A Prognostic Nomogram Combining Gene Expression Profiles and TNM Stage Predicts Survival in Gastric Cancer

Jian Huang
National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital &Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College
https://orcid.org/0000-0002-0363-2412

Dongcun Wang
National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital and Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Xiaoliang Wang
National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital and Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Xiaoxing Ye
National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital and Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Jiping Da (✉ Jiping_da@aliyun.com)
National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital and Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Research

Keywords: gastric cancer, risk score signature, TNM stage, nomogram, overall survival rate

Posted Date: August 9th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-764284/v1

Abstract

Background

Gastric carcinoma (GC) is a highly aggressive malignancy and is associated with high morbidity and mortality rates around the world, the current tumor-node-metastasis (TNM) staging system is inadequate to predict overall survival (OS) in GC patients. therefore, potential forecasting methods for prognosis are important to investigate.

Methods

Differentially expressed genes (DEGs) were screened using gene expression data from The Cancer Genome Atlas (TCGA). We then construct a risk score signature model by univariate Cox proportional hazards regression (CPHR) analysis, the Kaplan-Meier method☒KM☒and multivariate CPHR analysis. Using TNM stage, we developed a signature-based nomogram. Finally, we utilize an independent Gene Expression Omnibus dataset (GSE62254) validate the prognostic value of risk score signature model and nomogram.

Results

We identified five OS-related mRNAs among 1113 mRNAs that were differentially expressed between GC and normal samples in the TCGA dataset. We then constructed a five-mRNA signature model, which efficiently distinguished high-risk from low-risk patient in both cohort, and even viable in the TNM stage-Ⅲ, gender(male, female) and age(☒65-year-old, ≥65-year-old) subgroups (*P*<0.05). Utilizing TNM stage, we developed a signature-based nomogram, which performed better than use the TNM stage or five-mRNA signature alone for prognostic prediction in the TCGA and GSE62254 dataset.

Conclusions

These results suggest that both risk signature and nomogram were effective prognostic indicators for patients with GCs, and could potentially be used for individualized management of such patients.

Introduction

GC is the fifth leading cause of cancer and the third leading cause of cancer related deaths worldwide, accounting for up to 7% of cancer occurrences and 9% of deaths[1-3]. Patients with GC are rarely diagnosed at an early stage, and the prognosis of patients with late-stage GC remains extremely poor regarding recurrence and metastasis; >70% of patients eventually die from this disease[4,5]. Pretreatment evaluation may help identify patients with GC at high risk for recurrence, which may guide the future development of targeted treatment strategies and reduce mortality and recurrence rates. Therefore, it is essential to seek prognostic indicators and more effectively to screen the prognostic factors of patients with GC.

In recent years, the analysis of biological information, also known as bioinformatics, has attracted a great deal of attention and sustained breakthroughs in the search for oncogenic genes[6-10]. Various functions for molecular typing, prognostic prediction, new targeted drug development applications and biomarkers of prognosis have been confirmed[3,11-14]. Thus, we used bioinformatics to identify genes predictive of GC prognosis. In this study, we utilized a rigorous computational framework to mine mRNA expression profiles and clinical data from The Cancer Genome Atlas stomach adenocarcinoma (TCGA-STAD Project'). Then, we constructed a risk signature of mRNAs and a nomogram integrating the signature with clinical features, and the predictive performances of the signature and the nomogram were validated in GSE62254 dataset.

Methods

Data source and pre-processing

The raw counts of the RNA expression profiles and the clinical data for 375 GC patients and 32 normal control patients from the publicly available TCGA-STAD Project were downloaded directly from the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/, updated until March 31, 2021). All expression profiles were obtained as HT-seq raw read counts and were annotated with the Ensemble reference database(ftp://ftp.ensemble.org/pub/release-93/gtf/homo_sapiens). The mRNA expression profiles were normalized and variance stabilizing transformation was performed with the “DESeq2” package in R software. The present study was conducted in accordance with the publication guidelines and data access policies of TCGA (http://cancergenome.nih.gov/publications/publicationguidelines).

The GSE62254 microarray expression data were downloaded from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/), which contained data on 300 patients with GC and their associated clinical information.

Screening of differentially expressed mRNAs

Differentially expressed mRNAs (DEMs) between GC samples and normal control samples were detected with the “limma” package use R software in the TCGA dataset. We defined mRNAs with adjusted *P* values <0.01 and log2|fold change| values >2. Volcano plots and heatmaps were visualized with the “ggplot2” and “pheatmap” packages of R software, respectively.

Identification of OS-related mRNAs in GC patients

To identify prognostic mRNAs, we removed patients without accurate survival data, such as survival for less than 0 days. The association between DEM expression and OS was evaluated by univariate Cox proportional hazards regression (CPHR) analysis and the Kaplan-Meier method. Only DEM with logical consistency between their expression and prognostic effects were considered as candidate OS-related mRNAs. After excluding patients without defined clinical characteristics, the important clinicopathological characteristics of the patients in the TCGA and GSE62254 dataset are shown in Table 1. In the TCGA dataset, the candidate OS-related mRNAs were selected for multivariate CPHR analysis by R software. To optimize the fitting accuracy comprehensively with a moderate number of parameters, we computed the Akaike information criterion (AIC) and used it to estimate the relative quality of the statistical models for the given set of data. The best-fit predictive model with the lowest AIC was chosen.

Identification and assessment of the five-mRNA signature

After choosing the best-fit OS-related mRNAs through the above steps, we performed a multivariate CPHR analysis to calculate the coefficient of each mRNA in the TCGA dataset. We thereby constructed a risk score formula, weighted by the linear combination of the expression values of the best-fit OS-related mRNAs and their corresponding estimated regression coefficients. The formula is as follows: risk score = $X_1\alpha_1 + X_2\alpha_2 + X_3\alpha_3 +\ldots+ X_n\alpha_n$. Where X is mRNA expression level and α is the corresponding estimated regression coefficient from the multivariate CPHR analysis, the same formula was used to calculate the risk score of GSE62254 dataset. Using the median risk score of the TCGA dataset as the cut-off value, we divided patients into high-risk and low-risk groups. The Kaplan-Meier method and log-rank test were performed to assess the survival differences between the high-risk and low-risk groups in each dataset. Additionally, a stratified analysis was conducted to assess whether the association of the five-mRNA signature with OS was independent of the TNM stage and other clinical risk factors. To further evaluate the prognostic performance of the mRNA-based classifier, we plotted time-dependent receiver operating characteristic (ROC) curves and calculated the area under the time-dependent ROC curve (AUC) values in both datasets, with three and five years as the defining points.

Development of the mRNA signature-based prognostic nomogram

To identify independent predictors of OS, we tested conventional clinical risk factors and the mRNA-based signature through univariate and multivariate CPHR analyses of the TCGA and GSE62254 dataset. A prognostic nomogram was then established with the package “rms” in R software. Calibration curves were conducted to assess whether the predicted survival in the nomogram was agreement with the actual survival. The predictive performances of the prognostic model were also evaluated using AUC in the ROC analysis and concordance index (C-index) calculation.

Statistical analysis

R (https://www.r-project.org/) was used as the main tool for data analysis and mapping. Univariate CPHR analysis and the Kaplan-Meier method were used to obtain candidate OS-related mRNAs. Multivariate CPHR analysis was then performed to screen variables and determine the risk score formula. OS was analyzed using Kaplan-Meier survival curve analysis and the log-rank test. A time-dependent ROC curve was used to assess the specificity and sensitivity of the prognostic prediction at each time point. The nomogram incorporating both the mRNA signature and independent clinical risk factors was developed through a multivariate CPHR analysis and was validated with the C-index and calibration curves.

Results

Candidate OS-related mRNAs from GC patients

The overall design and flowchart of this study is presented in Figure 1. In total, 375 GC patients from the TCGA database were included, we compared mRNA expression profiles of the 375 GC samples with those of 32 normal samples. We identified 1113 differentially expressed mRNAs (DEMs) with a log2|fold change| >2 and an adjusted *P* value <0.01. Of the 1113 DEMs, 822 mRNAs were found to be upregulated and 291 were found to be downregulated in the GC patients. The volcano plots and heatmaps of DEMs were visualized with the “ggplot2”and “pheatmap” packages of R software, and are shown in Figure 2.

After the exclusion of 43 patients with insufficient survival data or availability of clinical data, 332 GC patients remained in our study. All 1113 DEMs were subjected to univariate CPHR analysis (P<0.01) and Kaplan-Meier analysis (P<0.05), with OS as the dependent variable and the mRNA level as the explanatory variable. As shown in Supplementary Table 1, 11 mRNAs were significantly associated with the OS of GC patients. Six of these 11 mRNAs (SERPINE1，MATN3，COL10A1，IGFBP1，CST2，VCAN) had hazard ratios (HRs) greater than 1, suggesting that their overexpression was associated with shorter OS. On the other hand, the HR of five mRNAs (PRKACG，MTBP，ARHGEF38，RAD54L，HSD17B3) was less than 1, with the opposite implications. The Kaplan-Meier analysis curves were consistent with the univariate CPHR analysis results (Supplementary Figure 1). Thus, we considered these dysregulated mRNAs as candidate OS-related mRNAs.

Identification and validation of a five-mRNA signature for survival prediction

To identify the best-fit OS-related mRNAs, we filtered these candidate mRNAs through a multivariate CPHR analysis (stepwise model). We used the AIC to avoid over-fitting. The five OS-related mRNAs with the largest likelihood ratios and lowest AIC values (MATN3, HSD17B3, PRKACG, SERPINE1 and IGFBP1) were selected from the stepwise model (Table 2) and integrated into a predictive signature based on their risk coefficients. The formula was as follows: Risk Score = $(0.338\times\text{Expression}_{\text{MATN3}})$ + $(-1.077 \times \text{Expression}_{\text{HSD7B3}})$ + $(-6.063 \times \text{Expression}_{\text{PRKACG}})$ + $(0.207\times \text{Expression}_{\text{SERPINE1}})$ + $(0.271\times \text{Expression}_{\text{IGFBP1}})$. Then, we calculated the five-mRNA-based risk score for each GC patient in the TCGA dataset. Using the median risk score as the cut-off value, we divided patients into high-risk and low-risk groups. The distributions of the OS statuses, mRNA-based risk scores and five mRNA expression profiles in the TCGA dataset are shown in Figure 3A. The pheatmap showed that 3 DEM (MATN3, SERPINE1 and IGFBP1) were expressed at higher levels in the high-risk group than in the low-risk group, and 2 DEM (HSD17B3 and PRKACG) was overexpressed in the low-risk group. Kaplan-Meier curve analysis clearly demonstrated that the high-risk group had a poorer prognosis than the low-risk group (*P*=1.867e-04) (Figure 3B). Subsequently, we constructed a time-dependent ROC curve with the TCGA dataset. As shown in Figure 3C, the AUC of the five-mRNA signature reached 0.700 at three years and 0.788 at five years (Figure 3C).

The performance of the five-mRNA signature for predicting survival was then validated with the GSE62254 dataset (n=300). When we used the five-mRNA signature and cut-off value derived from the TCGA dataset, the distributions of the OS statuses, five-mRNA-based risk scores and five-mRNA expression profiles in GSE62254 dataset were consistent with the findings described above (Figure 4A). Similar to the results in the TCGA dataset, a Kaplan-Meier curve analysis indicated that the survival time of GC patients was significantly shorter in the high-risk group (n=124) than in the low-risk group (n=176) (*P*=8.014e-03) (Figure 4B). The AUC of the five-mRNA signature was 0.600 at three years and 0.583 at five years (Figure 4C). Thus, the five-mRNA signature was great predictive value of GC patient.

The prognostic value of the five-mRNA signature was independent from those of conventional clinical risk factors

Next, we tested whether the prognostic performance of the five-mRNA signature was independent from those of conventional clinical risk factors. A multivariate CPHR analysis demonstrated that the HR of a high *vs.* low-risk score was 1.761 (*P* <0.001) in the TCGA dataset and 1.325 (*P*<0.001) in the GSE62254 dataset (Table 3 and Supplementary Table 2), indicating that the five-mRNA signature could independently predict the prognoses of GC patients.

Considering the number of GC patients, we combined the GC patients of TCGA and GSE62254 dataset for performed a risk-stratified analysis. The 632 GC patients were stratified into a stage-I subgroup (n=75), stage-II subgroup (n=206), stage-III subgroup (n=240) and stage-IV subgroup (n=111) based on their TNM stage. each subgroup was divided into a high-risk group and a low-risk group based on the risk scores proposed above. We found that the classification efficiency of the five-mRNA signature was limited when it was applied to certain subgroups. As shown in the Kaplan-Meier curves, for the stage-III, patients in the high-risk group had significantly poorer survival than those in the low-risk group (*P*=4.937e-03, log-rank test) (Figure 5E). However, the five-mRNA signature did not reach the threshold of significance in the stage-I, stage-II and stage-IV subgroup (Figure 5C, 5D, 5F). When a stratified analysis was carried out based on age (≥65-year-old or <65-year-old) and gender (Male or Female), all subgroup did the five-mRNA signature subdivide patients into a high-risk group and a low-risk group with significantly different survival (female subgroup, *P*=1.611e-04; male subgroup, *P*=2.775e-03; ≥65-year-old subgroup, *P*=4.19e-06; <65-year-old subgroup, *P*=1.69e-02) (Figure 5A, 5B, 5G and 5H), and there is no difference in the survival time of male and female in High risk and Low risk subgroup (Figure 5I and 5J). Thus, although the five-mRNA signature could be viewed as an independent prognostic predictor for GC patients, its performance was limited to specific subgroups.

Development of a nomogram combining the five-mRNA signature with TNM stage

Clinical risk factors such as the TNM stage are still vital predictors of OS in GC patients[9]. Therefore, we integrated these traditional risk factors with our five-mRNA signature to develop an efficient quantitative method of predicting OS. we evaluated the prognostic value of several clinical risk factors in univariate and multivariate CPHR analyses of the TCGA and GSE62254 dataset. We found that, in addition to the five-mRNA signature, age (≥65 *vs.* <65) and TNM stage (III-IV *vs.* I-II) were significantly associated with OS (*P*<0.05) (Table 3 and Supplementary Table 2). Ultimately, we selected 5-mRNA signature and TNM stage for multivariate CPHR analysis, and integrated them into nomogram. We then used this nomogram to predict the three-year and five-year survival of GC patients (Figure 6A). As shown in the nomogram, the five-mRNA signature contributed the most to the three- and five-year OS, followed closely by the TNM stage. This user-friendly graphical tool allowed us to determine the three- and five-year OS probability for each GC patient easily.

We then evaluated the discrimination and calibration abilities of the prognostic nomogram by using a C-index and calibration plots. A validation using a bootstrap with 1000 resamplings revealed that the nomogram performed well for discrimination: The C-index was 0.686 for the TCGA dataset, and 0.711 for GSE62254 dataset. The three-year and five-year OS probabilities generated by the nomogram were plotted against the observed outcomes, as shown in Figure 6B–6E.

We further assessed the prognostic performance of the nomogram in a time-dependent ROC curve analysis. The AUC of the nomogram was 0.704 at three years and 0.708 at five years in the TCGA dataset (Figure 7A). In the GSE62254 dataset, the AUC was 0.760 at three years and 0.752 at five years (Figure 7B).

Survival prediction power: comparison of the five-mRNA signature-based nomogram and other risk factors

To compare the predictive sensitivities and specificities of different prognostic factors, we used time-dependent ROC curves. As shown in Figure 7C, the predictive performance of the five-mRNA-based nomogram (AUC=0.704) was superior to the performance of the five-mRNA signature (AUC=0.700) and the TNM stage (AUC=0.595) in the TCGA dataset. In the GSE62254 dataset, the predictive value of the five-mRNA-based nomogram (AUC=0.760) was better than the five-mRNA signature (AUC=0.600) and the TNM stage (AUC=0.744) also (Figure 7D). Thus, the newly developed prognostic nomogram concentrated the advantages of the five-mRNA signature and TNM stage, improving their prognostic predictive efficiency for GC patients.

Discussion

Gastric cancer is a malignant cancer with a high mortality rate. Most patients with GC are diagnosed at the terminal stage of the disease due to its nonspecific clinical symptoms in the early stages, which also creates a challenge for treatment[15]. Prognostic prediction for GC patients largely relies on American Joint Committee on Cancer/Union for International Cancer Control (AJCC/UICC) on Cancer TNM staging system at daily clinical practice[16-18]. However, patients with similar TNM stage can exhibit variable responses to therapy. A series of genomic landscape discoveries have demonstrated that this phenomenon may be due to tumor heterogeneity and genomic heterogeneity[19,20]. Thus, a reliable prognostic model for GC is urgently required in the era of precision medicine.

Several studies have been found that mRNA involved in cycle regulation, cell adhesion, angiogenesis, and tumor carcinogenesis have been reported to play a crucial role in forecasting survival outcome of GC patients[21-24]. Other studies also successfully identified several lncRNA-, miRNA-, and mRNA-expression based risk signatures to prevent the development of various tumors[25-28]. In the present study, based on public high-throughput mRNA expression profiles and clinical data from TCGA-STAD Project, we discovered a novel five-mRNA signature that could effectively identify high-risk GC patients, and validation that patients with GC in the high-risk group had a shorter survival than those in the low-risk group. At the same time, the signature also showed a high prediction accuracy and robustness on calculating AUC in ROC analysis.

As interest in personalized medicine has grown, a few prognostic risk classifiers have been identified and found to enhance survival predictions in a variety of cancers[29-35]. However, most of these studies have focused only on statistical power in the screening of molecular markers, without regard for their clinical significance, in our study, we combined the TNM stage with molecular profiling. Ultimately, we constructed a five-mRNA signature-based nomogram to quantify an individual's probability of OS. The predictive performance of our proposed prognostic nomogram was superior to those of the five-mRNA signature or the traditional TNM stage alone. This objective probability scale should be simple for patients and clinicians to understand and use in clinical practice[36].

The most important convenience of our nomogram is its simplicity. Prognostic models are designed to identify the associations between risk factors and outcomes based on essential features, and should be accurate and parsimonious[14,37]. Our five-mRNA signature-based nomogram relies on routinely available variables, including genetic differences (the five-mRNA signature) and a histopathological characteristic (TNM stage). Therefore, clinicians can easily estimate outcomes and make clinical decisions for individual GC patients.

The most attractive biomarkers for clinical applications are those that provide accurate prognoses for patients, stratify patients into different risk groups and thus help clinicians choose the most effective treatment. In our study, the predictive capacity of our five-mRNA signature was independent from those of conventional clinical factors including age, TNM stage, lymph node metastasis and distant metastasis. In our stratified analysis, the five-lncRNA signature performed well for risk stratification in the male, female, stage-III, <65-year-old and ≥65-year-old subgroups. Notably, however, its application was limited in the stage-I, II and IV.

Although our newly proposed prognostic nomogram performed well in predicting survival for GC patients, the limitations of this study should be noted. One hand, the database of the TCGA and GEO databases lacks certain important pre- and postoperative parameters (e.g., chemotherapy, radiotherapy, immunotherapy), so we could not carry out a comprehensive survival analysis with these potential factors. On the other hand, we used data from an open-access published database and therefore our study design was retrospective. We are actively gathering samples and corresponding clinical data from a large number of GC patients to further validate our findings and to determine whether our nomogram improves the prediction of accuracy.

Conclusion

we successfully constructed a five-mRNA risk signature correlated with GC prognosis in the TCGA and GSE62254 cohort. The results indicated that the signature is a potent predictive indicator for patients with GC. Furthermore, we identified and validated a novel and robust nomogram incorporating the signature and clinical factors to predict the 3- and 5-year OS rates of patients with GC, which could aid in the individualized management of GC patients in the future.

Abbreviations

GC: Gastric carcinoma; TNM: Tumor-node-metastasis; OS: Overall survival; DEGs: Differentially expressed genes; TCGA: The Cancer Genome Atlas; CPHR: Univariate Cox proportional hazards regression; KM: the Kaplan-Meier method; GEO: Gene Expression Omnibus; DEMs: Differentially expressed mRNAs; AIC: Akaike information criterion; ROC: receiver operating characteristic; AUC: the area under the time-dependent ROC curve; C-index: concordance index; HRs: hazard ratios; AJCC: American Joint Committee on Cancer; UICC: Union for International Cancer Control

Declarations

Acknowledgements

This study used TCGA and GEO databases. We appreciate the data platform, and the authors have uploaded their data.

Authors’ contributions

All authors have made important contributions to data analysis, drafting the article or revising the article. The authors read and approved the final manuscript.

Funding

Supported by Sanming Project of Medicine in Shenzhen, Grant/Award Number: SZSM20181207

Competing of interests

The authors declare that there are no conflicts of interest.

Consent for publication

All authors have agreed to the publication of the work.

References

1. Zhang Z, Pi J, Zou D, et al. microRNA arm-imbalance in part from complementary targets mediated decay promotes gastric cancer progression. *Nat Commun.* 2019;10(1):4397.

2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015;65(2):87-108.

3. Cristescu R, Lee J, Nebozhyn M, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med.* 2015;21(5):449-456.

4. Allemani C, Weir HK, Carreira H, et al. Global surveillance of cancer survival 1995-2009: analysis of individual data for 25,676,887 patients from 279 population-based registries in 67 countries (CONCORD-2). *Lancet.* 2015;385(9972):977-1010.

5. Zhang E, He X, Zhang C, et al. A novel long noncoding RNA HOXC-AS3 mediates tumorigenesis of gastric cancer by binding to YBX1. *Genome Biol.* 2018;19(1):154.

6. Liao P, Li W, Liu R, et al. Genome-scale analysis identifies SERPINE1 and SPARC as diagnostic and prognostic biomarkers in gastric cancer. *Onco Targets Ther.* 2018;11:6969-6980.

7. Li L, Zhu Z, Zhao Y, et al. FN1, SPARC, and SERPINE1 are highly expressed and significantly related to a poor prognosis of gastric adenocarcinoma revealed by microarray and bioinformatics. *Sci Rep.* 2019;9(1):7827.

8. Cheng P. A prognostic 3-long noncoding RNA signature for patients with gastric cancer. *J Cell Biochem.* 2018;119(11):9261-9269.

9. Cui J, Wen Q, Tan X, Chen Z, Liu G. A Genomic-Clinicopathologic Nomogram Predicts Survival for Patients with Laryngeal Squamous Cell Carcinoma. *Dis Markers.* 2019;2019:5980567.

10. Fan F, Zhang H, Dai Z, et al. A comprehensive prognostic signature for glioblastoma patients based on transcriptomics and single cell sequencing. *Cell Oncol (Dordr).* 2021.

11. Wen P, Chidanguro T, Shi Z, et al. Identification of candidate biomarkers and pathways associated with SCLC by bioinformatics analysis. *Mol Med Rep.* 2018;18(2):1538-1550.

12. Xu Z, Zhou Y, Cao Y, Dinh TL, Wan J, Zhao M. Identification of candidate biomarkers and analysis of prognostic values in ovarian cancer by integrated bioinformatics analysis. *Med Oncol.* 2016;33(11):130.

13. Zhou L, Tang H, Wang F, et al. Bioinformatics analyses of significant genes, related pathways and candidate prognostic biomarkers in glioblastoma. *Mol Med Rep.* 2018;18(5):4185-4196.

14. Wang Y, Du L, Yang X, et al. A nomogram combining long non-coding RNA expression profiles and clinical factors predicts survival in patients with bladder cancer. *Aging (Albany NY).* 2020;12(3):2857-2879.

15. Kim GH, Bang SJ, Ende AR, Hwang JH. Is screening and surveillance for early detection of gastric cancer needed in Korean Americans? *Korean J Intern Med.* 2015;30(6):747-758.

16. Komatsu S, Otsuji E. Essential updates 2017/2018: Recent topics in the treatment and research of gastric cancer in Japan. *Ann Gastroenterol Surg.* 2019;3(6):581-591.

17. Deng J, Liu J, Wang W, et al. Validation of clinical significance of examined lymph node count for accurate prognostic evaluation of gastric cancer for the eighth edition of the American Joint Committee on Cancer (AJCC) TNM staging system. *Chin J Cancer Res.* 2018;30(5):477-491.

18. O'Sullivan B, Brierley J, Byrd D, et al. The TNM classification of malignant tumours-towards common understanding and reasonable expectations. *Lancet Oncol.* 2017;18(7):849-851.

19. Rocken C. Molecular classification of gastric cancer. *Expert Rev Mol Diagn.* 2017;17(3):293-301.

20. Sanjeevaiah A, Cheedella N, Hester C, Porembka MR. Gastric Cancer: Recent Molecular Classification Advances, Racial Disparity, and Management Implications. *J Oncol Pract.* 2018;14(4):217-224.

21. Akama Y, Yasui W, Yokozaki H, et al. Frequent amplification of the cyclin E gene in human gastric carcinomas. *Jpn J Cancer Res.* 1995;86(7):617-621.

22. Graziano F, Mandolesi A, Ruzzo A, et al. Predictive and prognostic role of E-cadherin protein expression in patients with advanced gastric carcinomas treated with palliative chemotherapy. *Tumour Biol.* 2004;25(3):106-110.

23. Tanigawa N, Amaya H, Matsumura M, Shimomatsuya T. Correlation between expression of vascular endothelial growth factor and tumor vascularity, and patient outcome in human gastric carcinoma. *J Clin Oncol.* 1997;15(2):826-832.

24. Sanz-Ortega J, Steinberg SM, Moro E, al. e. Comparative study of tumor angiogenesis and immunohistochemistry for p53, c-ErbB2, c-myc and EGFr as prognostic factors in gastric cancer. *Histol Histopathol* 2000;15:455-462.

25. Zhang Y, Han T, Li J, et al. Comprehensive analysis of the regulatory network of differentially expressed mRNAs, lncRNAs and circRNAs in gastric cancer. *Biomed Pharmacother.* 2020;122:109686.

26. Li Y, Weng Y, Pan Y, et al. A Novel Prognostic Signature Based on Metabolism-Related Genes to Predict Survival and Guide Personalized Treatment for Head and Neck Squamous Carcinoma. *Front Oncol.* 2021;11:685026.

27. Sun L, Li J, Li X, et al. A Combined RNA Signature Predicts Recurrence Risk of Stage I-IIIA Lung Squamous Cell Carcinoma. *Front Genet.* 2021;12:676464.

28. Liu J, Lu J, Li W. A Comprehensive Prognostic and Immunological Analysis of a New Three-Gene Signature in Hepatocellular Carcinoma. *Stem Cells Int.* 2021;2021:5546032.

29. Zhao QJ, Zhang J, Xu L, Liu FF. Identification of a five-long non-coding RNA signature to improve the prognosis prediction for patients with hepatocellular carcinoma. *World J Gastroenterol.* 2018;24(30):3426-3439.

30. Zhu X, Tian X, Yu C, et al. A long non-coding RNA signature to improve prognosis prediction of gastric cancer. *Mol Cancer.* 2016;15(1):60.

31. Lai J, Wang H, Pan Z, Su F. A novel six-microRNA-based model to improve prognosis prediction of breast cancer. *Aging (Albany NY).* 2019;11(2):649-662.

32. Li J, Chen Z, Tian L, et al. LncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma. *Gut.* 2014;63(11):1700-1710.

33. Tang J, Cui Q, Zhang D, Liao X, Zhu J, Wu G. An estrogen receptor (ER)-related signature in predicting prognosis of ER-positive breast cancer following endocrine treatment. *J Cell Mol Med.* 2019;23(8):4980-4990.

34. Sun J, Zhao H, Lin S, et al. Integrative analysis from multi-centre studies identifies a function-derived personalized multi-gene signature of outcome in colorectal cancer. *J Cell Mol Med.* 2019;23(8):5270-5281.

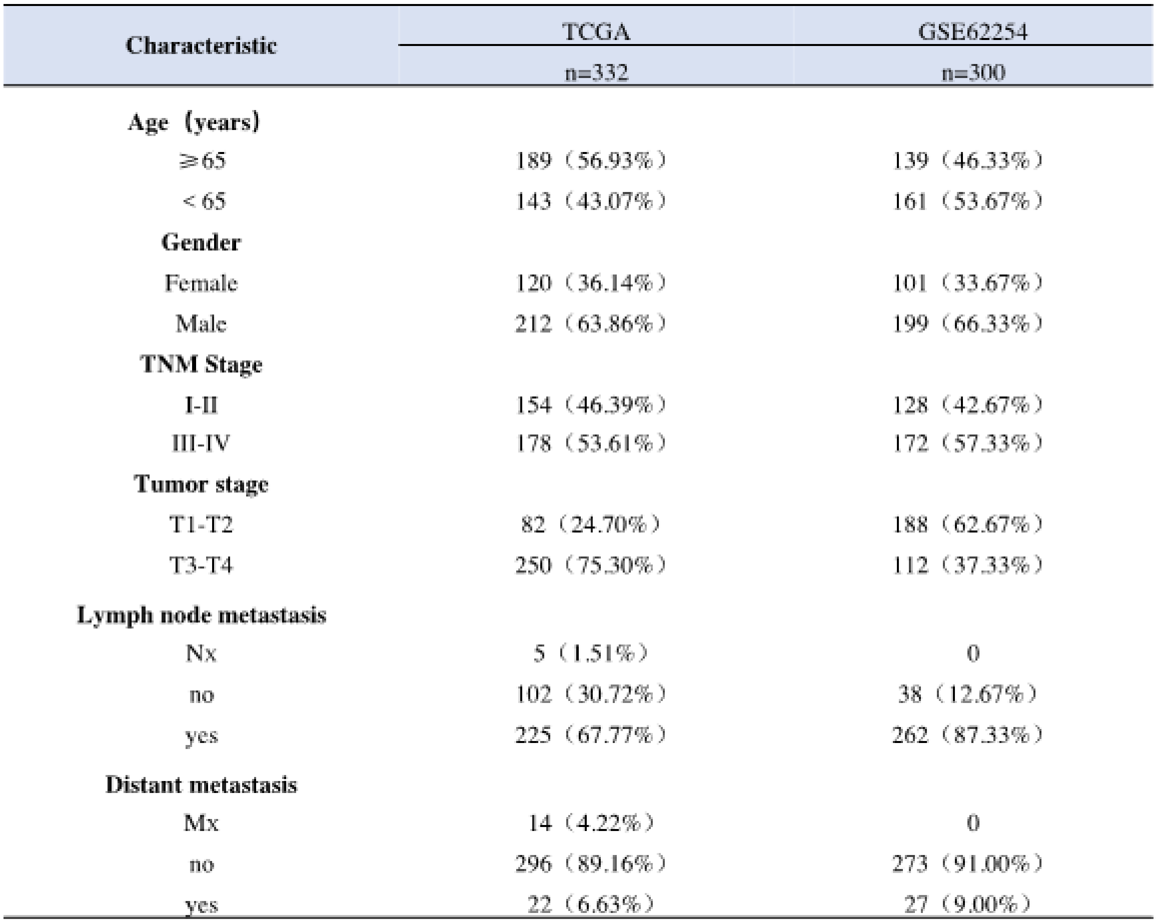
35. Yu P, Lan H, Song X, Pan Z. High Expression of the SH3TC2-DT/SH3TC2 Gene Pair Associated With FLT3 Mutation and Poor Survival in Acute Myeloid Leukemia: An Integrated TCGA Analysis. *Front Oncol.* 2020;10:829.

36. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *The Lancet Oncology.* 2015;16(4):e173-e180.

37. Zini L, Cloutier V, Isbarn H, et al. A simple and accurate model for prediction of cancer-specific mortality in patients treated with surgery for primary penile squamous cell carcinoma. *Clin Cancer Res.* 2009;15(3):1013-1018.

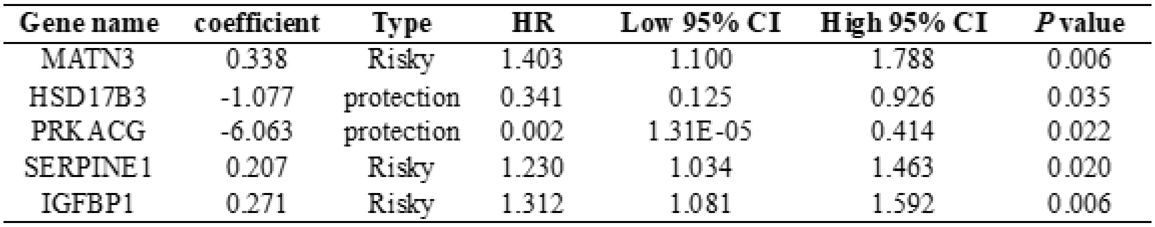
Tables

Table 1 Baseline clinical characteristics of gastric cancer cases involved in this study.



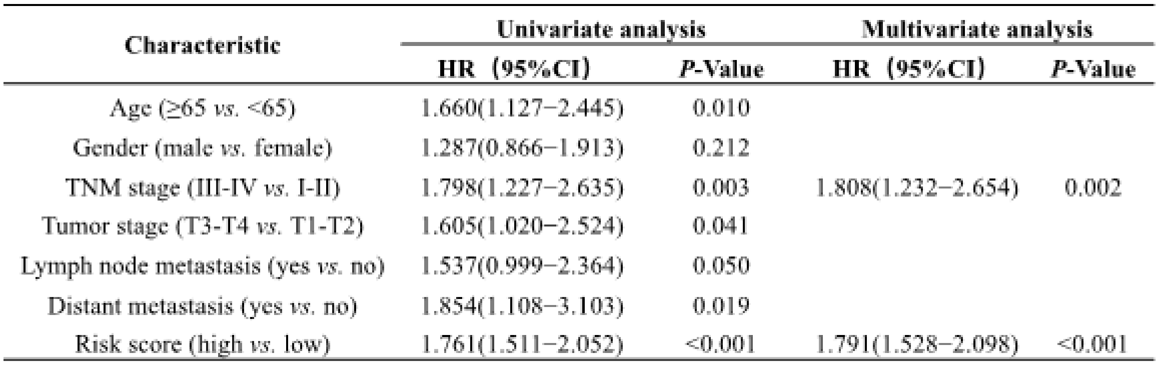
Characteristic	TCGA	GSE62254
	n=332	n=300
Age（years）		
≥65	189（56.93%）	139（46.33%）
< 65	143（43.07%）	161（53.67%）
Gender		
Female	120（36.14%）	101（33.67%）
Male	212（63.86%）	199（66.33%）
TNM Stage		
I-II	154（46.39%）	128（42.67%）
III-IV	178（53.61%）	172（57.33%）
Tumor stage		
T1-T2	82（24.70%）	188（62.67%）
T3-T4	250（75.30%）	112（37.33%）
Lymph node metastasis		
Nx	5（1.51%）	0
no	102（30.72%）	38（12.67%）
yes	225（67.77%）	262（87.33%）
Distant metastasis		
Mx	14（4.22%）	0
no	296（89.16%）	273（91.00%）
yes	22（6.63%）	27（9.00%）

Table 2 Five-mRNAs significantly associated with overall survival in the TCGA dataset. Abbreviations: HR, hazard ratio; CI, confidence interval.



Gene name	coefficient	Type	HR	Low 95% CI	High 95% CI	*P* value
MATN3	0.338	Risky	1.403	1.100	1.788	0.006
HSD17B3	-1.077	protection	0.341	0.125	0.926	0.035
PRKACG	-6.063	protection	0.002	1.31E-05	0.414	0.022
SERPINE1	0.207	Risky	1.230	1.034	1.463	0.020
IGFBP1	0.271	Risky	1.312	1.081	1.592	0.006

Table 3 Univariate and multivariate Cox proportional hazards regression analysis of five-mRNA signature and clinical risk factors in the TCGA dataset. Abbreviations: HR, hazard ratio; CI, confidence interval.



Characteristic	Univariate analysis		Multivariate analysis	
	HR（95%CI）	*P*-Value	HR（95%CI）	*P*-Value
Age (≥65 *vs.* <65)	1.660(1.127−2.445)	0.010		
Gender (male *vs.* female)	1.287(0.866−1.913)	0.212		
TNM stage (III-IV *vs.* I-II)	1.798(1.227−2.635)	0.003	1.808(1.232−2.654)	0.002
Tumor stage (T3-T4 *vs.* T1-T2)	1.605(1.020−2.524)	0.041		
Lymph node metastasis (yes *vs.* no)	1.537(0.999−2.364)	0.050		
Distant metastasis (yes *vs.* no)	1.854(1.108−3.103)	0.019		
Risk score (high *vs.* low)	1.761(1.511−2.052)	<0.001	1.791(1.528−2.098)	<0.001

Figures

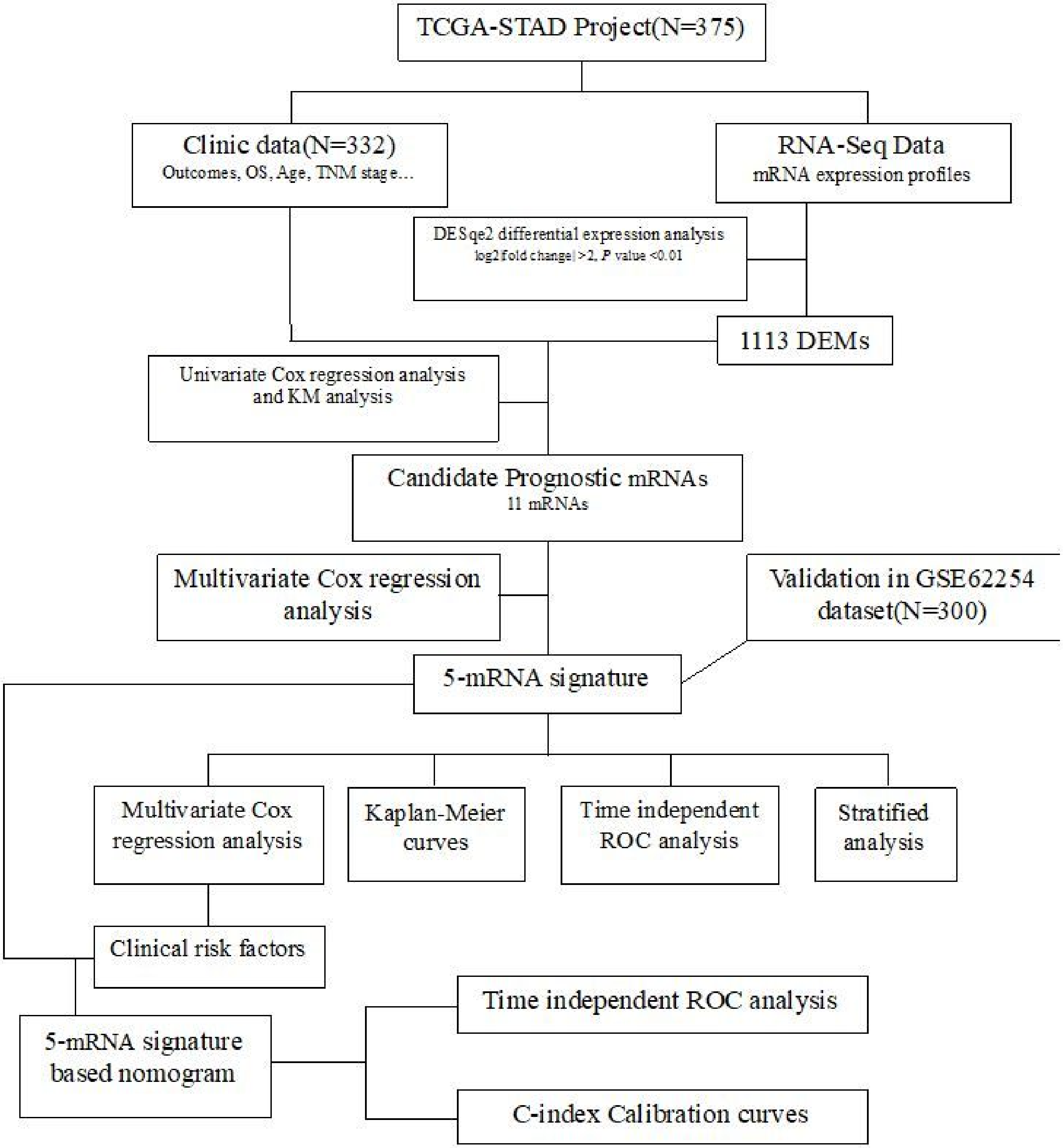

TCGA-STAD Project(N=375)
Clinic data(N=332)
Outcomes, OS, Age, TNM stag e…
RNA-Seq Data
mRNA expression profiles
DESqe2 differential expression analysis
log2fold change|>2, P value <0.01
1113 DEMs
Univariate Cox regression analysis and KM analysis
Candidate Prognostic mRNAs
11 mRNAs
Multivariate Cox regression analysis
Validation in GSE62254 dataset(N=300)
5-mRNA signature
Multivariate Cox regression analysis
Kaplan-Meier curves
Time independent ROC analysis
Stratified analysis
Clinical risk factors
5-mRNA signature based nomogram
Time independent ROC analysis
C-index Calibration curves

Figure 1

Flowchart of this study.

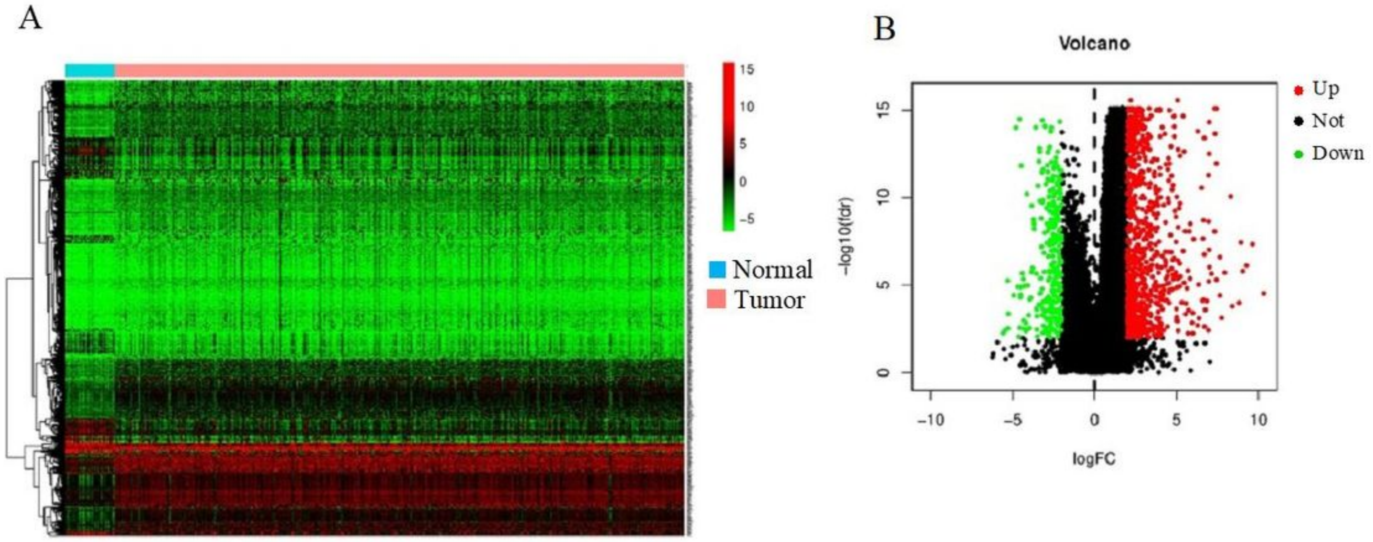


Figure 2

Volcano plot and heatmap of 1113 mRNAs in gastric cancer patients from TCGA -STAD Project. (A) Heatmap of 1113 mRNAs in gastric cancer patients from TCGA -STAD Project. (B)Volcano plot of 1113 mRNAs in gastric cancer patients from TCGA -STAD Project. Green and red indicate downregulated and upregulated mRNAs.

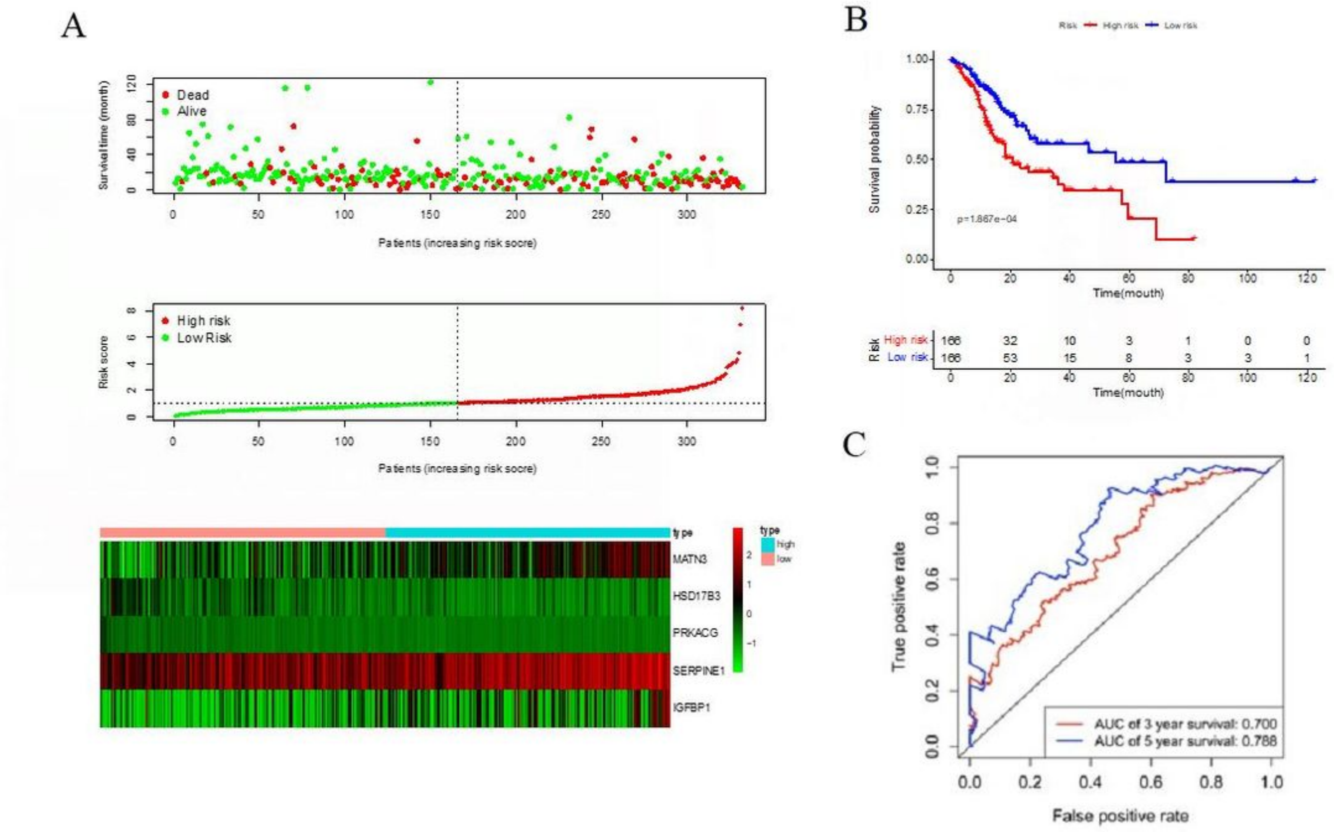


Figure 3

Identification and assessment of a five-mRNA signature to predict OS in the TCGA dataset. (A) The OS status, risk score and heatmap distribution of the five-mRNA signature. (B) Kaplan-Meier curves for OS based on the five-mRNA signature. The tick-marks on the curve represent the censored subjects. The number of patients at risk is listed below the curve. (C) Time-dependent ROC curve analysis of the five-mRNA signature for predicting OS.

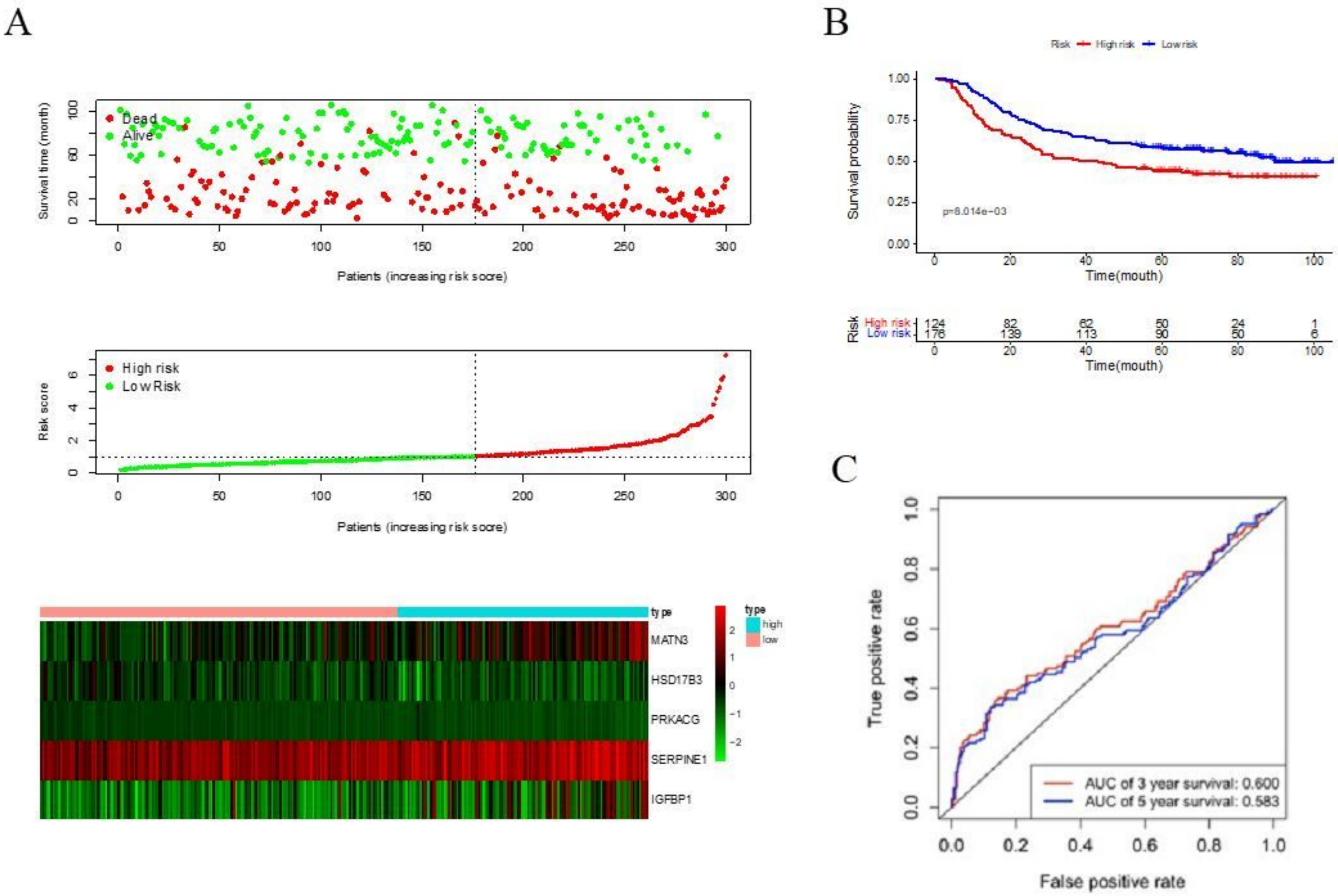


Figure 4

Identification and assessment of a five-mRNA signature to predict OS in the GSE62254 dataset. (A) The OS status, risk score and heatmap distribution of the five-mRNA signature. (B) Kaplan-Meier curves for OS based on the five-mRNA signature. The tick-marks on the curve represent the censored subjects. The number of patients at risk is listed below the curve. (C) Time-dependent ROC curve analysis of the five-mRNA signature for predicting OS.

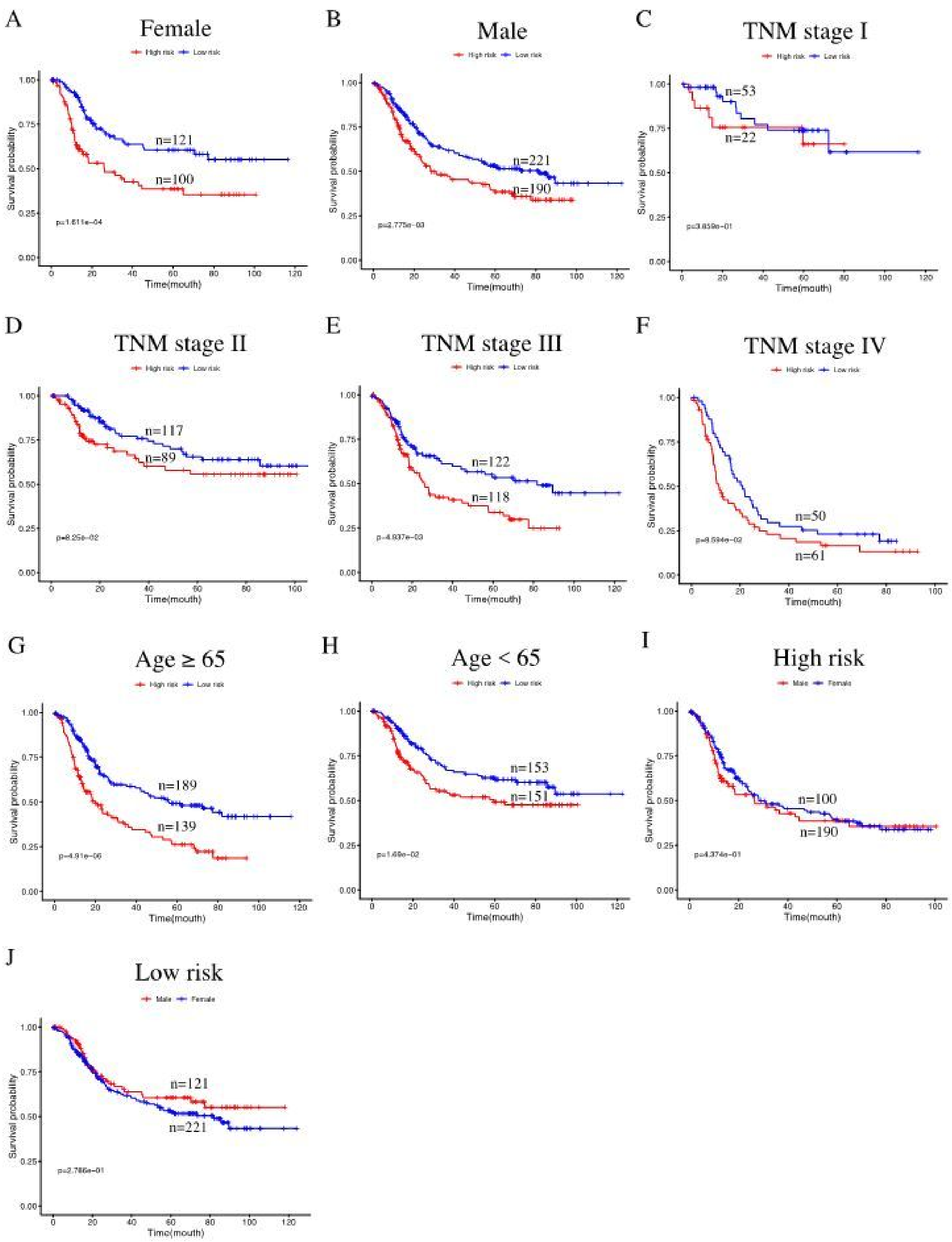


Figure 5

Risk-stratified analysis of the five-mRNA signature for gastric cancer patients. Kaplan-Meier analysis of patients in the Female subgroup (A), Male subgroup (B), stage-I subgroup (C), stage-II subgroup (D), stage-III subgroup (E), stage-IV subgroup (F), ≥65-year-old subgroup (G) and <65-year-old subgroup (H), Comparison of male and female risks in high-risk subgroup (I), Comparison of male and female risks in low-risk subgroup (J). The tick-marks on the curve represent the censored subjects. The differences between the two risk groups were assessed with two-sided log-rank tests.

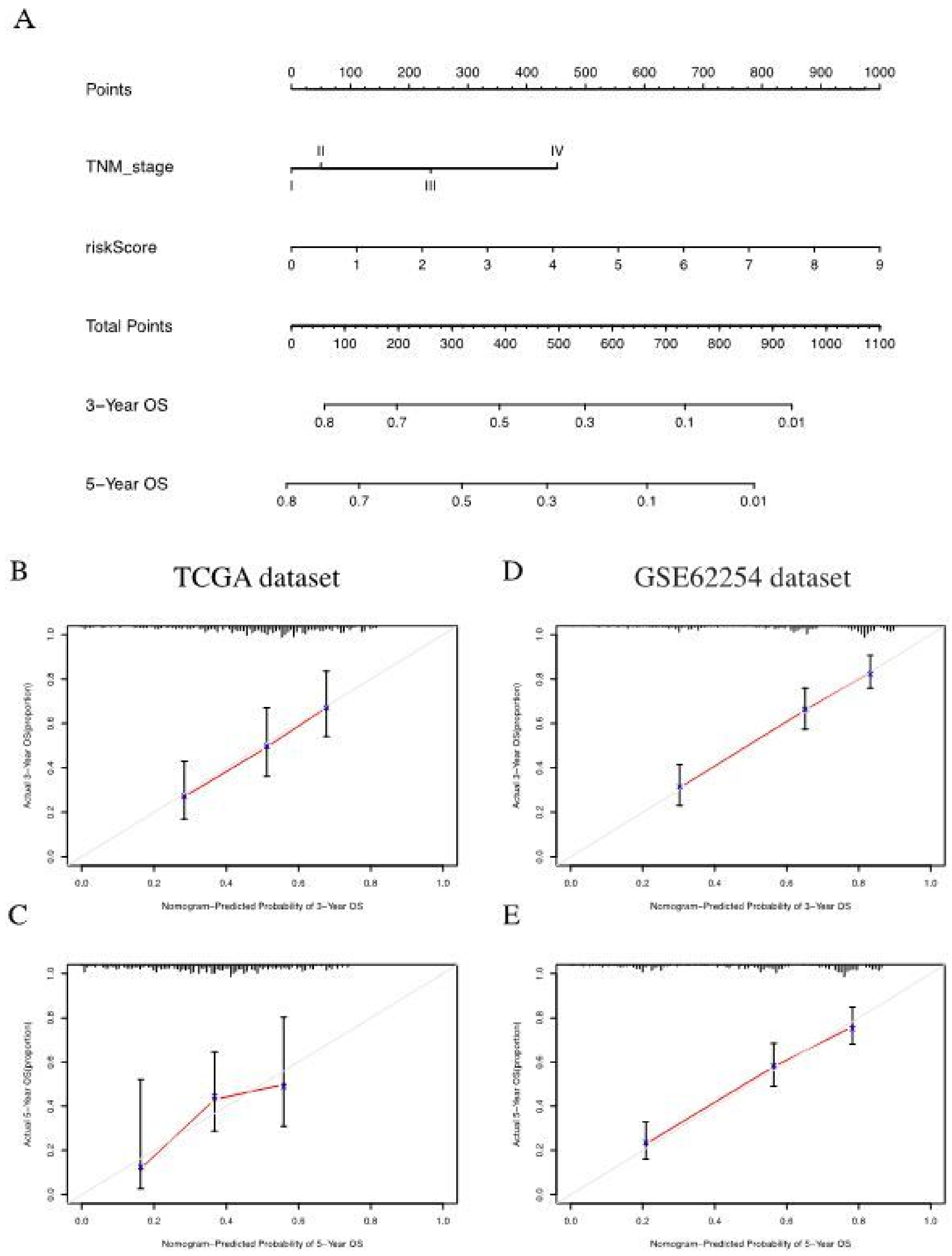


Figure 6

A five-mRNA signature-based nomogram to predict three- and five-year OS in gastric cancer patients. (A) Nomogram for predicting OS. Instructions: Locate each characteristic on the corresponding variable axis, and draw a vertical line upwards to the points axis to determine the specific point value. Repeat this process. Tally up the total points value and locate it on the total points axis. Draw a vertical line down to the three- or five-year OS to obtain the survival probability for a specific gastric cancer patient. (B–E) Calibration plots of the nomogram for predicting OS at three years (B) and five years (C) in the TCGA dataset, and at three years (D) and five years (E) in the GSE62254 dataset. The 45-degree dotted line represents a perfect prediction, and the red lines represent the predictive performance of the nomogram.

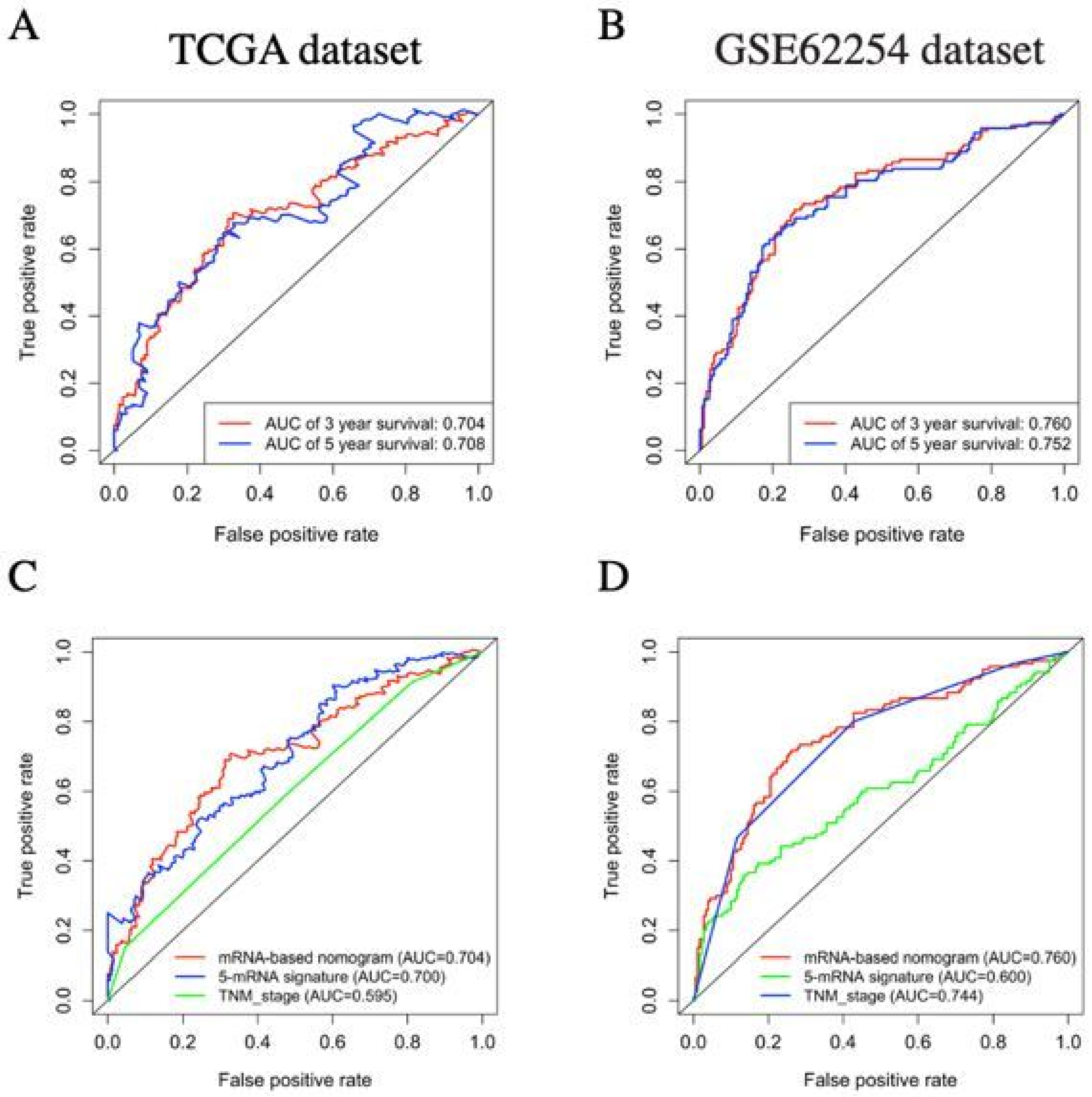


Figure 7

The prognostic value of the composite nomogram in comparison with TNM stage in the TCGA and GSE62254 dataset. Time-dependent ROC curves of the nomogram for predicting OS in the TCGA(A) and GSE62254(B) dataset. The prognostic accuracy of the three-mRNA-based prognostic nomogram compared with those of the three-mRNA signature and TNM stage in the TCGA dataset(C) and GSE62254(D) dataset.

Supplementary Files

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

- SupplementaryTable1andTable2.docx
- supp.fig1.jpg