Integrative bioinformatic characterization of driver genes and prognostic features in human gastric carcinoma

Jie Zheng
The first people's hospital of Jingmen
Lan Rao
The first people's hospital of Jingmen
Yiming Han
The first people's hospital of Jingmen
Li Ding
The first people's hospital of Jingmen
Jin'e Ao
The first people's hospital of Jingmen
Changqing Liu (suhandana303@163.com)
The first people's hospital of Jingmen https://orcid.org/0000-0001-7790-9210

Research article

Keywords: Gastric cancer, individualized therapies, driver gene, prognostic markers

Posted Date: March 27th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-19492/v1

Abstract

Background: Gastric cancer (GC) is the second most common cause of cancer death worldwide but could be more curable if diagnosed at an earlier stage. At present, the capability to predict the efficaciousness of molecular diagnosis for GC for each patient remains elusive. The purpose of this study was to identify tumor biomarkers through systems analysis of multigene predictors exploiting the available data resource.

Results: Our study showed that 61 shared DEGs were significantly involved in functions such as gastric acid secretion. COL1A1, MMP9, and SPP1 were highly connected nodes in the PPI network. Among the 61 common DEGs, 34 driver genes were identified, including CST1 and MMP9. A total of 34 drug-gene interactions were found, which involved 9 common DEGs. Moreover, CTHRC1 and INHBA were related to clinical outcome in gastric cancer patients.

Conclusion: The identification of new driver genes, such as MMP9, CTHRC1, and INHBA, may contribute to the understanding of the etiology of gastric cancer and the development of individualized therapies.

Background

Gastric carcinoma (GC), also known as stomach cancer, is among the most common type of malignant tumor in the digestive system and one of the leading causes of cancer mortality worldwide [1]. The 2018 global cancer statistics reported 1,033,701 cases of GC and 782,685 deaths due to GC [2]. Surgical resection remains the only curative option for patients with GC [3]. Although advances in diagnosis and treatment have improved the prognosis of patients with GC, advanced and recurrent GC patients have poor survival rates [4]. Increasing evidence suggests that molecular biomarkers may be useful for early identification of GC patients, especially those with high mortality risk, and for improved prognosis of GC patients [5]. Thus, the identification of molecular biomarkers and prognostic signatures of GC patients with high mortality risk would be extremely valuable and could improve therapeutic strategies.

Driver genes, which are generated by genomic alterations, are implicated in the process of oncogenesis and are considered therapeutic “druggable” targets [6]. A previous study evaluating GC-related driver genes found function-altering mutations in 30 oncogenes and 18 tumor suppressor genes [7]. Additionally, mutations in a number of other genes, such as tumor protein P53 (TP53) and human epidermal growth factor receptor 2 (HER2), have been identified to be drivers of GC [8], and a few candidate driver genes, including RHOA and CDH1, have been identified using next-generation sequencing technology [9]. Understanding the unique molecular profiles of GC patients and identification of putative new druggable targets may facilitate precision medicine for GC patients and improve the accuracy of cancer prognosis. However, the molecular mechanisms and driver genes in GC development and progression have not yet been completely elucidated.

The Gene Expression Omnibus (GEO) and Cancer Genome Atlas (TCGA) are two available public database repositories that offer the opportunity to study the genomic landscape of cancers by mining

gene sequencing data and associated large-scale clinical outcomes data [10]. In this study, we downloaded two datasets from the GEO database and TCGA database that are associated with GC and screened them for common differentially expressed genes (DEGs). Additionally, we performed functional enrichment analysis and protein-protein interaction (PPI) analysis using these common DEGs. Moreover, the identified candidate driver genes were then subjected to functional enrichment analysis, drug-gene interaction analysis, and survival analysis. The identification of novel candidate driver genes offers a more complete view of the genomic processes governing GC development and facilitates precision medicine for GC patients.

Results

Determining DEGs

We obtained a total of 359 DEGs from the GSE54129 dataset that met our screening criteria of p-value < 0.05 and $|\log_2 FC| > 2$, among which 154 were up-regulated and 205 were down-regulated. A total of 970 DEGs were obtained from TCGA STAD dataset. Thereinto, the expression of 427 genes were up-regulated and 543 genes down-regulate. Heat maps and volcano plots of DEGs identified in GSE54129 and TCGA STAD are shown in Figure 1A and 1B respectively. We found 61 common genes (20 up-regulated genes and 41 down-regulated genes) that were differentially expressed in both datasets (Figure 1C).

Functional enrichment analysis of common DEGs

Functional enrichment analysis showed that the common DEGs are enriched in 3 KEGG pathways, including metabolic pathways, glycolysis/gluconeogenesis, and gastric acid secretion (Figure 2A), and are involved in 5 GO BP terms, including digestion, potassium ion import, retinoid metabolic processes, oxidation-reduction processes, and cellular aldehyde metabolic processes.

PPI network of common DEGs

The PPI network of the 61 identified common DEGs is shown in Figure 2B, with a total of 59 nodes and 225 interactive relationship pairs. Nodes with a PPI degree greater than 10 are listed in Table 1, including notable genes such as Matrix Metallopeptidase 9 (*MMP9*, degree = 21), Secreted Phosphoprotein 1 (*SPP1*, degree = 20), Collagen Type I Alpha 1 Chain (*COL1A1*, degree = 20), and Biglycan (*BGN*, degree = 17), Inhibin Subunit Beta A (*INHBA,* degree = 14), and Collagen Triple Helix Repeat Containing 1 (*CTHRC1*, degree = 13).

Driver gene prediction

The scores of potential cancer-driven genes in each sample from TCGA are shown in Supplementary Table 1. The overall rank of each gene in all samples was determined in order to determine the overall strength of potential cancer drivers compared with other drivers using the condorcetRanking function. From the 61 common DEGs, we found 34 potential driver genes. As shown in Figure 3A, these 34

potential driver genes are enriched in 1 KEGG pathway (gastric acid secretion) and 8 GO BP terms (including negative regulation of canonical Wnt signaling pathway, positive regulation of monocyte chemotaxis, regulation of muscle contraction, positive regulation of vascular endothelial growth, cell differentiation, potassium ion import, and embryonic skeletal system development). The aggregate DawnRank scores of these 34 potential driver genes are shown in Supplementary Table 2, and the top 10 potential driver genes with the highest aggregate DawnRank scores are listed in Table 2, including Cystatin SN (*CST1*), *MMP9*, Gastrokine 1 (*GKN1*), Secreted Frizzled Related Protein 4 (*SFRP4*), and *CTHRC1*.

Drug-gene interaction prediction

Using DGIdb, we obtained 34 drug-gene interaction pairs, including 9 potential driver genes (3 up-regulated genes and 6 down-regulated genes) and 33 drugs. The specific interaction relationships are presented in a network diagram (Figure 3B). From this drug-gene interaction network, we found that ATPase H+/K+ Transporting Subunit Alpha (*ATP4A*, down-regulated), Somatostatin (*SST*, down-regulated), *MMP9* (up-regulated), and *COL1A1* (up-regulated) can each be targeted by several drugs.

Potential driver gene survival analysis

Only up-regulated potential driver genes were used in our survival analysis [6]. We screened two up-regulated driver genes related to survival, including *CTHRC1* and Inhibin Subunit Beta A (*INHBA*). A K-M survival curve is shown in Figure 4. The results show that higher expression of these two driver genes predicts shorter overall survival.

Discussion

In the present study, we identified 61 common DEGs in the two GC datasets, of which 34 were found to be candidate driver genes. These driver genes are enriched in the ‘potassium ion import’ GO term and the ‘gastric acid secretion’ KEGG pathway. Moreover, the potential driver genes, including the up-regulated gene MMP9, had higher degrees in the PPI network, higher aggregate DawnRank scores, and are predicted targets of more drugs. Furthermore, CTHRC1 and INHBA have prognostic value in GC patients.

Proteins in the matrix metalloproteinase (MMP) family are associated with the degradation of extracellular matrix in normal physiological processes. The presence of MMPs have been demonstrated in GC samples [11], and MMP9 expression was enhanced in GC compared to normal mucosa [12]. A previous study reported that the T allele of the MMP-9 promoter affected GC tumor progression and invasiveness [13]. Shan et al. showed that MMP-9 increased during the pathogenesis of GC through interaction with HER2 [14]. High MMP-9 mRNA expression levels were shown to be associated with poor survival in patients with metastatic GC [15]. In line with the previous study, our study found that MMP-9 is up-regulated and may be mutated to become a driver gene associated with GC progression. Future studies are needed to validate this.

Recent studies have demonstrated that the clonal architecture of driver genes has prognostic value in a number of cancers [16, 17]. In our study, we found that the overexpression of CTHRC1 and INHBA predicted shorter overall survival of GC patients. Recent studies showed that CTHRC1 was overexpressed after promoter demethylation in metastatic GC and was an independent prognostic marker in GC [18, 19]. Wang et al. found that INHBA up-regulation was associated with poor survival in GC [20]. Similarly, Oshima et al. showed that high INHBA gene expression was a useful independent outcome predictor and was associated with significantly poorer 5-year overall survival of GC patients after curative surgery [21]. Together, the expression of these two genes may be associated with the survival outcomes of GC patients, and may be useful for developing personalized GC treatment.

Conclusions

In conclusion, we identified driver genes, including MMP-9, CTHRC1, and INHBA, which may play critical roles in GC development. Furthermore, CTHRC1 and INHBA may be effective prognostic markers for GC. The present study contributes to our understanding of the detailed mechanisms of GC and offers a foundation for the development of personalized cancer therapies.

Methods

Collection of public datasets

Gene expression dataset GSE54129 (accession number) was got from the GEO repository, which is an online international resource for the retrieval of functional genomic data [22]. In total, GSE54129 (platform: GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array) includes data from 111 human GC tissues from 111 patients with GC that underwent subtotal gastrectomy and 21 noncancerous gastric tissues from 21 volunteers who underwent gastroscopy for a health examination.

TCGA database (http://firebrowse.org/) is a leading comprehensive repository of cancer genomic profiles that stores data about multi-dimensional major cancer-causing genomic alterations in various cancers [23]. We also downloaded the stomach adenocarcinoma (STAD) patients' clinical and RNA-Seq data (Level_3__RSEM_genes_normalized__data) from TCGA database, which includes a total of 415 GC samples and 35 controls. Copy number variation (CNV) data (CopyNumber_Gistic2.Level_4) was also obtained from TCGA database.

Data processing and identification

The R package affy [24] (version 1.50.0) and the limma package [25] (version 3.10.3l) were applied to process the data downloaded from the GEO database. The Robust Multi-array Average (RMA), background correction, normalization, and expression calculation was performed. The platform annotation file was used for probe annotation and a few probes that were not annotated with gene symbols were removed. In the case of multiple probes targeting the same gene, the mean expression value of those probes was considered as the final expression of the gene. For the data downloaded from

TCGA database, the expression profile data was presented as the processed RSEM value, which can be used directly with a log2 conversion.

Subsequently, an empirical Bayesian method was implemented in the limma package [25] for identification of DEGs between GC samples and controls in GSE54129 and the dataset from TCGA database, respectively. A gene was considered differentially expressed when this analysis satisfied the conditions of Benjamini & Hochberg (BH)-corrected p-value < 0.05 and |log fold change (FC)|> 2.

Finally, the VENNY2.1.0 (http://bioinfogp.cnb.csic.es/tools/venny/index.html) online analysis tool was used to select common DEGs identified in the two gene datasets.

Functional and pathway enrichment analysis

The commonly used enrichment analysis tool DAVID [26] (version 6.8) was applied to perform functional enrichment analysis of common DEGs, including Gene Ontology (GO) Biological Process (BP) enrichment analysis [27] and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis [28]. The cut-off criteria for included genes was a p-value<0.05 and a gene count ≥2.

PPI network analysis

We determined the PPI relationships of the identified common DEGs using the Search Tool for the Retrieval of Interacting Genes (STRING) (version: 10.0) [29]. The PPI score was set to medium confidence (0.4) and the organism was set to Homo sapiens. Cytoscape software was used to visualize the resulting complex networks (version: 3.2.0) [30]. Subsequently, the CytoNCA plug-in [31] (version 2.1.6) was used to perform PPI node topology analysis (parameter ='without weight'). Each node's ranked score was used to identify the key nodes involved in the protein interaction relationships in the PPI network.

Driver gene prediction

It is generally thought that nodes in a gene interaction network are more important and more likely to be key genes in a network if they are more closely related to other genes [32]. DawnRank [33] applies this theory in its algorithm and assigns high score values to genes that significantly affect the abnormal expression of downstream genes. A gene's rank score can then be used to determine which genes in the sample can be used as driver genes.

We extracted gene expression data from the common DEGs in the TCGA STAD cancer samples and control samples, as well as gene mutations in tumor samples. A PPI network was constructed as an adjacency matrix (e.g., node i and node j were connected, Aij was considered to be 1, otherwise Aij = 0) in order to predict and analyze cancer driver genes. The DawnRank algorithm was used to obtain the gene rank score for each gene in each patient. The higher the score, the more likely it is that the gene is a driver gene. Finally, functional enrichment analysis of the identified driver genes was conducted.

Prediction of small molecule drugs

The Drug-Gene Interaction database (DGIdb) mines multiple existing resources for drug–gene interactions and generates assumptions about how genes may be developed as therapeutic targets or prioritized for drug development [34]. In the present study, we used DGIdb2.0 [34] to predict drug-gene interactions for the identified GC driver genes. The preset filter was set to 'FDA Approved.' Cytoscape was then used to build a drug-gene network map.

Survival analysis

To perform survival analysis, we used TCGA clinical information, including overall survival (OS), survival status, and disease samples screened. The STAD samples were divided into high expression group and low expression group according to the mean gene expression. Log-ranktest was done and the statistical significance threshold set to $p < 0.05$, and the relationship between survival and driver genes were determined.

Abbreviations

ATP4A, ATPase H+/K+ Transporting Subunit Alpha

BG, Biglycan

BH, Benjamini & Hochberg

BP, Biological Process

CNV, Copy number variation

COL1A1, Collagen Type I Alpha 1 Chain

CST1, Cystatin SN

CTHRC1, Collagen Triple Helix Repeat Containing 1

DEGs, differentially expressed genes

DGIdb, Drug-Gene Interaction database

FC, fold change

GC, gastric carcinoma

GEO, Gene Expression Omnibus

GKN1, Gastrokine 1

GO, Gene Ontology

HER2, human epidermal growth factor receptor 2

INHBA, Inhibin Subunit Beta A

KEGG, Kyoto Encyclopedia of Genes and Genomes

MMP, matrix metalloproteinase

MMP9, Matrix Metallopeptidase 9

OS, overall survival

PPI, protein-protein interaction

RMA, Bust Multi-Array Average

SFRP4, Secreted Frizzled Related Protein 4

SPP1, Secreted Phosphoprotein 1

SST, Somatostatin

STAD, stomach adenocarcinoma

STRING, Search Tool for the Retrieval of Interacting Genes

TCGA, the Cancer Genome Atlas

TP53, tumor protein P53

Declarations

Acknowledgements

Not applicable.

Funding

Not Applicable. There is no funding to support our research.

Author contributions

Jie Zheng and Lan Rao were responsible for downloading and collating all data. Jie Zheng and Changqing Liu designed the study and assisted in writing the manuscript. Jie Zheng, Lan Rao, Yiming Han, Li Ding and Jin'e Ao completed data analysis together. All authors read and approved the final manuscript.

Ethics declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Data availability statement

All data, models, and code generated or used during the study appear in the submitted article.

References

1. Sitarz R, Skierucha M, Mielko J, Offerhaus GJA, Maciejewski R, Polkowski WP: **Gastric cancer: epidemiology, prevention, classification, and treatment**. *Cancer management and research* 2018, **10**:239.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: **Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries**. *CA: a cancer journal for clinicians* 2018, **68**(6):394-424.
3. Weledji EP: **The principles of the surgical management of gastric cancer**. *International journal of surgery Oncology* 2017, **2**(7):e11.
4. Zhang F, Li K, Yao X, Wang H, Li W, Wu J, Li M, Zhou R, Xu L, Zhao L: **A miR-567-PIK3AP1-PI3K/AKT-c-Myc feedback loop regulates tumour growth and chemoresistance in gastric cancer**. *EBioMedicine* 2019.
5. Matsuoka T, Yashiro M: **Biomarkers of gastric cancer: Current topics and future perspective**. *World journal of gastroenterology* 2018, **24**(26):2818.
6. Ohshima K, Hatakeyama K, Nagashima T, Watanabe Y, Kanto K, Doi Y, Ide T, Shimoda Y, Tanabe T, Ohnami S: **Integrated analysis of gene expression and copy number identified potential cancer driver genes with amplification-dependent overexpression in 1,454 solid tumors**. *Scientific reports* 2017, **7**(1):641.
7. Wang T, Liu Y, Zhao M: **Mutational analysis of driver genes with tumor suppressive and oncogenic roles in gastric cancer**. *PeerJ* 2017, **5**:e3585.
8. Nadauld LD, Ford JM: **Molecular profiling of gastric cancer: toward personalized cancer medicine**. *J Clin Oncol* 2013, **31**(7):838-839.

9. Lim B, Kim S-Y: **Identification of molecular subtypes and significantly mutated genes in gastric cancer using next-generation sequencing**. *Translational Cancer Research* 2016, **5**(1):S81-S83.
10. Sun C, Cheng X, Wang C, Wang X, Xia B, Zhang Y: **Gene expression profiles analysis identifies a novel two-gene signature to predict overall survival in diffuse large B-cell lymphoma**. *Bioscience reports* 2019, **39**(1):BSR20181293.
11. Murray GI, Duncan ME, Arbuckle E, Melvin WT, Fothergill JE: **Matrix metalloproteinases and their inhibitors in gastric cancer**. *Gut* 1998, **43**(6):791-797.
12. Sampieri CL, de la Pena S, Ochoa-Lara M, Zenteno-Cuevas R, Leon-Cordoba K: **Expression of matrix metalloproteinases 2 and 9 in human gastric cancer and superficial gastritis**. *World J Gastroenterol* 2010, **16**(12):1500-1505.
13. Matsumura S, Oue N, Nakayama H, Kitadai Y, Yoshida K, Yamaguchi Y, Imai K, Nakachi K, Matsusaki K, Chayama K: **A single nucleotide polymorphism in the MMP-9 promoter affects tumor progression and invasive phenotype of gastric cancer**. *Journal of cancer research and clinical oncology* 2005, **131**(1):19-25.
14. Shan Y, Ying R, Zhou C, Zhu A, Ye J, Zhu W, Ju T, Jin H: **MMP-9 is increased in the pathogenesis of gastric cancer by the mediation of HER2**. *Cancer gene therapy* 2015, **22**(3):101.
15. Al-Batran S-E, Pauligk C, Wirtz R, Werner D, Steinmetz K, Homann N, Schmalenberg H, Hofheinz R-D, Hartmann J, Atmaca A: **The validation of matrix metalloproteinase-9 mRNA gene expression as a predictor of outcome in patients with metastatic gastric cancer**. *Annals of oncology* 2011, **23**(7):1699-1705.
16. Nadeu F, Delgado J, Royo C, Baumann T, Stankovic T, Pinyol M, Jares P, Navarro A, Martín-García D, Beà S: **Clinical impact of clonal and subclonal TP53, SF3B1, BIRC3, NOTCH1, and ATM mutations in chronic lymphocytic leukemia**. *Blood* 2016, **127**(17):2122-2130.
17. Lan Y, Zhao E, Luo S, Xiao Y, Li X, Cheng S: **Revealing clonality and subclonality of driver genes for clinical survival benefits in breast cancer**. *Breast cancer research and treatment* 2019:1-14.
18. Wang P, Wang YC, Chen XY, Shen ZY, Cao H, Zhang YJ, Yu J, Zhu JD, Lu YY, Fang JY: **CTHRC1 is upregulated by promoter demethylation and transforming growth factor-β1 and may be associated with metastasis in human gastric cancer**. *Cancer science* 2012, **103**(7):1327-1333.
19. Gu L, Liu L, Zhong L, Bai Y, Sui H, Wei X, Zhang W, Huang P, Gao D, Kong Y: **Cthrc1 overexpression is an independent prognostic marker in gastric cancer**. *Human pathology* 2014, **45**(5):1031-1038.
20. Wang Q, Wen Y-G, Li D-P, Xia J, Zhou C-Z, Yan D-W, Tang H-M, Peng Z-H: **Upregulated INHBA expression is associated with poor survival in gastric cancer**. *Medical oncology* 2012, **29**(1):77-83.
21. Oshima T, Yoshihara K, Aoyama T, Hasegawa S, Sato T, Yamamoto N, Akito N, Shiozawa M, Yoshikawa T, Numata K: **Relation of INHBA gene expression to outcomes in gastric cancer after curative surgery**. *Anticancer research* 2014, **34**(5):2303-2309.
22. Wang Z, Lachmann A, Ma'ayan A: **Mining data and metadata from the gene expression omnibus**. *Biophysical reviews* 2019, **11**(1):103-110.

23. Tomczak K, Czerwińska P, Wiznerowicz M: **The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge**. *Contemporary oncology* 2015, **19**(1A):A68.
24. Gautier L, Cope L, Bolstad BM, Irizarry RA: **affy—analysis of Affymetrix GeneChip data at the probe level**. *Bioinformatics* 2004, **20**(3):307-315.
25. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential expression analyses for RNA-sequencing and microarray studies**. *Nucleic acids research* 2015, **43**(7):e47-e47.
26. Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, Guo Y, Stephens R, Baseler MW, Lane HC: **DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists**. *Nucleic acids research* 2007, **35**(suppl_2):W169-W175.
27. Consortium GO: **Gene ontology consortium: going forward**. *Nucleic acids research* 2014, **43**(D1):D1049-D1056.
28. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes**. *Nucleic acids research* 2000, **28**(1):27-30.
29. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP: **STRING v10: protein–protein interaction networks, integrated over the tree of life**. *Nucleic acids research* 2014, **43**(D1):D447-D452.
30. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks**. *Genome research* 2003, **13**(11):2498-2504.
31. Tang Y, Li M, Wang J, Pan Y, Wu F-X: **CytoNCA: a cytoscape plugin for centrality analysis and evaluation of protein interaction networks**. *Biosystems* 2015, **127**:67-72.
32. Yao X, Hao H, Li Y, Li S: **Modularity-based credible prediction of disease genes and detection of disease subtypes on the phenotype-gene heterogeneous network**. *BMC systems biology* 2011, **5**(1):79.
33. Hou JP, Ma J: **DawnRank: discovering personalized driver genes in cancer**. *Genome medicine* 2014, **6**(7):56.
34. Wagner AH, Coffman AC, Ainscough BJ, Spies NC, Skidmore ZL, Campbell KM, Krysiak K, Pan D, McMichael JF, Eldred JM: **DGIdb 2.0: mining clinically relevant drug–gene interactions**. *Nucleic acids research* 2015, **44**(D1):D1036-D1044.

Tables

Table 1 PPI nodes with degree ≥10

Gene	Degree	Betweenness	Closeness
MMP9	21	382.66336	0.5631068
SPP1	20	243.44894	0.5420561
COL1A1	20	258.4717	0.5471698
BGN	17	141.98326	0.51785713
LIPF	15	377.33154	0.51785713
INHBA	14	143.53568	0.50877196
SFRP4	14	93.93101	0.46774194
CTHRC1	13	47.303627	0.46031746
FAP	13	41.645958	0.48333332
THBS2	13	18.084356	0.46031746
SST	13	136.72327	0.48333332
COL10A1	13	47.679756	0.47933885
LTF	13	281.7638	0.48333332
GKN2	12	244.91818	0.5
FNDC1	12	39.763443	0.42647058
SULF1	11	70.82542	0.44615385
GKN1	11	248.30681	0.5043478
TFF2	11	202.62675	0.464

Table 2 The top 10 potential driver genes with higher aggregate DawnRank scores

Gene	The aggregate DawnRank scores
CST1	0.944344
MMP9	0.928709
GKN1	0.894126
SFRP4	0.762127
GKN2	0.752503
CTHRC1	0.743753
MAL	0.709131
PLA2G7	0.651946
VSTM2A	0.629134
SOSTDC1	0.591213

Figures

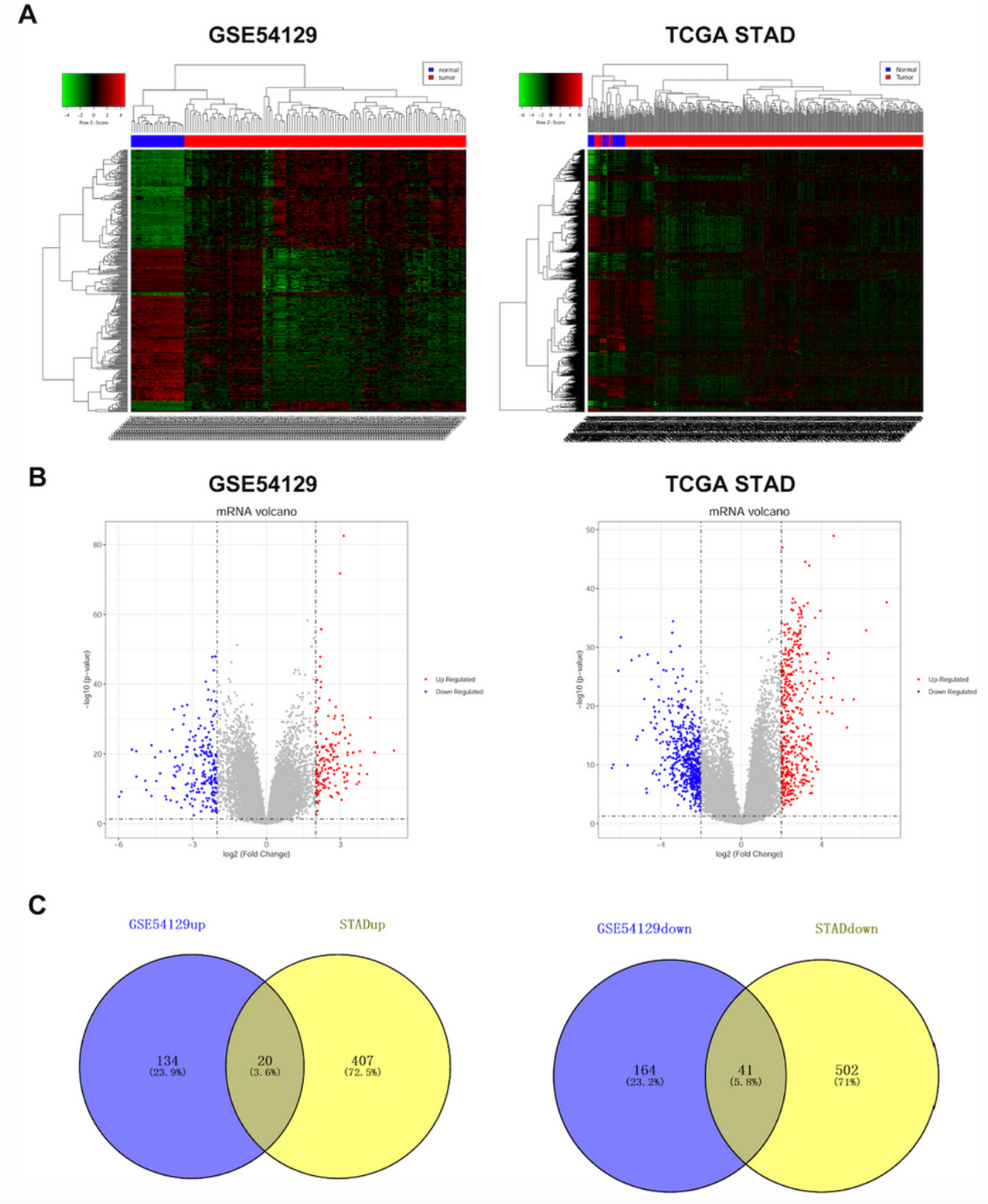

Figure 1

Differentially expressed genes (DEGs) in GC datasets. (A) Heat maps of DEGs identified in the GSE54129 and TCGA STAD datasets. (B) Volcano plots of DEGs identified in the GSE54129 and TCGA STAD datasets. (C) Venn diagram highlighting the 61 genes that were common to the two datasets.

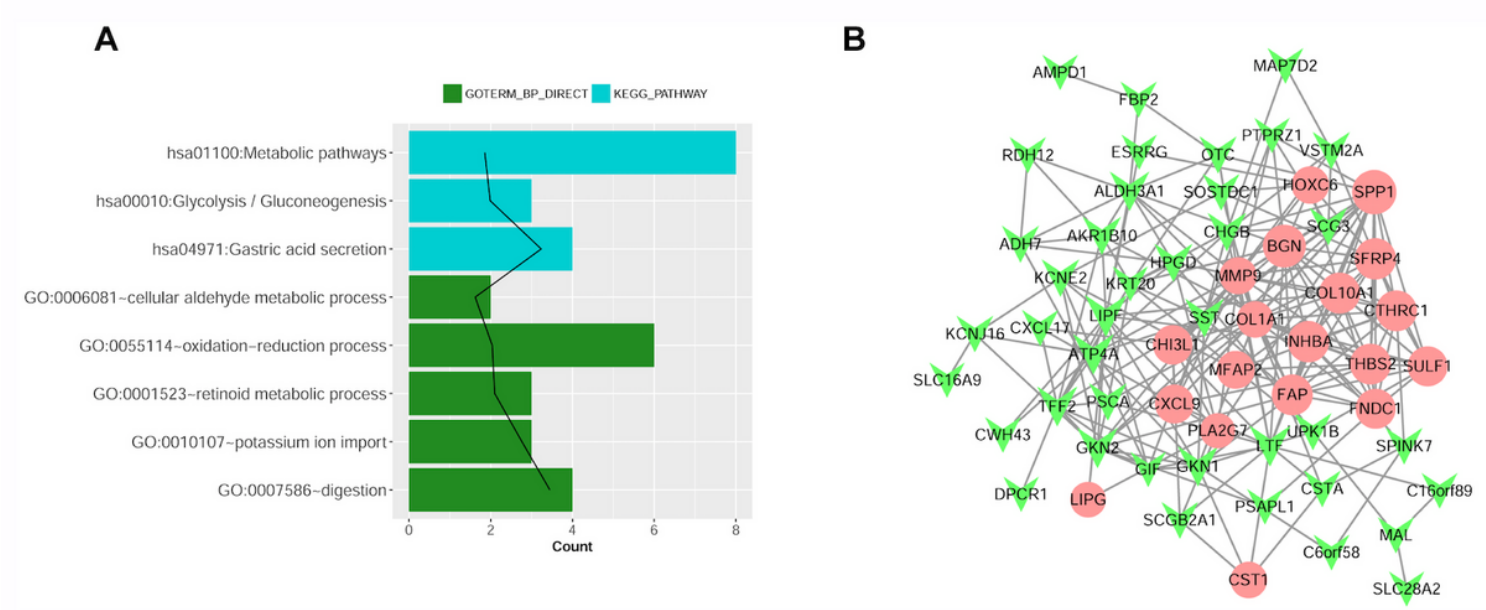

Figure 2

Functional enrichment analysis and PPI network of the 61 common DEGs. (A) GO BP terms and KEGG pathways enriched by common DEGs. The black line indicates -log10 (p-value). (B) PPI network of common DEGs. The red nodes represent up-regulated DEGs, the green inverted triangle nodes represent down-regulated DEGs, and the node size represents the degree value.

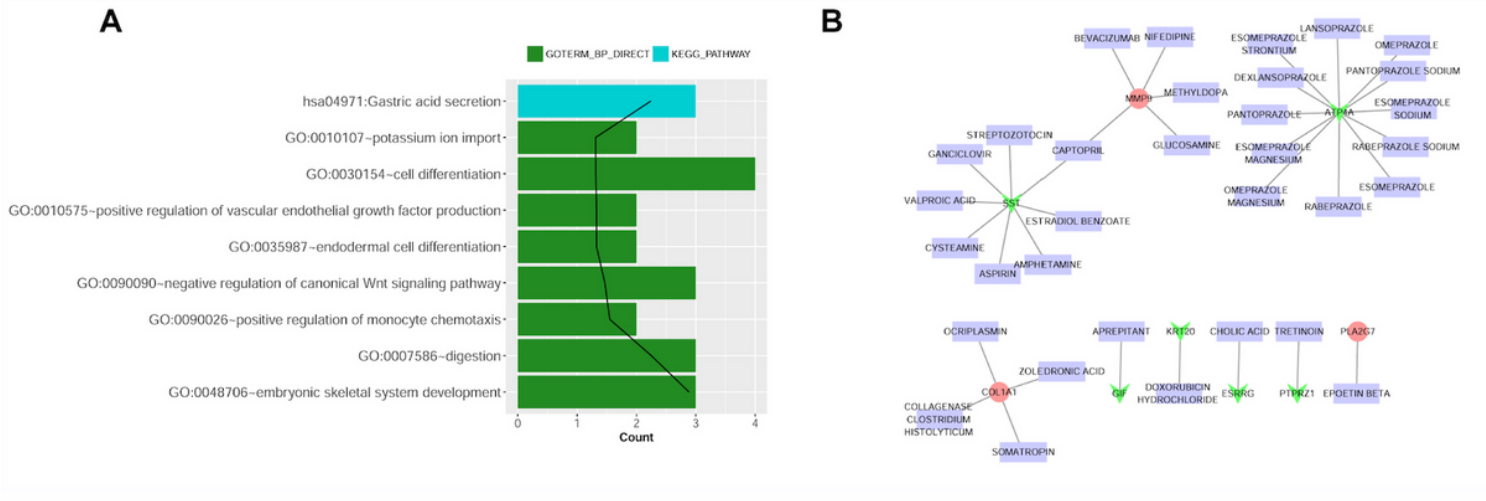

Figure 3

Functional enrichment and prediction of drug-driver gene interactions for potential driver genes. (A) GO BP terms and KEGG pathways enriched by the 34 identified potential driver genes. The black line indicates -log10 (p-value). (B) drug-driver gene interaction network. Red indicates up-regulated driver genes, green indicates down-regulated driver genes, and purple indicates drugs.

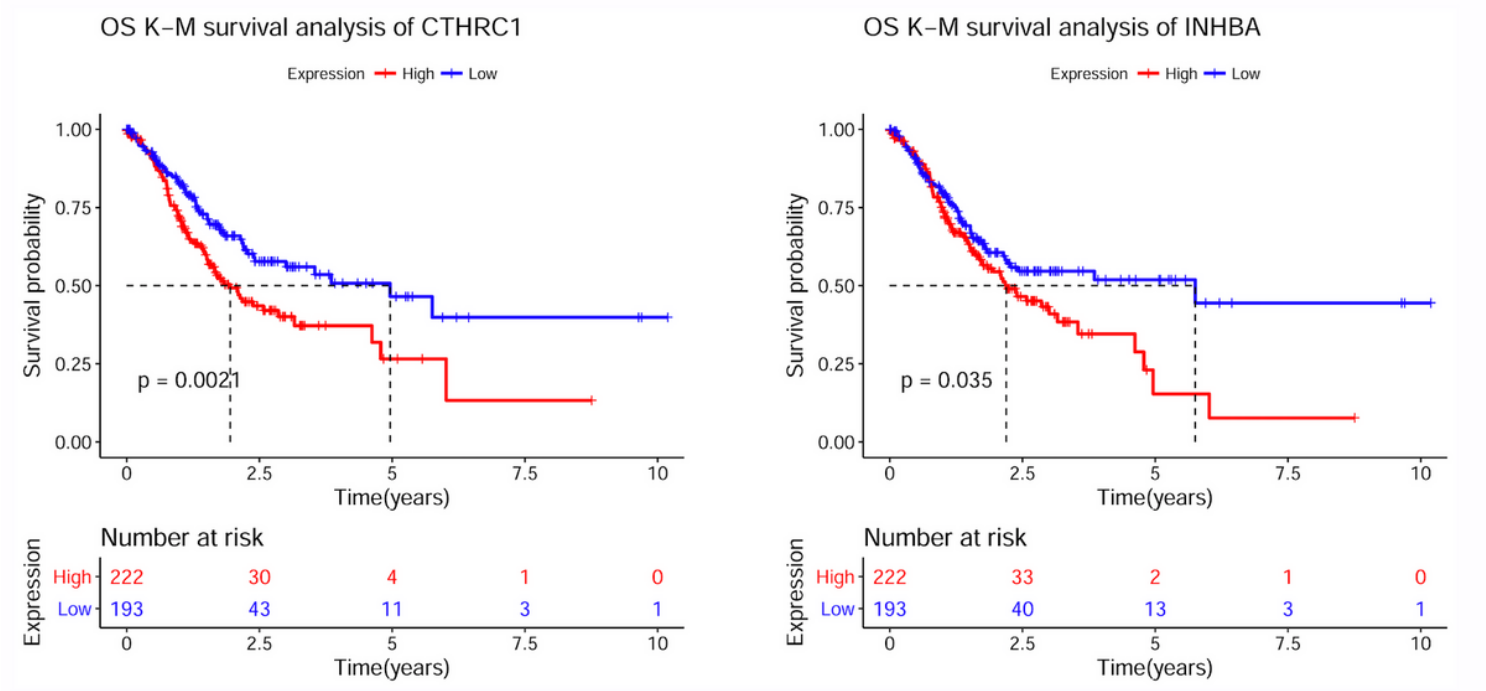

Figure 4

Survival analysis of candidate driver genes identified two potential prognostic factors.

Supplementary Files

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

- SupplementaryTable2.xlsx
- SupplementaryTable1.xlsx