




# Construction and Validation of a Nomogram for the Preoperative Prediction of Lymph Node Metastasis in Gastric Cancer

Shilong Li<sup>1,2</sup>, Zongxian Zhao<sup>1,2</sup>, Huaxiang Yang<sup>1,2</sup>, Daohan Wang<sup>1,2</sup>,  
Weilin Sun<sup>1,2</sup>, Shuliang Li<sup>1,3</sup>, Zhaoxiong Zhang<sup>1,2</sup>, and Weihua Fu, MD<sup>1,2</sup> 

## Abstract

**Background:** Increasing evidence indicated that the tumor microenvironment (TME) plays a critical role in tumor progression. This study aimed to identify and evaluate mRNA signature involved in lymph node metastasis (LNM) in TME for gastric cancer (GC).

**Methods:** Gene expression and clinical data were downloaded from The Cancer Genome Atlas (TCGA). The ESTIMATE algorithm was used to evaluate the TME of GC. The heatmap and Venn plots were applied for visualizing and screening out intersect differentially expressed genes (DEGs) involved in LNM in TME. Functional enrichment analysis, gene set enrichment analysis (GSEA) and protein-protein interaction (PPI) network were also conducted. Furthermore, binary logistic regression analysis were employed to develop a 4-mRNAs signature for the LNM prediction. ROC curves were applied to validate the LNM predictive ability of the riskscore. Nomogram was constructed and calibration curve was plotted to verify the predictive power of nomogram.

**Results:** A total of 88 LNM related DEGs were identified. Functional enrichment analysis and GSEA implied that those genes were associated with some biological processes, such as ion transportation, lipid metabolism and thiolester hydrolase activity. After univariate and multivariate logistic regression analysis, 4 mRNAs (RASSF2, MS4A2, ANKRD33B and ADH1B) were eventually screened out to develop a predictive model. ROC curves manifested the good performance of the 4-mRNAs signature. The proportion of patients with LNM in high-risk group was significantly higher than that in low-risk group. The C-index of nomogram from training and test cohorts were 0.865 and 0.765, and the nomogram was well calibrated.

**Conclusions:** In general, we identified a 4-mRNAs signature that effectively predicted LNM in GC patients. Moreover, the 4-mRNAs signature and nomogram provide a guidance for the preoperative evaluation and postoperative treatment of GC patients.

## Keywords

nomogram, gastric cancer, tumor microenvironment, lymph node metastasis, TCGA

Received October 26, 2020. Received revised April 05, 2021. Accepted for publication May 30, 2021.

## Introduction

Gastric cancer (GC) is a common cancer with high morbidity and mortality, both worldwide and in China.<sup>1,2</sup> The time of diagnosis of GC is also getting earlier and earlier with the attention to health and the improvement of diagnostic techniques and levels. And lymph node metastasis (LNM), which occurs in about 60%-80% of GC patients,<sup>3,4</sup> is the key to diagnosis and staging of GC and plays an indicator role in the survival and prognosis of patients.

<sup>1</sup> Department of General Surgery, Tianjin Medical University General Hospital, Tianjin, China

<sup>2</sup> Tianjin General Surgery Institute, Tianjin, China

<sup>3</sup> Department of Gastrointestinal Surgery, The Second People's Hospital of Liaocheng, Shandong, China

### Corresponding Author:

Weihua Fu, Department of General Surgery, Tianjin Medical University General Hospital, 154 Anshan Road, Heping District, Tianjin 300052, China.  
Email: tjmughgs\_fwh@163.com



Preoperative evaluation of LNM by endoscopic ultrasound (EUS), CT, PET/CT and multi-detector row CT (MDCT) can effectively improve clinical staging.<sup>5-9</sup> However, the diagnostic accuracy of EUS varies from operator to operator, ranging from 30% to 90% for N staging. In addition, sensitivity and specificity of EUS diagnosis for N staging range from 16.6% to 96.8% and 57.1% to 100%, respectively.<sup>10</sup> Besides, CT is a routine examination for preoperative evaluation of staging. The sensitivity and specificity of CT in the detection of LNM are 78% and 62%, respectively. However, PET/CT is also a detection measure. Comparing with CT, PET/CT has a relatively low detection rate with lower sensitivity (56%), but higher specificity (92%) in the detection of LNM.<sup>11</sup> Recently, MDCT is becoming a standard imaging modality for the staging of GC due to its superior spatial resolution.<sup>8</sup> In addition, although MDCT has sufficient predictive power to assess the status of lymph node involvement in serosa-invasive GC, its predictive power is very limited in non-serosa-invasive GC.<sup>9</sup> These detection methods are limited due to their sensitivity, specificity and unstable predictive accuracy. Although there are many ways to diagnose LNM, it is still difficult to accurately determine the status of lymph node involvement before surgery. Therefore, it is important to find a more objective and stable method to identify the status of LNM.

TME is a complex environmental conditions around the tumor, consisting of endothelial cells, cancer associated fibroblasts (CAFs), immune and inflammatory cells, mesenchymal cells, as well as the extracellular matrix (ECM).<sup>12</sup> The interaction between cells and microenvironment plays an important role in maintaining normal tissue homeostasis and tumor progression.<sup>13</sup> And various kinds of stromal cells were nested around the tumors, which promoted the growth and metastatic dissemination of tumors. However, tumor cells often disseminate to other microenvironments, such as lymph nodes and bone marrow, before metastasizing to future sites of metastasis.<sup>14</sup> Therefore, LNM is more likely to be a precursor of distant metastasis and an important indicator of tumor progression. However, the understanding of mechanism of the TME involved in LNM is far from enough. More recently, ESTIMATE algorithm is applied for the evaluation of TME in various tumors, such as acute myeloid leukemia (AML),<sup>15</sup> clear cell renal cell carcinoma (ccRCC)<sup>16</sup> and glioma.<sup>17</sup> Hence, we intend to apply this algorithm to evaluate and explore the connection between TME and LNM in GC.

In this article, the potential mechanisms involved in LNM in GC were revealed. More importantly, we identify a novel mRNAs signature associated with TME for LNM prediction and construct nomogram to predict the incidence of LNM before surgery in GC.

## Material and Methods

### *The GC Patients Dataset and TME Scores Calculation*

The gene transcriptome and clinical profiles of 343 stomach adenocarcinoma (STAD) patients from The Cancer Genome Atlas (TCGA) database were downloaded using GDC tool.

And we processed the gene expression data and extracted corresponding clinical information, such as age, tumor grade, AJCC stage, TNM staging and AJCC stage. Only those patients with N staging and mRNA expression data were enrolled in the study. After excluding 16 patients with N staging deletion, the remaining GC patients were included (n = 327). Scores of immune, stromal and Estimate were calculated by ESTIMATE algorithm according to those patients mRNA expression.

### *Acquisition of Differentially Expressed Genes (DEGs), Heatmaps and Clustering Analysis*

Those enrolled patients were divided into N<sup>-</sup> (without LNM, n = 102) and N<sup>+</sup> (with LNM, n = 225) group based on N staging. Meanwhile, the GC patients were also divided into high- and low-score groups according to the median immune and stromal scores, respectively. The “limma” package was used for the standardization of transcriptome data. Then genes that differed between the N<sup>-</sup> and N<sup>+</sup> groups and between the low- and high-score groups were also screened. In addition, the clustering analysis was applied to identify significant up and down gene sets between the subgroups of N staging. And heatmaps were plotted to illustrate the DEGs using “pheatmap” package.

### *Functional Enrichment Analysis and Protein-Protein Interaction (PPI) Network Construction*

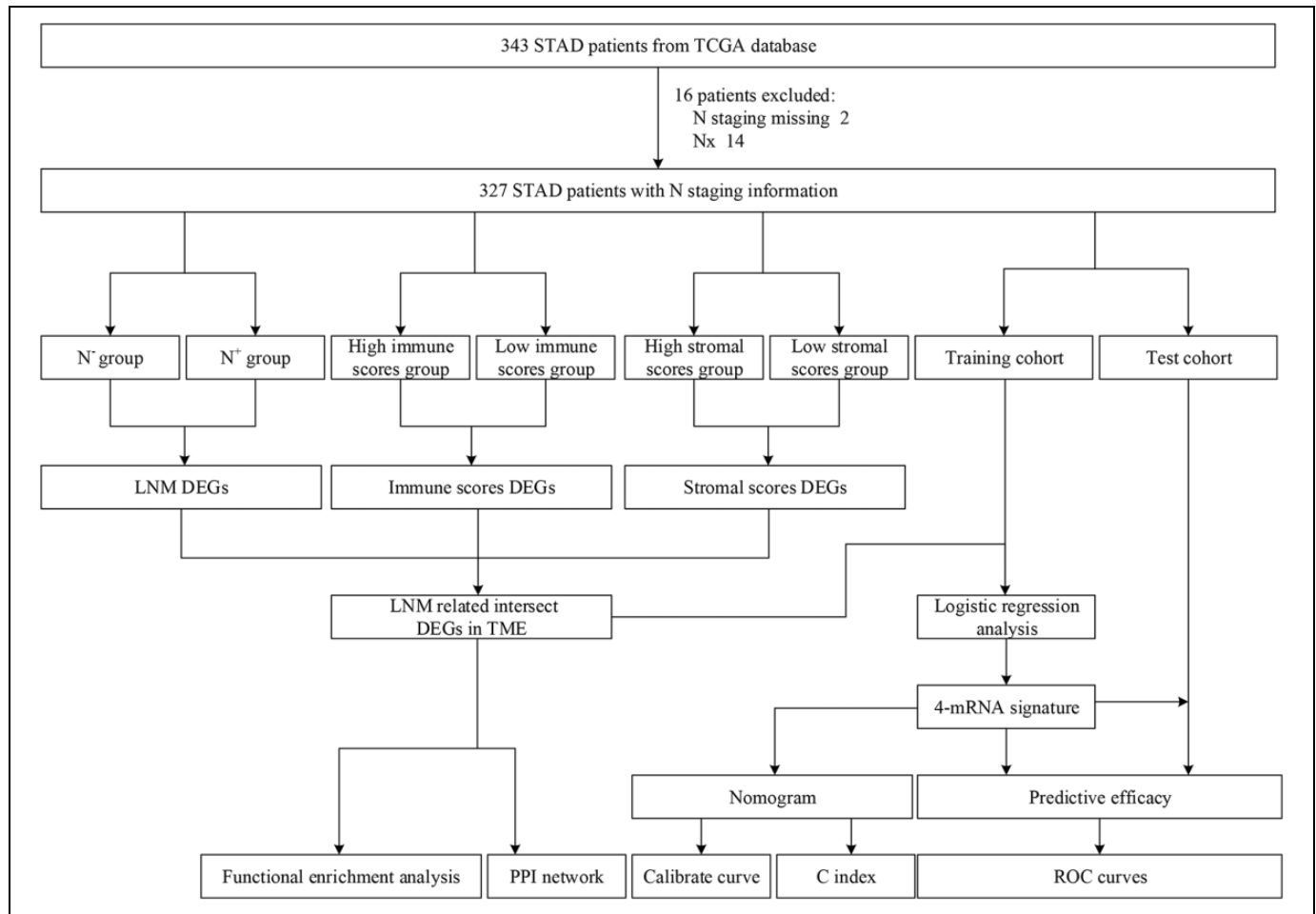
DEGs acquired from N staging were applied for functional enrichment analysis to explore potential LNM mechanism. And cellular components (CC), and molecular functions (MF) of Gene Ontology (GO) analysis were performed using “clusterProfiler,” “org.Hs.eg.db,” “enrichplot,” and “ggplot2” R packages. Besides, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was conducted based on the same packages as GO analysis. At the same time, the STRING database were used to construct the PPI network based on DEGs with medium confidence (0.400).

### *Gene Set Enrichment Analysis (GSEA)*

To further explore the underlying mechanisms involved in LNM, GSEA was employed to assess the related pathways and molecular mechanisms in GC. The gene sets with |NES| >1.5, P-value <0.05 and false discovery rate (FDR) <0.35 were considered statistically significant.

### *Acquisition of Intersect Genes, Logistic Regression Analysis and ROC Curves*

The intersect genes between DEGs based on N staging, stromal and immune scores were identified and visualized using “VennDiagram” package. The relationship between intersect genes and LNM was determined by univariate logistic regression analysis and those mRNAs with P < 0.05 were considered as closely related to LNM and applied for further multivariate logistic regression analysis. Subsequently, a formula was built



**Figure 1.** The flowchart of identifying the 4-mRNAs signature and construction and validation of the nomogram for lymph node metastasis.

to calculate the riskscore for each patient based on the expression level of mRNAs ( $Exp_i$ ) and the coefficients ( $Coe_i$ ). The ROC analysis was used to compare the predictive ability of riskscore and those mRNAs with  $P < 0.05$  in multivariate logistic regression.

$$\text{Riskscore(RS)} = \sum_{i=1}^n Exp_i \times Coe_i$$

### Nomogram Model Construction

The RASSF2, MS4A2, ANKRD33B and ADH1B were used to construct a nomogram model using “rms” R package based on training cohort. Calibration curves were applied to evaluate the consistency of the status of LNM between the prediction model and actual status for training and test cohorts. And C index was also calculated to assess the predictive power for both training and test cohorts.

### Statistical Analysis

IBM SPSS Statistics software (version 23.0) was applied for statistical analysis. Chi-square test and Fisher’s exact test was

applied for categorical variables, and Student’s t-test was applied for continuous variables. R software (version 4.0.2) was used to construct nomogram. And  $P < 0.05$  represented as statistically significant.

## Results

### Baseline Characteristics of Training and Test Cohorts of TCGA STAD Patients

The flowchart describing the entire process of this study was shown in Figure 1. We downloaded clinical information of 343 STAD patients from TCGA. Then 327 STAD patients with N staging information were included (16 were excluded, including 2 with N staging missing, 14 with Nx) and grouped into training ( $n = 163$ ) and test ( $n = 164$ ) cohort randomly. Eventually, after excluding those patients with missing clinical information, baseline characteristics were summarized for patients in training ( $n = 144$ ) and test ( $n = 151$ ) cohorts (Table 1). There were no significant differences in age, gender, grade, AJCC stage and TNM staging.

Subsequently, each cohort was divided into  $N^-$  and  $N^+$  groups according to the status of LNM. As shown in Table 2,

in the training cohort, there were no significant differences in age, gender and grade between  $N^-$  ( $n = 52$ ) and  $N^+$  ( $n = 92$ ) groups. However, LNM is closely related to AJCC staging ( $P < 0.001$ ), depth of tumor invasion ( $P = 0.001$ ), and distant metastasis ( $P = 0.049$ ). The proportion of stage I/II GC patients in

the  $N^-$  group (98.1%) was significantly higher than that in the  $N^+$  group (23.9%), and the proportion of stage III/IV GC patients in the  $N^-$  group (1.9%) was significantly lower than that in the  $N^+$  group (76.1%). Moreover, the proportion of T1 GC patients in the  $N^-$  group (9.6%) was significantly higher than that in the  $N^+$  group (1.1%), while the proportion of T2-4 GC patients in the  $N^-$  group (90.4%) was significantly lower than that in the  $N^+$  group (98.9%). Besides, comparing with  $N^-$  group, patients in  $N^+$  group were more likely to develop distant metastasis (7.6% vs. 0.0%). And similar results were observed in the test cohort. There were no significant differences in age, gender and grade between the  $N^-$  ( $n = 42$ ) and  $N^+$  ( $n = 109$ ) groups. Comparing with  $N^-$  group, the proportion of stage I/II GC patients in the  $N^+$  group was significantly lower (23.9% vs. 90.5%), while the proportion of stage III/IV GC patients in the  $N^+$  group was significantly higher (76.1% vs. 9.5%). And the proportion of T1 GC patients in the  $N^-$  group (19.0%) was significantly higher than that in the  $N^+$  group (1.8%), while the proportion of T2-4 GC patients in the  $N^-$  group (81.0%) was significantly lower than that in the  $N^+$  group (98.2%). Although, there was no statistical difference in M stage between the  $N^-$  and  $N^+$  groups ( $P = 0.113$ ). However, the proportion of M1 patients in  $N^+$  group (11.0%) was still higher than that in  $N^-$  group (2.4%).

**Table 1.** The Baseline Characteristics of Training and Test Cohorts.

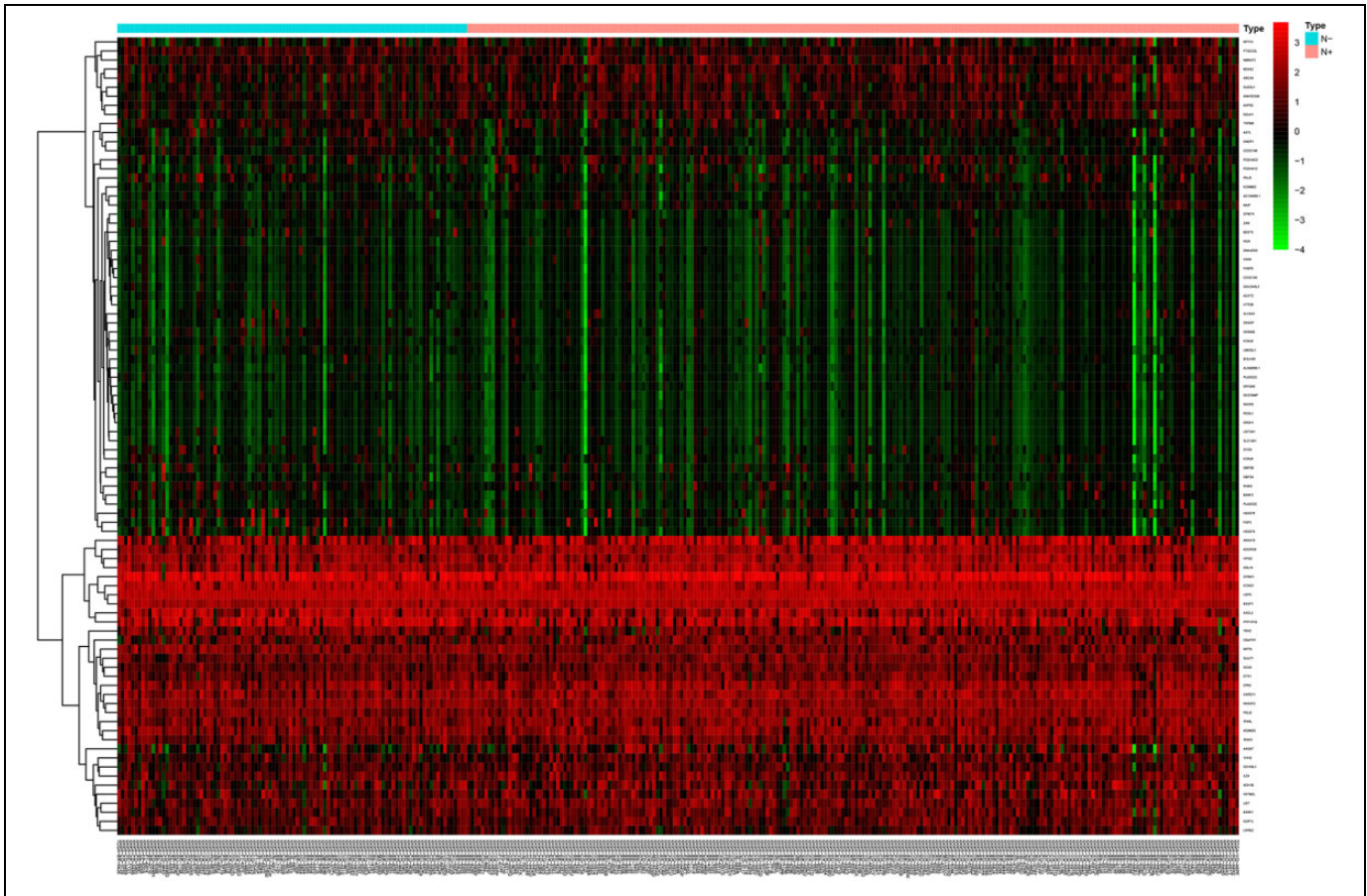
	Training cohort (n = 144)	Test cohort (n = 151)	P value
Age (yr)	66.6 ± 10.0	64.2 ± 11.4	0.060
Gender (male)	85 (59.0%)	94 (62.3%)	0.634
Grade			0.294
Highly differentiated	2 (1.4%)	3 (2.0%)	
Moderately differentiated	45 (31.3%)	59 (39.1%)	
Poorly differentiated	97 (67.4%)	89 (58.9%)	
AJCC stage			0.251
I	20 (13.9%)	23 (15.2%)	
II	53 (36.8%)	41 (27.2%)	
III	64 (44.4%)	74 (49.0%)	
IV	7 (4.9%)	13 (8.6%)	
T			0.545
1	6 (4.2%)	10 (6.6%)	
2	27 (18.8%)	35 (23.2%)	
3	74 (51.4%)	68 (45.0%)	
4	37 (25.7%)	38 (25.2%)	
N			0.190
0	52 (36.1%)	42 (27.8%)	
1	37 (25.7%)	39 (25.8%)	
2	33 (22.9%)	33 (21.9%)	
3	22 (15.3%)	37 (24.5%)	
M			0.201
0	137 (95.1%)	138 (91.4%)	
1	7 (4.9%)	13 (8.6%)	

### DEGs Between $N^-$ and $N^+$ Groups, Functional Enrichment Analysis and PPI Network

Those patients were grouped into  $N^-$  and  $N^+$  groups according to the status of LNM. The heatmap were shown in Figure 2. A total of 88 DEGs were identified with  $\log_2|\text{Fold Change}| > 0.5$

**Table 2.** The Baseline Characteristics of  $N^-$  and  $N^+$  Groups From Training and Test Cohorts.

	Training cohort			Test cohort		
	$N^-$ (n = 52)	$N^+$ (n = 92)	P value	$N^-$ (n = 42)	$N^+$ (n = 109)	P value
Age (yr)	66.1 ± 9.7	66.9 ± 10.2	0.653	66.2 ± 12.6	63.5 ± 10.8	0.188
Gender (male)	27 (51.9%)	58 (63.0%)	0.192	28 (66.7%)	66 (60.6%)	0.487
Grade			0.435			0.051
Highly differentiated	1 (1.9%)	1 (1.1%)		2 (4.8%)	1 (0.9%)	
Moderately differentiated	19 (36.5%)	26 (28.3%)		21 (50.0%)	38 (34.9%)	
Poorly differentiated	32 (61.5%)	65 (70.7%)		19 (45.2%)	70 (64.2%)	
AJCC stage			<0.001			<0.001
I	19 (36.5%)	1 (1.1%)		22 (52.4%)	1 (0.9%)	
II	32 (61.5%)	21 (22.8%)		16 (38.1%)	25 (22.9%)	
III	1 (1.9%)	63 (68.5%)		3 (7.1%)	71 (65.1%)	
IV	0 (0.0%)	7 (7.6%)		1 (2.4%)	12 (11.0%)	
T			0.001			<0.001
1	5 (9.6%)	1 (1.1%)		8 (19.0%)	2 (1.8%)	
2	14 (26.9%)	13 (14.1%)		14 (33.3%)	21 (19.3%)	
3	27 (51.9%)	47 (51.1%)		14 (33.3%)	54 (49.5%)	
4	6 (11.5%)	31 (33.7%)		6 (14.3%)	32 (29.4%)	
M			0.049			0.113
0	52 (100.0%)	85 (92.4%)		41 (97.6%)	97 (89.0%)	
1	0 (0.0%)	7 (7.6%)		1 (2.4%)	12 (11.0%)	



**Figure 2.** Comparison of gene expression profiles with the status of lymph node metastasis in GC. Heatmap was used to visualize differential expressed genes.  $N^-$  indicates GC patients without lymph node metastasis;  $N^+$ , GC patients with lymph node metastasis.

and  $P$ -value  $<0.05$  and applied for further functional enrichment analysis.

And 2 GO terms in CC and top 10 GO terms in MF were screened out (Figure 3A and B). For CC, those DEGs were only enriched 2 GO terms, including cation channel complex and basolateral plasma membrane (Figure 3A). Cation transmembrane transporter activity, metal ion transmembrane transporter activity and inorganic cation transmembrane transporter activity were top GO terms in MF (Figure 3B). In the KEGG pathway enrichment analysis, the top 9 pathways were shown in Figure 3C. Among them, pancreatic secretion, alpha-linolenic acid metabolism as well as linolenic acid metabolism were top KEGG pathways which 88 DEGs might be involved in. And PPI network was constructed based on 88 DEGs (Figure 3D).

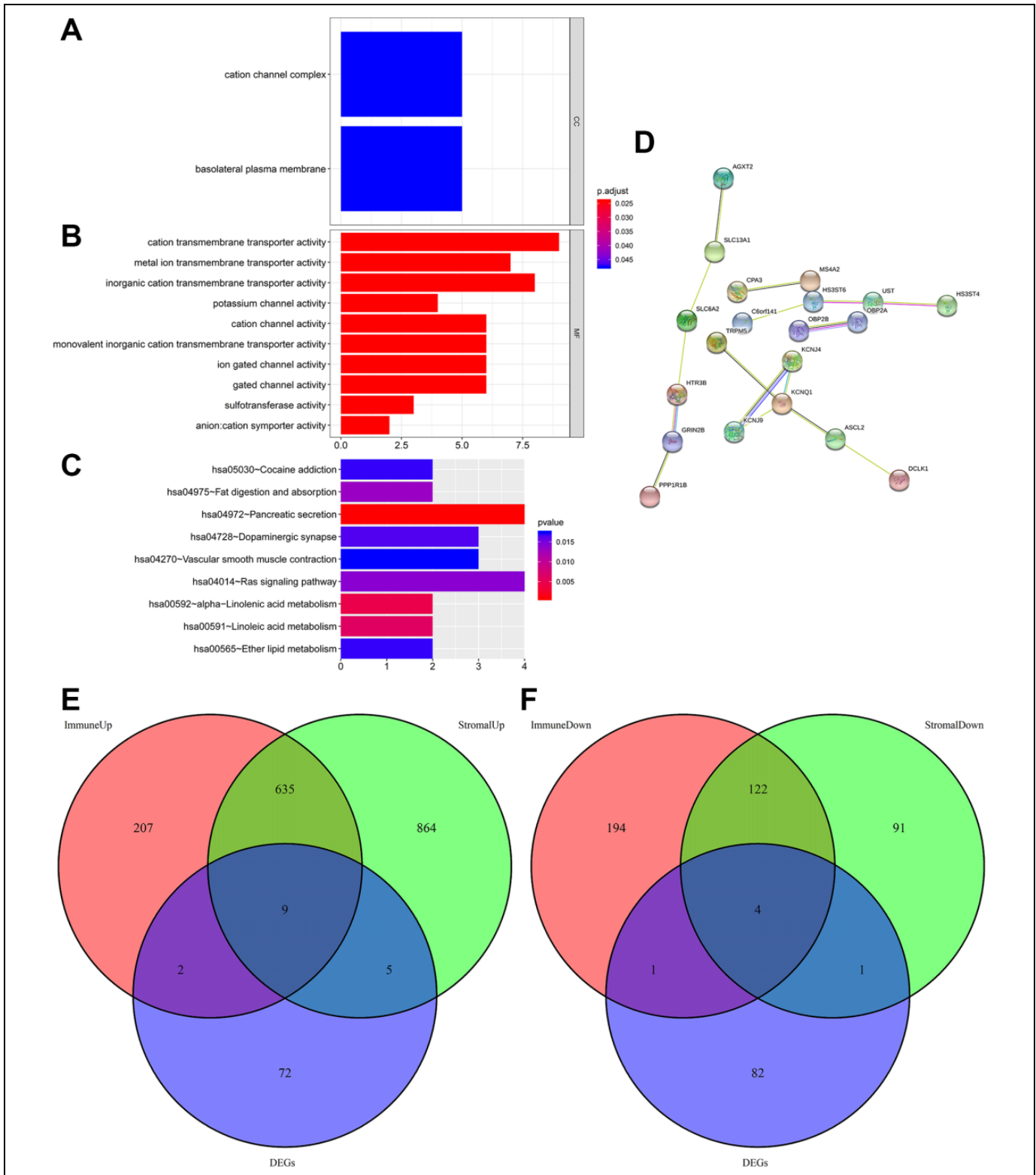
## GSEA

To further improve and supplement the results of functional enrichment analysis, GSEA was employed. The results of GSEA revealed that the DEGs were significantly enriched in 2 KEGG pathways and 1 GO term negatively related to LNM (Figure S1). In the GO analysis, those DEGs were only enriched in thiolester hydrolase activity (Figure S1A). And in

the KEGG analysis, those DEGs were enriched in 2 KEGG pathways, such as glyoxylate and dicarboxylate metabolism (Figure S1B) and peroxisome (Figure S1C).

## DEGs Analysis in Stromal Scores and Immune Scores and Acquisition of Intersect Genes

A total of 327 patients with GC from TCGA database were enrolled. The ESTIMATE algorithm was used to calculate stromal, immune scores. Then those STAD patients were divided into low- and high-group according to the median stromal and immune scores, respectively. In immune score groups, 853 highly expressed and 321 lowly expressed genes were identified from DEGs analysis (Figure 3E and F). Meanwhile, 1513 upregulated genes and 218 downregulated genes were identified from stromal score groups (Figure 3E and F). The Venn diagram was applied to distinguish 9 intersect genes between the LNM related DEGs and upregulated genes of immune and stromal scores (Figure 3E). Besides, 4 intersect genes between the LNM related DEGs and downregulated genes of immune and stromal scores were also identified (Figure 3F). Therefore, a total of 13 intersect genes were identified.



**Figure 3.** Functional enrichment analysis, PPI network and identification of intersect genes. Two GO terms in CC (A), and top 10 GO terms in MF (B) were performed for functional enrichment clustering analysis and visualized as bar chart. Top 9 KEGG pathways were identified and visualized as bar chart (C). Protein-protein interaction network was constructed (D). Venn plots were performed to visualize the number of up-regulated (E) and down-regulated intersect genes (F) in tumor microenvironment. GO indicates gene ontology; CC, cellular components; MF, molecular functions; KEGG, kyoto encyclopedia of genes and genomes.

**Table 3.** Summary of Univariate and Multivariate Logistic Regression Analysis.

	Univariate analysis			Multivariate analysis			
	OR	95% CI	P-value	Coefficient	OR	95% CI	P-value
OBP2B	0.854	0.681-1.071	0.172				
HS3ST6	0.883	0.721-1.080	0.226				
PLA2G2E	1.028	0.803-1.316	0.825				
PPP1R1B	0.676	0.522-0.876	0.003	-0.128	0.880	0.623-1.245	0.470
RASSF2	1.821	1.398-2.371	<0.001	1.820	6.169	2.603-14.620	<0.001
CPA3	1.431	1.176-1.740	<0.001	0.338	1.403	0.822-2.395	0.215
MS4A2	1.367	1.055-1.771	0.018	-1.252	0.286	0.085-0.961	0.043
ABCA8	1.448	1.141-1.839	0.002	-0.227	0.797	0.302-2.104	0.647
CLECL1	1.498	1.149-1.953	0.003	0.950	2.586	0.812-8.234	0.108
DTXI	1.408	1.116-1.777	0.004	-0.668	0.513	0.240-1.097	0.085
NAIP	1.283	0.977-1.685	0.073				
ANKRD33B	1.370	1.064-1.765	0.015	-1.351	0.259	0.091-0.735	0.011
ADH1B	1.486	1.227-1.799	<0.001	0.546	1.726	1.021-2.918	0.042

### Generation of 4-mRNAs Signature

Univariate analysis was performed to analyze the intersect mRNAs expression and the status of LNM of each patient from training cohort to identify the LNM-related mRNAs. A total of 9 mRNAs significantly correlated with LNM ( $P < 0.05$ ) were screened out and applied for further multivariate logistic regression analysis. Subsequently, 4 mRNAs (RASSF2, MS4A2, ANKRD33B and ADH1B) were singled out to construct a predictive model (Table 3). A riskscore analysis of the 4 mRNAs to calculate the riskscore for each patient based on the coefficients and expression level of the 4 mRNAs. Riskscore =  $1.820 * RASSF2 - 1.252 * MS4A2 - 1.351 * ANKRD33B + 0.546 * ADH1B$ . Among these mRNAs, 2 mRNAs with positive coefficients, including RASSF2 and ADH1B, which indicated that higher expression level of the 2 mRNAs was associated with higher risk of LNM. Meanwhile, the coefficients of the other 2 mRNAs (MS4A2 and ANKRD33B) were negative, which implied that higher expression level was associated with lower risk of LNM.

### Validation of the Validity of the 4-mRNAs Signature to Predict LNM

The riskscore for each patient of the training and test cohort was calculated. ROC curves were applied to determine the sensitivity and specificity of the 4-mRNAs signature and each mRNA. As depicted in Figure 4A, the AUC value of the 4-mRNAs signature was 0.800 in the training cohort which was significantly higher than that of each mRNA (RASSF2, 0.723; MS4A2, 0.606; ANKRD33B, 0.596; ADH1B, 0.707). The results indicated that the 4-mRNAs signature had good sensitivity and specificity for predicting LNM. The cutoff value was 1.016 which was determined by the ROC curve of training cohort. Then the training cohort was divided into low- ( $n = 69$ ) and high-risk ( $n = 94$ ) groups according to the cutoff value. As shown in Figure 4B, the distribution of riskscore in training cohort was plotted. Moreover, the proportion of patients with

LNM in the low-risk group was significantly lower than high-risk group ( $P < 0.001$ , Figure 4C).

To validate the predictive ability of the 4-mRNAs signature, ROC curves were also plotted to calculate the AUC value for test cohort. The AUC values of the 4-mRNAs signature in test cohorts were 0.742 (Figure 4D). And AUC values of the 4-mRNAs signature was higher than each mRNA in test cohort (Figure 4D). In addition, the test cohort was also grouped into low- ( $n = 92$ ) and high-risk ( $n = 72$ ) groups based on the same cutoff value. The distribution of riskscore in test cohort was also plotted (Figure 4E). And the proportion of patients with LNM in the high-risk group was significant higher than low-risk group ( $P < 0.001$ , Figure 4F).

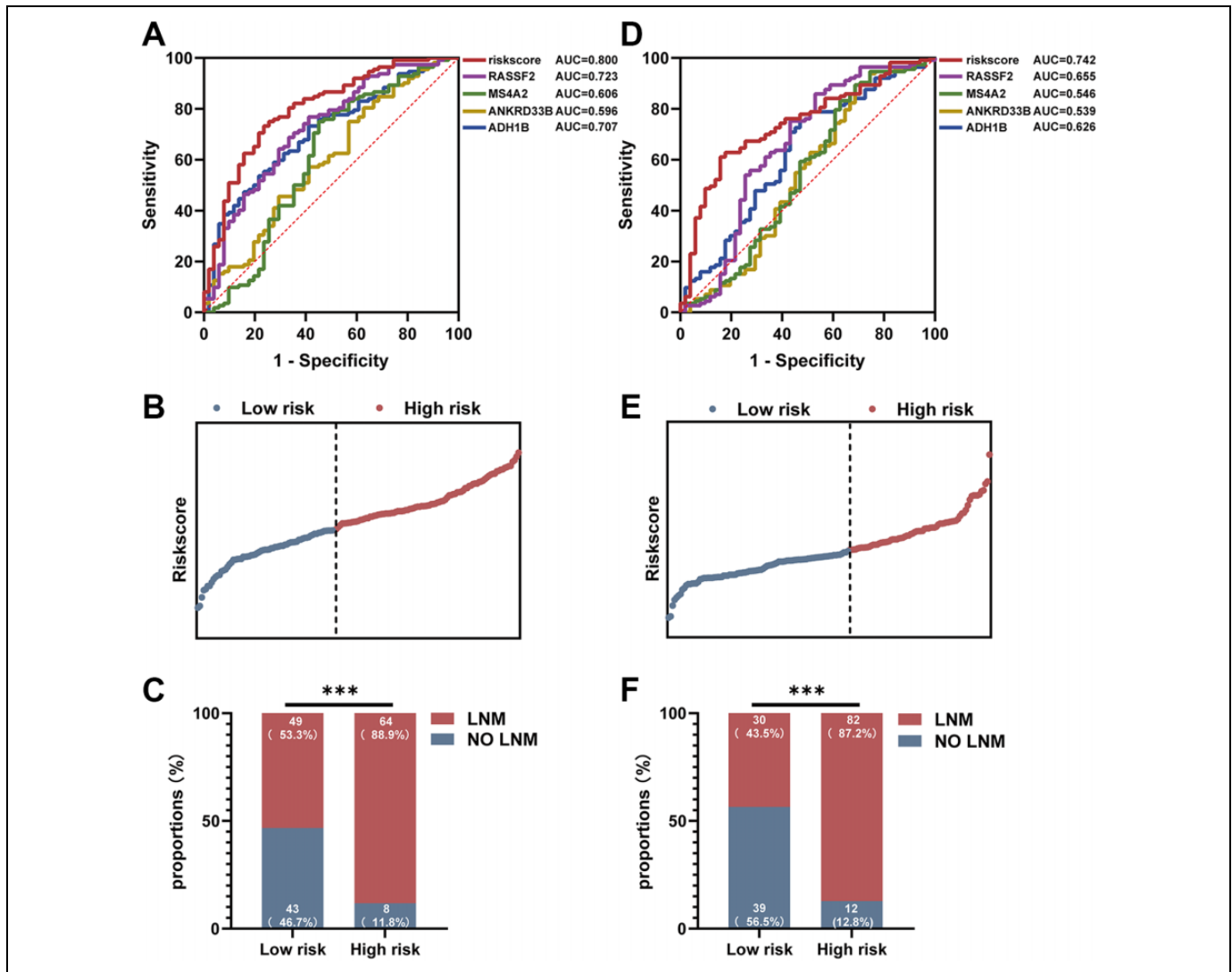
### Nomogram Model Construction and Prediction

To facilitate the 4-mRNAs signature application in clinical practice, nomogram was constructed based on training cohort (Figure 5A). A nomogram-based score is calculated for each patient based on 4 mRNAs on the point scale. The calibration curves of training (Figure 5B) and test cohorts (Figure 5C) imply that the nomogram model exhibits excellent performance for predicting LNM. The C-index of training and test cohorts were 0.856 and 0.756, respectively.

### Discussions

The prognosis of GC patients was evaluated mainly based on the TNM staging system. Among the TNM staging system, LNM is an important indicator. Moreover, the status of LNM is also vital to confirm the treatment regimens for GC. Hence, the identification of LNM-related biomarkers is beneficial to explore the underlying mechanisms involved in LNM and improve the prognosis of GC patients with lymph node involvement.

Combining the results of functional enrichment analysis and GSEA, we found that LNM is closely related to ion transport and some metabolic processes. Studies have shown that



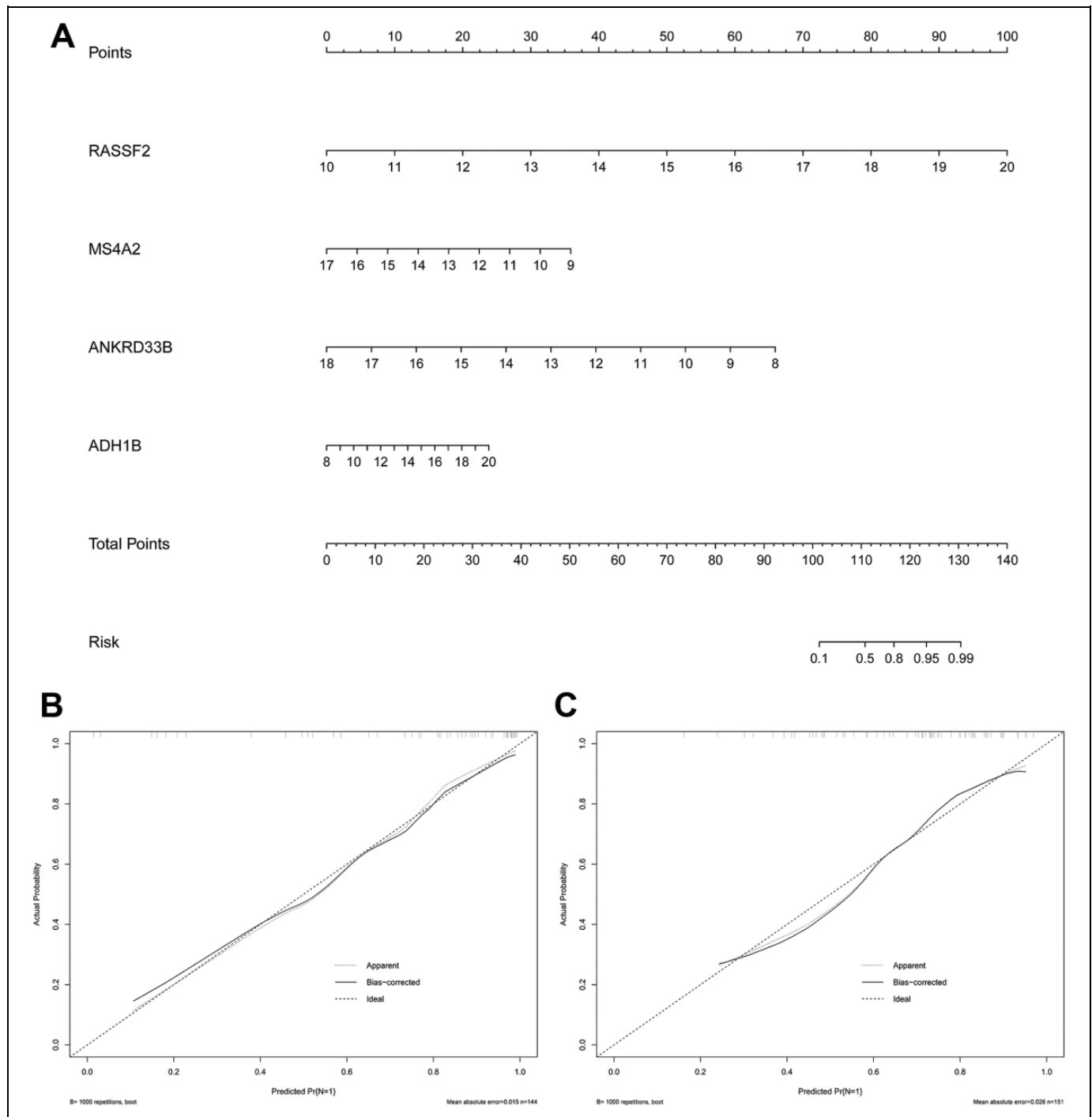
**Figure 4.** ROC curves, riskscore distribution and lymph node metastasis data of the training and test cohorts. ROC curves, riskscore distribution and the proportions of GC patients with lymph node metastasis in training cohort (A-C). ROC curves, riskscore distribution and the proportions of GC patients with lymph node metastasis in test cohort (D-F). LNM indicates lymph node metastasis.

transient receptor potential vanilloid 2 (TRPV2), a member of transient receptor potential (TRP)  $Ca^{2+}$  permeable channels, has shown carcinogenic activity in various cancers,<sup>18</sup> such as breast cancer,<sup>19</sup> esophageal squamous cell carcinoma,<sup>20,21</sup> hepatocarcinoma<sup>22-24</sup> and hematologic malignancies,<sup>25-27</sup> by controlling proliferation, migration, angiogenesis, and invasion. Moreover, according to the GO result of GSEA, thiolester hydrolase activity is negatively associated with LNM. And thiolester hydrolase can catalyze the hydrolysis of thioester bonds which can be found in acetyl-coenzyme A. Importantly, acetyl-coenzyme A is an important intermediate metabolite of 3 nutrients such as glucose, fat and protein, and it can finally produce a large amount of energy through the tricarboxylic acid (TCA) cycle which is beneficial to the metastasis of tumor.<sup>28</sup> In addition, pancreatic secretion, fat digestion and absorption and alpha-linolenic acid metabolism were top terms in the KEGG pathways analysis. Herein, lipid metabolism may

be involved in LNM of GC. Fatty acids are closely related to lipid metabolism in cancer, and the balance between the omega-3 and omega-6 families plays an important role in tumor metastasis.<sup>29</sup> Therefore, we suggested that LNM of GC was related to ion transmembrane transport, lipid metabolism and the decreased activity of thiolester hydrolase.

At present, the traditional diagnostic methods for LNM, such as EUS, CT, PET/CT and MDCT are subjective in some extent and have limited sensitivity, specificity and accuracy. Thus, we tried to build a prediction model based on gene expression level, which was quantifiable and more objective. We identified a 4-mRNAs signature with good predictive power for LNM in both training and test cohorts. The accuracy, sensitivity and specificity of the 4-mRNAs signature in training cohort were 80.0%, 73.2% and 76.5%, respectively. In test cohort, the accuracy, sensitivity and specificity of that were 74.2%, 61.1% and 84.3%, respectively. It has high accuracy





**Figure 5.** Construction and validation of nomogram. (A) The nomogram was constructed based on the training cohort. Calibration curves of the nomogram in the training (B) and test (C) cohorts.

in both training and test cohorts, indicating that the model has high stability in predicting LNM. To further elucidate the discriminate power of this model, patients were classified into low- and high-risk group based on the cutoff value of the risk-score for both training and test cohort. In both the training and test cohorts, the proportion of patients with LNM in the high-risk group was significantly higher than that in the low-risk group, and the proportion of patients with no LNM was

significantly lower. According to the published articles, we found that the sensitivity and specificity of CT for lymph node involvement were 78% and 62%, respectively. The Youden index of CT is 0.400. However, according to our data results, in the training cohort, the sensitivity and specificity of the 4-mRNAs signature constructed by us are 73.2% and 76.8%, respectively. And the sensitivity and specificity of that in test cohort are 61.1% and 84.3%, respectively. And the Youden

index of 4-mRNAs signature in training and test cohorts are 0.500 and 0.454, respectively. Obviously, the Youden index of 4-mRNAs signature is significantly higher than that of CT in both cohorts. Besides, the ability of MDCT to assess LNM in GC has also been recognized in recent years. And the AUC value of the LNM detection with MDCT in advanced GC is 74.4%, which is slightly lower than the 80.0% of the training cohort and similar to the 74.2% of the test cohort from the present study. According to another published articles, due to the low sensitivity (0.34), MDCT is insufficient for the detection of LNM in early GC. Therefore, we have reason to believe that this 4-mRNAs signature is not inferior or even superior to CT and MDCT.

To facilitate further clinical application, nomograms were constructed and calibrated. The C-index of training and test cohorts were 0.865 and 0.765, respectively. Furthermore, the calibration curves of training and test cohorts indicated that the nomogram was well calibrated. Nomograms were widely used to predict LNM of GC. However, there are some limitations to the application value as some factors are only available postoperatively. In this study, we applied the expression level of 4 mRNAs to predict LNM which could be obtained preoperatively. Moreover, compared with CT, the detection of the 4 mRNAs is simpler and radiation-free, which greatly saves the examination time of patients and brings them more convenience. However, this study has also some limitations. Firstly, the 4-mRNAs signature and nomogram were only constructed and validated in TCGA database. Secondly, this study did not directly compare the accuracy of the 4-mRNAs model and CT in the diagnosis of LNM in the same population. In the future, we will compare the predictive ability of the 4-mRNAs model with CT for LNM in our department, and evaluate whether the combination of the 2 can improve the accuracy in predicting LNM.

## Conclusions

In conclusion, we found that ion transmembrane transporter activity, lipid metabolism and thiolester hydrolase activity are closely related to LNM in GC. Besides, 4 LNM-related genes in TME were identified and applied for the construction of prediction model and nomogram. More importantly, the prediction model and nomogram can accurately predict LNM in preoperative and provide some reference for the evaluation of GC and the formulation of clinical treatment regimens.

## Abbreviations

AML, Acute Myeloid Leukemia; CAFs, Cancer Associated Fibroblasts; CC, Cellular Components; ccRCC, clear cell Renal Cell Carcinoma; DEGs, Differentially Expressed Genes; ECM, Extracellular Matrix; EUS, Endoscopic Ultrasound; GC, Gastric Cancer; GO, Gene Ontology; GSEA, Gene Set Enrichment Analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; LNM, Lymph Node Metastasis; MDCT, multi-detector row CT; MF, Molecular Functions; PPI, Protein-Protein Interaction; STAD, Stomach Adenocarcinoma; TCGA, The Cancer Genome Atlas; TME, Tumor Microenvironment; TRP, Transient Receptor Potential; TRPV2, Transient Receptor Potential Vanilloid 2.

## Authors' Note

Shilong Li and Zongxian Zhao contributed equally to this article.


## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iD

Weihua Fu, MD  <https://orcid.org/0000-0003-2576-865X>

## Supplemental Material

Supplemental material for this article is available online.

## References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394-424. doi:10.3322/caac.21492
2. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. *CA Cancer J Clin.* 2016;66(2):115-132. doi:10.3322/caac.21338
3. Ajani JA, D'Amico TA, Almhanna K, et al. Gastric cancer, version 3.2016, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw.* 2016;14(10):1286-1312. doi:10.6004/jnccn.2016.0137
4. Karpeh MS, Leon L, Klimstra D, Brennan MF. Lymph node staging in gastric cancer: is location more important than number? An analysis of 1,038 patients. *Ann Surg.* 2000;232(3):362-371. doi:10.1097/00000658-200009000-00008
5. Abdalla EK, Pisters PW. Staging and preoperative evaluation of upper gastrointestinal malignancies. *Semin Oncol.* 2004;31(4):513-529. doi:10.1053/j.seminoncol.2004.04.014
6. Kwee RM, Kwee TC. Imaging in local staging of gastric cancer: a systematic review. *J Clin Oncol.* 2007;25(15):2107-2116. doi:10.1200/JCO.2006.09.5224
7. Weber WA, Ott K. Imaging of esophageal and gastric cancer. *Semin Oncol.* 2004;31(4):530-541. doi:10.1053/j.seminoncol.2004.04.016
8. Kawanaka Y, Kitajima K, Fukushima K, et al. Added value of pretreatment (18)F-FDG PET/CT for staging of advanced gastric cancer: comparison with contrast-enhanced MDCT. *Eur J Radiol.* 2016;85(5):989-995. doi:10.1016/j.ejrad.2016.03.003
9. Luo M, Lv Y, Guo X, Song H, Su G, Chen B. Value and impact factors of multidetector computed tomography in diagnosis of preoperative lymph node metastasis in gastric cancer: a PRISMA-compliant systematic review and meta-analysis. *Medicine (Baltimore).* 2017;96(33):e7769. doi:10.1097/MD.0000000000007769
10. Cardoso R, Coburn N, Seevaratnam R, et al. A systematic review and meta-analysis of the utility of EUS for preoperative staging

- for gastric cancer. *Gastric Cancer*. 2012;15(suppl 1):S19-S26. doi:10.1007/s10120-011-0115-4
11. Chen J, Cheong JH, Yun MJ, et al. Improvement in preoperative staging of gastric adenocarcinoma with positron emission tomography. *Cancer*. 2005;103(11):2383-2390. doi:10.1002/cncr.21074
  12. Belli C, Trapani D, Viale G, et al. Targeting the microenvironment in solid tumors. *Cancer Treat Rev*. 2018;65:22-32. doi:10.1016/j.ctrv.2018.02.004
  13. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nat Med*. 2013;19(11):1423-1437. doi:10.1038/nm.3394
  14. Joyce JA, Pollard JW. Microenvironmental regulation of metastasis. *Nat Rev Cancer*. 2009;9(4):239-252. doi:10.1038/nrc2618
  15. Ni J, Wu Y, Qi F, et al. Screening the Cancer Genome Atlas Database for genes of prognostic value in acute myeloid leukemia. *Front Oncol*. 2019;9:1509. doi:10.3389/fonc.2019.01509
  16. Luo J, Xie Y, Zheng Y, et al. Comprehensive insights on pivotal prognostic signature involved in clear cell renal cell carcinoma microenvironment using the ESTIMATE algorithm. *Cancer Med*. 2020;9(12):4310-4323. doi:10.1002/cam4.2983
  17. Ni J, Liu S, Qi F, et al. Screening TCGA database for prognostic genes in lower grade glioma microenvironment. *Ann Transl Med*. 2020;8(5):209. doi:10.21037/atm.2020.01.73
  18. Cohen MR, Huynh KW, Cawley D, Moiseenkova-Bell VY. Understanding the cellular function of TRPV2 channel through generation of specific monoclonal antibodies. *PLoS One*. 2013;8(12):e85392. doi:10.1371/journal.pone.0085392
  19. Gogebakan B, Bayraktar R, Suner A, et al. Do fasudil and Y-27632 affect the level of transient receptor potential (TRP) gene expressions in breast cancer cell lines? *Tumour Biol*. 2014;35(8):8033-8041. doi:10.1007/s13277-014-1752-0
  20. Zhou K, Zhang SS, Yan Y, Zhao S. Overexpression of transient receptor potential vanilloid 2 is associated with poor prognosis in patients with esophageal squamous cell carcinoma. *Med Oncol*. 2014;31(7):17. doi:10.1007/s12032-014-0017-5
  21. Shiozaki A, Kudou M, Ichikawa D, et al. Esophageal cancer stem cells are suppressed by tranilast, a TRPV2 channel inhibitor. *J Gastroenterol*. 2018;53(2):197-207. doi:10.1007/s00535-017-1338-x
  22. Ma W, Li C, Yin S, et al. Novel role of TRPV2 in promoting the cytotoxicity of H<sub>2</sub>O<sub>2</sub>-mediated oxidative stress in human hepatoma cells. *Free Radic Biol Med*. 2015;89:1003-1013. doi:10.1016/j.freeradbiomed.2015.09.020
  23. Liu G, Xie C, Sun F, et al. Clinical significance of transient receptor potential vanilloid 2 expression in human hepatocellular carcinoma. *Cancer Genet Cytogenet*. 2010;197(1):54-59. doi:10.1016/j.cancergencyto.2009.08.007
  24. Liu LL, Fu D, Ma Y, Shen XZ. The power and the promise of liver cancer stem cell markers. *Stem Cells Dev*. 2011;20(12):2023-2030. doi:10.1089/scd.2011.0012
  25. Morelli MB, Liberati S, Amantini C, et al. Expression and function of the transient receptor potential ion channel family in the hematologic malignancies. *Curr Mol Pharmacol*. 2013;6(3):137-148. doi:10.2174/187446720603140415215431
  26. Boyd RS, Jukes-Jones R, Walewska R, Brown D, Dyer MJ, Cain K. Protein profiling of plasma membranes defines aberrant signaling pathways in mantle cell lymphoma. *Mol Cell Proteomics*. 2009;8(7):1501-1515. doi:10.1074/mcp.M800515-MCP200
  27. Santoni G, Farfariello V, Liberati S, et al. The role of transient receptor potential vanilloid type-2 ion channels in innate and adaptive immune responses. *Front Immunol*. 2013;4:34. doi:10.3389/fimmu.2013.00034
  28. Cai Z, Li CF, Han F, et al. Phosphorylation of PDHA by AMPK drives TCA cycle to promote cancer metastasis. *Mol Cell*. 2020;80(2):263-278 e7. doi:10.1016/j.molcel.2020.09.018
  29. Ferreri C, Sansone A, Ferreri R, Amezcaga J, Tueros I. Fatty acids and membrane lipidomics in oncology: a cross-road of nutritional, signaling and metabolic pathways. *Metabolites*. 2020;10(9):345. doi:10.3390/metabo10090345