

# Identification of Key Genes with Prognostic Value in Gastric Cancer by Bioinformatics Analysis

Zhiying Lin (✉ [lzy1191528410@163.com](mailto:lzy1191528410@163.com))

Jiangxi Provincial People's Hospital <https://orcid.org/0000-0002-6338-5628>

---

## Research Article

**Keywords:** gastric cancer, bioinformatics, key prognostic genes, PPI network

**Posted Date:** July 6th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-658577/v1>

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

[Read Full License](#)

---

# Abstract

**Background:** Gastric cancer (GC) is a digestive system tumor with high morbidity and mortality. It is urgently required to identify genes that are associated with GC prognosis and to elucidate the underlying molecular mechanisms. The aim of this study aims to identify the key genes which may affect the prognosis of GC patients and be a therapeutic strategy for GC patients by bioinformatic analysis.

**Methods:** The significant prognostic differentially expressed genes (DEGs) were screened out from TCGA and the GEO datasets. The protein-protein interaction (PPI) network was established by STRING and screen key gene by MCODE and CytoNCA plug-ins in Cytoscape. Function enrichment analysis, construction of prognostic risk model and nomogram verify key genes as potential therapeutic target.

**Results:** In total, 997 genes and 805 genes related to the prognosis of GC in the GSE84427 and TCGA datasets, respectively. We define the 128 genes shared by the two datasets as PRG. Then, the first 4 genes (MYLK, MYL9, LUM and CAV1) with great node importance in the PPI network of PRG were identified as key genes. Independent prognostic risk analysis found that patients with high PRG expression had poor prognosis, excluding their own age, gender and TNM stage. GO and KEGG enrichment analyses showed that key genes may exert influence through the PI3K-Akt pathway, in which, extracellular matrix organization and focal adhesion may play important roles in key genes influencing the prognosis of GC patients.

**Conclusion:** We identified that MYLK, MYL9, LUM and CAV1 are potential reliable biomarkers prognosis key genes and invasion and migration target in GC.

## Introduction

Gastric cancer (GC) is the fifth most common cancer and the third most common causation of cancer-related death in the world. The statistical results showed that there were more than 1 million new cases of GC in the world every year, and the number of GC-related death cases continuously increased, statistics for 2018 showed that the death toll has risen to 784000 [1]. Many interfering factors can cause the low survival rates of GC patients, among which, the diagnosis of GC patients usually occurs in the middle and late stages, easy recurrence and metastasis after an operation are the most common reasons for the poor prognosis of GC patients [2]. In the past ten years, a large number of studies have revealed that there were quite sensitive and effective biomarkers that can affect the occurrence and progression of GC. For example, Graziano F et al. found that methylation of CpG island in the promoter region of the CDH1 gene will lead to a change of CDH1 expression [3], which may play an important role in the occurrence and progression of diffuse GC, and CDH1 is likely to be one of the therapeutic targets of GC. Several previous studies [4] also found that HER2 expression is not only an independent risk factor affecting the prognosis of GC patients but also an effective target for the treatment of GC patients. These experiences provide the basis for the research on the occurrence, progression and treatment of GC. However, previous studies about biomarkers on the occurrence and progression of GC were based on the single-gene pattern, and

cancer is usually a disease involving multiple genes and mechanisms. Therefore, it is very important to comprehensively explain the specific mechanism of GC progression and identify significant biomarkers to improve the prognosis of GC patients.

Bioinformatics is a broad multidisciplinary field. Computational tools have been developed to analyze and manage the increasing amount of biological data [5]. Bioinformatics can be used to identify the key drivers of each specific cancer patient. Therefore, they have the potential to realize more personalized cancer treatment programs, paving the way for new targeted drugs targeting specific protein [6]. With the development of The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO), and the accumulation of cancer genetics knowledge have developed rapidly in recent ten years, tumor analysis based on database not only reveals the whole panorama of tumor-related genome changes, but also lays a foundation for the study of related tumor types [7].

In this study, bioinformatics methods and techniques were used to screening out prognostic differentially expressed genes (P-DEGs) of GC from GEO and TCGA. Further, we established a PPI network to identify the key genes in DEGs through module analysis and centrality analysis, constructed prognostic risk model and verified a unfavorable indicator. This study provides a reliable basis for exploring the molecular mechanisms of GC pathogenesis and identifying molecular targets for clinical diagnosis or treatment.

## Methods

### Data

The gene expression matrix data of GC patients were obtained from the data set (GSE84437) in the GEO database of the national bioinformatics center of the United States. The data set was composed of the gene chip expression profile data and the survival information of 433 GC patients, which were collected through the GPL6947 chip platform. Besides, 380 cases of GC tissue expression profile data and clinical information were downloaded from the TCGA database.

### Screening of prognosis related genes of GC patients

The gene expression matrix of GC tissues were obtained from the GEO database (n = 433) and TCGA (n = 380) respectively, and then the data were mined through R software. To obtain the standardized gene expression matrix of GC patients, the "impute" and "limma" packages in R were used to process the missing value estimation and logarithmic transformation of data. According to the K-M method, each gene in the gene expression matrix was divided into high or low expression group based on the median value of the gene expression. Subsequently, the survival difference between these two groups was evaluated and verified. The proportional hazards model was used for multivariate analyses and survival estimation to analyze, verify, filter and screen out these genes which were significantly correlated with the prognosis of GC patients ( $P < 0.05$ ). Finally, these genes filtered by the above survival analyses were

mutually verified in the two datasets obtained by GEO and TCGA. Then the common significant prognostic differentially expressed genes were identified as P-DEGs.

## **PPI network**

STRING <http://string-db.org> is an online tool, which is often used to predict protein-protein interactions [8]. Through STRING, the gene interaction analysis can be conducted, including physical and functional interactions. In this study, we used it to establish a PPI network of P-DEGs while the confidence score of connections in this network is required to be > 0.15 and the disconnected nodes in the network were excluded.

## **Module analysis and centrality analysis of PPI network**

The PPI network of P-DEGs is visualized by Cytoscape software. In this network, the functional modules and the interactions between genes were identified and measured through the MCODE plug-in [9]. The selection parameters in MCODE plug-in were as follows: find clusters = in whole network, degree cut-off = 2, clustering finding=haircut, Node score cut-off= 0.2, K-core= 2, max depth= 100. In all sub-modules, the higher the score was, the stronger the protein correlation in the sub-module was, and the sub-module with the highest score was considered as the result of MCODE analysis. The plug-in CytoNCA is used for centrality analysis, including three parameters: degree, betweenness, and eigenvector [10]. Degree is a measurement of the importance of a single node, which describes the number of sides of a connected node [11]. Betweenness is the shortest path to analyze a specific node [12]. While for the eigenvector, the importance of the node itself and its neighbors is considered [13]. The top 5% of the nodes under each parameter are considered as the important nodes of CytoNCA analysis, the genes which represented were considered as the result of centrality analysis. Finally, by combining the results of MCODE and CytoNCA plug-ins, the common genes were considered as the most important genes in the PPI network of P-DEGs and identified as key genes.

## **Prognostic analysis and validation of key genes**

Based on the gene expression matrix data of GC patients in GEO and TCGA databases, the median of key genes expression value was set as the cut-off value, and the key genes expression matrix of GC patients was divided into key genes high and low expression groups. By using the "survival" package in R, according to the K-M analysis and multivariate Cox regression test, the difference of overall survival events between the high and low expression groups of key genes was compared. Then, the survival rate and survival curve were analyzed and drawn. By using the "survival" package, according to univariate and multivariate Cox regression analysis, the hazard ratio (HR) and forest maps of independent prognostic analysis of single-gene and multiple-genes combination of key genes were analyzed and drawn. Finally, to precisely predict the survival rates of GC patients, the risk score of key genes and some clinic-pathological factors, such as age, gender and pathological stage, etc, were linked together. According to the risk ratio weighted key genes expression data, the key genes [14] was constructed as follows:

$$risk\ score = \sum_i^N (exp_i * HR_i)$$

Where N is the number of selected genes of key genes,  $exp_i$  is the expression value of each single-gene of key genes, and  $HR_i$  is the HR value of each single of multivariate Cox regression model. According to the median value of the risk score of key genes in the expression matrix of GC patients, GC patients were divided into the low-risk group and high-risk group, and the prognostic risk rates were measured by K-M analysis. Subsequently, based on the multivariate Cox regression analysis, the nomogram is established and drawn through the "RMS" package in R language.

### GO and KEGG analysis of key genes

According to the median value of each key gene, GC patients in GEO and TCGA databases were divided into high and low expression groups of each gene respectively. The differentially expressed genes sets (DEGs) between high and low expression groups of each key genes were identified and the corresponding GO and KEGG functional enrichment analysis of each DEGS conducted through "limma", "cluster profile", "org.Hs.eg.db", "enrichplot", "ggplot2" and other R software packages.  $|\log_2FC| > 0.5$ ,  $P < 0.05$  and adjusted  $P < 0.05$  were considered as the cut-off criteria.

### Statistical analysis

R language (version 4.0.1) was used for data statistical analysis: K-M analysis, univariate Cox regression analysis and multivariate Cox regression analysis were used to identify the key genes; Survival curve and forest maps of survival analysis and independent prognostic analysis of single-gene or multiple-genes combination of key genes were drawn with R language through the "survival" package.  $P < 0.05$  and adjusted  $P < 0.05$  were considered as the cut-off criteria.

## Result

### Identification of P-DEGs

To explore the key genes affecting the prognosis of GC patients and the roles of these genes play in the mechanism of GC progression. The gene expression matrix data obtained from GEO and TCGA databases were used to conducted multivariate analyses and survival estimation to screen out the genes that were significantly correlated with the prognosis of GC patients ( $P < 0.05$ ). Subsequently, we obtained 997 and 805 genes related to the prognosis of GC in the GSE84427 and TCGA gene expression matrix data sets, respectively. Therefore 128 common P-DEGs were obtained by mutual validation between the two datasets, which means 128 of 997 genes in GSE84427 and 128 of 805 genes in TCGA databases (Fig. 1A).

### Module analysis and centrality analysis of P-DEGs related PPI network

In order to study the molecular mechanism which can affect the prognosis of GC patients from a systematic perspective, we established a PPI network of P-DEGs to explore the molecular mechanism. The results showed that there were 124 nodes and 819 edges in the PPI network. Furthermore, we used the MCODE plug-in in Cytoscape software to analyze the modules existing for exploring more closely related genes in the PPI network. The results showed that there were four modules and 1 non-module in the PPI network, and scores of the four modules were as follows: 8.667 (module 1), 7.455 (module 2), 4.111 (module 3), and 2.667 (module 4), respectively. We found that the first module (module 1) was the most interactive area in the PPI network, which is located in the center of the whole network, including 16 nodes and 65 edges (Fig. 1B). Therefore, the protein interactions in module 1, which ranks the first, maybe the strongest, and most important part of the whole network. Module 1 was considered as the final result of the MCODE analysis. At the same time, to obtain GC prognosis related key genes in this complex PPI network, we use the centrality analysis method to analyze the PPI network. First, we used the CytoNCA plug-in to analyze the score of three parameters of each gene in the PPI network, which were degree, betweenness, and eigenvector. Then, we selected the genes with its score ranked top 5% in three parameters. Finally, we selected these genes which ranked top 5% in three parameters and showed in module 1 as key genes, which were MYLK, MYL9, LUM and CAV1, and they were all in module 1 with high centrality (Fig. 1C).

## Prognostic Value Of Key Genes In Gc Patients

To analysis the role of key genes in the progression of GC, the survival analyses of four genes of key genes were further analyzed through the K-M method. According to the median expression of the genes matrix, GC patients were divided into the high expression group and low expression group. The survival curve showed that the expression of MYLK, MYL9, LUM and CAV1 were significantly correlated with the 5-year survival rate and overall survival time of GC patients in GEO and TCGA databases ( $P < 0.05$ ). According to the survival analyses, the median survival time of GC patients with lower expression of MYLK, MYL9, LUM and CAV1 were 1.37, 1.41, 1.35 and 1.42 years, with higher expression of MYLK, MYL9, LUM and CAV1 were 1.06, 1.08, 1.15 and 1.06 years in TCGA database, respectively. Compared with GC patients with lower expression of MYLK, MYL9, LUM and CAV1 (GEO,  $n = 217$ ; TCGA,  $n = 190$ ), these patients with high expression of key genes ( $n = 216$ , GEO;  $n = 190$ , TCGA) had significantly poorer prognosis ( $P < 0.05$ , Fig. 2). The results were verified through GEO gene matrix once again. According to the univariate and multivariate Cox regression analyses, the results of independent prognosis of key genes in the GEO and TCGA databases showed that the HR of MYLK, MYL9, LUM and CAV1 all presented as  $HR > 1$ , which were 1.15, 1.18, 1.19 and 1.31, respectively ( $P < 0.05$ ). These results indicated that key genes can independently affect the prognosis of GC patients (Fig. 3). The influence of key genes is of great significance and has potential value as prognostic biomarkers and therapeutic targets for GC patients.

## Go And Kegg Enrichment Analyses

To better elucidate the mechanisms of key genes affecting GC prognosis, we performed GO and KEGG enrichment analyses. Results of GO analyses showed that most GO-terms were significantly enriched in the extracellular matrix organization, extracellular structure organization, cell-substrate adhesion, tissue migration, muscle contraction, muscle tissue development, mesenchymal development, etc. (Fig. 4). Besides, the results of KEGG analyses showed that the related pathways were significantly enriched in focal adhesion, PI3K-Akt signaling pathway, ECM receptor interaction, cell adhesion molecules, proteoglycans in cancer, protein digestion and absorption, Cell cycle, calcium signaling pathway, etc (Fig. 5). These results indicated that key genes affect the prognosis of GC patients mainly through influencing the invasion, migration and cell cycle functions of GC cells.

### **Construction and validation of prognostic risk model of key genes**

Based on multivariate Cox regression analysis, key genes (MYLK, MYL9, LUM and CAV1) were integrated, and prognostic risk model of key genes was established according to GEO and TCGA data respectively. The risk score of key genes was calculated using the formula mentioned in the method, processes were as follows: risk score =  $(HR_{(MYLK)} \times MYLK \text{ expression level}) + (HR_{(MYL9)} \times MYL9 \text{ expression level}) + (HR_{(LUM)} \times LUM \text{ expression level})$ , risk score and clinical features of GC patients can be estimated based on the total points. The results showed that the risk score is very high in both GEO and TCGA databases (risk score is about 10 in GEO and 9 in TCGA, Fig. 4), which confirmed that key genes may be potential therapeutic targets for prognosis.

To verify the reliability of key genes, GC patients were divided into the low-risk group and high-risk group according to the median risk score in TCGA and GEO databases, respectively. The survival curves showed that the prognosis of the high-risk group was worse than that of the low-risk group (Fig. 5,  $P < 0.05$ ). With the risk score increasing, the number of patients' death increases (Fig. 5). Univariate and multivariate Cox regression analyses were performed based on the gene matrix data, the results of which showed that the risk score of key genes was independently correlated with the overall survival rate of GC patients (Table 1,  $P < 0.05$ ). These results indicated that the key genes can be a significant reference to the prognosis of GC patients. The key genes can be used to guide the next step of treatment after surgery or/and chemoradiotherapy treatment. MYLK, MYL9, LUM and CAV1 can be potential targets to improve the prognosis of GC patients.

## **Discussion**

GC is one of the most common and malignant tumors. Although the main treatment methods of GC such as surgery, radiotherapy, and chemotherapy have made progress, the incidence rate and mortality rate of GC patients remains stubbornly high [15, 16]. More than 90% of the GC patients were in the late stage when diagnosed, which was related to the unclear symptoms in early-stage of GC patients and unclear influential factors of GC prognosis to a large extent [17, 18]. The occurrence and progression of GC is a multi-stage, slow-moving pathological process, in which genetic mutations, epigenetic changes, and abnormal molecular signal transduction pathways all can participate in the occurrence, diffusion, and

metastasis of GC [19]. Therefore, it is very important to find specific prognostic biomarkers of GC to develop therapeutic strategies for malignant behaviors of tumors. These problems highlight the necessity of finding prognostic markers of GC. Nowadays, high-throughput platforms for detecting gene expression have been developed rapidly in the processes of disease progression, which lays the foundation for the discovery of new targets that can be used to predict, diagnose and treat cancer. In this study, we integrated GEO and TCGA databases, using bioinformatics analysis methods, to mine and analyze high-throughput data to conduct module and centrality analysis of the PPI network. Which helped us screen out key genes that has an important impact on the prognosis of GC patients and can be considered as a biomarker and potential therapeutic target for GC prognosis. Then the establishment a prognostic risk model of key genes further explained the kernel roles the key genes may play in the development of GC.

Module analysis (MCODE) and centrality analysis (CytoNCA) in PPI network play important roles in screening molecular markers, these genes appear in the modules with the highest scores, and also rank higher in centrality analysis results are the key genes that can affect the occurrence of diseases [10]. Studies have shown that module analysis can help to screen key genes in cancers more accurately such as cervical cancer [20], glioblastoma [21], and head and neck squamous cell carcinoma [22]. While CytoNCA can analyze the centrality degree of each node in the whole PPI network and can exhibit the nodes with important connections, to help select key genes [23]. Combined with these two methods, key genes (MYLK, MYL9, LUM and CAV1) with important value in the whole PPI network was obtained. Some studies also elucidated the impact of key genes on various tumors.

Liang X et al., and Luo Z et al., indicated that Caveolin 1 (CAV1) played an important role in the occurrence and progression of varieties of malignant tumors, especially in the malignant progression of GC by promoting epithelial-mesenchymal transition (EMT) function. Under the conditions of extracellular matrix integrin interaction and Tyr-14 phosphorylation, CAV1 enhanced melanoma cells will migrate, invade, and migrate to the lung [24, 25]. The results of Jin g-h et al., showed that compared with normal gastric mucosa, Myosin Light Chain 9 (MYL9) was abnormally up-regulated in GC patients' tumor tissues, and it could affect the prognosis of GC patients through adhesion plaque and leukocyte cross endothelial migration [26]. As an important part of extracellular matrix, lumbican (LUM) can be expressed in many organs and tissues of human body. Its abnormal expression can affect the formation of extracellular matrix. The process of tumor proliferation, invasion and migration is often accompanied by the synthesis and degradation of extracellular matrix. Therefore, LUM can play an important role in tumor metastasis and invasion through extracellular matrix [27]. Myosin Light Chain Kinase (MYLK) can catalyze the phosphorylation of myosin light chain and regulates the invasion and metastasis of some malignant tumors [28, 29].

In the past few years, there has been more and more evidence of the key role of the extracellular matrix in mediating different cell processes (including cell adhesion, polarity, migration, differentiation, proliferation, and survival), and tumor cells are closely related to it [30]. Focal adhesion is a strong adhesion of the sub-cellular structure to the extracellular matrix. It also acts as a scaffold for many signal transduction pathways involving integral protein or mechanical force exerted on cells [31]. Focal

adhesion dysfunction is considered to be an essential pathway in tumor invasion and migration [32, 33]. Many cellular processes in cancer are attributed to kinase signaling networks. Akt, as a serine/threonine kinase, also known as protein kinase B, is a carcinogenic protein that can regulate cell survival, proliferation, growth, apoptosis, and glycogen metabolism. Over-expression of Akt is a common molecular feature of human malignant tumors. Many tumor tissues and tumor cells are accompanied by Activation of PI3K/Akt signaling pathway [34]. In this study, we explored the relationship between key genes and classical carcinogenic signaling pathways by GO and KEGG enrichment analysis. Results showed that key genes can promote the development of GC by regulating various signaling pathways, many of which have been proved to play important roles in the occurrence and progression of cancer. In particular, focal adhesion and PI3K/Akt signaling pathways may be the main signaling pathways involved in the effect of key genes on GC prognosis, and their influences could not be divorced from the extracellular matrix. Therefore, in these GC patients whose key genes expression significantly increased, we can try to further study the specific treatment involving key genes.

Over this study, we not only used the integrated multi-step analysis (including network data mining and enrichment analysis) to find the key genes that have an important impact on the prognosis of GC patients, namely key genes (MYLK, MYL9, LUM, CAV1), but also established a risk model with predictive value based on key genes, which provides a basis for exploring the molecular mechanism affecting the prognosis of GC patients.

## Conclusion

The integrative analyses of gene expression matrix identified 128 common P-DEGs. The 4 key genes (MYLK, MYL9, LUM, CAV1) of P-DEGs may be predictive biomarkers or therapeutic targets for GC prognosis. These predictions should be verified through experiments validation, though this study provided new insights into the development of individualized therapeutic targets for GC.

## Abbreviations

GC

Gastric cancer

P-DEGs

Prognostic differentially expressed genes

PPI

Protein-protein interaction

TCGA

The Cancer Genome Atlas

GEO

Gene Expression Omnibus

HR

Hazard ratio

# **Declarations**

## **Acknowledgements**

Not applicable.

## **Funding**

Not applicable.

## **Availability of data and materials**

Not applicable.

## **Abbreviations**

GC: Gastric cancer

P-DEGs : Prognostic differentially expressed genes

PPI: Protein-protein interaction

TCGA: The Cancer Genome Atlas

GEO: Gene Expression Omnibus

HR: Hazard ratio

## **Authors' contributions**

ZYL designed this paper and performed critical revision of the manuscript.

## **Notes**

### **Ethics approval and consent to participate**

This study was conducted in accordance with the Declaration of Helsinki. This study was conducted with approval from the Ethics Committee of Jiangxi Provincial People's Hospital. Written informed consent was obtained from all participants.

### **Consent for publication**

Not applicable.

### **Competing interests**

The authors declare that they have no competing interests.

## References

1. Smyth EC, Nilsson M, Grabsch HI, van Grieken NCT, Lordick F. Gastric cancer. *The Lancet*. 2020;396 10251:635 – 48; doi: 10.1016/s0140-6736(20)31288-5.
2. Fang H, Li R, Gu Y, Fei Y, Jin K, Chen Y, et al. Intratumoral interleukin-9 delineates a distinct immunogenic class of gastric cancer patients with better prognosis and adjuvant chemotherapeutic response. *Oncoimmunology*. 2020;9 1:1856468; doi:10.1080/2162402X.2020.1856468.
3. Graziano F, Arduini F, Ruzzo A, Mandolesi A, Bearzi I, Silva R, et al. Combined analysis of E-cadherin gene (CDH1) promoter hypermethylation and E-cadherin protein expression in patients with gastric cancer: implications for treatment with demethylating drugs. *Ann Oncol*. 2004;15 3:489–92. doi:10.1093/annonc/mdh108.
4. Digkila A, Wagner AD. Advanced gastric cancer: Current treatment landscape and future perspectives. *World J Gastroenterol*. 2016;22 8:2403–14. doi:10.3748/wjg.v22.i8.2403.
5. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, et al. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res*. 2010;38 Web Server issue:W695–9; doi:10.1093/nar/gkq313.
6. Zhang J, Yang PL, Gray NS. Targeting cancer with small molecule kinase inhibitors. *Nat Rev Cancer*. 2009;9 1:28–39. doi:10.1038/nrc2559.
7. Cancer Genome Atlas. Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*. 2013;45 10:1113–20; doi:10.1038/ng.2764.
8. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguéz P, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic acids research*. 2011;39 Database issue:D561-D8; doi: 10.1093/nar/gkq973.
9. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*. 2003;4:2. doi:10.1186/1471-2105-4-2.
10. Tang Y, Li M, Wang J, Pan Y, Wu FX. CytoNCA: a cytoscape plugin for centrality analysis and evaluation of protein interaction networks. *Biosystems*. 2015;127:67–72. doi:10.1016/j.biosystems.2014.11.005.
11. Luo X, Guo L, Dai XJ, Wang Q, Zhu W, Miao X, et al. Abnormal intrinsic functional hubs in alcohol dependence: evidence from a voxelwise degree centrality analysis. *Neuropsychiatr Dis Treat*. 2017;13:2011–20. doi:10.2147/NDT.S142742.
12. Li Y, Li W, Tan Y, Liu F, Cao Y, Lee KY. Hierarchical Decomposition for Betweenness Centrality Measure of Complex Networks. *Sci Rep*. 2017;7:46491. doi:10.1038/srep46491.
13. Negre CFA, Morzan UN, Hendrickson HP, Pal R, Lisi GP, Loria JP, et al. Eigenvector centrality for characterization of protein allosteric pathways. *Proc Natl Acad Sci U S A*. 2018;115 52:E12201-E8. doi:10.1073/pnas.1810452115.

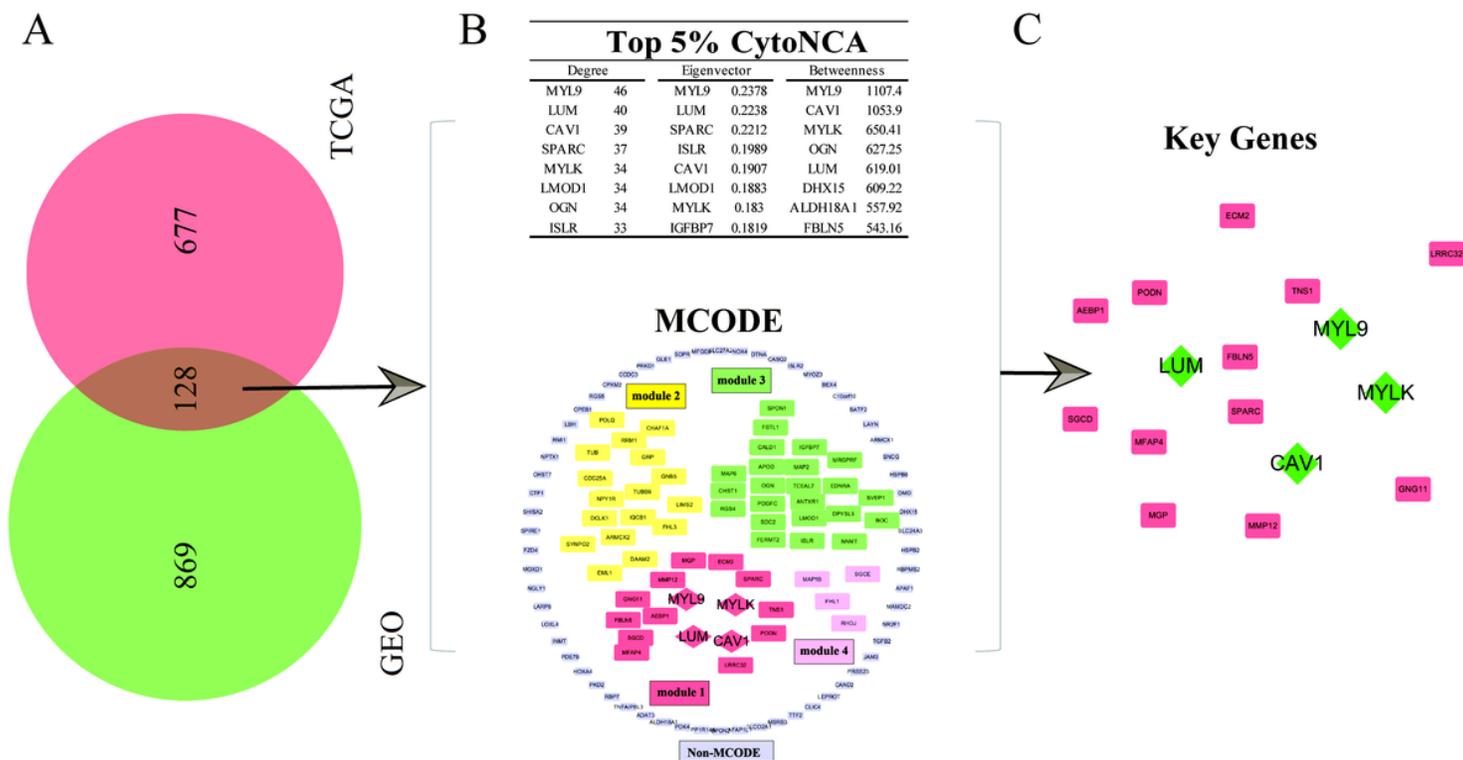
14. Lin T, Fu Y, Zhang X, Gu J, Ma X, Miao R, et al. A seven-long noncoding RNA signature predicts overall survival for patients with early stage non-small cell lung cancer. *Aging*. 2018;10 9:2356–66. doi:10.18632/aging.101550.
15. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer*. 2015;136(5):E359-E86. doi:10.1002/ijc.29210.
16. Li Y, Guo X-B, Wang JS, Wang H-C, Li L-P. Function of fibroblast growth factor 2 in gastric cancer occurrence and prognosis. *Mol Med Rep*. 2020;21(2):575–82. doi:10.3892/mmr.2019.10850.
17. Yan HHN, Siu HC, Law S, Ho SL, Yue SSK, Tsui WY, et al. A Comprehensive Human Gastric Cancer Organoid Biobank Captures Tumor Subtype Heterogeneity and Enables Therapeutic Screening. *Cell stem cell*. 2018;23 6; doi:10.1016/j.stem.2018.09.016.
18. Huang Y-K, Wang M, Sun Y, Di Costanzo N, Mitchell C, Achuthan A, et al. Macrophage spatial heterogeneity in gastric cancer defined by multiplex immunohistochemistry. *Nature communications*. 2019;10 1:3928; doi:10.1038/s41467-019-11788-4.
19. Shan C, Zhang Y, Hao X, Gao J, Chen X, Wang K. Biogenesis, functions and clinical significance of circRNAs in gastric cancer. *Mol Cancer*. 2019;18 1:136; doi:10.1186/s12943-019-1069-0.
20. Xia L, Su X, Shen J, Meng Q, Yan J, Zhang C, et al. functions as a key candidate gene in cervical cancer as determined by integrated bioinformatic analysis. *Cancer management research*. 2018;10:663–70. doi:10.2147/CMAR.S162813.
21. Yang Q, Wang R, Wei B, Peng C, Wang L, Hu G, et al. Candidate Biomarkers and Molecular Mechanism Investigation for Glioblastoma Multiforme Utilizing WGCNA. *BioMed research international*. 2018;2018:4246703. doi:10.1155/2018/4246703.
22. Yang B, Chen Z, Huang Y, Han G, Li W. Identification of potential biomarkers and analysis of prognostic values in head and neck squamous cell carcinoma by bioinformatics analysis. *OncoTargets therapy*. 2017;10:2315–21. doi:10.2147/OTT.S135514.
23. Lu Y, Li A, Lai X, Jiang J, Zhang L, Zhong Z, et al. Identification of differentially expressed genes and signaling pathways using bioinformatics in interstitial lung disease due to tyrosine kinase inhibitors targeting the epidermal growth factor receptor. *Invest New Drugs*. 2019;37 2:384–400. doi:10.1007/s10637-018-0664-z.
24. Liang X, Chen W, Shi H, Gu X, Li Y, Qi Y, et al. PTBP3 contributes to the metastasis of gastric cancer by mediating CAV1 alternative splicing. *Cell death disease*. 2018;9 5:569. doi:10.1038/s41419-018-0608-8.
25. Luo Z, Rong Z, Zhang J, Zhu Z, Yu Z, Li T, et al. Circular RNA circCCDC9 acts as a miR-6792-3p sponge to suppress the progression of gastric cancer through regulating CAV1 expression. *Mol Cancer*. 2020;19 1:86; doi:10.1186/s12943-020-01203-8.
26. Jin G-H, Xu W, Shi Y, Wang L-B. Celecoxib exhibits an anti-gastric cancer effect by targeting focal adhesion and leukocyte transendothelial migration-associated genes. *Oncology letters*. 2016;12 4:2345–50.

27. Chen X, Li X, Hu X, Jiang F, Shen Y, Xu R, et al. LUM Expression and Its Prognostic Significance in Gastric Cancer. *Frontiers in oncology*. 2020;10:605. doi:10.3389/fonc.2020.00605.
28. Tan X, Chen M. MYLK and MYL9 expression in non-small cell lung cancer identified by bioinformatics analysis of public expression data. *Tumor Biology*. 2014;35 12:12189–200. doi:10.1007/s13277-014-2527-3.
29. Lin J, He Y, Chen L, Chen X, Zang S, Lin W. MYLK promotes hepatocellular carcinoma progression through regulating cytoskeleton to enhance epithelial–mesenchymal transition. *Clinical Experimental Medicine*. 2018;18 4:523–33. doi:10.1007/s10238-018-0509-2.
30. Moreira AM, Pereira J, Melo S, Fernandes MS, Carneiro P, Seruca R, et al. The Extracellular Matrix: An Accomplice in Gastric Cancer Development and Progression. *Cells*. 2020;9 2; doi:10.3390/cells9020394.
31. Burridge K. Focal adhesions: a personal perspective on a half century of progress. *FEBS J*. 2017;284 20:3355–61. doi:10.1111/febs.14195.
32. Carragher NO, Frame MC. Focal adhesion and actin dynamics: a place where kinases and proteases meet to promote invasion. *Trends in cell biology*. 2004;14 5:241–9.
33. Paluch EK, Aspalter IM, Sixt M. Focal Adhesion-Independent Cell Migration. *Annu Rev Cell Dev Biol*. 2016;32:469–90.
34. Song M, Bode AM, Dong Z, Lee M-H. AKT as a Therapeutic Target for Cancer. *Cancer research*. 2019;79 6:1019–31. doi:10.1158/0008-5472.CAN-18-2738.

## Tables

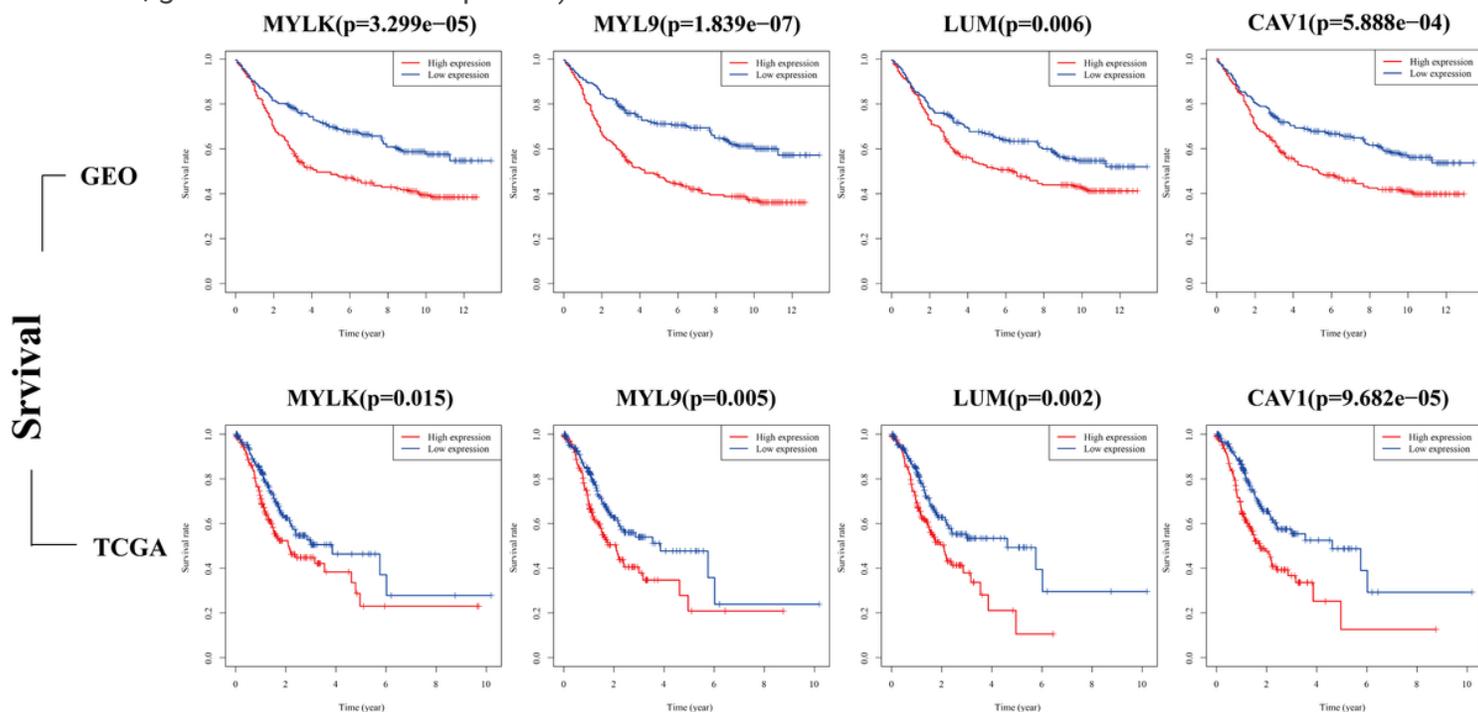
Due to technical limitations, table 1 is only available as a download in the Supplemental Files section.

## Figures



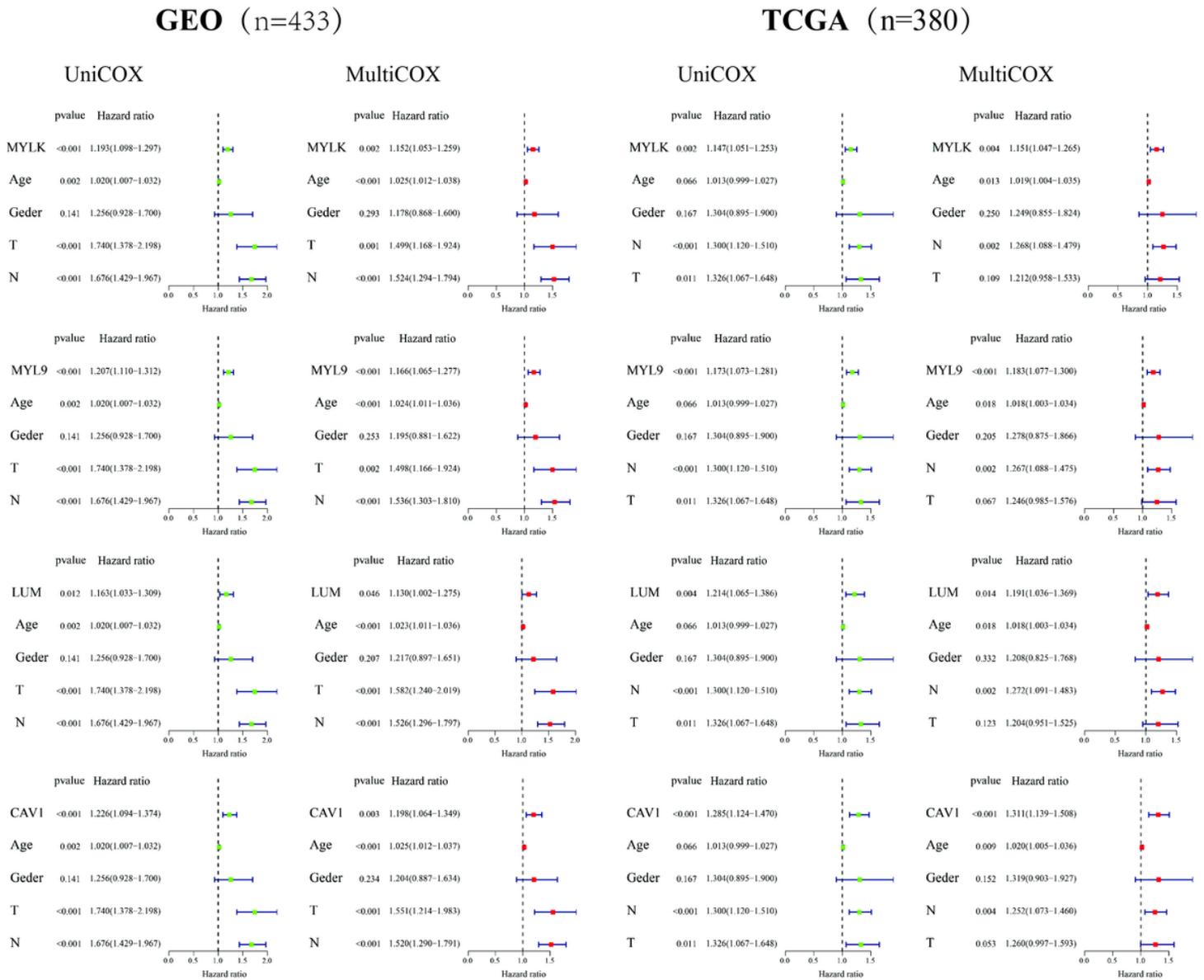
**Figure 1**

Selection of key genes for GC patients. A: 128 common P-DEGs were obtained from the intersection of TCGA and GEO datasets. B: Four modules (namely module 1-4 and one non-MCODE module) and score ranked top 5% in three parameters from CytoNCA centrality analysis. C: Key genes (MYLK, MYL9, LUM and CAV1, green diamond in the picture) was obtained.



**Figure 2**

Survival analyses of key gene (MYLK, MYL9, LUM and CAV1) Patients with high expression of key genes have poor prognosis ( $P < 0.05$ ). The median survival time of GC patients with lower expression groups of MYLK, MYL9, LUM and CAV1 were 6.58, 7.00, 6.42 and 6.75 years, with higher expression groups of MYLK, MYL9, LUM and CAV1 were 3.54, 3.58, 4.25 and 4.08 years in GEO database, while the median survival time of lower expression groups of MYLK, MYL9, LUM and CAV1 were 1.37, 1.41, 1.35 and 1.42 years, with higher expression groups of MYLK, MYL9, LUM and CAV1 were 1.06, 1.08, 1.15 and 1.06 years in TCGA database, respectively.



**Figure 3**

Independent prognostic analysis of key genes

# GeneOntology

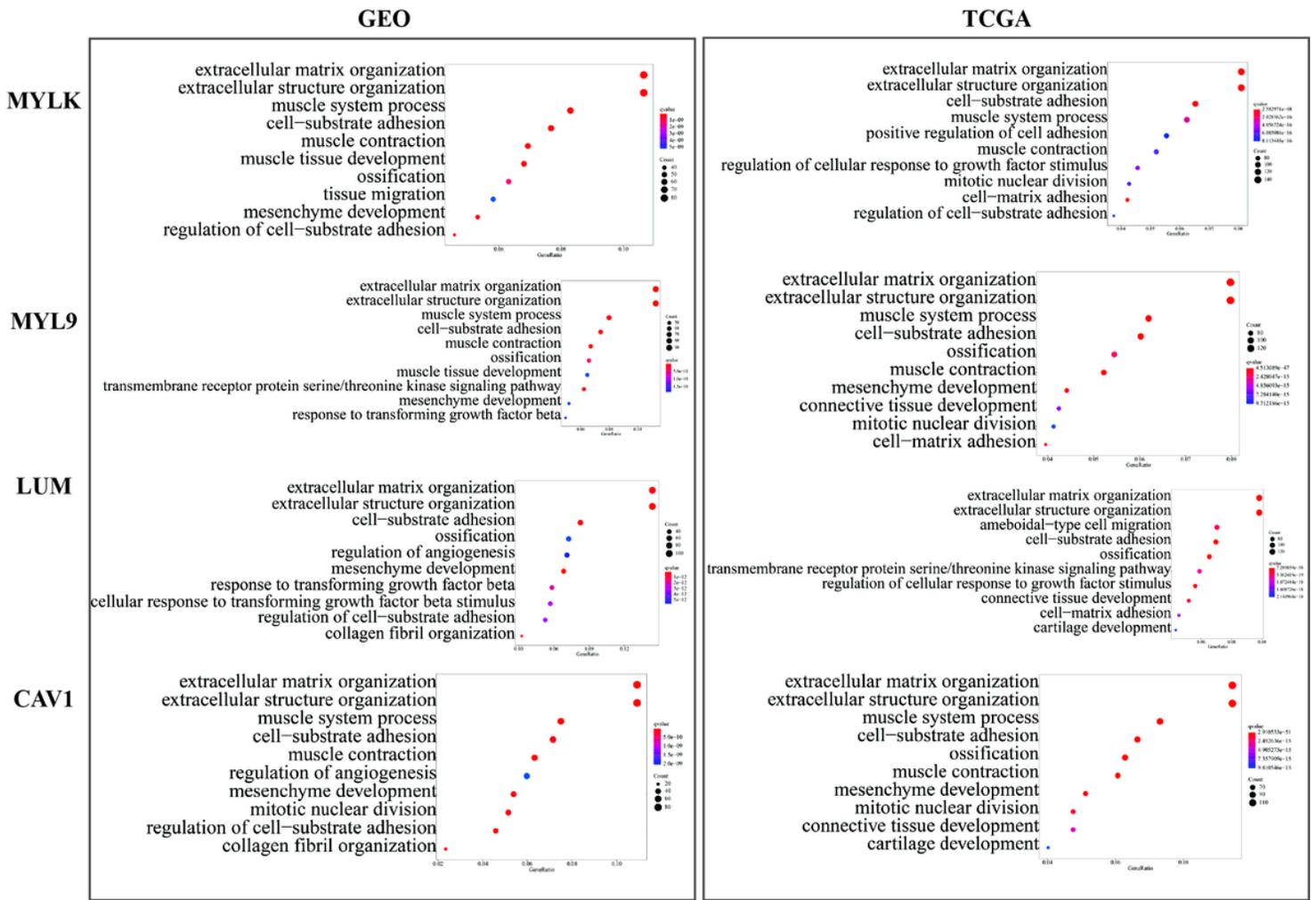


Figure 4

GO enrichment analyses of key genes

# Kyoto Encyclopedia of Genes and Genomes

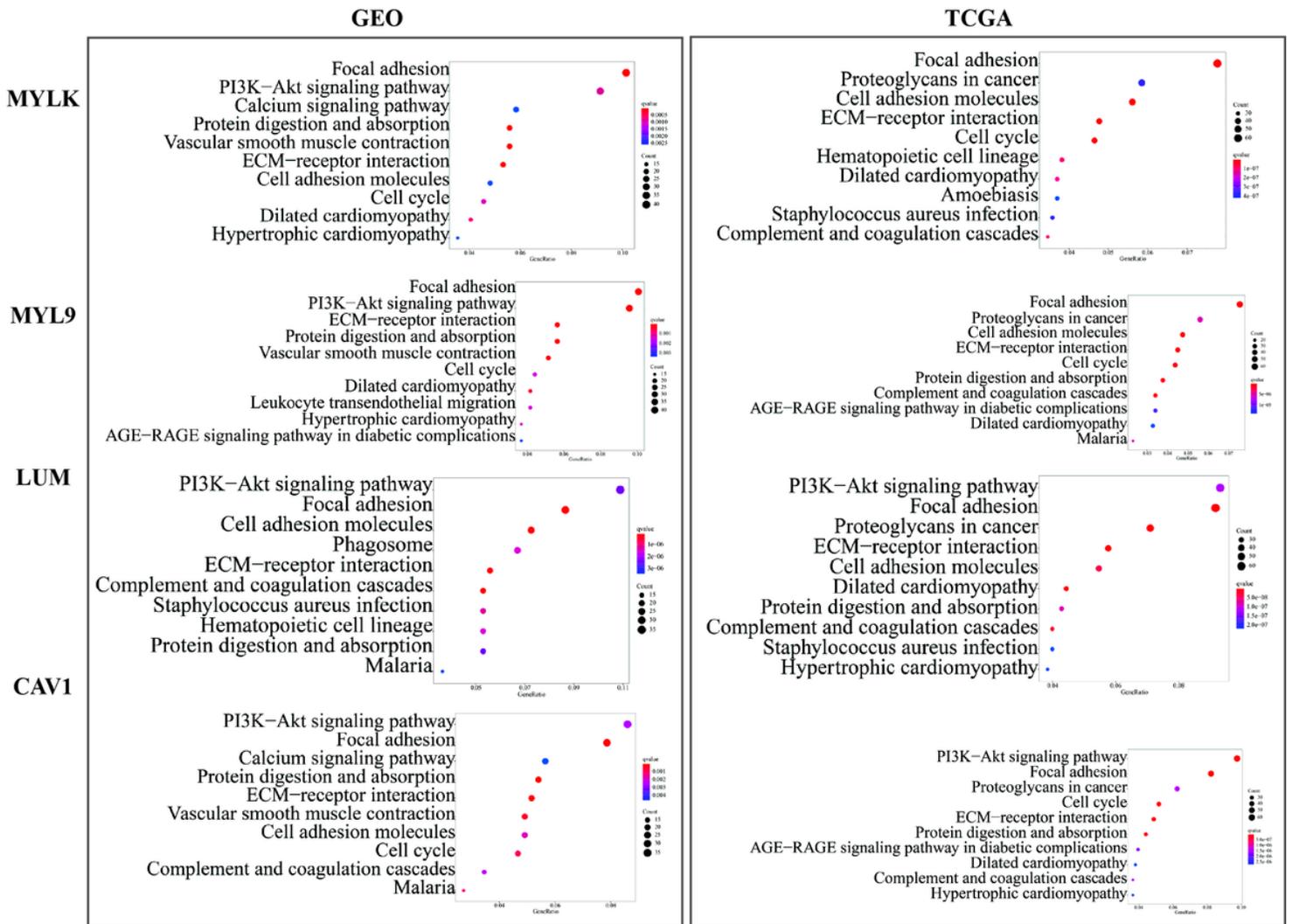
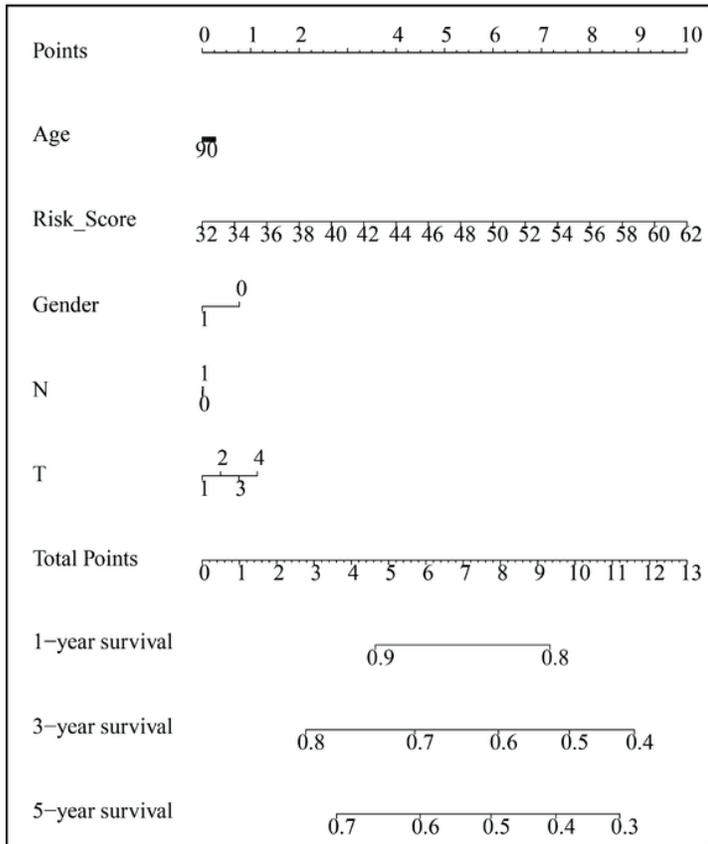


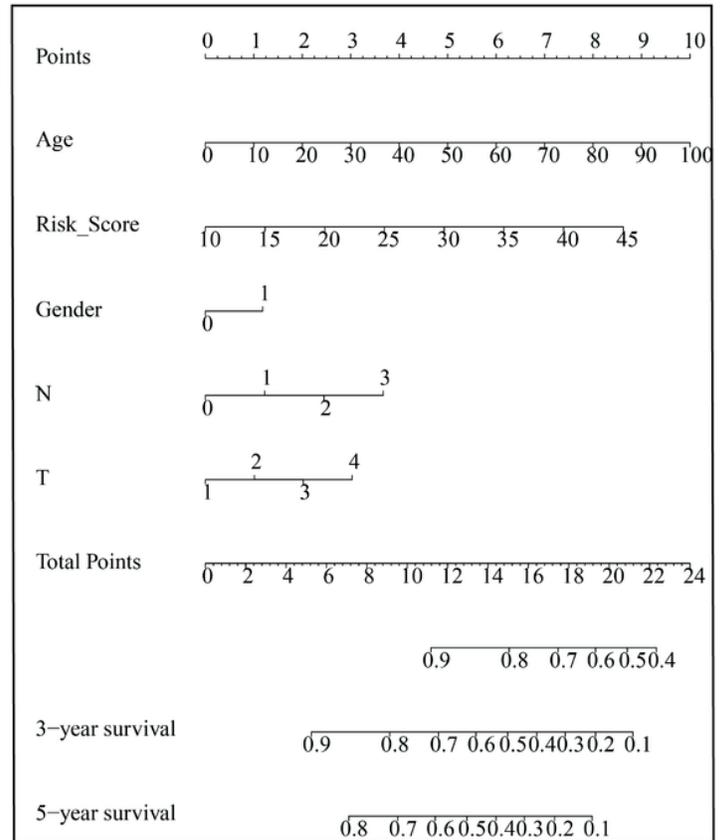
Figure 5

KEGG enrichment analyses of key genes

## GEO



## TCGA

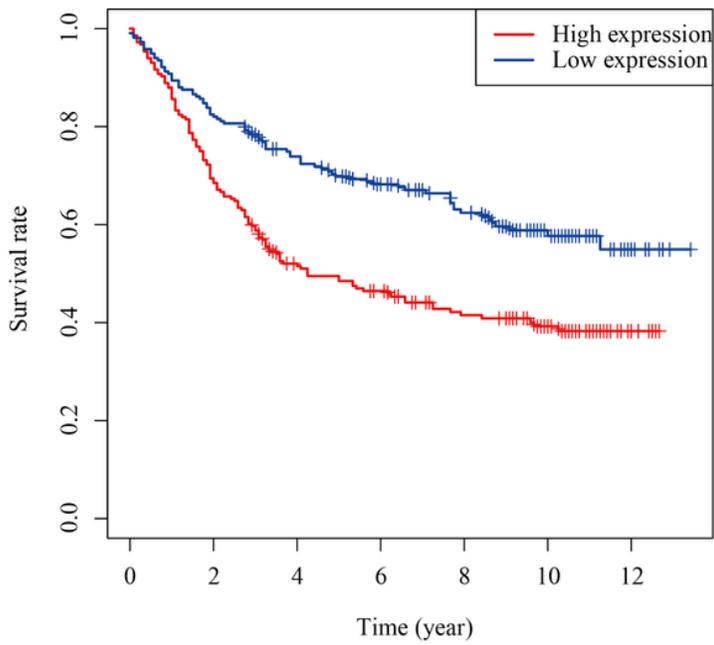


**Figure 6**

Nomogram Through the construction of a prognostic risk model, the risk score of key genes was calculated. Each influencing factor was assigned a score based on their regression coefficient in the multivariate Cox proportional hazards model. The score of each factor can be read according to the points scale in the figure (first line) and the total score obtained by adding score of influencing factors.

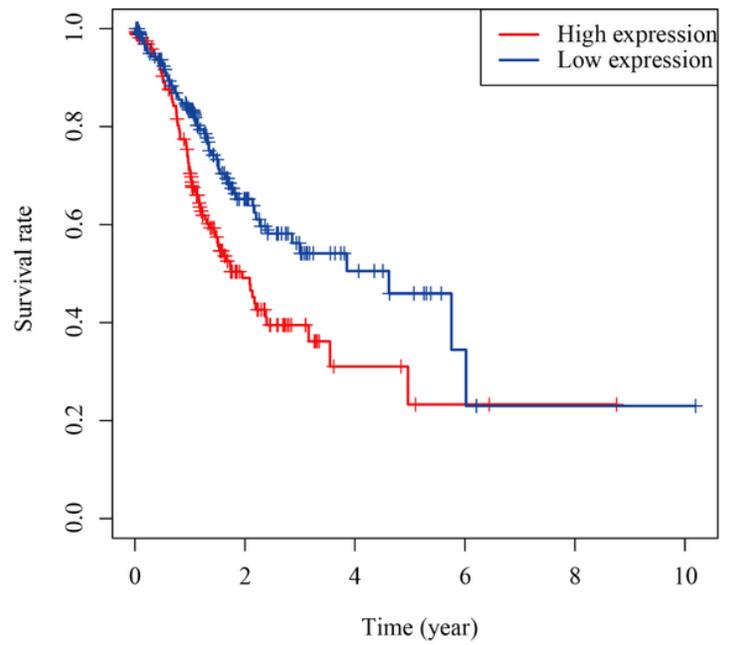
## GEO

Risk\_Score( $p=1.596e-05$ )



## TCGA

Risk\_Score( $p=0.003$ )



**Figure 7**

Survival analyses of risk score of key genes in GEO and TCGA

## Supplementary Files

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

- [Table1.xlsx](#)