Bioinformatics analysis identifies COL1A1, THBS2, and SPP1 as potential predictors of patient prognosis and immunotherapy response in gastric cancer

Yali Wang¹, Kun Zheng¹, Xiuqiong Chen¹, Rui Chen¹, Yanmei Zou¹

¹Department of Oncology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China.

Corresponding Author:

Yanmei Zou¹

Jiefang Road, Wuhan, Hubei, 430030, China

Email address: zouyanmei0101@126.com

Abstract

Background

This study aimed to use bioinformatics tools to explore pivotal genes associated with the occurrence of gastric cancer (GC) and assess their prognostic significance, and link with...
clinicopathological parameters. We also investigated the predictive role of COL1A1, THBS2, and SPP1 in immunotherapy.

Materials and methods

We identified differential genes (DEGs) that were up- and down-regulated in the three datasets (GSE26942, GSE13911, and GSE118916) and created protein-protein interaction (PPI) networks from the overlapping DEGs. We then investigated the potential functions of the hub genes in cancer prognosis using PPI networks, and explored the influence of such genes in the immune environment.

Results

Overall, 268 overlapping DEGs were identified, of which 230 were up-regulated and 38 were down-regulated. CytoHubba selected the top ten hub genes, which included SPP1, TIMP1, SERPINE1, MMP3, COL1A1, BGN, THBS2, CDH2, CXCL8, and THY1. With the exception of SPP1, survival analysis using the Kaplan-Meier database showed that the levels of expression of these genes were associated with overall survival. Genes in the most dominant module explored by MCODE, COL1A1, THBS2, and SPP1, were primarily enriched for two KEGG pathways. Further analysis showed that all three genes could influence clinicopathological
parameters and immune microenvironment, and there was a significant correlation between

COL1A1, THBS2, SPP1, and PD-L1 expression, thus indicating a potential predictive role for

GC response to immunotherapy.

Conclusion

ECM-receptor interactions and focal adhesion pathways are of great significance in the

progression of GC. COL1A1, THBS2 and SPP1 may help predict immunotherapy response in

GC patients.

Keywords: Bioinformatics, gastric cancer, hub genes, immunotherapy

Subjects: Bioinformatics, Cancer, Diagnostics & Biomarkers, Gene Expression & Regulation

Introduction

Gastric cancer (GC) is the fourth most common human malignant tumor worldwide and the

second leading cause of cancer death worldwide [1]. It has an insidious onset and is mostly
detected at an intermediate to advanced stage [2]. Therefore, even though multiple therapies such

as surgery, chemotherapy, radiation therapy, and targeted therapies alone or in combination have
played a role in the treatment of GC, the overall 5-year survival rate is still less than 20% [3].

Such low survival rate is mainly due to the heterogeneity of GC [4]. Genetic mutations are known to play a key role in the development, progression, and prognosis of a variety of diseases, and genetic biomarkers have been widely used in the diagnosis and targeted treatment of various diseases [5-7]. In this context, identifying the mutated genes in GC and developing effective therapeutic strategies is a crucial factor in improving patient prognosis.

In recent years, the analysis of biological information, also known as bioinformatics, has attracted a great deal of attention and sustained breakthroughs in the search for oncogenic genes. Various functions for molecular typing, prognostic prediction, new targeted drug development applications and biomarkers of prognosis have been confirmed [8-10]. Thus, we used bioinformatics to identify genes predictive of GC prognosis. In this study, we analyzed three mRNA gene chip datasets (GSE26942, GSE13911, and GSE118916), and identified overlapping DEGs between tumor and normal samples. The prognostic value of the identified genes was assessed by survival analysis. The efficacy of immune checkpoint inhibitors was influenced by a combination of factors such as tumor genomics, in vivo PD-L1 levels, and characteristics of the tumor microenvironment. Therefore, we selected core genes to explore their effects on the
immune microenvironment and PD-L1 expression in GC. The genes identified in this study may be potential prognostic biomarkers and therapeutic targets for GC. Further, the association of these genes with the GC immune environment could be further explored.

**Materials and Methods**

**Microarray Data**

To obtain gene expression datasets for GC in this study, we downloaded three datasets from the GEO database ([http://www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/)): GSE26942, GSE13911, and GSE118916. These RNA profiles were based on GPL6947 (Illumina HumanHT-12 V3.0 expression bead chip), GPL570 platform (Affymetrix Human Genome U113 Plus 2.0 array), and GPL15207 (Affymetrix Human Gene Expression Array), and contained 258 GC tissues and 58 normal tissues. The GSE26942 profile contains 205 GC tissues and 12 normal tissues, the GSE13911 profile 38 GC tissues and 31 normal tissues, and the GSE118916 profile 15 GC tissues and 15 normal tissues.
DEGs Screening

We obtained and exclusively selected differential genes (DEGs) by GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r/) analysis from the GEO database. Genes with P value <0.05 and | logFC | > 1 were considered as DEGs in this dataset. A Venn diagram was used to visualize overlapping DEGs between the above datasets.

Function and Pathway Enrichment Analysis

Overlapping DEGs were used to map possible biological functions, and gene ontology (GO) enrichment analysis was performed in terms of biological processes (BP), cellular components (CC), and molecular function (MF) [11, 12]. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was then performed to investigate potential enrichment for signaling pathways among the overlapping DEGs. The KEGG database is widely used to explore information about biological pathways, genomes, diseases, chemicals, and to identify functional and metabolic pathways [13]. GO and KEGG
pathway enrichment analysis were performed using the DAVID website. Bubble charts were created using R language.

Establishment of the PPI network, modules selection and identification of hub genes

To investigate the protein interactions of DEGs, we submitted them to the Interaction Gene Search tool (STRING) [14]. We then used Cytoscape software to integrate and visualize the protein-protein interaction (PPI) network. Molecular complex detection (MCODE) was applied to screen the modules of the PPI network, and the core modules were selected. In addition, hub genes in the network were identified using the cytoHubba application in Cytoscape software. The maximum group centrality of each gene in the network was calculated by the maximal clique centrality (MCC) score. We regarded the top 10 genes as hub genes.

Relationship between hub genes and survival analysis

We explored the main functions and pathway enrichment analysis of the hub genes by searching the DAVID website (https://david.ncifcrf.gov/) [15]. To further validate the reliability of the hub genes, their expression was first examined using the Gene Expression Profiling Interaction
Analysis (GEPIA) database analysis [16]. The relationship between hub genes and patient prognosis was then analyzed by Kaplan-Meier curves (http://kmplot.com/) [17], and \( P < 0.05 \) was considered to be statistically significant.

Expression and mutation analysis of hub genes

The relationship between gene expression and mutations of hub genes was carried out using the online c-BioPortal database (http://cbioportal.org) [18]. A total of 478 patients/samples were selected for further analysis. OncoPrint was obtained using an online database at c-BioPortal.

Link between Hub genes expression levels and GC clinicopathological parameters

ONCOMINE (https://www.oncomine.org/) [19] was used to explore the gene expression levels of potential biomarkers in GC tumor tissue compared with normal tissue. UALCAN (http://ualcan.path.uab.edu/) [20], an online database for the analysis and mining of cancer data, was used to explore the relationship between mRNA expression levels of potential biomarkers and the clinicopathological parameters of GC patients. \( P \) value < 0.05 was considered statistically significant. Immunohistochemistry was performed using the Human Protein Atlas.
We evaluated the expression levels of hub genes between normal gastric tissues and GC tissues from THPA.

Relationship between hub genes and immune infiltrates

The correlation between hub genes and immune infiltration (B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages, and dendritic cells) in The Cancer Genome Atlas (TCGA) of Stomach Adenocarcinoma (STAD) was carried out using the Tumor Immune Estimation Resource (TIMER) platform.

The expression between hub genes and PD-L1 gene expression

The relationship between each of the three hub genes and PD-L1 gene expression in the clinical data of 580 GC patients in TCGA database are available from UCSC (https://genome-cancer.ucsc.edu/).

Results

Identification of DEGs
After a comprehensive analysis of the three microarray datasets, we identified 268 overlapping DEGs, of which 230 genes were upregulated and 38 genes were downregulated. A Venn diagram for the three datasets is shown in Figure 1.

GO and KEGG pathway enrichment analyses of DEGs

GO and KEGG pathway enrichment analyses were performed using the DAVID website. The top 10 terms of BP, CC, MF, and KEGG pathways are shown in Figure 2. For biological processes, results from the GO analysis revealed an enrichment for genes linked to processes such as digestion, oxidation-reduction process, cellular response to cadmium ion, xenobiotic metabolic process, cellular response to zinc ion, negative regulation of growth, multicellular organismal water homeostasis, cell adhesion, cellular aldehyde metabolic process, and extracellular matrix organization (Figure 2A). Cellular component analysis showed that the DEGs were particularly enriched for categories such as extracellular space, extracellular exosome, apical plasma membrane, extracellular region, basolateral plasma membrane, organelle membrane, platelet alpha granule, external side of plasma membrane, perinuclear region of cytoplasm, and sodium channel complex (Figure 2B). Regarding molecular functions, analysis
of DEGs revealed a significantly enrichment for functions such as oxidoreductase activity,
creatine kinase activity, retinal dehydrogenase activity, ligand-gated sodium channel activity,
aromatase activity, heparin binding, WW domain binding, NADP binding, heme binding, and
protease binding (Figure 2C). Additionally, the results of KEGG pathway analysis revealed that
the DEGs were mostly linked to the metabolism of xenobiotics by cytochrome P450, chemical
carcinogenesis, mineral absorption, drug metabolism - cytochrome P450, gastric acid secretion,
mineral absorption, metabolic pathways, ECM-receptor interaction, retinol metabolism,
glycolysis / gluconeogenesis, fructose and mannose metabolism (Figure 2D).

Construction of PPI network and identification of hub genes

A PPI network from the STRING database was established to predict the potential interactions
of the DEGs at the protein level (Figure 3A). By opening the PPI network from the STRING
database with Cytoscape software, we constructed a co-expression network consisting of 213
nodes and 489 edges (Figure 3B). MCODE was applied to screen modules of the PPI network.
The most pivotal module in the PPI network was selected (Figure 3C). In addition, hub genes in
the network were identified using the cytoHubba application in Cytoscape software. The top 10
genes were selected as hub genes, including SPP1, TIMP1, SERPINE1, MMP3, COL1A1, BGN, THBS2, CDH2, CXCL8, and THY1 (Figure 3D).

Functional analysis of hub genes and survival analysis

Most hub genes were involved in the core module. Enrichment analysis using the DAVID website revealed that hub genes were mainly enriched for the following GO terms: extracellular region, cell adhesion, extracellular matrix organization, extracellular space, extracellular matrix, and extracellular matrix disassembly (Table 1). A P-value <0.05 was set as the cut-off value. KEGG results included ECM-receptor interaction and focal adhesion (Table 1). Based on the GEPIA website, all hub genes were differentially expressed between gastric tumor tissues and normal tissues (Figure 4), which further indicated that these genes may contribute to the occurrence and development of GC. The Kaplan-Meier plotter database confirmed that higher expression levels of these genes were related to lower overall survival, except for SPP1, MMP3, and CXCL8. In fact, the mRNA expression of SPP1 was independent of overall survival. Higher levels of MMP3 and CXCL8 expression had positive impacts on overall survival (Figure 5).

Mutational analysis of hub genes
In the mutational analysis of the hub genes, we found that mutations were mainly present in THBS2, COL1A1, and CDH2 (Figure 6). Among all mutation types, amplification was the most common.

**Correlation between hub genes expression levels and GC clinicopathological parameters**

COL1A1, THBS2, and SPP1 were found to participate in both ECM-receptor interaction and focal adhesion (Table 1). Furthermore, we used ONCOMINE to explore the expression levels of potential biomarkers in GC tumor tissue compared with normal tissue. We found 400, 447, and 431 studies on COL1A1, THBS2, and SPP1, respectively (Figure 7). To further validate their prognostic significance of the three genes, we looked at their levels of expression in normal gastric tissues compared to GC tissues in THPA. Immunohistochemistry indicated that the expression levels of COL1A1 and SPP1 are upregulated in GC tissues. THBS2 protein did not show significant differences in immunohistochemistry between GC and normal gastric tissues (Figure 8). We further explored the relationships between COL1A1, THBS2, and SPP1 expression levels and clinicopathological parameters of GC patients using UALCAN analysis. As shown in Figure 9 A-C, the mRNA transcription levels of the three genes, considered...
potential biomarkers of GC, were positively correlated with individual cancer stages, and nodal metastasis status in GC, but independent of H. pylori infection and sex. Notably, the levels of expression of the three genes differed significantly among different human populations (Figure 9).

**Correlation between hub genes and immune infiltrates**

The efficacy of immune checkpoint inhibitors is influenced by a combination of factors such as tumor genomics, in vivo PD-L1 levels, and characteristics of the tumor microenvironment. To provide further insight into the management of immunotherapy in GC, this study analyzed the correlation between hub genes and tumor immune infiltration via the TIMER platform (Figure 10). The expression levels of COL1A1 showed a significant positive correlation with macrophages and CD4+ T cells (cor = 0.357, P<0.0001), and a negative correlation with B cells (cor = -0.214, P<0.0001). THBS2 was highly correlated with macrophages (cor = 0.526, P<0.0001) and dendritic cells (cor = 0.393, P<0.0001). In addition, it also had some correlation with CD8+ T cells and CD4+ T cell expression (P<0.005). SPP1 showed the most significant negative correlation with B cells (cor = -0.328, p<0.0001), and no correlation with CD8+ T cells (P > 0.05). With the exception of CD4+ T cells, SPP1 showed a certain degree of positive correlation with the remaining immune infiltrating cells.
Association between hub genes and PD-L1 gene expression in GC patients

After reviewing the literature, we chose to explore the association between gene mutations and PD-L1 gene expression to assess gene signature and immune relevance. Results from the UALCAN database showed a significant increase in PD-L1 expression in GC (Figure 11A) and that the three hub genes exist in the COL1A1/PD-L1 axis, THBS2/PD-L1 axis, and SPP1/PD-L1 axis (Figure 11B-C).

Discussion

In recent years, immunotherapy has been extensively studied as a promising strategy for the treatment of cancer [23]. Programmed cell death ligand 1 (PD-L1) is an important immunomodulatory molecule that is highly expressed in many human cancer types [24, 25] and can inhibit physiological and pathological pathways by interacting with its receptor programmed cell death 1 (PD-1) [26]. Despite the success of immune checkpoint therapy in melanoma and lung cancer, the efficacy of PD-1 or PD-L1 blockade in GC is not good. Clinically, the overall remission rate in GC patients treated with PD-1/PD-L1 inhibitors is only about 20% [27]. Further, PD-L1 is not an ideal biomarker for gastroesophageal cancer. The efficacy of immune
checkpoint inhibitor single agents is influenced by tumor genomics, in vivo PD-L1 levels, and
the tumor microenvironment [27-29]. For example, in GC, microsatellite instability and EBV
status are predictive of the efficacy of immunosuppressive drugs [30, 31]. Similarly, there are
many factors related to the immune microenvironment that also predict whether a patient will
respond to single-agent PD-1 inhibition. Such factors include high tumor antigen load, dendritic
cells, CD4+T cell infiltration, CD8+T cell infiltration, and pro-inflammatory cytokines, among
others [32-35]. The most well-known cytokine responsible for PD-L1 upregulation in the
gastrointestinal tract is the regulatory cytokine interleukin 10 (IL10) [36]. In addition,
single-agent effects may also be associated with altered PD-L1 expression due to genetic
alterations and activation of the oncogenic signaling pathways PI3K/AKT, ERK, and
JAK/STAT3 pathways [28, 29].

An increasing number of reports have revealed oncogenic genes and other genetic marker
relationships in GC. In the context of precision medicine, tailoring treatments based on a
patient’s genotype is becoming increasingly important. Therefore, we selected hub genes to
explore the assessment of gene signatures and their correlations with the tumor
immunoenvironment.
Our results showed that the hub genes COL1A1, THBS2, and SPP1 were also able to alter cellular components in the tumor microenvironment, and that they had an axial relationship with PD-L1 expression. COL1A1, which encodes the most abundant protein of the collagen family, is a major component of the extracellular matrix and can influence cell behavior and tissue structure [37]. COL1A1 is considered to effectively suppress gene expression and inhibit the proliferation, migration, and invasion of GC cells [38]. THBS2 was identified as a potent inhibitor of angiogenesis and tumor growth, while promoting cell adhesion and migration [39]. THBS2 also affects the proteolysis of tumor cytoplasm, thereby contributing to certain proteins in the PI3K / AKT signaling pathway. Previous studies have observed that the expression of TGF-β, COL1A1, and THBS2 in GC cells is associated with the survival of GC patients in a time-dependent manner [40], suggesting that COL1A1 and THBS2 may affect PI3K by upregulating the TGF-β signaling pathway. COL1A1 has also been observed to play a regulatory role in the JAK pathway [41]. Interaction of SPP1 and its receptor CD47 further inhibits angiogenesis by antagonizing nitric oxide signaling in endothelial and vascular smooth muscle cells which, in turn, affects the tumor microenvironment [42]. SPP1 can directly regulate interleukin 6 [43]. SPP1 is also associated with the expression of genes related to the PI3K/ AKT
pathway and epithelial-mesenchymal transition (EMT) [44]. We explored that COL1A1, THBS2, and SPP1 were mainly involved in ECM receptor interactions and adhesion plaque pathways using KEGG pathway analysis. Further, they affect the immune microenvironment and upregulate the expression of PD-L1 through PI3K/AKT signaling pathway, JAK signaling pathway, and TGF-β signaling pathway. In addition, it is possible that more basic experiments are needed to elucidate the mechanism.

**Data Availability Statement**

The authors certify that all the original data in this research could be obtained from public database. The clinical data of GC patients was obtained from the TCGA database (https://tcga-data.nci.nih.gov/tcga/). The data of training groups are available at NCBI GEO (GSE26942 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE26942, GSE13911 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE13911, and GSE118916 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE26942). All data generated or analysed during the present study are included in this published article and its supplementary file information files.
Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

This work was supported by National Natural Science Foundation of China [grant number 81502118].

Author Contribution

YW conducted the data screening, statistical analysis and writing of the manuscript. KZ provided technical support. XC and RC provided comments on the structure of the article. YZ had read and provided final approval of the version to be submitted. The final manuscript was approved by all the authors above.

Acknowledgements

We thank the TCGA and GEO databases for the availability of the data. We would like to thank Editage (www.editage.cn) for English language editing.
Abbreviations

GC: Gastric cancer
DEGs: Differential genes
PPI: Protein-protein interaction
MCODE: Molecular complex detection
GO: Gene ontology
BP: Biological processes
CC: Cellular components
MF: Molecular function
KEGG: Kyoto Encyclopedia of Genes and Genomes
MCC: Maximal clique centrality
GEPIA: Gene Expression Profiling Interaction Analysis
TCGA: The Cancer Genome Atlas
STAD: Stomach Adenocarcinoma
TIMER: The Tumor Immune Estimation Resource
PD-L1: Programmed cell death ligand 1
EMT: Epithelial-mesenchymal transition

IL10: Interleukin 10

Figures Legends

Figure 1. A Venn diagram of DEGs. A comparison between three datasets with 433, 3327, and 1888 DEGs, revealed 268 common DEGs between GC and normal tissues.

Figure 2. Functional enrichment analysis and KEGG pathway analysis of DEGs. (A) Functional enrichment analysis of BP; (B) Functional enrichment analysis of CC; (C) Functional enrichment analysis of MF; (D) KEGG pathway analysis.

Figure 3. Selection of hub genes. (A) PPI network by STRING analysis; (B) PPI network was produced using Cytoscope Software; (C) Core modules were designed using MCODE; (D) Top 10 genes network. The top 10 genes of the MMC method were chosen using the CytoHubba plug-in. The more forward ranking is represented by a redder color.

Figure 4. Expression of hub genes in normal and GC tissues. (A)SPP1; (B)TIMP1; (C)SERPINE1; (D)MMP3; (E)COL1A1; (F)BGN; (G)THBS2; (H)CDH2; (I)CXCL8; (J)THY1.
Figure 5. Relationship between the expression of hub genes and overall survival estimated using the Kaplan-Meier plotter. (A) SPP1; (B) TIMP1; (C) SERPINE1; (D) MMP3; (E) COL1A1; (F) BGN; (G) THBS2; (H) CDH2; (I) CXCL8; (J) THY1.

Figure 6. Genetic mutation analysis of 10 hub genes in GC. The mutation rate is 7% for THBS2; 6% for COL1A1 and CDH2; 5% for SERPINE1, MMP3, and BGN; and less than 5% for SPP1, CXCL8, TIMP1, and THY1.

Figure 7. mRNA transcription levels of potential GC biomarkers using ONCOMINE. Numbers in each cell indicate the number of studies that met the target genes screening criteria. The depth of color indicates the level of gene expression, with red indicating overexpression and blue indicating downregulation.

Figure 8. Immunohistochemistry from THPA was used to explore differences in the gene expression of three genes between normal and GC tissues. (A) COL1A1; (B) THBS2; (C) SPP1.

Figure 9. Relationship between clinicopathological parameters and hub genes. (A) Clinicopathological parameters of COL1A1; (B) Clinicopathological parameters of THBS2; (C) Clinicopathological parameters of SPP1.
Figure 10. Correlation analysis between immune infiltrates and expression of COL1A1, THBS2, and SPP1 genes. The expression levels of COL1A1 showed a significant positive correlation with macrophages and CD4 + T cells, and it was a definite negative correlation with B cells. THBS2 was highly correlated with macrophages and dendritic cells. SPP1 showed the most significant negative correlation with B cells, and no correlation with CD8 + T cells. With the exception of CD4 + T cells, SPP1 showed a certain degree of positive correlation with the remaining immune infiltrating cells.

Figure 11. Correlation between expression of hub genes and PD-L1 expression. (A) Expression of PD-L1 in normal versus GC tissues; (B) Co-identification of PD-L1 and COL1A1 in GC tissues; (C) Co-identification of PD-L1 and THBS2 in GC tissues. (D) Co-identification of PD-L1 and SPP1 in GC tissues.

Table legends

Table 1. Functional enrichment analysis and KEGG pathway analysis of hub genes.
References

1. Siegel RL, Miller KD, Jemal A: *Cancer statistics, 2019*. *CA Cancer J Clin* 2019, 69(1).
2. Kumar V, Soni P, Garg M, Kamholz S, Chandra AB: *Emerging Therapies in the Management of Advanced-Stage Gastric Cancer*. *Front Pharmacol* 2018, 9:404.
3. Shah MA, Kelsen DP: *Gastric cancer: a primer on the epidemiology and biology of the disease and an overview of the medical management of advanced disease*. *J Natl Compr Canc Netw* 2010, 8(4):437-447.
4. Oh SC, Sohn BH, Cheong J-H, Kim S-B, Lee JE, Park KC, Lee SH, Park J-L, Park Y-Y, Lee H-S et al: *Clinical and genomic landscape of gastric cancer with a mesenchymal phenotype*. *Nat Commun* 2018, 9(1):1777.
5. Lech G, Słotwiński R, Słodkowski M, Krasnodębski IW: *Colorectal cancer tumour markers and biomarkers: Recent therapeutic advances*. *World J Gastroenterol* 2016, 22(5):1745-1755.
6. Seijo LM, Peled N, Ajona D, Boeri M, Field JK, Sozzi G, Pio R, Zulueta JJ, Spira A, Massion PP et al: *Biomarkers in Lung Cancer Screening: Achievements, Promises, and Challenges*. *J Thorac Oncol* 2019, 14(3):343-357.
7. Dasari S, Wudayagiri R, Valluru L: *Cervical cancer: Biomarkers for diagnosis and treatment*. *Clin Chim Acta* 2015, 445.
8. Wen P, Chidanguro T, Shi Z, Gu H, Wang N, Wang T, Li Y, Gao J: *Identification of candidate biomarkers and pathways associated with SCLC by bioinformatics analysis*. *Mol Med Rep* 2018, 18(2):1538-1550.
9. Xu Z, Zhou Y, Cao Y, Dinh TLA, Wan J, Zhao M: *Identification of candidate biomarkers and analysis of prognostic values in ovarian cancer by integrated...*
bioinformatics analysis. Med Oncol 2016, 33(11):130.

10. Zhou L, Tang H, Wang F, Chen L, Ou S, Wu T, Xu J, Guo K: Bioinformatics analyses of significant genes, related pathways and candidate prognostic biomarkers in glioblastoma. Mol Med Rep 2018, 18(5):4185-4196.

11. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT et al: Gene ontology: Tool for the unification of biology. Nature Genetics 2000, 25(1):25-29.

12. The Gene Ontology (GO) project in 2006. Nucleic Acids Res 2006, 34(Database issue):D322-D326.

13. Kanehisa M, Goto S: KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 2000, 28(1):27-30.

14. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP et al: STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2014, 43(D1):D447-D452.

15. Jiao X, Sherman BT, Huang DW, Stephens R, Baseler MW, Lane HC, Lempicki RA: DAVID-WS: a stateful web service to facilitate gene/protein list analysis. Bioinformatics 2012, 28(13):1805-1806.

16. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z: GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res 2017, 45(W1):W98-W102.

17. Ranstam J, Cook JA: Kaplan-Meier curve. Br J Surg 2017, 104(4):442.

18. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E et al: The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012, 2(5):401-404.

19. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincead-Beal C, Kulkarni P et al: Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia 2007, 9(2):166-180.

20. Chandrareshkar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravartti BVSK, Varambally S: UALCAN: A Portal for Facilitating Tumor
Subgroup Gene Expression and Survival Analyses. *Neoplasia* 2017, 19(8):649-658.

21. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, Zwahlen M, Kampf C, Wester K, Hober S et al: Towards a knowledge-based Human Protein Atlas. *Nature biotechnology* 2010, 28(12):1248-1250.

22. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B, Liu XS: TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res* 2017, 77(21):e108-e110.

23. Postow MA, Callahan MK, Wolchok JD: Immune Checkpoint Blockade in Cancer Therapy. *J Clin Oncol* 2015, 33(17):1974-1982.

24. Marin-Acevedo JA, Dholaria B, Soyano AE, Knutson KL, Chumsri S, Lou Y: Next generation of immune checkpoint therapy in cancer: new developments and challenges. *Journal of Hematology & Oncology* 2018, 11(1):39.

25. Wang J, Yuan R, Song W, Sun J, Liu D, Li Z: PD-1, PD-L1 (B7-H1) and Tumor-Site Immune Modulation Therapy: The Historical Perspective. *Journal of hematology & oncology* 2017, 10(1):39.

26. Topalian SL, Drake CG, Pardoll DM: Immune checkpoint blockade: a common denominator approach to cancer therapy. *Cancer Cell* 2015, 27(4):450-461.

27. Muro K, Chung HC, Shankaran V, Geva R, Catenacci D, Gupta S, Eder JP, Golant T, Le DT, Burtness B et al: Pembrolizumab for patients with PD-L1-positive advanced gastric cancer (KEYNOTE-012): a multicentre, open-label, phase 1b trial. *Lancet Oncol* 2016, 17(6):717-726.

28. Vassilakopoulou M, Avgeris M, Velcheti V, Kotoula V, Rampias T, Chatzopoulos K, Perisanidis C, Kontos CK, Giotakis AI, Scorilas A et al: Evaluation of PD-L1 Expression and Associated Tumor-Infiltrating Lymphocytes in Laryngeal Squamous Cell Carcinoma. *Clin Cancer Res* 2016, 22(3):704-713.

29. Koemans WJ, Chalabi M, van Sandick JW, van Dieren JM, Kodach LL: Beyond the PD-L1 horizon: In search for a good biomarker to predict success of immunotherapy in gastric and esophageal adenocarcinoma. *Cancer Letters* 2019, 442:279-286.

30. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 2014, 513(7517):202-209.

31. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, Skora AD, Luber
BS, Azad NS, Laheru D et al: PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N Engl J Med 2015, 372(26):2509-2520.

32. Thompson ED, Zahurak M, Murphy A, Cornish T, Cuka N, Abdelfatah E, Yang S, Duncan M, Ahuja N, Taube JM et al: Patterns of PD-L1 expression and CD8 T cell infiltration in gastric adenocarcinomas and associated immune stroma. Gut 2017, 66(5):794-801.

33. Knief J, Lazar-Karsten P, Hummel R, Wellner U, Thorns C: PD-L1 expression in carcinoma of the esophagogastric junction is positively correlated with T-cell infiltration and overall survival. Pathol Res Pract 2019, 215(6):152402.

34. Velcheti V, Schalper KA, Carvajal DE, Anagnostou VK, Syrigos KN, Sznol M, Herbst RS, Gettinger SN, Chen L, Rimm DL: Programmed death ligand-1 expression in non-small cell lung cancer. Lab Invest 2014, 94(1):107-116.

35. Liu X, Cao Y, Li R, Gu Y, Chen Y, Qi Y, Lv K, Wang J, Yu K, Lin C et al: Poor clinical outcomes of intratumoral dendritic cell-specific intercellular adhesion molecule 3-grabbing non-integrin-positive macrophages associated with immune evasion in gastric cancer. Eur J Cancer 2020, 128:27-37.

36. Zhang H, Li R, Cao Y, Gu Y, Lin C, Liu X, Lv K, He X, Fang H, Jin K et al: Poor Clinical Outcomes and Immunoevasive Contexture in Intratumoral IL-10-Producing Macrophages Enriched Gastric Cancer Patients. Ann Surg 2020.

37. Sengupta P, Xu Y, Wang L, Widom R, Smith BD: Collagen alpha1(I) gene (COL1A1) is repressed by RFX family. J Biol Chem 2005, 280(22):21004-21014.

38. Li J, Ding Y, Li A: Identification of COL1A1 and COL1A2 as candidate prognostic factors in gastric cancer. World J Surg Oncol 2016, 14(1):297.

39. Sun R, Wu J, Chen Y, Lu M, Zhang S, Lu D, Li Y: Down regulation of Thrombospondin2 predicts poor prognosis in patients with gastric cancer. Mol Cancer 2014, 13:225.

40. Kim H, Watkinson J, Varadan V, Anastassiou D: Multi-cancer computational analysis reveals invasion-associated variant of desmoplastic reaction involving INHBA, THBS2 and COL11A1. BMC Med Genomics 2010, 3:51.

41. Barcus CE, O’Leary KA, Brockman JL, Rugowski DE, Liu Y, Garcia N, Yu M, Keely PJ, Eliceiri KW, Schuler LA: Elevated collagen-I augments tumor progressive signals, intravasation and metastasis of prolactin-induced estrogen receptor alpha positive
mammary tumor cells. *Breast Cancer Res* 2017, 19(1):9.

42. Bornstein P: *Thrombospondins function as regulators of angiogenesis. J Cell Commun Signal* 2009, 3(3-4):189-200.

43. Kupcova Skalnikova H, Cizkova J, Cervenka J, Vodicka P: *Advances in Proteomic Techniques for Cytokine Analysis: Focus on Melanoma Research. Int J Mol Sci* 2017, 18(12).

44. Song S-Z, Lin S, Liu J-N, Zhang M-B, Du Y-T, Zhang D-D, Xu W-H, Wang H-B: Targeting of SPP1 by microRNA-340 inhibits gastric cancer cell epithelial-mesenchymal transition through inhibition of the PI3K/AKT signaling pathway. *J Cell Physiol* 2019, 234(10):18587-18601.
| Analysis Type by Cancer     | Cancer vs. Normal | Cancer vs. Normal | Cancer vs. Normal |
|-----------------------------|-------------------|-------------------|-------------------|
|                             | COL1A1            | THBS2             | SPP1              |
| Bladder Cancer              | 1                 | 1                 | 1                 |
| Brain and CNS Cancer        | 6                 | 1                 | 2                 |
| Breast Cancer               | 15                | 5                 | 8                 |
| Cervical Cancer             | 4                 | 3                 | 3                 |
| Colorectal Cancer           | 11                | 12                | 10                |
| Esophageal Cancer           | 4                 | 3                 | 3                 |
| Gastric Cancer              | 8                 | 9                 | 4                 |
| Head and Neck Cancer        | 3                 | 5                 | 8                 |
| Kidney Cancer               | 4                 | 4                 | 1                 |
| Leukemia                    | -                 | -                 | -                 |
| Liver Cancer                | 4                 | 3                 | 5                 |
| Lung Cancer                 | 6                 | 9                 | 12                |
| Lymphoma                    | 7                 | 4                 | 4                 |
| Melanoma                    | -                 | 1                 | 1                 |
| Myeloma                     | -                 | -                 | 1                 |
| Other Cancer                | 7                 | 3                 | 3                 |
| Ovarian Cancer              | 2                 | 1                 | 2                 |
| Pancreatic Cancer           | 2                 | 5                 | 1                 |
| Prostate Cancer             | -                 | -                 | 1                 |
| Sarcoma                     | 7                 | -                 | 1                 |
| **Significant Unique Analyses** | **86**          | **62**            | **72**            |
| **Total Unique Analyses**   | **400**           | **447**           | **431**           |
Expression of CD274 in STAD based on Sample types

A) TCGA samples

B) Col1A1 ~ CD274
Pearson's rho
r = 0.1263 (p = 0.007324)
Spearman's rank rho
rho = 0.1403 (p = 0.002862)

C) THBS2 ~ CD274
Pearson's rho
r = 0.1603 (p = 0.0006448)
Spearman's rank rho
rho = 0.1510 (p = 0.001313)

D) SPP1 ~ CD274
Pearson's rho
r = 0.1097 (p = 0.01988)
Spearman's rank rho
rho = 0.1775 (p = 0.0001541)
| Category             | Term                                                                 | Count | PValue         |
|----------------------|----------------------------------------------------------------------|-------|----------------|
| GOTERM_BP_DIRECT     | GO:0007155-cell adhesion                                            | 5     | 6.23E-05       |
| GOTERM_BP_DIRECT     | GO:0030198-extracellular matrix organization                         | 4     | 1.25E-04       |
| GOTERM_BP_DIRECT     | GO:0022617-extracellular matrix disassembly                          | 3     | 7.13E-04       |
| GOTERM_BP_DIRECT     | GO:0001525-angiogenesis                                             | 3     | 0.00594        |
| GOTERM_BP_DIRECT     | GO:004344-cellular response to fibroblast growth factor stimulus     | 2     | 0.01597        |
| GOTERM_BP_DIRECT     | GO:0043434-response to peptide hormone                               | 2     | 0.02334        |
| GOTERM_BP_DIRECT     | GO:0030574-collagen catabolic process                                | 2     | 0.03379        |
| GOTERM_BP_DIRECT     | GO:0030336-negative regulation of cell migration                     | 2     | 0.04979        |
| GOTERM_CC_DIRECT     | GO:0005576-extracellular region                                      | 8     | 1.27E-06       |
| GOTERM_CC_DIRECT     | GO:0005615-extracellular space                                      | 6     | 2.14E-04       |
| GOTERM_CC_DIRECT     | GO:0031012-extracellular matrix                                     | 4     | 3.31E-04       |
| GOTERM_CC_DIRECT     | GO:0070062-extracellular exosome                                     | 6     | 0.00636        |
| GOTERM_CC_DIRECT     | GO:0005578-proteinaceous extracellular matrix                        | 3     | 0.00725        |
| GOTERM_CC_DIRECT     | GO:0031093-platelet alpha granule lumen                              | 2     | 0.02684        |
| GOTERM_CC_DIRECT     | GO:0005604-basement membrane                                        | 2     | 0.03835        |
| GOTERM_CC_DIRECT     | GO:0005581-collagen trimer                                           | 2     | 0.04454        |
| GOTERM_MF_DIRECT     | GO:0050840-extracellular matrix binding                              | 2     | 0.01378        |
| GOTERM_MF_DIRECT     | GO:0005515-protein binding                                          | 9     | 0.02599        |
| KEGG_PATHWAY         | hsa04512:ECM-receptor interaction                                    | 3     | 0.00421        |
| KEGG_PATHWAY         | hsa04510:Focal adhesion                                             | 3     | 0.02219        |