



Identification of an EMT-Related Gene Signature for Predicting Overall Survival in Gastric Cancer

Weiyu Dai^{††}, Yizhi Xiao^{††}, Weimei Tang¹, Jiaying Li¹, Linjie Hong¹, Jieming Zhang¹, Miaomiao Pei¹, Jianjiao Lin^{1,2}, Side Liu^{1,2}, Xiaosheng Wu^{1*}, Li Xiang^{2*} and Jide Wang^{1,2*}

¹ Guangdong Provincial Key Laboratory of Gastroenterology, Department of Gastroenterology, Nanfang Hospital, Southern Medical University, Guangzhou, China, ² Department of Gastroenterology, Longgang District People's Hospital, Shenzhen, China

OPEN ACCESS

Edited by:

Juan Caballero,
European Bioinformatics Institute
(EMBL-EBI), United Kingdom

Reviewed by:

Ugur Sezerman,
Sabanci University, Turkey
Carolyn Elisabeth Behrendt,
City of Hope, United States

*Correspondence:

Xiaosheng Wu
wuxiaosheng92@163.com
Li Xiang
shellyxiang@sina.com
Jide Wang
jidewang55@163.com

^{††} These authors have contributed
equally to this work and share first
authorship

Specialty section:

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

Received: 30 January 2021

Accepted: 01 June 2021

Published: 24 June 2021

Citation:

Dai W, Xiao Y, Tang W, Li J,
Hong L, Zhang J, Pei M, Lin J, Liu S,
Wu X, Xiang L and Wang J (2021)
Identification of an EMT-Related Gene
Signature for Predicting Overall
Survival in Gastric Cancer.
Front. Genet. 12:661306.
doi: 10.3389/fgene.2021.661306

Background: It has been widely reported that epithelial-mesenchymal transition (EMT) is associated with malignant progression in gastric cancer (GC). Integration of the molecules related to EMT for predicting overall survival (OS) is meaningful for understanding the role of EMT in GC. Here, we aimed to establish an EMT-related gene signature in GC.

Methods: Transcriptional profiles and clinical data of GC were downloaded from The Cancer Genome Atlas (TCGA). We constructed EMT-related gene signature for predicting OS by using univariate Cox regression and least absolute shrinkage and selection operator (LASSO) regression analyses. Time-dependent receiver operating characteristic (ROC), Kaplan-Meier analysis were performed to assess its predictive value. A nomogram combining the prognostic signature with clinical characteristics for OS prediction was established. And its predictive power was estimated by concordance index (C-index), time-dependent ROC curve, calibration curve and decision curve analysis (DCA). GSE62254 dataset from Gene Expression Omnibus (GEO) was used for external validation. Quantitative real-time PCR (qRT-PCR) was used to detect the mRNA expression of the five EMT-related genes in human normal gastric mucosal and GC cell lines. To further understand the potential mechanisms of the signature, Gene Set Enrichment Analysis (GSEA), pathway enrichment analysis, predictions of transcription factors (TFs)/miRNAs were performed.

Results: A novel EMT-related gene signature (including ITGAV, DAB2, SERPINE1, MATN3, PLOD2) was constructed for OS prediction of GC. With external validation, ROC curves indicated the signature's good performance. Patients stratified into high- and low-risk groups based on the signature yielded significantly different prognosis. Univariate and multivariate Cox regression suggested that the signature was an independent prognostic variable. Nomogram for prognostication including the signature presented better predictive accuracy and clinical usefulness than the similar model without risk score to some extent with external validation. The qRT-PCR assays suggested that high expression of the five EMT-related genes could be found in human GC cell lines compared with normal gastric mucosal cell line. GSEA and pathway

enrichment analysis revealed that focal adhesion and ECM-receptor interaction might be the two important pathways to the signature.

Conclusion: Our EMT-related gene signature may have practical application as an independent prognostic factor in GC.

Keywords: EMT, gastric cancer, LASSO, prognosis, signature

INTRODUCTION

Gastric cancer (GC) is the third leading cause of cancer-related death and the fifth most common malignancy worldwide, with over 1,000,000 new patients and an estimated 783,000 deaths in 2018 (Bray et al., 2018). 39% of GC patients were identified with metastatic diseases (Thomassen et al., 2014). Patients with metastasis tended to have poor survival (Sleeman and Steeg, 2010). The increased chemotherapy in patients with metastasis didn't increase the population-based overall survival (OS) (Bernards et al., 2013). To date, the TNM stage system is widely regarded as a guideline for survival estimate. But wide variation in prognosis exists among GC patients with the same TNM stage on account of the inherent heterogeneity (Jiang et al., 2017, 2018). Hence, novel strategies are needed to improve the survival prediction and further guide individual treatment in GC.

Epithelial-mesenchymal transition (EMT) is a reversible process in which epithelial cells can transdifferentiate into motile mesenchymal cells, and it is vital to embryogenesis, wound healing and the tumorigenic process (Dongre and Weinberg, 2019). EMT is a complicated process which can be driven by key transcriptional factors like SNAIL, zinc-finger E-box-binding (ZEB) and basic helix-loop-helix (bHLH) transcriptional factors (Peinado et al., 2007). And reprogramming of gene expression, lots of pathways such as transforming growth factor- β (TGF- β) family signaling, PI3K-AKT, ERK-MAPK, p38 MAPK and JUN N-terminal kinase (JNK) pathways, etc., are involved in EMT (Lamouille et al., 2014). EMT is associated with stemness, initiation, invasion, metastasis and chemo-resistance in GC, and the status of EMT is a critical prognosticator for GC (Murai et al., 2014; Huang L. et al., 2015). Due to the convenient access to transcriptional data from online data hubs, establishing the gene signature underlying the mechanism of cancer is an area of active research (Wang et al., 2019b; Zhao et al., 2019; Cao et al., 2020). Considering that EMT status has been previously shown to be prognostic in GC, biomarkers related to EMT represent a promising source for assembling an independently significant prognostic signature for GC.

In this study, we constructed an EMT-related gene signature for predicting OS based on the transcriptional profiles of GC from The Cancer Genome Atlas (TCGA). Univariate Cox regression and least absolute shrinkage and selection operator (LASSO) regression were conducted to identify the prognostic five-gene signature. Receiver operating characteristic (ROC) curve and survival analysis were used to estimate it. Then, a nomogram was built by combining the risk score and clinical parameters to predict OS in GC. Concordance index (C-index), ROC curve, calibration curve and decision curve analysis (DCA) were

performed to assess the nomogram. Besides, the prognostic value of the nomogram was verified by an external validation. Collectively, our finding highlights the functional role of EMT-related gene signature and nomogram in predicting OS for GC.

MATERIALS AND METHODS

Data Collection

RNA-sequencing and clinical information of GC samples in TCGA were obtained from the Genomic Data Commons Data Portal (GDC¹). "HTSeq-FPKM" workflow type of transcriptome profiling for TCGA-STAD (stomach adenocarcinoma) project was download, including 375 cancer tissues samples' and 32 normal samples' gene expression profiles. Clinical information of 443 GC tissues from TCGA-STAD project was downloaded with the format of "bcr xml." The TCGA-STAD cohort was assigned as the training cohort. The external validation cohort GSE62254 was acquired from Gene Expression Omnibus (GEO²) (Cristescu et al., 2015). GSE62254 was conducted by GPL570 platform (Affymetrix Human Genome U133 Plus 2.0 Array), consisting of 300 GC samples with corresponding clinical information. The normalized expression matrix GSE62254 was used directly for subsequent analyses. All the data was obtained in March 2020. Patients who met the following criteria were included in the subsequent analyses: (a) sufficient gene expression information, (b) survival time no less than 30 days, (c) sufficient clinical information of age, gender, TNM stage, T stage, N stage, M stage, number of lymph nodes examined and number of positive nodes. Increasing evidences have revealed that lymph node ratio (LNR, the ratio of the positive lymph nodes positive to lymph nodes examined) was an important prognostic factor in GC, so it was considered in our study and the LNR values of all patients included were calculated (Zhao et al., 2016; Lee et al., 2017). Thus, 278 patients (278 tumor samples and 26 normal samples) in TCGA-STAD and 298 patients (298 tumor samples) in GSE62254 were included in our study with the accompanying information above. For our study was based on the de-identified data from the TCGA and GEO databases, institutional review, institutional approval and informed consent were not required.

Identification of Differentially Expression EMT-Related Genes (DEEGs)

Gene members from gene set HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION from Molecular

¹<https://portal.gdc.cancer.gov/>

²<https://www.ncbi.nlm.nih.gov/geo/>

Signatures Database v7.0 (MSigDB³) were selected as candidate EMT-related genes (Subramanian et al., 2005; Liberzon et al., 2015). The R package “sva” was applied to eliminate the batch effect among the datasets (Leek et al., 2019). DEEGs were identified in training cohort TCGA-STAD by using an R package “limma” (Ritchie et al., 2015). A false discovery rate (FDR) adjusted p -value < 0.05 and an absolute value of \log_2 (fold change) > 0.5 were considered as the criteria for DEEGs identification. Heatmaps were conducted by using R package “pheatmap” (Kolde, 2019).

Construction of EMT-Related Gene Signature

Firstly, the prognostic DEEGs were screened out by using univariate Cox regression analyses for OS. Then LASSO regression was applied to construct a multi-gene signature with the prognostic DEEGs based on lambda.min. The optimal value of lambda was identified through tenfold cross-validations. Univariate Cox regression and LASSO regression were performed in R with “survival” and “glmnet” package (Therneau and Grambsch, 2000; Friedman et al., 2010; Therneau, 2020). Risk score of each patients was calculated based on the signature, using the formula as follows:

$$\text{Risk score} = \sum_{i=1}^n \beta_i \times \text{Exp}_i$$

in which the *Exp* represents the expression of gene and the β is the LASSO coefficient of gene. All samples were separated to high- and low-risk groups based on the optimal cut-off value determined by the “surv_cutpoint” function of the R package “survminer” (Kassambara et al., 2019), which uses the maxstat (maximally selected rank statistics) statistic to determine the optimal cutpoint for continuous variables.

Assessment and Validation of EMT-Related Gene Signature

Receiver operating characteristic curve was performed to qualify the discrimination of the signature by measuring the area under the curve (AUC). ROC curve was plotted with R package “survivalROC” (Heagerty et al., 2000). Kaplan-Meier curve combined with a log-rank test for OS was performed to evaluate the predictive value of the signature by using the R package “survival” (Therneau and Grambsch, 2000; Therneau, 2020). Univariate and multivariate Cox regression were performed to identify whether risk score was an independent prognostic factor for OS. To validate the signature, the same methods were performed in the external validation cohort GSE62254.

Correlation Between the Signature and Clinical Characteristics

To investigate the predictive ability of the prognostic signature in different clinical characteristics, all the patients were divided

into subgroups according to age, gender, TNM stage, T stage, N stage, M stage and LNR. Survival analysis and investigation of risk score were performed in each subgroups. The relationship between risk levels and clinical characteristics was measured using chi-square test.

Public Database Mining of Genes in EMT-Related Gene Signature

Oncomine⁴ was used to investigate the expression profile of the EMT-related gene in GC (Rhodes et al., 2007). The genomic alterations and co-expressed genes of the EMT-related gene were explored by using cBioportal⁵ (Gao et al., 2013). We used The Human Protein Atlas⁶ to study the expression profile of the EMT-related gene at a translational level (Uhlen et al., 2005, 2017). The networks between EMT-related genes and transcription factors (TFs) or miRNAs were predicted by NetworkAnalyst⁷ and drawn with Cytoscape 3.7.0 (Shannon et al., 2003; Lachmann et al., 2010; Hsu et al., 2014; Zhou et al., 2019). The prediction of TFs was based on ChEA database while the data of miRNAs prediction was collected from miRTarBase via NetworkAnalyst platform. Co-expressed genes with a Spearman correlation ≥ 0.4 or < -0.4 were submitted to Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis by using R package “clusterProfiler” with a p -value < 0.05 and a q -value < 0.05 (Kanehisa and Goto, 2000; Yu et al., 2012). The enrichment analysis was visualized by R package “enrichplot” (Yu, 2019).

Cell Culture

The human GC cell lines AGS and NCI-N87 and the human normal gastric mucosal cell line GES-1 were obtained from the Cell Bank of the Chinese Academy of Sciences (Shanghai, China). The human GC cell lines Hs-746T and SNU-719 were purchased from Procell (Wuhan, China). The human GC cell line SNU-5 was obtained from ATCC. Cells were cultured in RPMI-1640 medium (Gibco, Grand Island, NY, United States) supplemented with 10% fetal bovine serum (FBS) (NBCS) (PAA Laboratories, Inc., Pasching, Austria) at 37°C in an atmosphere of 5% CO₂.

RNA Isolation and Quantitative Real-Time PCR (qRT-PCR)

Total RNAs were extracted from cells by using Trizol reagent (Invitrogen, Carlsbad, CA, United States), and qRT-PCR was performed by using the PrimerScript RT Master Mix (Takara Bio, Inc., Shiga, Japan) and TB Green Premix Ex Taq (Takara Bio, Inc., Shiga, Japan) according to the manufacturer’s instructions. GAPDH was used as gene internal control and the final data were analyzed with the $2^{-\Delta\Delta Ct}$ method. The specific sense primers for ITGAV, DAB2, SERPINE1, MATN3, PLOD2 and GAPDH are listed in **Supplementary Table 1**.

⁴<https://www.oncomine.org/>

⁵<https://www.cbioportal.org/>

⁶<https://www.proteinatlas.org/>

⁷<https://www.networkanalyst.ca/>

³<https://www.gsea-msigdb.org/gsea/msigdb/index.jsp>

Establishment and Assessment of Signature-Based Nomogram

A nomogram for OS prediction was formulated based on the result of multivariate Cox regression by using the R package “rms” (Harrell, 2019). The C-index, ROC curve, calibration curve and DCA were used to assess the nomogram. The C-index and AUC of ROC curve were calculated to evaluate the discriminatory of the nomogram. The calibration curve was performed to compare the predicted survival outcome with the actual outcome by a bootstrap method with 1000 resamples. DCA was preformed to assess clinical utility of the nomogram by comparing the net benefit of the nomogram with all or none strategies (Vickers and Elkin, 2006). Akaike information criterion (AIC) was used to test the goodness of fit for models. The same methods were used to validate the results in the external validation cohort GSE62254.

Gene Set Enrichment Analysis

Gene Set Enrichment Analysis (GSEA) was performed to study the different KEGG pathways between high- and low-risk groups in TCGA-STAD by using the GSEA software (v4.0.3⁸) (Subramanian et al., 2005). The reference sets for calculating Enrichment Score (ES) were c2.cp.kegg.v7.0.symbols.gmt. Gene sets were considered to be significantly enriched when p -value < 0.05 and FDR < 0.25 after performing 1000 permutations. The GSEA figures were plotted with the R package “ggplot2” (Wickham, 2016).

Survival Analysis

The optimal cut-off value for high and low expression of gene or LNR was determined by the “surv_cutpoint” function of the R package “survminer” (Kassambara et al., 2019). Survival analysis was used to identify the difference of OS between the high and low expression groups of the EMT-related gene or the high- and low-risk groups of the signature. All the survival analyses mentioned above were performed by Kaplan-Meier curve with a two-side log-rank test.

Statistical Analyses

All the analyses were performed in R v3.6.0 (The R Foundation for Statistical Computing, Vienna, Austria) and GraphPad Prism v7.00 (GraphPad Software Inc., United States). Boxplot was analyzed using Mann-Whitney U test. The R package “corrplot” was applied to draw the correlation plot of prognostic EMT-related genes and the Spearman test was used to analyze the correlation (Wei and Simko, 2017). Comparisons between qRT-PCR results from different cell lines were performed using one-way ANOVA and Dunnett’s T3 multiple comparison test and the results were presented as mean \pm SD. p < 0.05 was considered statistically significant.

⁸<https://www.gsea-msigdb.org/gsea/index.jsp>

RESULTS

Construction of the EMT-Related Gene Signature in GC

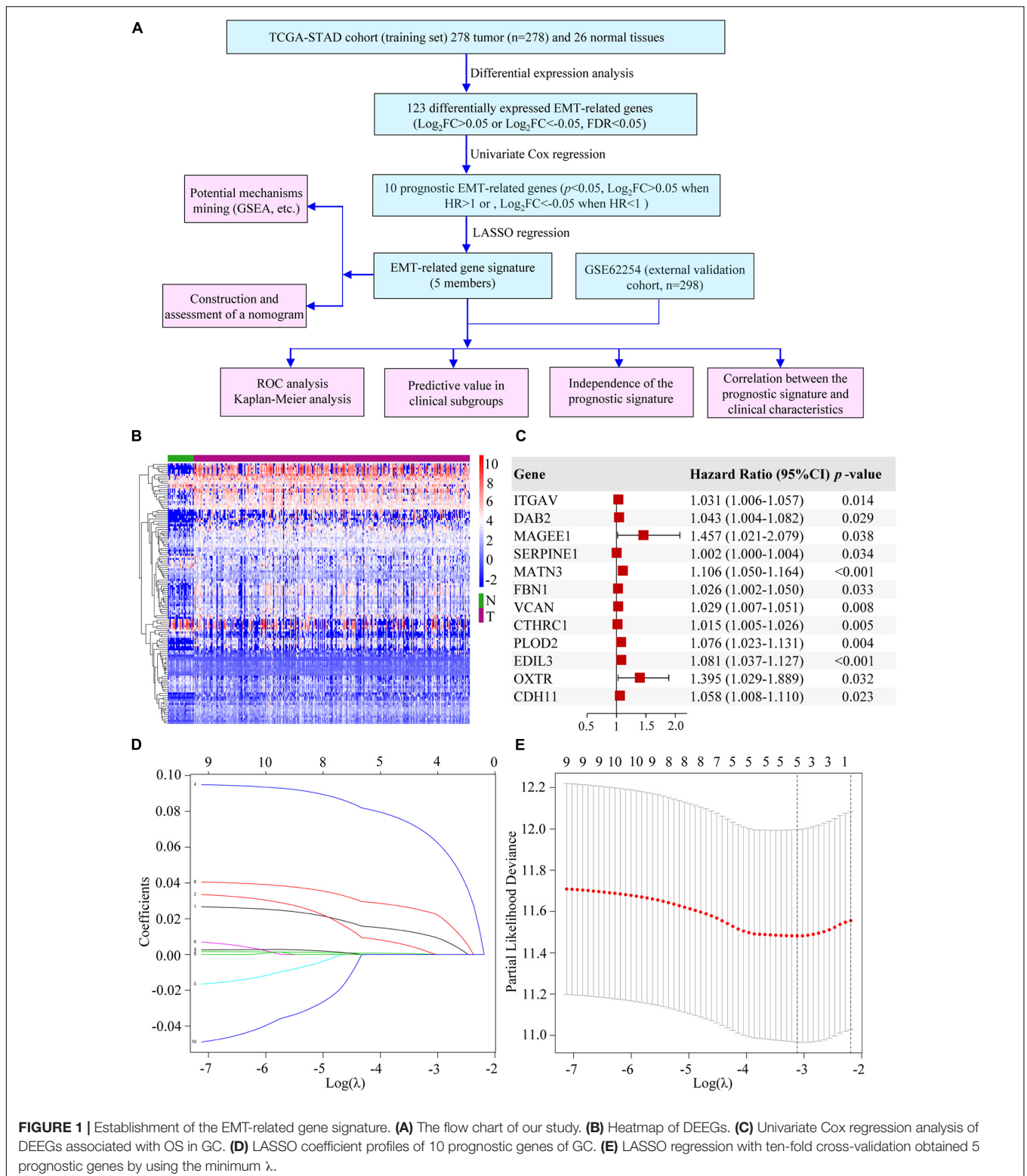
We conducted our study as illustrated in **Figure 1A**. A total of 123 significantly differentially expressed EMT-related genes were identified from TCGA-STAD cohort, of which 89 were upregulated and 34 were downregulated in GC (**Figure 1B** and **Supplementary Table 2**). After an initial screening of EMT-related genes associated with OS by using univariate Cox regression analysis, 12 prognostic genes were found (**Figure 1C**). *MAGEE1* and *EDIL3* were excluded from our analysis because they were downregulated genes with HR > 1. Considering that there were correlations among 10 prognostic genes (**Supplementary Figure 1**), all of them were selected to the LASSO modeling to reduce multicollinearity (**Figures 1D,E**). The prognostic risk score of the signature was identified: risk score = 0.010325 \times (expression level of *ITGAV*) + 0.000891 \times (expression level of *DAB2*) + 0.000183 \times (expression level of *SERPINE1*) + 0.065772 \times (expression level of *MATN3*) + 0.023410 \times (expression level of *PLOD2*). It was indicated that they were all risk factors for OS. The absolute values of coefficients indicated that *MATN3* had the most influence on OS prediction, yet *SERPINE1* had the least.

Estimation and Validation of EMT-Related Gene Signature

Based on the “surv_cutpoint” function of the R package “survminer,” we calculated that the optimal cut-off value was 0.318659. Patients in the TCGA cohort were divided into high- and low-risk groups according to the optimal cut-off. The patients’ risk score distribution, survival status and gene expression levels of EMT-related gene signature were presented in **Figure 2A**. To validate the predictive value of the EMT-related gene signature, risk scores for patients in GSE62254 were calculated with the same formula. And patients were separated into high- and low-risk groups according to the same cut-off. Risk score distribution, survival status, and gene expression levels of the signature were also shown (**Figure 2B**). The AUCs for 1-, 3-, and 5-year OS were 0.655, 0.696, and 0.784 in the TCGA cohort (**Figure 2C**). And in the validation cohort, the AUCs for 1-, 3- and 5-year OS were 0.640, 0.658, and 0.635, respectively, showing the good prognostic discrimination of the EMT-related gene signature (**Figure 2D**). The survival analysis showed that the OS of low risk group was better than that of high risk group (TCGA-STAD, p < 0.001; GSE62254, p < 0.001) (**Figures 2E,F**). Together, by modeling with training cohort and external validation, our results indicated that the EMT-related gene signature performed well for OS prediction.

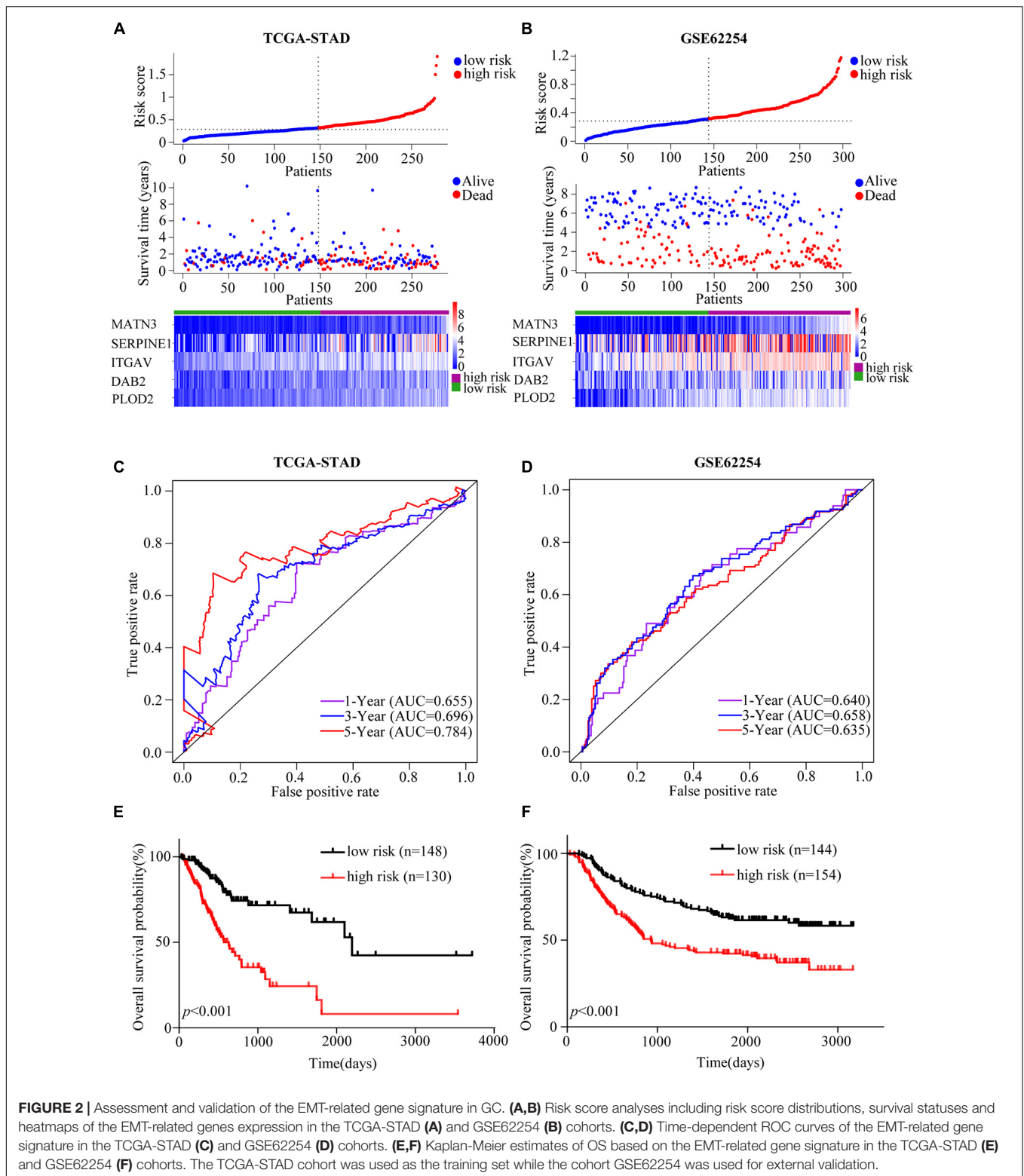
EMT-Related Gene Signature in Different Clinical Subgroups

Patients were divided into different subgroups according to age, gender, TNM stage, T stage, N stage, M stage and LNR. Kaplan-Meier analyses of the EMT-related gene signature in



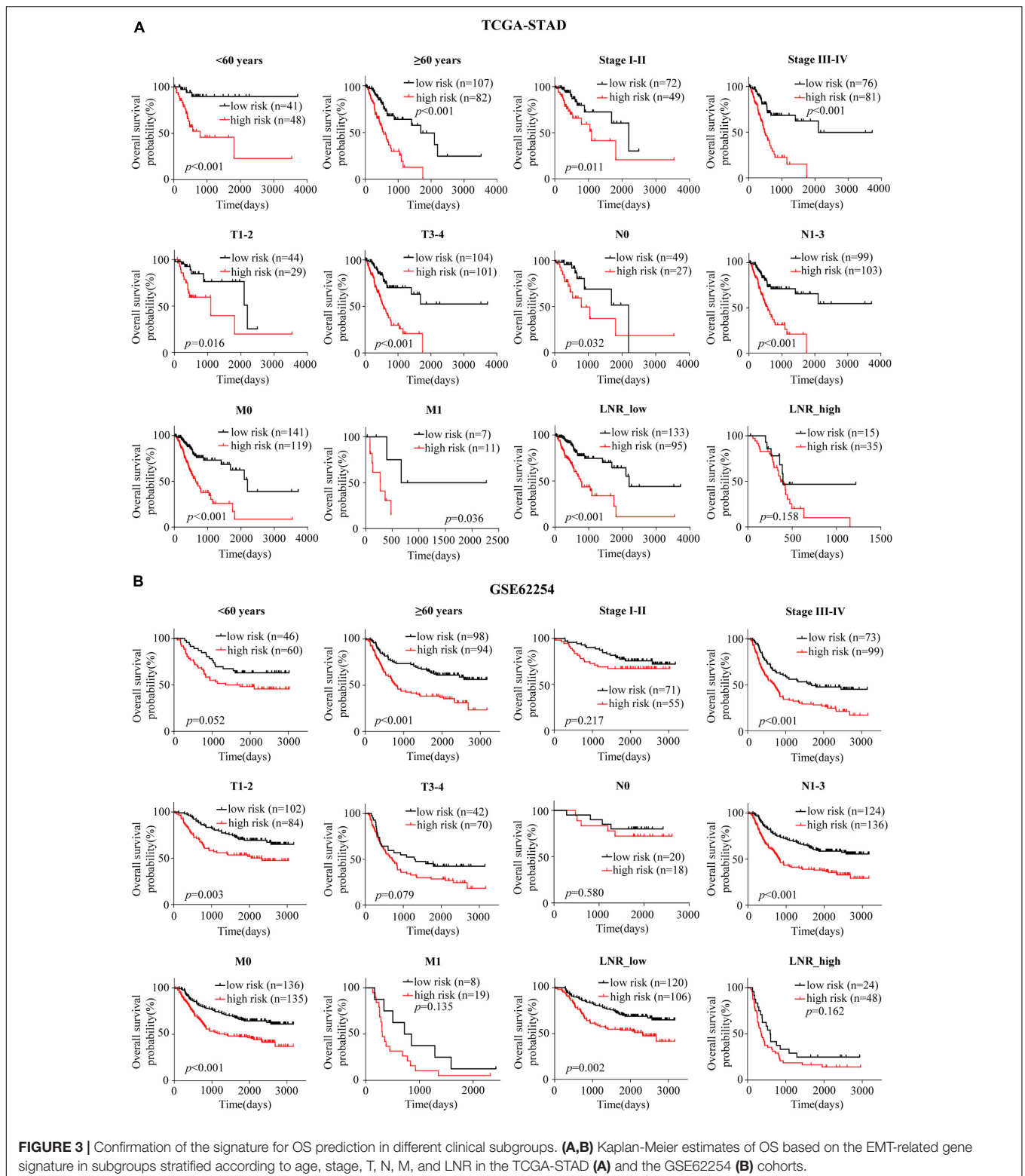
subgroups showed that in the TCGA cohort, patients with high-risk had worse OS than patients with low-risk in < 60 years ($p < 0.001$), ≥ 60 years ($p < 0.001$), female ($p < 0.001$), male ($p < 0.001$), stage I-II ($p = 0.011$), stage III-IV ($p < 0.001$), T1-2

($p = 0.016$), T3-4 ($p < 0.001$), N0 ($p = 0.032$), N1-3 ($p < 0.001$), M0 ($p < 0.001$), M1 ($p = 0.036$) and LNR_low ($p < 0.001$) subgroups (**Figure 3A** and **Supplementary Figure 2A**). Similar results could be obtained in subgroups such as ≥ 60 years



($p < 0.001$), female ($p < 0.001$), male ($p = 0.005$), stage III-IV ($p < 0.001$), T1-2 ($p = 0.003$), N1-3 ($p < 0.001$), M0 ($p < 0.001$), and LNR_low ($p = 0.002$) for the external validation cohort **(Figure 3B and Supplementary Figure 2B)**. We performed

univariate Cox regression based on the factors including age, stage, T, N, M, number of lymph nodes examined, number of positive nodes, LNR and risk score. As factors such as T, N and M are not independent of stage by definition and number



of lymph nodes examined and number of positive nodes are components of LNR, they were omitted from multivariate Cox regression. So multivariate Cox regression was performed based on the factors including age, gender, stage, LNR and risk score

(**Supplementary Table 3**). But, gender was finally excluded from our multivariate Cox regression model (shown in **Tables 1, 2**) as it was not a significant prognostic factor according to the multivariate Cox regression analysis in **Supplementary Table 3**.

TABLE 1 | Univariate and multivariate Cox regression analyses of clinical characteristics and risk score in TCGA-STAD.

Characteristics	Number	Univariate Cox regression		Multivariate Cox regression	
		Hazard Ratio (95%CI)	p-value	Hazard Ratio (95%CI)	p-value
Age	278	1.016 (0.996–1.036)	0.121	1.029 (1.008–1.052)	0.008
Gender					
Male/Female	175/103	1.549 (0.988–2.429)	0.056		
Tumor stage					
II/I	82/39	1.349 (0.615–2.962)	0.455	1.182 (0.532–2.626)	0.681
III/I	127/39	1.953 (0.954–3.999)	0.067	1.258 (0.552–2.867)	0.585
IV/I	30/39	4.363 (1.984–9.595)	<0.001	3.050 (1.235–7.533)	0.016
T					
T3-4/T1-2	205/73	1.548 (0.951–2.521)	0.079		
N					
N1-3/N0	202/76	1.568 (0.958–2.566)	0.073		
M					
M1/M0	18/260	2.547 (1.317–4.925)	0.005		
Number of lymph nodes examined	278	0.994 (0.981–1.006)	0.323		
number of positive lymph nodes	278	1.053 (1.034–1.072)	<0.001		
LNR	278	4.548 (2.514–8.226)	<0.001	2.987 (1.412–6.322)	0.004
Risk score	278	4.447 (2.400–8.242)	<0.001	4.185 (2.039–8.588)	<0.001

Univariate and multivariate Cox regression suggested that the five-gene prognostic signature was an independent OS predictor for GC in the TCGA (Table 1) cohort and the GEO cohort (Table 2).

Correlation Between the EMT-Related Gene Signature and Clinical Characteristics

Distribution of several clinical parameters varied between high- and low-risk groups. It was illustrated that there were more N1-3

or LNR_high cases in the high-risk group than in the low-risk group of the TCGA cohort (N, $p = 0.021$; LNR, $p < 0.001$) (Figure 4A and Supplementary Table 4). In the validation cohort, more stage III-IV, T3-4, M1 or LNR_high cases could be found in the high-risk group (stage, $p = 0.018$; T, $p = 0.004$; M, $p = 0.042$; LNR, $p = 0.003$) (Figure 4B and Supplementary Table 5). What's more, in the TCGA cohort, patients with N1-3 ($p = 0.012$), LNR_high ($p = 0.004$) tended to have higher risk scores (Figure 4C). And patients with stage III-IV ($p = 0.024$), T3-4 ($p = 0.012$), M1 ($p = 0.0496$), LNR_high ($p < 0.001$) yielded higher risk scores in the GEO cohort (Figure 4D).

Expression Profiles and Survival Analyses of the Five Members

To make a complete analysis of the contributions of EMT-related gene signature members in GC, the expression profiles and OS predictive values of ITGAV, DAB2, SERPINE1, MATN3 and PLOD2 were investigated. As shown in Figure 5A, all of them were significantly upregulated in GC samples in the TCGA cohort. The EMT-related gene signature members were retrieved using the Oncomine database. It was indicated that EMT-related gene signature members acted as oncogenes in most types of cancer (Figure 5B). There was 1 dataset for SERPINE1 suggesting its upregulation in GC. The images of immunohistochemistry (IHC) staining showed the protein expression of IAGAV, DAB2, SERPINE1 and PLOD2 in GC (Figure 5C). However, we did not find the protein expression images of MATN3 in the database. In the TCGA cohort, all of the EMT-related gene signature members were significantly associated with unfavorable OS outcome (Figure 5D) (ITGAV, $p < 0.001$; DAB2, $p = 0.002$; SERPINE1, $p = 0.002$; MATN3, $p < 0.001$; PLOD2, $p = 0.002$). Similarly, all the genes except ITGAV ($p = 0.167$) were observed as significantly unfavorable prognostic genes in GSE62254 (Figure 5E) (DAB2, $p < 0.001$; SERPINE1, $p = 0.014$; MATN3, $p < 0.001$; PLOD2, $p < 0.001$). Furthermore, we detected the mRNA levels of the EMT-related genes in human GC cell lines (AGS, SNU-5, Hs-746T, NCI-N87, SNU-719) and normal gastric mucosal cell line GES-1. Elevated expression of ITGAV was found in AGS, Hs-746T, NCI-N87 and SNU-719 compared with GES-1. The expression of DAB2 in AGS, SNU-5 and SNU-719 was higher than that in GES-1. The expression level of SERPINE1 in Hs-746T and SNU-719 and the expression of MATN3 in SNU-719 were 2-fold higher than those in GES-1. Expression of PLOD2 was significantly higher in SNU-5, Hs-746T and NCI-N87 compared with GES-1 (Figure 6).

Potential Mechanisms Mining of the Five Members

In order to explore the potential mechanisms of the five members in GC, we conducted the GSEA analysis for the prognostic signature, and investigated the enriched KEGG pathways, TFs/miRNA predictions, genomic alterations of the EMT-related signature members. GSEA analysis revealed that the high-risk group might be involved in KEGG pathways such as calcium signaling pathway, ECM receptor interaction, focal adhesion, gap junction and other pathways (Figure 7A). What's

TABLE 2 | Univariate and multivariate Cox regression analyses of clinical characteristics and risk score in GSE62254.

Characteristics	Number	Univariate Cox regression		Multivariate Cox regression	
		Hazard Ratio (95%CI)	p-value	Hazard Ratio (95%CI)	p-value
Age	298	1.011 (0.996–1.026)	0.157	1.028 (1.012–1.044)	0.001
Gender					
Male/Female	197/101	0.917 (0.656–1.282)	0.611		
Tumor stage					
II/I	96/30	1.955 (0.759–5.034)	0.165	1.678 (0.649–4.338)	0.285
III/I	95/30	4.064 (1.623–10.176)	0.003	2.669 (1.030–6.920)	0.043
IV/I	77/30	9.737 (3.907–24.265)	<0.001	4.837 (1.755–13.328)	0.002
T					
T3-4/T1-2	112/186	2.365 (1.719–3.255)	<0.001		
N					
N1-3/N0	260/38	2.847 (1.450–5.589)	0.002		
M					
M1/M0	27/271	3.806 (2.460–5.888)	<0.001		
Number of lymph nodes examined	298	0.987 (0.976–0.998)	0.023		
number of positive lymph nodes	298	1.072 (1.056–1.087)	<0.001		
LNR	298	25.430 (13.312–48.578)	<0.001	7.030 (2.803–17.631)	<0.001
Risk score	298	5.205 (2.750–9.853)	<0.001	4.059 (2.097–7.855)	<0.001

more, the KEGG enrichment analyses for co-expressed genes associated with EMT-related signature members were presented in **Figure 7B**. The KEGG pathways such as focal adhesion, ECM-receptor interaction, PI3K-Akt signaling pathway and proteoglycans in cancer were the top significant pathways related to all the members. It was interesting that focal adhesion and ECM-receptor interaction were the two pathways that both appeared in the results of GSEA and pathway enrichment analysis, which we thought were important to our signature. The TFs and miRNAs connected with EMT-related

gene signature members were investigated by NetworkAnalyst (**Figures 7C,D**). Explored by using the cBioportal database, the genomic alterations of EMT-related gene signature members in GC varied from 6 to 13% (**Figure 7E**) (ITGAV, 13%; DAB2, 12%; SERPINE1, 8%; MATN3, 6%; PLOD2, 9%).

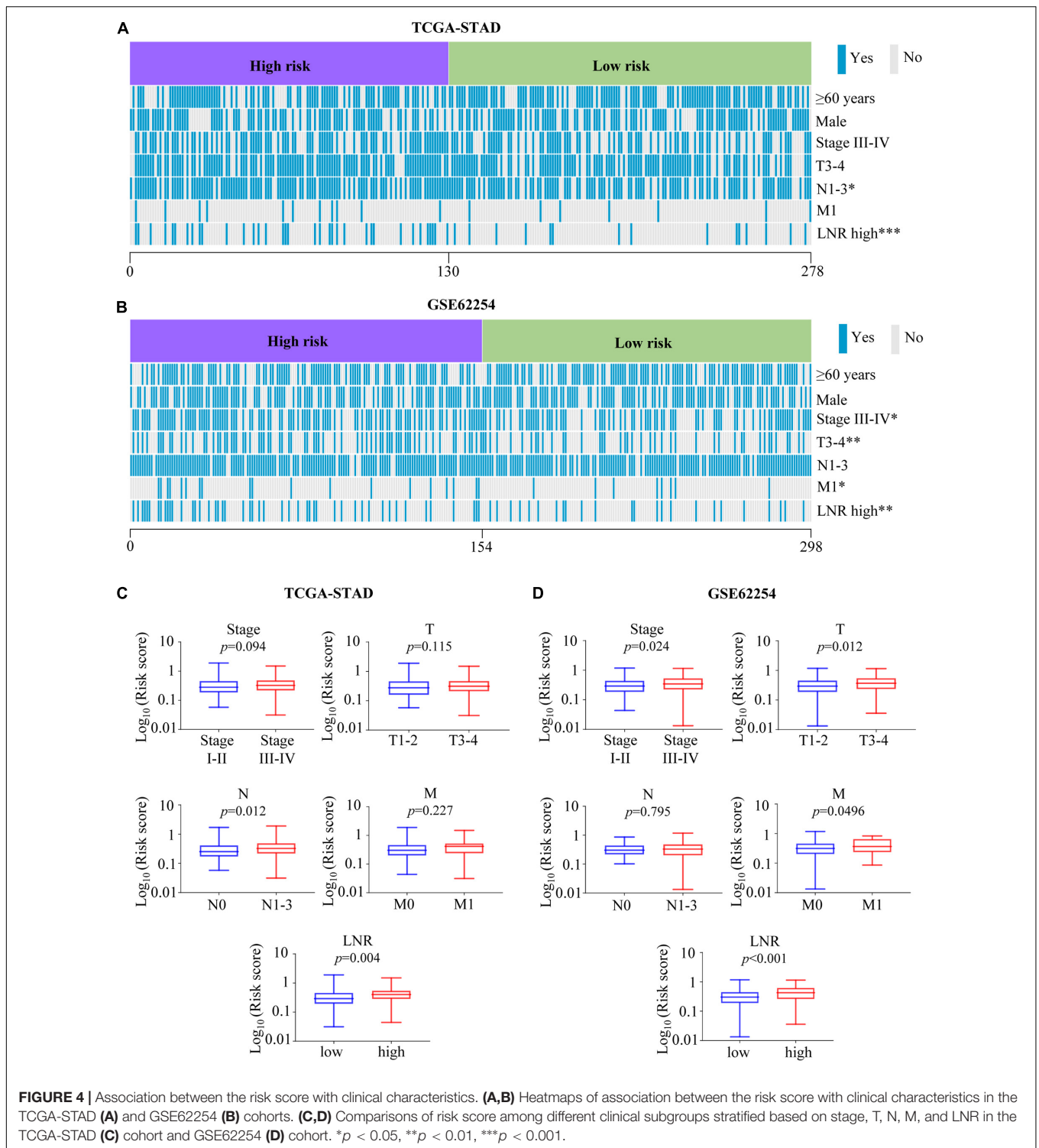
Construction and Validation of the Signature-Based Nomogram

A nomogram integrating the risk score, age, TNM stage and LNR for OS prediction of the patients with GC was shown in **Figure 8A**. The nomogram was built based on the variables applied to the final multivariable Cox regression of the training cohort above (**Table 1**). The C-index, ROC curve, calibration curve and DCA were used to quantify the model's discrimination, calibration and clinical usefulness.

In the training cohort, the C-index of the nomogram for OS prediction was 0.702 and the AUCs of the ROC for 1-, 3- or 5-year OS of the nomogram were 0.718, 0.751, and 0.727, respectively (**Figure 8D**). Compared with age, TNM stage, LNR and the age + TNM stage + LNR model, the combined nomogram yielded largest AUC for 1-, 3-year OS but not for 5-year OS. The validation cohort GSE62254 was used to test the predictive accuracy of the nomogram. The C-index in validation cohort was 0.730 and the AUC values of ROC were 0.826 at 1 year, 0.787 at 3 years and 0.764 at 5 years (**Figure 8E**). We could find that in the validation cohort, the AUCs of the nomogram (age + TNM + LNR + risk score model) were better than age + TNM + LNR model, or age, TNM stage, LNR alone for 1-, 3- and 5-year OS, suggesting that the nomogram presented better discrimination when including risk score into model for prognostication in GC to a certain extent, and we could explain the modest improvement by the EMT risk score. The calibration curve and DCA curves of the training cohort TCGA-STAD were presented in **Figures 8B,F**. In the validation cohort GSE62254, the calibration curves for the probabilities of 1-, 3- or 5-year OS demonstrated good agreement between prediction by nomogram and actual observation (**Figure 8C**). Shown by the DCA curves of validation set, the combined nomogram yielded modest additional net benefit for 3- or 5-year OS probability from using the nomogram instead of clinical model without EMT risk score, illustrating that the combined nomogram had potential for clinical utility and the modest additional net benefit for 3- or 5-year OS probability might be explained by the EMT risk score (**Figure 8G**). We used AIC to test the goodness of fit for models including our nomogram (using age, TNM stage and LNR as well as EMT risk score) and the models combining two or three of the factors including age, TNM stage, LNR and risk score (**Supplementary Table 6**). We could find that the nomogram might be the optimal model based on the smallest value for the AIC statistic. Collectively, our combined nomogram performed well for OS prediction in GC.

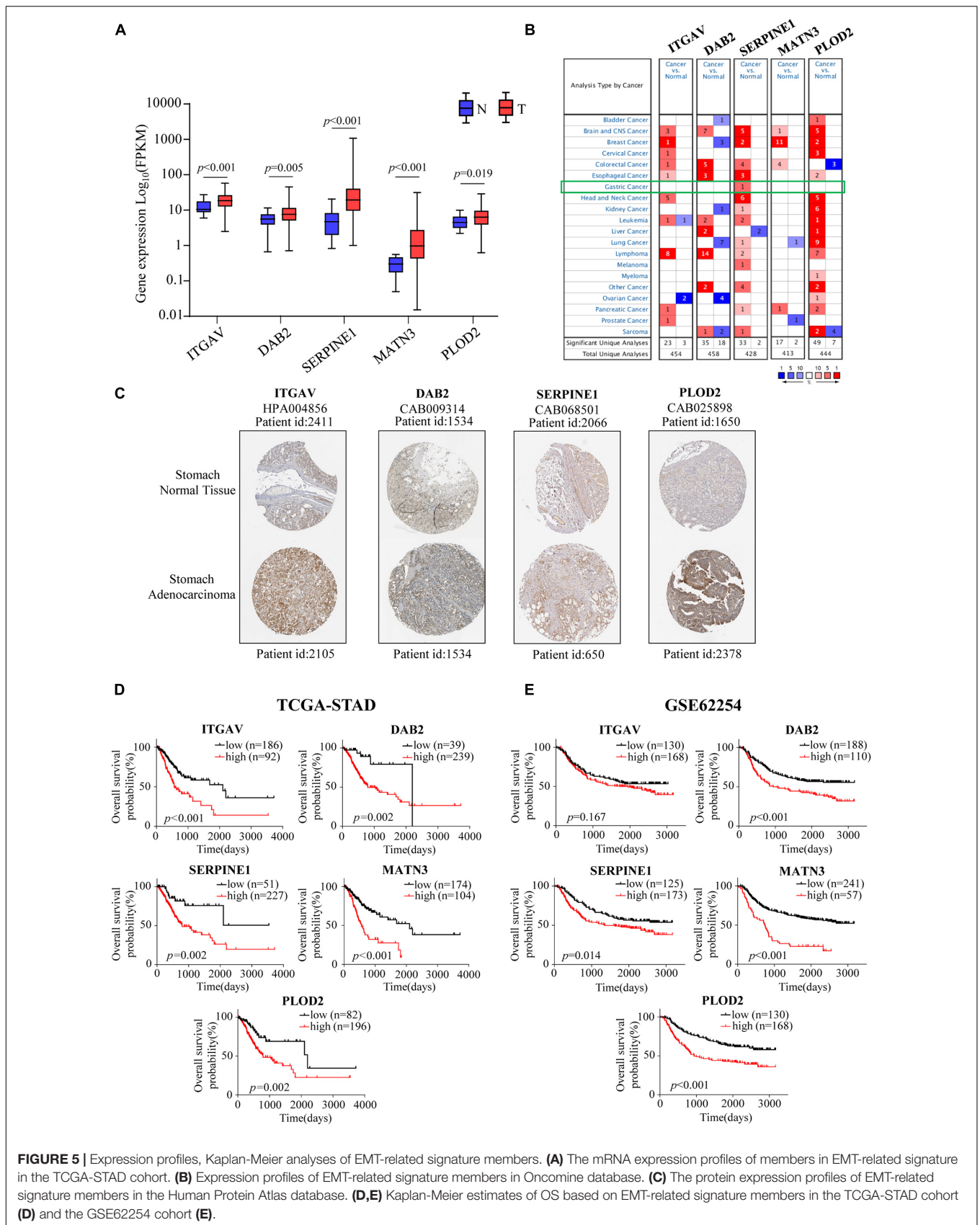
DISCUSSION

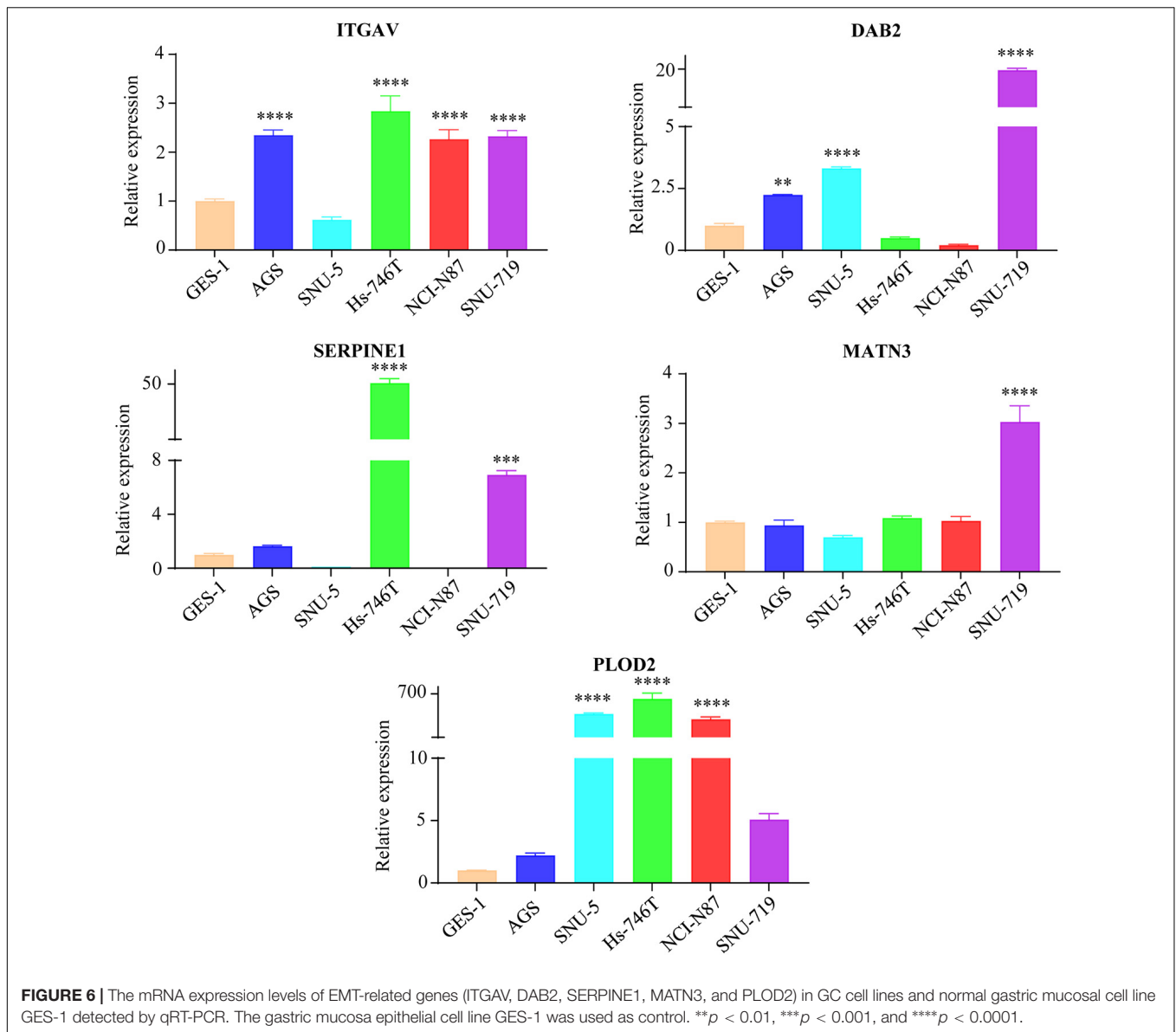
Gastric cancer remains a great challenge for public health worldwide and its OS is still not satisfactory. More and more



attention was paid to the role of EMT in OS prediction (Tan et al., 2014; Cao et al., 2020). It has been revealed that EMT could lead to drug resistance in breast cancer, lung cancer and GC, and metastasis in bladder cancer and GC, which may be the reasons why EMT could contribute to a worse OS (Huang J. et al., 2015; Tao et al., 2020; Tian et al., 2020; Wang et al., 2020;

Zhang et al., 2020). Many single potential prognostic genes that associated with EMT in GC have been reported by researchers. But as we know, EMT is a complex process that is triggered by many genes. So integration of these genes tends to be significant for understanding the process of EMT. Thanks to the rapid improvements in sequencing techniques, mining the



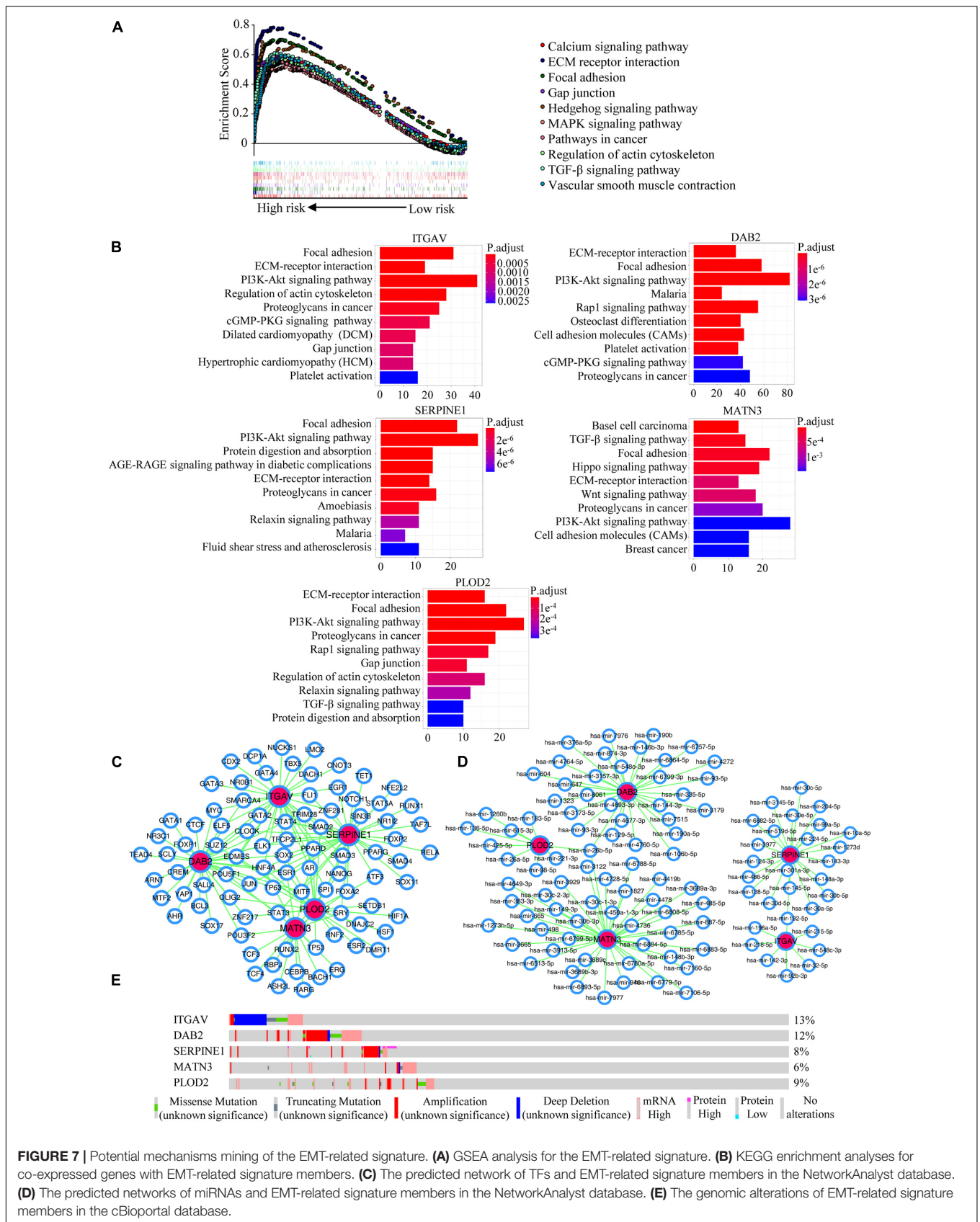


gene signatures from transcriptional profiles for individual risk stratification of patients with cancer has flourished. Combining the gene signature with clinical parameters has been highlighted when predicting the survival and considering individualized treatment for patients.

Similar to our study, Cao et al. have built an EMT-related gene signature that might facilitate risk stratification of patients and personalized treatment in bladder cancer (Cao et al., 2020). Besides, Tan et al. developed a generic EMT signature to estimate extent of EMT in several kinds of tumors, showing that EMT is linked to OS in ovarian cancer, glioblastoma and GC, but given that GC is a specific cancer different from others, a signature specific to GC may be a better choice for OS prediction (Tan et al., 2014). Zhu et al. constructed another prognostic and predictive classifier for GC, and intriguingly, the high-score group was related to EMT subtype, suggesting the importance of EMT in

risk stratification (Zhu et al., 2018). Therefore, a comprehensive signature of EMT-related gene was necessary to be built for outcome prediction of GC patients.

In this study, we developed a novel five-gene signature related to EMT which included ITGAV, DAB2, SERPINE1, MATN3 and PLOD2. EMT is a complicated and sophisticated biological process involving many pathways. The GSEA and pathway enrichment analysis in our study revealed that focal adhesion and ECM-receptor interaction might be the two important significantly enriched EMT-related pathways to the signature. Focal adhesion signaling events play essential roles in reorganizing the actin cytoskeleton, changing cell shape and motility, and regulating cell proliferation, differentiation and survival (Petit and Thiery, 2000). Extracellular matrix (ECM), constituting the main part of the extracellular microenvironment, can directly interact with cells, regulating cell growth, migration,



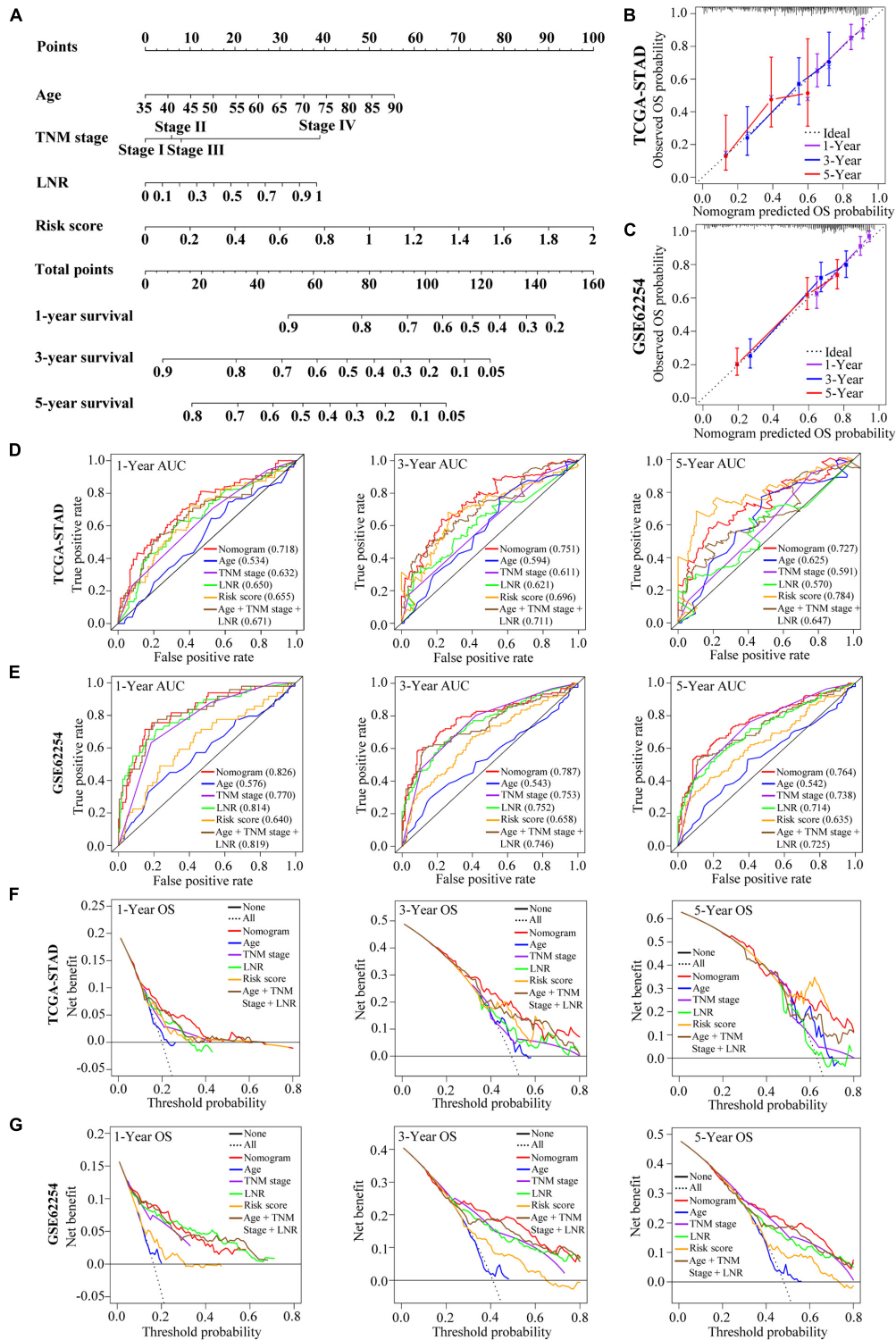


FIGURE 8 | Construction and validation of a nomogram for OS prediction in GC. **(A)** The nomogram consisting of age, TNM stage, LNR and the risk score. **(B,C)** Calibration curves of the nomogram for 1-, 3-, and 5-year OS prediction in the TCGA cohort **(B)** and the GEO cohort **(C)**. The X-axis represents the nomogram-predicted outcome while Y-axis represents the observed outcome. The 45° line represents the best prediction. **(D,E)** Time-dependent ROC curves of the nomogram, age, TNM stage, LNR, risk score and age + TNM stage + LNR model for 1-, 3-, and 5-year OS prediction in the TCGA cohort **(D)** and the GEO cohort **(E)**. **(F,G)** DCA curves of the nomogram, age, TNM stage, LNR, risk score and age + TNM stage + LNR model for 1-, 3-, and 5-year OS prediction in the TCGA cohort **(F)** and the GEO cohort **(G)**. The plots show the expected net benefits at different threshold probability. The black line “None” represents the assumption that event will happen in no patients while the dash line “All” represents the assumption that event will happen in all patients.

proliferation, differentiation, metabolism, and function by integrin or other cell surface receptors (Yang et al., 2020). The results of GSEA and pathway enrichment analysis suggested the potential mechanisms involved in our signature, providing us direction for further experiment research in the future. All of members in the signature were negative predictors of OS in our signature, and they all have been reported in cancers. ITGAV belongs to the integrin family of extracellular matrix receptors, functioning in cell surface adhesion and signaling. Suppression of ITGAV inhibited cell growth, invasion, and self-renewal of breast cancer by altering BCL2 and PXN levels (Cheuk et al., 2020). Evidences have been reported that it could promote growth, migration, and invasion of GC cells, and was positively associated with lymph node metastasis (Wang et al., 2019a). DAB2, initially known as DOC-2, was considered to be a tumor suppressor because of its absence in 85% of ovarian cancer (Fazili et al., 1999). However, Chao et al. suggested that upregulation of DAB2 could promote EMT by inhibiting E-cadherin while stimulating vimentin and phospho-FAK, indicating the significance of DAB2 in EMT (Chao et al., 2012). In human gastric carcinomas, DAB2⁺ tumor-associated macrophages correlated with a poor clinical outcome (Marigo et al., 2020). SERPINE1, an inhibitor of tissue plasminogen activator and urokinase, is a fibrinolytic inhibitor. It was validated that SERPINE1 could promote migration and invasion by regulating EMT in GC (Yang et al., 2019). What's more, it was identified as prognostic biomarker for GC by bioinformatics, consistent with our study (Li et al., 2019; Xu et al., 2019). MATN3 encodes a protein which belongs to von Willebrand factor A domain containing protein family related to the formation of filamentous networks in the extracellular matrices of various tissues (Wagener et al., 1997). It was verified that MATN3 protein was upregulated in gastric adenocarcinoma, acting as a predictor of poor prognosis (Wu et al., 2018). MATN3 has been used for previous prognostic models to predict recurrence for GC patients, indicating the vital performance of MATN3 in GC (Lee et al., 2014; Zhou et al., 2018). But the mechanism of MATN3 in GC is not yet clear. PLOD2 is a kind of enzyme that catalyzes the hydroxylation of lysyl residues in collagen-like peptides (Qi and Xu, 2018). PLOD2 has been shown to promote metastasis in cancer such as breast cancer, biliary tract cancer and lung cancer (Du et al., 2017; He et al., 2018; Okumura et al., 2018), etc. Besides, PLOD2 was reported to play an important role in peritoneal dissemination of GC, and it was regulated by hypoxia-inducible factor-1 (HIF-1) and involved in extracellular matrix remodeling, alignment and mechanical properties (Kiyozumi et al., 2018). According to the investigations above, five genes in the EMT-related signature have an important impact on the carcinogenesis and tumor progression. In our study, by using qRT-PCR assays, we could find the high mRNA expression of the five EMT-related genes in the GC cell lines. Probably, if the mechanisms of five genes in GC process are explored deeply and widely, they can better serve as biomarkers for GC.

The EMT-related gene signature with five prognostic genes was constructed by applying univariate Cox regression and LASSO regression. LASSO regression is a method which can reduce the risk of overfitting in the model, and it was used to

improve the predictor selection in our signature. Then, patients were divided into high- and low-risk groups based on the signature. With external validation, the ROC curve and survival analysis showed that the signature performed well and the high-risk patients had poorer OS. Univariate and multivariate Cox regression indicated that the signature could be an independent factor to predict OS. In order to improve the signature's ability of OS prediction, we built a nomogram which combined the signature with clinical parameters according to the variables of multivariate Cox regression above. And it was assessed by C-index, ROC curve, calibration curve, and DCA. External validation was conducted to verify the prognostic value of the combined nomogram. In validation cohort, ROC curves showed that the nomogram (using age, TNM stage, and LNR as well as EMT risk score) had a better discrimination than age + TNM stage + LNR model without EMT risk score. Perfect agreement could be seen when comparing predictive survival outcome with the actual outcome in the calibration curve. DCA curves indicated that the nomogram might have good clinical usefulness for 3-, 5-year OS prediction and the modest additional net benefit for 3- or 5-year OS probability from using the nomogram instead of clinical model without EMT risk score might be explained by the EMT risk score. What's more, based on the smallest value for the AIC statistic, the nomogram (combining age, TNM stage, LNR and EMT risk score) might be the optimal model. Thus, combining the prognostic signature and clinical characteristics may improve prognostication for GC to some extent, suggesting the prognostic signature's and nomogram's potential application values for individual risk stratification in clinic. What's more, it provides a new perspective for covering the insufficiency of current staging system.

Several limitations should also be noticed in our study. Firstly, our study was a retrospective study based on two public datasets in which most patients are Asian and White, and because of geographically variation, extending our findings to more other ethnic cohorts is necessary. Secondly, though evidences were provided by our study that the five-gene signature was a significantly predictor for GC survival, underlying mechanisms between genes of the signature and GC are not clear enough. Further experiment researches of five-gene signature in lab are crucial before clinical use. Thirdly, more independent cohorts are needed to validate the prognostic signature and nomogram. Fourthly, TCGA-STAD dataset recorded cases' original staging, which, over time, reflected AJCC different editions. Because of the incomplete detailed descriptions for staging, the standardization for TNM staging was difficult. We hope that this concern will be resolved in the future for more accurate modeling. Fifthly, resection quality at the time of surgery is an important prognostic factor in GC, but insufficient information on resection quality of cohorts in our study resulted in our omission with this consideration. Thus, further well-designed, prospective, international studies are necessary to verify our findings.

In summary, EMT is vital to malignant progression and associated with poor OS of patients with GC. Here, we identified an EMT-related gene signature and a combined nomogram to predict OS of GC, which can add clinical value to traditional staging system for predicting OS, and might

facilitate individualized treatment and clinical decision-making for GC patients.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: TCGA database: <https://portal.gdc.cancer.gov/>; GEO database: <https://www.ncbi.nlm.nih.gov/geo/>.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

JW, LX, and XW designed and conceived this study. WD, YX, and JZ collected and analyzed the data. WT, JLin, LH, MP, JLi, and SL performed statistical analyses. WD and YX contributed to writing and revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant/Award numbers: 81772964, 81974448, and 82073066), National Major New Drug Creation Science and Technology Major Special Fund Funding Project (Grant/Award number: 2020ZX09201017), Science and Technology Planning Project of Guangdong Province (Grant/Award number: 2017B020209003), Guangdong Medical

Research Foundation (Grant/Award number: B2019126), and Shenzhen Science and Technology Innovation Commission (Grant/Award number: JCYJ20180306170328854).

ACKNOWLEDGMENTS

We acknowledge the generous support of the Guangdong Provincial Key Laboratory of Gastroenterology, Department of Gastroenterology, Nanfang Hospital, Southern Medical University and the Department of Gastroenterology, Longgang District People's Hospital, Shenzhen.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.661306/full#supplementary-material>

Supplementary Figure 1 | Correlation plot of 10 prognostic EMT-related genes in the TCGA cohort.

Supplementary Figure 2 | Confirmation of the signature for OS prediction in clinical subgroups stratified according to gender. **(A,B)** Kaplan-Meier estimates of OS based on the EMT-related gene signature in subgroups classified by gender in the TCGA-STAD **(A)** and the GSE62254 **(B)** cohorts.

Supplementary Table 1 | List of the primers used for qRT-PCR.

Supplementary Table 2 | 123 significantly differentially expressed EMT-related genes.

Supplementary Table 3 | Multivariate Cox regression of TCGA-STAD and GSE62254 cohorts with variables including age, gender, stage, LNR and risk score.

Supplementary Table 4 | Clinical characteristics of TCGA-STAD cohort patients in different risk groups.

Supplementary Table 5 | Clinical characteristics of GSE62254 cohort patients in different risk groups.

Supplementary Table 6 | The profiles of the EMT-related gene signature based nomogram and models with different factors' combinations.

REFERENCES

- Bernards, N., Creemers, G. J., Nieuwenhuijzen, G. A., Bosscha, K., Puijck, J. F., and Lemmens, V. E. (2013). No improvement in median survival for patients with metastatic gastric cancer despite increased use of chemotherapy. *Ann. Oncol.* 24, 3056–3060. doi: 10.1093/annonc/mdt401
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492
- Cao, R., Yuan, L., Ma, B., Wang, G., Qiu, W., and Tian, Y. (2020). An EMT-related gene signature for the prognosis of human bladder cancer. *J. Cell Mol. Med.* 24, 605–617. doi: 10.1111/jcmm.14767
- Chao, A., Lin, C. Y., Lee, Y. S., Tsai, C. L., Wei, P. C., Hsueh, S., et al. (2012). Regulation of ovarian cancer progression by microRNA-187 through targeting Disabled homolog-2. *Oncogene* 31, 764–775. doi: 10.1038/onc.2011.269
- Cheuk, I. W., Siu, M. T., Ho, J. C., Chen, J., Shin, V. Y., and Kwong, A. (2020). ITGAV targeting as a therapeutic approach for treatment of metastatic breast cancer. *Am. J. Cancer Res.* 10, 211–223.
- Cristescu, R., Lee, J., Nebozhyn, M., Kim, K. M., Ting, J. C., Wong, S. S., et al. (2015). Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat. Med.* 21, 449–456. doi: 10.1038/nm.3850
- Dongre, A., and Weinberg, R. A. (2019). New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nat. Rev. Mol. Cell Biol.* 20, 69–84. doi: 10.1038/s41580-018-0080-4
- Du, H., Chen, Y., Hou, X., Huang, Y., Wei, X., Yu, X., et al. (2017). PLOD2 regulated by transcription factor FOXA1 promotes metastasis in NSCLC. *Cell Death Dis.* 8:e3143. doi: 10.1038/cddis.2017.553
- Fazili, Z., Sun, W., Mittelstaedt, S., Cohen, C., and Xu, X. X. (1999). Disabled-2 inactivation is an early step in ovarian tumorigenicity. *Oncogene* 18, 3104–3113. doi: 10.1038/sj.onc.1202649
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33, 1–22.
- Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* 6:11. doi: 10.1126/scisignal.2004088
- Harrell, F. E. Jr. (2019). *rms: Regression Modeling Strategies. R package version 5.1-4*. Available online at: <https://CRAN.R-project.org/package=rms> (accessed March 25, 2020).

- He, J. Y., Wei, X. H., Li, S. J., Liu, Y., Hu, H. L., Li, Z. Z., et al. (2018). Adipocyte-derived IL-6 and leptin promote breast Cancer metastasis via upregulation of Lysyl Hydroxylase-2 expression. *Cell Commun. Signal.* 16:100. doi: 10.1186/s12964-018-0309-z
- Heagerty, P. J., Lumley, T., and Pepe, M. S. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 56, 337–344. doi: 10.1111/j.0006-341x.2000.00337.x
- Hsu, S. D., Tseng, Y. T., Shrestha, S., Lin, Y. L., Khaleel, A., Chou, C. H., et al. (2014). miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Res.* 42, D78–D85. doi: 10.1093/nar/gkt1266
- Huang, J., Li, H., and Ren, G. (2015). Epithelial-mesenchymal transition and drug resistance in breast cancer (Review). *Int. J. Oncol.* 47, 840–848. doi: 10.3892/ijo.2015.3084
- Huang, L., Wu, R. L., and Xu, A. M. (2015). Epithelial-mesenchymal transition in gastric cancer. *Am. J. Transl. Res.* 7, 2141–2158.
- Jiang, Y., Li, T., Liang, X., Hu, Y., Huang, L., Liao, Z., et al. (2017). Association of adjuvant chemotherapy with survival in patients with stage II or III gastric cancer. *JAMA Surg.* 152:e171087. doi: 10.1001/jamasurg.2017.1087
- Jiang, Y., Zhang, Q., Hu, Y., Li, T., Yu, J., Zhao, L., et al. (2018). ImmunoScore signature: a prognostic and predictive tool in gastric cancer. *Ann. Surg.* 267, 504–513. doi: 10.1097/SLA.0000000000002116
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27
- Kassambara, A., Kosinski, M., and Pieck, P. (2019). *survminer: Drawing Survival Curves using ggplot2*. R package version 0.4.6. Available online at: <https://CRAN.R-project.org/package=survminer> (accessed March 17, 2020).
- Kiyozumi, Y., Iwatsuki, M., Kurashige, J., Ogata, Y., Yamashita, K., Koga, Y., et al. (2018). PLOD2 as a potential regulator of peritoneal dissemination in gastric cancer. *Int. J. Cancer* 143, 1202–1211. doi: 10.1002/ijc.31410
- Kolde, R. (2019). *heatmap: Pretty Heatmaps*. R package version 1.0.12. Available online at: <https://CRAN.R-project.org/package=heatmap> (accessed March 16, 2020).
- Lachmann, A., Xu, H., Krishnan, J., Berger, S. I., Mazloom, A. R., and Ma'ayan, A. (2010). ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics.* 26, 2438–2444. doi: 10.1093/bioinformatics/btq466
- Lamouille, S., Xu, J., and Derynck, R. (2014). Molecular mechanisms of epithelial-mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* 15, 178–196. doi: 10.1038/nrm3758
- Lee, J., Sohn, I., Do, I. G., Kim, K. M., Park, S. H., Park, J. O., et al. (2014). Nanostring-based multigene assay to predict recurrence for gastric cancer patients after surgery. *PLoS One* 9:e90133. doi: 10.1371/journal.pone.0090133
- Lee, Y. C., Yang, P. J., Zhong, Y., Clancy, T. E., Lin, M. T., and Wang, J. (2017). Lymph node ratio-based staging system outperforms the seventh AJCC system for gastric cancer: validation analysis with National Taiwan University Hospital Cancer Registry. *Am. J. Clin. Oncol.* 40, 35–41. doi: 10.1097/COC.0000000000000110
- Leek, J. T., Johnson, W. E., Parker, H. S., Fertig, E. J., Jaffe, A. E., Storey, J. D., et al. (2019). *sva: Surrogate Variable Analysis*. R package version 3.34.0. Available online at: <http://www.bioconductor.org/packages/release/bioc/html/sva.html> (accessed March 16, 2020)
- Li, L., Zhu, Z., Zhao, Y., Zhang, Q., Wu, X., Miao, B., et al. (2019). FN1, SPARC, and SERPINE1 are highly expressed and significantly related to a poor prognosis of gastric adenocarcinoma revealed by microarray and bioinformatics. *Sci. Rep.* 9:7827. doi: 10.1038/s41598-019-43924-x
- Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1, 417–425. doi: 10.1016/j.cels.2015.12.004
- Marigo, I., Trovato, R., Hofer, F., Ingangi, V., Desantis, G., Leone, K., et al. (2020). Disabled Homolog 2 controls prometastatic activity of tumor-associated macrophages. *Cancer Discov.* 10, 1758–1773. doi: 10.1158/2159-8290.CD-20-0036
- Murai, T., Yamada, S., Fuchs, B. C., Fujii, T., Nakayama, G., Sugimoto, H., et al. (2014). Epithelial-to-mesenchymal transition predicts prognosis in clinical gastric cancer. *J. Surg. Oncol.* 109, 684–689. doi: 10.1002/jso.23564
- Okumura, Y., Noda, T., Eguchi, H., Sakamoto, T., Iwagami, Y., Yamada, D., et al. (2018). Hypoxia-induced PLOD2 is a key regulator in epithelial-mesenchymal transition and chemoresistance in biliary tract cancer. *Ann. Surg. Oncol.* 25, 3728–3737. doi: 10.1245/s10434-018-6670-8
- Peinado, H., Olmeda, D., and Cano, A. (2007). Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? *Nat. Rev. Cancer.* 7, 415–428. doi: 10.1038/nrc2131
- Petit, V., and Thiery, J. P. (2000). Focal adhesions: structure and dynamics. *Biol. Cell* 92, 477–494. doi: 10.1016/s0248-4900(00)01101-1
- Qi, Y., and Xu, R. (2018). Roles of PLODs in collagen synthesis and cancer progression. *Front. Cell Dev. Biol.* 6:66. doi: 10.3389/fcell.2018.00066
- Rhodes, D. R., Kalyana-Sundaram, S., Mahavisno, V., Varambally, R., Yu, J., Briggs, B. B., et al. (2007). OncoPrint 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* 9, 166–180. doi: 10.1593/neo.07112
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Sleeman, J., and Steeg, P. S. (2010). Cancer metastasis as a therapeutic target. *Eur. J. Cancer* 46, 1177–1180. doi: 10.1016/j.ejca.2010.02.039
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102
- Tan, T. Z., Miow, Q. H., Miki, Y., Noda, T., Mori, S., Huang, R. Y., et al. (2014). Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO Mol. Med.* 6, 1279–1293. doi: 10.15252/emmm.201404208
- Tao, L., Shu-Ling, W., Jing-Bo, H., Ying, Z., Rong, H., Xiang-Qun, L., et al. (2020). MiR-451a attenuates doxorubicin resistance in lung cancer via suppressing epithelial-mesenchymal transition (EMT) through targeting c-Myc. *Biomed. Pharmacother.* 125:109962. doi: 10.1016/j.biopha.2020.109962
- Therneau, T. M. (2020). *A Package for Survival Analysis in R*. R package version 3.1-11. Available online at: <https://CRAN.R-project.org/package=survival> (accessed March 17, 2020).
- Therneau, T. M., and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. New York, NY: Springer.
- Thomassen, I., van Gestel, Y. R., van Ramshorst, B., Luyer, M. D., Bosscha, K., Nienhuijs, S. W., et al. (2014). Peritoneal carcinomatosis of gastric origin: a population-based study on incidence, survival and risk factors. *Int. J. Cancer* 134, 622–628. doi: 10.1002/ijc.28373
- Tian, S., Peng, P., Li, J., Deng, H., Zhan, N., Zeng, Z., et al. (2020). SERPINH1 regulates EMT and gastric cancer metastasis via the Wnt/beta-catenin signaling pathway. *Aging (Albany NY)* 12, 3574–3593. doi: 10.18632/aging.102831
- Uhlen, M., Bjorling, E., Agaton, C., Szegedy, C. A., Amini, B., Andersen, E., et al. (2005). A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell Proteomics.* 4, 1920–1932. doi: 10.1074/mcp.M500279-MCP200
- Uhlen, M., Zhang, C., Lee, S., Sjostedt, E., Fagerberg, L., Bidkhori, G., et al. (2017). A pathology atlas of the human cancer transcriptome. *Science* 357:eaan2507. doi: 10.1126/science.aan2507
- Vickers, A. J., and Elkin, E. B. (2006). Decision curve analysis: a novel method for evaluating prediction models. *Med. Decis. Making* 26, 565–574. doi: 10.1177/0272989X06295361
- Wagner, R., Kobbe, B., and Paulsson, M. (1997). Primary structure of matrilin-3, a new member of a family of extracellular matrix proteins related to cartilage matrix protein (matrilin-1) and von Willebrand factor. *FEBS Lett.* 413, 129–134. doi: 10.1016/s0014-5793(97)00895-8
- Wang, H., Chen, H., Jiang, Z., Lin, Y., Wang, X., Xiang, J., et al. (2019a). Integrin subunit alpha V promotes growth, migration, and invasion of gastric cancer cells. *Pathol. Res. Pract.* 215:152531. doi: 10.1016/j.prp.2019.152531
- Wang, H., Wu, X., and Chen, Y. (2019b). Stromal-immune score-based gene signature: a prognosis stratification tool in gastric cancer. *Front. Oncol.* 9:1212. doi: 10.3389/fonc.2019.01212

- Wang, X., Xu, Z., Sun, J., Lv, H., Wang, Y., Ni, Y., et al. (2020). Cisplatin resistance in gastric cancer cells is involved with GPR30-mediated epithelial-mesenchymal transition. *J. Cell Mol. Med.* 24, 3625–3633. doi: 10.1111/jcmm.15055
- Wei, T., and Simko, V. (2017). R package “corrplot”: Visualization of a Correlation Matrix. Version 0.84. Available online at: <https://github.com/taiyun/corrplot> (accessed March 17, 2020).
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag New York.
- Wu, P. L., He, Y. F., Yao, H. H., and Hu, B. (2018). Matriline-3 (MATN3) overexpression in gastric adenocarcinoma and its prognostic significance. *Med. Sci. Monit.* 24, 348–355. doi: 10.12659/msm.908447
- Xu, B., Bai, Z., Yin, J., and Zhang, Z. (2019). Global transcriptomic analysis identifies SERPINE1 as a prognostic biomarker associated with epithelial-to-mesenchymal transition in gastric cancer. *PeerJ* 7:e7091. doi: 10.7717/peerj.7091
- Yang, J. D., Ma, L., and Zhu, Z. (2019). SERPINE1 as a cancer-promoting gene in gastric adenocarcinoma: facilitates tumour cell proliferation, migration, and invasion by regulating EMT. *J. Chemother.* 31, 408–418. doi: 10.1080/1120009X.2019.1687996
- Yang, X., Chen, L., Mao, Y., Hu, Z., and He, M. (2020). Progressive and prognostic performance of an extracellular matrix-receptor interaction signature in gastric cancer. *Dis. Markers* 2020:8816070. doi: 10.1155/2020/8816070
- Yu, G. (2019). *enrichplot: Visualization of Functional Enrichment Result*. R package version 1.6.1. Available online at: <https://github.com/GuangchuangYu/enrichplot> (accessed March 21, 2020).
- Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118
- Zhang, M., Du, H., Wang, L., Yue, Y., Zhang, P., Huang, Z., et al. (2020). Thymoquinone suppresses invasion and metastasis in bladder cancer cells by reversing EMT through the Wnt/beta-catenin signaling pathway. *Chem. Biol. Interact.* 320:109022. doi: 10.1016/j.cbi.2020.109022
- Zhao, L., Jiang, L., He, L., Wei, Q., Bi, J., Wang, Y., et al. (2019). Identification of a novel cell cycle-related gene signature predicting survival in patients with gastric cancer. *J. Cell Physiol.* 234, 6350–6360. doi: 10.1002/jcp.27365
- Zhao, L. Y., Li, C. C., Jia, L. Y., Chen, X. L., Zhang, W. H., Chen, X. Z., et al. (2016). Superiority of lymph node ratio-based staging system for prognostic prediction in 2575 patients with gastric cancer: validation analysis in a large single center. *Oncotarget* 7, 51069–51081. doi: 10.18632/oncotarget.9714
- Zhou, G., Soufan, O., Ewald, J., Hancock, R. E. W., Basu, N., and Xia, J. (2019). NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res.* 47, W234–W241. doi: 10.1093/nar/gkz240
- Zhou, Y. Y., Kang, Y. T., Chen, C., Xu, F. F., Wang, H. N., and Jin, R. (2018). Combination of TNM staging and pathway based risk score models in patients with gastric cancer. *J. Cell Biochem.* 119, 3608–3617. doi: 10.1002/jcb.26563
- Zhu, X., Tian, X., Sun, T., Yu, C., Cao, Y., Yan, T., et al. (2018). GeneExpressScore Signature: a robust prognostic and predictive classifier in gastric cancer. *Mol. Oncol.* 12, 1871–1883. doi: 10.1002/1878-0261.12351

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Dai, Xiao, Tang, Li, Hong, Zhang, Pei, Lin, Liu, Wu, Xiang and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.