

Development and Validation of a Gene-panel-based Nomogram for Prediction of Lymph Node Metastasis in Esophageal Squamous Cell Carcinoma

Xuejie Wu

Second Affiliated Hospital of Soochow University

Donglai Chen

Shanghai Pulmonary Hospital, Tongji University

Wenjia Wang

Second Affiliated Hospital of Soochow University

Fei Ye

Hai'an Hospital Affiliated to Nantong University

Yonghua Sang

Second Affiliated Hospital of Soochow University

Yongbing Chen (✉ chentongt@sina.com)

Second Affiliated Hospital of Soochow University <https://orcid.org/0000-0002-7595-092X>

Wentao Yang

Second Affiliated Hospital of Soochow University

Primary research

Keywords: genomic, esophageal cancer, lymph node metastasis, database

Posted Date: August 27th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-63589/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Esophageal squamous cell carcinoma (ESCC) is one of the main histological subtypes of esophageal cancer. This study aimed to develop a gene-panel-based nomogram for identification of lymph node metastasis (LNM) in ESCC patients.

Methods: RNA sequencing profiles of ESCC patients were obtained from the Gene Expression Omnibus (GEO) database and The Cancer Genome Atlas (TCGA) database. A bioinformatic approach was employed to investigate differentially expressed genes (DEGs) between ESCC patients with LNM and those without. A 4-DEGs panel was eventually identified and integrated with clinical characteristics to construct a nomogram for predicting LNM. Predictive performance of the nomogram was further evaluated by calibration curves and concordance index (C-index).

Results: A total of 179 ESCC patients with 32059 genes from the GEO dataset were included. Among these genes, 3524 DEGs were correlated with lymph node involvement. Meanwhile, TCGA dataset containing 93 ESCC patients was obtained, in which 82 DEGs were selected out of 18416 genes. Among the 11 communal DEGs, four genes were identified to be (ALG3, CPOX, LMLN, PSMD2) associated with LNM. A nomogram was established by integrating the four-gene panel and three clinical characteristics including T stage, G stage and tumor location. The nomogram exhibited good performance with C-indices of 0.710 and 0.693 in the GEO and TCGA datasets, respectively.

Conclusion: Our novel 4-gene-based nomogram displayed its value in prediction of LNM in ESCC patients, which may be helpful in determining treatment approach for early-stage ESCC patients.

Background

Esophageal cancer is the seventh commonest malignancy and the sixth leading cause of cancer-related mortality worldwide(1, 2). Esophageal cancer consists of two main subtypes, esophageal squamous cell carcinoma(ESCC) and esophageal adenocarcinoma. Nowadays, ESCC accounted for the predominant histological subtype in eastern Asia, especially in China(3–5).

Over the decades, esophagectomy with lymph node (LN) dissection has remained the most effective treatment modality for patients with early-stage ESCC(6, 7). Adequate lymph nodes harvested, esophagectomy might increase postoperative morbidity and mortality which is probably unsuitable for patients with suspected T1N0M0 disease(8–10). In recent years, endoscopic resection (ER), however, exhibits great potential in treating small-sized T1 disease, which might be alternative to conventional esophagectomy(11). Notably, the increasing prevalence of ER highlights the importance of assessing the status of regional lymph nodes (LN) in ESCC(12). However, it is hard to accurately evaluate the lymph node status using conventional radiological examinations(13, 14). Since lymph node metastasis (LNM) is a significant predictor of poor prognosis in ESCC(15–19), availability of a predictive tool for LNM will be helpful in improving treatment modalities and circumventing excessive LN dissection.

In the present study, we performed a genome-wide discovery of genes associated with LNM in ESCC by analyzing datasets from the GEO and TCGA databases. Four LNM-associated genes identified, a gene-panel-based nomogram was established and validated, which displayed good performance in predicting LNM in ESCC patients.

Materials And Methods

Patient cohorts and transcriptomic data

Transcriptomic data and the corresponding clinical variables of ESCC samples were downloaded from the GEO database (GSE53625) and TCGA database. The exclusion criteria were as follows: (1) ESCC samples of which the genomic information were absent; (2) Samples lacking information on clinicopathological characteristics, especially nodal stage. (3) Samples with distant metastasis (M1 stage).

Identification of communal differentially expressed genes (DEGs)

GSE53625 including mRNA sequencing data of ESCC patients were preprocessed and annotated. Differential expression analysis was performed between ESCC samples with LNM and those without using the R package limma(version 3.6.2)(20). DEGs were determined by significance criteria (false discovery rate ((FDR) < 0.05 , $|\logFC| > 1$, $p < 0.05$). Similar criteria ($|\logFC| > 1$, $p < 0.05$) were applied to TCGA dataset to identify DEGs associated with presence of LNM. Considering the limited number of overlapped DEGs between the two datasets, we subsequently assessed the association of the communal DEGs for the identification of LNM. The communal DEGs was defined as intersection of the DEGs between the two datasets and was visualized by the Venn Diagram package (version 1.6.20).

Multivariate Logistic regression analysis and calculation of Pearson correlation coefficients

The Logistic regression model was applied to identify the independent predictors of the presence of LNM among the 11 communal DEGs. Four of the 11 communal DEGs were eventually confirmed to be associated with LNM, and the correlation coefficients of the four DEGs were calculated with Pearson correlation.

Construction and validation of the gene-panel-based nomogram

The predictive power of the four communal DEGs was initially assessed. Integrating clinical variables to improve the predictive performance, we established a nomogram for predicting LNM by using the GEO dataset as the primary cohort. TCGA dataset was used as the validation cohort. In addition, we

constructed another cohort comprised of patients with T1-2 disease from the two cohorts as the early-stage cohort. Calibration curves were plotted to assess the calibration of the gene-panel-based nomogram. To quantify the discrimination performance of the nomogram, Harrell's C-index was measured. Meanwhile, the predicted probability of LNM was also evaluated by Receiver Operating characteristic (ROC) curves with area under curves (AUC) calculated. Finally, the predictive power of our nomogram was verified in the early-stage cohort.

Statistical Analysis

Pearson χ^2 test was used for categorical variables, and two-sample t test was used for continuous variables. A two-sided P-value of <0.05 was considered as statistically significant. All the statistical analyses were performed using R(version 3.6.2) and SPSS statistical software version 23 (IBM Corp, Armonk, NY).

Results

Baseline information and study design

The baseline characteristics of the two datasets are summarized in Table 1. A total of 179 ESCC samples in the GEO database and 93 samples in TCGA database were included in our study. There were 96 patients with LNM and 83 patients without LNM in the GEO cohort, while 38 patients with LNM and 55 without LNM in TCGA cohort. In the early-stage cohort, there are 34 patients with LNM and 41 patients without LNM (Supplementary Table 2).The design of our study are illustrated using a flow chart in Fig. 1.

Table 1
Baseline Characteristics of Patients with ESCC in each database

	GEO database(n = 179)			TCGA database(n = 93)		
Validation	LN(+)	LN(-)	<i>P</i> value	LN(+)	LN(-)	<i>P</i> value
Age,years			0.503			0.024
≤ 65	72	58		34	38	
> 65	24	25		4	17	
Sex			0.565			0.264
Female	16	17		4	11	
Male	80	66		34	44	
Race			NA			0.856
White	0	0		18	25	
Non-white	96	83		20	30	
Tumor Location			0.025			0.156
Upper thoracic	13	7		4	2	
Middle thoracic	43	54		14	30	
Lower thoracic	40	22		20	23	
T stage			0.720			0.202
T1-T2	22	17		13	27	
T3-T4	74	66		25	28	
Differentiation			0.455			0.080
Well differentiated; G1	16	16		2	14	
Moderately differentiated; G2	50	48		24	25	
Poorly differentiated; G3	30	19		8	11	
Unknown; Gx	0	0		4	5	
Abbreviations: LN, lymph node; NA, not applicable.						

Identification of communal DEGs

The gene expression profiles of patients with LNM and those without were compared to screen DEGs. Ultimately, a total of 3524 DEGs were identified from 32059 transcripts in the GEO cohort ((FDR) < 0.05 |logFC| > 1, $p < 0.05$), in which 1082 genes were up-regulated and 2442 genes were down-regulated. Meanwhile, 82 DEGs were found in TCGA cohort (|logFC| > 1, $p < 0.05$) with 5 genes up-regulated and 77 genes down-regulated. Heat map and volcano map approaches were used to demonstrate the expression levels and upregulation/downregulation of DEGs in the GEO cohort (Suppl Fig. 1A-B) and TCGA cohort (Suppl Fig. 1C-D). A Venn diagram was employed to visualize the intersection of DEGs between the two datasets (Suppl Fig. 2). As shown in Fig. 2A-B, the expression levels of the 11 communal DEGs (ALG3,AP2M1,CPOX,CRHR2,LMLN,MAP6D1,MRPL147,PARL,PSMD2,SLC15A2,SMYD3) in the two datasets were illustrated using heat maps.

Identification of gene signatures associated with LNM and development of a four-gene panel.

As shown in Table 2, among the 11 communal genes, ALG3(odds ratio (OR), 0.772; 95% CI: 0.66–0.904; $p = 0.0015$), CPOX(OR, 1.155; 95% CI: 0.621–1.06; $p = 0.038$), LMLN(OR, 0.861; 95% CI: 0.754–0.983; $p = 0.028$) and PSMD2(OR, 1.64; 95% CI: 1.186–2.267; $p = 0.003$) were associated with the presence of LNM in multivariate Logistic regression analysis. The expression sites of 4 genes, biological functions, KEGG pathways and AUC of individual genes predicting the probability of LNM are shown in Supplementary Table 1. The absolute pair-wise Pearson correlation coefficients among the four communal DEGs genes were calculated to show their independence (Suppl Fig. 3A). A four-gene panel was thereafter built to predict LNM. Surprisingly, the ROC curve for predicting probability of LNM indicated unsatisfactory performance with the AUC merely 0.547(Suppl Fig. 3B).

Table 2
Logistic regression analysis for intersection genes in
ESCC tissues with LNM and NLNM

Gene	Multivariate	
	OR(95%CI)	P -value
ALG3	0.772(0.66–0.904)	0.00155
AP2M1	0.812(0.621–1.06)	0.12746
CPOX	1.155(1.009–1.322)	0.03859
CRHR2	1.001(0.897–1.117)	0.98875
LMLN	0.861(0.754–0.983)	0.02840
MAP6D1	0.941(0.827–1.071)	0.35660
MRPL47	1.191(0.993–1.52)	0.16198
PARL	0.991(0.781–1.257)	0.93986
PSMD2	1.64(1.186–2.267)	0.00317
SLC15A2	0.992(0.934–1.054)	0.80537
SMYD3	0.933(0.883–1.045)	0.23428
Abbreviations: OR, odds ratio; CI, confidence interval.		

Construction and validation of a gene-panel-based nomogram for Predicting LNM

To improve the predictive ability of the aforementioned panel, we developed a nomogram integrating clinical variables and the four communal DEGs. Previous studies have identified clinicopathologic factors including T stage, G stage and tumor location, as predictors of the risk of LNM in patients with ESCC(14, 21–25). And some studies also aimed to construct LNM prediction models based on the aforementioned clinical variables(26, 27). Ultimately, the nomogram consisting of 7 variables including 4 DEGs, T stage, G stage and tumor location was established to predict the probability of LNM (Fig. 3). Calibration curves regarding the model performance in the two datasets are shown in Fig. 4. With a C-index of 0.710 and that of 0.693 in respective dataset, the nomogram displayed good discrimination in both the primary and validation cohorts, which outperformed the four-gene panel itself ($p < 0.01$). Similarly, the nomogram also exhibited good predictive potential in patients with T1-2 disease with a C-index of 0.755 in the early-stage cohort.

Discussion

In this study, we performed a comprehensive RNA sequencing-based gene expression profiling analysis of ESCC patients with and without LNM to establish a gene expression panel for identification of LNM. After extraction of the communal LNM-associated gene signatures, we subsequently evaluated the robustness of our gene-panel-based nomogram in the two public datasets available. We demonstrated that the nomogram was valuable in both the primary and validation cohorts to identify patients with LNM.

From a clinical standpoint, the LN status is vitally important for therapeutic decision-making, especially in ESCC patients with relatively early T stage. Without LN metastasis, Tis and T1 ESCC patients can be successfully treated with ER(28). However, if the tumor is most likely to involve LNs, esophagectomy with LN dissection is often required to obtain better curability. Even for cases that apparently are not suitable for ER, a more precise information on LN status can significantly contribute to decision making for neoadjuvant treatment or definite chemoradiation, as recommended by the National Comprehensive Cancer Network Guidelines. Therefore, availability of robust biomarkers that facilitate categorization of patients with ESCC based on the LN status will permit personalized treatment modality.

Although researches on seeking robust biomarkers for predicting LNM has been ongoing over the past decades, few studies available reported gene panels associated with the LN status(3, 29). It is known to us that any single gene is not reliable enough to identify LNM since its expression level can be affected by many confounding factors. Notably, the recent advancements in whole genome and transcriptome sequencing have allowed us to characterize the associations between LNM and gene profiles from a new perspective. In recent years, attention has been focused on the LNM prediction model for gastrointestinal tumors(30). However, the comprehensive prediction model of LNM for ESCC is still blank. Consequently, our nomogram derived from transcriptomic data analyses demonstrated that integrating gene panel and clinical variables might be a promising approach to predict LN status.

Admittedly, there are several limitations in our study. First, not all patients underwent neoadjuvant therapy followed by esophagectomy and may have influenced the effectiveness of the gene-signature. Second, our nomogram could not provide relevant information on location of the metastatic lymph nodes. Notably, a previous study reported that the prognosis of patients with LNs involvement limited to the peritumor areas was significantly better than that of patients with LNs involvement farther away(17). In addition, there are difference in the sequence data between the GEO dataset and TCGA dataset. More specifically, GEO employs gene chip technology to detect gene expression in patient tissues while TCGA uses the RNA sequencing technology, which may cause potential errors.

Conclusion

In conclusion, our gene-panel-based nomogram can help predict the status of LNM in ESCC patients, and may determine the suitability of ER or esophagectomy for ESCC patients.

Abbreviations

ESCC: Esophageal squamous cell carcinoma; TCGA: The Cancer Genome Atlas; GEO: Gene Expression Omnibus; LNM: Lymph node metastasis; DEGs: Differentially expressed genes; C-index: Concordance index; LN: lymph nodes; FDR: false discovery rate; ROC: Receiver operating characteristic; AUC: Area under the curve; HR: Hazard ratio.

Declarations

Acknowledgements

We appreciate the free use of TCGA and GEO databases

Authors' contributions

WXJ, CDL, and WWJ: conception and design; Provision of study materials or patients; Collection and assembly of data; Data analysis and interpretation. XY, YFH: Provision of study materials or patients; Collection and assembly of data; Data analysis and interpretation. SYH and YWT: Administrative support; Provision of study materials or patients. YWT and CYB: Conception and design; Administrative support; Provision of study materials or patients

Funding

Supported by the projects from Suzhou Key Laboratory of Thoracic Oncology (SZS201907), Suzhou Key Discipline for Medicine (SZXK201803), Municipal Program of People's Livelihood Science and Technology in Suzhou (SS2019061) and Major Project for Social Development, Jiangsu Provincial Department of Science and Technology (BE2020653).

Availability of data and materials

The TCGA-ESCC dataset used in this study could be obtained from TCGA database (<https://cancergenome.nih.gov/>). The GEO datasets (GSE53625) used in this study could be obtained from GEO database (<https://www.ncbi.nlm.nih.gov/geo/>)

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests

References

1. Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012. *CA-Cancer J Clin*. 2015;65:87–108.
2. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA-Cancer J Clin*. 2018;68:394–424.
3. Sonohara F, Gao F, Iwata N, et al. Genome-wide discovery of a novel gene-expression signature for the identification of lymph node metastasis in esophageal squamous cell carcinoma. *Ann Surg*. 2019;269:879–86.
4. Qiu ML, Lin JB, Li X, et al. Current state of esophageal cancer surgery in China: a national database analysis. *BMC Cancer*. 2019;19:6912–2.
5. Arnold M, Soerjomataram I, Ferlay J, et al. Global incidence of oesophageal cancer by histological subtype in 2012. *Gut*. 2015;64:381–7.
6. Steffen T, Dietrich D, Schnider A, et al. Recurrence patterns and long-term results after induction chemotherapy, chemoradiotherapy, and curative surgery in patients with locally advanced esophageal cancer. *Ann Surg*. 2019;269:83–7.
7. Yuan B, Liu L, Huang H, et al. Comparison of the short-term and long-term outcomes of surgical treatment versus endoscopic treatment for early esophageal squamous cell neoplasia larger than 2cm: a retrospective study. *Surg Endosc*. 2019;33:2304–12.
8. Takahashi H, Arimura Y, Masao H, et al. Endoscopic submucosal dissection is superior to conventional endoscopic resection as a curative treatment for early squamous cell carcinoma of the esophagus. *Gastrointest Endosc*. 2010;72:255–64.
9. Marino KA, Sullivan JL, Weksler B. Esophagectomy versus endoscopic resection for patients with early-stage esophageal adenocarcinoma: A National Cancer Database propensity-matched study. *J Thorac Cardiovasc Surg*. 2018;155:2211–8.
10. Nealis TB, Washington K, Keswani RN. Endoscopic therapy of esophageal premalignancy and early malignancy. *J Natl Compr Cancer Netw*. 2011;9:890–9.
11. Ajani JA, D'Amico TA, Bentrem DJ, et al. Esophageal and esophagogastric junction cancers, Version 2.2019. *J Natl Compr Cancer Netw*. 2019;17:855–83.
12. Gockel I, Sgourakis G, Lyros O, et al. Risk of lymph node metastasis in submucosal esophageal cancer: a review of surgically resected patients. *Expert Rev Gastroenterol Hepatol*. 2011;5:371–84.
13. Bergeron EJ, Lin J, Chang AC, et al. Endoscopic ultrasound is inadequate to determine which T1/T2 esophageal tumors are candidates for endoluminal therapies. *J Thorac Cardiovasc Surg*. 2014;147:765–71.

14. Shin S, Kim HK, Choi YS, et al. Clinical stage T1-T2N0M0 oesophageal cancer: accuracy of clinical staging and predictive factors for lymph node metastasis. *Eur J Cardiothorac Surg.* 2014;46:274–9.
15. Darling G. The role of lymphadenectomy in esophageal cancer. *J Surg Oncol.* 2009;99:189–93.
16. Tachibana M, Kinugasa S, Hirahara N, et al. Lymph node classification of esophageal squamous cell carcinoma and adenocarcinoma. *Eur J Cardiothorac Surg.* 2008;34:427–31.
17. Mariette C, Piessen G, Briez N, et al. The number of metastatic lymph nodes and the ratio between metastatic and examined lymph nodes are independent prognostic factors in esophageal cancer regardless of neoadjuvant chemoradiation or lymphadenectomy extent. *Ann Surg.* 2008;247:365–71.
18. Rice TW, Blackstone EH, Rybicki LA, et al. Refining esophageal cancer staging. *J Thorac Cardiovasc Surg.* 2003;125:1103–13.
19. Eloubeidi MA, Desmond R, Arguedas MR, et al. Prognostic factors for the survival of patients with esophageal carcinoma in the US - The importance of tumor length and lymph node status. *Cancer.* 2002;95:1434–43.
20. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.
21. Gamboa AM, Kim S, Force SD, et al. Treatment allocation in patients with early-stage esophageal adenocarcinoma: Prevalence and predictors of lymph node involvement. *Cancer.* 2016;122:2150–7.
22. Rampado S, Cassaro M, Battaglia G, et al. Prediction of lymph node status in superficial esophageal carcinoma. *Dig Liver Dis.* 2008;40:22-S3.
23. Akutsu Y, Uesato M, Shuto K, et al. The overall prevalence of metastasis in T1 esophageal squamous cell carcinoma a retrospective analysis of 295 patients. *Ann Surg.* 2013;257:1032–8.
24. Merkow RP, Bilimoria KY, Keswani RN, et al. Treatment trends, risk of lymph node metastasis, and outcomes for localized esophageal cancer. *JNCI-J Natl Cancer Inst.* 2014;106:766–76.
25. Duan XF, Tang P, Shang XB, et al. The prevalence of lymph node metastasis for pathological T1 esophageal cancer: a retrospective study of 143 cases. *Surg Oncol-Oxf.* 2018;27:1–6.
26. Ma DW, Jung DH, Kim JH, et al. Predicting lymph node metastasis for endoscopic resection of superficial esophageal squamous cell carcinoma. *J Thorac Cardiovasc Surg.* 2019;157:397–402.
27. Xu W, Liu XB, Li SB, et al. Prediction of lymph node metastasis in superficial esophageal squamous cell carcinoma in Asia: a systematic review and meta-analysis. *Diseases of the esophagus: official journal of the international society for Dis Esophagus.* 2020;05:13.
28. Zeng Y, Liang W, Liu J, He J. Endoscopic treatment versus esophagectomy for early-stage esophageal cancer: A population-based study using propensity score matching. *Journal of gastrointestinal surgery: official journal of the Society for Surgery of the Alimentary Tract.* 2017;21:1977–83.
29. Mao Y, Fu Z, Zhang Y, et al. A seven-lncRNA signature predicts overall survival in esophageal squamous cell carcinoma. *Sci Rep.* 2018;8:8823.

30. Daisuke I. Feng, et al. A genomewide transcriptomic approach identifies a novel gene expression signature for the detection of lymph node metastasis in patients with early stage gastric cancer. Ebiomedicine. 2019;41:268–75.

Figures

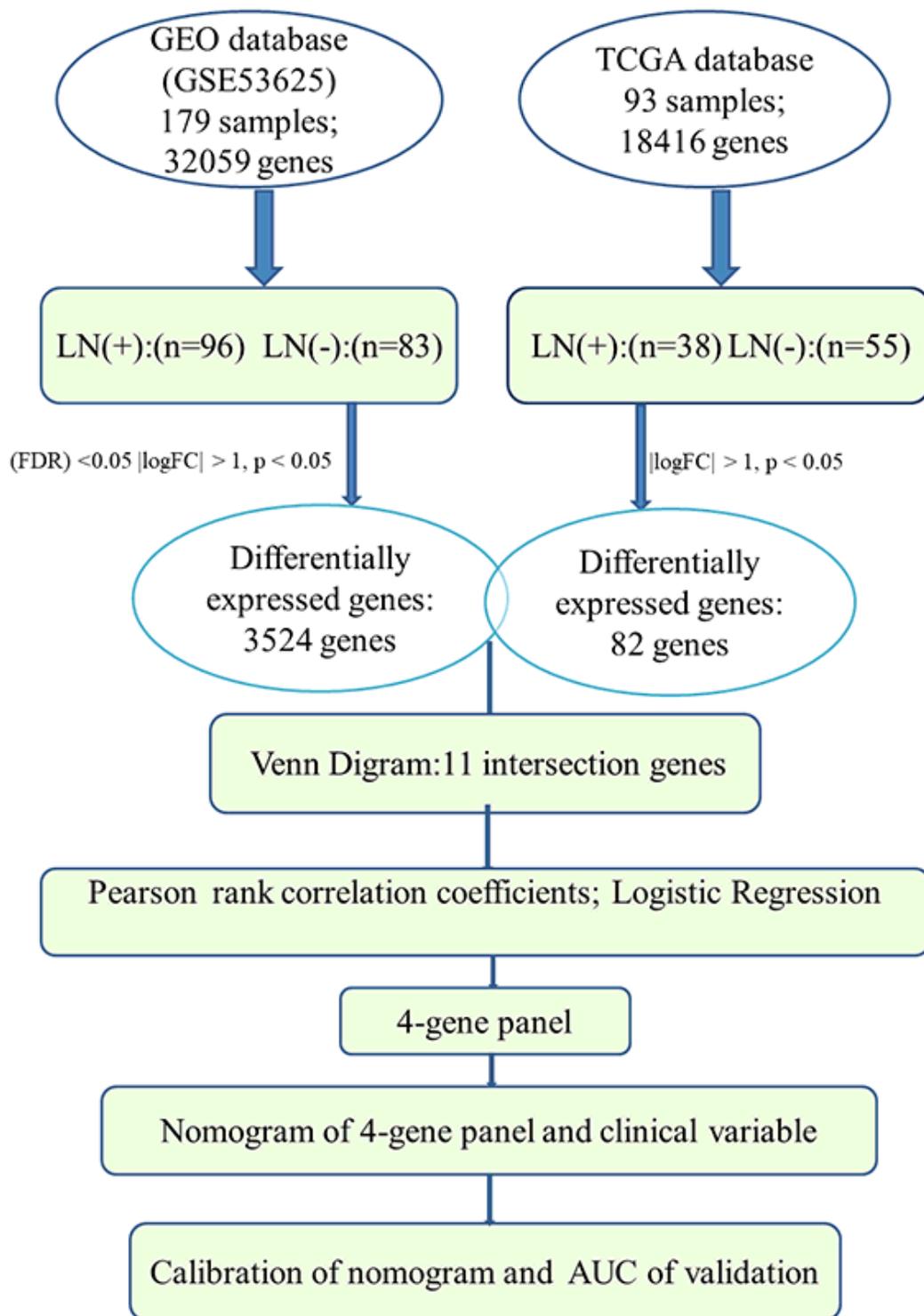


Figure 1

Flow Chart

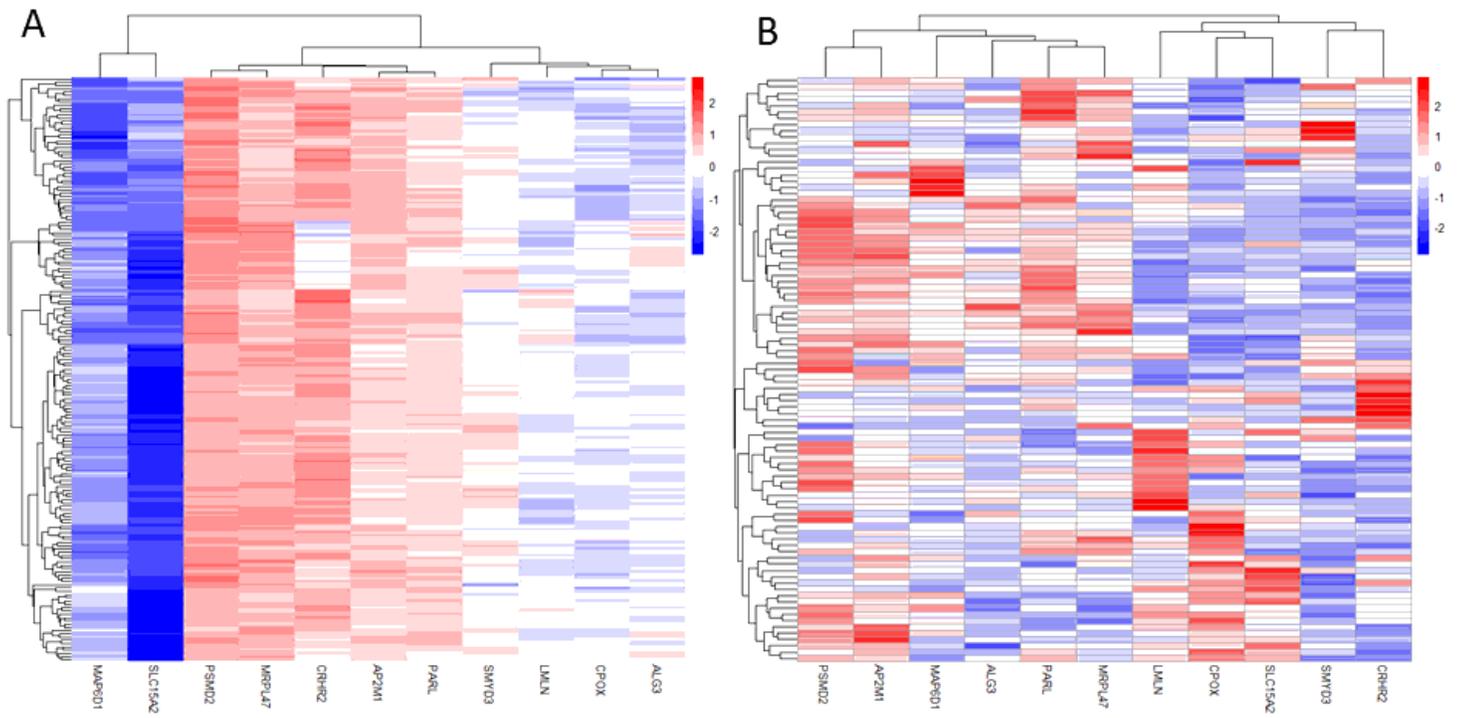


Figure 2

A-B the expression levels of the 11 communal DEGs

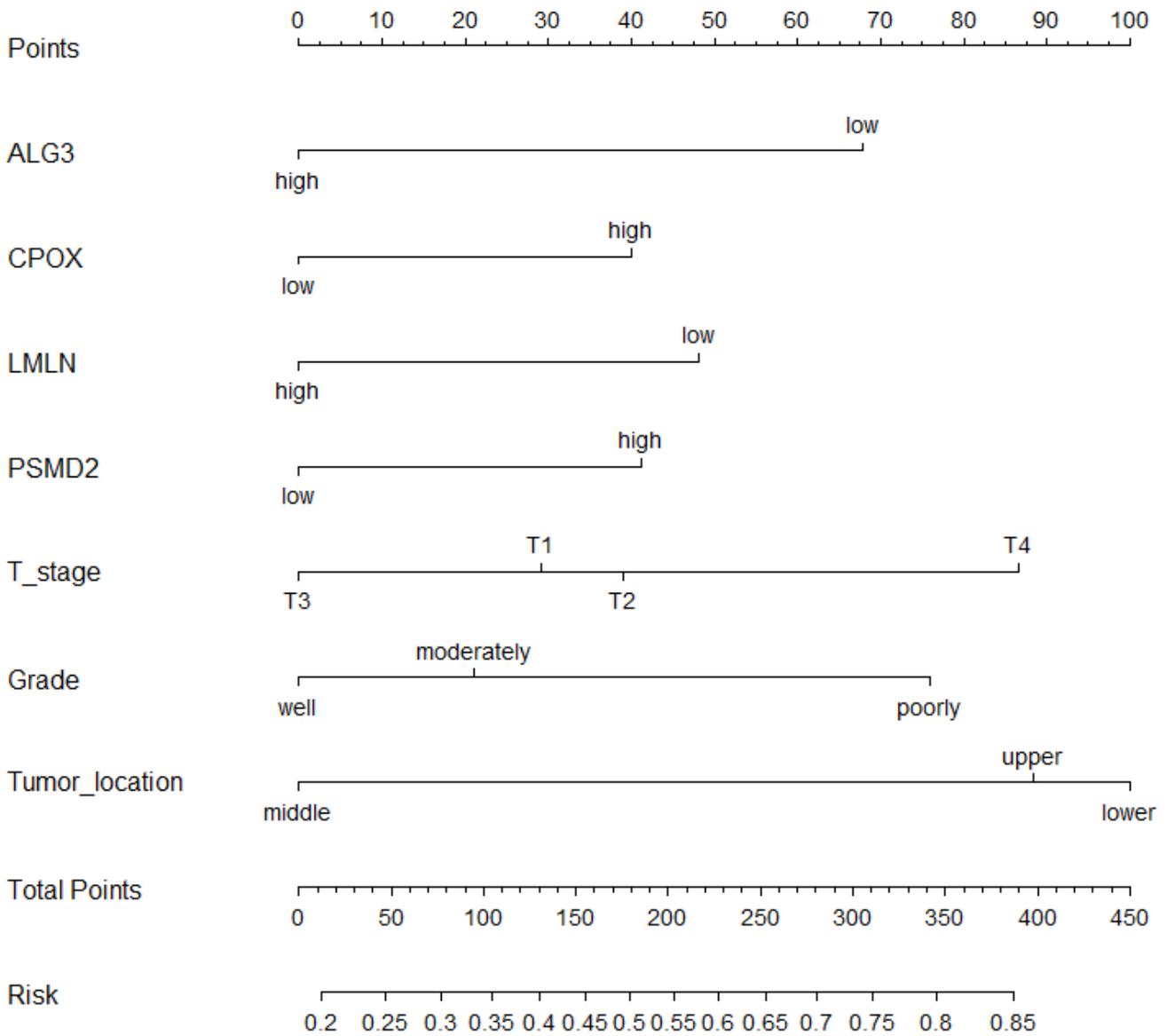


Figure 3

the nomogram consisting of 7 variables including 4 DEGs, T stage, G stage and tumor location was established to predict the probability of LNM

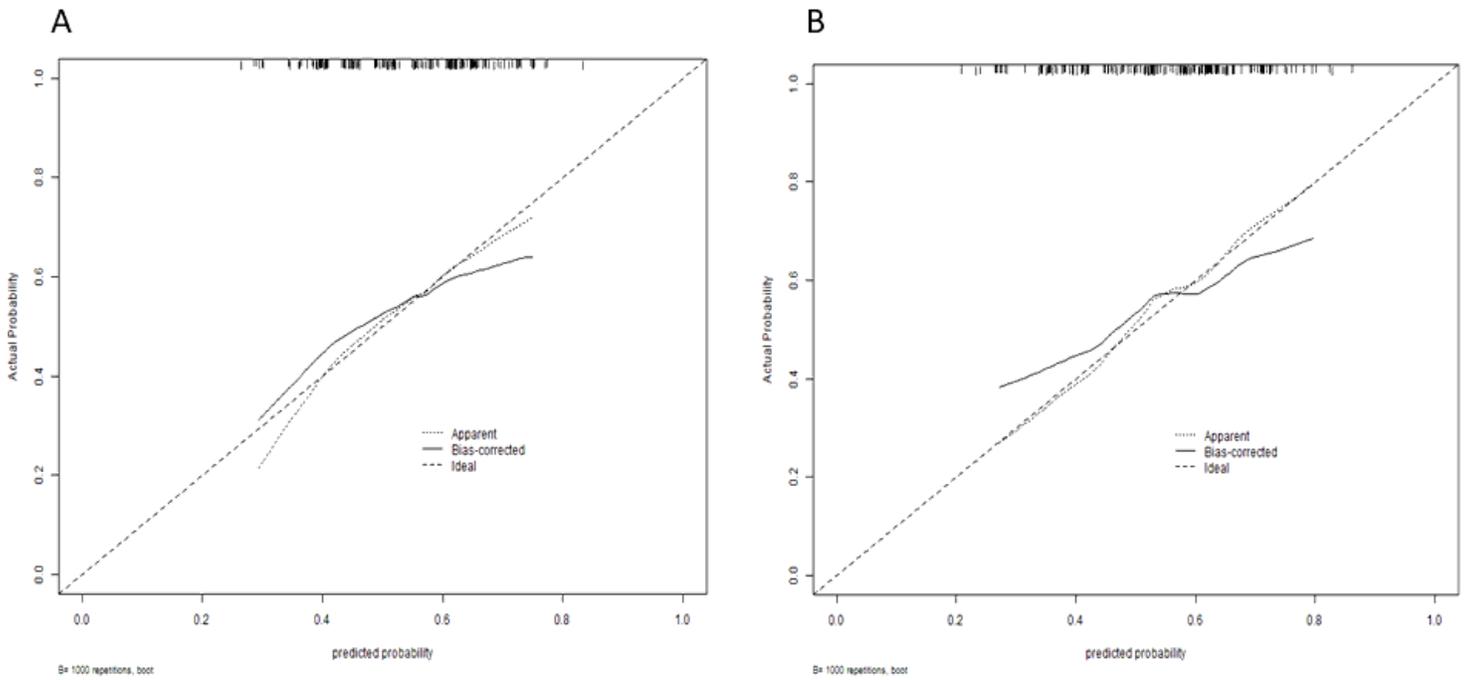


Figure 4

Calibration curves regarding the model performance in the two datasets

Supplementary Files

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

- [S3.tif](#)
- [S2.tif](#)
- [S1.tif](#)
- [SupplementaryTables.doc](#)