


SHORT COMMUNICATION

Prognostic value of gastric cancer-associated gene signatures: Evidence based on a meta-analysis using integrated bioinformatics methods

Jun Wang | Peng Gao | Yongxi Song | Jingxu Sun | Xiaowan Chen | Hong Yu |
Yu Wang | Zhenning Wang 

Department of Surgical Oncology and General Surgery, The First Hospital of China Medical University, Heping District, Shenyang, China

Correspondence: Zhenning Wang, Department of Surgical Oncology and General Surgery, The First Hospital of China Medical University, 155 North Nanjing Street, Heping District, Shenyang 110001, China (josieon826@sina.cn).

Funding information

National Key R&D Program of China, Grant/Award Number: MOST-2016YFC1303200, MOST-2016YFC1303202, MOST-2017YFC0908300, MOST-2017YFC0908305

Abstract

Selecting differentially expressed genes (DEGs) based on integrated bioinformatics analyses has been used in previous studies to explore potential biomarkers in gastric cancer (GC) with microarray and RNA sequencing data. However, the genes obtained may be inaccurate because of noisy data and errors, as well as insufficient clinical sample sizes. Thus, we aimed to find robust and strong DEGs with prognostic value for GC, where the robust rank aggregation method was employed to select significant DEGs from eight Gene Expression Omnibus data sets with a total of 140 up-regulated and 206 down-regulated genes. Network data mining was then used to screen hub genes, and 11 genes were filtered using Fisher's exact test. Based on these results, we built a prognostic signature with seven genes (*FBN1*, *MMP1*, *PLAU*, *SPARC*, *COL1A2*, *COL2A1* and *ATP4A*) using stepwise multivariate Cox proportional hazard regression. According to the risk score for each patient, we found that high-risk group patients had significantly worse survival results compared with those in the low-risk group (log-rank test P -value < 0.001). This seven-gene signature was then validated with an external data set. Thus, we established a signature based on seven DEGs with prognostic value for GC patients using multi-steps bioinformatics methods, which may provide novel insights and potential biomarkers for prognosis, as well as possibly serving as new therapeutic targets in clinical applications.

KEYWORDS

gastric cancer, meta-analysis, network mining, prognosis, robust rank aggregation

1 | INTRODUCTION

A previous study estimated that 951 600 new GC cases and 723 100 deaths occurred in 2012, especially in Eastern Asia.¹ Although a dramatic worldwide decline in the incidence and

mortality rates of GC has occurred, GC still has a poor 5-year survival rate.² Therefore, molecular biomarkers have attracted much attention because of diagnosing and evaluating the prognosis in GC.

Microarray and RNA sequencing technologies, as well as gene profiling data sets such as The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO), have been used to identify various DEGs and significant biological pathways in different

Jun Wang and Peng Gao contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2018 The Authors. Journal of Cellular and Molecular Medicine published by John Wiley & Sons Ltd and Foundation for Cellular and Molecular Medicine.

cancers. Several recent studies of DEGs associated with GC³⁻⁵ have employed integrated bioinformatics analyses to explore the patterns of gene expression. However, biased gene expression results may be obtained using a single data set because of data outliers, noise, and errors, as well as insufficient sample sizes. The robust rank aggregation (RRA) method has been employed for selecting differentially expressed microRNA (miRNA) profiles based on multiple data sets in various cancers, which is robust to these noises. However, no previous study of GC has identified DEGs using the RRA method, especially to detect prognostic gene signatures, which motivated this study.

In this study, we performed multi-step analysis to examine prognostic gene signatures in order to determine whether the RRA method can be used for selecting DEGs from a variety of GEO data sets and for identifying prognostic biomarkers in GC.

2 | MATERIALS AND METHODS

2.1 | Gene expression omnibus data set selection and data generation

Eight independent GC gene expression microarray data sets were downloaded from the GEO database. A summary of the detailed series information is shown in Table S1. The filter of DEGs according to the criteria of: $|\log_2 \text{fold-change}| > 1$ and $\text{adj.}P\text{-value} < 0.05$.

2.2 | Robust rank aggregation method for meta-analysis

In order to avoid inconsistent results among different studies and to identify robust DEGs based on the GC data sets, the RRA method⁶ was applied to the lists of genes, which employs a probabilistic model for aggregation. The RRA method is robust to noise, and it facilitates the calculation of significance probabilities to all of the elements in the final ranking. The *P*-values were subjected to Bonferroni's correction to avoid false-positive results.

2.3 | Network data mining from DEGs

To determine more accurate and robust DEGs with prognostic value in GC based on above genes we obtained. The PPI networks were built by the Human Integrated Protein-Protein Interaction rEference (HIPPIE), and Fisher's exact test was used to select hub genes according to a *P*-value cut-off < 0.01 and mapping number > 5 .

2.4 | Prognostic gene signature risk scoring system based on DEGs

Stepwise multivariate Cox proportional hazard regression was performed to obtain the regression coefficient for each gene. The area under the time-dependent receiver operating characteristic (ROC) curve (AUC) was determined to predict the 5-year survival, and high- and low-risk groups were according to the median-risk score. The

Kaplan-Meier curve was plotted to compare the survival outcomes in different groups. Gene Expression Profiling Interactive Analysis (GEPIA)⁷ is a web server that used to analyse the gene expression patterns in different TNM stages of GC.

3 | RESULTS

In this study, we employed a multi-step strategy to obtain a signature for DEGs with prognostic value in GC patients (Figure 1). We first downloaded eight GEO data sets of GC with 493 tumour and 213 normal samples. Significant DEGs were then filtered out from each GEO data set (Table S2). The RRA method⁶ was next applied to screen out precise and robust DEGs with 140 significantly up-regulated and 206 down-regulated (Table S3). GO processes and pathways enrichment results are shown in Figure S1.

Based on above method, we have obtained plentiful DEGs of GC. However, these genes may not all play significant biological roles. To find robust and strong DEGs from above genes, we combined various network data mining methods. The PPI networks were built (Figure S2A) based on a confidence score > 0.6 . Fisher's exact test (Table S5) was used to select hub genes according to a *P*-value < 0.01 and mapping number > 5 . Therefore, 11 hub genes were filtered (Table S6).

However, to find key DEGs with prognostic role in GC, we next built a prognostic signature with seven genes (*FBN1*, *MMP1*, *PLAU*, *SPARC*, *COL1A2*, *COL2A1* and *ATP4A*) using stepwise multivariate Cox proportional hazard regression. The AUC was 0.816 for predicting the 5-year survival (Figure 2A). According to the median risk score of each patient, we found that the patients from the high-risk group had significantly poorer overall survival results compared with those in the low-risk group (log-rank test *P*-value < 0.001) (Figure 2B). The seven-gene signature risk score distributions, patient survival results and expression heatmap are shown in Figure 2C. *SPARC*, *COL1A2* and *FBN1* were differentially expressed in various TNM stages of GC (Figure 2D-F).

Finally, this seven-gene signature was then validated with an independent data set (GSE62254, *n* = 300 samples) (Figure S4A,B). The Kaplan-Meier curves indicated that there were significant differences between the high- and low-risk groups (log-rank test *P*-value < 0.001). Therefore, this seven-gene signature may be useful for prediction of the survival and prognosis of GC.

4 | DISCUSSION

Molecular biomarkers such as genes have attracted much attention because they can be useful for the pathogenesis of GC. Thus, several recent studies have investigated the DEGs in GC. One study has reported that using 26 paired GC samples and microarray analysis, 2371 differential mRNAs were detected.⁵ In addition, several studies have investigated prognostic gene signatures in GC. A group

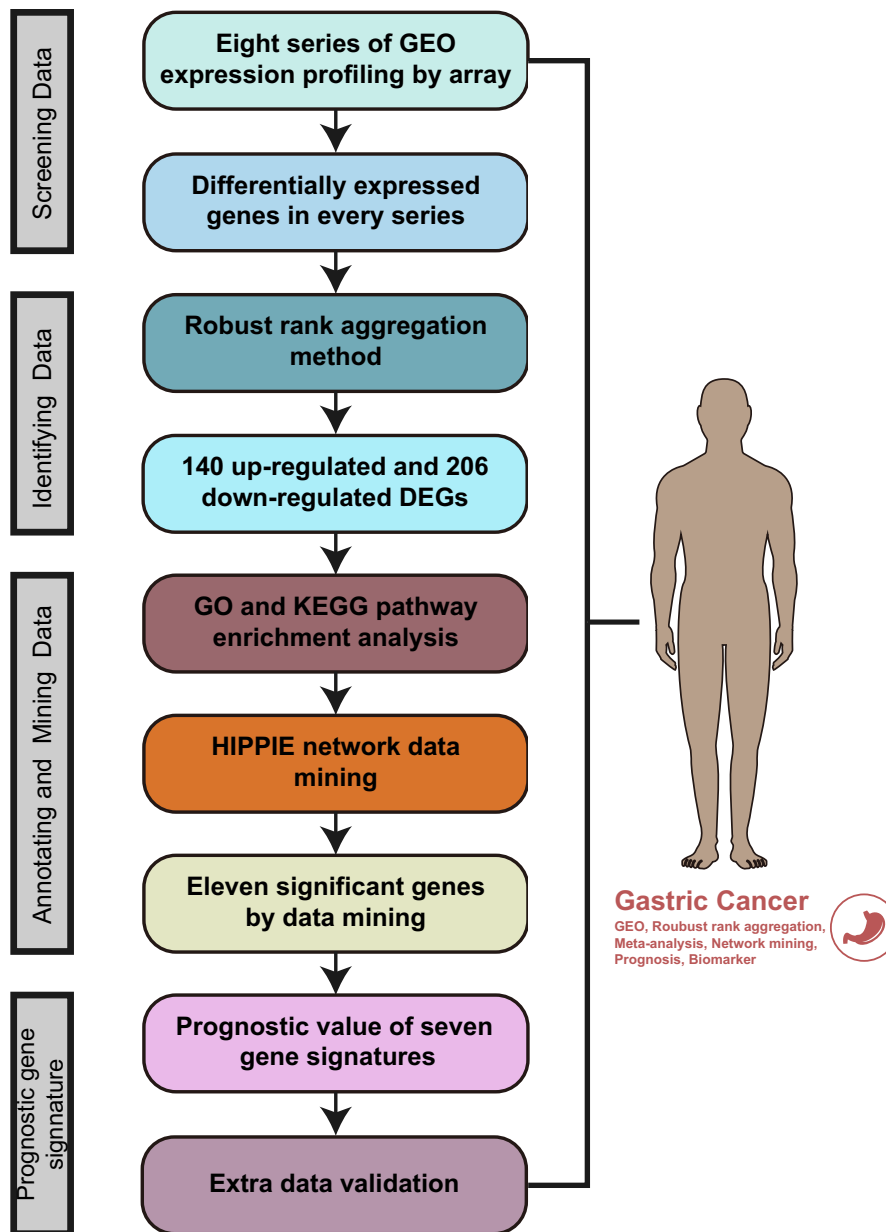


FIGURE 1 Workflow of our multi-step strategy for identifying a gene signature with prognostic value in gastric cancer (GC)

expression signature based on five genes was established using univariate survival analysis and the LASSO method.⁸ However, there have been no previous reports of the detection of DEGs in GC using the RRA method.

In this study, we not only selected significant DEGs using integrated and robust bioinformatics methods comprising various GEO series, the RRA method and network data mining. But also we developed a seven-gene signature with prognostic value. The RRA method can avoid results error of a mass of DEGs we obtained. In our network data mining, we used HIPPIE and Fisher's exact test to generate significant hub genes, which were more accurate and robust. Moreover, based on above DEGs, we establish a seven-gene signature which was validated with an external data set independently and accurately.

As for these seven genes, six of them play important roles in the molecular mechanism of GC progression. *SPARC* has been shown associated with cancer progression.⁹ Moreover, *SPARC*,¹⁰ *COL1A2*¹¹ and *ATP4A*¹² were identified as DEGs in GC, which were consistent with our results. *PLAU* was involved in the prediction of GC patient survival¹³ and *MMP1* is mostly associated with genetic polymorphisms.^{14,15} However, no previous study has considered the possible role of *FBN1* in GC.

In this study, we not only find robust and strong DEGs in GC using integrated multi-step analysis including RRA method, enrichment analysis and network data mining, but also build a seven-gene signature with prognostic value for GC based on above DEGs. Our findings may provide novel insights and potential biomarkers for GC prognosis.

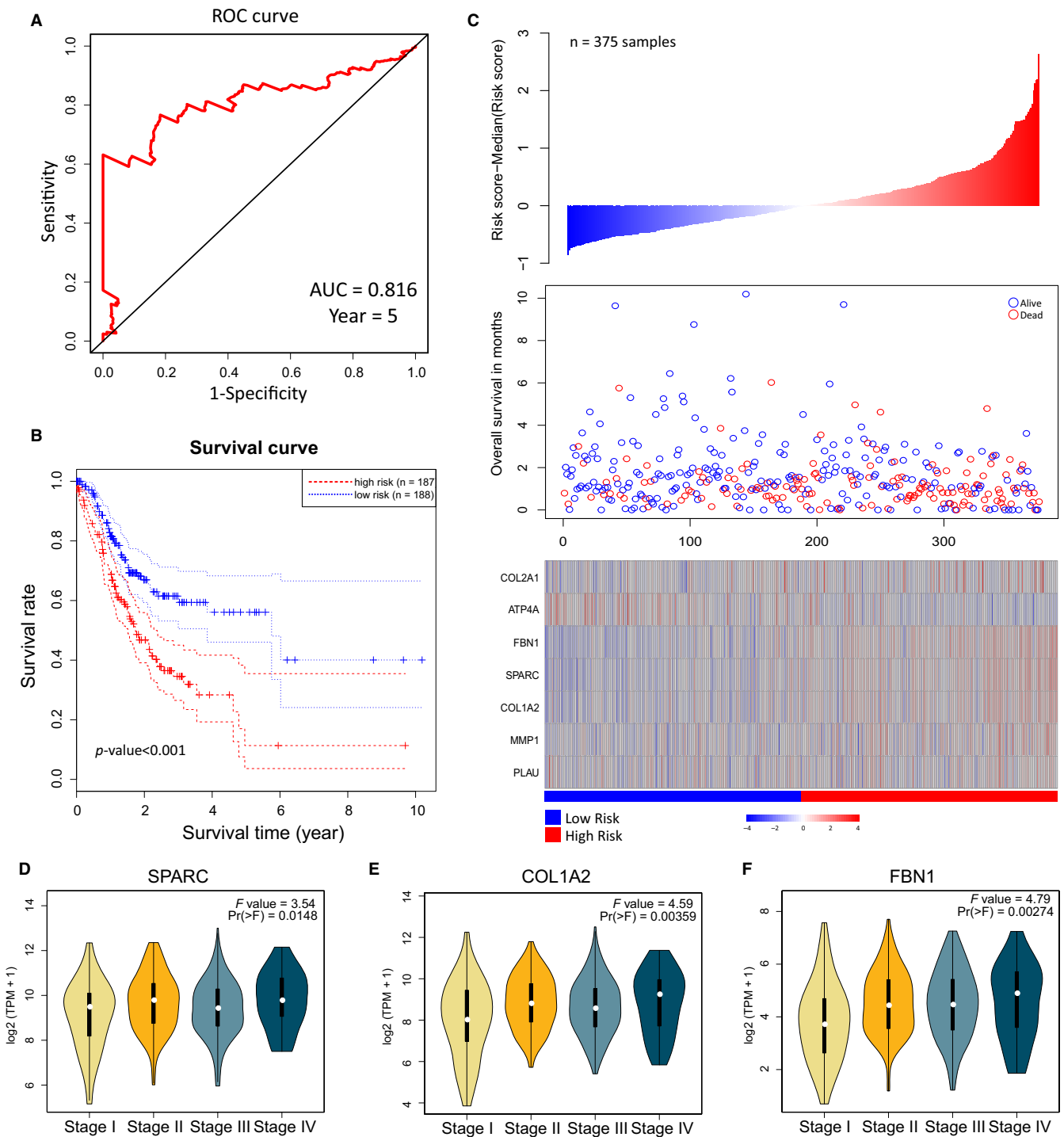


FIGURE 2 Establishment of a seven-gene signature prognostic risk scoring system based on above DEGs. A, Time-dependent ROC curve for predicting the 5-y survival. B, Kaplan-Meier curve for the seven-gene signature (log-rank test P -value < 0.001). The two dotted lines in each group are the level for a two-sided confidence interval on the survival curve. C, The seven-gene signature-based risk score distributions, patient survival results and expression heatmap. D-F, Expression profiles of *SPARC*, *COL1A2* and *FBN1* in different TNM stages of GC

ACKNOWLEDGEMENTS

We gratefully thank the department of Surgical Oncology of The First Hospital of China Medical University for technical assistance. We thank International Science Editing (<http://www.international-science.com>) for editing this manuscript. This work was supported by National Key R&D Program of China (MOST-2016YFC1303200, MOST-2016YFC1303202, MOST-2017YFC0908300, MOST-2017YFC0908305).

ncediting.com) for editing this manuscript. This work was supported by National Key R&D Program of China (MOST-2016YFC1303200, MOST-2016YFC1303202, MOST-2017YFC0908300, MOST-2017YFC0908305).

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

ORCID

Zhenning Wang  <http://orcid.org/0000-0003-0557-3097>

REFERENCES

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin*. 2015;65:87-108.
2. Karimi P, Islami F, Anandasabapathy S, Freedman ND, Kamangar F. Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention. *Cancer Epidemiol Biomarkers Prev*. 2014;23:700-713.
3. Li H, Liu JW, Liu S, Yuan Y, Sun LP. Bioinformatics-based identification of methylated-differentially expressed genes and related pathways in gastric cancer. *Dig Dis Sci*. 2017;62:3029-3039.
4. Li X, Dong W, Qu X, et al. Molecular dysexpression in gastric cancer revealed by integrated analysis of transcriptome data. *Oncol Lett*. 2017;13:3177-3185.
5. Li H, Yu B, Li J, et al. Characterization of differentially expressed genes involved in pathways associated with gastric cancer. *PLoS ONE*. 2015;10:e0125013.
6. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics*. 2012;28:573-580.
7. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017;45:W98-W102.
8. Hou JY, Wang YG, Ma SJ, Yang BY, Li QP. Identification of a prognostic 5-Gene expression signature for gastric cancer. *J Cancer Res Clin Oncol*. 2017;143:619-629.
9. Tai IT, Tang MJ. SPARC in cancer biology: its role in cancer progression and potential for therapy. *Drug Resist Updat*. 2008;11:231-246.
10. Gao Y, Yin SP, Xie XS, Xu DD, Du WD. The relationship between stromal cell derived SPARC in human gastric cancer tissue and its clinicopathologic significance. *Oncotarget*. 2017;8:86240-86252.
11. Zhuo C, Li X, Zhuang H, et al. Elevated THBS2, COL1A2, and SPP1 expression levels as predictors of gastric cancer prognosis. *Cell Physiol Biochem*. 2016;40:1316-1324.
12. Zang S, Guo R, Xing R, et al. Identification of differentially-expressed genes in intestinal gastric cancer by microarray analysis. *Genomics Proteomics Bioinformatics*. 2014;12:276-283.
13. Xu ZY, Chen JS, Shu YQ. Gene expression profile towards the prediction of patient survival of gastric cancer. *Biomed Pharmacother*. 2010;64:133-139.
14. Fang WL, Liang WB, Gao LB, Zhou B, Xiao FL, Zhang L. Genetic polymorphisms in Matrix Metalloproteinases -1 and -7 and susceptibility to gastric cancer: an association study and meta-analysis. *Iran J Allergy Asthma Immunol*. 2013;12:203-210.
15. Yang TF, Guo L, Wang Q. Meta-analysis of associations between four polymorphisms in the matrix metalloproteinases gene and gastric cancer risk. *Asian Pac J Cancer Prev*. 2014;15:1263-1267.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Wang J, Gao P, Song Y, et al. Prognostic value of gastric cancer-associated gene signatures: Evidence based on a meta-analysis using integrated bioinformatics methods. *J Cell Mol Med*. 2018;22:5743-5747. <https://doi.org/10.1111/jcmm.13823>