Research article

Machine learning-based prognosis signature for survival prediction of patients with clear cell renal cell carcinoma

Siteng Chen [a,1], Tuanjie Guo [b,1], Encheng Zhang [b,1], Tao Wang [b], Guangliang Jiang [c], Yishuo Wu [d], Xiang Wang [b,*], Rong Na [e,**], Ning Zhang [c,***]

[a] Department of Urology, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[b] Department of Urology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[c] Department of Urology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[d] Department of Urology, Huashan Hospital, Fudan University, Shanghai, China
[e] Department of Surgery, Queen Mary Hospital, The University of Hong Kong, Hong Kong SAR, China

ARTICLE INFO

ABSTRACT

The sole clinicopathological characteristic is not enough for the prediction of survival of patients with clear cell renal cell carcinoma (ccRCC). However, the survival prediction model constructed by machine learning technology for patients with ccRCC using clinicopathological features is rarely reported yet. In this study, a total of 5878 patients diagnosed as ccRCC from four independent patient cohorts were recruited. The least absolute shrinkage and selection operator analysis was implemented to identify optimal clinicopathological characteristics and calculate each coefficient to construct the prognosis model. In addition, weighted gene co-expression network and gene enrichment analysis associated with risk score were also carried out. Three clinicopathologic features were selected for the construction of the prognosis risk score model as the prognostic factors of ccRCC, including tumor size, tumor grade, and tumor stage. In the CPTAC (Clinical Proteomic Tumor Analysis Consortium) cohort, the General cohort, the SEER (Surveillance, Epidemiology, and End Results) cohort, and the Huashan cohort, patients with high-risk score had worse clinical outcomes than patients with low-risk score (hazard ratio 5.15, 4.64, 3.96, and 5.15, respectively). Further functional enrichment analysis demonstrated that our machine learning-based risk score was significantly connected with some cell proliferation-related pathways, consisting of DNA repair, cell division, and cell cycle. In summary, we developed and validated a machine learning-based prognosis prediction model, which might contribute to clinical decision-making for patients with ccRCC.

1. Introduction

As a highly aggressive carcinoma, the renal tumor has become one of the most lethal diseases of urological cancers. In 2022, about 79,000 new cancer cases and 13,920 cancer deaths related to kidney and renal pelvis are predicted to be found in America [1]. As the most common solid lesion in the kidney, renal cell carcinoma (RCC) occupies approximately 90% of all renal malignancies [2].

The Fuhrman grading system is currently recognized as the most predictive grading system for clear cell renal cell carcinoma (ccRCC) [3], which has also been proved to be a prognostic factor for ccRCC [4, 5].

The tumor stage is also one of the important clinical characteristics for evaluating the clinical outcomes of patients with ccRCC [6]. However, the decision of the optimal surgical procedure for patients is mainly based on tumor size [7]. For patients with high-grade ccRCC, the associated risk increased by 13% for every 1 cm of tumor enlargement [8]. Intricate relationships could also be found among these clinicopathologic features in ccRCC. Until now, some transcriptome-based prognostic markers have also been reported to predict survival outcomes for patients with ccRCC [9, 10, 11]. However, transcriptome sequencing costs a lot of manpower and financial resources, so their potential applications in clinical practices are limited. Therefore, it is still urgent to develop an

* Corresponding author.
** Corresponding author.
*** Corresponding author.
E-mail addresses: xiang.wang1@shgh.cn (X. Wang), narong.hs@gmail.com (R. Na), zn12235@rjh.com.cn (N. Zhang).
[1] Equal contributors and co-first authors.

economical and convenient survival prediction model to improve clinical practicability.

Machine learning is the science that gets the computer to learn without being explicitly programmed. As a promising technology, machine learning is becoming widespread in studies among multiple malignant tumors, including skin cancer [12], breast carcinoma [13], and neurologic tumors [14]. Machine learning is widely accepted to bring about dramatic changes in the individualized diagnosis and treatment of patients [15]. Currently, studies using machine learning to predict the classification, nuclear grade, and prognosis of RCC have been reported using data from radiomics [16, 17, 18, 19]. The identification of mortality-risk-associated missense variants in clear cell renal cell carcinoma using deep learning has also been well studied [20]. However, studies using machine learning to predict the prognosis of ccRCC patients with more accessible data such as clinicopathological characteristics have not been reported. In our research, we developed and validated a prognosis risk score model based on the clinicopathologic characteristics of patients with ccRCC from 4 independent patient cohorts using machine learning algorithm, which could help to make up for the lack of current clinical prognosis prediction for patients with ccRCC. The workflow of this study is shown in Figure 1.

2. Materials and methods

2.1. Patient cohorts and data resources

A total of 5878 patients diagnosed with ccRCC from 4 independent patient cohorts were recruited for analysis in this study. All included patients should meet the following inclusion criteria: (a) pathological evidence to diagnose a single type of primary ccRCC; (b) complete clinical and pathological characteristics, including age, gender, tumor size, tumor grade, and tumor stage; (c) with access to clinical follow-up information for more than three years after surgical treatment.

After eliminating the participants unqualified, 314 patients, who were diagnosed with ccRCC in Shanghai General Hospital from January 2012 to December 2018 were included in the General cohort. The records of 137 patients from Huashan hospital, who underwent surgery surgical from October 2012 to March 2015 were also retrospectively reviewed and defined as the Huashan cohort. Additionally, in this study another 98 patients from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) [21] and 5329 patients from the Surveillance, Epidemiology, and End Results (SEER) program [22] met the inclusion criteria were also recruited.
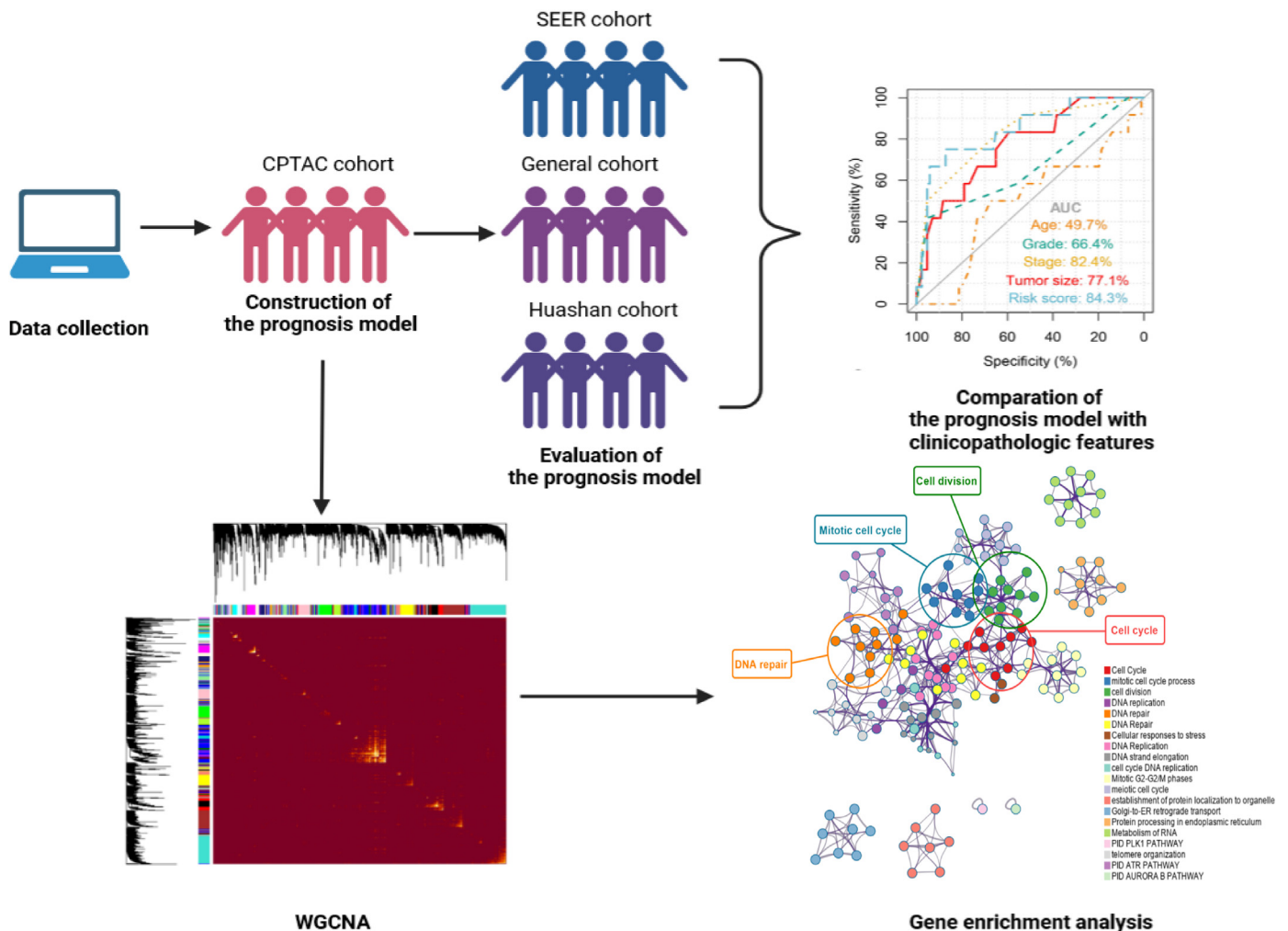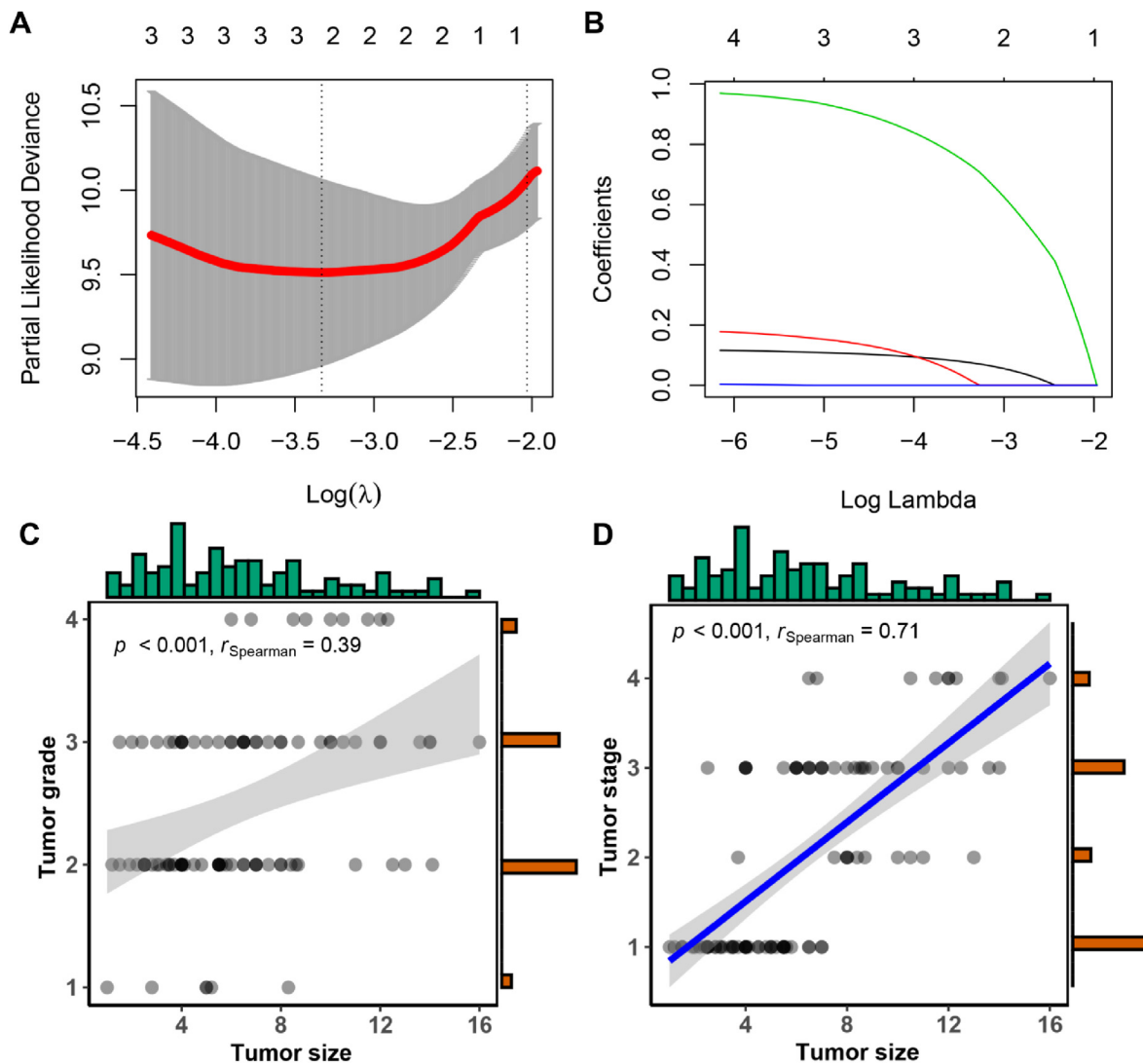

Figure 1. The workflow of this study.

Figure 2. Construction of the machine learning-based prognosis model in the CPTAC cohort. (A–B) the tenfold cross-validated error and respective coefficients at varying levels of penalization plotted against the log (lambda) sequence for the least absolute shrinkage and selection operator analysis, respectively. (C) Correlation analysis of the tumor size and tumor grade for ccRCC patients, bar plot on the top and the right represent the proportion of tumor size and grade, respectively. (D) Correlation analysis of the tumor size and tumor stage for ccRCC patients, bar plot on the top and the right represent the proportion of tumor size and stage, respectively. CPTAC, Clinical Proteomic Tumor Analysis Consortium; ccRCC, clear cell renal cell carcinoma.

2.2. Construction of the prognosis model

In this study, the least absolute shrinkage and selection operator (LASSO) via *glmnet* package in R [23] was carried out to identify optimal clinicopathological characteristics and calculate each coefficient to construct the prognostic model in the CPTAC cohort. The lambda value was set as 1000 to ensure the robustness of the LASSO model. The alpha value was set as 1, and other hyperparameters were set as default values. Then, the machine learning-based risk score is calculated by accumulating the product of the selected eigenvalue values and their respective coefficients.

2.3. Weighted gene co-expression network and gene enrichment analysis

Normalized transcriptomic data of ccRCC patients were acquired from the CPTAC cohort [21]. Genes with less than 70% transcriptome value were excluded from the total sample for further analysis. We conducted weighted gene co-expression network analysis (WGCNA) based on valid 17067 genes by *WGCNA* package in R [24] to construct co-expression gene networks in ccRCC. When the soft-thresholding

power of β value was defined as 6, which was recommended by the function of *pickSoftThreshold*, 17067 genes were hierarchically clustered into 26 gene modules. Correlation analysis between gene modules and the clinicopathologic feature was further performed to identify the optimal gene module with the highest correlation with the machine learning-based risk score. Subsequently, gene enrichment analysis was carried out to explore the potential biological mechanisms in which the risk score might be involved via Metascape [25].

2.4. Statistical analysis

In this study, the software of R 3.6.2 was used for data analyses and visualization. Kaplan–Meier (KM) curve analysis with hazard ratio (HR) and 95% confidence interval (CI) were implemented to compare different overall survival (OS) and disease-free survival (DFS) outcomes through the log-rank test. The cut-off value for identifying high or low risk was set as the median value in each patient cohort. The evaluation of the machine learning-based prognosis prediction model was performed using receiver operating characteristic curve (ROC) analysis with the area under curve (AUC) value of the 3-year survival prediction.
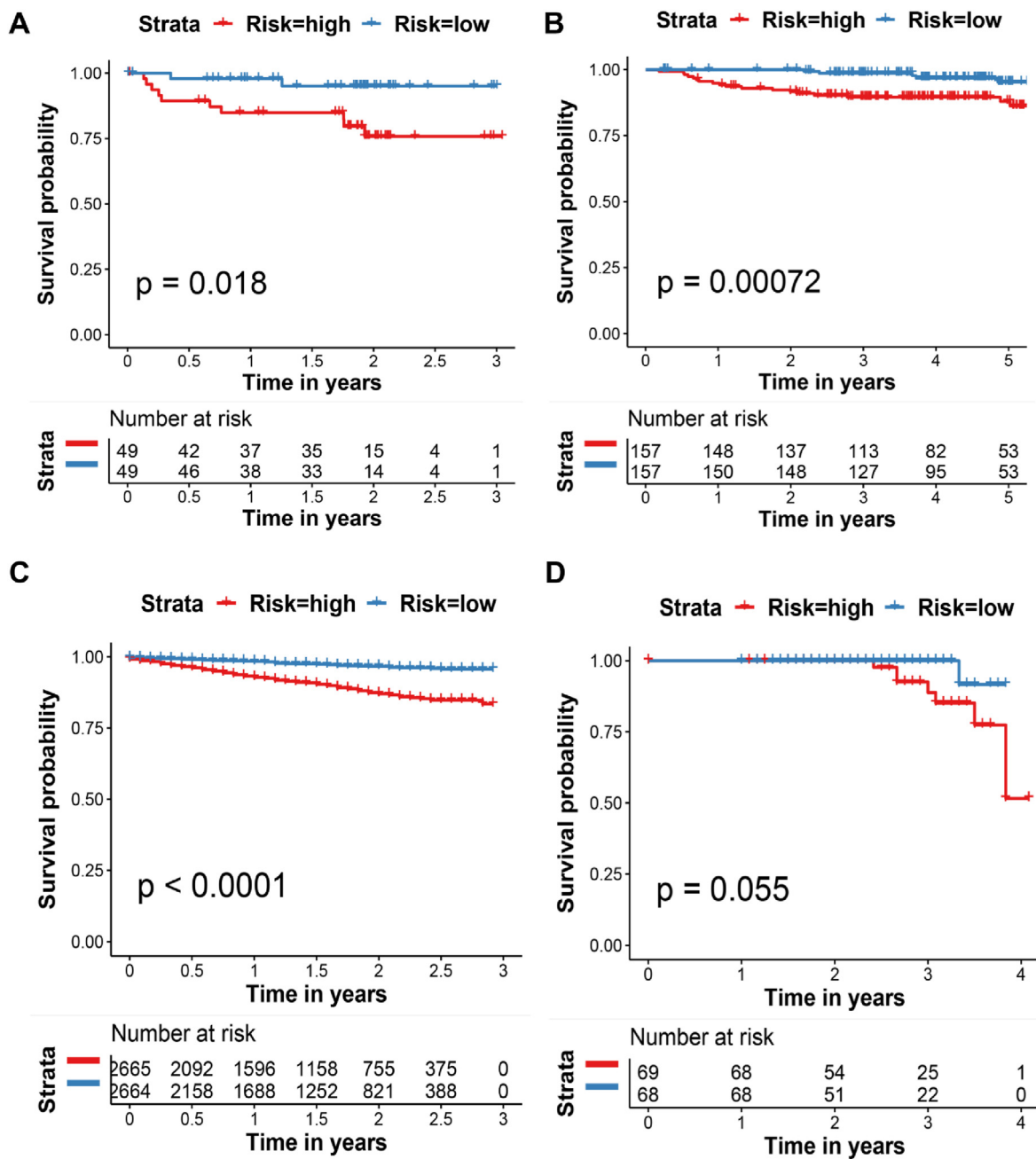
Figure 3. Survival analysis of the machine learning-based prognosis model in multiple patient cohorts. (A. B. D) Kaplan-Meier curve analysis of disease-free survival comparation in the CPTAC cohort, the General cohort, and the Huashan cohort, respectively. (C) Kaplan-Meier curve analysis of overall survival comparation in the SEER cohort. CPTAC, Clinical Proteomic Tumor Analysis Consortium; SEER, Surveillance, Epidemiology, and End Results.

3. Results

3.1. Construction of the machine learning-based prognosis risk score model for patients with ccRCC

Basic clinical characteristics of 5878 patients from the General cohort, Huashan Cohort, CPTAC cohort, and SEER cohort were shown in Table S1. Four clinicopathological characteristics, including age, tumor size, tumor grade, and tumor stage, were used for LASSO analysis in the CPTAC cohort. As illustrated in Figure 2A, the left vertical line was equal to the minimum ten-fold cross-validation error arrived at 3, which means that 3 features were screened out as the most important prognostic factors for ccRCC patients, including tumor size, tumor grade, and tumor stage. The

regression coefficients for each selected features were also acquired from Figure 2B (coefficients $_{tumor\ size}$ = 0.073198238, coefficients $_{tumor\ grade}$ = 0.008798867, coefficients $_{tumor\ stage}$ = 0.723105654). Then, the machine learning-based risk score was calculated by accumulating the product of the selected feature values and their respective coefficients. Correlation analysis revealed that tumor size had a positive correlation with tumor grade (Figure 2C) and tumor stage (Figure 2D).

3.2. Evaluation of the machine learning-based prognosis risk score model in clinical practice

To evaluate the machine learning-based prognosis model for patients with ccRCC, we performed a KM curve survival analysis in the
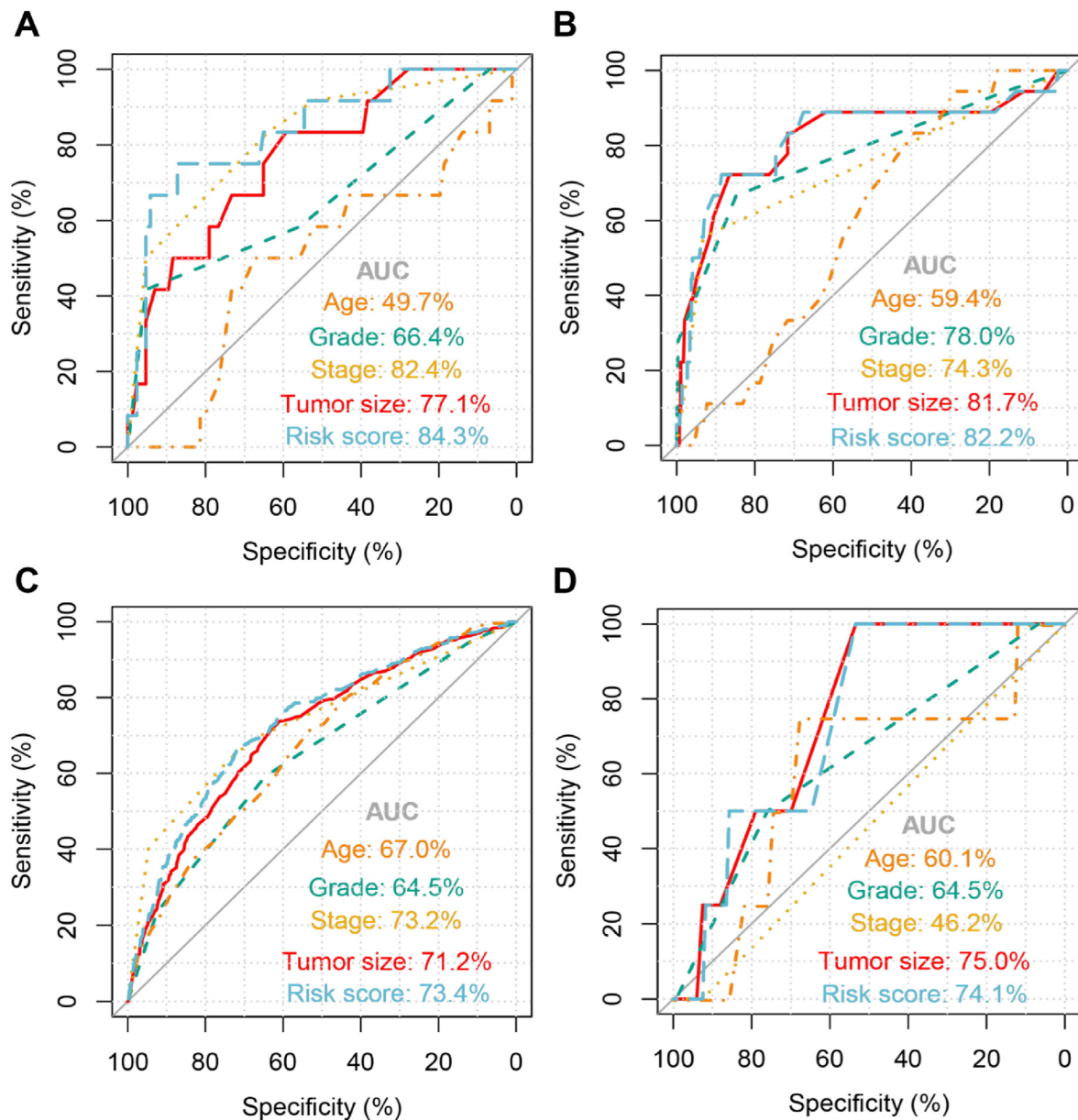
Figure 4. Evaluation of the machine learning-based prognosis prediction model through receiver operating characteristic curve analysis. (A) Comparing the area under curve value of 3-year disease-free survival prediction among the prognostic model and major clinicopathologic features in the CPTAC cohort. (B) Comparing the area under curve value of 3-year disease-free survival prediction among the prognostic model and major clinicopathologic features in the General cohort. (C) Comparing the area under curve value of 3-year overall survival prediction among the prognostic model and major clinicopathologic features in the SEER cohort. (D) Comparing the area under curve value of 3-year disease-free survival prediction among the prognostic model and major clinicopathologic features in the Huashan cohort. The P-value was acquired by comparing the AUCs between risk score and other indicators. AUC, area under curve; CPTAC, Clinical Proteomic Tumor Analysis Consortium; SEER, Surveillance, Epidemiology, and End Results.

CPTAC cohort. As shown in Figure 3A, compared with patients with lower risk scores, patients with higher risk scores have significantly worse clinical survival outcomes (HR = 5.15, 95% CI: 1.66–15.96, p = 0.018). External validation in General cohort (HR = 4.64, 95% CI: 2.15–10.02, p = 0.0007, Figure 3B), SEER cohort (HR = 3.96, 95% CI: 3.14–5.00, p < 0.0001, Figure 3C), and Huashan cohort (HR = 6.02, 95% CI: 1.50–24.15, p = 0.055, Figure 3D) also shows that patients with higher risk scores have a significantly worse prognosis. Further cox regression analysis indicated that our machine learning-based risk score could be used as an independent prognostic factor for patients with ccRCC patients in the SEER cohort and the General cohort (Table S2).

3.3. Comparison of the machine learning-based risk score and traditional clinicopathologic features

We further explore whether the prediction performance had been improved in the risk score compared with traditional clinicopathologic features by using ROC curve analysis in each patient cohort. The results indicated that the risk score achieved AUC values of 84.3%, 82.2%, 73.4%, and 74,1% in the CPTAC cohort (Figure 4A), the General cohort (Figure 4B), the SEER cohort (Figure 4C), and the Huashan cohort (Figure 4D), respectively. The risk score displayed slightly higher accuracy than some traditional clinicopathologic features (Table S3), even though without significant difference, which might be due to the limited sample size in each patient cohort.
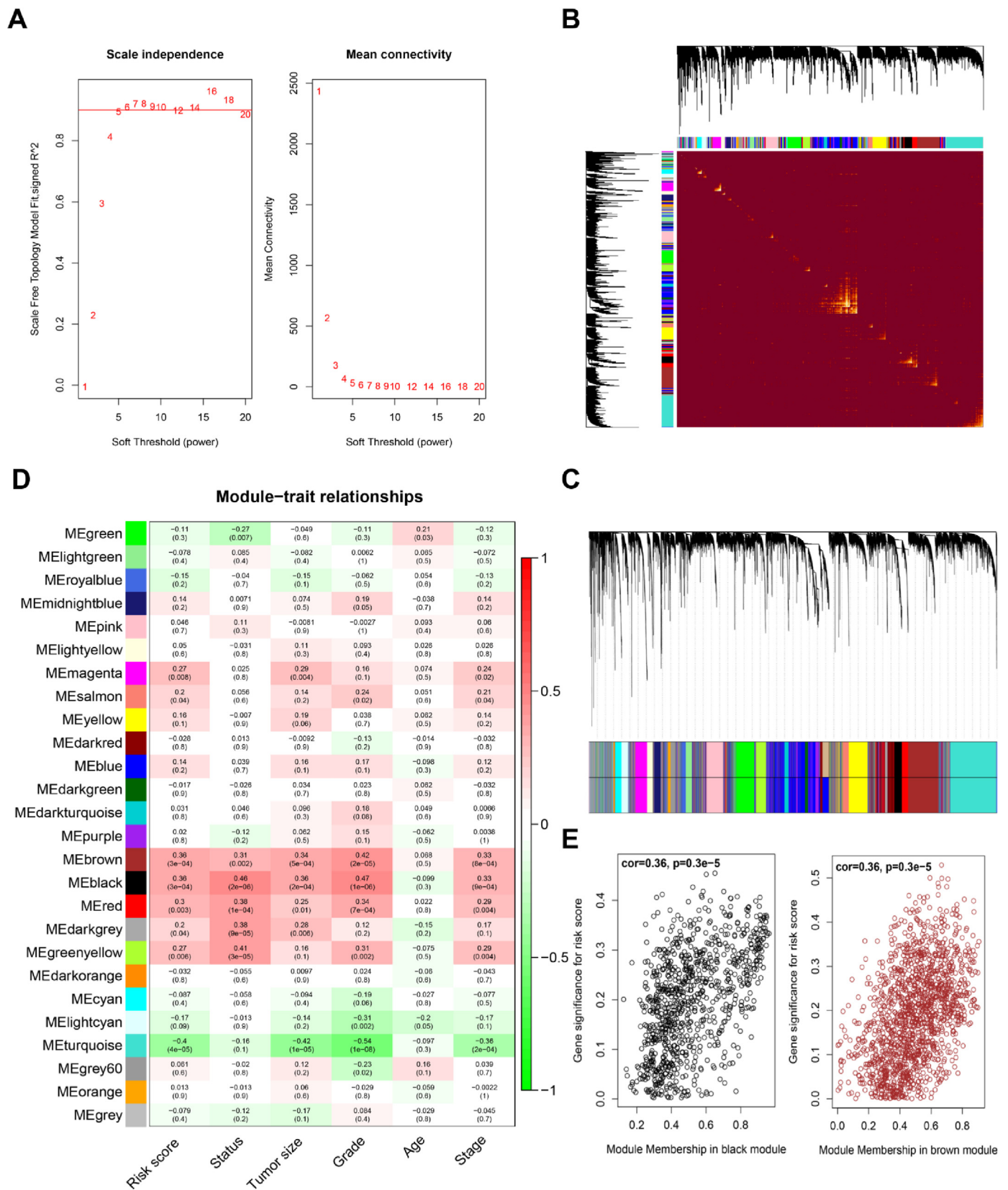
Figure 5. Weighted gene co-expression network analysis in the CPTAC cohort. (A) The estimation of soft threshold power for weighted gene co-expression network analysis. (B) Topological overlap matrix showing the gene network using a heatmap plot. (C) The merged dendrogram with different colors revealing the different modules identified by network analysis. (D) The relationship between gene modules and clinical characteristics. (E) Correction analysis of the selected gene modules and the risk score.
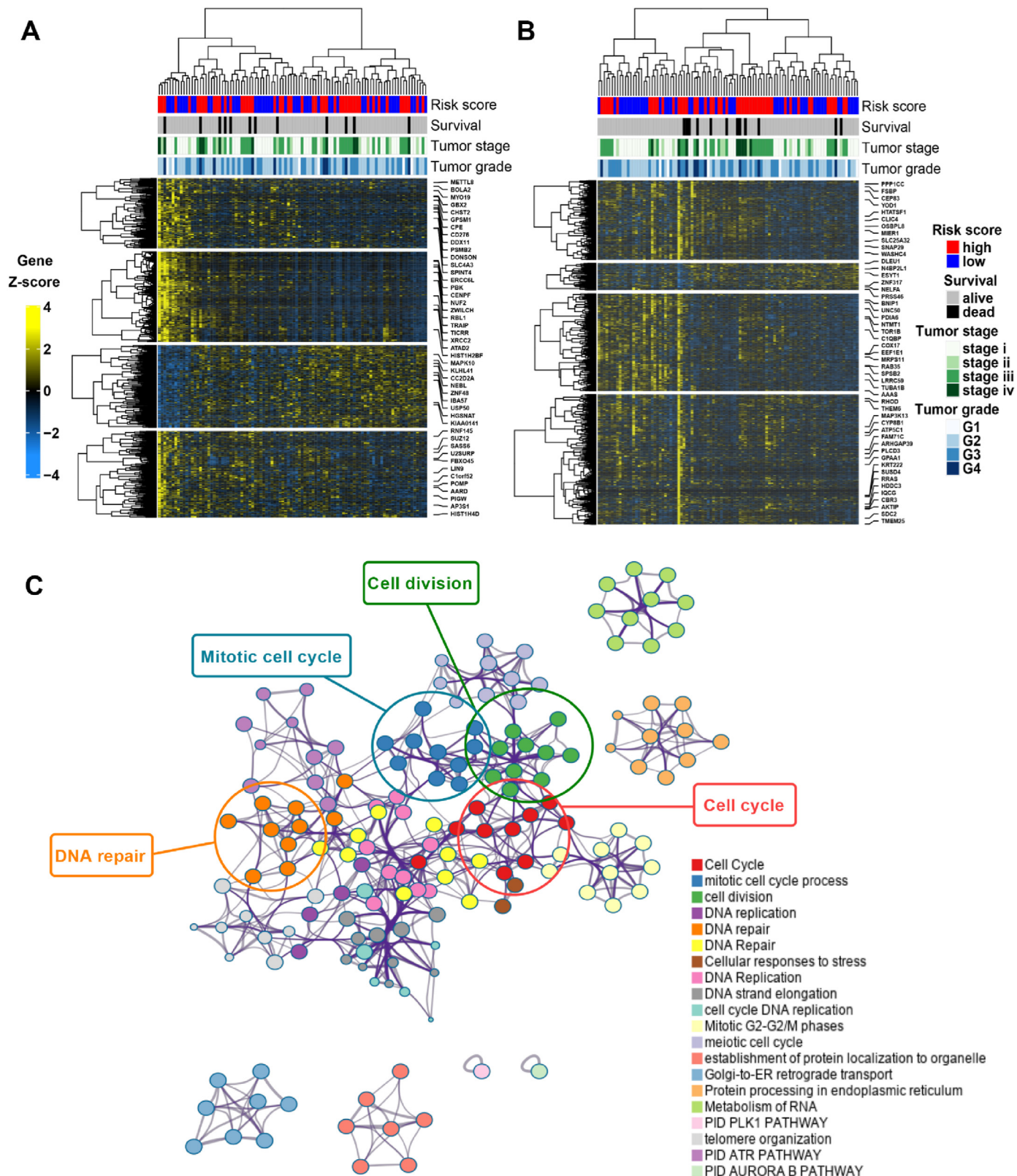
Figure 6. Potential mechanism analysis from co-expressed genes associated with the risk score. (A) Visualization of the expressions of the co-expressed genes in the black module. (B) Visualization of the expressions of the co-expressed genes in the brown module. (C) Potentially enriched pathways of the co-expressed genes associated with the risk score in the black and brown modules.

3.4. Association of the risk score and transcriptomic pathway

We carried out WGCNA and identified 26 independent modules in ccRCC patients (Figure 5A-C). The relationship between gene modules

and clinicopathologic features was shown in Figure 5D. The black module and the brown module were found to perform higher correlations to the risk score (Figure 5E). Further enrichment analysis based on 2352 genes in the black module (Figure 6A) and brown module (Figure 6B)

indicated that several cell proliferation-related pathways were statistically enriched. The top four enriched pathways associated with the risk score included cell cycle, mitotic cell cycle, cell division, and DNA repair (Figure 6C).

4. Discussion

It has been known for many years that the grade of ccRCC is associated with the prognosis of patients. However, the use of tumor grade alone to predict prognosis has significant drawbacks. For example, many studies have found no significant difference in prognosis between two adjacent grades [26]. Therefore, we cannot solely rely on the grade of ccRCC to predict the prognosis of patients.

In recent years, machine learning has made great progress in the fields of medicine. Machine learning technology is used by some scholars for predicting the tumor grade of RCC with imaging data [27, 28]. However, the use of machine learning technology based on clinicopathological data to predict the prognosis of patients with ccRCC is rarely reported now.

In this study, we screened several clinicopathological characteristics, and finally developed the survival prediction model based on tumor stage, tumor grade, and tumor size via using machine learning algorithm. The machine learning-based prognostic model exhibited consummate performance in differentiating patients with high survival risk, which could also be used as an independent prognostic factor for patients with ccRCC. Functional enrichment analysis also indicated that our machine learning-based risk score was significantly associated with some biological processes, including cell cycle, cell division, and DNA repair, which have been shown to be related to the occurrence and development of ccRCC [29, 30].

The size of the tumor is an essential factor in assessing the patient's prognosis, which is also the most intuitive and easy-to-measure attribute of the tumor. The size of RCC can not only be used to stage the tumor, but it can also predict the prognosis of the patients. As the tumor grows, the prognosis of the patient becomes worse. Some scholars have proposed using 3.0 cm as a cutting point. Patients with RCC within 3.0 cm have a better prognosis, but when it is greater than 3.0 cm, the prognosis will worsen [31]. In addition, the size of the tumor is also related to the synchronous and asynchronous metastasis of RCC, but when the primary tumor is less than 3.0 cm, the risk of metastasis is negligible [32]. Studies have also found that tumor size is related to malignant potential. As the tumor size increases, the degree of malignancy increases [33].

Of all the clinicopathologic data, the stage of RCC is the strongest predictor of patients outcome [6]. The TNM staging system for RCC is the most commonly used and important clinical grading system for the prognosis of patients. It has been modified many times, and the latest version is AJCC in 2017 [34]. With the increasingly refined classification of RCC, its guiding role in the treatment of RCC patients is becoming increasingly sophisticated. The use of the tumor stage alone for prognostic analysis is also inadequate. The TNM staging of RCC only takes into account the size of the tumor and does not consider whether the tumor is necrotic or not. In this study, our risk score based on machine learning performed well in predicting the 3-year survival status of ccRCC patients, which could act as new prognostic features with cost neutrality.

Our model not only shows a good role in predicting prognosis but also has convenient and practical value. In clinical practice, clinicians can only evaluate the prognosis of patients according to the size, grade, and stage of the tumor by using our prognosis model, without sequencing and radiomics analysis of the tumor. This model also has a certain guiding significance for clinical decision-making and individualized treatment. As the survival curves showed, patients with higher risk scores had significantly worse clinical survival outcomes compared to patients with lower risk scores, which suggested that patients with higher risk scores may need the timely intervention of extra treatments except for surgery.

Although our prediction model is perfectly constructed, there are still some limitations in our present scenario. Firstly, despite a large number of patients included in this study, there was a large difference in the number of patients between the cohorts, which would inevitably cause deviation. In addition, our research is a retrospective study and will be affected by unknown confounding factors. To verify our model more accurately, further prospective research needs to be carried out.

5. Conclusions

Through retrospective analysis of multicenter clinical data, we developed and validated a prediction model based on machine learning algorithm, which may contribute to clinical decision-making for patients with ccRCC. Further functional enrichment analysis demonstrated that our machine learning-based risk score was significantly connected with some cell proliferation-related pathways, consisting of DNA repair, cell division, and cell cycle.

Declarations

Author contribution statement

Xiang Wang; Ning Zhang; Rong Na: Conceived and designed the experiments.

Siteng Chen; Tuanjie Guo; Encheng Zhang: Performed the experiments; Wrote the paper.

Tao Wang; Guangliang Jiang; Yishuo Wu: Analyzed and interpreted the data.

Data availability statement

Data included in article/supp. material/referenced in article.

Declaration of interest's statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2022.e10578.

References

[1] R.L. Siegel, K.D. Miller, H.E. Fuchs, A. Jemal, Cancer statistics, CA Cancer J. Clin. 72 (1) (2022) 7–33.

[2] G. Kovacs, M. Akhtar, B.J. Beckwith, P. Bugert, C.S. Cooper, B. Delahunt, J.N. Eble, S. Fleming, B. Ljungberg, L.J. Medeiros, H. Moch, V.E. Reuter, E. Ritz, G. Roos, D. Schmidt, J.R. Srigley, S. Störkel, E. van den Berg, B. Zbar, The Heidelberg classification of renal cell tumours, J. Pathol. 183 (2) (1997) 131–133.

[3] S.A. Fuhrman, L.C. Lasky, C. Limas, Prognostic significance of morphologic parameters in renal cell carcinoma, Am. J. Surg. Pathol. 6 (7) (1982) 655–663.

[4] H. Borgmann, M. Musquera, A. Haferkamp, A. Vilaseca, T. Klatte, S.F. Shariat, A. Scavuzzo, M.A. Jimenez Rios, I. Wolff, U. Capitanio, P. Dell'Oglio, L.M. Krabbe, E. Herrmann, T. Ecke, D. Vergho, N. Huck, N. Wagener, S. Pahernik, S. Zastrow, M. Wirth, C. Surcel, C. Mirvald, K. Prochazkova, G. Hutterer, R. Zigeuner, L. Cindolo, M. Hora, C.G. Stief, M. May, S.D. Brookman-May, Prognostic significance of Fuhrman grade and age for cancer-specific and overall survival in patients with papillary renal cell carcinoma: results of an international multi-institutional study on 2189 patients, World J. Urol. 35 (12) (2017) 1891–1897.

[5] J.J. Patard, E. Leray, N. Rioux-Leclercq, L. Cindolo, V. Ficarra, A. Zisman, A. De La Taille, J. Tostain, W. Artibani, C.C. Abbou, B. Lobel, F. Guille, D.K. Chopin, P.F. Mulders, C.G. Wood, D.A. Swanson, R.A. Figlin, A.S. Belldegrun, A.J. Pantuck, Prognostic value of histologic subtypes in renal cell carcinoma: a multicenter experience, J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 23 (12) (2005) 2763–2771.

[6] J.B. Thrasher, D.F. Paulson, Prognostic factors in renal cancer, Urol. Clin. 20 (2) (1993) 247–262.

[7] B. Ljungberg, K. Bensalah, S. Canfield, S. Dabestani, F. Hofmann, M. Hora, M.A. Kuczyk, T. Lam, L. Marconi, A.S. Merseburger, P. Mulders, T. Powles,

M. Staehler, A. Volpe, A. Bex, EAU guidelines on renal cell carcinoma: 2014 update, Eur. Urol. 67 (5) (2015) 913–924.

[8] J. Rothman, B. Egleston, Y.N. Wong, K. Iffrig, S. Lebovitch, R.G. Uzzo, Histopathological characteristics of localized renal cell carcinoma correlate with tumor size: a SEER analysis, J. Urol. 181 (1) (2009) 29–33, discussion 33-4.

[9] T.H. Ho, P. Kapur, J.E. Eckel-Passow, A. Christie, R.W. Joseph, D.J. Serie, J.C. Cheville, R.H. Thompson, F. Homayoun, V. Panwar, J. Brugarolas, A.S. Parker, Multicenter validation of enhancer of zeste homolog 2 expression as an independent prognostic marker in localized clear cell renal cell carcinoma, J. Clin. Oncol. 35 (32) (2017) 3706–3713.

[10] G. Wu, Q. Wang, Y. Xu, Q. Li, L. Cheng, A new survival model based on ferroptosis-related genes for prognostic prediction in clear cell renal cell carcinoma, Aging (Albany NY) 12 (14) (2020) 14933–14948.

[11] S.A. Brooks, A.R. Brannon, J.S. Parker, J.C. Fisher, O. Sen, M.W. Kattan, A.A. Hakimi, J.J. Hsieh, T.K. Choueiri, P. Tamboli, J.K. Maranchie, P. Hinds, C.R. Miller, M.E. Nielsen, W.K. Rathmell, ClearCode34: a prognostic risk predictor for localized clear cell renal cell carcinoma, Eur. Urol. 66 (1) (2014) 77–84.

[12] T.J. Brinker, A. Hekler, A.H. Enk, C. Berking, S. Haferkamp, A. Hauschild, M. Weichenthal, J. Klode, D. Schadendorf, T. Holland-Letz, C. von Kalle, S. Fröhling, B. Schilling, J.S. Utikal, Deep neural networks are superior to dermatologists in melanoma image classification, Eur. J. Cancer 119 (2019) 11–17.

[13] B. Ehteshami Bejnordi, M. Veta, P. Johannes van Diest, B. van Ginneken, N. Karssemeijer, G. Litjens, J. van der Laak, M. Hermsen, Q.F. Manson, M. Balkenhol, O. Geessink, N. Stathonikos, M.C. van Dijk, P. Bult, F. Beca, A.H. Beck, D. Wang, A. Khosla, R. Gargeya, H. Irshad, A. Zhong, Q. Dou, Q. Li, H. Chen, H.J. Lin, P.A. Heng, C. Haß, E. Bruni, Q. Wong, U. Halici, M. Öner, R. Cetin-Atalay, M. Berseth, V. Khvatkov, A. Vylegzhanin, O. Kraus, M. Shaban, N. Rajpoot, R. Awan, K. Sirinukunwattana, T. Qaiser, Y.W. Tsang, D. Tellez, J. Annuscheit, P. Hufnagl, M. Valkonen, K. Kartasalo, L. Latonen, P. Ruusuvuori, K. Liimatainen, S. Albarqouni, B. Mungal, A. George, S. Demirci, N. Navab, S. Watanabe, S. Seno, Y. Takenaka, H. Matsuda, H. Ahmady Phoulady, V. Kovalev, A. Kalinovsky, V. Liauchuk, G. Bueno, M.M. Fernandez-Carrobles, I. Serrano, O. Deniz, D. Racoceanu, R. Venâncio, Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer, JAMA 318 (22) (2017) 2199–2210.

[14] P. Kickingereder, F. Isensee, I. Tursunova, J. Petersen, U. Neuberger, D. Bonekamp, G. Brugnara, M. Schell, T. Kessler, M. Foltyn, I. Harting, F. Sahm, M. Prager, M. Nowosielski, A. Wick, M. Nolden, A. Radbruch, J. Debus, H.P. Schlemmer, S. Heiland, M. Platten, A. von Deimling, M.J. van den Bent, T. Gorlia, W. Wick, M. Bendszus, K.H. Maier-Hein, Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study, Lancet Oncol. 20 (5) (2019) 728–740.

[15] R. Suarez-Ibarrola, S. Hein, G. Reis, C. Gratzke, A. Miernik, Current and future applications of machine and deep learning in urology: a review of the literature on urolithiasis, renal cell carcinoma, and bladder and prostate cancer, World J. Urol. 38 (10) (2020) 2329–2347.

[16] M. Nazari, I. Shiri, H. Zaidi, Radiomics-based machine learning model to predict risk of death within 5-years in clear cell renal cell carcinoma patients, Comput. Biol. Med. 129 (2021), 104135.

[17] H. Coy, K. Hsieh, W. Wu, M.B. Nagarajan, J.R. Young, M.L. Douek, M.S. Brown, F. Scalzo, S.S. Raman, Deep learning and radiomics: the utility of Google TensorFlow™ Inception in classifying clear cell renal cell carcinoma and oncocytoma on multiphasic CT, Abdom Radiol (NY) 44 (6) (2019) 2009–2020.

[18] F. Lin, E.-M. Cui, Y. Lei, L.-P. Luo, CT-based machine learning model to predict the Fuhrman nuclear grade of clear cell renal cell carcinoma, Abdom Radiol (NY) 44 (7) (2019) 2528–2534.

[19] A. Baghdadi, N.A. Aldhaam, A.S. Elsayed, A.A. Hussein, L.A. Cavuoto, E. Kauffman, K.A. Guru, Automated differentiation of benign renal oncocytoma and chromophobe renal cell carcinoma on computed tomography using deep learning, BJU Int. 125 (4) (2020) 553–560.

[20] J.-B. Chen, H.-S. Yang, S.-H. Moi, L.-Y. Chuang, C.-H. Yang, Identification of mortality-risk-related missense variant for renal clear cell carcinoma using deep learning, Ther Adv Chronic Dis 12 (2021), 2040622321992624.

[21] D.J. Clark, S.M. Dhanasekaran, F. Petralia, J. Pan, X. Song, Y. Hu, F. da Veiga Leprevost, B. Reva, T.M. Lih, H.Y. Chang, W. Ma, C. Huang, C.J. Ricketts, L. Chen, A. Krek, Y. Li, D. Rykunov, Q.K. Li, L.S. Chen, U. Ozbek, S. Vasaikar, Y. Wu, S. Yoo, S. Chowdhury, M.A. Wyczalkowski, J. Ji, M. Schnaubelt, A. Kong, S. Sethuraman, D.M. Avtonomov, M. Ao, A. Colaprico, S. Cao, K.C. Cho, S. Kalayci, S. Ma, W. Liu, K. Ruggles, A. Calinawan, Z.H. Gümüş, D. Geiszler, E. Kawaler, G.C. Teo, B. Wen, Y. Zhang, S. Keegan, K. Li, F. Chen, N. Edwards, P.M. Pierorazio, X.S. Chen, C.P. Pavlovich, A.A. Hakimi, G. Brominski, J.J. Hsieh, A. Antczak, T. Omelchenko, J. Lubinski, M. Wiznerowicz, W.M. Linehan, C.R. Kinsinger, M. Thiagarajan, E.S. Boja, M. Mesri, T. Hiltke, A.I. Robles, H. Rodriguez, J. Qian, D. Fenyö, B. Zhang, L. Ding, E. Schadt, A.M. Chinnaiyan, Z. Zhang, G.S. Omenn, M. Cieslik, D.W. Chan, A.I. Nesvizhskii, P. Wang, H. Zhang, Integrated proteogenomic characterization of clear cell renal cell carcinoma, Cell 179 (4) (2019) 964–983, e31.

[22] B.F. Hankey, L.A. Ries, B.K. Edwards, The surveillance, epidemiology, and end results program: a national resource, Cancer Epidemiol. Biomarkers Prev. 8 (12) (1999) 1117–1121.

[23] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, J. Stat. Software 33 (1) (2010) 1–22.

[24] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinf. 9 (2008) 559.

[25] Y. Zhou, B. Zhou, L. Pache, M. Chang, A.H. Khodabakhshi, O. Tanaseichuk, C. Benner, S.K. Chanda, Metascape provides a biologist-oriented resource for the analysis of systems-level datasets, Nat. Commun. 10 (1) (2019) 1523.

[26] M.Y. Teo, K. Seier, I. Ostrovnaya, A.M. Regazzi, B.E. Kania, M.M. Moran, C.K. Cipolla, M.J. Bluth, J. Chaim, H. Al-Ahmadie, A. Snyder, M.I. Carlo, D.B. Solit, M.F. Berger, S. Funt, J.D. Wolchok, G. Iyer, D.F. Bajorin, M.K. Callahan, J.E. Rosenberg, Alterations in DNA damage response and repair genes as potential marker of clinical benefit from PD-1/PD-L1 blockade in advanced urothelial cancers, J. Clin. Oncol. 36 (17) (2018) 1685–1694.

[27] M. Nazari, I. Shiri, G. Hajianfar, N. Oveisi, H. Abdollahi, M.R. Deevband, M. Oveisi, H. Zaidi, Noninvasive Fuhrman grading of clear cell renal cell carcinoma using computed tomography radiomic features and machine learning, Radiol. Med. 125 (8) (2020) 754–762.

[28] A. Stanzione, C. Ricciardi, R. Cuocolo, V. Romeo, J. Petrone, M. Sarnataro, P.P. Mainenti, G. Improta, F. De Rosa, L. Insabato, A. Brunetti, S. Maurea, MRI radiomics for the prediction of fuhrman grade in clear cell renal cell carcinoma: a machine learning exploratory study, J. Digit. Imag. 33 (4) (2020) 879–887.

[29] S. Sakano, Y. Hinoda, N. Okayama, Y. Kawai, Y. Korenaga, S. Eguchi, K. Nagao, C. Ohmi, K. Naito, The association of DNA repair gene polymorphisms with the development and progression of renal cell carcinoma, Ann. Oncol. 18 (11) (2007) 1817–1827.

[30] C.R. Thoma, A. Toso, K.L. Gutbrodt, S.P. Reggi, I.J. Frew, P. Schraml, A. Hergovich, H. Moch, P. Meraldi, W. Krek, VHL loss causes spindle misorientation and chromosome instability, Nat. Cell Biol. 11 (8) (2009).

[31] B. Bhindi, C.M. Lohse, R.J. Mason, M.E. Westerman, J.C. Cheville, M.K. Tollefson, S.A. Boorjian, R.H. Thompson, B.C. Leibovich, Are we using the best tumor size cut-points for renal cell carcinoma staging? Urology 109 (2017) 121–126.

[32] R.H. Thompson, J.R. Hill, Y. Babayev, A. Cronin, M. Kaag, S. Kundu, M. Bernstein, J. Coleman, G. Dalbagni, K. Touijer, P. Russo, Metastatic renal cell carcinoma risk according to tumor size, J. Urol. 182 (1) (2009) 41–45.

[33] R.H. Thompson, J.M. Kurta, M. Kaag, S.K. Tickoo, S. Kundu, D. Katz, L. Nogueira, V.E. Reuter, P. Russo, Tumor size is associated with malignant potential in renal cell carcinoma cases, J. Urol. 181 (5) (2009) 2033–2036.

[34] S.R. Williamson, K. Taneja, L. Cheng, Renal cell carcinoma staging: pitfalls, challenges, and updates, Histopathology 74 (1) (2019) 18–30.