EBioMedicine 41 (2019) 268–275

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

EBioMedicine
Published by THE LANCET

A genomewide transcriptomic approach identifies a novel gene expression signature for the detection of lymph node metastasis in patients with early stage gastric cancer

Daisuke Izumi [a,b,c,1], Feng Gao [d,1], Shusuke Toden [a], Fuminori Sonohara [a,e], Mitsuro Kanda [e], Takatsugu Ishimoto [b,f], Yasuhiro Kodera [e], Xin Wang [d,g,**], Hideo Baba [b], Ajay Goel [a,*]

[a] Center for Gastrointestinal Research, Baylor Scott & White Research Institute and Charles A. Sammons Cancer Center, Baylor University Medical Center, Dallas, TX, USA
[b] Department of Gastroenterological Surgery, Graduate School of Medical Sciences, Kumamoto University, Kumamoto, Japan
[c] Department of Surgery, Kumamoto General Hospital, Kumamoto, Japan
[d] Department of Biomedical Sciences, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Hong Kong, China
[e] Department of Gastroenterological Surgery, Nagoya University Graduate School of Medicine, Nagoya, Japan
[f] The International Research Center for Medicine Sciences, Kumamoto University, Kumamoto, Japan
[g] Shenzhen Research Institute, City University of Hong Kong, Shenzhen, China

ARTICLE INFO

Article history:
Received 1 October 2018
Received in revised form 29 January 2019
Accepted 30 January 2019
Available online 13 February 2019

Keywords:
Gastric cancer
Gene signature
Lymph node metastasis
Prediction
Early stage

ABSTRACT

Background: Although identification of lymph node (LN) metastasis is a well-recognized strategy for improving outcomes in patients with gastric cancer (GC), currently there is lack of availability of adequate molecular biomarkers that can identify such metastasis. Herein we have developed a robust gene-expression signature for detecting LN metastasis in early stage GC by using a transcriptome-wide biomarker discovery and subsequent validation in multiple clinical cohorts.

Methods: A total of 532 patients with pathological T1 and T2 GC from 4 different cohorts were analyzed. Two independent datasets ($n = 96$, and $n = 188$) were used to establish a gene signature for the identification of LN metastasis in GC patients. The diagnostic performance of our gene-expression signature was subsequently assessed in two independent clinical cohorts using qRT-PCR assays ($n = 101$, and $n = 147$), and subsequently compared against conventional tumor markers and image-based diagnostics.

Findings: We established a 15-gene signature by analyzing multiple high throughput datasets, which robustly distinguished LN status in both training (AUC = 0.765, 95% CI 0.667–0.863) and validation cohorts (AUC = 0.742, 95% CI 0.630–0.852). Notably, the 15-gene signature was significantly superior compared to the conventional tumor markers, CEA ($P = .04$) and CA19–9 ($P = .005$), as well as computed tomography-based imaging ($P = .04$).

Interpretation: We have established and validated a 15-gene signature for detecting LN metastasis in GC patients, which offers a robust diagnostic tool for potentially improving treatment outcomes in gastric cancer patients.

Fund: NIH: CA72851, CA181572, CA14792, CA202797, CA187956; CPRIT: RP140784: Baylor Sammons Cancer Center polot grants (AG), VPRT: 9610337, CityU 21101115, 11102317, 11103718; JCYJ20170307091256048 (XW).

1. Introduction

Lymph node (LN) metastasis is one of the major factors which influences poor prognosis in gastric cancer (GC) patients [1]. Therefore, accurate identification of LN status prior to treatment, particularly in early stages (mucosal and submucosal), is considered critical for improving treatment strategies and survival outcomes in these patients. Currently, diagnosis for LN metastasis is primarily made through various imaging modalities such as computed tomography (CT) and positron emission tomography with CT (PET-CT). However, these imaging-based

* Correspondence to: Ajay Goel, Center for Gastrointestinal Cancer Research, Center for Translational Genomics and Oncology, Baylor Scott & White Research Institute and Charles A. Sammons Cancer Center, Baylor University Medical Center, 3410 Worth Street, Suite 610, Dallas, TX 75246, USA.
** Correspondence to: Xin Wang, Department of Biomedical Sciences, City University of Hong Kong, 1B-102, 1/F, Block 1, 31 To Yuen Street, Kowloon Tong, Hong Kong.
E-mail addresses: xin.wang@cityu.edu.hk (X. Wang), ajay.goel@BSWHealth.org (A. Goel).
[1] Contributed equally to this work.

https://doi.org/10.1016/j.ebiom.2019.01.057

Research in context

Evidence before this study

Lymph node (LN) metastasis is one of the major factors which influences poor prognosis in gastric cancer (GC) patients. Therefore, accurate identification of LN status prior to treatment, particularly in early stages (mucosal and submucosal), is considered critical for improving treatment strategies and survival outcomes in these patients. Although current imaging diagnostic methods are limited, emerging evidence indicate that molecular diagnostic methods for identification of patients with LN metastasis appear to be promising.

Added value of this study

We conducted a comprehensive bioinformatics analysis in two independent datasets (TCGA and ACRG [GSE62254]), to establish a 15-gene signature for the identification of LN metastasis in GC patients. The diagnostic performance of this gene-expression signature was subsequently validated in two independent clinical cohorts, and the gene signature was significantly superior compared to the conventional tumor markers, CEA and CA19–9, as well as computed tomography-based imaging.

Implications of all the available evidence

Our study showed that a novel 15-gene signature can be used for the identification of LN metastasis in patients with early stage GC. Although further validation in prospective clinical cohorts are necessary, these markers offer promising potential for the identification of LN metastasis in gastric cancer patients.

diagnostic approaches are often inadequate in clinical settings for LN identification, and are often accompanied with high false negative rates, which can be as high as up to 45% [2]. Therefore, availability of detection methodologies that are more accurate and robust for the identification of LN metastasis are much needed, as these may lead to significant improvement in prognosis of GC patients.

The majority of early stage GCs can be successfully treated by endoscopic resections, by utilizing procedures such as endoscopic submucosal dissection and mucosal resection [3]; especially in primary lesions that are confined to mucosal or submucosal layers within the gastric epithelial lining (T1). In current clinical practice, patients that are found positive for at least one of the pathological risk factors, such as tumor ulceration, undifferentiated tumor type and lymphovascular invasion, are often recommended to undergo gastrectomy with LN dissection. Unfortunately however, pathological evaluation of the post-surgical gastrectomy tissues, especially from early stage (T1) GC patients, often reveals the presence of LN metastasis in only ~20% of patients [4,5]; indicating that a large majority of patients experienced overtreatment and un-necessary gastrectomies due to the inadequacy of histopathological risk-assessment criteria currently used in the clinic. These data underscore the need to develop more sensitive and specific molecular biomarkers that can facilitate LN metastasis detection in early stage GC patients for reducing excessive treatment burden and improving the overall survival in these patients.

In recent decades a few gene expression-based classifiers have been developed for various types of cancers as diagnostic and predictive tools [6–8]. For example Oncotype DX, is a commercially available test that encompasses analysis of a panel of genes for predicting the risk of recurrence in breast and colorectal cancers [9,10]. Similar prognostic and predictive gene-expression based classifications have also been developed for GC patients [11–13]; however, they have limited clinical usefulness. Furthermore, to the best of our knowledge, thus far no transcriptomic biomarkers have been developed for the identification of LN metastasis in GC patients. Consequently, availability of molecular biomarkers that can more accurately (via-a-vis current histopathologic factors) identify patients with LN metastasis, will very likely reduce the burden of unnecessary overtreatment in GC patients in the near future.

In this study, we for the first time undertook a systematic and comprehensive transcriptome-wide biomarker discovery approach to develop a gene expression signature for the identification of LN metastasis status in GC patients. By analyzing multiple independent patient cohorts, we discovered and validated a novel, 15-gene signature that robustly identifies LN metastasis status in GC patients. In addition, we illustrate that our gene-expression based signature was superior to the conventional tumor markers (CEA and CA-19-9) as well as CT-based imaging; highlighting its potential clinical significance in detecting LN metastasis in GC patients, which may lead to more personalized treatment approaches in future.

2. Materials and methods

2.1. Clinical specimens and data sources

The two publicly available gene expression datasets, TCGA and ACRG (GSE62254), were used to identify candidate genes for the identification of LN metastasis in GC patients. TCGA data (level 3 RNA-Seq data) was downloaded from the Broad GDAC Firehose portal (http://gdac.broadinstitute.org, accessed on Mar 21, 2016). The RNA-seq data measured expression levels for 20,531 genes with scaled estimates in the gene-level RSEM files, which were converted to TPM (transcripts per million) by multiplying by 10^6 and then log2-transformed. The ACRG (GSE62254) dataset was downloaded from the GEO database directly in its processed form, using the Bioconductor package 'GEOquery' in R.

For clinical validation, we examined 248 tissue specimens from two independent patient cohorts with T1 and T2 early-stage GCs, which were referred to as clinical cohort-1 (or testing cohort) and clinical cohort-2 (validation cohort). These cohorts included 50 and 198 specimens without pretreatment from LN-positive (LNP) and LN-negative (LNN) patients, respectively. We established a gene-signature based on the candidate genes identified from the high-throughput dataset-based discovery phase, using clinical cohort-1. Subsequently these genes were validated in the clinical cohort-2 patients. The detailed patient demographics and clinicopathological characteristics are shown in Table 1 and Supplemental materials and methods.

Since GC diagnosis and treatment decision-making is primarily decided following endoscopic resections, we also included T2 lesions considering that these lesions can be underestimated during endoscopy. Lymphovascular invasion was diagnosed after pathological review of surgical tissues, and data for serum levels of carcinoembryonic antigen (CEA) and cancer antigen (CA) 19–9 were collected from each participating institution.

2.2. Ethics statement

Written informed consent was obtained from all patients, and the study was approved by the institutional review boards of all the participating institutions.

2.3. Study design and participants

Our study design included the following two major phases: a biomarker discovery and a clinical validation phase. Based on RNA-Seq data for T1 patient specimens in the TCGA dataset, we first prioritized 15 genes differentially expressed between 5 LNP and 13 LNN patients with GC. Using the same set of specimens for training, we built a multivariate logistic regression model using the 15 genes as covariates, and

Table 1
Demographic, clinical characteristics and tumor markers for clinical cohort 1 and 2[a].

Characteristics	Clinical cohort-1 (n = 101)		Clinical cohort-2 (n = 147)	
	LN Positive	LN Negative	LN Positive	LN Negative
	$n = 24$	$n = 77$	$n = 26$	$n = 121$
Preoperative factors				
Age (y.o)	65.3 ± 2.1	69.1 ± 1.2	67.7 ± 2.6	63.3 ± 1.2
Sex				
Male	18	63	14	67
Female	6	14	12	44
CEA (ng/ml)	3.39 ± 2.49[b]	3.69 ± 7.91[b]	2.13 ± 0.36[c]	2.16 ± 0.17[c]
positive	3	4	0	5
negative	17	51	6	7
CA19–9 (U/ml)[d]	93.9 ± 36.7	27.8 ± 22.0	12.2 ± 3.61	15.8 ± 1.76
positive	5	7	0	4
negative	14	46	4	8
Clinical N stage (CT)				
positive	NA	NA	6	5
negative	NA	NA	20	116
Postoperative factors				
Tumor size (mm)	41.4 ± 4.6	41.3 ± 2.6	42.7 ± 4.2	28.1 ± 1.9
Pathological T stage				
1	10	46	26	121
2	14	31	0	0
Final Stage				
I	9	70	21	121
II	12	6	4	0
III	1	0	1	0
IV	2	1	0	0
Lymphatic invasion				
positive	16	44	18	7
negative	8	33	8	144
Venous invasion				
positive	9	29	11	15
negative	15	48	15	106

[a] Plus-minus values are means ± SE. NA denotes not available.
[b] Cutoff value is 5 ng/ml.
[c] Cut off value is 3.4 ng/ml.
[d] Cut off value is 37 U/ml. LN, lymph node.

subsequently derived a LN risk scoring formula. To demonstrate the robustness of this panel as a diagnostic marker, and its applicability to patients with T2 stage GC, we first evaluated its performance in the training set of T1 patients. This was followed by in silico validation in an expanded TCGA dataset involving 96 T1/T2 patients (LNP 49, LNN 47), and another independent set of 188 T2 patients (LNP 157, LNN 31) from the ACRG cohort.

In the clinical validation phase, two large, independent patient cohorts were analyzed to validate the 15-gene signature identified during the discovery phase. Using qRT-PCR data derived from 101 T1/T2 patient (LNP 24, LNN 77) specimens in the clinical cohort-1 as the testing set, we conducted a multivariate logistic regression analysis for qRT-PCR, from which a LN risk scoring formula was derived. The diagnostic performance of the 15-gene signature subsequently evaluated using an independent qRT-PCR dataset from 147 (LNP 26, LNN 121) T1 specimens from the clinical validation cohort-2. To demonstrate the clinical significance of our data, we benchmarked our gene signature against the conventional tumor markers, CEA and CA19–9. Computed tomography (CT) was performed before surgery in all patients belonging to the clinical cohort-2, and the imaging results were evaluated by board certified radiologists. When the size on the short axis of the regional LN was >10 mm, clinical LN status was deemed to be positive.

2.4. RNA isolation and quantitative reverse-transcription PCR

Total RNA extraction from tissue specimens was performed using miRNeasy RNA isolation kits (Qiagen, Hilden, Germany). Synthesis of complementary DNA (cDNA) was conducted on 1 μg of total RNA using the High Capacity cDNA Reverse Transcription Kit (Invitrogen, Carlsbad, CA, USA). Quantitative real-time reverse transcription analysis (qRT-PCR) was performed using the SensiFAST™ SYBR® Lo-ROX Kit (Bioline, London, UK) on the Quantstudio 7 Real Time PCR System (Life Technologies, Carlsbad, CA, USA). The average expression levels of target genes were normalized against beta-actin using the comparative CT method [14]. To ensure consistent measurements throughout all assays, for each PCR amplification reaction, three independent cDNA samples were loaded as internal controls to account for any plate-to-plate variation, and the results from each plate were normalized against internal normalization controls.

2.5. Statistical analysis

Wilcoxon's signed-rank tests, Mann-Whitney *U* tests and Kruskal-Wallis tests were used to analyze gene expression data, as appropriate. The Benjamini-Hochberg method was used to correct for multiple hypotheses testing, wherever applicable. Risk scores derived from the 15-gene multivariate logistic regression model were used to plot receiver-operating-characteristic (ROC) curves and calculate area under the curves (AUCs). Confidence intervals for the ROC curves were calculated using the method of *DeLong* [15] as well as the statistical significance of comparison two ROC curves. Univariate and multivariate logistic regression models were employed to evaluate the statistical significance of clinicopathological variables and the 15-gene model in diagnosing LN metastasis status. All statistical analyses were performed using Medcalc V.12.3.0 (Broekstraat 52, 9030; Mariakerke, Belgium), the GraphPad Prism V5.0 (GraphPad Software, San Diego, California, USA) and R (3.3.3, R Development Core Team, https://cran.r-project.org/).

3. Results

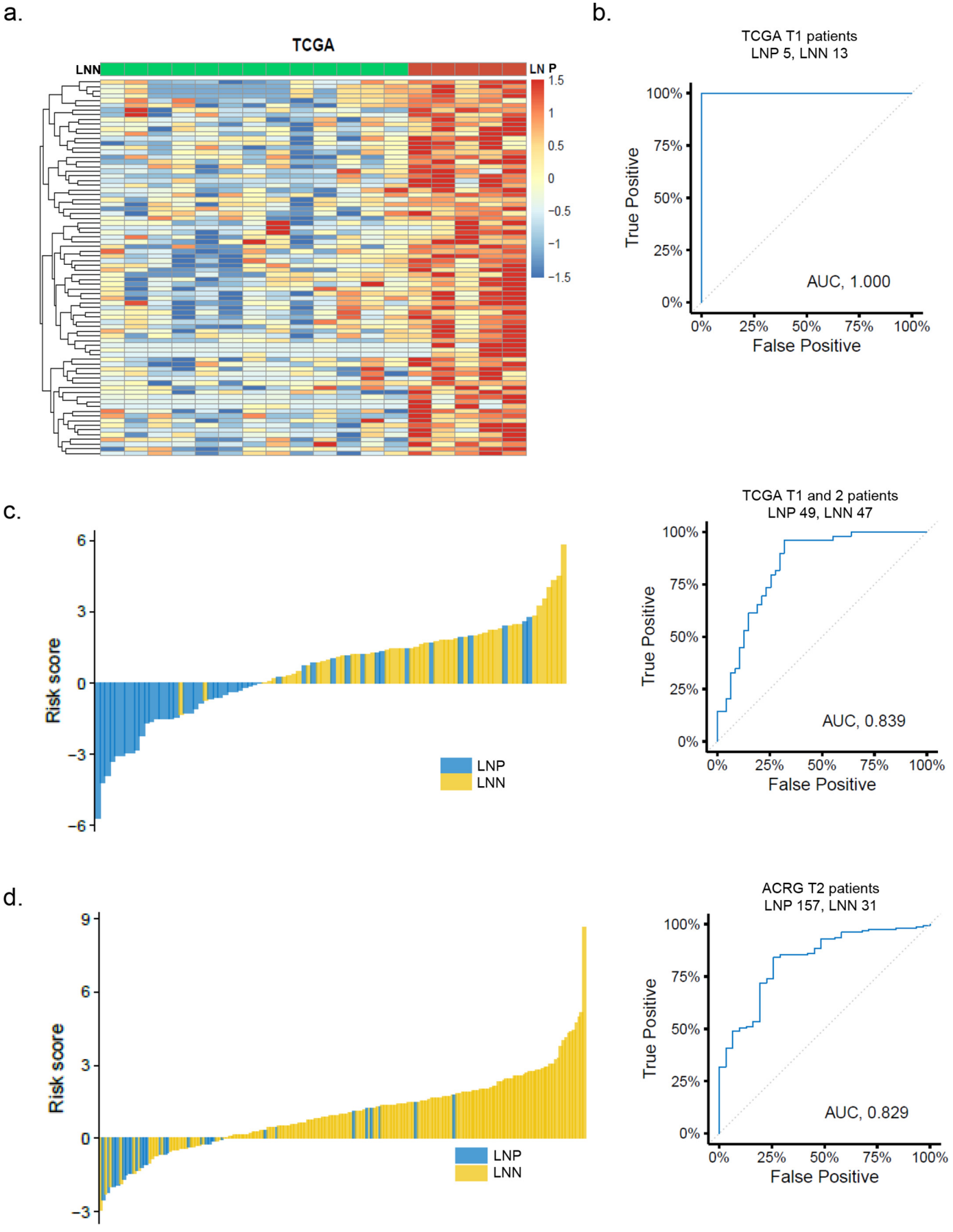
3.1. Genome-wide discovery of a novel gene expression signature to detect lymph node metastasis in early stage gastric cancer

To identify a panel of genes that can help diagnose patients with lymph node metastasis, we first analyzed RNA-seq expression profiling data from 18 patients with early stage T1 cancers, which were either LN metastasis positive or negative. Among a total of 20,531 genes, 84 genes were differentially expressed between 5 lymph node positive (LNP) and 13 negative (LNN) patients ($P < .01$ [Wilcoxon signed-rank test], log2 fold change >1.5; Fig. 1a and S2). To identify a robust candidate gene signature, we further narrowed down the gene list to 15 by filtering out lowly expressed genes (average expression level < 3 log2-transformed TPM). Using multivariate logistic regression analysis, we found the 15 candidate genes were able to successfully distinguish LNP from LNN GC patients in the training set (AUC = 1.000, 95% CI 1.000–1.000; Fig. 1b).

In view of the availability of multiple public datasets consisting of T2 GC patients, we next investigated whether our T1 lymph node metastasis GC gene signature could also identify LN status in these additional patient cohorts. Intriguingly, our genes were able to distinguish LNP from LNN patients in an expanded set of 96 T1 and T2 patients in the TCGA cohort (AUC = 0.839, 95% CI 0.757–0.921; Fig. 1c), as well as in the ACRG cohort of 188 T2 patients (AUC = 0.829, 95% CI 0.752–0.906; Fig. 1d). These data highlight the diagnostic potential of our novel 15-genes in identifying LN metastasis in early-stage gastric cancer patients.

3.2. Validation and establishment of the 15-gene signature for detecting Lymph node status in gastric cancer patients

Next, we assessed the diagnostic accuracy of the 15 gene-panel by qRT-PCR in 24 LNP and 77 LNN tissue specimens in the clinical cohort-1 ($n = 101$). While individual genes had limited predictive power (AUCs varying between 0.506 and 0.605), the combination of


a.
TCGA
LNN
LN P
1.5
1
0.5
0
−0.5
−1
−1.5
b.
TCGA T1 patients
LNP 5, LNN 13
True Positive
False Positive
AUC, 1.000
c.
Risk score
LNP
LNN
TCGA T1 and 2 patients
LNP 49, LNN 47
AUC, 0.839
d.
Risk score
LNP
LNN
ACRG T2 patients
LNP 157, LNN 31
AUC, 0.829

various genes demonstrated significant detecting power in identifying LN metastasis in GC patients. Using a multivariate logistic regression analysis, we obtained a risk scoring formula for the 15-gene signature as follows, risk score = 1.300 - (0.233 × C5AR1) - (0.274 × CD83) + (0.048 × ETV4) - (0.213 × FAM13A) + (0.207 × FKBP10) - (0.227 × GPRIN3) + (0.052 × LCK) + (0.124 × NR4A2) + (0.402 × PRSS21) + (0.035 × RGPD1) - (0.196 × SLC2A3) - (0.212 × SLFN2) + (0.174 × TMEM86B) - (0.118 × TRPV4) - (0.075 × YBX2). Although the performance of each individual genes was not significant, risk scores for LN metastasis determined using this formula for patients in this testing cohort demonstrated a very encouraging and robust diagnostic performance (AUC = 0.765, 95% confidence interval [CI], 0.667–0.862; OR = 23.61, 95% CI, 3.034–183.6; Fig. 2a and Table 2).

Of interest, our 15-gene signature was significantly superior compared to the conventional tumor markers, CEA ($P = .033$ [DeLong]) and CA19–9 ($P = .044$ [Delong]; Fig. 2b) in identifying LNP patients. To further validate the diagnostic efficiency of this 15-gene signature, we next examined its performance in an independent validation cohort comprising of 26 LNP and 121 LNN T1 GC patients using the multivariate logistic regression analysis. In line with results from our testing cohort, the 15-gene signature was once again able to robustly distinguish LNP from LNN early stage GC patients (AUC = 0.742, 95% CI, 0.631–0.852; OR = 6.563, 95% CI, 2.585–16.66; Fig. 2b and Table 2).

3.3. The 15-gene signature outperformed conventional diagnostic approaches for LN metastasis detection in gastric cancer patients

Using multivariate analysis, we demonstrated that our 15-gene signature was able to successfully detect LN metastasis, independent of pre-operative clinical factors such as age, gender, tumor markers and clinical LN status determined by computed tomography (Table 3). To further evaluate the significance of this signature, we next compared the diagnostic potential of our gene signature versus various pre-operative clinical factors including tumor markers and clinical LN status. We found the diagnostic value of our 15-gene signature was significantly superior via-a-vis conventional tumor markers including the levels of circulating CEA (AUC = 0.520, 95% CI, 0.429–0.610; $P = .044$ [DeLong]) and CA19–9 (AUC = 0.518, 95% CI, 0.427–0.608; $P = .0047$ [DeLong]; Fig. 2b and Table 2).

In addition, to evaluate the performance of the 15-gene signature, we compared its performance against the clinical N stage determined by preoperative diagnostic CT scans in the patients from the validation cohort. Interestingly, our gene expression signature demonstrated a significantly superior accuracy (0.803, 95% CI, 0.510–0.905) compared to CT imaging data for identifying the presence of LN metastasis (AUC = 0.742; $P = .038$ [DeLong]; Fig. 2c and Table 2), which only achieved an AUC of 0.595 (95% CI, 0.511–0.675).

3.4. A combination of the 15-gene signature together with other clinicopathological features further improves the diagnostic accuracy for LN metastasis detection in gastric cancer patients

We next asked whether a combination of our 15-gene signature together with currently used clinicopathological factors (e.g. age, gender, tumor markers, and clinical N stage using multivariate logistic regression analysis) might further enhance the diagnostic accuracy of our panel. It was interesting to observe that indeed integration of our gene expression signature with the clinical N stage, significantly improved the discriminative accuracy of our biomarker panel further in identifying LNP gastric cancer cases (AUC = 0.789, 95% CI, 0.706–0.857) compared to use of the 15-gene signature alone (Fig. 2d and Table 2).

4. Discussion

As we usher into the new era of precision-medicine, tailoring individualized treatments are definitely going to serve as cornerstones for a more effective cancer care. Currently early stage GC patients, which are deemed to be high-risk for lymph node (LN) metastasis based upon various pre-surgical histopathological features are frequently over-treated, due to the lack of availability of adequate molecular markers that can more robustly identify such metastasis prior to the surgery. In this study, we undertook a systematic and comprehensive, genome-wide transcriptomic biomarker discovery, and developed a panel of genes for the identification of LN metastasis in patients with early stage gastric cancers (T1 and T2), using independent, publicly-available gene expression datasets. Subsequently, a 15-gene signature was optimized for qRT-PCR based analysis using the clinical testing cohort-1 by logistic regression analysis, followed by validation in an independent patient cohort. Finally, we performed a head-to-head comparison between the 15-gene signature with the conventional tumor markers (CEA and CA19–9) as well as CT-based imaging, and demonstrated its superiority for identifying GC patients with LN metastasis.

Recent advancements in high-throughput sequencing technologies have resulted in comprehensive molecular characterization of GC [16]. Similar to other major malignancies, multiple molecular subtypes of GC have been proposed based on integrative analysis of transcriptome wide gene expression profiles [17]. Accordingly, several gene expression-based cancer biomarkers utilizing multiple genes have been suggested over the years [7,9]. Because RNA-sequencing provides molecular insights into tumor heterogeneity and the disease process, in this study we focused on establishing a gene expression based-signature for the diagnosis of LN metastasis in early stage GC patients using a transcriptomic-wide analysis of T1 tumors from GC patients. We identified a cluster of 15 highly expressed genes, several of which are functionally relevant and GC-associated genes including *C5AR1*, *CD83*, *NR4A2*, *ETV4,* and *TRPV4*. In gastric cancer, C5AR1 has been shown to promote motility and invasiveness of cancer by activating RhoA, and its expression is reported to be associated with prognosis of GC patients [18]. CD83 is a molecular marker for mature dendritic cells. In gastric cancer, decreased density of CD83 (+) dendritic cells and increased density of FOXP3 (+) regulatory T cells, are observed in the primary tumor and metastatic lymph nodes of GC, and has been shown to inversely correlate with prognosis of GC patients [19]. An in vitro study has shown that the expression of orphan nuclear receptor NR4A2 in GC cells attenuates 5-fluorouracil-induced apoptosis and affect chemoresistance, and predicts an unfavorable postoperative survival of GC patients with chemotherapy [20]. Another putative oncogene, PEA3/ETV4 has been shown to be upregulated at both mRNA and protein levels in GC tissue and the increased expression correlates with the expression of their downstream metastasis associated target gene, MMP-1 and high expression of PEA3/ETV4 was associated with poor prognosis in GC [21]. Similarly, TRPV4 was shown to be a gene required for cancer cell invasion and trans-endothelial migration and its expression in GC correlated with poor clinical outcomes [22]. Furthermore, as a gene signature, three genes, NR4A2, FAM13A and PRSS21 had a significant contribution to our gene signature, suggesting that these genes play an important mechanistic role in GC LN metastasis.

Fig. 1. Genome-wide discovery of a gene expression signature for the identification of lymph node metastasis status in early gastric cancers (GC). (A) Heatmap illustrating the expression levels of the genes expressed between patients with lymph node-positive (LNP) and lymph node-negative (LNN) gastric cancers. Of these, 84 genes were differentially expressed between LNP and LNN patients in the training dataset of 18 TCGA T1 patients. (B) ROC curve shows the diagnostic performance of the 15-gene signature for discriminating LNP from LNN TCGA T1 patients. (C) Waterfall plot shows the LN risk scores by LN status in the TCGA T1 and T2 cohort, and the ROC curve demonstrates the diagnostic performance in the expanded set of TCGA T1 and T2 patients. (D) Waterfall plot illustrates the risk scores by LN status and the its diagnostic potential in an independent set of ACRG T2 patients.

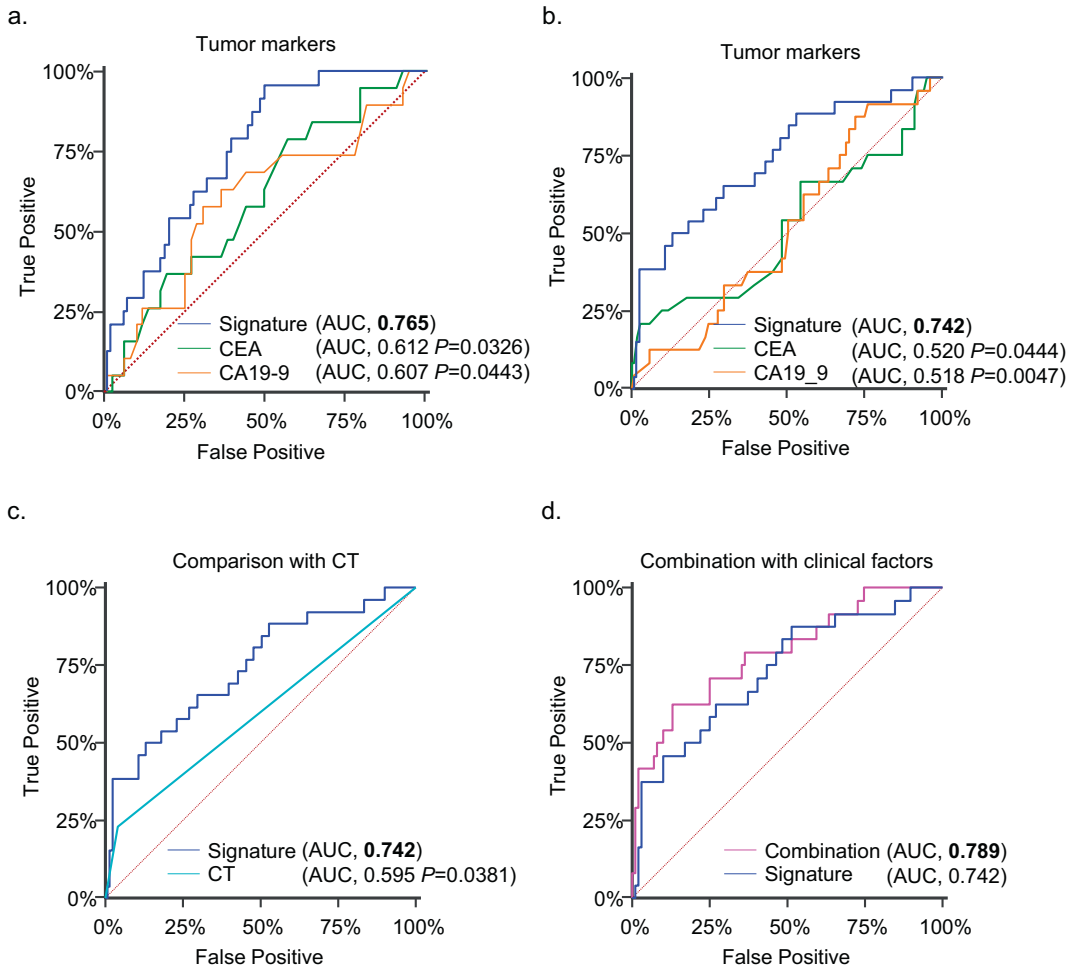


Fig. 2. Clinical validations of the 15-gene signature in identification of lymph node metastasis status in early GC. ROC curves show that our novel 15-gene signature had a higher diagnostic value for identification of LN metastasis over CEA and CA19_9 in (A) clinical cohort-1 (testing) and (B) clinical cohort-2 (validation), respectively. (C) ROC curves illustrate that the 15-gene signature had a higher performance compared to the clinical LN status determined by CT in clinical cohort-2. Comparison of AUC values were conducted by DeLong test (D) ROC curves illustrate that the combinatorial model integrating the 15-gene signature and clinical N stage further improved the predictive accuracy in clinical cohort 2.

In the past, a few studies have attempted to identify gene-expression-based biomarkers that may facilitate identification of LN metastasis in GC patients using cDNA microarrays [23–25]. However, to the best of our knowledge, ours is the first study to perform a systematic and comprehensive biomarker discovery from multiple RNA-Seq based datasets. Second, we performed validation of our discovered biomarkers in multiple, independent, datasets from publicly available resources, followed by confirmation of our results in in-house, clinical patient cohorts. Third, we focused our biomarker discovery and validation effort specifically in early-stage cancers, because excessive surgical treatment in these individuals have long-term consequences with adverse quality of life. Fourth, we compared the performance of our biomarkers with various tumor markers and CT imaging results, and successfully demonstrated the superiority of our signature over these with currently used modalities in the clinical settings.

One of the potential limitations of our study is that retrospective clinical cohorts were used for the development of the gene panel. In addition, one of the limitations of the present study is that we used frozen tissue and FFPE-derived RNA from resected tissues. Considering that this gene-signature will be examined in pre-surgical biopsy specimens in a clinical setting, further prospective trials are required to examine the robustness and performance of our 15-gene signature in fresh biopsy tissues. Furthermore, another limitation was that the sample size for biomarker discovery was limited. Since one of the primary objectives of our study was to identify biomarkers for early-stage gastric cancers, we focused on patients with T1 cases with LN metastasis, which further reduced the total number of patients during the discovery phase. Consequently, we were limited to deriving our gene-signature with the limited sample size, and could not fully utilize appropriate power calculations for biomarker discovery. Therefore, we would like to acknowledge this potential limitation of our study, that our effort would have been more comprehensive, if we had an access to larger cohorts of patients with T1 LN metastasis, which will likely require a multi-institutional effort given the rarity of this disease. Nevertheless, the reassuring aspect of our study is that regardless of this concern, 15-gene signature was successfully able to identify LN in GC patients, and was superior to both currently used tumor markers (CEA and CA-19-9) as well as CT imaging. Although the further clinical validation is required using a large prospective cohort, our gene signature was able to discriminate LNP patients from LNN patients using surgically resected sample.

In conclusion, we have developed a novel 15-gene signature, which can potentially be used for the identification of LN metastasis in patients with early stage GC. Pending further validation in prospective clinical cohorts, these markers offer promising potential for the identification of LN metastasis in gastric cancer patients.

Funding source

The present work was supported by the grants CA72851, CA181572, CA184792, CA202797 and CA187956 from the National Cancer Institute; a grant (RP140784) from the Cancer Prevention Research Institute of Texas (CPRIT), pilot grants from the Baylor Sammons Cancer Center, as

Table 2
Statistical evaluation of the performance of individual signature genes in the clinical cohorts[a]

Gene	Clinical cohort 1				Clinical cohort 2			
	Odds ratio (95% CI)	AUC (95% CI)	Sensitivity	Specificity	Odds ratio (95% CI)	AUC (95% CI)	Sensitivity	Specificity
C5AR1	0.848 (0.661–1.088)	0.526 (0.424–0.626)	0.583	0.533	0.899 (0.784–1.031)	0.541 (0.457–0.624)	0.269	0.934
CD83	1.081 (0.814–1.435)	0.522 (0.420–0.623)	0.458	0.701	0.827 (0.583–1.173)	0.545 (0.460–0.627)	0.346	0.835
ETV4	0.994 (0.792–1.246)	0.515 (0.413–0.616)	0.417	0.727	1.057 (0.918–1.216)	0.602 (0.518–0.681)	0.615	0.653
FAM13A	1.167 (0.850–1.601)	0.552 (0.450–0.652)	0.958	0.208	1.079 (0.978–1.190)	0.610 (0.526–0.689)	0.731	0.488
FKBP10	0.992 (0.793–1.241)	0.506 (0.405–0.607)	0.833	0.312	1.070 (0.860–1.330)	0.531 (0.447–0.614)	0.769	0.405
GPRIN3	0.942 (0.727–1.219)	0.532 (0.430–0.632)	0.875	0.247	1.096 (0.767–1.566)	0.526 (0.442–0.609)	0.423	0.835
LCK	0.911 (0.722–1.150)	0.590 (0.488–0.687)	0.458	0.779	1.007 (0.901–1.126)	0.539 (0.455–0.622)	0.808	0.388
NR4A2	0.923 (0.683–1.248)	0.546 (0.444–0.645)	0.417	0.792	0.878 (0.782–0.989)	0.596 (0.512–0.676)	0.308	0.934
PRSS21	0.756 (0.627–0.985)	0.605 (0.503–0.701)	0.625	0.623	0.983 (0.910–1.063)	0.509 (0.426–0.593)	0.346	0.785
RGPD1	0.946 (0.711–1.259)	0.526 (0.424–0.626)	0.208	0.974	0.899 (0.767–1.055)	0.626 (0.543–0.705)	0.577	0.653
SLC2A3	1.116 (0.932–1.337)	0.597 (0.495–0.693)	0.542	0.714	0.973 (0.792–1.195)	0.553 (0.469–0.635)	0.846	0.380
SLFN2	1.040 (0.775–1.395)	0.506 (0.404–0.607)	0.833	0.260	0.975 (0.886–1.073)	0.504 (0.420–0.587)	0.846	0.058
TMEM86B	0.839 (0.689–1.021)	0.538 (0.436–0.638)	0.333	0.896	0.915 (0.734–1.141)	0.530 (0.446–0.613)	0.962	0.182
TRPV4	1.041 (0.814–01.331)	0.529 (0.427–0.629)	0.375	0.831	1.062 (0.985–1.146)	0.603 (0.520–0.683)	0.846	0.388
YBX2	1.048 (0.911–1.207)	0.547 (0.444–0.646)	0.500	0.675	1.005 (0.929–1.087)	0.545 (0.461–0.627)	0.846	0.289
Risk score	147.7 (8.676–2514)	0.765 (0.670–0.844)	0.958	0.506	602.2 (29.48–12,298)	0.742 (0.663–0.810)	0.868	0.500

[a] AUC, area under the ROC curve.

Table 3
Stratified analysis using preoperative clinical factors and the 15-gene signature in clinical cohort-2.

MicroRNA	Univariate analysis		Multivariate analysis	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Age	1.029 (0.992–1.067)	N.S.	1.046 (0.993–1.102)	N.S.
Sex	0.940 (0.402–2.201)	N.S.	0.625 (0.210–1.859)	N.S.
CEA	0.987 (0.762–1.279)	N.S.	0.950 (0.677–1.335)	N.S.
CA19–9	0.986 (0.954–1.018)	N.S.	0.985 (0.953–1.018)	N.S.
Clinical LN status	6.960 (1.939–24.99)	0.003	8.125 (1.521–43.40)	0.014
Risk score	23.61 (3.034–183.6)	<0.0001	6.563 (2.585–16.66)	0.0001

Abbreviation: OR, odds ratio.

well as funds from the Baylor Scott & White Research Institute awarded to Ajay Goel, and a VPRT grant (9610337) from the City University of Hong Kong, grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 21101115, 11102317, 11103718), a grant from The Science Technology and Innovation Committee of Shenzhen Municipality (JCYJ20170307091256048) awarded to Xin Wang, and was partially supported by the Shenzhen Research Institute, City University of Hong Kong.

Declaration of interest

None of the authors has any potential conflicts to disclose.

Author contribution

Study concept and design: DI, FG, XW, AG, ST; Specimen providers: DI, TI, HB, FS, MK, YK; Acquisition of clinical data: DI, TI, HB, FS, MK, YK; Analysis and interpretation of data and statistical analysis: DI, FG, XW, AG; Drafting of the manuscript: DI, ST, FG, FS, XW and AG.

Acknowledgements

We thank Preethi Ravindranathan for a critical reading of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2019.01.057.

References

[1] Bonenkamp JJ, Songun I, Hermans J, Sasako M, Welvaart K, Plukker JT, et al. Randomised comparison of morbidity after D1 and D2 dissection for gastric cancer in 996 Dutch patients. Lancet 1995;345(8952):745–8.
[2] Saito T, Kurokawa Y, Takiguchi S, Miyazaki Y, Takahashi T, Yamasaki M, et al. Accuracy of multidetector-row CT in diagnosing lymph node metastasis in patients with gastric cancer. Eur Radiol 2015;25(2):368–74.
[3] Japanese Gastric Cancer A. Japanese gastric cancer treatment guidelines 2014 (ver. 4). Gastric Cancer 2017;20(1):1–19.
[4] Gotoda T, Yanagisawa A, Sasako M, Ono H, Nakanishi Y, Shimoda T, et al. Incidence of lymph node metastasis from early gastric cancer: estimation with a large number of cases at two large centers. Gastric Cancer 2000;3(4):219–25.
[5] Hirasawa T, Gotoda T, Miyata S, Kato Y, Shimoda T, Taniguchi H, et al. Incidence of lymph node metastasis and the feasibility of endoscopic resection for undifferentiated-type early gastric cancer. Gastric Cancer 2009;12(3):148–52.
[6] Sun LL, Wu JY, Wu ZY, Shen JH, Xu XE, Chen B, et al. A three-gene signature and clinical outcome in esophageal squamous cell carcinoma. Int J Cancer 2015;136(6):E569–77.
[7] O'Connell MJ, Lavery I, Yothers G, Paik S, Clark-Langone KM, Lopatin M, et al. Relationship between tumor gene expression and recurrence in four independent studies of patients with stage II/III colon cancer treated with surgery alone or surgery plus adjuvant fluorouracil plus leucovorin. J Clin Oncol 2010;28(25):3937–44.

[8] Hoshida Y, Villanueva A, Kobayashi M, Peix J, Chiang DY, Camargo A, et al. Gene expression in fixed tissues and outcome in hepatocellular carcinoma. N Engl J Med 2008;359(19):1995–2004.
[9] Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004; 351(27):2817–26.
[10] Gray RG, Quirke P, Handley K, Lopatin M, Magill L, Baehner FL, et al. Validation study of a quantitative multigene reverse transcriptase-polymerase chain reaction assay for assessment of recurrence risk in patients with stage II colon cancer. J Clin Oncol 2011;29(35):4611–9.
[11] Chen CN, Lin JJ, Chen JJ, Lee PH, Yang CY, Kuo ML, et al. Gene expression profile predicts patient survival of gastric cancer after surgical resection. J Clin Oncol 2005;23 (29):7286–95.
[12] Cho JY, Lim JY, Cheong JH, Park YY, Yoon SL, Kim SM, et al. Gene expression signature-based prognostic risk score in gastric cancer. Clin Cancer Res 2011;17 (7):1850–7.
[13] Busuttil RA, George J, Tothill RW, Ioculano K, Kowalczyk A, Mitchell C, et al. A signature predicting poor prognosis in gastric and ovarian cancer represents a coordinated macrophage and stromal response. Clin Cancer Res 2014;20(10):2761–72.
[14] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Delta Delta C(T)) Method. Methods 2001;25(4): 402–8.
[15] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44(3):837–45.
[16] The Cancer Genome Atlas Research N, Analysis Working Group: Dana-Farber Cancer I, Institute for Systems B, University of Southern C, Memorial Sloan Kettering Cancer C, Agency BCC, et al. Comprehensive molecular characterization of gastric adenocarcinoma. Nature 2014;513:202–9.
[17] Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med 2015;21(5):449–56.
[18] Kaida T, Nitta H, Kitano Y, Yamamura K, Arima K, Izumi D, et al. C5a receptor (CD88) promotes motility and invasiveness of gastric cancer by activating RhoA. Oncotarget 2016;7(51):84798–809.
[19] Kashimura S, Saze Z, Terashima M, Soeta N, Ohtani S, Osuka F, et al. CD83(+) dendritic cells and Foxp3(+) regulatory T cells in primary lesions and regional lymph nodes are inversely correlated with prognosis of gastric cancer. Gastric Cancer 2012;15(2):144–53.
[20] Han Y, Cai H, Ma L, Ding Y, Tan X, Chang W, et al. Expression of orphan nuclear receptor NR4A2 in gastric cancer cells confers chemoresistance and predicts an unfavorable postoperative survival of gastric cancer patients with chemotherapy. Cancer 2013;119(19):3436–45.
[21] Keld R, Guo B, Downey P, Cummins R, Gulmann C, Ang YS, et al. PEA3/ETV4-related transcription factors coupled with active ERK signalling are associated with poor prognosis in gastric adenocarcinoma. Br J Cancer 2011;105(1):124–30.
[22] Lee WH, Choong LY, Mon NN, Lu S, Lin Q, Pang B, et al. TRPV4 regulates breast cancer cell extravasation, stiffness and actin cortex. Sci Rep 2016;6:27903.
[23] Weiss MM, Kuipers EJ, Postma C, Snijders AM, Siccama I, Pinkel D, et al. Genomic profiling of gastric cancer predicts lymph node status and survival. Oncogene 2003;22(12):1872–9.
[24] Teramoto K, Tada M, Tamoto E, Abe M, Kawakami A, Komuro K, 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):225–36.
[25] Marchet A, Mocellin S, Belluco C, Ambrosi A, DeMarchi F, Mammano E, et al. Gene expression profile of primary gastric cancer: towards the prediction of lymph node status. Ann Surg Oncol 2007;14(3):1058–64.