# Development and validation of an immune-related gene-based prognostic signature in gastric cancer

**Penglei Ge** ( doc677@126.com )
First Affiliated Hospital of Zhengzhou University
**Xiaofang Chen**
First Affiliated Hospital of Zhengzhou University
**Yang Wu**
First Affiliated Hospital of Zhengzhou University
**Yubin Fu**
First Affiliated Hospital of Zhengzhou University
**Chunbo Li**
First Affiliated Hospital of Zhengzhou University
**Zhengkai Feng**
First Affiliated Hospital of Zhengzhou University
**Jiahao Xue**
First Affiliated Hospital of Zhengzhou University
**Lin Li**
First Affiliated Hospital of Zhengzhou University
**Gong Zhang**
First Affiliated Hospital of Zhengzhou University
**Zhiqiang Gao**
First Affiliated Hospital of Zhengzhou University
**Xiaowei Dang**
First Affiliated Hospital of Zhengzhou University

**Research**

**Keywords:** gastic cancer, immune, gene expression, biomarker, prognostic signature

**Posted Date:** February 6th, 2020

**DOI:** https://doi.org/10.21203/rs.2.22778/v1

# Abstract

Background: Gastric cancer is a common lethal cancer worldwide. We aimed to develop a reliable, individualized, immune-related prognostic signature that can be used to stratify and estimate prognosis in patients with gastric cancer. Methods: This retrospective study analyzed the gene expression profiles of gastric cancer with tumor tissue samples from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) cohorts, which included 676 cases in total. Immune genes from the InnateDB database were selected to develop and validate an immune-related prognostic model for gastric cancer patients. Results: An immune-related gene pair (IRGP) model was constructed that enabled us to stratify patients into high- and low-risk immune risk groups in the training set. Patients with a low risk score had a significantly longer median survival time than those with a high risk score. Further, we compared the predictive accuracy of the IRGP model with clinical characteristics, including TNM, grade, age, and stage. The results showed that the model had the highest mean C-index (0.69) compared with grade (0.55) or stage (0.60) in survival prediction. Then, we constructed a nomogram that integrated the IRGP model with independent clinical characteristics, which showed the best prognostic accuracy compared with other signatures. Conclusion: A clinical-immune signature based on IRGP is a promising prognostic biomarker in gastric cancer. Prospective studies are needed to further validate its accuracy and to test its clinical utility in individualized treatment.


# Background

Gastric cancer is the fifth most common malignancy and the second leading cause of cancer-related death in the world [1]. Surgical resection is currently the most effective therapeutic method for gastric cancer. Although the 5-year overall survival in early gastric cancer can exceed 90% with early diagnosis and multidisciplinary treatment, long-term outcomes for patients with advanced gastric cancer remain unsatisfactory [2–3]. The TNM staging and histological subtype systems are the most used clinicopathological variables for routine prognostic prediction and treatment of gastric cancer. However, large clinical studies have shown that the present TNM staging system does not provide accurate prognostic value and cannot be used to stratify high-risk patients [4–5]. Therefore, novel prognostic biomarkers for gastric cancer are urgently needed to identify patients with worse survival and higher mortality and to optimize treatment strategies for them.

Immunotherapy, as an emerging treatment, has shown significant therapeutic promise for some cancers [6–7]. In-depth studies on immune response are increasingly common. The various components of the immune system are recognized as determinants for cancer development and progression [8]. Evading immune damage has been recognized as a new hallmark of cancer [9–10]. Increasing evidence has suggested that the immune-related components of the tumor microenvironment not only reflect the immune response and chemotherapy benefit [11–12], but also correlate with clinical outcomes in various cancers, including gastric cancer [13]. Therefore, understanding the immune features of gastric cancer may offer clinicians new tools to treat the disease and predict prognosis. Publicly available large-scale

gene expression datasets provide a way for us to conduct a more generalized prognostic analysis based on immune-related gene expression signatures.

In this study, we integrated multiple cohorts with gene expression data to develop and validate individualized prognostic features of gastric cancer with immune-related gene pairs (IRGPs). To exploit the complementary value of molecular and clinical characteristics, we further performed an integrated analysis to improve the predicted accuracy of gastric cancer.

# Methods

### Gene expression and clinical data from TCGA and GEO cohorts

Transcriptome expression profiles and corresponding clinical information of gastric cancer were downloaded from the Genomic Data Commons Data Portal of The Cancer Genome Atlas (TCGA, https://www.cancergenome.nih.gov/). The expression data was FPKM (fragments per kilobase million) type, which contained 376 gastric cancer cases as of December 2018. Samples without clinical information or with overall survival time <30 days were excluded. Genes with FPKM of zero in more than half of samples were also eliminated. Only expression profiles of immune-related genes were included.

For the second dataset, the microarray data and clinical information of GSE62254 were downloaded from Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/), which contained 300 gastric cancer samples. Only primary tumor data and expression profiles of immune-related genes were included, and samples with overall survival time <30 days were removed.

### Immune-related gene pair construction

We constructed a prognostic signature based on immune-related genes from data downloaded from the InnateDB database (https://www.innatedb.com/). The data included various endogenous immune-related genes (IRGs) of several species supported by the literature and manually corrected [14]. In this study, human IRGs were collected, and a total of 1039 genes were eligible after removing genes with repeated names. The list of immune-related genes is given in Additional File 1.

To assess the difference between RNA-seq data from TCGA and microarray data from GEO, we used IRG and IRGP values to perform correlation clustering analyses for samples from the two cohorts, respectively.

First, a number of IRGPs were constructed, with the number being (1039 × 1038)/2. The IRGP values were calculated based on the specific gene expression level to generate a score. The IRGP score was assigned as 1 if IRG1 was less than IRG2; otherwise, it was 0. Then, the IRGP score of each sample in both the TCGA and GEO cohorts was calculated. Because of platform factors and biologically preferential transcription [15], some IRGPs had constant values (0 or 1). To improve the accuracy of the results, we removed IRGP scores with constant values (0 or 1) for all samples in both cohorts.

### Integrating IRGPs from different datasets and grouping

To build a prognostic signature, the IRGPs shared by the two datasets were extracted to divide them into different sets. The TCGA data were divided into a training set and validation set. The IRGPs of GEO were used as an independent testing set.

To prevent the bias of random allocation from affecting the stability of the subsequent model, samples in the TCGA cohort were repeatedly grouped 100 times in different randomizations. The group sampling was in accordance with a training set:validation set ratio of 0.5:0.5. Criteria for selecting the most appropriate training set and testing set should be that (1) the two groups were similar in age distribution, clinical stage, follow-up time, and patient death rate, and (2) the number of the two classifications of the two data sets randomly clustered was close.

### Construction and validation of an immune-related prognostic model

We developed the immune-related prognostic model using the training set. The significant different IRGPs were defined by univariate Cox regression analysis with $p$ value = 0.05. Because a traditional Cox regression model could not be used directly to select highly correlated genes, least absolute shrinkage and selection operator (LASSO) was used to shrink the regression coefficients; LASSO is a type of biased estimation with complex collinearity data and it better solves the multicollinearity problem in regression analysis [16-17]. LASSO L1-penalized Cox analysis was performed by using glmnet R package, and a relatively small number of potential parameters were selected. Finally, an immune-related prognostic model was constructed using the regression coefficients derived from multivariate Cox regression analysis. Then, the risk score of each sample was calculated. To separate patients into low- or high-risk groups, the median risk score was determined as the cut-off value. The prognostic value of the IRGP model was evaluated using receiver operating characteristic (ROC) regression and Kaplan-Meier survival analysis in different cohorts. Then, the immune-related prognostic model was validated in other data sets.

### Functional annotation and enrichment analysis of the immune-related genes in the model

To gain biological understanding of the IRGPs, functions of all the immune-related genes in IRGPs were explored. Then, enrichment analysis was conducted using Gene Ontology (GO). The immunological pathways of genes in Kyoto Encyclopedia of Genes and Genomes (KEGG) were also determined using an R package (clusterProfiler, v3.8).

### Construction and validation of a composite immune-clinical prognostic model

Prognostic risk models were constructed using clinical features of TNM, grade, age, and stage from the TCGA dataset. Then, we integrated traditional prognostic clinical factors and the IRGP risk model to create a composite immune-clinical prognostic index by applying Cox regression in the TCGA dataset using the R package rms.

To individualize the predicted survival probability for 1, 3, 5, or 7 years, we constructed a nomogram based on the clinical characters and RiskScore results of the multivariate analysis. The nomogram, which uses the length of a line to indicate the degrees of different factors affecting the result, is a method that can display the results of a risk model more intuitively and effectively. It was convenient to apply in the prediction of outcomes. We constructed the nomogram model with the clinical features TNM, stage, age, and RiskScore.

A concordance index (C-index) was used to determine the discrimination of the nomogram, which was assessed by a bootstrap approach with 1000 resamples [18]. Furthermore, the predictive accuracies of the nomogram and other clinical characteristics were compared by using the C-index. ROC curve analysis was also assessed to predict the prognostic values of models by age, stage, and IRGPs. Then, we validated the accuracy of the nomogram in predicting prognosis using different datasets.

### Statistical analysis

All analyses were conducted using R software (version 3.5.2; https://www.r-project.org/). Univariate Cox proportional hazards regression analysis of the relationship of IRGP and clinical factors with overall survival was assessed using log-rank test. LASSO regression analysis was used to further condense the quantity of factors significantly associated with overall survival in univariate analyses. Survival R package was used for Kaplan-Meier curve analysis. The C-index was estimated using the R rms package, which was also used to build the nomogram, including significant clinical characteristics and calibration plots. Survival rate of ROC analysis was calculated by time ROC R package. All statistical tests were two-sided, and $p$ values < 0.05 were considered statistically significant.

## Results

### Data from TCGA and GEO cohorts

A flowchart of the analysis procedure for the study is shown in Fig. 1. A total of 336 patients diagnosed with gastric cancer in the TCGA cohort remained. After preprocessing the initial data, we further divided the TCGA dataset into a training set and a validation set, which contained, respectively, 167 and 169 patients (Additional Files 2, 3). The GEO dataset contained 266 patients and was considered the independent testing set. Detailed clinical characteristics of all patients are listed in Table 1.

Table 1. Clinical information of all the patients from TCGA and GEO cohorts.

| TCGA | | Training set | Validation set | Testing set |
|---|---|---|---|---|
| Event | | | | |
| Alive | 197 | 92 | 105 | 148 |
| Dead | 139 | 75 | 64 | 118 |
| T | | | | |
| T1 | 15 | 4 | 11 | |
| T2 | 74 | 37 | 37 | 169 |
| T3 | 156 | 72 | 84 | 78 |
| T4 | 87 | 51 | 36 | 19 |
| TX | 4 | 3 | 1 | |
| N | | | | |
| N0 | 99 | 49 | 50 | 36 |
| N1 | 91 | 41 | 50 | 116 |
| N2 | 68 | 34 | 34 | 67 |
| N3 | 68 | 35 | 33 | 47 |
| NX | 10 | 8 | 2 | |
| M | | | | |
| M0 | 302 | 143 | 159 | 243 |
| M1 | 22 | 16 | 6 | 23 |
| MX | 12 | 8 | 4 | |
| Stage | | | | |
| I | 45 | 17 | 28 | 28 |
| II | 106 | 52 | 54 | 87 |
| III | 137 | 70 | 67 | 81 |
| IV | 34 | 20 | 14 | 68 |
| X | 14 | 8 | 6 | 2 |
| Grade | | | | |
| G1 | 9 | 4 | 5 | |
| G2 | 119 | 54 | 65 | |

| TCGA | | Training set | Validation set | Testing set |
|---|---|---|---|---|
| G3 | 199 | 104 | 95 | |
| GX | 9 | 5 | 4 | |
| Age | | | | |
| 0 ~ 50 | 28 | 13 | 15 | |
| 50 ~ 60 | 80 | 37 | 43 | |
| 60 ~ 70 | 100 | 51 | 49 | |
| 70 ~ 80 | 109 | 58 | 51 | |
| 80 ~ 100 | 19 | 8 | 11 | |

### Consistency results of datasets from the two cohorts

We used IRG and IRGP data to make correlation clustering analyses for samples from the TCGA and GEO datasets, respectively. The results showed that IRG data were obviously separated from the two platforms (Additional Figure 1). The IRGP data could also separate data from the two platforms (Additional Figure 2), but the difference was significantly decreased and the correlation was closer (Additional Figure 3). This analysis showed that platform differences were largely eliminated after the two datasets had been transformed by IRGP. This meant that the scores calculated entirely by gene expression profile of a tumor sample could be used to compare data between different platforms without the need for normalization.

### Construction of an immune-related prognostic model and evaluation of its predictive ability

The training and validation sets had, respectively, 167 and 169 samples (Additional Files 4, 5). Detailed information of the two groups is given in Table 1.

A total of 358,255 IRGPs from the TCGA cohort were divided into training and validation sets; 9620 IRGPs that were significantly different by univariate Cox regression analysis were selected for the training set (Additional File 6). The relationship between *p* value and hazard ratio (HR) of the 9620 genes is shown in Figure 2. The HR values corresponding to the IRGPs with significant *p* values usually deviated from 1. Because there were too many correlated covariates, we applied Cox-proportional hazards analysis to select the greatest prognostic value of IRGPs basing on LASSO estimation. Then, 11 IRGPs were selected to construct an IRGP model by using L1-penalized Cox proportional hazards regression on the training set, which was S100A8_vs_TREML2*-1.70 + CD36_vs_NOD1*1.09 + PTGS2_vs_IL1A*-1.37 + HDAC11_vs_C4BPA*-0.89 + MID2_vs_TRAF3*1.13 + ZBP1_vs_TRIM62*-2.29 + AAMP_vs_ITGAV*-0.56 + CCNT1_vs_TRAF2*0.50 + CTCF_vs_CSK*0.38 + ARF6_vs_ITGAV*-0.31 + IL4R_vs_CSK*-0.16.

Considering that the integral distribution of overall survival time of the sample was >2 years, we evaluated the prediction effects of the model for 1, 3, and 5 years using the four datasets. The average areas under the curve (AUC) of the training set, validation set, testing set, and TCGA+GEO were 0.89, 0.74,

0.64, and 0.67, respectively (Figure 3). The model significantly stratified patients into high- and low-risk (Risk-H and Risk-L, respectively) immune risk groups in the training set. The distribution of Risk-H and Risk-L group samples in different overall survival times is shown in Figure 3. We also found that the quantity of samples in Risk-H and Risk-L group in 0 year and 1 year was not much different. However, the number of samples in the Risk-H group gradually became less than those in the Risk-L group 5 years later, and the tendency became more significant as overall survival time prolonged (Figure 3).

We predicted Kaplan-Meier survival curves for the Risk-H and Risk-L groups based on the 11-IRGP predictive model on the training set, validation sed, and testing set, TCGA+GEO dataset. There was a significant difference in overall survival time between the two groups in all datasets, and patients with a low risk score had a significantly longer median survival than those with a high risk score (Figure 4).

### Pathway enrichment analysis of immune-related genes in the prognostic model

Enrichment analysis for the 20 unique genes in the 11 IRGPs identified 12 significant genes ($p < 0.05$) (Table 2). The first three obviously enriched genes belonged to the gene families histone deacetylases, P-TEFb complex, and TNF receptor-associated factors. The most significant function and pathway enrichments were positive regulation of nuclear factor kappa-B (NF-κB) transcription factor activity and interleukin (IL)-17 signaling pathway (Figure 5).

Table 2. Results of enrichment analysis for the 20 unique genes in the 11 IRGPs.

| GeneFamily | Genes | pvalue | padj |
|---|---|---|---|
| Histone deacetylases, class IV | HDAC11 | 0.00 | 0.03 |
| P-TEFb complex | CCNT1 | 0.00 | 0.09 |
| TNF receptor associated factors | TRAF2 | 0.01 | 0.14 |
| Zinc fingers | TRAF3 | 0.01 | 0.27 |
| Integrin alpha subunits | ITGAV | 0.02 | 0.33 |
| S100 calcium binding proteins | S100A8 | 0.02 | 0.43 |
| NLR family | NOD1 | 0.02 | 0.44 |
| Scavenger receptors | CD36 | 0.02 | 0.48 |
| ARF GTPase family | ARF6 | 0.03 | 0.55 |
| Interleukin receptors | IL4R | 0.04 | 0.73 |
| Interleukins | IL1A | 0.04 | 0.75 |
| Sushi domain containing | C4BPA | 0.05 | 0.98 |
| X-linked mental retardation | MID2 | 0.07 | 1 |
| Tripartite motif containing | TRIM62 | 0.08 | 1 |
| SH2 domain containing | CSK | 0.08 | 1 |
| V-set domain containing | TREML2 | 0.13 | 1 |
| WD repeat domain containing | AAMP | 0.21 | 1 |
| Zinc fingers C2H2-type | CTCF | 0.47 | 1 |
| unknown | PTGS2/ZBP1 | 1 | 1 |

### Comparison of the immune-related model with clinical characteristics

As the significant correlation between clinical features of TNM, stage, and prognostic results, there were significant prognostic differences between the samples grouped by clinical features from the TCGA dataset (Additional Figure 4). We compared the predictive accuracy of the immune-related prognostic model with clinical characteristics including TNM, grade, age, and stage. The model had the highest mean C-index (0.69) compared with grade (0.55) and stage (0.60) (Figure 6). These results indicated that the 11-gene-based risk model performed better than conventional characteristics in predicting survival.

### Integrating the immune-related prognostic model with clinical characteristics

To provide clinicians with a quantitative approach to predicting the prognosis of gastric cancer patients, we constructed a nomogram that integrated the immune-related prognostic model with independent clinical characteristics [19]. Because the M stage had no significant correlation with grade and prognosis, only the TN+Stage+Age+RiskScore nomogram was constructed (Figure 7). In the nomogram, a point scale was used to assign points to these variables. Then, the sum of points for all variables was recorded as the total points. The probability of patient survival at 1, 3, 5, and 7 years is assessed by drawing a vertical line from the total point axis straight downward to the relevant outcome axis. The RiskScore was found to contribute the most risk points, ranging from 0 to 100, compared with other clinical factors.

We compared the predictive accuracy of this nomogram in training set, whose performance (C-index: 0.73) (Figure 6) was better than that of Stage+Age+RiskScore (C-index: 0.72), RiskScore (C-index: 0.69), or Stage (C-index: 0.60). Then, we used the datasets from validation set and GSE62254 to further validate the accuracy of the nomogram in predicting prognosis. The time-dependent AUC indicated that the nomogram had relatively accurate prediction for 1-, 3-, 5-, and 7-year survival rate, which were 0.75, 0.81, 0.81, 0.75 in validation set (Figure 8A), and 0.85, 0.81, 0.78, 0.79 in GSE62254 set (Figure 8B), respectively. In sum, these findings suggested that the nomogram was better for predicting survival time in gastric cancer patients than other clinical prognostic factors.

# Discussion

Gastric cancer remains one of the deadliest malignant tumors worldwide because of its complicated cellular and molecular heterogeneity [3]. Patients suffering from gastric cancer are at substantial risk for recurrence and death, even after radical surgery. To improve the overall survival time of these patients, other therapeutic methods, such as chemotherapy, radiotherapy, targeted therapy, and immunotherapy, should be performed. Because of tumor heterogeneity, traditional treatments have not achieved very good results. Treatment options are lacking for all the patients to treat with the similar methods. Thus, stable and accurate prognostic signatures are urgently needed to identify those patients with refractory disease and worse survival. As an emerging treatment, immunotherapy has shown encouraging results in many cancers, including gastric cancer [20]. It is very important for clinicians to identify patients at a high risk of recurrence and who may benefit from additional treatment. Therefore, reliable prognostic biomarkers are critically needed to select these patients and administer appropriate treatment. Some prognostic models do currently exist for gastric cancer [21–22], but their accuracy at estimating survival remains limited.

Large-scale public cohorts with gene expression data have allowed researchers to establish more generalized prognostic signatures for gastric cancer. In view of the biological heterogeneity among different data sets and the technical bias across measurement platforms, gene expression data should be appropriately normalized to make a predictive model robust. To overcome this problem, we used a relative ranking method based on gene expression levels to eliminate the difficulties in pre-processing data, a method that has shown good performance in non-small-cell lung cancer [23].

In this study, we combined immune-related genes in the tumor microenvironment with clinical markers, developed a prognostic model based on 11 immune-related gene pairs for gastric cancer, and validated it in independent data sets. Our results showed that the immune-related model could predict prognosis and stratify patients into high- and low-risk subgroups better than traditional clinical characteristics. We further compared the prognostic values of immune-related genes and significant clinical characteristics and found that models integrating both had a more accurate estimation.

Studies have indicated that the tumor microenvironment, mainly comprising immune cells, plays a critical role in tumor growth and progression [24–25]. Accumulating evidence indicates that innate and adaptive immune systems are crucial factors in tumor initiation and progression [26]. Evading immune damage had been recognized as a new hallmark of cancer. Understanding the immune status and functions of infiltrated immune cells in the tumor microenvironment could help to improve the response rate of immunotherapy. Certain histopathological patterns, such as intratumoral infiltration of cytotoxic lymphocytes, are associated with better prognosis for several cancer types [27–28]. The InnateDB database supplied human innate immune genes. In this study, we identified 11 IRGPs consisting of 20 immune-related genes. Most genes in our signature were related to cytokines, cytokine receptors, and antivirals. Studies have demonstrated that cytokines and cytokine receptors play important roles in inflammatory processes, chemotaxis, and angiogenesis [29]. They can directly cause programmed death of tumor cells, increase the number and activity of immune effector molecules, and improve the ability of the immune system to recognize tumor cells [30–31].

The two most significant gene ontology were positive regulation of NF-κB transcription factor activity and regulation of multi-organism process. The former can activate T cells by upregulating a variety of cytokines, such as tumor necrosis factor (TNF)-α and interferon-γ, which are necessary for the immune response. Then, NF-κB can regulate diverse processes such as innate immunity, inflammation, and cell proliferation and apoptosis [32]. The pathways of the genes including IL-17 signaling pathway, TNF signaling pathway, necroptosis and so on, all of which could perform a role in inflammatory response and immune regulation.

Clinical stage has been the traditional standard for distinguishing the degree of malignancy of cancers. Patients with cancers of different stages had different prognoses, but the accuracy of the prognosis was still controversial [33]. A risk model based on immune-related genes could more accurately reflect patients' immune status and prognosis. In each clinical stage subgroup, the Risk-H and Risk-L samples were generally different, which showed that the model could better evaluate the risk of patients and indicated that the prognosis of patients with gastric cancer was highly correlated with their clinical features.

To improve the survival time of patients suffering from gastric cancer, it is necessary to identify patients at a high risk of recurrence through prognostic models and to develop personalized treatments. However, traditional clinical pathological stage cannot accurately stratify the patients. Our prognostic predictive model has several strengths. First, only immune-related genes that have clearly physiological functions

were selected. This is more practical, especially for patients undergoing immunotherapy. Second, compared with other signatures and clinical features, the model was more robust in stratifying patients and predicting survival. Third, to eliminate the difficulty of data preprocessing from different platforms and normalizing them, we used a relative ordering method that previously showed robust performance in non-small-cell lung cancer [23].

# Limitations

Although our research provides new insights into the immune microenvironment and immune-related therapies of gastric cancer, it has some limitations. First, it is a retrospective study, so the results must be confirmed by prospective studies. Second, the data sets are not large enough. Third, the model is developed from numerous immune-related genes, and further experiments of their function and mechanism in gastric cancer is warranted to support their clinical application.

# Conclusion

A clinical-immune signature based on immune-related gene pairs is a promising prognostic biomarker in gastric cancer, and it could be used to stratify and predict patients' survival. Prospective studies using larger clinical and gene expression data sets are needed to validate its accuracy and test its clinical utility in individualized treatment of gastric cancer.

# Abbreviations

IRGPs: Immune-related gene pairs; TCGA: The cancer genome atlas; FPKM: Fragments per kilobase million; GEO: Gene Expression Omnibus; LASSO: Least absolute shrinkage and selection operator; ROC: Receiver operating characteristic; GO: Gene ontolgy; KEGG: Kyoto Encyclopedia of Genes and Genomes; C-index: Concordance index

# Declarations


### Acknowledgements

We thank Louise Adam, ELS(D), from Edanz Editing (www.liwenbianji.cn/ac) for editing the English text of a draft of this manuscript.


### Authors' contributions

PLG and XFC carried out data management and drafted the manuscript. XFC and YW helped with data management. YBF, ZKF and JHX searched the literature. CBL and LL helped with the statistical analysis.

GZ, ZQG and XWD interpreted the findings. PLG performed project administration. All authors designed the study, read and approved the final manuscript.


## Funding

None.


## Availability of data and materials

All data generated or analysed during this study are included in this published

article.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

All authors have read and approved the content and agree to submit for consideration for publication in the journal.

## Competing interests

All authors have declared that there are no conflicts of interest related to the contents of this article.

## Author details


[1]Department of Hepatobiliary and Pancreatic Surgery, The First Affiliated Hospital of Zhengzhou University, No. 1 Jianshe East Road, Zhengzhou, Henan Province, China.

[2]Department of Medical Insurance, The First Affiliated Hospital of Zhengzhou University, No. 1 Jianshe East Road, Zhengzhou, Henan Province, China.

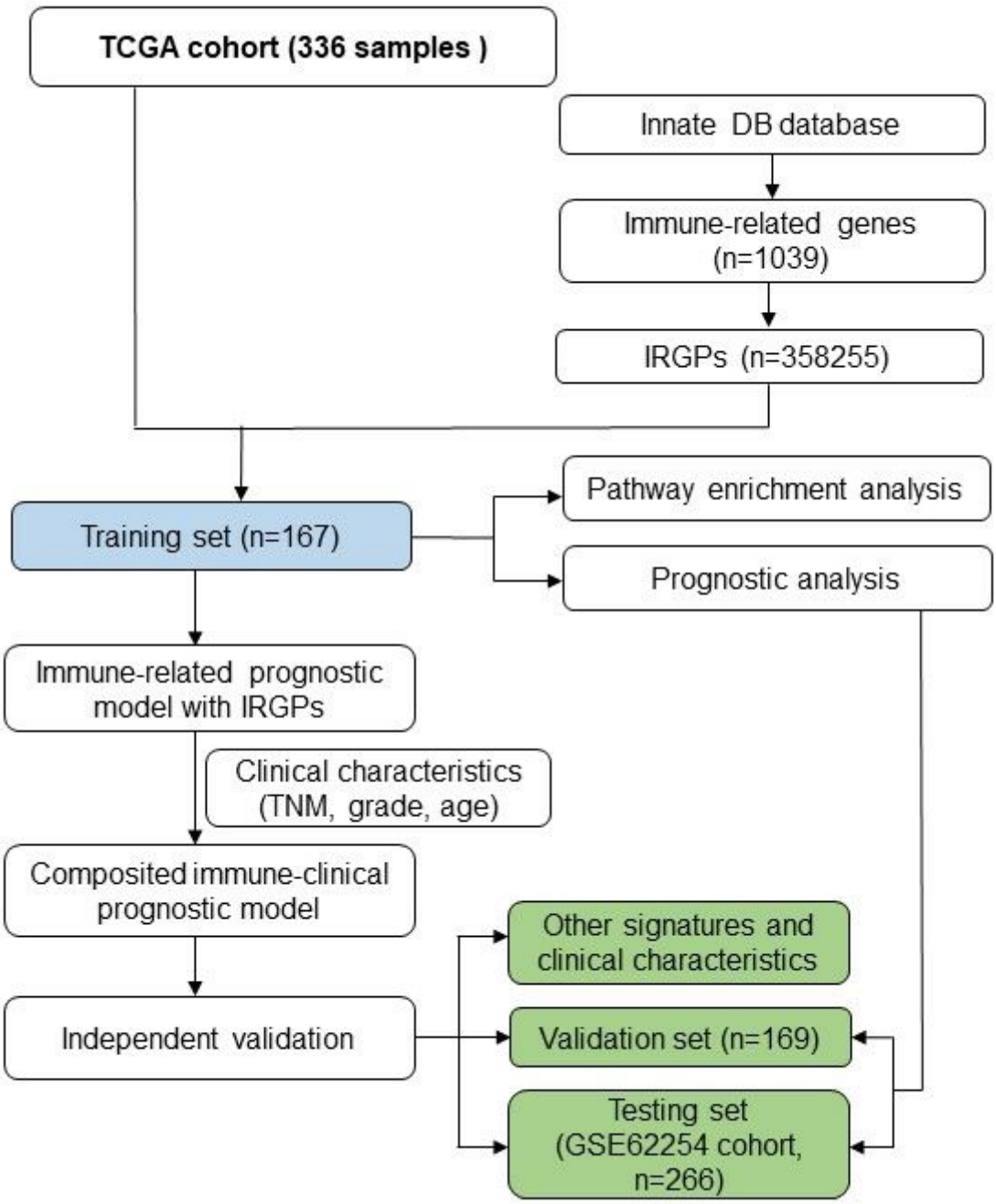
# References


1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68:394-424.
2. GASTRIC (Global Advanced/Adjuvant Stomach Tumor Research International Collaboration) Group, Oba K, Paoletti X, Bang YJ, Bleiberg H, Burzykowski T, Fuse N, Michiels S, Morita S, Ohashi Y, et al. Role of chemotherapy for advanced/recurrent gastric cancer: an individual-patient-data meta-analysis. Eur J Cancer. 2013;49:1565-77.
3. Van Cutsem E, Sagaert X, Topal B, Haustermans K, Prenen H. Gastric cancer. Lancet. 2016;388:2654-64.
4. McLean MH, El-Omar EM. Genetics of gastric cancer. Nat Rev Gastroenterol Hepatol. 2014;11:664-74.
5. Wadhwa R, Song S, Lee JS, Yao Y, Wei Q, Ajani JA. Gastric cancer-molecular and clinical dimensions. Nat Rev Clin Oncol. 2013;10:643-55.
6. Sullivan RJ, Flaherty KT. Immunotherapy: Anti-PD-1 therapies-a new first-line option in advanced melanoma. Nat Rev Clin Oncol. 2015;12:170-1.
7. Esteva FJ, Hubbard-Lucey VM, Tang J, Pusztai L. Immunotherapy and targeted therapy combinations in metastatic breast cancer. Lancet Oncol. 2019;20:175-86.
8. 8 Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, Nair VS, Xu Y, Khuong A, Hoang CD, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med. 2015;21:938-45.
9. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646-74.
10. Gentles AJ, Bratman SV, Lee LJ, Harris JP, Feng W, Nair RV, Shultz DB, Nair VS, Hoang CD, West RB, et al. Integrating tumor and stromal gene expression signatures with clinical indices for survival stratification of early-stage non–small cell lung cancer. J Natl Cancer Inst. 2015;107:djv211.
11. Jiang Y, Zhang Q, Hu Y, Li T, Yu J, Zhao L, Ye G, Deng H, Mou T, Cai S, et al. ImmunoScore Signature: A Prognostic and Predictive Tool in Gastric Cancer. Ann Surg. 2018;267:504-513.
12. Turley SJ, Cremasco V, Astarita JL. Immunological hallmarks of stromal cells in the tumour microenvironment. Nat Rev Immunol. 2015;15:669-82.
13. Nishino M, Ramaiya NH, Hatabu H, Hodi FS. Monitoring immune-checkpoint blockade: response evaluation and biomarker development. Nat Rev Clin Oncol. 2017;14:655-68.
14. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, Winsor GL, Hancock RE, Brinkman FS, Lynn DJ. InnateDB: systems biology of innate immunity and beyond–recent updates and continuing curation. Nucleic Acids Res. 2013;41:D1228-33.
15. Kim S, Lin CW, Tseng GC. MetaKTSP: a meta-analytic top scoring pair method for robust cross-study validation of omics prediction analysis. Bioinformatics. 2016;32:1966-73.
16. Gui J, Li H Penalized. Cox regression analysis in the high-dimensional and lowsample size settings, with applications to microarray gene expression data. Bioinformatics. 2005;21:3001-8.

17. Tibshirani R, Bien J, Friedman J, Hastie T, Simon N, Taylor J, Tibshirani RJ. Strong rules for discarding predictors in lasso-type problems. J R Stat Soc Series B Stat Methodol. 2012;74:1-22.
18. Kiran M, Chatrath A, Tang X, Keenan DM, Dutta AA. A prognostic signature for lower grade Gliomas based on expression of Long non-coding RNAs. Mol Neurobiol. 2019;56:4786-98.
19. Long J, Wang A, Bai Y, Lin J, Yang X, Wang D, Yang X, Jiang Y, Zhao H. Development and validation of a TP53-associated immune prognostic model for hepatocellular carcinoma. EbioMedicine. 2019;42.
20. Gambardella V, Tleitas T, Cervantes A. Understanding mechanisms of primary resistance to checkpoint inhibitors will lead to precision immunotherapy of advanced gastric cancer. Annals of Oncology. 2019;30.
21. Koo DH, Ryoo BY, Kim HJ, Ryu MH, Lee SS, Moon JH, Chang HM, Lee JL, Kim TW, Kang YK. A prognostic model in patients who receive chemotherapy for metastatic or recurrent gastric cancer: validation and comparison with previous models. Cancer Chemother and Pharmacol. 2011;68:913-21.
22. Wang J, Qu J, Li Z, Che X, Zhang J, Liu J, Teng Y, Jin B, Zhao M, Liu Y, et al. A prognostic model in metastatic or recurrent gastric cancer patients with good performance status who received first-line chemotherapy. Transl Oncol. 2016;9:256-61.
23. Li B, Cui Y, Diehn M, Li R. Development and validation of an individualized immune prognostic signature in early-stage nonsquamous non–small cell lung cancer. JAMA Oncol. 2017;3:1529-37.
24. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. Nat Med. 2013;19:1423-1437.
25. Lim YZ, South AP. Tumour-stroma crosstalk in the development of squamous cell carcinoma. Int J Biochem Cell Biol. 2014;53:450-8.
26. Ogino S, Galon J, Fuchs CS, Dranoff G. Cancer immunology-analysis of host and tumor factors for personalized medicine. Nat Rev Clin Oncol. 2011;8:711-9.
27. Gartrell RD, Marks DK, Hart TD, Li G, Davari DR, Wu A, Blake Z, Lu Y, Askin KN, Monod A, et al. Quantitative Analysis of Immune Infiltrates in Primary Melanoma. Cancer Immunol Res. 2018;6:481-93.
28. Dienstmann R, Villacampa G, Sveen A, Mason MJ, Niedzwiecki D, Nesbakken A, Moreno V, Warren RS, Lothe RA, Guinney J. Relative contribution of clinicopathological variables, genomic markers, transcriptomic subtyping and microenvironment features for outcome prediction in stage II/III colorectal cancer. Ann Oncol. 2019;30:1622-9.
29. Lin WW, Karin M. A cytokine-mediated link between innate immunity, inflammation, and cancer. J Clin Invest. 2007;117:1175-83.
30. Pesce S, Greppi M, Tabellini G, Rampinelli F, Parolini S, Olive D, Moretta L, Moretta A, Marcenaro E. Identification of a subset of human natural killer cells expressing high levels of programmed death 1: A phenotypic and functional characterization. J Allergy Clin Immunol. 2017;139:335-46.

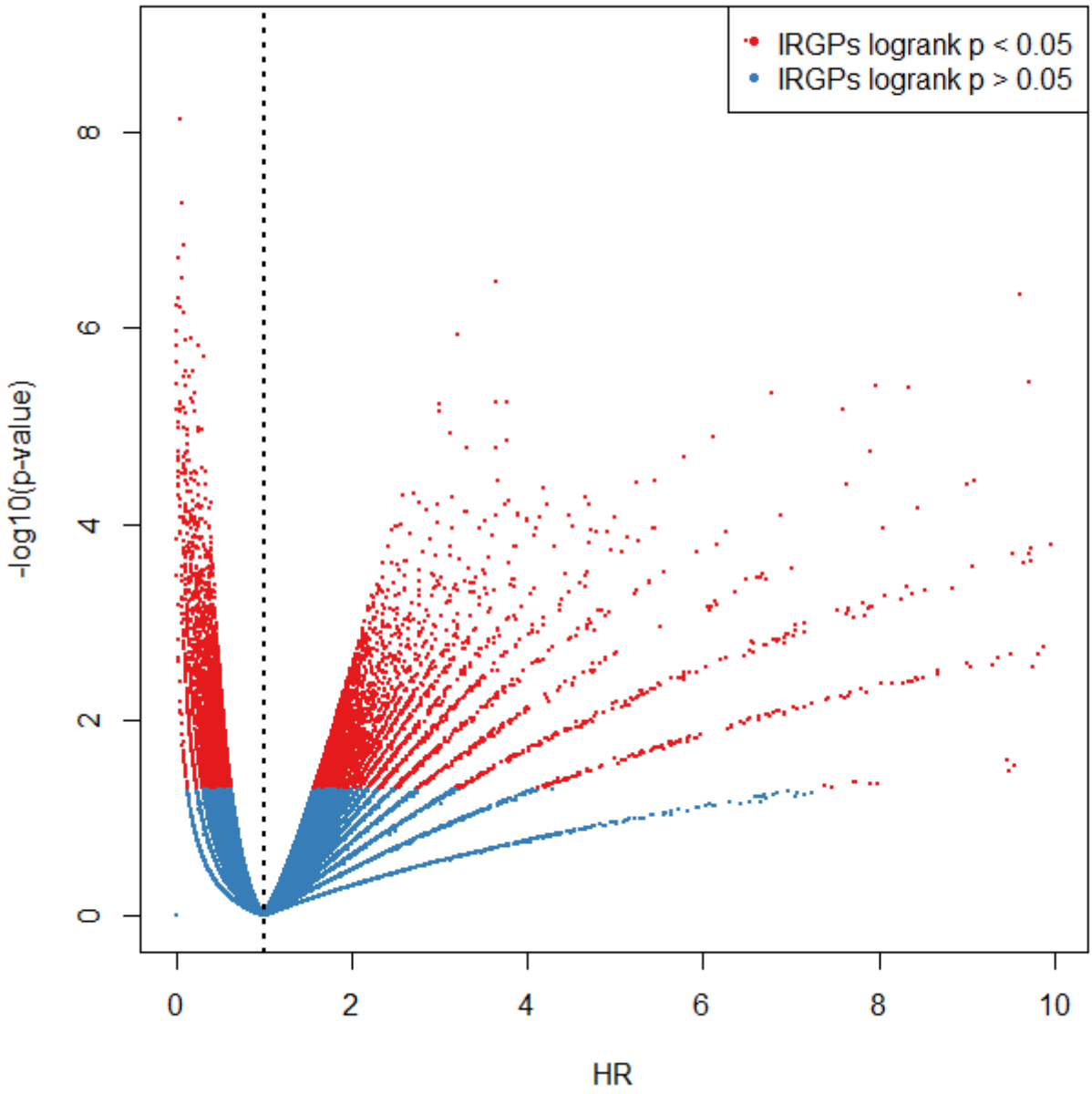
31. Liu R, Lu Z, Gu J, Liu J, Huang E, Liu X, Wang L, Yang J, Deng Y, Qian J, et al. MicroRNAs 15A and 16-1 activate signaling pathways that mediate chemotaxis of immune regulatory B cells to colorectal tumors. Gastroenterology. 2018;154:637-651.
32. Basseres DS, Baldwin AS. Nuclear factor-kappaB and inhibitor of kappaB kinase pathways in oncogenic initiation and progression. Oncogene 2006;25:6817-30.
33. Marrelli D, Morgagni P, de Manzoni G, Coniglio A, Marchet A, Saragoni L, Tiberio G, Roviello F, Italian Research Group for Gastric Cancer (IRGGC). Prognostic value of the 7th AJCC/UICC TNM classification of noncardia gastric cancer: analysis of a large series from specialized Western centers. Ann Surg. 2012;255:486-491.

# Figures



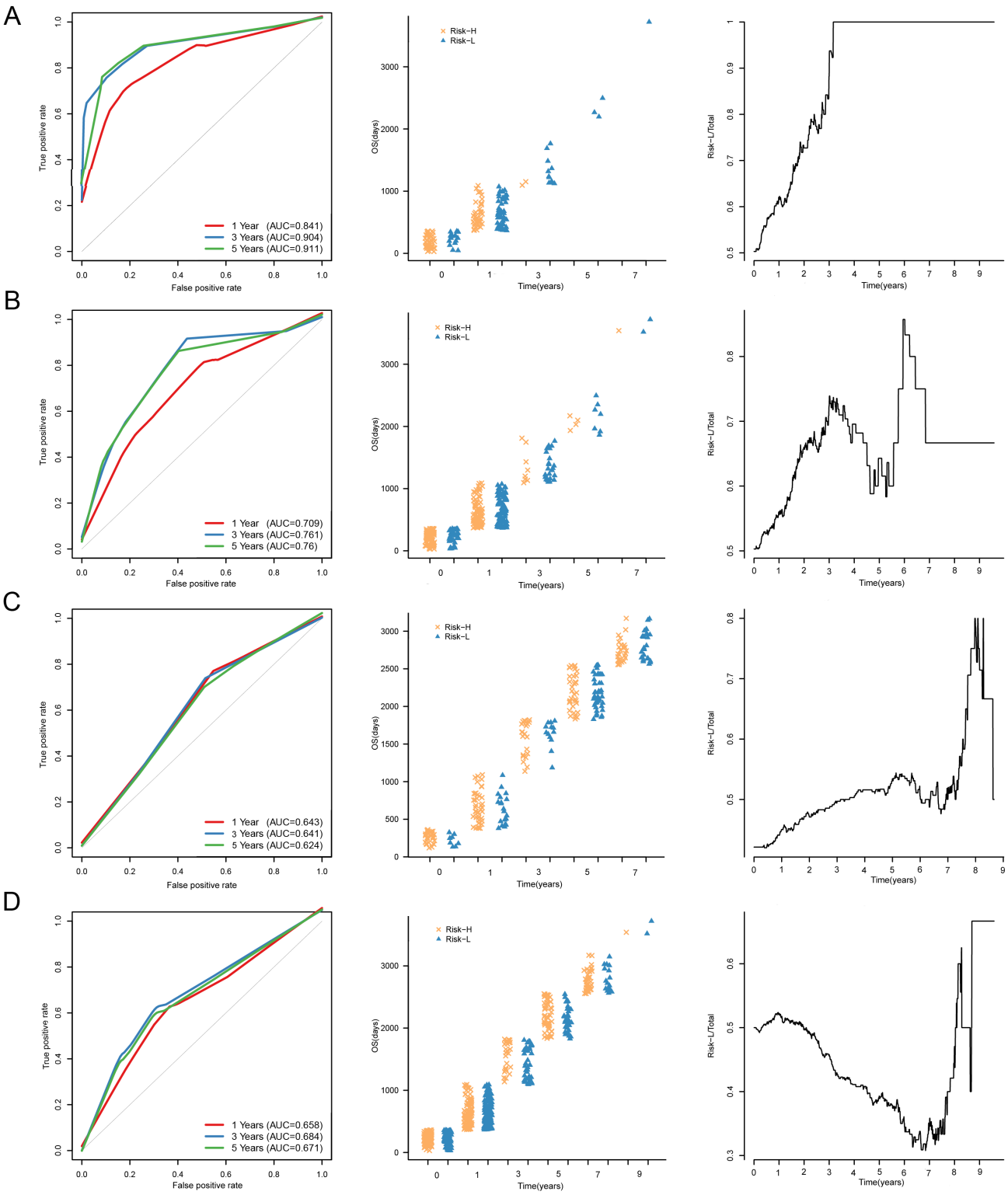

## Figure 1

Flowchart describing the procedure of developing and validating the prognostic values of immune-related gene-based model.



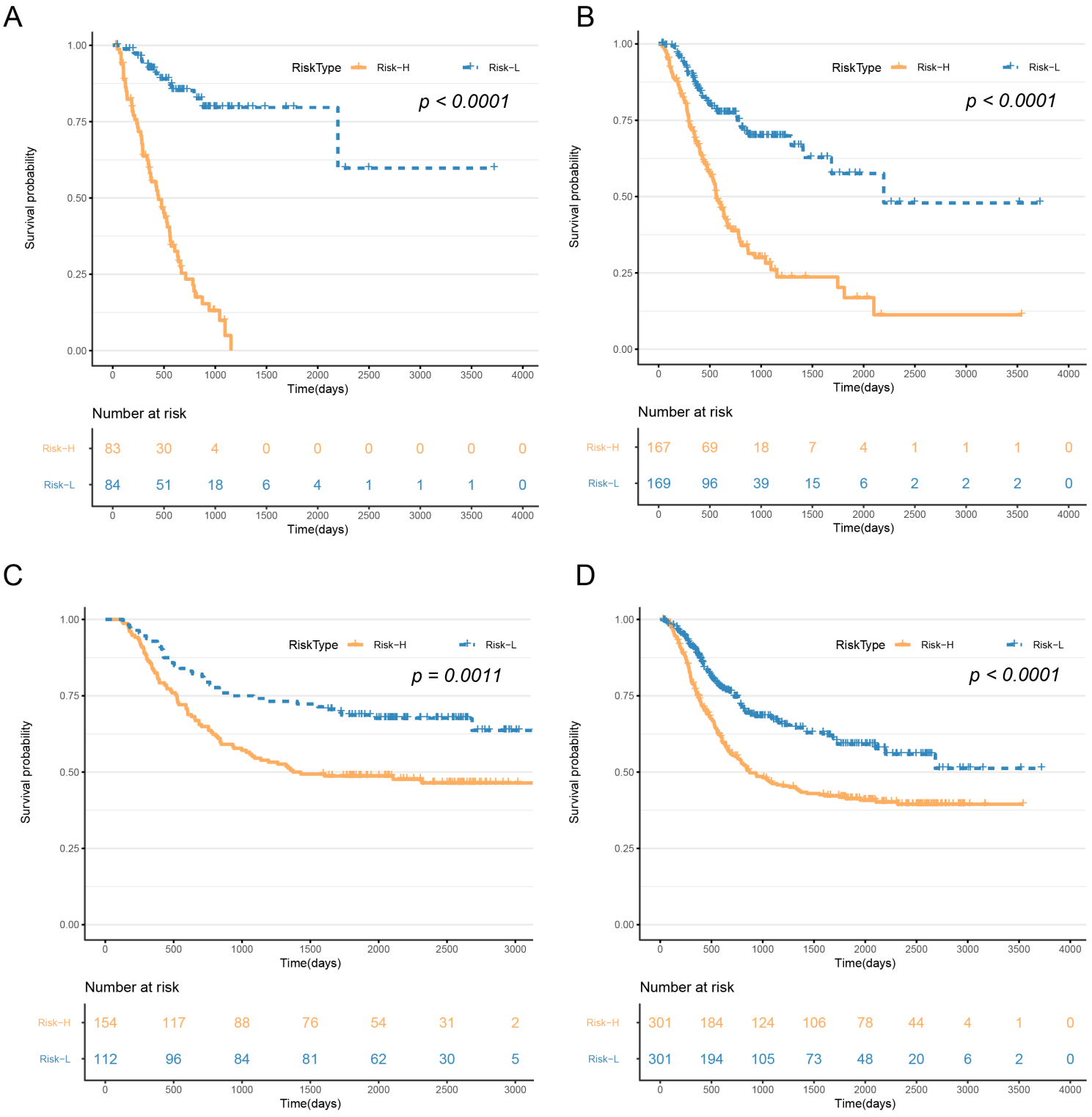

## Figure 2

Relationship between p value and HR of IRGPs in training set. Red dot represented 9620 genes with log (rank p)<0.05, which usually deviated from 1. HR, hazard ratio.



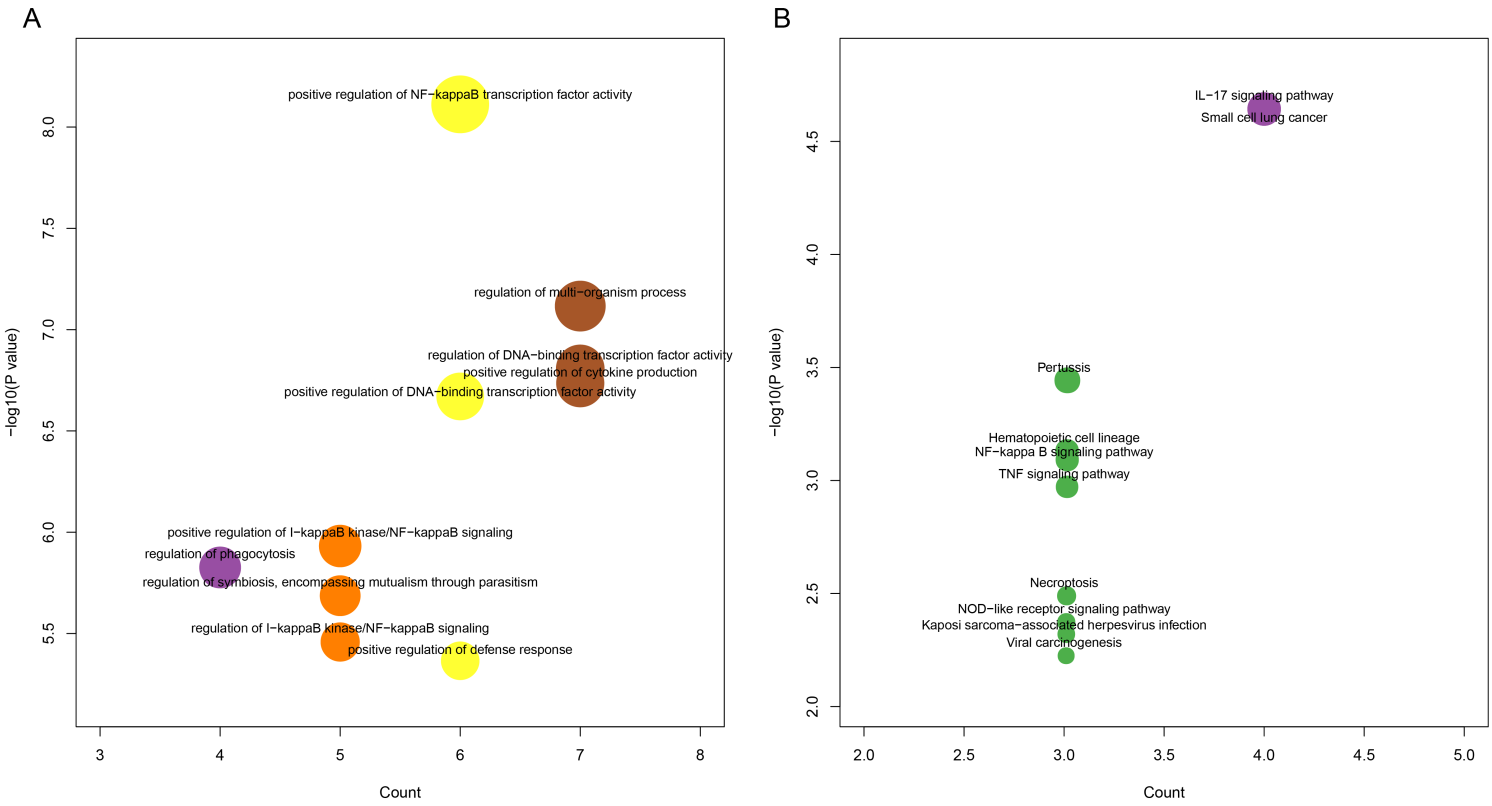

## Figure 3

Time-dependent ROC curve analysis of the immune-related prognostic model, sample distribution of Risk-H and Risk-L groups in different OS, proportion of Risk-L to total samples varying with different OS in the (A) training set, (B) validation set, (C) testing set, and (D) TCGA+GEO. OS, overall survival times.




## Figure 4

Kaplan-Meier survival curves for Risk-H and Risk-L groups based on the immune-related prognostic model on the (A) training set, (B) validation set, (C) testing set, and (D) TCGA+GEO. The curves show overall survival based on the relative high- and low-risk patients divided by the optimal cut-off point.




## Figure 5

(A) GO and (B) KEGG enrichment analysis for the 20 genes in the immune-related prognostic model. Circle size represented p value, and circle color represented gene count.

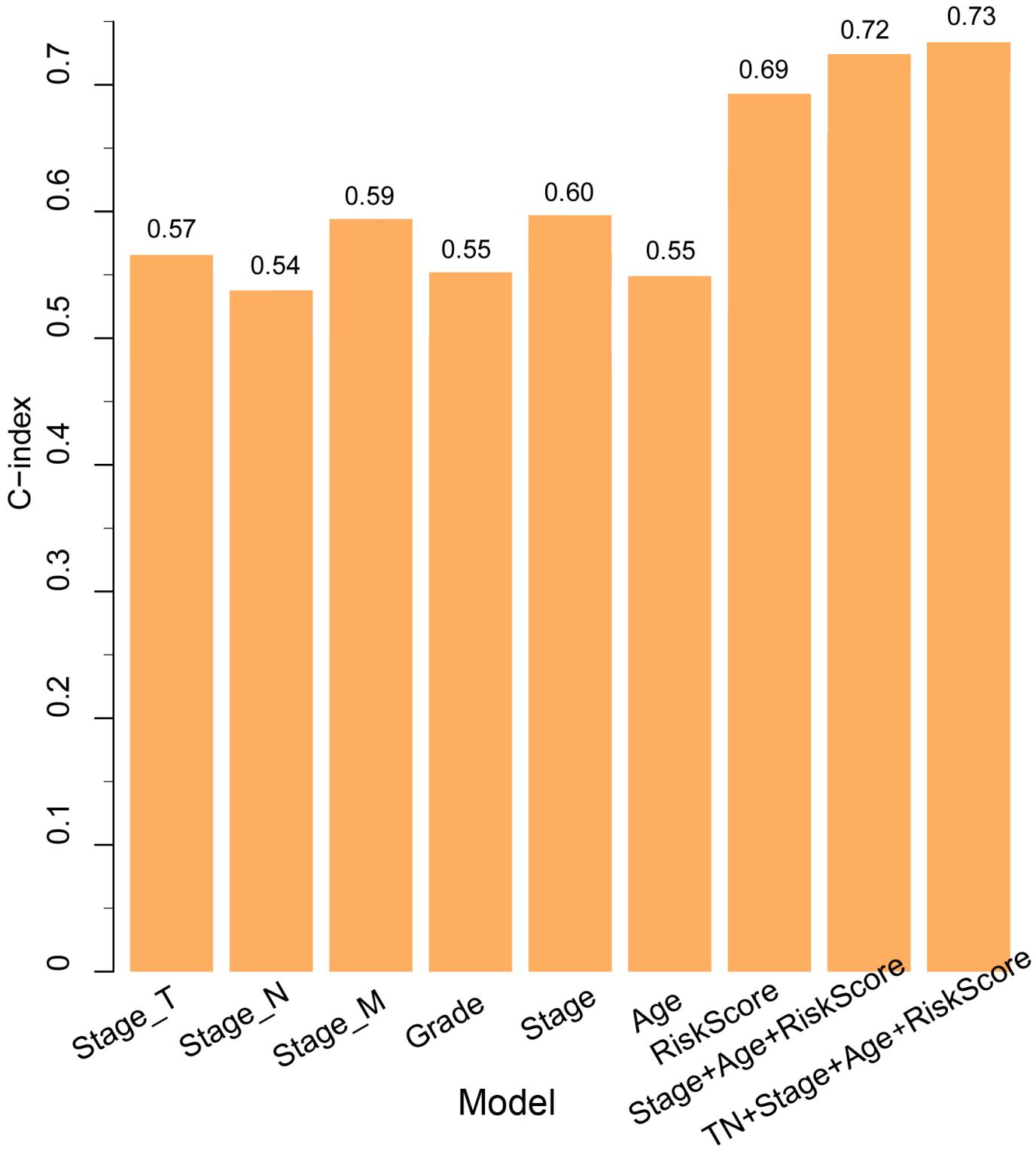



**Figure 6**

Comparision of the predictive accuracy of the immune-related prognostic model with clinical characteristics including TNM, grade, stage and age. TN+Stage+Age+RiskScore had the highest C-index (0.73) compared with immune-related prognostic model (0.69), grade (0.55) and stage (0.60).

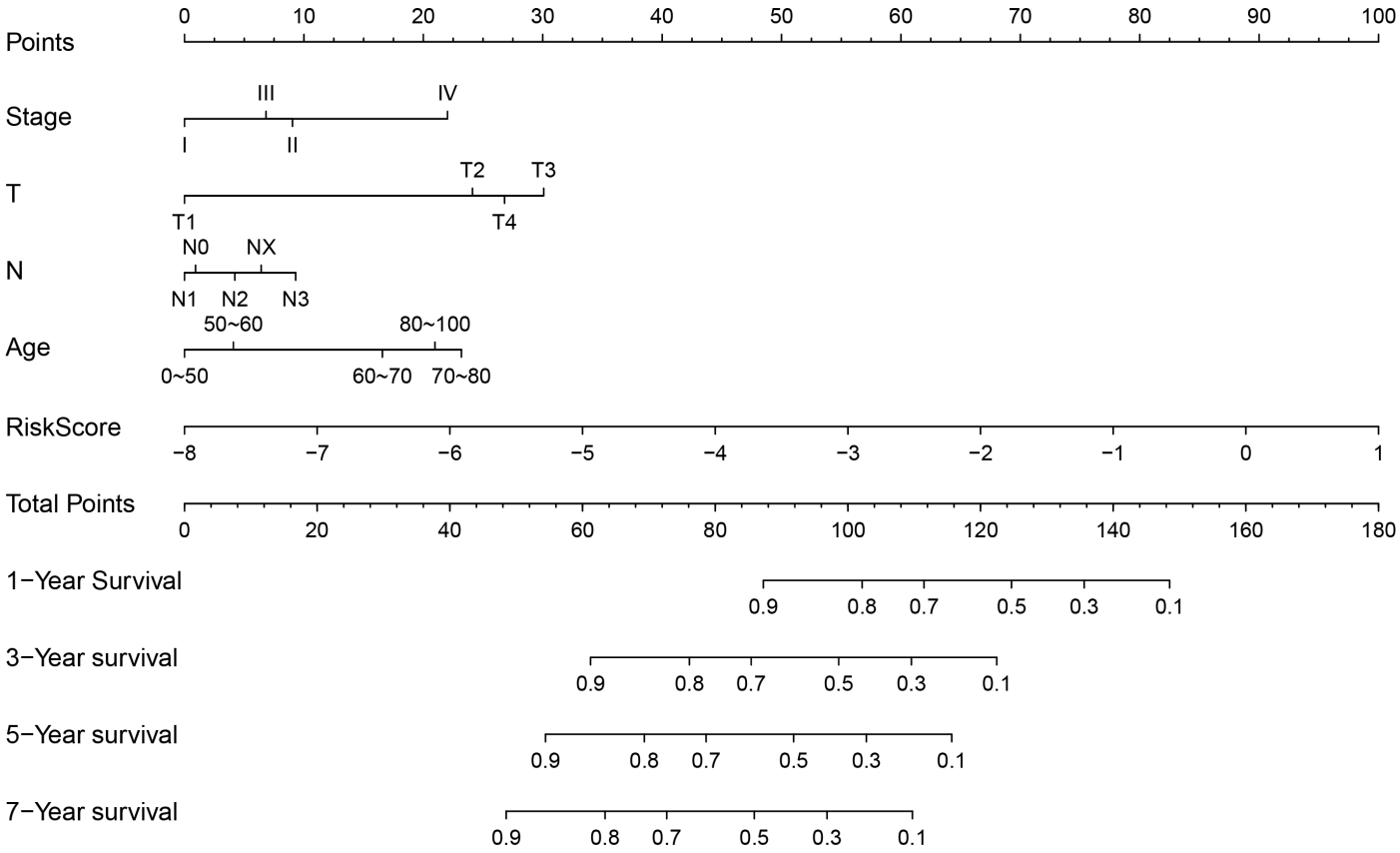



**Figure 7**

Nomogram predicting 1-, 3-, 5- and 7-year overall survival for patients with gastric cancer. The nomogram is assessed by adding up points identified on the point scales for each variable. The total points projected on the bottom scales indicate the probability of 1-, 3-, 5- and 7-year overall survival.

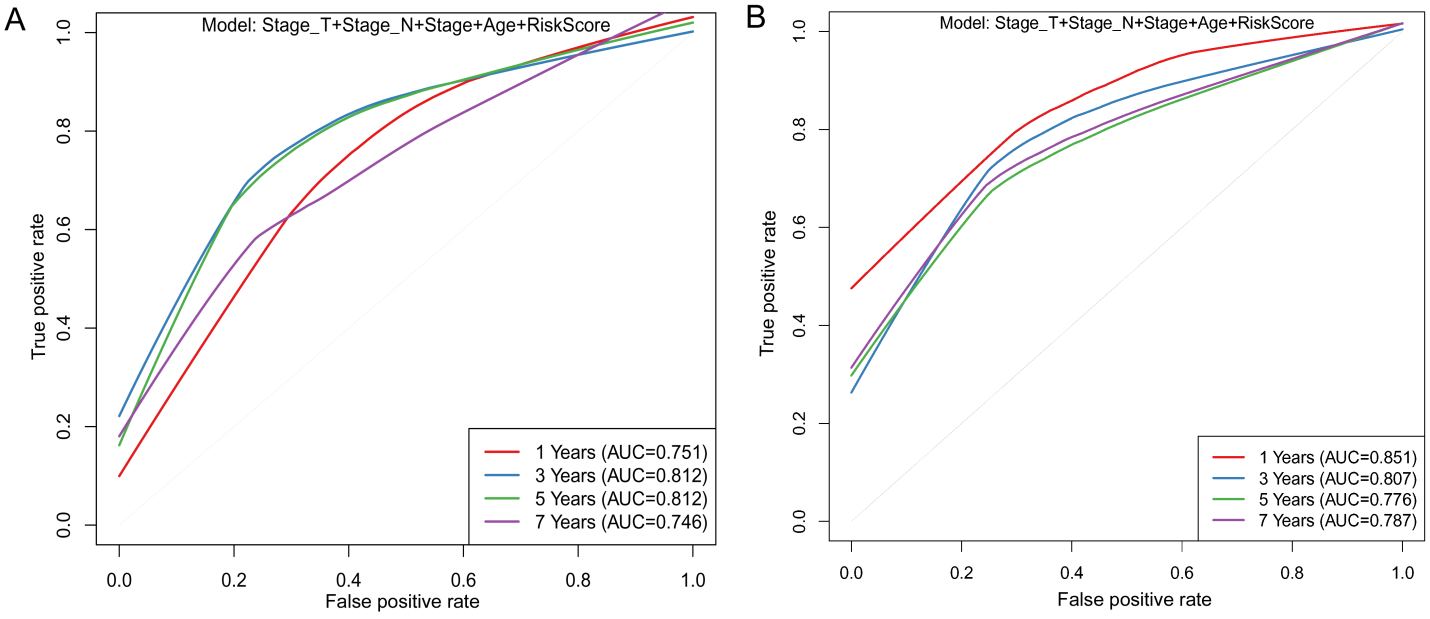



**Figure 8**

The time-dependent ROC curves for the prognostic model in the (A) validation set and (B) GSE62254 cohort. The ROC curves indicate relatively accuracy of the prognostic model.

## Supplementary Files

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

- SupplementaryFigure4.tif
- Additionalfile3.txt
- Additionalfile5.txt
- Additionalfile2.txt
- SupplementaryFigure1.tif
- Additionalfile4.txt
- SupplementaryFigure2.tif
- Additionalfile6.xlsx
- SupplementaryFigure3.tif
- Additionalfile1.txt