A novel gene expression-based prognostic scoring system to predict survival in gastric cancer

Pin Wang[1,*], Yunshan Wang[2,3,*], Bo Hang[2], Xiaoping Zou[1], Jian-Hua Mao[2]

[1]Department of Gastroenterology, Drum Tower Clinical Medical School Of Nanjing Medical University, Nanjing, Jiangsu 210008, China

[2]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

[3]International Biotechnology R&D Center, Shandong University School of Ocean, Weihai, Shandong 264209, China

[*]These authors have contributed equally to this work

Correspondence to: Xiaoping Zou, **email:** zouxiaoping795@hotmail.com
        Jian-Hua Mao, **email:** JHMao@lbl.gov

ABSTRACT

Analysis of gene expression patterns in gastric cancer (GC) can help to identify a comprehensive panel of gene biomarkers for predicting clinical outcomes and to discover potential new therapeutic targets. Here, a multi-step bioinformatics analytic approach was developed to establish a novel prognostic scoring system for GC. We first identified 276 genes that were robustly differentially expressed between normal and GC tissues, of which, 249 were found to be significantly associated with overall survival (OS) by univariate Cox regression analysis. The biological functions of 249 genes are related to cell cycle, RNA/ncRNA process, acetylation and extracellular matrix organization. A network was generated for view of the gene expression architecture of 249 genes in 265 GCs. Finally, we applied a canonical discriminant analysis approach to identify a 53-gene signature and a prognostic scoring system was established based on a canonical discriminant function of 53 genes. The prognostic scores strongly predicted patients with GC to have either a poor or good OS. Our study raises the prospect that the practicality of GC patient prognosis can be assessed by this prognostic scoring system.

INTRODUCTION

Gastric cancer (GC) is a malignant tumor initiated from the epithelial cells of gastric mucosa. GC has been one of the most common malignant tumors in the world and ranks fifth in the incidence rate, following lung cancer, breast cancer, colorectal cancer and prostate cancer [1, 2]. Despite of the slightly reduced overall incidence and mortality of GC over the past decade, to date, the incidence and mortality of GC still remain very high. Moreover, the number of people suffering from GC follows an upward trend, and there are about one million of new cases each year [1, 2]. With the advances in science and biotechnology, the level of early diagnosis for GC has been improved to certain extent, which, in turn, significantly improves its five-year survival rate. Even so, the five-year survival rate of advanced GC is only about 29.3%, which is due to that GC is prone to relapse and metastasis [1-4].

GC is a polygenic disease, where the interactions of various cancer genes with the microenvironment *in vivo* lead the early lesions of gastric mucosa to the dysplasia, and ultimately to the development of GC. The characteristically differential expression of related genes can be observed throughout the whole process. In clinical practice, there has been a lack of corresponding molecular markers for the distinguishment of GC staging and degree of differentiation. Through analyzing the GC and adjacent normal tissues using the microarray technology, Cui *et al* [5] first established a gene expression profile related to GC staging and differentiation. A total of 19 genes were involved in the expression profile, which was reported to be able to distinguish the highly differentiated and poorly differentiated GC with a relatively high accuracy rate. In addition, they also discovered a series of gene expression profiles capable of identifying GC staging.

In recent years, researchers also proposed that the molecular biological characteristics in GC composition

play an important role in the prognosis. Currently, the overexpression of *HER2* gene is closely associated with the prognosis and lymph node metastasis of GC [6]. p53 is a broadly studied tumor suppressor associated with malignancy, and the accumulation of this protein in GC has also been confirmed and appears to be negatively correlated with the prognosis. By determining p53 in plasma and stomach tissue, Mattioni *et al* [7] found that the survival rate of patients with positive anti-p53 expression in plasma was significantly higher than those with the negative result. The transcription factor hypoxia-inducible factor 1 alpha (HIF-1α) is highly expressed in GC cells and exhibits an even higher expression in patients with GC at the early stage as identified by TNM classification. Therefore, HIF-1α may be related to the early development of GC and understanding its function may be helpful in exploring GC origin [8].

In the current cancer research, there are still certain difficulties in analyzing the biological significance of most genes. The gene expression profiling technology is of great significance for the investigation of different subtypes and their prognosis, and the construction of genes into a network with the help of gene expression profiling technology proves to be critical for the understanding of cancer initiation and development. Based on the analysis on the transcriptional profiles of GC at different stages, Takeno *et al.* [9] constructed a GC regulatory network with CDKNIA as the node and screened out seven genes related to GC occurrence (*i.e.*, MMP7, SPARC, SOD2, INHBA, IGFBP7, NEK6 and LUM); through dividing these seven genes into two groups according to their correlation with expression levels and stages, the results showed that these seven genes were activated as the disease progressed, indicating that these genes may be associated with cancer development.

In this study, we tested the hypothesis that molecular features of GC are a key driver of tumor behavior, which can be used to establish prognostic scoring system that can improve prediction of clinical outcome. We first established a multi-step analytic approach to identify a comprehensive panel of gene biomarkers using publically available and well-characterized dataset and TCGA (The Cancer Genome Atlas), and then we employed different multivariate clustering techniques to identify the key genes for prognostic classification. Based on these analyses, we created a 53-gene expression prognostic scoring system and successfully applied it to predict overall survival (OS) in the TCGA GC as well as the GSE15459 data.

RESULTS

Identification of robust differentially expressed genes in gastric cancers

We developed a multi-step strategy to identify a critical gene signature that is able to distinguish good and bad prognosis for GC patients using publically available datasets (Figure 1). Firstly, we sought to identify significantly differentially expressed genes through comparing gene expression between normal and GC tissues using two datasets: TCGA that was generated by RNA sequencing [10] and GSE30727 that was generated by Affymetrix microarray (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30727). A total of 688 and 3239 genes reached our criteria (2 fold changes and adjusted p-value <0.05) in TCGA and GSE30727, respectively (Supplementary Table S1). Importantly, 276 genes were found to be overlapping between TCGA and GSE30727 datasets. Of which, 57 genes were found to be downregulated while 219 genes were found to be upregulated in GCs (Supplementary Table S2).

Evaluation of prognostic impact of differentially expressed genes in gastric cancers

To further assess the importance of the above 276 genes in GC development, we next evaluated their prognostic value for GC patients in a large public clinical microarray database using the Kaplan-Meier plotter (http://kmplot.com/analysis/index.php?p=service&cancer=gastric) [11]. Each of these genes was divided into two groups based on its expression value. Subsequently, the effects of high or low expression level of these genes on the overall survival (OS) were evaluated using Cox regression analysis, the Kaplan-Meier survival curve and log-rank test. 249 out of 276 genes were found to be significantly associated with OS (Supplementary Table S3). This result suggested that these molecular markers may provide a prediction for the prognosis of GC patients. Finally we ranked the genes according to their log-rank test p-values derived from univariate analysis (Supplementary Table S3), which served as the criteria for the choice of genes into canonical discriminant function (see below). Figure 2 showed the Kaplan-Meier curves for top six genes in GCs.

Creation of gene co-expression networks for 249 genes in gastric cancer

In order to better reveal the molecular mechanism underlying GC development, we computationally mapped the 249 genes to biological functions, pathways and upstream transcriptional regulators using the Database for Annotation, Visualization and Integrated Discovery (DAVID), and observed that these genes are significantly enriched for regulating cell proliferation, adhesion and migration, RNA/non-coding (nc) RNA process, extracellular matrix organization, vasculature development, response to oxidative stress, *etc.* (Figure 3A), all of which are hallmarks of cancer. Analysis of the upstream regulators of these 249 genes suggests that the NMYC, STAT3, GATA1 and p53 pathways play a role in GC (Figure 3B).

Sets of genes that exhibit correlated expression patterns often share a common function or are part of the same physical structure. We next used a network analysis approach to identify functionally related groups of genes using TCGA data, which contain 265 patients with GC (Supplementary Table S4) [10]. We started by representing TCGA GC expression as a network where significantly correlated genes are drawn as nodes connected by an edge (FDR<0.05 and |r| ≥0.7; details see Materials and Methods). We then identified fully connected gene sets (cliques) that were enriched for functions (Figure 4). These sub-networks were enriched for genes representing hallmarks of cancer as discovered in the Figure 3A. Interestingly, the subnetworks for cell cycle, RNA/ncRNA process and acetylation are highly connected with each other, in contrast to the subnetwork for extracellular matrix (Figure 4).

Development of a 53-gene prognostic scoring system

We designed a strategy to develop a prognostic scoring system (Figure 5A). Firstly 65 good (status is alive and time of follow-up ≥ 15 months) and 43 bad (status= deceased and the time of survival < 15 months) prognostic patients (total 108 patients) were selected as a training set from 253 of 269 GC patients who have the information of OS and OS status in a TCGA-based study (Supplementary Table S4) [10]. Then we applied a stepwise canonical discriminant analysis to identify a gene signature that is able to classify 108 patients into good or bad prognosis with 100% accuracy. To determine the optimal number of genes in a given sub-network used for building the signature, combinations of genes were tested by adding one gene at a time according to their
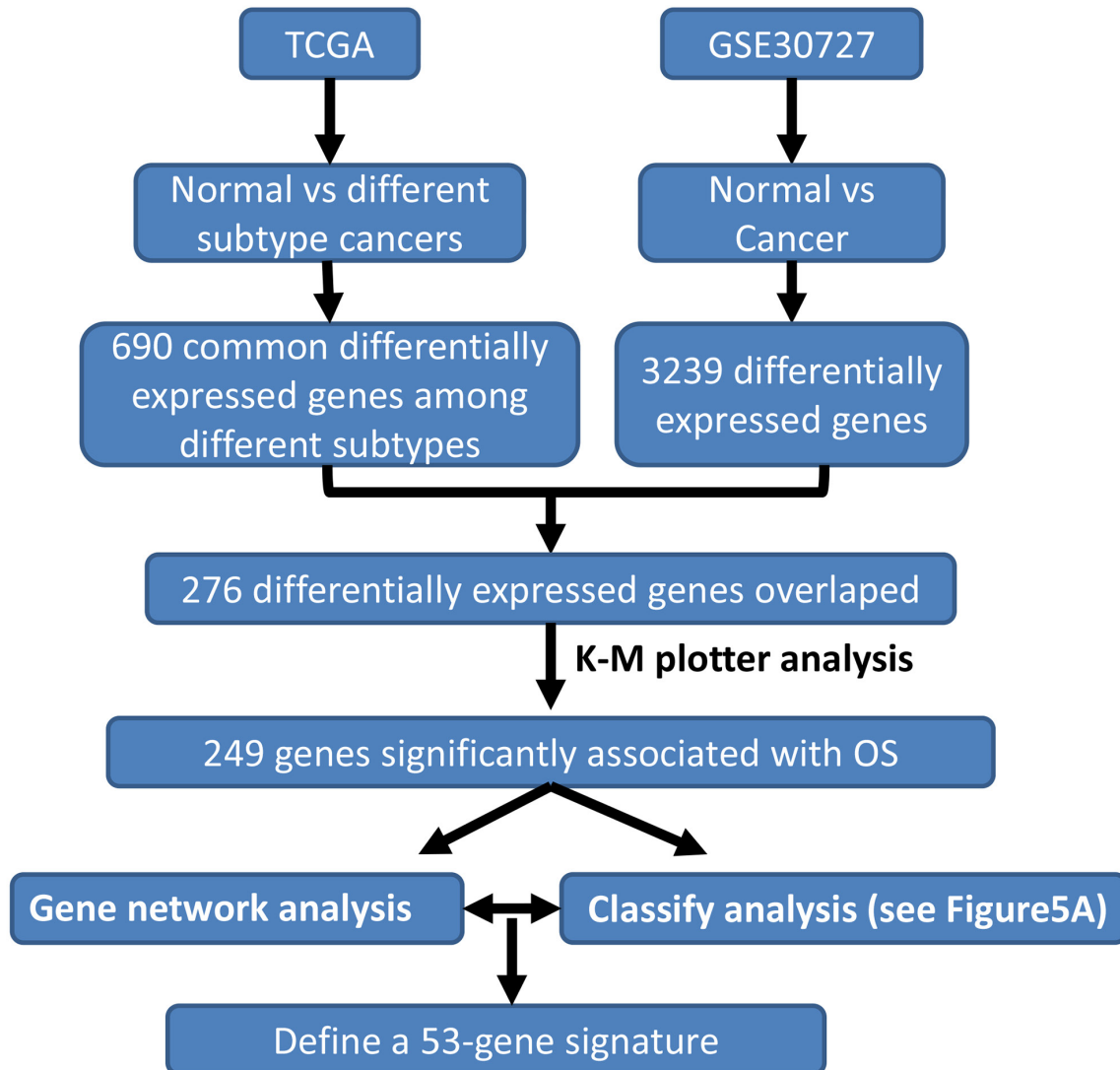

Figure 1: Schematic diagram for a multi-step strategy to identify gene signature for prognosis in gastric cancer. The results for each step have been summarized.
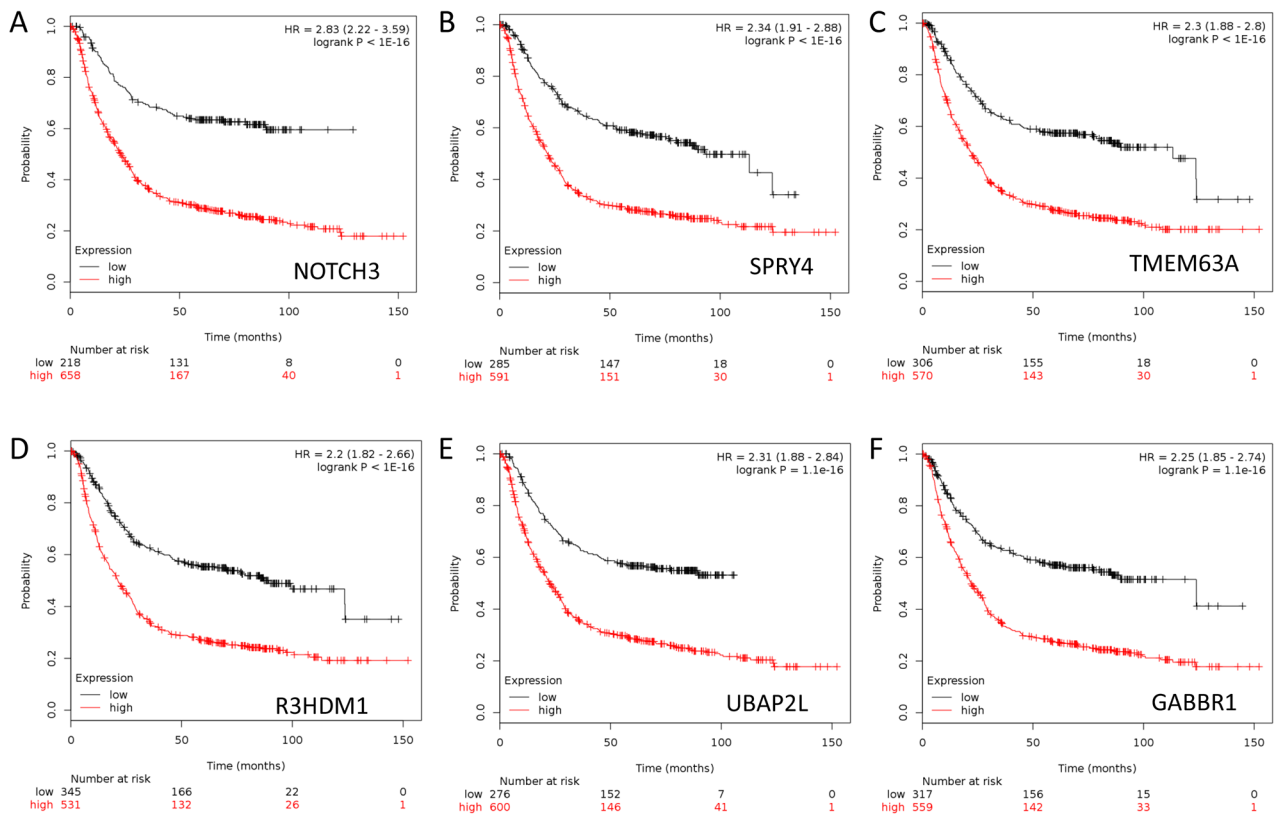
Figure 2: Kaplan-Meier survival curves for gastric cancer patients according to tumor expression of top rank 6 genes are presented. A. NOTCH3. **B.** SPRY4. **C.** TMEM63A. **D.** R3HDM1. **E.** UBAP2L. **F.** GABBR1.The p values were obtained from a log-rank test between two groups.
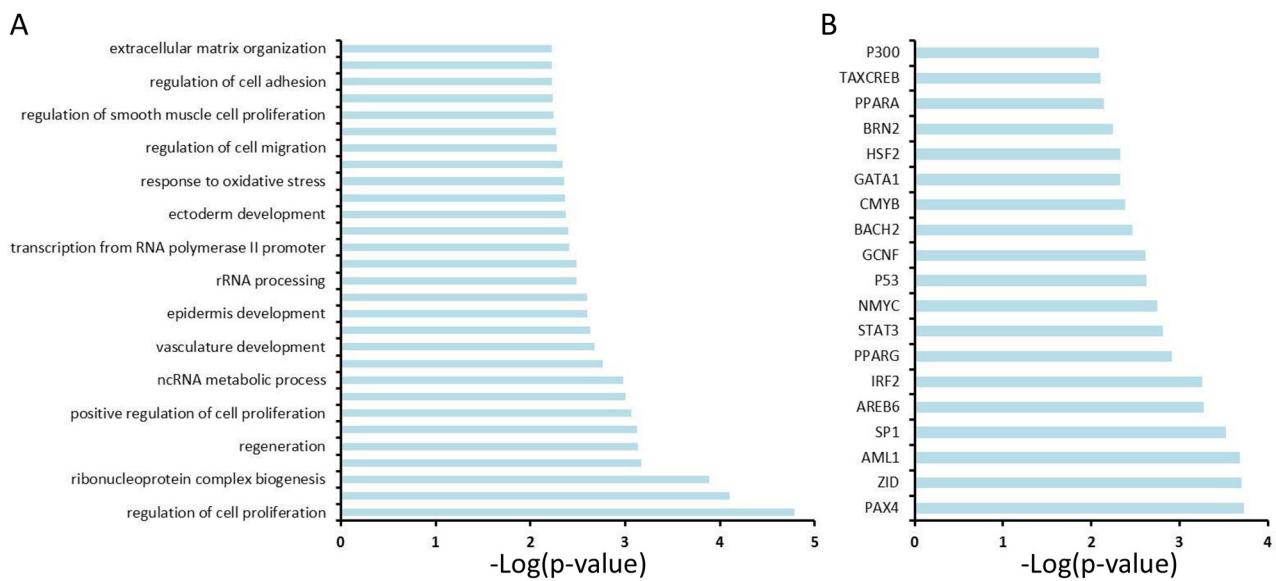

Figure 3: Functional annotation analysis of 249 genes using the Database for Annotation, Visualization and Integrated Discovery (DAVID). A. DAVID analysis reveals the potential signaling pathways that are enriched among 249 genes. **B.** DAVID analysis reveals upstream transcriptional factors that regulate the expression of 249 genes.

ranks given above. The number of significant genes that gave the highest discriminative ability between good and bad prognostic groups was considered optimal. The same procedure was executed for those genes that were not in the gene co-expression network. Finally we identified a 53-gene signature to yield 100% accuracy to separate 108 patients into good or bad prognosis. A prognostic score for a patient was used to calculate a patient's risk of death and was defined as the linear combination of logarithmically transformed gene expression levels weighted by canonical discriminant function coefficients (Supplementary Table S5). The distributions of prognostic score in good and bad prognostic patients was clearly discriminative (Figure 5B), indicating that this prognostic scoring system has its discriminative ability to distinguish good prognostic patients from bad prognostic patients.

Prediction of overall survival in gastric cancer patients

The above-developed prognostic scoring system was applied to all 253 GC patients who have the information of OS and OS status in a TCGA-based study [10]. Prognostic score was used to predict survival probability for each individual patient. We divided patients into two groups named good and bad signature based on prognostic score. If the prognostic score is ≤ -2, we defined that the patient had good signature; if the prognostic score is >-2, we defined the patient as bad signature. As shown in Figure 5C, the patients with good signature had significantly longer overall survival

than those with bad signature. More than 50% of patients with good signature still survived after 100 months while all patients with bad signature died before 80 months. Next, we validated this prognostic scoring system using the GSE15459 public dataset (Supplementary Table S6) [12]. Although the GC tissues were profiled by Affymetrix microarray, which caused the difference in baseline and expression level scale, the prognostic score still predicted prognosis, the patients with lower score (good signature, the 1st quartile) significantly survived longer than the others (bad signature, the 2nd – 4th quartiles) (Figure 5D). These results raise the prospect that the practicality of gastric cancer patient prognosis can be assessed by this prognostic scoring system.

DISCUSSION

GC is one of the leading causes of death among women worldwide [1, 2]. A growing body of evidence has demonstrated that GC is a complex and heterogeneous disease with substantial variation in their molecular and clinical characteristics. Microarray and next generation sequencing technologies have been invaluable tools to deconvolute the heterogeneity and complexity of somatic GC genetics, providing tremendous information to define new biomarkers for diagnosis, prognosis and prediction of therapeutic response, and to identify new potential therapeutic targets. Several molecular characterization studies have been conducted in GC [10, 13–15], including few that have attempted to identify the gene signature for
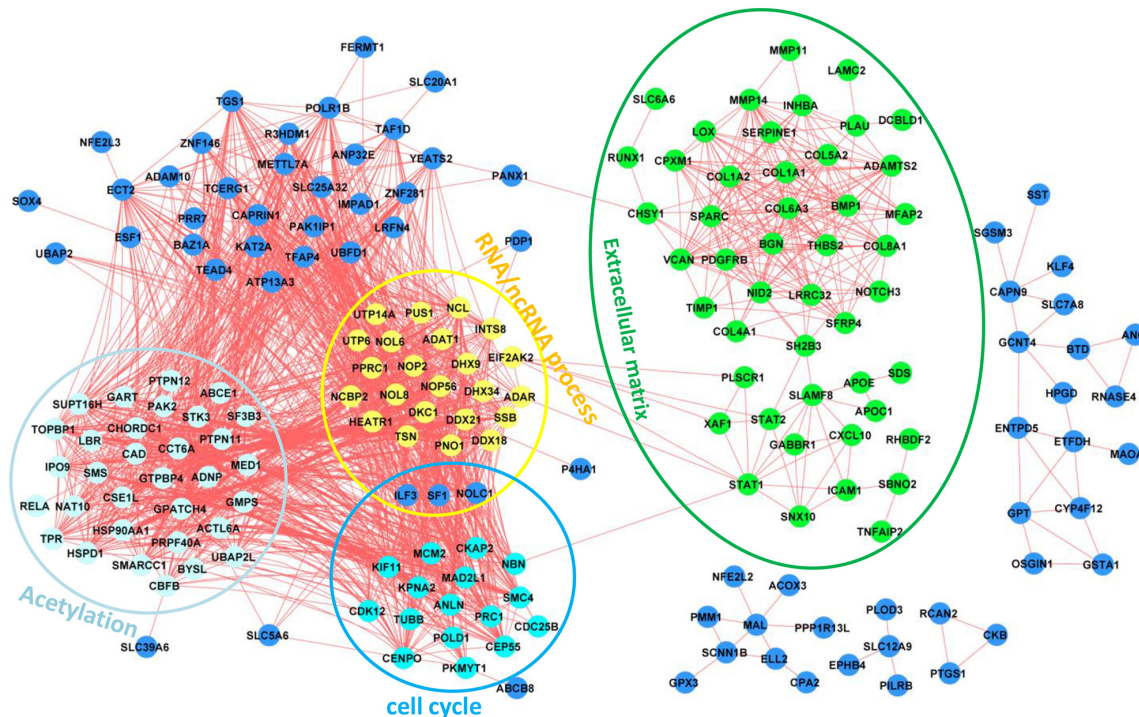

Figure 4: Gene correlation networks of the 249 genes. Functional annotations have been indicated for different subnetworks.

prognosis in GC [12, 16–18]. However, there have been no reports on a prognostic scoring system that can be practically used in preclinical and clinical research. In this study, we have developed a multi-step strategy to define a 53-gene signature for prediction of overall survival for GC patients using the multi-omics data, and for the first time established a prognostic scoring system based on such a 53-gene signature. We also showed that the prognostic scores are able to distinguish patients with good prognosis from those with bad prognosis.

It is not surprising to find that among 53 genes, some of them have already been implicated in GC in previously published studies, including providing useful prognostic information about the survival. These genes include *TNFAIP2* [19], *FGFR4* [20–22], *CXCL10* [23], *CEP55* [24], *CXCL1* [25, 26], *LIMK1* [27], *LAMC2* [28], *APOE* [29], *INHBA* [30], *OSMR* [31, 32], *APOC1* [33], *KLF4* [34], *MMP14* [35], *ADH1C* [36], *COL6A3* [37, 38], *CCT2* [39], *NOL8* [40], *EPHB4* [41] and *MCM2* [42, 43]. Ye *et al* reported that high expression of FGFR4 protein is associated with a poor prognosis in patients with advanced GC and expedites the progress of advanced

GC [20]. Moreover, Shen *et al* found that the FGFR4 Gly388Arg polymorphism is a useful prognostic marker for GC patients when the tumor is relatively small, well differentiated, or at an early clinical stage [21]. CEP55 is one of the centrosome family proteins and functions in cell cycle regulation, which is one of the cancer hallmarks. Knockdown of CEP55 led to reduced proliferation and colony formation in SGC7901 and BGC823 cell lines through affecting the PI3K/AKT signal pathway and the expression of cyclin-related proteins, suggesting that CEP55 might be a potential therapeutic target for GC [24]. Consistent with our finding, some studies showed that MCM2 expression levels predict diagnosis and prognosis in GC [42, 43]. miRNAs are a vital and evolutionarily ancient component of gene regulation. SNPs in miRNA binding sites could affect its binding affinity to target genes. A recent study showed that the TNFAIP2 miRNA binding site rs8126 T>C SNP may be a marker for susceptibility to GC [19].

Although the function and role of some other genes in the 53 genes selected have not been reported for their association with GC, their importance as
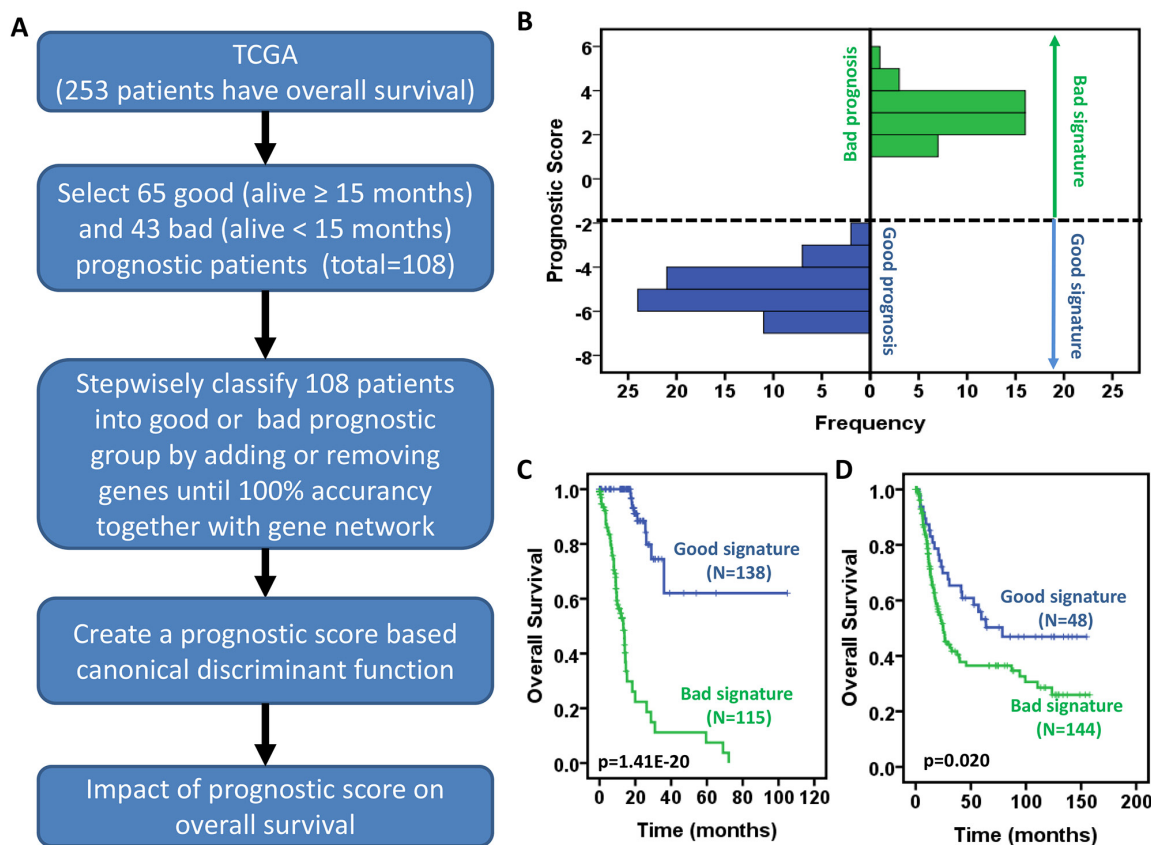

Figure 5: Development of a prognostic scoring system for gastric cancer patients. A. Schematic diagram for a multi-step strategy to develop a prognostic scoring system for gastric cancer patients using the TCGA data. B. Distribution of prognostic score between patients with good and bad prognosis. C-D. Prognostic scores are significantly associated with overall survival of gastric cancer patients in TCGA GC data (C) and GSE15459 (D). Kaplan-Meier survival curves for gastric cancer patients according to prognostic scores. The p values were obtained from a log-rank test between two groups.

cancer genes have been demonstrated in many studies with other types of human cancers. For example, Chemokine (C-X-C motif) ligand (CXCL1) plays a critical role in tumor metastasis and is demonstrated to be significantly associated with Snail expression. It has been reported that expression of *CXCL1* is associated with hepatocellular carcinoma survival [25]. Our study showed that *CXCL1* is implicated in GC overall survival. ABCE1 (ATP-binding cassette E1) plays a crucial role in the metastasis and progress of lung cancers, and therefore it has been suggested as a valuable therapeutic target for the management of these cancers [44]. We have demonstrated in this study that ABCE1 is elevated in GC and is an indicator for the prognosis of GC.

In summary, using the available multi-omics data, we have generated a prognostic scoring system that has been demonstrated successful prediction of patient overall survival in TCGA and a microarray dataset. Knowing the accurate prognosis of a patient with GC is very important for the determination of a most suitable clinical therapy for the patient. Clearly further studies are needed to establish the clinical application of this prognostic scoring system for GC.

MATERIALS AND METHODS

Datasets used in this study

The differentially expressed genes were assessed in the TCGA [10] and microarray dataset (GSE30727) profiled with Human Exon 1.0 ST Array (HuEx-1_0-st). The process data for GSE15459 from GEO website were downloaded for analysis. The expression levels of 276 genes for the set of samples in each TCGA study were obtained from the cBioPortal for cancer genome (Supplementary Table S4) [45, 46].

Gene co-expression network construction

A network of 249 genes was constructed based on the Expression Correlation software tool (http://baderlab.org/Software/ExpressionCorrelation). Correlation coefficients exceeding a threshold ($|r| \geq 0.7$) and false discover rate (FDR <0.05) were displayed as 'edges' between two 'nodes' (where nodes represent genes), and this approach was used to define the 249 genes co-expression network. The gene co-expression network figure was generated using Cytoscape version 2.8.0 (www.cytoscape.org). The transcriptional network of the 249 genes was assessed using the data from TCGA [10].

Statistics analysis

GEO2R was used to identify the differentially expressed genes between normal and tumor tissues in GSE30272 (adjusted p-value ≤0.05 and fold changes ≥2). The list of differentially expressed genes in TCGA was obtained from the published study [10].

The association of 276 genes with overall survival of GC patients was assessed using the Kaplan-Meier plotter (http://kmplot.com/analysis/index.php?p=service&cancer=gastric). The log-rank test p-values derived from Kaplan-Meier analysis were used to rank the gene.

108 patients in TCGA study [10] were selected as a training set using two-step stratified sampling methods. In the first step, patients were divided into two groups based on their survival status (alive *VS* deceased), and in the second step, patients in each group were further divided based on their survival time (<15 months or ≥15 months).

After identifying the relevant genes and their influence upon the prognostic model in 108 training samples, we established a prognostic scoring system as following:

$$\textbf{Prognostic score} = \sum_{i=1}^{53} (Canonical\ Discriminant\ Function\ Coefficient) * (gene\ i\ expression\ level)$$

And then the individuals of the populations were assigned a prognostic score to stratify them into two subgroups of prognostic relevance. In this preliminary work, the groups were assigned according to prognostic score as following:

If the prognostic score ≤ -2, we defined the patient with good signature; if the prognostic score > -2, we defined the patient with bad signature. This helped to ultimately stratify the cohorts of patients in a Kaplan–Meier curve. All statistical analyses were performed using the Statistical Package for the Social Sciences version 11.5 (SPSS, Inc., Chicago, IL).

CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

REFERENCES

1. Torre LA, Siegel RL, Ward EM, Jemal A. Global Cancer Incidence and Mortality Rates and Trends-An Update. Cancer Epidemiol Biomarkers Prev. 2016; 25:16-27.

2. 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.

3. R. Wadhwa, S. Song, J. Lee, Y. Yao, Q. Wei, J.A. Ajani. Gastric cancer – molecular and clinical dimensions Nat. Rev. Clin. Oncol, 2013; 10: 643–655

4. Lu N, Li R, Liu Q, Hu B, Xu X, Ji C, Han X, Wang P, Liu B. Antitumor and antimetastatic effects of pemetrexed-loaded targeted nanoparticles in B16 bearing mice. Drug delivery. 2015; 26:1-9.

5. Cui J, Li F, Wang G, Fang X, Puett JD, Xu Y. Gene-Expression Signatures Can Distinguish Gastric Cancer Grades and Stages. PLoS ONE. 2011; 6: e17819.

6. Hecht JR, Bang YJ, Qin SK, Chung HC, Xu JM, Park JO, Jeziorski K, Shparyk Y, Hoff PM, Sobrero A, Salman P, Li J, Protsenko SA, et al. Lapatinib in Combination With Capecitabine Plus Oxaliplatin in Human Epidermal Growth Factor Receptor 2-Positive Advanced or Metastatic Gastric, Esophageal, or Gastroesophageal Adenocarcinoma: TRIO-013/LOGiC-A Randomized Phase III Trial. J Clin Oncol. 2016; 34: 443-51.

7. Mattioni M, Soddu S, Porrello A, D'Alessandro R, Spila A, Guadagni F. Serum anti-p53 antibodies as a useful marker for prognosis of gastric carcinoma. The International journal of biological markers. 2006; 22:302-6.

8. Cabuk D, Basaran G, Celikel C, Dane F, Yumuk PF, Iyikesici MS, Ekenel M, Turhal NS. Vascular endothelial growth factor, hypoxia-inducible factor 1 alpha and CD34 expressions in early-stage gastric tumors: relationship with pathological factors and prognostic impact on survival. Oncology. 2007; 72:111-7.

9. Takeno A, Takemasa I, Doki Y, Yamasaki M, Miyata H, Takiguchi S, Fujiwara Y, Matsubara K, Monden M. Integrative approach for differentially overexpressed genes in gastric cancer by combining large-scale gene expression profiling and network analysis. British journal of cancer. 2008; 99:1307-15.

10. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014; 513: 202-9.

11. Győrffy B, Surowiak P, Budczies J, Lánczky A. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. PloS one. 2013; 8: e82241.

12. Ooi CH, Ivanova T, Wu J, Lee M, Tan IB, Tao J, Ward L, Koo JH, Gopalakrishnan V, Zhu Y, Cheng LL, Lee J, Rha SY, et al. Oncogenic pathway combinations predict clinical prognosis in gastric cancer. PLoS Genet. 2009; 5: e1000676.

13. Hara R, Kikuchi H, Setoguchi T, Miyazaki S, Yamamoto M, Hiramatsu Y, Kamiya K, Ohta M, Baba S, Konno H. Microarray analysis reveals distinct gene set profiles for gastric and intestinal gastrointestinal stromal tumors. Anticancer Res. 2015; 35: 3289-98.

14. Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, Liu J, Yue YG, Wang J, Yu K, Ye XS, Do IG, Liu S, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med. 2015; 21:449-56.

15. Wong SS, Kim KM, Ting JC, Yu K, Fu J, Liu S, Cristescu R, Nebozhyn M, Gong L, Yue YG, Wang J, Ronghua C, Loboda A, et al. Genomic landscape and genetic heterogeneity in gastric adenocarcinoma revealed by whole-genome sequencing. Nat Commun. 2014;5: 5477.

16. Sun H. Identification of key genes associated with gastric cancer based on DNA microarray data. Oncology Letters. 2016; 11:525-30.

17. Zhao X, Cai H, Wang X, Ma L. Discovery of signature genes in gastric cancer associated with prognosis. Neoplasma. 2016; 63:239-45.

18. Yin Y, Zhuo W, Zhao Y, Chen S, Li J, Wang L, Zhou T, Si JM. Converting a microarray signature into a diagnostic test: a trial of custom 74 gene array for clarification and prediction the prognosis of gastric cancer. PLoS One. 2013; 8:e81561.

19. Xu Y, Ma H, Yu H, Liu Z, Wang L-E, Tan D, Muddasani R, Lu V, Ajani JA, Wang Y, Wei Q. The miR-184 binding-site rs8126 T> C polymorphism in TNFAIP2 is associated with risk of gastric cancer. PLoS One 2013; 8:e64973.

20. Ye YW, Zhang X, Zhou Y, Wu J, Zhao C, Yuan L, Wang G, Du C, Wang C, Shi Y. The correlations between the expression of FGFR4 protein and clinicopathological parameters as well as prognosis of gastric cancer patients. Journal of surgical oncology. 2012; 106:872-9.

21. Shen YY, Lu YC, Shen DP, Liu YJ, Su XY, Zhu GS, Yin XL, Ni XZ. Fibroblast growth factor receptor 4 Gly388Arg polymorphism in Chinese gastric cancer patients. World journal of gastroenterology: WJG. 2013; 19: 4568-75.

22. Murase H, Inokuchi M, Takagi Y, Kato K, Kojima K, Sugihara K. Prognostic significance of the co-overexpression of fibroblast growth factor receptors 1, 2 and 4 in gastric cancer. Molecular and clinical oncology. 2014; 2: 509-17.

23. Rajkumar T, Vijayalakshmi N, Gopal G, Sabitha K, Shirley S, Raja UM, Ramakrishnan SA. Identification and validation of genes involved in gastric tumorigenesis. Cancer cell international. 2010; 10:45.

24. Tao J, Zhi X, Tian Y, Li Z, Zhu Y, Wang W, Xie K, Tang J, Zhang X, Wang L, Xu Z. CEP55 contributes to human gastric carcinoma by regulating cell proliferation. Tumor Biology. 2014; 35:4389-99.

25. Xiang Z, Jiang DP, Xia GG, Wei ZW, Chen W, He Y, Zhang CH. CXCL1 expression is correlated with Snail

expression and affects the prognosis of patients with gastric cancer. Oncology letters. 2015; 10:2458-64.

26. Junnila S, Kokkola A, Mizuguchi T, Hirata K, Karjalainen-Lindsberg ML, Puolakkainen P, Monni O. Gene expression analysis identifies over-expression of CXCL1, SPARC, SPP1, and SULF1 in gastric cancer. Genes, Chromosomes and Cancer. 2010; 49:28-39.

27. You T, Gao W, Wei J, Jin X, Zhao Z, Wang C, Li Y. Overexpression of LIMK1 promotes tumor growth and metastasis in gastric cancer. Biomedicine & Pharmacotherapy. 2015; 69: 96-101.

28. Yamamoto H, Kitadai Y, Yamamoto H, Oue N, Ohdan H, Yasui W, Kikuchi A. Laminin gamma2 mediates Wnt5a-induced invasion of gastric cancer cells. Gastroenterology. 2009; 137:242.

29. Sakashita K, Tanaka F, Zhang X, Mimori K, Kamohara Y, Inoue H, Sawada T, Hirakawa K, Mori M. Clinical significance of ApoE expression in human gastric cancer. Oncology reports. 2008; 20:1313-9.

30. Oshima T, Yoshihara K, Aoyama T, Hasegawa S, Sato T, Yamamoto N, Akito N, Shiozawa M, Yoshikawa T, Numata K, Rino Y, Kunisaki C, Tanaka K, et al. Relation of INHBA gene expression to outcomes in gastric cancer after curative surgery. Anticancer research. 2014; 34:2303-9.

31. Junnila S, Kokkola A, Karjalainen-Lindsberg ML, Puolakkainen P, Monni O. Genome-wide gene copy number and expression analysis of primary gastric tumors and gastric cancer cell lines. BMC cancer. 2010; 10:73.

32. Deng G, Kakar S, Okudiara K, Choi E, Sleisenger MH, Kim YS. Unique methylation pattern of oncostatin m receptor gene in cancers of colorectum and other digestive organs. Clinical Cancer Research. 2009;15:1519-26.

33. Oue N, Hamai Y, Mitani Y, Matsumura S, Oshimo Y, Aung PP, Kuraoka K, Nakayama H, Yasui W. Gene Expression Profile of Gastric Carcinoma Identification of Genes and Tags Potentially Involved in Invasion, Metastasis, and Carcinogenesis by Serial Analysis of Gene Expression. Cancer research. 2004; 64:2397-405.

34. Zhang L, Kim S, Ding W, Tong Y, Zhang X, Pan M, Chen S. Arsenic sulfide inhibits cell migration and invasion of gastric cancer *in vitro* and in *vivo*. Drug design, development and therapy. 2015; 9:5579-90.

35. Dong Y, Chen G, Gao M, Tian X. Increased expression of MMP14 correlates with the poor prognosis of Chinese patients with gastric cancer. Gene. 2015; 563:29-34.

36. Duell EJ, Sala N, Travier N, Muñoz X, Boutron-Ruault MC, Clavel-Chapelon F, Barricarte A, Arriola L, Navarro C, Sánchez-Cantalejo E, Quirós JR, Krogh V, Vineis P, et

al. Genetic variation in alcohol dehydrogenase (ADH1A, ADH1B, ADH1C, ADH7) and aldehyde dehydrogenase (ALDH2), alcohol consumption and gastric cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort. Carcinogenesis. 2012; 33:361-7.

37. Xie X, Liu X, Zhang Q, Yu J. Overexpression of collagen VI α3 in gastric cancer. Oncology letters. 2014; 7:1537-43.

38. Hu K, Chen F. Identification of significant pathways in gastric cancer based on protein-protein interaction networks and cluster analysis. Genetics and molecular biology. 2012; 35:701-8.

39. Malta-Vacas J, Nolasco S, Monteiro C, Soares H, Brito M. Translation termination and protein folding pathway genes are not correlated in gastric cancer. Clinical Chemistry and Laboratory Medicine. 2009; 47:427-31.

40. Jinawath N, Furukawa Y, Nakamura Y. Identification of NOL8, a nucleolar protein containing an RNA recognition motif (RRM), which was overexpressed in diffuse-type gastric cancer. Cancer science. 2004; 95:430-5.

41. Liersch-Löhn B, Slavova N, Buhr HJ, Bennani-Baiti IM. Differential protein expression and oncogenic gene network link tyrosine kinase ephrin B4 receptor to aggressive gastric and gastroesophageal junction cancers. International Journal of Cancer. 2016; 138:1220-31.

42. Liu M, Li JS, Tian DP, Huang B, Rosqvist S, Su M. MCM2 expression levels predict diagnosis and prognosis in gastric cardiac cancer. Histology and histopathology. 2013; 28:481-92.

43. Yang C, Wen Y, Li H, Zhang D, Zhang N, Shi X, Jiang B, Ma X, Yang P, Tang H, Peng Z, Yang Y. Overexpression of minichromosome maintenance 2 predicts poor prognosis in patients with gastric cancer. Oncology reports. 2012; 27:135-42.

44. Tian Y, Tian X, Han X, Chen Y, Song CY, Jiang WJ, Tian DL. ABCE1 plays an essential role in lung cancer progression and metastasis. Tumor Biology. 2016 Jan 5. [Epub ahead of print].

45. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer discovery. 2012; 2:401-4.

46. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Science signaling. 2013; 6:pl1.