique clusters, distinguished by their immunological expression profiles. This approach not only sheds light on the heterogeneity of ccRCC but also paves the way for a more nuanced understanding of its immune landscape. Building upon this classification, we developed a prognostic scoring system that leverages the differential genes among these clusters. This system allows for the calculation of risk scores crucial for guiding prognostic analyses and tailoring treatment strategies for ccRCC patients, underscoring the potential of precision medicine in oncology.

The significance of predicting oncological outcomes cannot be overstated, as it holds profound implications for patient care and clinical decision-making [5]. The ability to accurately forecast

survival outcomes through immune-related prognostic signatures and models represents a significant advancement in the field[6]. The work corroborates our findings, reinforcing the utility of immune genes as reliable predictors for ccRCC prognosis[7]. These investigations underscore a growing agreement on the significance of immunogenomic analysis for forecasting outcomes in kidney cancer, indicating a critical transition towards therapeutic strategies that are both more precise and efficacious.

The model's genes, associated with immune infiltration and checkpoint genes, shed light on immunotherapy's potential. Key findings include the tumor-suppressing roles of EMX2, PDK4's potential suppressive function[6], TOX3's inhibition of ccRCC cell migration and invasion[7], SEMA3 family's dual roles in tumorigenesis[8], RNASE2's significance in immune responses[9], LIPA's prognostic impact[10], and PSAT1's expression linked to poor differentiation and metastases[11]. These genes' roles in tumor development and KIRC prognosis need further validation.

Our study contributes to this growing body of evidence by providing a robust prognostic model based on immunomodulatory genes, which has been rigorously validated against external datasets. This validation not only confirms the model's predictive accuracy but also its practical applicability in clinical settings, offering a valuable tool for enhancing the prognostic assessment of ccRCC patients.

However, our study is not without its limitations. The reliance on data from public databases, while invaluable, necessitates further validation through multicenter sequencing or experimental data to ensure the generalizability and reliability of our findings. Moreover, the intricate relationships among the genes within our model warrant deeper investigation to unravel their biological interactions and contributions to ccRCC progression. This exploration could potentially unveil novel therapeutic targets and pathways for intervention.

To sum up, this research emphasizes the crucial importance of immunomodulatory genes in determining the prognosis for ccRCC, while also illustrating how genetic profiling can enhance the precision of risk stratification and the formulation of treatment strategies. As we move towards an era of precision oncology, it is imperative that future research continues to explore and validate these promising avenues, with the ultimate goal of improving patient outcomes in ccRCC and beyond.

Acknowledgements

Our team extends sincere thanks for the generous financial backing from the Science Technology Bureau of Shaoxing (Grant/Award Number: 2023A14035) and the support from the Zhejiang Province Traditional Chinese Medicine Science and Technology Plan Project (Grant/Award Number: 2024ZF168).

References

[1] Melo, A.A.M., Gallegos, J.A.O., and Merchan, J.R. (2023). Abstract C121: Epidemiological characteristics and patterns of recurrence of renal cell carcinoma in Hispanics: A single US center cohort study. Cancer Epidemiology, Biomarkers & Prevention, 32, C121-C121. 10.1158/1538-7755.Disp22-c121.

[2] Schiavoni, V., Campagna, R., Pozzi, V., Cecati, M., Milanese, G., Sartini, D., Salvolini, E., Galosi, A.B., and Emanuelli, M. (2023). Recent Advances in the Management of Clear Cell Renal Cell Carcinoma: Novel Biomarkers and Targeted Therapies. Cancers (Basel), 15. 10.3390/cancers15123207.

[3] Lebacle, C., Pooli, A., Bessede, T., Irani, J., Pantuck, A.J., and Drakaki, A. (2019). Epidemiology, biology and treatment of sarcomatoid RCC: current state of the art. World J Urol, 37, 115-123. 10.1007/s00345-018-2355-y.

[4] Liu, F., Davaro, F., Wong, R., Siddiqui, S., and Hamilton, Z. (2022). Young age is associated with decreased recurrence for renal cell carcinoma. Can J Urol, 29, 11142-11149.

[5] Correa, A.F., Jegede, O., Haas, N.B., Flaherty, K.T., Pins, M.R., Messing, E.M., Manola, J., Wood, C.G., Kane, C.J.,

Jewett, M.A.S., et al. (2019). Predicting Renal Cancer Recurrence: Defining Limitations of Existing Prognostic Models with Prospective Trial-Based Validation. J Clin Oncol, 37, 2062-2071. 10.1200/jco.19.00107.
[6] Choiniere, J., Wu, J., and Wang, L. (2017). Pyruvate Dehydrogenase Kinase 4 Deficiency Results in Expedited Cellular Proliferation through E2F1-Mediated Increase of Cyclins. Mol Pharmacol, 91, 189-196. 10.1124/mol. 116. 106757.
[7] Jiang, B., Chen, W., Qin, H., Diao, W., Li, B., Cao, W., Zhang, Z., Qi, W., Gao, J., Chen, M., et al. (2019). TOX3 inhibits cancer cell migration and invasion via transcriptional regulation of SNAI1 and SNAI2 in clear cell renal cell carcinoma. Cancer Lett, 449, 76-86. 10.1016/j.canlet.2019.02.020.
[8] Toledano, S., Nir-Zvi, I., Engelman, R., Kessler, O., and Neufeld, G. (2019). Class-3 Semaphorins and Their Receptors: Potent Multifunctional Modulators of Tumor Progression. Int J Mol Sci, 20. 10.3390/ijms20030556.
[9] Xiang, Y., Zhou, S., Hao, J., Zhong, C., Ma, Q., Sun, Z., and Wei, C. (2020). Development and validation of a prognostic model for kidney renal clear cell carcinoma based on RNA binding protein expression. Aging (Albany NY), 12, 25356-25372. 10.18632/aging.104137.
[10] Zhang, Y., Tang, M., Guo, Q., Xu, H., Yang, Z., and Li, D. (2022). The value of erlotinib related target molecules in kidney renal cell carcinoma via bioinformatics analysis. Gene, 816, 146173. 10.1016/j.gene.2021.146173.
[11] Li, M.K., Liu, L.X., Zhang, W.Y., Zhan, H.L., Chen, R.P., Feng, J.L., and Wu, L.F. (2020). Long non-coding RNA MEG3 suppresses epithelial-to-mesenchymal transition by inhibiting the PSAT1-dependent GSK-3β/Snail signaling pathway in esophageal squamous cell carcinoma. Oncol Rep, 44, 2130-2142. 10.3892/or.2020.7754.