A Genomic-Clinicopathologic Nomogram for the Preoperative Prediction of Lymph Node Metastasis in Gastric Cancer

Xin Zhong

Zhejiang University School of Medicine Sir Run Run Shaw Hospital <https://orcid.org/0000-0001-9944-0484>

Feichao Xuan

Zhejiang University School of Medicine Sir Run Run Shaw Hospital

Yun Qian

Zhejiang University School of Medicine Sir Run Run Shaw Hospital

Junhai Pan

Zhejiang University School of Medicine Sir Run Run Shaw Hospital

Suihan Wang

Zhejiang University School of Medicine Sir Run Run Shaw Hospital

Wenchao Chen

Zhejiang University School of Medicine Sir Run Run Shaw Hospital

Tianyu Lin

Zhejiang University School of Medicine Sir Run Run Shaw Hospital

Hepan Zhu

Zhejiang University School of Medicine Sir Run Run Shaw Hospital

Xianfa Wang (✉ 3195011@zju.edu.cn)

Zhejiang University School of Medicine Sir Run Run Shaw Hospital

Guanyu Wang

Zhejiang University School of Medicine Sir Run Run Shaw Hospital

Primary research

Keywords: Gastric cancer, Lymph node metastasis, Gene signature, nomogram

Posted Date: August 11th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-55337/v1>

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

[Read Full License](#)

Abstract

Background: Preoperative evaluation of the lymph node (LN) state is of pivotal significance for therapy strategy decisions in gastric cancer (GC) patients. However, there is lack of noninvasive method that can identify such status preoperatively. We aimed at developing a genomic biosignature based model to forecast the possibility of LN metastasis in GC patients.

Methods: We employed the RNA profile retrieving strategy and conducted RNA expression profiling in a large GC cohort (GSE62254, $n = 300$) from Gene Expression Ominus (GEO). In the exploratory stage, 300 GC patients from GSE62254 were involved and the differentially expressed RNAs (DERs) for LN-status were determined using R software. The GC samples in GSE62254 were randomly divided into a learning set ($n = 210$) and a verification set ($n = 90$). By performing the Least absolute shrinkage and selection operator (LASSO) regression approach, a set of 23-RNA signature was established and the signature based nomogram was subsequently built for distinguishing LN condition. The diagnostic effectiveness of this model was assessed, as well as the clinical performance subsequently assessed using the decision curve analysis (DCA). Metascape was used for bioinformatic analysis of the DERs

Results: Based on this genomic signature, we established a nomogram which robustly distinguished LN status in the learning (AUC = 0.916, 95% CI 0.833–0.999) and verification sets (AUC = 0.775, 95% CI 0.647–0.903). DCA demonstrated the clinical value of this nomogram. Functional enrichment analysis of the DERs was conducted using bioinformatics methods which posited that these DERs were involved with several lymphangiogenesis-correlated cascades.

Conclusions: Here, we present a genomic signature based nomogram that integrates the 23-RNA biosignature based scores and Lauren classification. This model can be readily utilized to estimate the possibility of LN metastasis with good performance in GC. The uncovering of the DERs reveals the prospective biogenesis of LN metastasis in GC.

Background

Gastric cancer (GC) constitutes the 5th most frequent cancer and the 3rd top cause of cancer-associated mortality globally [1]. Some studies demonstrated that LN metastasis is an independent risk index for poor prognosis of GC [2, 3]. Precise and exact preoperative identification of LN involvement is of vital significance for treatment strategy decisions in patients with GC [4, 5]. Several studies suggested that clinicopathologic factors like lymphatic invasion or pathological differentiation are associated with LN metastasis, however they can hardly be obtained preoperatively [6, 7]. The current preoperative prediction of LN metastasis relies primarily on morphological features of the lymph nodes referring to CT, leading to unfavorable sensitivity [8]. Meanwhile, present tumor biosignatures, including carcinoembryonic antigen (CEA), as well as carbohydrate antigen 199(CA-199) have shown poor performance for predicting LN metastasis in GC. Therefore, novel diagnostic biomarkers are needed to improve on current strategies for predicting LN metastasis in patients with GC. Currently, gene expression reports have been conducted to

elucidate the distinct molecular biosignatures for LN metastases. Daisuke Izumi et al proposed a 15-gene signature for detecting LN metastasis in GC [9]. Song et al developed a co-expression network of RNAs for assessing LN metastasis in patients with GC [10]. These studies showed genes had strong power of prediction value for detecting LN metastasis. However, clinicopathologic factors reported to be associated with LN status was not involved in these studies [11–13]. Nomogram is a visual predictive tool and has been developed to quantify risk factors of LN metastasis in several carcinomas [14, 15], including in early GC [16]. The current nomogram however only integrates clinical and postsurgical factors, which would restrict their clinical value. Therefore, we aimed to generate and verify a nomogram that integrated both the gene biosignature and preoperative clinicopathologic parameters for individually predicting LN metastasis in GC.

Methods

Protocol approval

Gene expression information and sample data from GSE62254 dataset in this research was retrieved from GEO (<http://www.ncbi.nlm.nih.gov/geo/>). Clinical information of these samples were directly downloaded from the authors website (<https://www.nature.com/articles/nm.3850>). The study dataset obtained from GEO database was fully anonymous and therefore the ethics approval was not required.

GC datasets preparation

The GSE62254 dataset was retrieved from GEO in its processed format, using the package ‘GEOquery’(version 2.56.0,) in R. The samples in GSE62254 data were randomly clustered into a learning set and a verification set.

Differential Expression Analysis

Human gene annotation files (GRCh38.p12) was derived from the Ensembl repository (<https://asia.ensembl.org/index.html>) for RNAs annotation. Samples in the GSE62254 dataset were divided into LN-negative and LN-positive arms referring to the source information. The expression data of RNAs were examined using the package limma (version 3.34.7) [17]. DEs were distinguished between the two groups according to the false discovery rate (FDR) < 0.05. Based on the R package pheatmap (version 1.0.8), hierarchical clustering analysis was performed [18]. A volcano plot was drawn by the ggplot2 package.

Development of the 23-RNA Signature

The LASSO regression approach which is applicable for the regression of high-dimensional data was performed by using R package “glmnet” (version 3.5.3) [19]. The most estimated genes from the learning set were selected using the LASSO regression method. A risk score computed for each patient using the linear combination of selected genes that were weighted by their corresponding coefficients. The risk score of the GC for each sample was computed using the following formula:

Risk score (RS) = $\sum$ (coefficients *ExpRNA) + Intercept

The R package “OptimalCutpoints” was applied for measuring the optimum cutoff point of risk score. The optimum cutoff was employed to cluster the patients into high- or low-risk classes obtained when the Youden index in ROC curve predicting LN metastasis reached its maximum in the learning set. Samples were clustered into high- or low-risk clusters utilizing the optimum cutoff.

Construction and assessment of genomic signature based Model

Candidate predictors including age, gender, Borrmann classification, Lauren classification, tumor location and the risk scores were embedded into logistic regression analysis to design a diagnostic model for predicting LN metastasis in the learning set. Using the likelihood ratio test with Akaike’s information criterion as the stopping rule, we applied the backward step-wise selection [20, 21]. The forest plot of clinical features and risk scores was drawn using forestplot package. To offer the clinician a quantitative technique to predict individual likelihood of LN metastasis, nomogram prediction model was constructed based on the independent risk factors using the R package rms (version 5.1-2) [22]. By using the “pROC” package, receiver operating characteristic (ROC) assessment was conducted to inspect the sensitivity and specificity of the nomogram. We draw the ROC curve using the “ggplot2” package. The calibration curve was subsequently utilized to examine the effectiveness of the nomogram with additional 1000 bootstrap samples to decrease the over fit bias. Decision curve analysis (DCA) was applied to inspect the clinical application of the gene signature based mode [23].

Functional enrichment analysis

Metascape (<http://metascape.org/gp/index.html>) was applied to predict the potential biological functions of the differentially expressed genes [24].

Statistical analyses

The chi-square test was applied in categorical variables analysis between the two groups. Statistical analyses were performed by SPSS software (version 24) or R software (version 3.5.3). All tests were dual-sided and $p < .05$ signified statistical significance.

Results

Patient characteristics

The samples in GSE62254 data were randomly clustered into a learning set ($n = 210$, Additional file 1) and a verification set ($n = 90$, Additional file 2). The baseline features of all patients are indicated in Table 1. The rates of LN metastasis were 88.1% in the learning set and 85.6% in the verification sets respectively. There was no difference in patients’ characteristics between the two sets.

Table 1
Baseline features of all subjects

	Learning set(n = 210)	Verification set(n = 90)	pValue
Gender			0.539754
Male	137	62	
Female	73	28	
Age			0.667548
>=65	111	50	
< 65	99	40	
Borrmann			0.560256
B-I	11	5	
B-II	70	34	
B-III	98	43	
B-IV	31	8	
Lauren			0.216635
Intestinal	91	51	
Mixed	13	4	
Diffuse	101	35	
Location			0.787408
Antrum	108	47	
Body	73	34	
Cardia	24	8	
Whole	5	1	
N			0.544463
Negative	25	13	
Positive	185	77	
Categorical variables were compared by Chi squared test or Fisher's exact test as appropriate			
LN, lymph node;			

Differential Expression Analysis

Overall, 14,651 mRNAs, 840 lncRNAs, and 111 miRNAs were annotated from the GSE62254 datasets. The 300 GC samples in the GSE62254 dataset were divided into LN-negative (38 samples) and LN-positive (262 samples) groups. 186 DERs were screened out under the defined thresholds between the LN-positive and the LN-negative groups. Among the 186 DERs, 70 DERs are upregulated and 116 DERs are downregulated. Based on the expression of the DERs, The heatmap and volcano plot are shown, respectively, in Figs. 1 and 2.

Construction of 23-RNA Signature based Risk Score

A total of 23 DERs were found to be significantly associated to LN status using the LASSO regression method(Fig. 3a-b).The risk score formula was subsequently constructed based on the 23 DERs, and this formula was as follow:

$$\begin{aligned} RS = & 0.3370*ExpTRAPPC10+(-0.6895)*ExpRHOA + 0.0452*ExpIGFBP2 + 1.4984*ExpC11orf80+ \\ & (-0.0937)*ExpZNF74+(-0.9888)*ExpFOXN2 + 0.6580*ExpGOLGA8A + 0.9803*ExpRSRP1+ \\ & (-0.4094)*ExpUSP10 + 0.3896*ExpCLTB+(-1.2924)*ExpPIK3R1 + 1.5335*ExpPABPN1+ \\ & (-0.3669)*ExpCLCN4+(-1.4978)*ExpPAR6B + 0.0329*ExpTRPA1+(-0.0174)*ExpBAG3 + \\ & 0.4511*ExpZNF26 + 0.0381*ExpGDPD3 + 1.1286*ExpSPTBN5 + 2.3647*ExpKLHL28 + \\ & 1.0420*ExpGTPBP8 + 2.5667*ExpTXNDC11 + 0.1489*ExpTMEM163 + Intercept \end{aligned}$$

Distribution of risk score between LN-negative and LN-positive groups with significant difference ($p < 0.05$) were shown respectively in Fig. 4a-b. The cutoff of the risk scores was calculated, and the samples were separately clustered into high or low risk classes in both the learning and verification sets. The cutoff value (1.3806) was acquired when the ROC curve reached optimum sensitivity (94.05%) and specificity (88.00%) for predicting LN metastasis. Patients with a risk score higher than 1.3806 were assigned to the high-risk group ($n = 177$) and the rest ($n = 33$) were assigned to the low-risk arm in the learning set.

Construction and Verification of genomic signature based Model

By using logistic regression analysis, Lauren classification (odds ratio [OR] = 2.2493, 95% CI, 1.0001–5.0585; $P < 0.05$) and risk scores (OR = 155.2005, 95%CI, 32.6914-736.8059; $P < 0.05$) were confirmed as independent risk factors for LN metastasis (Table 2, Fig. 5). Basing on these two independent predictive factors, a nomogram model was subsequently built (Fig. 6). LN metastasis probability could be easily calculated based on their Lauren classification and risk scores. ROC evaluation was applied to examine sensitivity and specificity of the nomogram. The result showed that this nomogram had an optimum sensitivity of 94.1% and specificity of 88.0% for predicting LN metastasis in the learning set, and an optimum sensitivity of 74% and specificity of 76.9% in the verification set respectively. The area under curve (AUC) were 0.916(95% CI: 0.833–0.999) for learning set and 0.775(95% CI: 0.647–0.903) for the verification set respectively, which suggested that the nomogram had good utility (Fig. 7). Additionally, the predicted probability of LN metastasis was further compared with the authentic probability by the calibration curve in the learning set. The results revealed consistence between the estimated likelihood of

LN metastasis and the authentic probability (Fig. 8). The DCA for genomic-clinicopathologic nomogram demonstrated that if the threshold ranged from 0.20 to 0.95, the nomogram model is more beneficial relative to either the treat-all-patients scheme or the treat-none scheme (Fig. 9).

Table 2 Multivariate evaluations to evaluate potential predictive factors for LN metastasis

Variable	Logistic regression analysis		
	OR	95%CI	pValue
Gender	0.3918	0.0794-1.935	0.2502
Age	0.8887	0.2263-3.4897	0.8658
Borrmann	1.1643	0.4917-2.7565	0.7295
Lauren	2.2493	1.0001-5.0585	0.0499
Location	1.0608	0.4435-2.5371	0.8945
Risk score	155.2005	32.6914-736.8059	<0.0001

LN, lymph node; OR, odds ratio; CI, confidence interval

Functional enrichment analysis

Metascape was applied for cascade and process enrichment assessments of these DERs. The top 15 groups with their illustrative enriched terms were present in Fig. 10. A sub-cluster of the enriched terms had been selected and regarded as a network plot, in which the terms with a similarity > 0.3 were connected by edges (Figs. 11 and 12). Specifically, the DERs enriched were associated with several pathways, such as Signaling by platelet derived growth factor (PDGF) and Intrinsic Pathway for Apoptosis.

Discussion

LN metastasis is among the primary conditions that influence the prognosis of GC [2, 3]. Precise preoperative determination of LN participation in GC is pivotal for clinical decision-making. Less invasive treatment such as endoscopic submucosal resection (ESD) can be carried out effectively for LN negative patients in early GC. However ESD should be avoided for early GC patients with high risk of LN metastasis [25, 26]. For localized LN negative patients of GC, limited LN resection is recommended to reduce postoperative complication. Surgical resection with extensive lymphadenectomy is necessary for advanced GC patients with LN metastasis [5]. Therefore, it is of crucial significance to accurately determine the extent and degree of LN metastasis for treatment strategy decisions. With the development of high throughput sequencing (HTS) technologies, molecular portrait of GC have been comprehensively analyzed by gene-expression profile [27, 28]. As RNA-sequencing technology provides molecular insights into tumor biology process, we focused on building a genomic signature based Nomogram for predicting

LN metastasis in GC. By using cDNA microarrays, several studies have reported certain gene expression-based biomarkers for predicting LN metastasis in GC [29–31]. However, these studies lack clinical characteristics which were reported to be associated with LN status in GC [11–13].

GC could be grouped into intestinal or diffuse kinds referring to the Lauren classification [32]. It is reported that intestinal type of GC stems from premalignant lesions via an initial *Helicobacter pylori* (*H. pylori*) triggered chronic gastritis and successive atrophic and metaplastic gastritis [33]. On the other hand, the diffuse form of GC are directly triggered by bioactive inflammation of the gastric mucosa [34, 35]. Diffuse forms are normally prevalent in younger patients with an elevated risk of LN metastasis compared with intestinal types [36–38]. Our study demonstrated that Lauren classification was an independent risk index for LN metastasis and diffuse type was related with elevated risk of LN metastasis relative to the intestinal form.

Here, we constructed and verified a diagnostic, genomic biosignature based nomogram as a noninvasive strategy for preoperative estimation of LN metastasis in GC. This nomogram incorporates two items of genomic signature based risk scores and Lauren classification. The effectiveness of this estimation model showed a good agreement between prediction and actual probability according to its calibration plot. The areas under the ROC curve for learning and verification set suggested that the nomogram had good utility. In the view of clinicians, clinical performance of a model is more valuable than discrimination ability, such as false negative and false positive. Therefore, decision analysis of curve was performed to quantify different strategies and determine an optimal threshold range. This LN metastasis prediction model would help clinicians to balance quality of life and the aggressive resection.

Providing insights into the potential biological process, “metascape” was conducted for the functional and enrichment analysis of DERs. Taken together, the DERs were enriched in three signaling pathways including Signaling by PDGF, Interleukin-7 signaling and Intrinsic Pathway for Apoptosis. The PDGF receptor cascade constitutes a signaling network, which is essential for the growth of cells of mesenchymal parentage. Dysregulation of this pathway can lead to extracellular matrix reconstruction in a tumor-enhancement manner to promote the migration, infiltration, angiogenesis, and likely also lymphangiogenesis [39, 40]. Interleukin-7 (IL-7) has been detailed to be required for both B-cell and T-cell proliferation and its lack would result in immature immune cell arrest. IL-7 can modulate cell growth, as well as apoptosis and modulate cancer lymphangiogenesis [41–42]. Apoptosis is a form of programmed cell death. Insufficient apoptosis is associated with neoplastic diseases [43]. However, the specific molecular mechanism between the differentially expressed genes and potential pathways remained unclear, which provide new clues and molecular targets for identification and specific treatment of GC with LN metastasis.

Compared with previous nomograms [14–16], our model incorporates both preoperative clinical characteristics and genomic signature based risk scores. This model showed high diagnostic accuracy for predicting LN metastasis. However, there are some limitations in this study. First, we did not perform external verification using data from another institution for this model. Second, clinicopathological

factors, such as CEA level and CT-reported LN status, were not available in GSE62254 dataset. These important clinical features thus, could not be examined in this study. Finally, studies should be conducted to elucidate the function of DERs for illuminating the mechanism of LN metastasis.

Conclusions

In conclusion, this nomogram incorporating both the genomic signature based risk score and Lauren classification can be readily applied to estimate LN metastasis in GC preoperatively.

Abbreviations

GC: gastric cancer; LN: lymph node; GEO: the Gene Expression Omnibus; DERs: differentially expressed RNAs; LASSO : Least absolute shrinkage and selection operator; DCA: decision curve analysis; AUC :The area under curve; ROC: receiver operating characteristic; FDR :false discovery rate; OR: odds ratio; CI: confidence interval; PDGF: platelet derived growth factor; ESD :endoscopic submucosal resection; IL-7:Interleukin-7;

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The dataset analyzed in this study is provided as the additional document at the end of the current article.

Competing interests

The authors disclose that they have no competing interests.

Funding

This work was funded by the National Nature Science Foundation of China (Grant Nos.81272493 and 81472213) and the Zhejiang Provincial Natural Science Foundation of China(Grant No. LQ19H160044).

Authors' contributions

XZ, XFW and GYW designed the study. XZ analyzed, as well as interpreted the data. XZ, FCX and YQ drafted the manuscript. JHP, SHW, WCC, TYL and HPZ helped to revise the manuscript. All authors read and ratified the final draft.

Acknowledgments

We appreciate GEO database for providing the original study data. We thank QWZ for her invaluable contribution in biostatistics analysis.

Author details

¹Department of General Surgery, Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University, East Qing chun Road 3, 310016, Hangzhou, Zhejiang, China.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68:394-424. <https://doi.org/10.3322/caac.21492>
2. Lo SS, Wu CW, Chen JH, et al. Surgical results of early gastric cancer and proposing a treatment strategy. *Ann Surg Oncol*. 2007;14:340-347. <https://doi.org/10.1245/s10434-006-9077-x>. doi: 10.1245/s10434-006-9077-x
3. Mu GC, Huang Y, Liu ZM, Wu XH, Qin XG, Chen ZB. Application value of nomogram and prognostic factors of gastric cancer patients who underwent D2 radical lymphadenectomy. *BMC Gastroenterol*. 2019;19:188. <https://doi.org/10.1186/s12876-019-1098-6>
4. Lou N, Zhang L, Chen XD, et al. A novel scoring system associating with preoperative platelet/lymphocyte and clinicopathologic features to predict lymph node metastasis in early gastric cancer. *J Surg Res*. 2017;209:153-161. <https://doi.org/10.1016/j.jss.2016.10.011>
5. Japanese Gastric Cancer Association. Japanese gastric cancer treatment guidelines 2010 (ver. 3). *Gastric Cancer*. 2011;14(2):113-123. <https://doi.org/10.1007/s10120-011-0042-4>
6. Chu YN, Yu YN, Jing X, et al. Feasibility of endoscopic treatment and predictors of lymph node metastasis in early gastric cancer. *World J Gastroenterol*. 2019;25(35):5344-5355. <https://doi.org/10.3748/wjg.v25.i35.5344>
7. Huang Q, Cheng Y, Chen L, et al. Low risk of lymph node metastasis in 495 early gastric cardiac carcinomas: a multicenter clinicopathologic study of 2101 radical gastrectomies for early gastric carcinoma. *Mod Pathol*. 2018;31(10):1599-1607. <https://doi.org/10.1038/s41379-018-0063-1>
8. Kim AY, Kim HJ, Ha HK. Gastric cancer by multidetector row CT: preoperative staging. *Abdom Imaging*. 2005;30(4):465-472. <https://doi.org/10.1007/s00261-004-0273-5>
9. Izumi D, Gao F, Toden S, 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;41:268-275. <https://doi.org/10.1016/j.ebiom.2019.01.057>
10. Song Z, Zhao W, Cao D, Zhang J, Chen S. Elementary screening of lymph node metastatic-related genes in gastric cancer based on the co-expression network of messenger RNA, microRNA and long non-coding RNA. *Braz J Med Biol Res*. 2018;51(4):e6685. <https://doi.org/10.1590/1414-431x20176685>
 11. Oka S, Tanaka S, Kaneko I, et al. Advantage of endoscopic submucosal dissection compared with EMR for early gastric cancer. *GastrointestEndosc*. 2006;64:877-883. <https://doi.org/10.1016/j.gie.2006.03.932>
 12. Ajani JA, Bentrem DJ, Besh S, et al. Gastric cancer, version 2.2013: featured updates to the NCCN Guidelines. *J Natl ComprCancNetw*. 2013;11:531-546. <https://doi.org/10.6004/jnccn.2013.0070>
 13. Hyung WJ, Cheong JH, Kim J, Chen J, Choi SH, Noh SH. Application of minimally invasive treatment for early gastric cancer. *J Surg Oncol*. 2004;85:181-186. <https://doi.org/10.1002/jso.20018>
 14. Klar M, Jochmann A, Foeldi M, et al. The MSKCC nomogram for prediction the likelihood of non-sentinel node involvement in a German breast cancer population. *Breast Cancer Res Treat*. 2008;112:523-531. <https://doi.org/10.1007/s10549-007-9884-1>
 15. Briganti A, Larcher A, Abdollah F, et al. Updated nomogram predicting lymph node invasion in patients with prostate cancer undergoing extended pelvic lymph node dissection: the essential importance of percentage of positive cores. *Eur Urol*. 2012;61(3):480-487. <https://doi.org/10.1016/j.eururo.2011.10.044>
 16. Zheng Z, Zhang Y, Zhang L, et al. A nomogram for predicting the likelihood of lymph node metastasis in early gastric patients. *BMC Cancer*. 2016;16:92. Published 2016 Feb 12. <https://doi.org/10.1186/s12885-016-2132-5>
 17. Smyth GK. limma: Linear models for microarray data. In: Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S, editors. *Bioinformatics & Computational Biology Solutions Using R & Bioconductor*. New York, NY: Springer (2011). p.397–420. doi: 10.1007/0-387-29362-0_23
 18. Wang L, Cao C, Ma Q, et al. RNA-seq analyses of multiple meristems of soybean: novel and alternative transcripts, evolutionary and functional implications. *BMC Plant Biol*. 2014;14:169. Published 2014 Jun 17. <https://doi.org/10.1186/1471-2229-14-169>
 19. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med*. 2007;26(30):5512-5528. <https://doi.org/10.1002/sim.3148>
 20. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015;350:g7594. Published 2015 Jan 7. <https://doi.org/10.1136/bmj.g7594>
 21. Sauerbrei W, Boulesteix AL, Binder H. Stability investigations of multivariable regression models derived from low- and high-dimensional data. *J Biopharm Stat*. 2011;21(6):1206-1231. <https://doi.org/10.1080/10543406.2011.629890>

22. Eng KH, Schiller E, Morrell K. On representing the prognostic value of continuous gene expression biomarkers with the restricted mean survival curve. *Oncotarget*. 2015;6(34):36308-36318.<https://doi.org/10.18632/oncotarget.6121>
23. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak*. 2008;8:53. Published 2008 Nov 26. <https://doi.org/10.1186/1472-6947-8-53>
24. Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019;10(1):1523. Published 2019 Apr 3. <https://doi.org/10.1038/s41467-019-09234-6>
25. ASGE TECHNOLOGY COMMITTEE, Kantsevoy SV, Adler DG, et al. Endoscopic mucosal resection and endoscopic submucosal dissection. *GastrointestEndosc*. 2008;68(1):11-18. <https://doi.org/10.1016/j.gie.2008.01.037>
26. ASGE Standards of Practice Committee, Gan SI, Rajan E, et al. Role of EUS. *GastrointestEndosc*. 2007;66(3):425-434. <https://doi.org/10.1016/j.gie.2007.05.026>
27. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. 2014;513(7517):202-209.<https://doi.org/10.1038/nature13480>
28. Cristescu R, Lee J, Nebozhyn M, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med*. 2015;21(5):449-456. <https://doi.org/10.1038/nm.3850>
29. Weiss MM, Kuipers EJ, Postma C, et al. Genomic profiling of gastric cancer predicts lymph node status and survival. *Oncogene*. 2003;22(12):1872-1879. <https://doi.org/10.1038/sj.onc.1206350>
30. Teramoto K, Tada M, Tamoto E, et al. Prediction of lymphatic invasion/lymph node metastasis, recurrence, and survival in patients with gastric cancer by cDNA array-based expression profiling. *J Surg Res*. 2005;124(2):225-236. <https://doi.org/10.1016/j.jss.2004.10.003>
31. Marchet A, Mocellin S, Belluco C, et al. Gene expression profile of primary gastric cancer: towards the prediction of lymph node status. *Ann Surg Oncol*. 2007;14(3):1058-1064. <https://doi.org/10.1245/s10434-006-9090-0>
32. LAUREN P. THE TWO HISTOLOGICAL MAIN TYPES OF GASTRIC CARCINOMA: DIFFUSE AND SO-CALLED INTESTINAL-TYPE CARCINOMA. AN ATTEMPT AT A HISTO-CLINICAL CLASSIFICATION. *Acta Pathol Microbiol Scand*. 1965;64:31-49.<https://doi.org/10.1111/apm.1965.64.1.31>
33. Correa P. Human gastric carcinogenesis: a multistep and multifactorial process-First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. *Cancer Res*. 1992;52(24):6735-6740.
34. Watanabe M, Kato J, Inoue I, et al. Development of gastric cancer in nonatrophic stomach with highly active inflammation identified by serum levels of pepsinogen and Helicobacter pylori antibody together with endoscopic rugal hyperplastic gastritis. *Int J Cancer*. 2012;131(11):2632-2642. <https://doi.org/10.1002/ijc.27514>

35. Nardone G, Rocco A, Malfertheiner P. Review article: helicobacter pylori and molecular events in precancerous gastric lesions. *Aliment Pharmacol Ther.* 2004;20(3):261-270.
<https://doi.org/10.1111/j.1365-2036.2004.02075.x>
36. Adachi Y, Yasuda K, Inomata M, Sato K, Shiraishi N, Kitano S. Pathology and prognosis of gastric carcinoma: well versus poorly differentiated type. *Cancer.* 2000;89(7):1418-1424.
[https://doi.org/10.1002/1097-0142\(20001001\)89:7<1418::aid-cnrcr2>3.0.co;2-a](https://doi.org/10.1002/1097-0142(20001001)89:7<1418::aid-cnrcr2>3.0.co;2-a)
37. Ribeiro MM, Sarmento JA, SobrinhoSimões MA, Bastos J. Prognostic significance of Lauren and Ming classifications and other pathologic parameters in gastric carcinoma. *Cancer.*1981;47(4):780-784.[https://doi.org/10.1002/1097-0142\(19810215\)47:4<780::aid-cnrcr2820470424>3.0.co;2-g](https://doi.org/10.1002/1097-0142(19810215)47:4<780::aid-cnrcr2820470424>3.0.co;2-g)
38. Lee T, Tanaka H, Ohira M, et al. Clinical impact of the extent of lymph node micrometastasis in undifferentiated-type early gastric cancer. *Oncology.* 2014;86(4):244-252.
<https://doi.org/10.1159/000358803>
39. Ehnman M, Östman A. Therapeutic targeting of platelet-derived growth factor receptors in solid tumors. *Expert OpinInvestig Drugs.* 2014;23(2):211-226.
<https://doi.org/10.1517/13543784.2014.847086>
40. Andrae J, Gallini R, Betsholtz C. Role of platelet-derived growth factors in physiology and medicine. *Genes Dev.* 2008;22(10):1276-1312. <https://doi.org/10.1101/gad.1653708>
41. Lin J, Zhu Z, Xiao H, et al. The role of IL-7 in Immunity and Cancer. *Anticancer Res.* 2017;37(3):963-967. <https://doi.org/10.21873/anticancer.11405>
42. Jian M, Yunjia Z, Zhiying D, Yanduo J, Guocheng J. Interleukin 7 receptor activates PI3K/Akt/mTOR signaling pathway via downregulation of Beclin-1 in lung cancer. *Mol Carcinog.* 2019;58(3):358-365.
<https://doi.org/10.1002/mc.22933>
43. Matsuura K, Canfield K, Feng W, Kurokawa M. Metabolic Regulation of Apoptosis in Cancer. *Int Rev Cell Mol Biol.* 2016;327:43-87. <https://doi.org/10.1016/bs.ircmb.2016.06.006>

Figures

Figure 1

Volcano plot: The volcano plot(the red and blue dots represent up- and down-regulation of differentially expressed RNAs respectively, false discovery rate< 0.05);

Figure 3

Selection of the genes through the LASSO approach. (a) Selection of tuning parameter (λ) via 10-fold cross-verification with minimum criteria. The area under curve was plotted versus $\log(\lambda)$. Dotted vertical lines were drawn at the optimal values using the minimum criteria and the 1 standard error of the minimum criteria (the 1-SE criteria). The optimal λ value of 0.033, with $\log(\lambda) = -3.411$ was chosen based on minimum criteria. (b) LASSO coefficient profiles of the 186 differentially expressed RNAs. A coefficient profile plot was generated against the $\log(\lambda)$ sequence. Vertical line was drawn at the optimal value where optimal λ led to 23 nonzero coefficients.

Figure 4

Violin plot: (a) Distribution of risk score in learning set; (b) Distribution of risk score in verification set.

Figure 5

Forest plot demonstrating the findings of the logistic regression model for studying the potential risk factors of lymph node metastasis in the learning set.

Figure 6

Developed a genomic signature based nomogram. The nomogram was designed in the learning set, with the 23-mRNA biosignature based risk score and Lauren classification integrated.

Figure 7

Assessment of the nomogram for predicting lymph node metastasis in the learning set and the verification set. The area under curve of nomogram in learning set was 0.916 (95% CI: 0.833-0.999) after 500 repetitions of bootstrap (DeLong). Area under curve of nomogram in verification set was 0.775 (95% CI: 0.647-0.903).

Calibration Curve

Figure 8

Calibration plot in the learning set. After 1000 repetitions of bootstrap, the bias-corrected calibration curve (solid line) was close to the ideal curve (dashed line), indicating a good agreement between the estimated and authentic likelihood (mean absolute error=0.021).

Figure 9

Decision curve analysis for the genomic signature based nomogram. The y-axis represents the net benefit. Decision curve analysis revealed that if the threshold likelihood is over 20%, using the nomogram in the current study to estimate lymph node metastases is more beneficial than the treat-all-patients scheme or the treat-none scheme.

Figure 10

Bar graph of enriched terms across the differentially expressed RNAs, colored by p-values.

Figure 11

Network of enriched terms: Colored by the cluster-ID, in which the nodes with similar cluster ID are frequently close to each other.

Figure 12

Network of enriched terms: Colored by p-value, in which the terms with more genes tend to have a more remarkable p-value.

Supplementary Files

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

- [Verificationset.xls](#)
- [Learningset.xls](#)