Immune Gene Data-based Molecular Subtyping of Rectal Adenocarcinoma Related to Prognosis

Xia shu sen
North Sichuan Medical University https://orcid.org/0000-0003-3521-6971
Hong-peng Tian
North Sichuan Medical University
Zuo-liang Liu
North Sichuan Medical University
Zai-hua Yan
North Sichuan Medical University
Xian-yan Wang
North Sichuan Medical University
Chang-yuan Meng
North Sichuan Medical University
Li-fa Li
North Sichuan Medical University
Xue-gui Tang
North Sichuan Medical University
GuangJun Zhang (✉ zhanggj1977@126.com)
https://orcid.org/0000-0002-6934-638X

Research article

Keywords: READ, TCGA, Immune, Prognosis

Posted Date: February 17th, 2020

DOI: https://doi.org/10.21203/rs.2.23763/v1

Abstract

Background: The morbidity and mortality of rectal adenocarcinoma (READ) is increasing, which is considered as an aggressive type of colorectal malignancy. A great deal of evidence has suggested the significant association between the progression of READ and the immunophenotype of tumor cells (i.e. the expression of intracellular immune-related genes).

Methods: Samples retrieved from the TCGA and ImmPort database were investigated to identify immune-related genes specifically impacting the prognosis of READ patients. Several typical ones were then selected to construct the prognostic prediction model of READ patients through Lasso algorithm. The training and test cohorts were incorporated into the model, respectively. We stratified READ patients to evaluate the accuracy, efficiency and stability of the model in the prediction and classification of patient prognosis according to the value of median RiskScore (Risk-H and Risk-L). GO and KEGG signaling pathway enrichment analysis were conducted among the nine selected immune-related genes.

Results: A total of 57 immune-related displaying marked correlated with patient prognosis were identified and nine most typical genes could be majorly enriched into several pathways with close correlation with READ and the corresponding immune response. the distribution of nine immune-related genes were examined in the samples from both Risk-H and –L groups. Finally, the connection of the RiskScore value with the clinical characteristics of sample and the related signaling pathways were investigated.

Conclusions: The prognostic prediction model of RiskScore constructed based on the expression profiling of the nine immune-associated genes exhibited high prediction accuracy and stability to identify the relevant immune features. This model could contribute to the guidance for clinicians in diagnosing and predicting the prognosis for various immunophenotypes. Meanwhile, it also can offer various therapeutic targets for precise treatment of READ in clinical practice according to the identified immune molecules specific to different subtypes.

Background

Rectal adenocarcinoma (READ) is defined as a type of rectal cancer located above the dentate line and between the sigmoid colon and the transitional part of the rectum, accounting for 75% ~ 85% of all types of colorectal cancer[1]. At present, the pathogenesis of READ remains unclear, which might be associated with dietary factors, rectal adenoma, hereditary precancerous lesions, and inflammatory diseases of the large intestine[2]. Moreover, READ is a highly complex and heterogeneous malignancy. The heterogeneity of READ results in numerous clinical subtypes, with diverse degrees of sensitivity to chemotherapy and targeted therapy, thus leading to various prognoses[3, 4]. Approximately 2/3 READ patients are diagnosed at advanced stage in clinical practice[5].

The current therapeutic strategy of rectal cancer commonly includes a combination of surgery, chemotherapy, radiotherapy or targeted therapy[6]. Under normal conditions, a specific therapy is established according to the general assessment of the patient status, as well as cancer location and stage. However, only limited success rate could be achieved from the conventional approaches for READ, such as chemotherapy and radiotherapy, which might be due to the toxicity and non-specificity of these traditional therapy[7]. Notably, certain large-scale clinical studies do no recommend systemic adjuvant therapy as a postoperative treatment in the majority of

READ patients at early stage, largely due to the fact that the therapy-triggered toxicity on human body is far greater than the survival benefit of patients[8]. In high-risk patients, however, the absence of systemic adjuvant therapy would lead to rapid recurrence of the disease, which may even progress into infiltration and distant metastasis in the para-carcinoma tissues. To this end, it is of crucial necessity to identify the survival risks in patients during the early diagnosis. Moreover, patients should be stratified into different subgroups, and additional adjuvant systemic therapy should be prescribed for high-risk patients[9].

READ could be currently categorized into three types according to the differentiation degree of cancer cells: highly differentiated READ, moderately differentiated READ, and poorly differentiated READ[10]. Of the three different pathological types, poorly differentiated READ the most aggressive. However, this classification method certainly cannot be applied in the prognosis of patients at early stage. Biomarkers allow for reliable estimation of disease prognosis and patient survival, which are, therefore, valuable in decision-making of the clinical treatment of READ.[11] Recently, accumulative studies have indicated the application of gene expression in the prediction and stratification of the survival prognosis for READ patients[12]. Unfortunately, this proposal has not been adopted as the routine clinical practice yet, due to the small sample size, excessive data fitting or inadequate evidence in most studies. On this account, the open and accessible large-scale databases containing data of gene expression, such as the TCGA[13] and ImmPort databases[14], allow for mining the potentially more reliable biomarkers to predict and classify the prognosis of READ patients.

Immune cell infiltration, postulated to manifest active tumor responses, has been detected in most human solid tumors, and lymphocytic infiltration of READ has been shown to confer a survival advantage[15]. In addition, immune escape has also been validated as a novel cancer marker. Recently, immunotherapy-targeted specific immune checkpoints (including PD-1/PD-L1) therapy has given rise to surprising therapeutic effects in READ patients[16, 17]. However, the prominent molecular events concerning the tumor cell-immunocyte interaction within the READ microenvironment should be further exploited and summarized, which can help to determine their potential roles in predicting the prognosis of READ patients.

In this study, we mainly constructed and validated a prognostic prediction model for READ on the basis of the immune-related genes according to the clinical characteristics of patients recruited from the TCGA and ImmPort databases. Our present outcomes might potentially help the clinicians to evaluate the therapeutic effect, to predict the prognosis, and to select proper therapy for READ.

Methods

Pre-processing of preliminary sample data and initial screening of immune-related genes in READ

The latest clinical data on follow-up were extracted by using TCGA GDC API. Altogether 171 RNA-Seq data samples were retrieved (shown in Table S1), 161 of them were tumor tissues. Moreover, the immune-related gene set was also derived from the ImmPort database, involving a total of 1811 genes, as presented in Table S2.

First of all, 156 of the retrieved RNA-seq data (Table S3) were subject to pro-processing. Consequently, 156 samples involving 876 genes were adopted into further modeling analysis. Due to the small sample size of TCGA-READ and the limited number of death samples (26), events of disease progression were combined with

death cases, which yielded to 46 after combination, aiming to improve the accuracy of the model. The data of processed samples were shown in Table 1.

Secondly, in consideration of the relatively small sample size of TCGA-READ, 80% samples of TCGA-READ datasets were assigned to the training sets, and all samples were considered as test sets, instead of establishing training sets and test sets with an average distribution of 0.5:0.5 among all samples. According to the above-described method, 156 samples were classified as the training set and the test set, respectively. Data in the final training set and test set were shown in Table S4 and Table S5, respectively. Meanwhile, the clinical information of samples from both training and test sets were shown in Table 1. As a result, there was no significant difference the training set and the test set, as indicated by P-value, revealing reasonable sample grouping.

Univariate survival analysis of the immune-related genes in the training set

Univariate Cox proportional hazards regression model was used to analyze all immune-related genes and the survival data by utilizing the survival coxph function of R package[18]. A p value <0.05 was considered as statistical significance.

Screening of immune-related genes specific to the prognosis of READ, and the establishment of the prognostic prediction model

At first, the R software package glmnet[19] was adopted for lasso regression analysis, which finally gave rise to the risk model based on specific immune-related genes. The formula were shown as follows:

RiskScore=FYN*0.0667786+FGF18*0.075671487+TPT1*8.70E-05+ERAP2*-0.006313662+NFKB1*-0.03711375+TAP1*-0.002024+RARG*0.032456265+ADIPOR2*-0.009543411+HSPA1A*0.001132712

Afterwards, related gene expression profiles were selected from both the training and test sets, which were subsequently substituted into the as-proposed model to compute the RiskScore value of each sample. Of note, the median RiskScore value was employed as the cutoff value to classify the samples from the high risk group (Risk-H) or low risk group (Risk-L), respectively. Finally, to comprehensively assess the efficiency, accuracy and stability of the model to in the prediction and classification of the prognosis of READ patients, ROC analysis, KM analysis and gene clustering analysis were performed.

Functional annotations and signaling pathway enrichment of the immune-related genes specific to prognosis

The gene families of the nine eventually-selected genes were annotated based on the human gene classification in the HGNC database[20]. Notably, KEGG and GO enrichment analyses were performed by using the clusterProfile[21] of the R software package for the nine identified immune-related, prognosis-specific genes in this study.

Correlation of the RiskScore value with the signaling pathways and the clinical characteristics of samples

To begin with, the score of KEGG functional enrichment analysis was analyzed using the ssGSEA function of the R software package GSVA[22]. Meanwhile, the correlation with the RiskScore value was also computed,

followed by clustering analysis in accordance with the enrichment scores of each pathway in all samples. Moreover, we analyzed the correlations of related factors (including T, N, M stages, age and gender) with the RiskScore value. Finally, the nomogram model and forest plot were performed based on the related clinical characteristics along with the RiskScore value, followed by assessment of the correlations between the RiskScore value as well as clinical characteristics and patient survival.

Results

Identification of specific immune-related genes according to survival and prognosis outcomes in READ patients

At first, relevant data were collected from the TCGA and ImmPort databases, which were subsequently subject to pre-processing. Afterwards, univariate Cox proportional hazards regression model was used to analyze all immune-related genes and survival data by using the R package survival coxph function, with the significance level set at $P<0.05$ (shown in Table S6). Finally, a total of 57 prognosis-significant immune-related genes were identified. The associations between the p-values of these 57 genes and the HRs as well as the expression quantities were shown in Fig.1.

Screening of prognosis-specific immune-related genes and construction of the prognostic prediction model for READ

Although 57 immune-related genes were identified, most of these genes were not good enough for clinical detection. Thus, we aimed to narrow the immune-related gene scope while maintain the high accuracy. To this end, these 57 genes were further compressed using lasso regression for reduction of the number of genes in the risk model. The Lasso algorithm, a type of shrinkage estimate, could be used to construct a penalty function to acquire a relatively refined model, so that it could compress some coefficients, which would be set as 0. Consequently, the advantages of subset shrinkage were retained. Moreover, as a biased estimate for the multicollinearity data processing, it could estimate parameters while be proper for realizing variable selection, and could solve the problem of multicollinearity in regression analysis to a better extent. In the present study, the glmnet of R software package was used for lasso regression analysis. Briefly, 57 genes were compressed into nine genes (shown in Figure S1 and Table S7), and the formula were summarized in Materials and methods Section.

Thereafter, all samples from the training set were substituted into the formula to compute the RiskScore values. Typically, the median RiskScore value was employed as the cutoff value to classify samples into high risk (Risk-H) or low risk (Risk-L) groups. Moreover, ROC analysis concerning prognostic classification for RiskScore was also conducted. The OS duration of all samples was approximately two years (Figure S2). As a result (Figure 2A), we evaluated the 1-3 year survival prediction efficiency of the model in this study, with the average AUC reaching 0.823. Additionally, the sample distribution of both Risk-H and Risk-L groups under various OS were displayed in Figure 2B, revealing no statistical significance in the sample size between the 0- and 1-year Risk-H and Risk-L group. Moreover, the sample size in Risk-H group after the 3rd year was markedly decreased compared with that in Risk-L group. Of note, such variations became more obvious along with the extension of OS (Figure 2C). The clustering outcomes of the samples from the training set were shown in Figure 2D. It was clear that the above-mentioned nine genes could be markedly clustered into the high and low expression groups,

respectively, while samples in the training set could also be assigned into two groups. Moreover, we also compared the RiskScore values of the two subclasses (Figure 2E).

To validate the reliability of the prognostic prediction model, the expression profiles of the above-mentioned nine genes were retrieved from the test set as well, which were be substituted into the validation model. Meanwhile, we computed the RiskScore values of all samples. Similarly, data extracted from the test set were adopted to evaluate the 1~3-year survival prediction efficiencies, as shown in Figure 3A. Additionally, sample distribution in both Risk-H and Risk-L group under diverse OS was shown in Figure 3B, which suggested not obvious difference in the sample size between the 0~1-year Risk-H and Risk-L groups. Moreover, the sample size in Risk-H group after the 2^{nd} year was obviously decreased in comparison with that in Risk-L group, which became more obvious with the extension of OS (Figure 3C). The clustering outcomes for samples in the test set, and the difference in the RiskScore value between two subgroups, were shown in Figure 3D and E, respectively.

The KM survival curves of the risk model constructed based on the nine genes to classify the Risk-H and Risk-L groups in both training and test sets were shown in Figure 4. To be specific, the KM survival curve of the training set ($p<0.0001$) and test set ($p<0.001$) was shown in Figure 5A and Figure 5B, respectivley.

In our present study, we found that the prognosis model established based on the expression profile data of the above nine immune genes showed great prediction accuracy and stability in identifying immune features.

Functional annotations of the immune-related genes and signaling pathway enrichment specific to patient prognosis

At first, the gene families of the nine obtained genes were annotated according to the human gene classification in HGNC database (Table 2). As a result, these nine immune-related genes were significantly enriched in six gene families ($p < 0.05$), and the detailed information of gene function annotation was displayed in Table S8. In addition, the expression levels of these nine genes of the samples were significantly different between the Risk-H and Risk-L group (Figure 5). Meanwhile, the clusterProfile of the R software package was used to enrichment analyses of the above-mentioned nine prognosis specific immune related genes. Results of GO enrichment analysis and KEGG pathway enrichment analysis were shown in Figure 6A and Figure 6B, respectively, and relevant data were shown in Table S9 and Table S10. As was shown, most genes were enriched into multiple immune-related, cancer-associated biological processes and signaling pathways.

Correlation of the RiskScore value with the signaling pathways as well as clinical characteristics of samples

To begin with, the R software package GSVA was utilized to analyze the KEGG functional enrichment scores with the facilitation of ssGSEA function. Moreover, the correlations with the RiskScore value was further determined based on the enrichment scores of all pathways among various samples. A total of 34 related KEGG pathways were obtained, as shown in Table S11. Afterwards, all the 34 pathways were chosen to conduct clustering analysis in accordance with their enrichment scores of all samples (Figure 7). Besides, the correlation between the enrichment score and the RiskScore value was investigated by selecting the top five pathways with the highest GSEA enrichment score (including the prion diseases, ABC transporters, primary immunodeficiency, spliceosome and pendocytosis) to analyze the distribution between Risk-H and Risk-L groups. As shown in Figure 8, the scores acquired from these pathways were significantly different between Risk-H and Risk-L group.

Afterwards, the correlations between various factors (such as T, N, M stages, age and gender) and the RiskScore value were analyzed as well, as shown in Figure 9. Clearly, there was no obvious association of other features with the RiskScore value ($p>0.05$), with the exception of N ($p=0.0035$), suggesting that the dependence of RiskScore model on regional lymph node invasion to certain extent.

Eventually, the RiskScore value was combined with the clinical characteristics for the establishment of the nomogram model. Nomogram, an approach to intuitively and effectively present the outcomes of a certain risk model, could be conveniently applied in outcome prediction. In the nomogram, the length of a straight line represented the effects of different variables and their values on the outcome. In this study, due to the inconsistency between TNM and Stage, nomograms for the combination of TNM+RiskScore or Stage+RiskScore were constructed, respectively, as shown in Figure 10. The RiskScore features were obviously associated with the highest impact on the prediction of survival rate, suggesting the good performance of the established risk model based on the nine genes on prognostic prediction.

Finally, the forest plot was established by utilizing both RiskScore and clinical characteristics. As shown in Figure 11, in the forest plot established by TNM+RiskScore and Stage+RiskScore, the HRs of RiskScore were around 3, with a p-value <0.05. The multivariate cox-regression analyses of diverse clinical characteristics and RiskScore were shown in Table S12.

Discussion

The current standard therapeutic options for READ include surgical resection alone for earlystage READ, and surgical resection followed by adjuvant radiochemotherapy for the advanced stage READ. However, the effect of surgical resection is frequently restricted due to the local invasion of adjacent tissues by cancer cells or distant metastasis[23, 24]. Meanwhile, radiochemotherapy is also limited by its toxicity on normal tissues in the body[25]. Conventional anti-cancer therapies exert great burden on the body in while obtain therapeutic benefits[26]. Thus, no consistent therapeutic benefit can be achieved among these patients through the clinical medication, which might be partially due to the potential toxic, side effects and tumor heterogeneity[27]. Nevertheless, the application of postoperative systemic adjuvant therapy is still controversial in clinical practice. As a result, it is crucial to retrieve the potential biomarkers for READ for prognosis and recurrence prediction, to implement and to benefit from the early adjuvant therapy for high-risk patients.

The relationship between the immune system and the pathogenesis and progression of malignant tumors has attracted great attention in recent years, which has shed novel light on READ treatment, thus promoting the continuous development of anti-cancer treatment[16]. Starting from the tumor origin, namely, the immune system of the human body, to control and kill tumor cells via the regulation of the immune system of the body and enhancement of the anti-tumor immunity in the tumor microenvironment, has ushered in a new way for anti-tumor therapy. Therefore, screening the novel and meaningful READ prognosis-specific immune-related genes is of great value in the prediction of patient prognosis and mining novel therapeutic targets[28, 29]. In addition, the classification of READ based on the prognosis-specific immune-related genes would definitely contribute to the accurate prediction of patient outcome as well as identification of READ patients with high or low risk of postoperative recurrence.

In the present research, a total of nine prognosis-specific, immune-related genes were identified by means of big data mining, statistics and sorting including the TCGA and ImmPort databases. Afterwards, we also constructed the prognostic prediction model according to the expression of these nine immune-related genes, followed by calculation of the RiskScore values of patients. Moreover, both prediction and verification of the model were performed. Typically, the presently-proposed prognosis model in this study, which was established based on the expression profiles of specific immune genes, rendered further classification of patients with clear clinical stage into various subgroups according to the predictive survival outcomes.

During the implementation of this project, samples extracted from the TCGA and ImmPort databases were investigated to identify immune-related genes specifically impacting the prognosis of READ patients. As a result, 57 immune-related genes with remarkable association with patient prognosis were selected, which were subsequently subject to shrinkage estimate. Among them, nine most typical ones showing evident association with patient prognosis were further chosen for the establishment of prognostic prediction model for READ patients through Lasso algorithm. Subsequently, samples in both training and test sets were incorporated into the constructed model, respectively. Meanwhile, to evaluate the efficiency, accuracy and stability of the model in the prediction and classification of patient prognosis, READ patients were stratified according to the value of median RiskScore (Risk-H and Risk-L) to assess the. Thereafter, functional annotations, GO and KEGG signaling pathway enrichment analyses were conducted among the nine selected immune-related genes. Our findings indicated that, all of the above-mentioned nine genes could be majorly enriched into several pathways with close association with READ and the corresponding immune response. In addition, the distributions of nine immune-related genes were examined in the samples in Risk-H and -L groups. Finally, the connection between the RiskScore value and the sample clinical characteristics as well as the related signaling pathways was investigated.

Among the nine genes, seven (NFKB1, FYN, ADIPOR2, TAP1, HSPA1A, ERAP2 and FGF18) out of them have been previously reported to be involved in the pathogenesis, progression, malignant transformation, and pathological process of immune microenvironment of READ, which are significantly correlated with patient survival and prognosis[30–35]. These previous findings have been verified the high reliability and accuracy of the bioinformatic mining results in this study. However, the correlations of the remaining two genes (RARG and TPT1) with READ have not been validated in neither basic nor clinical studies, which we are the most interested in. RARG has been confirmed to participate in the regulation of the proliferation and invasion of multiple malignant cancer cells[36], affecting production and release of cytokines and growth factors[37], modulating the innate immune response to viruses and pathogens[38]. Meanwhile, TPT1 has been validated to be involved in lymphocyte proliferation and activation, which can regulate the sensitivity of tumor cells to immunotherapy[39].

At present, the correlations of the expression of numerous gene with the survival of READ patients are increasingly mined, however, most of the existing studies are only verified on animal model, in vitro cell model or small-scale patient samples. Nonetheless, more comprehensive, large-scale population analysis is warranted due to the complexity of READ microenvironment. Fortunately, the rapid development of genome-wide sequencing renders the free development of high-throughput tumor databases, including TCGA, which have made it possible for the bioinformatic big data analysis of the large-scale READ population.

Conclusion

To sum up, our present findings could contribute to the identification of the novel markers for READ in clinic. Additionally, our proposed risk model based on the nine prognosis-specific immune-related genes would thereby provide multiple targets for the precise treatment of READ and facilitate the classification of READ patients at molecular subtype level. Furthermore, this established model is a promising tool to offer guidance for clinicians in prognostic prediction, clinical diagnosis and medication for READ patients with distinct immunophenotypes to some extent.

Abbreviations

READ: rectal adenocarcinoma; TCGA: the cancer genome atlas; ROC: Receiver Operating Characteristic Curves; HGNC: hugo gene nomenclature committee; KEGG: Kyoto Encyclopedia of Genes and Genomes.

Declarations

Acknowledgments

Not Applicable.

Authors' contributions

XXS, THP, WXY, LLF, YZH and LZL performed experiments; XSS, ZGJ,and TXG designed research and wrote the paper; XXS and MCY analyzed data. All authors read and approved the final Manuscript.

Fundings

This study was supported by Sichuan Youth Science and Technology Foundation (2017JQ0039), Scientific Research Project of Nanchong Municipal Science and Technology Bureau (16YFZJ0128, 18SXHZ0575, 18SXHZ0527), Scientific Research Project of Sichuan Provincial Health and Family Planning Commission (19ZD005), and PhD research startup foundation of North Sichuan Medical College (CBY17-QD07).

Availability of data and materials

All data generated or analysed during this study and materials used are

available in the manuscript and supplementary information file.

Ethics approval and consent to participate

Not Applicable.

Consent for publication

Not Applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]The Second Department of General Surgery, The Affiliated Hospital of North Sichuan Medical College, Nanchong, Sichuan, People's Republic of China.

[2]Institute of Hepatobiliary, Pancreatic and Intestinal Disease, North Sichuan Medical College, Nanchong, Sichuan, People's Republic of China.

[3] Anorectal Department of Integrated Traditional Chinese and Western Medicine, North Sichuan Medical College, Nanchong, Sichuan, People's Republic of China.

References

[1] J.A.W. Hagemans, J. Rothbarth, G.H.W. van Bogerijen, E. van Meerten, J. Nuyttens, C. Verhoef, J.W.A. Burger, Treatment of Inguinal Lymph Node Metastases in Patients with Rectal Adenocarcinoma, Annals of surgical oncology, 26 (2019) 1134-1141.

[2] R.S. Shinde, N. Katdare, N.A.N. Kumar, R. Bhamre, A. Desouza, V. Ostwal, R. Engineer, A. Saklani, Impact of histological subtype on treatment outcomes in locally advanced rectal adenocarcinoma treated with neoadjuvant chemoradiation, Acta oncologica, 57 (2018) 1721-1723.

[3] F. Letaief, M. Nasri, M. Ayadi, K. Meddeb, A. Mokrani, Y. Yahyaoui, N. Chraiet, H. Raies, A. Mezlini, Potential predictive factors for pathologic complete response after the neoadjuvant treatment of rectal adenocarcinoma: a single center experience, Cancer biology & medicine, 14 (2017) 327-334.

[4] L.S. Lino-Silva, R.A. Salcedo-Hernandez, E.B. Ruiz-Garcia, A.M. Leon-Takahashi, L. Garcia-Perez, Outcome of young patients with rectal adenocarcinoma, Journal of gastrointestinal oncology, 8 (2017) 96-101.

[5] A. Merchea, S.M. Ali, S.R. Kelley, E. Duchalais, J.Y. Alabbad, E.J. Dozois, D.W. Larson, Long-Term Oncologic Outcomes of Minimally Invasive Proctectomy for Rectal Adenocarcinoma, Journal of gastrointestinal surgery : official journal of the Society for Surgery of the Alimentary Tract, 22 (2018) 1412-1417.

[6] P. Renz, R.E. Wegner, S. Hasan, R. Brookover, G. Finley, D. Monga, M. Raj, J. McCormick, A. Kirichenko, Survival Outcomes After Surgical Management of the Primary Tumor With and Without Radiotherapy for Metastatic Rectal Adenocarcinoma: A National Cancer Database (NCDB) Analysis, Clinical colorectal cancer, (2019).

[7] S. Ahmed, C. Eng, Role of Chemotherapy in the Neoadjuvant/Adjuvant Setting for Patients With Rectal Adenocarcinoma Undergoing Chemoradiotherapy and Surgery or Radiotherapy and Surgery, Current oncology reports, 20 (2018) 3.

[8] R.Y. Tay, M. Jamnagerwalla, M. Steel, H.L. Wong, J.J. McKendrick, I. Faragher, S. Kosmider, I. Hastie, J. Desai, M. Tacey, P. Gibbs, R. Wong, Survival Impact of Adjuvant Chemotherapy for Resected Locally Advanced Rectal Adenocarcinoma, Clinical colorectal cancer, 16 (2017) e45-e54.

[9] M.S. Chiu, V. Verma, N.R. Bennion, A.R. Bhirud, J. Li, M.E. Charlton, C. Are, C. Lin, Comparison of outcomes between rectal squamous cell carcinoma and adenocarcinoma, Cancer medicine, 5 (2016) 3394-3402.

[10] W.M. Mendenhall, W.R. Rout, R.A. Zlotecki, S.E. Mitchell, R.D. Marsh, E.M. Copeland, 3rd, Conservative treatment of rectal adenocarcinoma, Hematology/oncology clinics of North America, 15 (2001) 303-319.

[11] Y. Hua, X. Ma, X. Liu, X. Yuan, H. Qin, X. Zhang, Identification of the potential biomarkers for the metastasis of rectal adenocarcinoma, APMIS : acta pathologica, microbiologica, et immunologica Scandinavica, 125 (2017) 93-100.

[12] P. Kapur, Predictive biomarkers for response to therapy in advanced colorectal/rectal adenocarcinoma, Critical reviews in oncogenesis, 17 (2012) 361-372.

[13] J. McCain, The cancer genome atlas: new weapon in old war?, Biotechnology healthcare, 3 (2006) 46-51B.

[14] S. Bhattacharya, P. Dunn, C.G. Thomas, B. Smith, H. Schaefer, J. Chen, Z. Hu, K.A. Zalocusky, R.D. Shankar, S.S. Shen-Orr, E. Thomson, J. Wiser, A.J. Butte, ImmPort, toward repurposing of open access immunological assay data for translational and clinical research, Scientific data, 5 (2018) 180015.

[15] D. Caputo, M. Caricato, A. Coppola, V. La Vaccara, M. Fiore, R. Coppola, Neutrophil to Lymphocyte Ratio (NLR) and Derived Neutrophil to Lymphocyte Ratio (d-NLR) Predict Non-Responders and Postoperative Complications in Patients Undergoing Radical Surgery After Neo-Adjuvant Radio-Chemotherapy for Rectal Adenocarcinoma, Cancer investigation, 34 (2016) 440-451.

[16] S. Sadahiro, T. Suzuki, Y. Maeda, A. Tanaka, A. Kamijo, C. Murayama, Y. Nakayama, T. Akiba, Effects of preoperative immunochemoradiotherapy and chemoradiotherapy on immune responses in patients with rectal adenocarcinoma, Anticancer research, 30 (2010) 993-999.

[17] M. Hecht, M. Buttner-Herold, K. Erlenbach-Wunsch, M. Haderlein, R. Croner, R. Grutzmann, A. Hartmann, R. Fietkau, L.V. Distel, PD-L1 is upregulated by radiochemotherapy in rectal adenocarcinoma patients and associated with a favourable prognosis, European journal of cancer, 65 (2016) 52-60.

[18] R. Liang, M. Wang, G. Zheng, H. Zhu, Y. Zhi, Z. Sun, A comprehensive analysis of prognosis prediction models based on pathwaylevel, genelevel and clinical information for glioblastoma, International journal of molecular medicine, 42 (2018) 1837-1846.

[19] J.Y. Hou, Y.G. Wang, S.J. Ma, B.Y. Yang, Q.P. Li, Identification of a prognostic 5-Gene expression signature for gastric cancer, Journal of cancer research and clinical oncology, 143 (2017) 619-629.

[20] B. Braschi, P. Denny, K. Gray, T. Jones, R. Seal, S. Tweedie, B. Yates, E. Bruford, Genenames.org: the HGNC and VGNC resources in 2019, Nucleic acids research, 47 (2019) D786-D792.

[21] Y. Xu, J. Chen, Z. Yang, L. Xu, Identification of RNA Expression Profiles in Thyroid Cancer to Construct a Competing Endogenous RNA (ceRNA) Network of mRNAs, Long Noncoding RNAs (lncRNAs), and microRNAs (miRNAs), Medical science monitor : international medical journal of experimental and clinical research, 25 (2019) 1140-1154.

[22] S. Hanzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq data, BMC bioinformatics, 14 (2013) 7.

[23] M.A. Amante, I.O. Real, G. Bermudez, Thyroid metastasis from rectal adenocarcinoma, BMJ case reports, 2018 (2018).

[24] T. Nishikawa, S. Ishihara, T. Ushiku, K. Hata, K. Sasaki, K. Murono, K. Yasuda, K. Otani, T. Tanaka, T. Kiyomatsu, K. Kawai, H. Nozawa, T. Watanabe, Anal metastasis from rectal adenocarcinoma, Clinical journal of gastroenterology, 9 (2016) 379-383.

[25] C. Weissenberger, G. Von Plehn, F. Otto, A. Barke, F. Momm, M. Geissler, Adjuvant radiochemotherapy of stage II and III rectal adenocarcinoma: role of CEA and CA 19-9, Anticancer research, 25 (2005) 1787-1793.

[26] M. Demes, S. Scheil-Bertram, H. Bartsch, A. Fisseler-Eckhoff, Signature of microsatellite instability, KRAS and BRAF gene mutations in German patients with locally advanced rectal adenocarcinoma before and after neoadjuvant 5-FU radiochemotherapy, Journal of gastrointestinal oncology, 4 (2013) 182-192.

[27] W. Daskalaki, E. Wardelmann, M. Port, K. Stock, J. Steinestel, S. Huss, J. Sperveslage, K. Steinestel, S. Eder, Expression levels of hnRNP K and p21WAF1/CIP1 are associated with resistance to radiochemotherapy independent of p53 pathway activation in rectal adenocarcinoma, International journal of molecular medicine, 42 (2018) 3269-3277.

[28] B. Zhang, B. Cheng, F.S. Li, J.H. Ding, Y.Y. Feng, G.Z. Zhuo, H.F. Wei, K. Zhao, High expression of CD39/ENTPD1 in malignant epithelial cells of human rectal adenocarcinoma, Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine, 36 (2015) 9411-9419.

[29] B. Zhang, B. Song, X. Wang, X.S. Chang, T. Pang, X. Zhang, K. Yin, G.E. Fang, The expression and clinical significance of CD73 molecule in human rectal adenocarcinoma, Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine, 36 (2015) 5459-5466.

[30] T.I. Kopp, V. Andersen, A. Tjonneland, U. Vogel, Polymorphisms in NFKB1 and TLR4 and interaction with dietary and life style factors in relation to colorectal cancer in a Danish prospective case-cohort study, PloS one, 10 (2015) e0116394.

[31] Q. Wang, J. Qian, F. Wang, Z. Ma, Cellular prion protein accelerates colorectal cancer metastasis via the Fyn-SP1-SATB1 axis, Oncology reports, 28 (2012) 2029-2034.

[32] L. Zhou, H.F. Zhang, W. Ning, X. Song, X. Liu, J.X. Liu, Associations of adiponectin receptor 2 (AdipoR2) gene polymorphisms and AdipoR2 protein expression levels with the risk of colorectal cancer: A case-control study, Molecular medicine reports, 16 (2017) 3983-3993.

[33] A. Ling, A. Lofgren-Burstrom, P. Larsson, X. Li, M.L. Wikberg, A. Oberg, R. Stenling, S. Edin, R. Palmqvist, TAP1 down-regulation elicits immune escape and poor prognosis in colorectal cancer, Oncoimmunology, 6 (2017) e1356143.

[34] S.B. Nadin, F.D. Cuello-Carrion, M.L. Sottile, D.R. Ciocca, L.M. Vargas-Roig, Effects of hyperthermia on Hsp27 (HSPB1), Hsp72 (HSPA1A) and DNA repair proteins hMLH1 and hMSH2 in human colorectal cancer hMLH1-deficient and hMLH1-proficient cell lines, International journal of hyperthermia : the official journal of European Society for Hyperthermic Oncology, North American Hyperthermia Group, 28 (2012) 191-201.

[35] Z.N. Erdem, S. Schwarz, D. Drev, C. Heinzle, A. Reti, P. Heffeter, X. Hudec, K. Holzmann, B. Grasl-Kraupp, W. Berger, M. Grusch, B. Marian, Irinotecan Upregulates Fibroblast Growth Factor Receptor 3 Expression in Colorectal Cancer Cells, Which Mitigates Irinotecan-Induced Apoptosis, Translational oncology, 10 (2017) 332-339.

[36] M.D. Long, P.K. Singh, J.R. Russell, G. Llimos, S. Rosario, A. Rizvi, P.R. van den Berg, J. Kirk, L.E. Sucheston-Campbell, D.J. Smiraglia, M.J. Campbell, The miR-96 and RARgamma signaling axis governs androgen signaling and prostate cancer progression, Oncogene, 38 (2019) 421-444.

[37] M.T. Mizwicki, M. Fiala, L. Magpantay, N. Aziz, J. Sayre, G. Liu, A. Siani, D. Chan, O. Martinez-Maza, M. Chattopadhyay, A. La Cava, Tocilizumab attenuates inflammation in ALS patients through inhibition of IL6 receptor signaling, American journal of neurodegenerative disease, 1 (2012) 305-315.

[38] I.G. Ovsyannikova, N. Dhiman, I.H. Haralambieva, R.A. Vierkant, M.M. O'Byrne, R.M. Jacobson, G.A. Poland, Rubella vaccine-induced cellular immunity: evidence of associations with polymorphisms in the Toll-like, vitamin A and D receptors, and innate immune response genes, Human genetics, 127 (2010) 207-221.

[39] A. Crisa, F. Ferre, G. Chillemi, B. Moioli, RNA-Sequencing for profiling goat milk transcriptome in colostrum and mature milk, BMC veterinary research, 12 (2016) 264.

Tables

Table1　Training set and test set sample information

Clinical Features	TrainingSet	TestingSet	*P-value*
OS	124	156	0.7721719
Event	124	156	0.2126965
Alive	106	130	
Dead	18	26	
T	123	155	0.42926
T1	4	9	
T2	21	28	
T3	87	105	
T4	11	13	
TX	1	1	
N	122	153	0.3620972
N0	64	79	
N1	32	43	
N2	26	31	
NX	2	3	
M	112	141	0.3711134
M0	93	118	
M1	19	23	
MX	12	15	
Stage	117	147	0.4516387
I	21	30	
II	41	46	
III	35	47	
IV	20	24	
X	7	9	
NewEvent	124	156	0.3799357
0	98	126	
1	26	30	
Age	124	156	0.4646033
0~50	15	19	
50~60	24	29	
60~70	36	52	
70~100	49	56	
Gender	124	156	0.2081067
FEMALE	55	69	
MALE	69	87	
X	7	9	
NewEvent	124	156	0.3799357
0	98	126	
1	26	30	
Age	124	156	0.4646033
0~50	15	19	
50~60	24	29	
60~70	36	52	
70~100	49	56	
Gender	124	156	0.2081067
FEMALE	55	69	
MALE	69	87	

Table 2 9 gene function annotation results

GeneFamily	Genes	pvalue	padj
NF-kappa B complex subunits	NFKB1	0.002328488	0.020956396
Src family tyrosine kinases	FYN	0.003878135	0.034903219
Progestin and adipoQ receptor family	ADIPOR2	0.004652156	0.041869407
ATP binding cassette subfamily B	TAP1	0.004652156	0.041869407
TIM23 complex	HSPA1A	0.005038966	0.045350696
M1 metallopeptidases	ERAP2	0.005425642	0.048830782
Nuclear hormone receptors	RARG	0.019257335	0.173316011
unknown	FGF18/TPT1	1	1

Figures

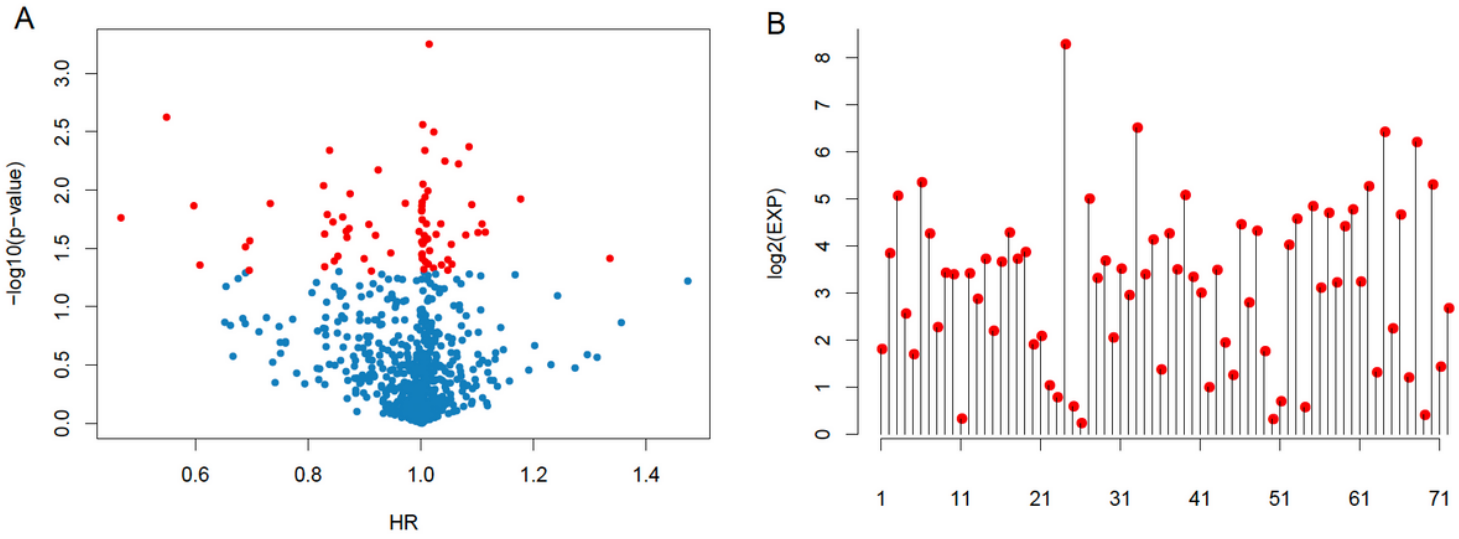

Figure 2

The relationships of the p-values of 57 genes and the HR, expression quantities. (A) The relationships of the p-values of 57 genes and the HR is displayed. (B) The relationships of the p-values of 57 genes and the expression levels. Red dots represents significantly different immune-related genes (p□0.05) regarding prognosis.

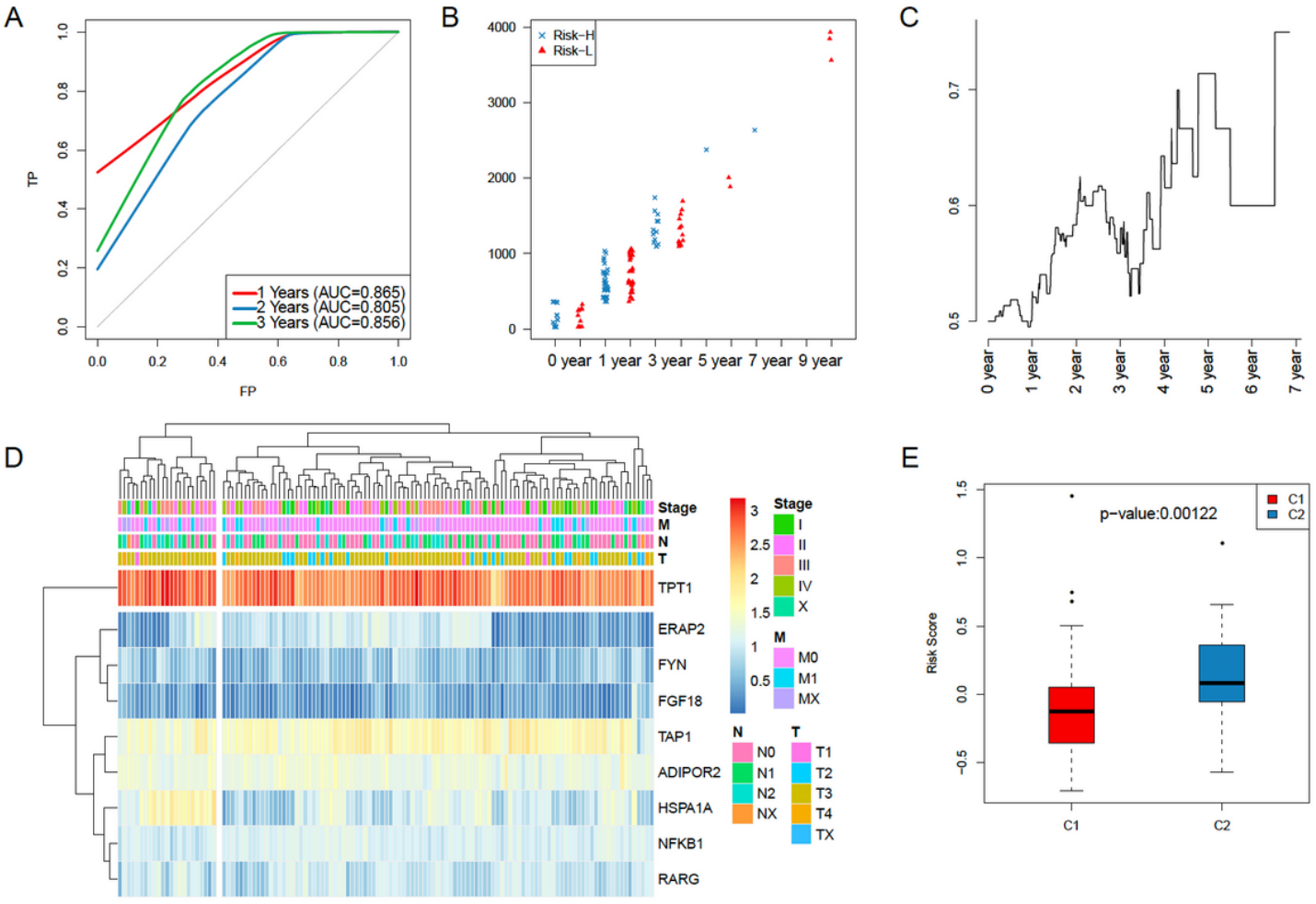

Figure 4

Verify the stability of the prognosis prediction model included 9 immune-related genes for READ patients in training set. (A) The survival predicted ROC curves of 9-gene risk model in training set. (B) The distribution of samples in Risk-H and Risk-L groups of training set divided through 9-gene risk model under different OS. (C) The level of Risk-L group/Total sample size with the extension in OS in the traning set. (D) The clustering results of training set samples. (E) Difference in the RiskScore between the two subgroups which had been clustered by the expression of 9 genes of training set samples.

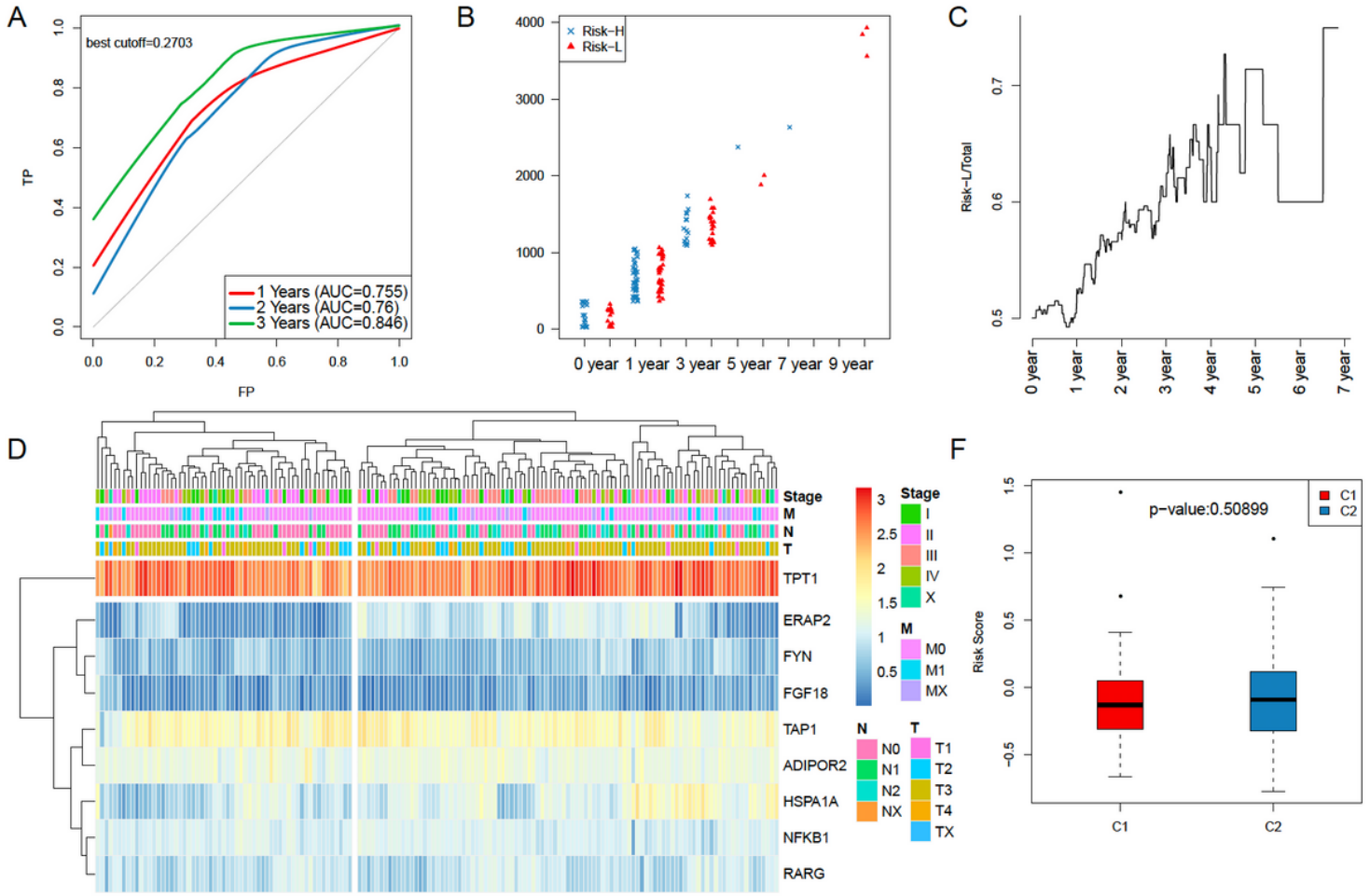

Figure 6

Verify the reliability of the prognosis prediction model included 9 immune-related genes for READ patients in test set. (A) The survival predicted ROC curves of 9-gene risk model in test set. (B) The distribution of samples in Risk-H and Risk-L groups of test set divided through 9-gene risk model under different OS. (C) The level of Risk-L group/Total sample size with the extension in OS in the test set. (D) The clustering results of test set samples. (E) Difference in the RiskScore between the two subgroups which had been clustered by the expression of 9 genes of test set samples.

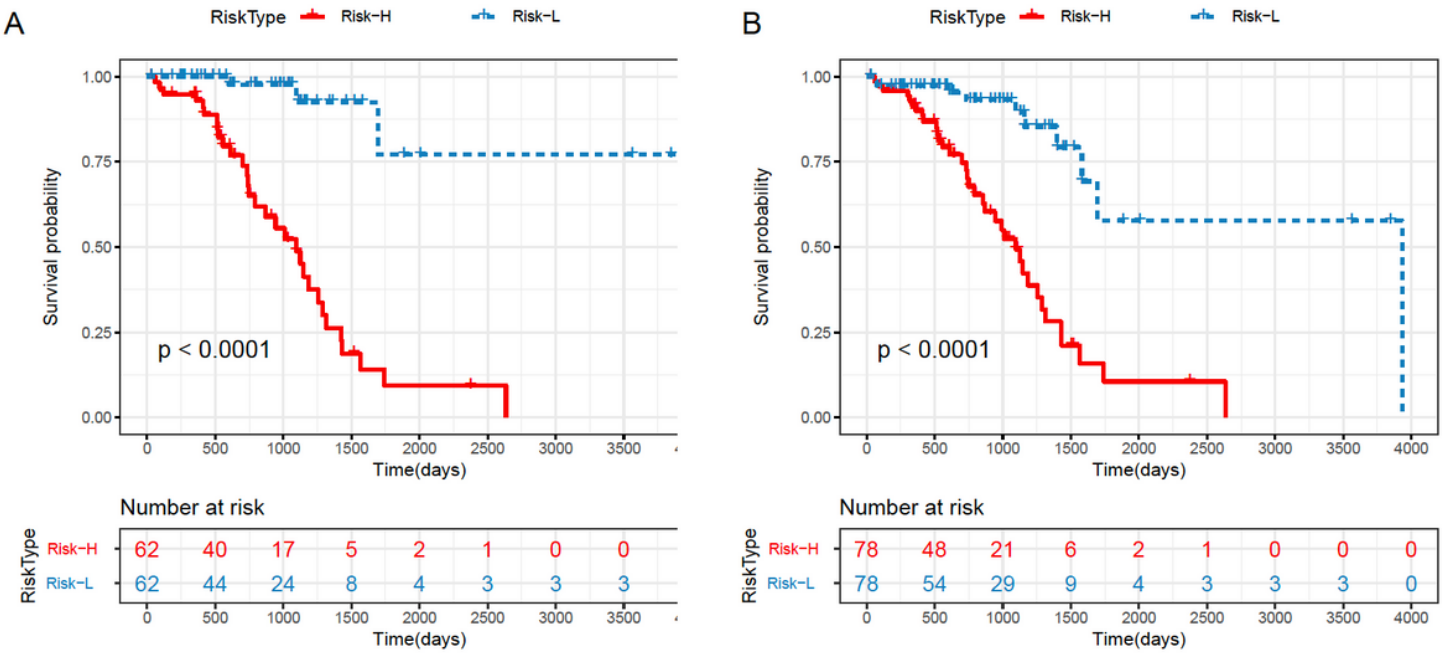

Figure 8

The KM survival curve of the 9-gene-based risk model in predicting the Risk-H and Risk-L groups on the training set (A) and test set (B).

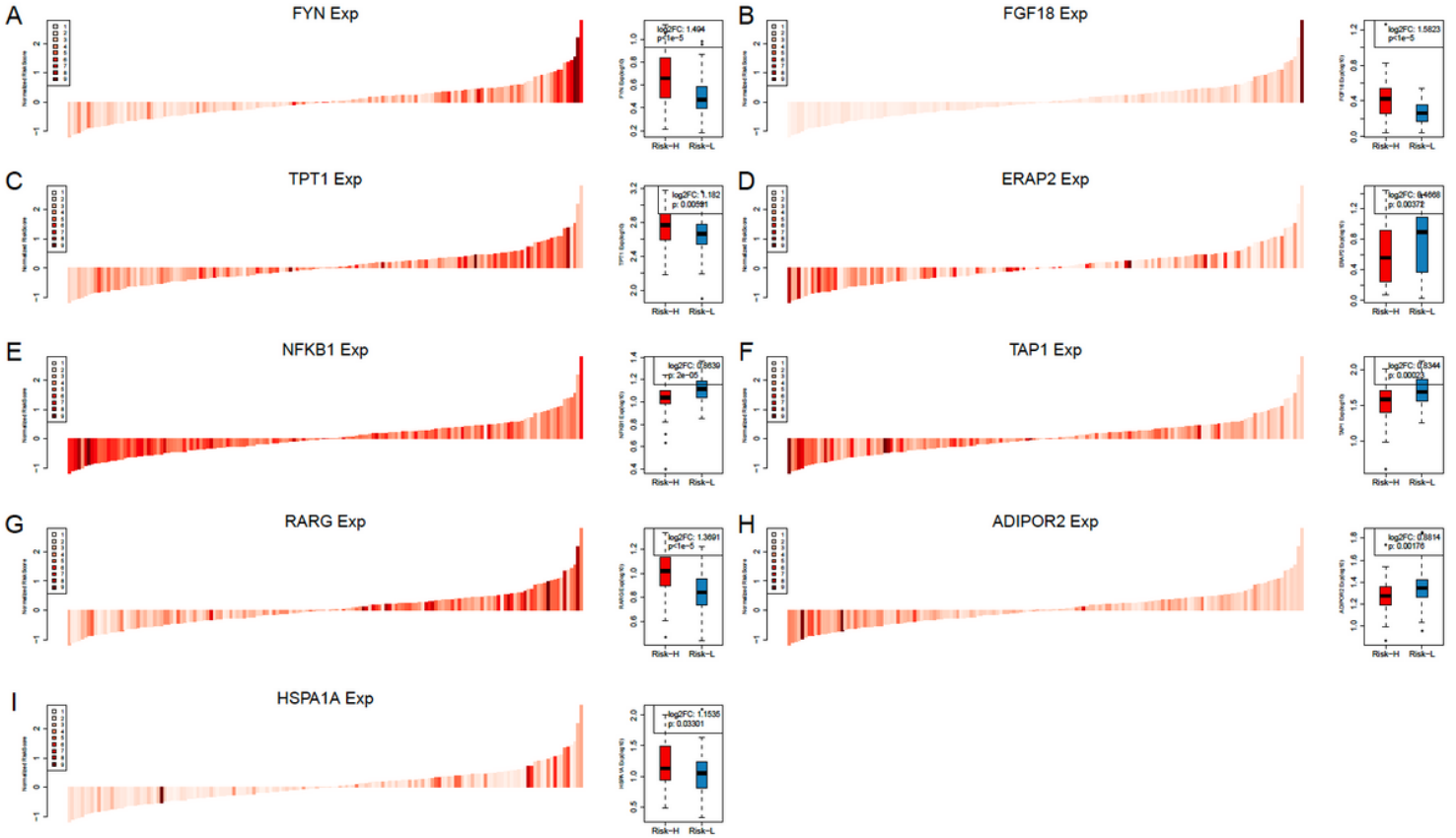

Figure 10

The expression differences of the FYN (A), FGF18 (B), TPT1 (C), ERAP2 (D), NFKB1 (E), TAP1 (F), RARG (G), ADIPOR2 (H) and HSPA1A (I) between the Risk-H and Risk-L groups.

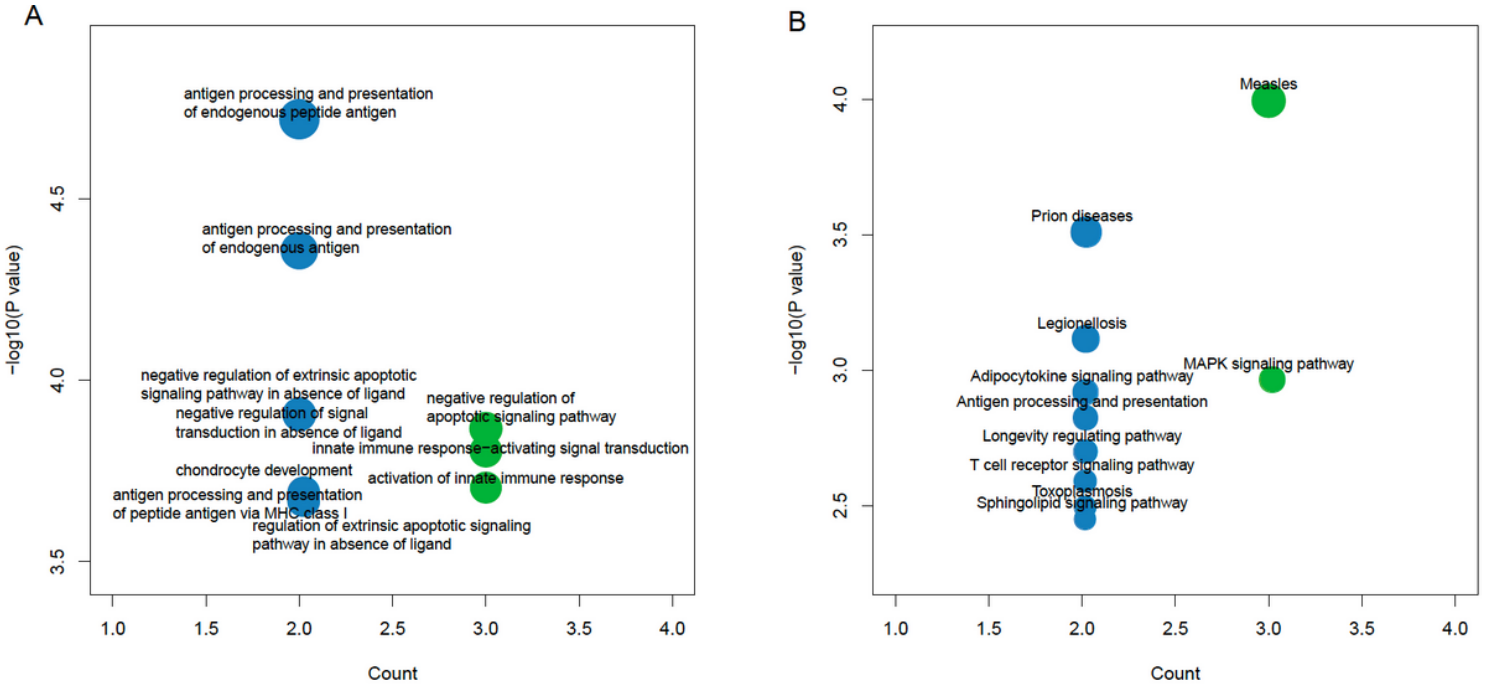

Figure 12

The GO (A) and KEGG pathway (B) enrichment analysis of the 9 specific immune-related genes.

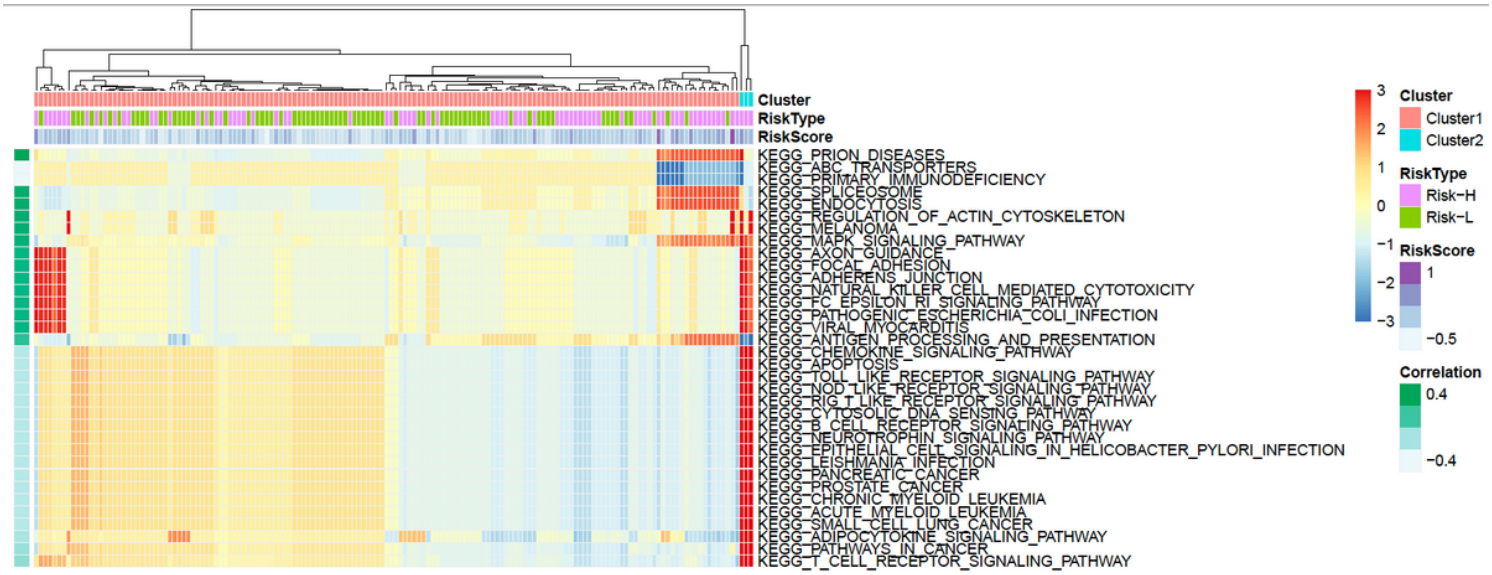

Figure 14

Correlation of RiskScore with signaling pathways. KEGG functional enrichment score of each sample was analyzed, the correlation with RiskScore was calculated, respectively, based on the enrichment score of each pathway in each sample, and all the 34 pathways related KEGG pathways were shown. Clustering analysis had to be carried out according to the enrichment score.

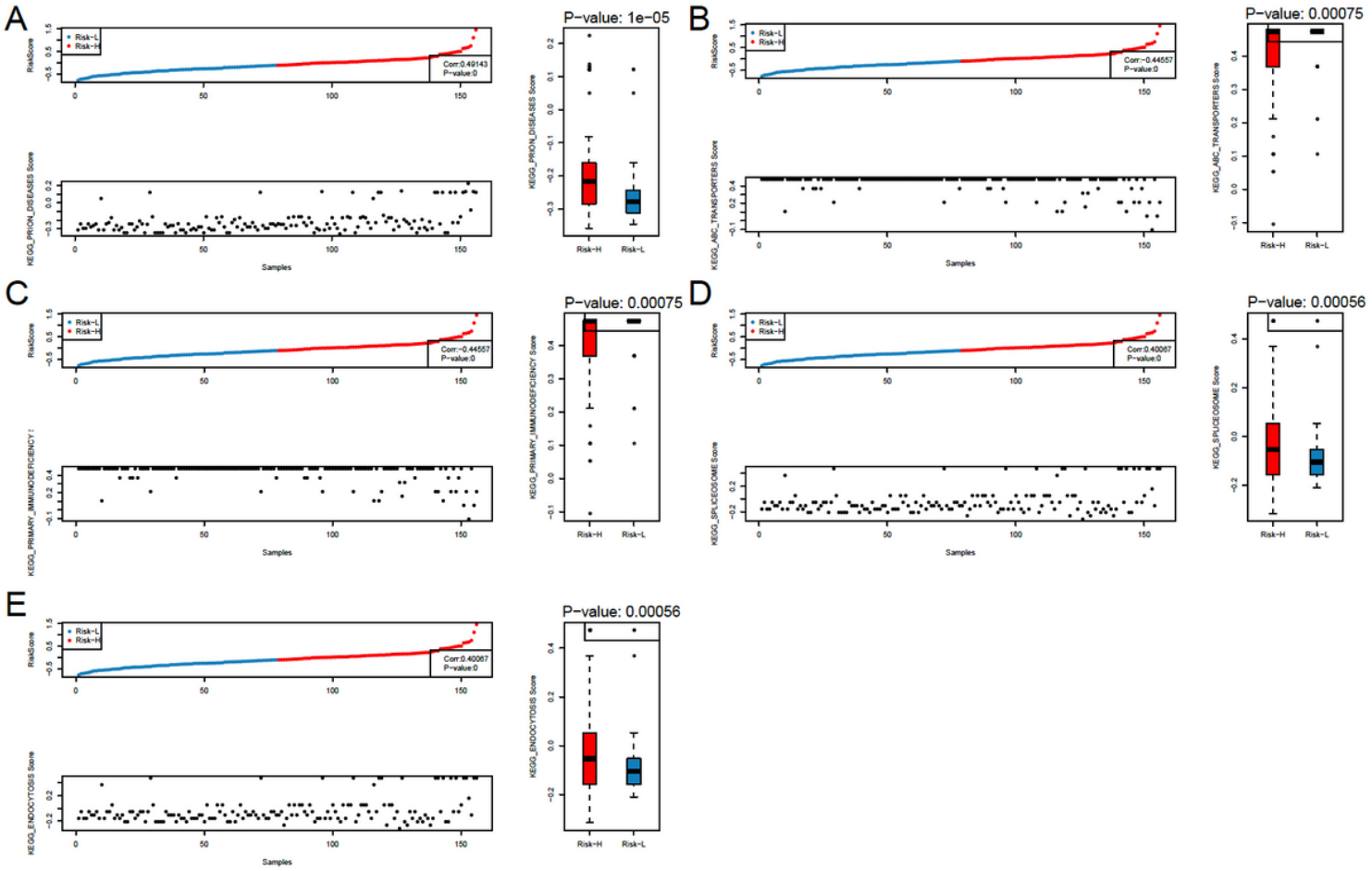

Figure 16

The relationships of enrichment score of prion diseases (A), ABC transporters (B), primary immunodeficiency (C), spliceosome (D) and pendocytosis (E) with RiskScore for each sample.

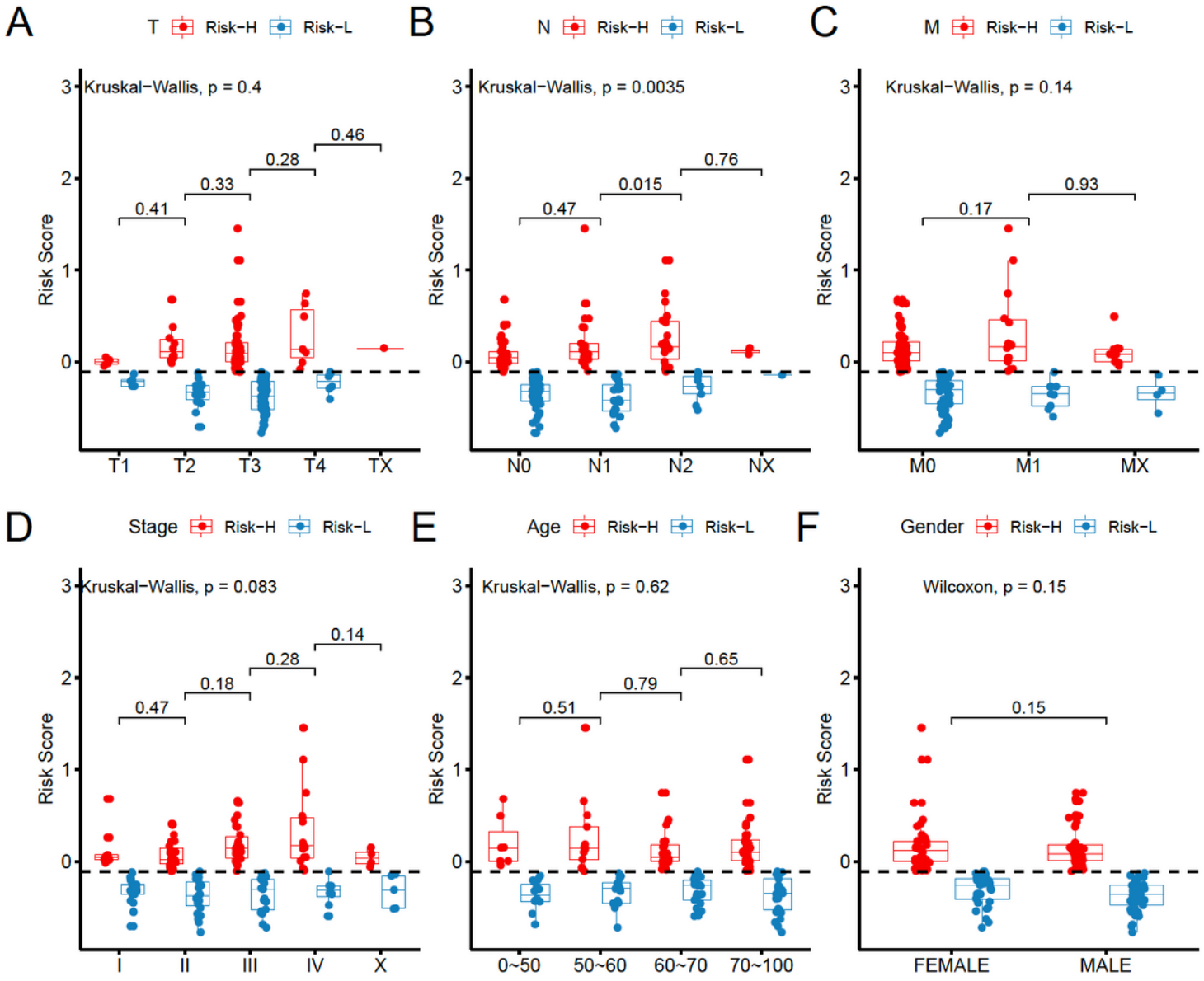

Figure 18

The relationships of different clinical factors with RiskScore for READ. Comparison of RiskScore among different T (A), N (B), M (C), stage (D), age (E) and gender (F). The horizontal axis represents the different clinical factors, and the vertical axis represents RiskScores. The red dots represent Risk-H samples, the blue dots represent Risk-L samples, and the dotted line represents the median RiskScore.

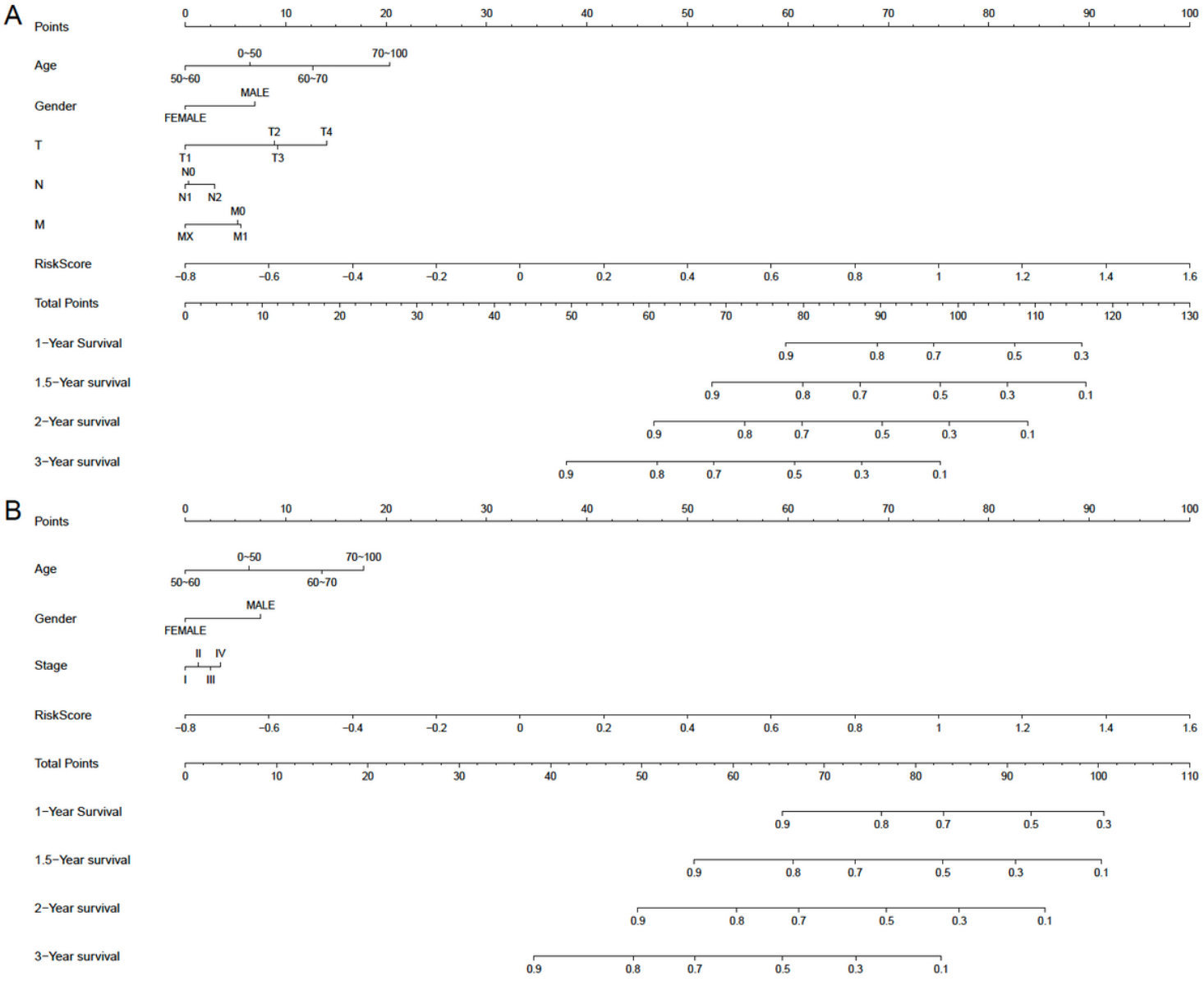

Figure 20

The nomogram model constructed by combining the TNM+age+gender (A) or stage+age+gender (B) with RiskScore for READ patients

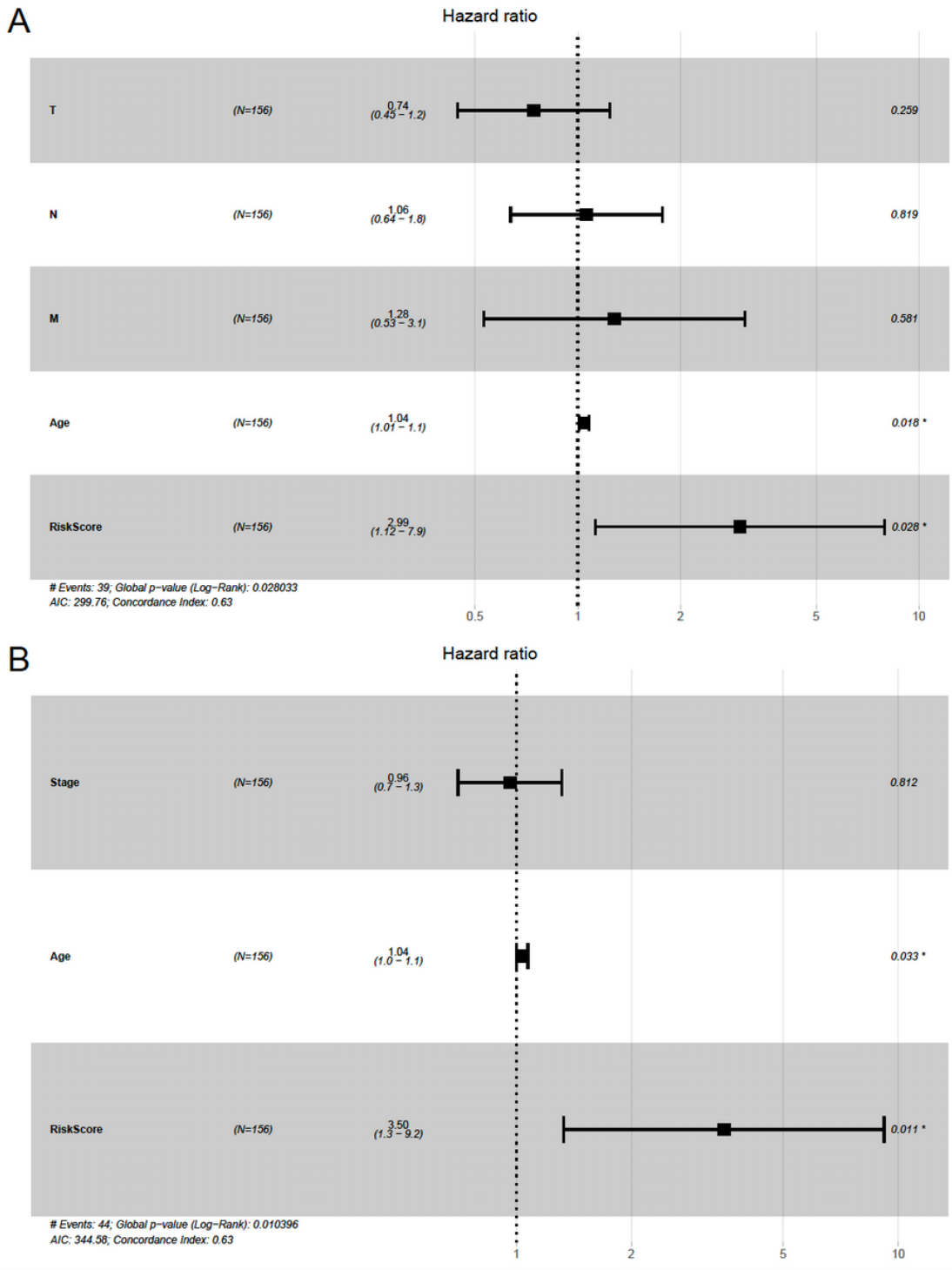

Figure 22

The forest plot constructed by combining the TNM+age (A) or stage+age (B) with RiskScore for READ patients.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- S6S12Table.xlsx

- S6S12Table.xlsx
- S2Table.txt
- S3Table.txt
- S4Table.txt
- S5Table.txt
- S3Table.txt
- S4Table.txt
- S5Table.txt
- S1Figure.pdf
- S1Table.txt
- S2Figure.pdf
- S2Table.txt
- S1Table.txt
- S2Figure.pdf
- S1Figure.pdf