

Identification of the Differentially Expressed Genes and Prognostic Associations in Single-Cell RNA Sequencing Data of Gastric Cancer

Shuang Gao

Fifth Affiliated Hospital of Zhengzhou University

Fazhan Li

Zhengzhou University School of medical sciences

Minghai Zhao

Henan Cancer Hospital

Wanqing Wu

Fifth Affiliated Hospital of Zhengzhou University

Yuming Fu

Fifth Affiliated Hospital of Zhengzhou University

Youcai Tang

Fifth Affiliated Hospital of Zhengzhou University

Pengyuan Zheng

Fifth Affiliated Hospital of Zhengzhou University

Yang Mi (✉ miyang198@126.com)

Zhengzhou University

Research

Keywords: gastric cancer, single-cell RNA sequencing, differentially expressed genes, prognosis

Posted Date: January 21st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-150872/v1>

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

[Read Full License](#)

Abstract

Background: Due to the lack of effective drugs, gastric cancer(GC) has a high mortality rate among other cancers, with a low 5-year survival rate and an inferior prognosis. Thus, screening of meaningful tumor biomarkers or therapeutic targets could play a vital role in the diagnosis, treatment, prognosis, and follow-up of GC.

Methods: Gene expression profiles and comprehensive clinical information of 407 patients with GC were downloaded from The Cancer Genome Atlas (TCGA) database. GC-related single-cell RNA sequencing data from the GSE118916 dataset was downloaded from the Gene Expression Omnibus (GEO) database. The differentially expressed genes (DEGs) were screened from transcriptomic data in GC and normal samples by R language. The DAVID database was also used to analyze the functions and pathways of DEGs. After combining differential genes with patient survival information, target genes were identified. The interaction of DEGs in the protein-protein interaction (PPI) network was also studied.

Results: Our study identified a total of 209 differential genes, which might be positively related to GC. Gene Ontology (GO) analysis indicated numerous enrichment of DEGs in the extracellular matrix organization, extracellular structure organization, and muscle contraction. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis showed that the DEGs were mainly enriched in focal adhesion, protein digestion and absorption, AGE-RAGE signaling pathway in diabetic complications. Further analysis showed the higher expression of Carboxypeptidase vitellogenic-like gene (CPVL) was related to the better prognosis of GC patients in both TCGA and the GEO database. FAM3 metabolism regulating signaling molecule D (FAM3D) and oxidized low-density lipoprotein receptor 1 (OLR1) were significantly associated with GC patients' prognosis only in the GEO database. Lastly, the PPI network shows the gene expression proteins that interact most closely with CPVL protein.

Conclusion: Our study revealed that CPVL gene could be a promising target for the diagnosis and treatment of GC, which has a great significance for the future research on GC. In addition, we were the first to find a close relationship between FAM3D and GC.

1. Introduction

Gastric cancer (GC) is a major health burden worldwide and the third leading cause of cancer-related death worldwide. Around 989,600 new cases and 783,000 GC related deaths were reported worldwide in 2018¹⁻³. The average 5-year survival rate of 90% of GC patients with adenocarcinomas is less than 20%, and GC patients have a poor prognosis⁴. Primary prevention is still a challenge⁵ because GC is mostly asymptomatic before it enters the advanced stage. Early detection with effective screening methods is crucial to reduce the mortality rate of GC⁶.

Biomarkers based detection is a relatively non-invasive and convenient diagnostic method, which has been widely used in clinical practice. However, the sensitivity and specificity of existing GC biomarkers are

insufficient⁷⁻⁹. Currently, effective targeted therapies for GC are still focusing on the anti-human epidermal growth factor receptor-2 (HER2)¹⁰, and anti-vascular endothelial growth factor (VEGF) pathways¹¹⁻¹³, more deep research studies and development of new molecular targeting drugs still need to be implemented. The study of biomarkers has a great potential in oncology research and plays an essential role in diagnosis, treatment, and prognosis.

However, heterogeneity of tumor cells was rarely distinguished from traditional RNA-seq data based on GC tissue. Single-cell RNA sequencing (scRNA-seq) is a high-throughput sequencing of the transcriptome of a large number of single cells in a single sample, enabling researchers to screen out essential genes responsible for the intercellular differences in the mixed cell population¹⁴⁻¹⁹.

In the current study, scRNA-seq data were analyzed to identify differentially expressed genes (DEGs) related to GC in both conventional transcriptome data and scRNA-seq data. We used The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO) to incorporate transcriptome data and scRNA-seq data into the analyses and selected common DEGs. Then the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functions were enriched for DEGs. The correlation between differential genes and survival rate was analyzed based on the clinical data of the GC patients. Finally, protein-protein interaction (PPI) analysis was carried out for the prognostic genes. These analyses' results may provide new targets for further detailed studies on GC.

2. Materials And Methods

2.1. Microarray data

Gene expression profiles and complete clinical information of 407 patients with GC were downloaded from TCGA database. From Gene Expression Omnibus (GEO), NCBI's public genome database, and high-throughput Gene Expression data, we downloaded the information of 15 GC and 15 normal samples from the GSE118916 dataset, and also downloaded the information of 524 GC and 401 normal cells from the scRNA-seq data of GC in the GSE112302 dataset.

2.2. Data pre-processing and DEGs screening

After pre-processing, such as data standardization, we used the R computing environment of the SVA software package for batch normalization²⁰. Next, we performed the differential analysis ($|\text{Log}_2\text{FC}| > 1$, adjusted p-value < 0.05) by comparing tumor to normal samples in the R computing environment using limma package²¹. Then we did a Venn diagram to pick the genes that were different among GC and normal samples, and a total of 209 DEGs were selected.

2.3. Gene ontology and functional enrichment analysis

The Gene Ontology (GO) is composed of biological process (BP), cellular component (CC), and molecular function (MF). To find out the function of DEGs, the clusterProfile²² was used to conduct GO enrichment

analysis²³ and KEGG enrichment analysis²⁴ of 209 DEGs in the R computing environment.

2.4. Survival analysis

In GC samples from the TCGA database, survival analysis in the R environment was performed using TCGA biolinks to determine whether these 209 genes were associated with prognosis. The Kaplan–Meier survival analysis was also used to assess the relationship between genes and GC patients' prognosis in the GEO database. We used clinical information to map the survival curves of patients with high and low expression of differentially expressed genes ($p < 0.05$) to screen out genes related to survival and prognosis time.

2.5. Analysis of protein interaction network

Search Tool²⁵ was used to search for known protein interactions online. The confidence score > 0.9 . Then, we imported the protein interaction data into the Cytoscape software²⁶.

3. Results

3.1 Data download and difference analysis

The differentially expressed genes (DEGs) between tumor and normal samples were selected, and the volcanic map was drawn to show the DEGs. 209 DEGs were obtained from the TCGA database ($|\log_2\text{fold change}| > 1, P < 0.05$) (Fig. 1A). Then, 4258 DEGs were obtained from GSE118916 data set of GEO database ($|\log_2\text{fold change}| > 0.5, P < 0.05$) (Fig. 1B). 1500 DEGs were obtained from the scRNA-seq data from the GSE112302 data set, among which GIF, ATP4B, GKN1, REG3A, REG1B, PI3, PGA5, APOE, C1QB, TYROBP were the group of genes with the highest coefficient of variation between tumor and normal cells (Fig. 1C). Finally, 209 common DEGs in the three groups were selected (Fig. 1D).

3.2. Functional enrichment analysis

To further analyze the functions of the DEGs, functional enrichment analysis of GO and KEGG were used, and we found that in GO, as to biological process, the DEGs were significantly enriched in extracellular matrix organization, extracellular structure organization, muscle system process. About cellular component, DEGs significantly enriched in collagen-containing extracellular matrix, fibrillar collagen trimer, banded collagen fibril. For molecular function, DEGs significantly enriched in the extracellular matrix structural constituent, extracellular matrix structural constituent conferring tensile strength, platelet-derived growth factor binding (Fig. 2A&B). In KEGG, the DEGs were mainly enriched in focal adhesion, protein digestion and absorption, AGE-RAGE signaling pathway in diabetic complications (Fig. 2C&D).

3.3. Survival analysis

TCGA biolinks were used to draw the survival curve in the R environment, and we picked the six results with the lowest P value (Fig. 3A-F). It can be seen that the higher expression of CPVL was significantly correlated with the better prognosis of GC patients ($P < 0.05$). To further evaluate the correlation between the six genes and GC patients' prognosis, the Kaplan-Meier Plotter Database was used in the GEO Database. Then we found that CPVL, FAM3D, and OLR1 were associated with GC patients' prognosis in the GSE15459 dataset (Fig.S1).

3.4. PPI network construction

Since CPVL was significantly correlated with GC patients' prognosis in both TCGA and the GEO databases, we wanted to know the PPI network of CPVL. The DEGs are drawn on String for the PPI network, and the results are decorated in Cytoscape. Up-regulated and down-regulated genes were labeled in red and blue (Fig.4A), and genes related to CPVL were extracted (Fig.4B). It can be seen that CPVL interacts with proteins expressed by FGL2, PLBD1, APOE, CTSB, PGC, CES2, NEXN to a certain extent. These genes may play an important role in the pathogenesis of GC.

4. Discussion

At present, the etiology and molecular mechanism of GC need to be further studied, and the effect of early screening and diagnosis also needs to be improved. In recent years, bioinformatical analysis has increasingly been used to find new therapeutic targets and diagnostic markers for various cancers, which has been proven to be a vital step for the early prevention and treatment of cancers²⁷.

In this study, at first we downloaded the gene expression profiles and complete clinical information of 407 patients with GC from TCGA database. Then we downloaded 15 GC and 15 normal samples in the GSE118916 dataset, and 524 GC and 401 normal cells from scRNA-seq of GC in the GSE112302 Database. The limma package was used to analyze the differences between the tumor and normal samples, and 209 DEGs were obtained. The intercellular differences in GC cells cannot be detected by a large number of RNA sequences because conventional transcriptome detection is the average of all cells in the sample. We assumed that the mechanism and prognosis might be only related to certain cells in which certain expression of the genes are highly varies than the normal cells. Using the scRNA-seq data to test highly variable genes (HVG) allows researchers to screen out key genes that cause differences between cells in a mixed cell population¹⁷⁻¹⁹. Thus, the 209 DEGs could be the key genes closely related to the mechanism and prognosis of GC.

We conducted functional enrichment analysis on 209 DEGs in GO and KEGG. We found substantial enrichment of DEGs in the extracellular matrix organization, extracellular structure organization, and muscle contraction in GO. In KEGG, DEGs were mainly concentrated in focal adhesion, protein digestion and absorption, AGE-RAGE signaling pathway in diabetic complications. Changes in ECM (extracellular matrix) affect cell-ECM adhesion and cell-cell adhesion, reduce the epithelial characteristics of gastric epithelial cancer cells, and promote the proliferation of cancer cells and reduce the effect of

chemotherapy^{28,29}. Increased ECM density in the tumor microenvironment leads to increased integrin clustering and enhanced FAK (focal adhesion kinase) phosphorylation, which promotes cell invasion and proliferation³⁰. Studies have shown that the AGE-RAGE signaling pathway enhances metastasis of human GC³¹. By combining the screened DEGs with the clinical information of GC patients in both TCGA and the GEO databases, it was further found that the higher expression of CPVL was related to the better prognosis of GC patients. Previous studies have shown that CPVL can digest phagocytic particles in lysosomes and participate in inflammatory reactions³², and its expression also increased in GC³³. However, only a few studies have reported its other roles in GC. Our study confirmed the role of CPVL gene in GC survival. These results suggest that CPVL gene may be a promising therapeutic target for GC, which provides significance for future research on GC. Then we analyzed the relationship between the other five genes and prognosis in the GEO database. We found that FAM3D, OLR1 were near related to GC patients' prognosis in the GSE15459 dataset, while they did not correlate with prognosis in TCGA database. As the P value of FAM3D in TCGA was close to 0.05 and no studies have shown that FAM3D is associated with GC, the relationship between FAM3D and GC was worthy of further study. Studies have shown that OLR1 is positively correlated with lymph node metastasis of GC³⁴.

According to the PPI network diagram, it can be seen that CPVL interacts with proteins expressed by FGL2, PLBD1, APOE, CTSB, PGC, CES2, NEXN. Some studies have shown that overexpression of FGL2 (fibrinogen-like protein 2) induces epithelial to mesenchymal transformation in colorectal cancer and promotes tumor progression³⁵. One study showed that PLBD1 (Phospholipase B domain containing 1) was highly expressed in various pancreatic cancer cells and could be a promising biomarker for pancreatic cancer³⁶. Katsuya Sakashita et al. found that the survival time of patients with high expression of ApoE (apolipoprotein E) was shorter than that of patients with low expression³⁷. CTSB (Cathepsin B) is overexpressed in many tumors such as GC and colon cancer³⁸. PGC (peroxisome proliferator-activated receptor γ coactivator) degrades the extracellular matrix and promotes tumor invasion and metastasis³⁹. In regard to CES2 (Carboxylesterase-2), low expression of CES2 has been associated with a higher survival rate in colon cancer⁴⁰. NEXN (Nexilin) deficiency increases the expression of adhesion molecules and inflammatory cytokines⁴¹. So far, there are very few other studies on these genes and identifications of CPVL gene as a target for GC prognosis and diagnosis. These associated genes are the novel findings of our study.

5. Conclusion

Our study elucidated the relationship between CPVL and GC. We found that the higher expression of CPVL gene was associated with a better prognosis of GC, suggesting that CPVL may be a new target for the treatment of GC. CPVL interacts with proteins expressed by FGL2, PLBD1, APOE, CTSB, PGC, CES2, NEXN to a certain extent. Furthermore, our study is the first to find that FAM3D is closely related to GC, which has great research potential.

Abbreviations

GC

Gastric cancer

TCGA

The Cancer Genome Atlas

GEO

Gene Expression Omnibus

DEGs

Differentially expressed genes

PPI

Protein-protein interaction

GO

Gene Ontology

KEGG

Kyoto Encyclopedia of Genes and Genomes

CPVL

Carboxypeptidase vitellogenic-like gene

OLR1

Oxidized low-density lipoprotein receptor 1

HER2

Human epidermal growth factor receptor-2

VEGF

Vascular endothelial growth factor

ScRNA-seq

Single-cell RNA sequencing

BP

Biological process

CC

Cellular component

MF

Molecular function

HVG

Highly variable genes

ECM

Extracellular matrix

FAK

Focal adhesion kinase

FGL2

Fibrinogen-like protein 2

PLBD1

Phospholipase B domain containing 1

ApoE

Apolipoprotein E

CTSB

Cathepsin B

PGC

Peroxisome proliferator-activated receptor γ coactivator

CES2

Carboxylesterase-2

NEXN

Nexilin

Declarations

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Funding

The design of the study was supported by the funding of Zhengzhou Major Collaborative Innovation Project (No.18XTZX12003); the collection, analysis, and interpretation of data were supported by the funding of Key projects of discipline construction in Zhengzhou University (No.XKZDJC202001).

Author information

Affiliation

Henan Key Laboratory for Helicobacter pylori & Microbiota and GI cancer, Marshall Medical Research Center, The Fifth Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China

Shuang Gao, Fazhan Li, Pengyuan Zheng, and Yang Mi

Zhengzhou University School of Medical Sciences, Zhengzhou, Henan, China

Shuang Gao, Fazhan Li, Pengyuan Zheng, and Yang Mi

Department of Gastroenterology, The Fifth Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China

Shuang Gao, Fazhan Li, Pengyuan Zheng, and Yang Mi

Department of Gastrointestinal Surgery, The Fifth Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China.

Wanqing Wu and Yuming Fu

Department of Gastrointestinal Surgery, Henan Cancer Hospital, The Affiliated Cancer Hospital of Zhengzhou University, Zhengzhou, Henan, China

Minghai Zhao

Department of Pediatrics, The Fifth Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China

Youcai Tang

Contributions

Y M, S G and FZ L: Conceptualization, Writing- Original Draft; S G: Completing the manuscript; FZ L: Analyzing the data; Y M, PY Z, MH Z, WQ W, YM F, and YC T: Revising the manuscript. All authors read and approved the final manuscript.

Corresponding author

Correspondence to Yang Mi.

Ethics declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors report no declarations of interest.

Acknowledgments

The authors gratefully acknowledge all the participants.

References

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015;136(5):E359-86.
2. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin*. 2011;61(2):69–90.
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019;69(1):7–34.
4. Miyahara R, Niwa Y, Matsuura T, et al. Prevalence and prognosis of gastric cancer detected by screening in a large Japanese population: data from a single institute over 30 years. *J Gastroenterol Hepatol*. 2007;22(9):1435–42.
5. Zhang XY, Zhang PY. Gastric cancer: somatic genetics as a guide to therapy. *J Med Genet*. 2017;54(5):305–12.
6. Wagner AD, Syn NL, Moehler M, et al. Chemotherapy for advanced gastric cancer. *Cochrane Database Syst Rev*. 2017;8(8):Cd004064.
7. Kanda M, Kodera Y. Molecular mechanisms of peritoneal dissemination in gastric cancer. *World J Gastroenterol*. 2016;22(30):6829–40.
8. Tatsuta M, Itoh T, Okuda S, Yamamura H, Baba M, Tamura H. Carcinoembryonic antigen in gastric juice as an aid in diagnosis of early gastric cancer. *Cancer*. 1980;46(12):2686–92.
9. Kim DY, Kim HR, Shim JH, Park CS, Kim SK, Kim YJ. Significance of serum and tissue carcinoembryonic antigen for the prognosis of gastric carcinoma patients. *J Surg Oncol*. 2000;74(3):185–92.
10. Gravalos C, Jimeno A. HER2 in gastric cancer: a new prognostic factor and a novel therapeutic target. *Ann Oncol*. 2008;19(9):1523–9.
11. Zhou F, Li N, Jiang W, et al. Prognosis significance of HER-2/neu overexpression/amplification in Chinese patients with curatively resected gastric cancer after the ToGA clinical trial. *World J Surg Oncol*. 2012;10:274.
12. Kim JY, Jeon TJ, Bae BN, et al. The prognostic significance of growth factors and growth factor receptors in gastric adenocarcinoma. *Apmis*. 2013;121(2):95–104.
13. Ye YW, Zhang X, Zhou Y, et al. The correlations between the expression of FGFR4 protein and clinicopathological parameters as well as prognosis of gastric cancer patients. *J Surg Oncol*. 2012;106(7):872–9.
14. Shalek AK, Satija R, Adiconis X, et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*. 2013;498(7453):236–40.
15. Wills QF, Livak KJ, Tipping AJ, et al. Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments. *Nat Biotechnol*. 2013;31(8):748–52.
16. Kolodziejczyk AA, Kim JK, Tsang JC, et al. Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. *Cell Stem Cell*. 2015;17(4):471–85.
17. Islam S, Kjällquist U, Moliner A, et al. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res*. 2011;21(7):1160–7.

18. Tang F, Barbacioru C, Wang Y, et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods*. 2009;6(5):377–82.
19. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10(1):57–63.
20. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;28(6):882–3.
21. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
22. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–7.
23. Harris MA, Clark J, Ireland A, et al. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*. 2004;32(Database issue):D258-61.
24. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 1999;27(1):29–34.
25. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015;43(Database issue):D447-52.
26. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504.
27. Jiang P, Liu XS. Big data mining yields novel insights on cancer. *Nat Genet*. 2015;47(2):103–4.
28. Burute M, Thery M. Spatial segregation between cell-cell and cell-matrix adhesions. *Curr Opin Cell Biol*. 2012;24(5):628–36.
29. Al-Kilani A, de Freitas O, Dufour S, Gallet F. Negative feedback from integrins to cadherins: a micromechanical study. *Biophys J*. 2011;101(2):336–44.
30. Levental KR, Yu H, Kass L, et al. Matrix crosslinking forces tumor progression by enhancing integrin signaling. *Cell*. 2009;139(5):891–906.
31. Deng R, Mo F, Chang B, et al. Glucose-derived AGEs enhance human gastric cancer metastasis through RAGE/ERK/Sp1/MMP2 cascade. *Oncotarget*. 2017;8(61):104216–26.
32. Mahoney JA, Ntolosi B, DaSilva RP, Gordon S, McKnight AJ. Cloning and characterization of CPVL, a novel serine carboxypeptidase, from human macrophages. *Genomics*. 2001;72(3):243–51.
33. Ran X, Xu X, Yang Y, et al. A quantitative proteomics study on olfactomedin 4 in the development of gastric cancer. *Int J Oncol*. 2015;47(5):1932–44.
34. Ma C, Xie J, Luo C, et al. OxLDL promotes lymphangiogenesis and lymphatic metastasis in gastric cancer by upregulating VEGF-C expression and secretion. *Int J Oncol*. 2019;54(2):572–84.
35. Qin WZ, Li QL, Chen WF, et al. Overexpression of fibrinogen-like protein 2 induces epithelial-to-mesenchymal transition and promotes tumor progression in colorectal carcinoma. *Med Oncol*. 2014;31(9):181.

36. Makawita S, Smith C, Batruch I, et al. Integrated proteomic profiling of cell line conditioned media and pancreatic juice for the identification of pancreatic cancer biomarkers. *Mol Cell Proteomics*. 2011;10(10):M111.008599.
37. Sakashita K, Tanaka F, Zhang X, et al. Clinical significance of ApoE expression in human gastric cancer. *Oncol Rep*. 2008;20(6):1313–9.
38. Lin L, Aggarwal S, Glover TW, Orringer MB, Hanash S, Beer DG. A minimal critical region of the 8p22-23 amplicon in esophageal adenocarcinomas defined using sequence tagged site-amplification mapping and quantitative polymerase chain reaction includes the GATA-4 gene. *Cancer Res*. 2000;60(5):1341–7.
39. Dos Reis ST, Pontes J Jr, Villanova FE, et al. Genetic polymorphisms of matrix metalloproteinases: susceptibility and prognostic implications for prostate cancer. *J Urol*. 2009;181(5):2320–5.
40. Shaojun C, Li H, Haixin H, Guisheng L. Expression of Topoisomerase 1 and carboxylesterase 2 correlates with irinotecan treatment response in metastatic colorectal cancer. *Cancer Biol Ther*. 2018;19(3):153–9.
41. Hu YW, Guo FX, Xu YJ, et al. Long noncoding RNA NEXN-AS1 mitigates atherosclerosis by regulating the actin-binding protein NEXN. *J Clin Invest*. 2019;129(3):1115–28.

Figures

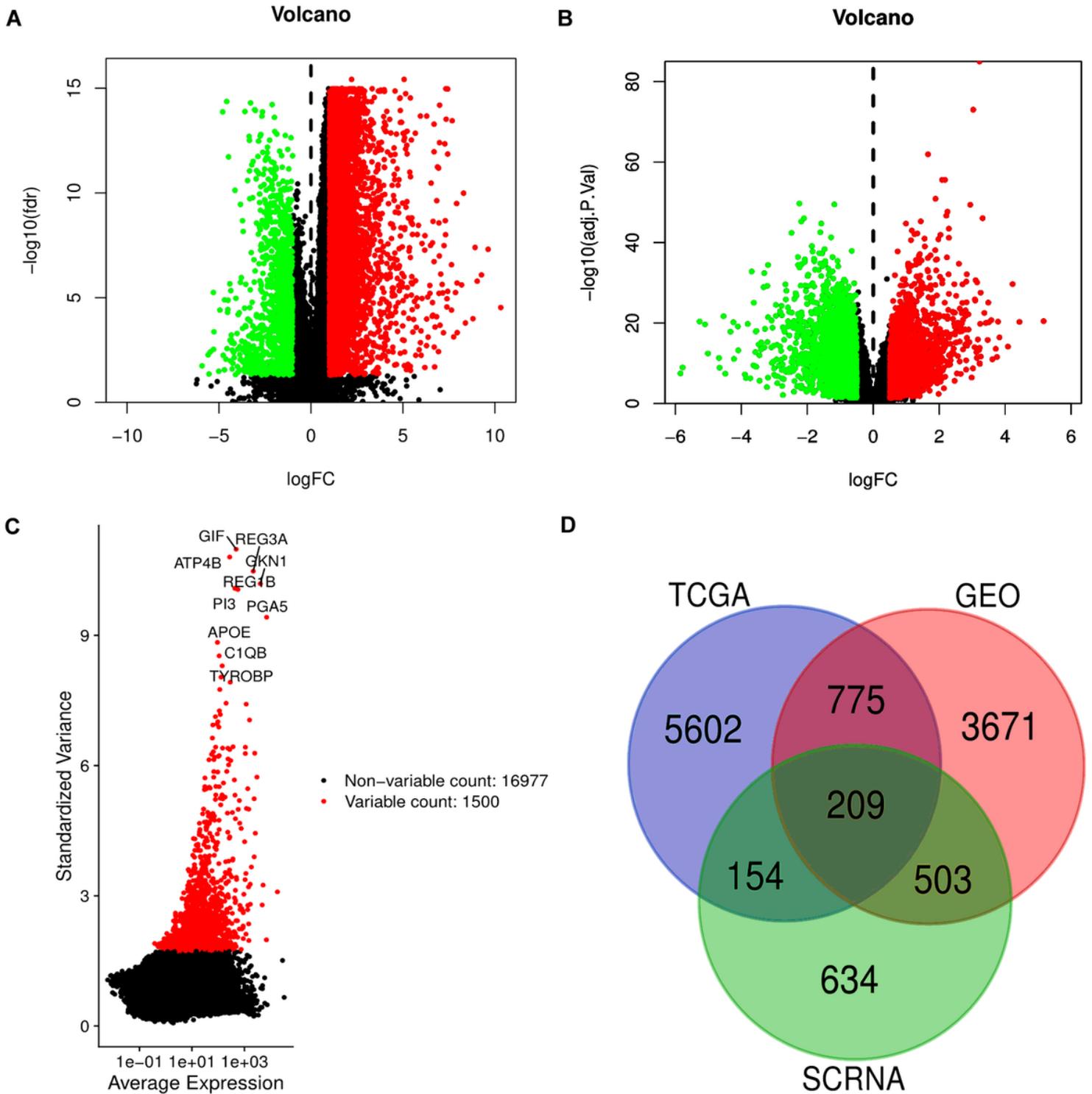


Figure 1

DEGs expression profile in TCGA and the GEO dataset. (A) A volcano plot of 6740 DEGs in TCGA. $|\log_2\text{fold change}| > 1$, $P < 0.05$. (B) A volcano plot of 5158 DEGs in GEO. $|\log_2\text{FC}| > 0.5$, $P < 0.05$. (C) 1500 high variable genes (HVG) between tumor cells and normal cells using scRNA-seq data. (D) Venn diagrams of DEGs.

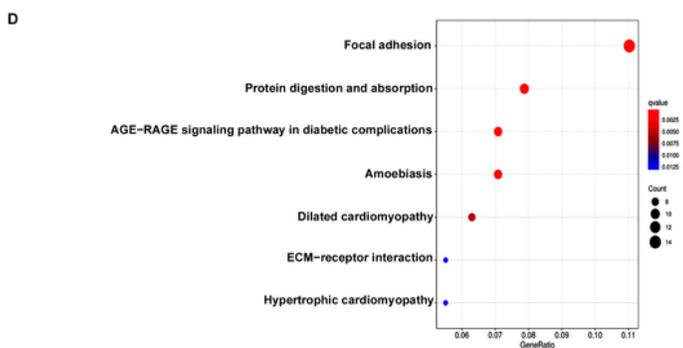
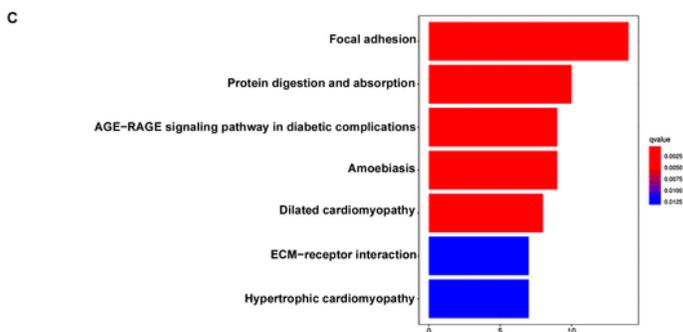
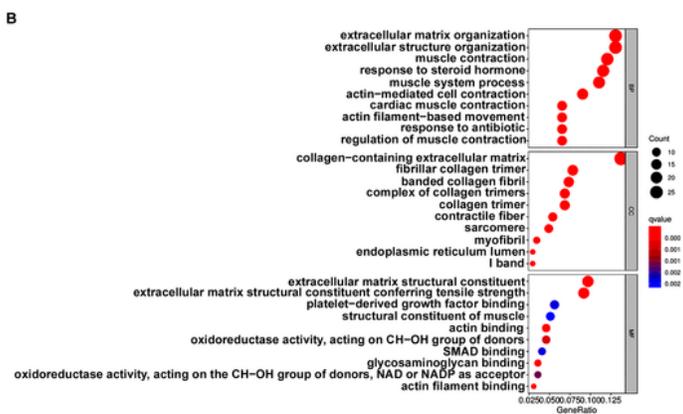
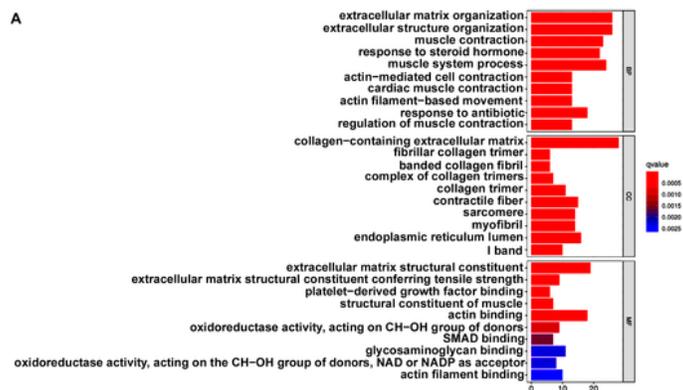


Figure 2

The Gene Ontology and KEGG pathway analysis results for DEGs. (A&B) The results of GO enrichment analysis. BP: biological process; CC: cellular component; MF: molecular function. (C&D) The results of KEGG enrichment analysis.

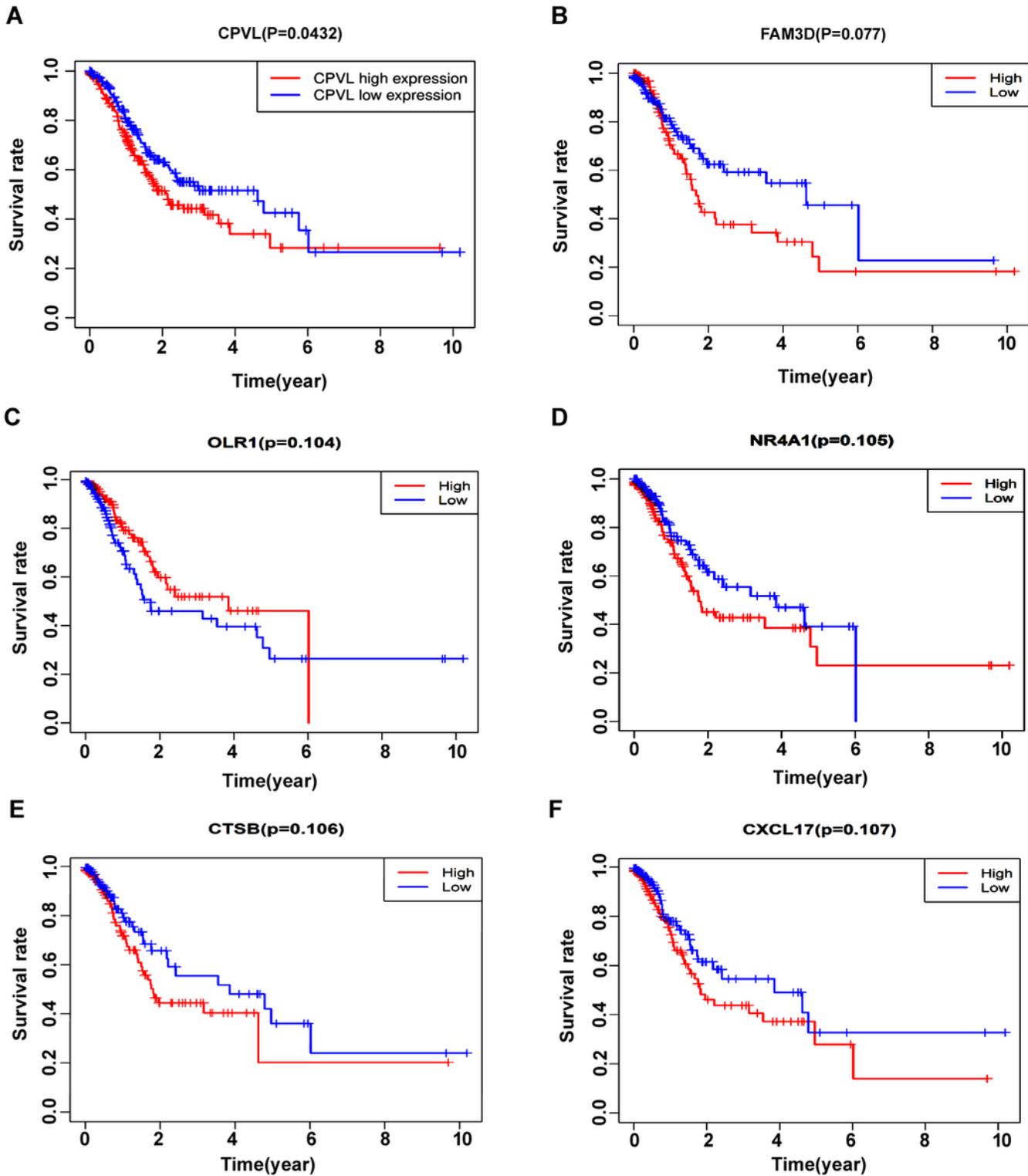


Figure 3

The survival curves by mRNA expression levels of DEGs of 407 GC patients in TCGA. $P < 0.05$ was considered statistically significant. (A) CPVL; (B) FAM3D; (C) OLR1; (D) NR4A1; (E) CTSB; (F) CXCL17.

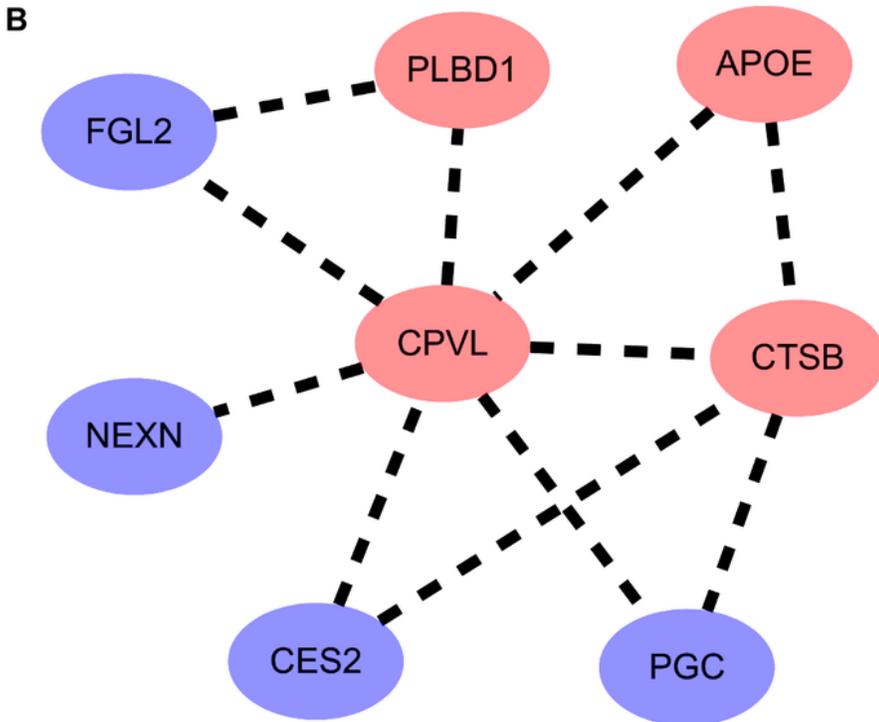
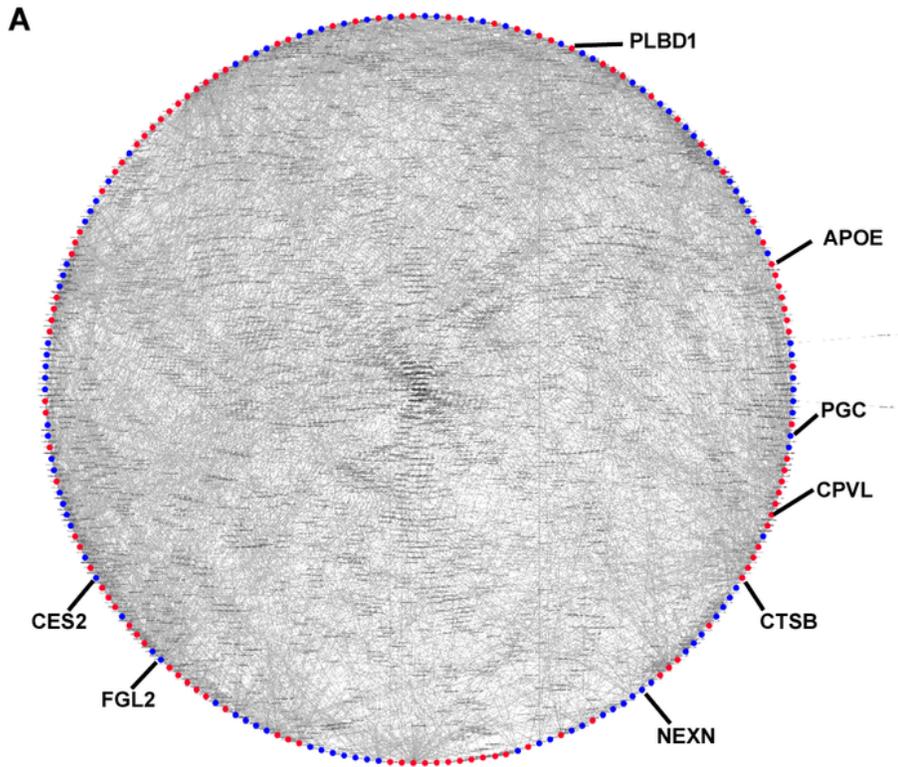


Figure 4

Protein-protein interaction (PPI) network analysis of DEGs. (A) module 1; (B) module 2. Each node is a DEG. Red nodes represent up-regulated genes. Blue nodes represent down-regulated genes.

Supplementary Files

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

- [figureS1.tif](#)