Comprehensive analysis reveals CTHRC1, SERPINE1, VCAN and UPK1B as the novel prognostic markers in gastric cancer

Zhipeng Zhu\(^1\), Jiuhua Xu\(^2\), Lulu Li\(^1\), Weipeng Ye\(^2\), Borong Chen\(^1\), Junjie Zeng\(^1\), Zhengjie Huang\(^1,2\)

\(^1\)Department of Gastrointestinal Surgery, Xiamen Cancer Center, The First Affiliated Hospital of Xiamen University, Xiamen, China; \(^2\)Department of clinical medicine, Fujian Medical University, Fuzhou, China

Contributions: (I) Conception and design: Z Zhu; (II) Administrative support: Z Huang; (III) Provision of study materials or patients: Z Zhu; (IV) Collection and assembly of data: L Li, J Xu; (V) Data analysis and interpretation: W Ye, B Chen, J Zeng; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

\(^\ast\)These authors contributed equally to this work.

Correspondence to: Dr. Zhengjie Huang. Department of Gastrointestinal Surgery, Xiamen Cancer Center of The First Affiliated Hospital of Xiamen University, 55 Zhen Hai Road, Si Ming District, Xiamen, China. Email: huangzhengjie@xmu.edu.cn.

Background: Gastric cancer (GC) is one of the most common malignant diseases worldwide, the incidence and mortality for GC is still high, thus it is urgently important to identify the effective and reliable biomarkers to evaluate GC and the underlying molecular events.

Methods: The study integrated four Gene Expression Omnibus (GEO) profile datasets and The Cancer Genome Atlas (TCGA) dataset to screen differentially expressed genes (DEGs), screened key genes by performing the Kaplan-Meier analysis, univariate and multivariate-cox analysis. Further analysis were performed to evaluate and validate the prognostic value of the key genes based on TCGA database and online websites. In addition, mechanism analysis of the key genes was performed thought biological processes and KEGG pathway analysis.

Results: In the study, 192 DEGs (92 up-regulated and 100 down-regulated) were identified from the GEO and TCGA datasets. Next, gene ontology (GO) for DEGs focused primarily on cell adhesion, extracellular region and extracellular matrix structural constituent. Then four significant key genes were screened by performed the Kaplan-Meier analysis, univariate and multivariate-cox analysis. By using Kaplan-Meier plotter and OncoLnc, the expression level was associated with a worse prognosis. In addition, the area under curve (AUC) for time-dependent receiver operating characteristic (ROC) indicated a moderate diagnostic value. Furthermore, the expression of collagen triple helix repeat containing 1 (CTHRC1), serpin family E member 1 (SERPINE1), Versican (VCAN) was associated with tumor size, Uroplakin 1B (UPK1B) expression was associated with distant metastasis. Finally, multiple biological processes and signaling pathway associated with key genes revealed the underlying mechanism in GC.

Conclusions: Taken together, CTHRC1, SERPINE1, VCAN, UPK1B were novel potential prognostic molecular markers for GC, which acted as oncogene to promote the development of GC.

Keywords: Gastric cancer (GC); data mining; prognostic markers

Submitted Jan 06, 2020. Accepted for publication Jun 05, 2020.
doi: 10.21037/tcr-20-211

View this article at: http://dx.doi.org/10.21037/tcr-20-211

Introduction

The incidence and mortality for gastric cancer (GC) have been appreciably declining for several decades. However, GC is still the fourth most common cancer and the second leading cause of cancer deaths worldwide (1-3). In China alone, there were about 679/100,000 of new GC cases and 798/100,000 of death GC cases and accounting for
the third of malignant tumor incidence and mortality in 2015 (4). The pathogenesis of GC is multifactorial, including genetic susceptibility and environmental factors, cell cycle, DNA repair, metabolism, cell-to-cell and cell-to-matrix interactions, apoptosis, angiogenesis, and immune surveillance contribute to cancer development (5). However, although there have been extensive previous studies on the molecular mechanism of GC formation and progression, the molecular mechanism of GC is not yet clear. Due to high morbidity and mortality in GC, it is urgently important to reveal the causes and the underlying molecular mechanisms. Thus, identifying novel diagnostic and prognostic biomarkers remains critical importance for stomach cancer.

In this work, we have downloaded four original microarray datasets GSE79973 (6), GSE63089 (7), GSE54129 (8), GSE63089 (9), from NCBI-Gene Expression Omnibus database (NCBI-GEO), there are total of 262 GC cases and 88 normal cases available. differentially expressed genes (DEGs) between cancer tissues and normal tissues were obtained from GEO and TCGA gene expression profile, respectively. Then, we overlapped the four GEO and TCGA gene expression data according to the gene symbol. Next, gene differential expression analysis between normal cases and tumor cases was performed by using limma package in the Bioconductor package from GEO and TCGA gene expression profile, with corrected P value <0.05 and absolute log fold change (FC) >1 being considered as the cutoff criterion. Finally, Gene set enrichment analysis (GSEA) was performed to further investigate pathways of four key genes that may be associated with GC.

Methods

Identification and processing of microarray data

We used the “GC” OR “gastric carcinoma” keyword to search gene expression profiles from GEO database (http://www.ncbi.nlm.nih.gov/geo/), and four qualified gene expression profiles (GSE54129, GSE79973, GSE63089, GSE26899) were identified with platform and series matrix file(s) being downloaded as TXT files, type of data were RMA signal intensity and standardized, and log2 transformed. The dataset information was presented in Table 1.

Identification of DEGs and overlapped genes

R annotation package was performed to convert the probe into gene symbol. Next, SVA package was used for background correction, merge package was applied to combine the four gene expression data according to the gene symbol. Then, gene differential expression analysis between normal cases and tumor cases was performed by using limma package in the Bioconductor package from GEO and TCGA gene expression profile, with corrected P value <0.05 and absolute log fold change (FC) >1 being considered as the cutoff criterion. Finally, overlapped genes were identified from four GEO and TCGA gene expression profiles.

Overlapped genes enrichment analysis

The DAVID database (https://david.ncifcrf.gov/) is an essential foundation for the success of any high-throughput gene function analysis. We used DAVID to perform GO annotations analysis on overlapped genes.
Identification and validation of clinically relevant hub genes

The Kaplan-Meier analysis was performed to screen the survival-related genes, univariate and multivariate-cox analysis for OS was conducted to identify the key genes from the survival-related genes. To evaluate and validate the prognostic value of the key genes, we performed the Kaplan-Meier analysis based on the online website including Kaplan-Meier plotter (http://kmplot.com/) and OncoLnc (http://www.oncolnc.org/). The Kaplan-Meier analysis for disease free survival (DFS) based on TCGA dataset, univariate and multivariate-cox analysis for DFS by mining TCGA dataset, the receiver operating characteristic (ROC) analysis for OS and DFS, the correlation analysis between TMN and expression of key genes based on TCGA data. The gene expression level ≤ median was regarded as low expression, otherwise was regarded as high expression.

Biological processes and signaling pathway analysis for the co-expressed genes associated with GC

To explore the potential mechanisms for the key genes, we identified the co-expressed genes associated with key genes by using Coexpedia (http://www.coexpedia.org/), biological processes and KEGG-signaling pathway for the co-expressed genes associated with GC were predicted by R software.

Gene set enrichment analysis

To further investigate pathways of four key genes that may be associated with GC, GSEA was performed using the JAVA program (http://www.broadinstitute.org/gsea) with TCGA dataset. Expression of each key gene was set to annotated phenotypes, 1,000 times were performed for gene set permutations. The nominal P value <0.05 was used to sort the pathways enriched in each phenotype.

Results

The DEG of GEO gene expression profiles

We performed background correction on the GEO expression profiles. The result was shown in Figure 1. Then, we analyzed the DEGs of integrated GEO and TCGA gene expression profiles by using the limma package (FDR <0.05, absolute log FC >1), 219 up-regulated genes and 179 down-regulated genes were obtained from GSE26899, GSE54129, GSE63089 and GSE79973, 1,110 up-regulated genes and 1566 down-regulated genes were obtained from TCGA dataset. After using Venny, 92 up-regulated genes and 100 down-regulated genes were overlapped across four GEO and TCGA datasets (Figure 2).

GO and KEGG enrichment analysis

Enrichment analysis of the overlapped genes was performed using the DAVID online site (corrected P value<0.05). The enrichment analysis was divided into three functional groups, including biological processes, cell composition and molecular function, biological processes. In the biological processes group, the differential genes were mainly enriched in cell adhesion and biological adhesion. In the cell composition, the differential genes were mainly enriched in the extracellular region and the extracellular region part. In the molecular function, the differential genes were mainly enriched in the extracellular matrix structural constituent and pattern binding (Figure 3).

Identification of four key genes from overlapped genes

Twenty-three survival-related genes were identified by performing the Kaplan-Meier analysis, and high expression level was associated with a poorer OS (Table S1). Then, we identified four significant key genes by conducting univariate and multivariate-Cox analysis for OS, including CTHRC1, SERPINE1, UPK1B, VCAN, with HR >1 (P<0.05) (Figure 4).

Prognostic significance for the four genes

The gene expression of the cancer group was higher than the normal group from TCGA dataset for CTHRC1 (Figure 5A), SERPINE1 (Figure 5B), UPK1B (Figure 5C) and VCAN (Figure 5D). Meantime, the gene expression of the cancer group was higher than paracancerous group for CTHRC1 (Figure 5E), SERPINE1 (Figure 5F), UPK1B (Figure 5G) and VCAN (Figure 5H). By using OncoLnc, it indicated high gene expression was significantly associated with a shorter OS (Figure 6A,B,C,D). Then Kaplan Meier plotter revealed the same trend, high expression presented worse OS (Figure 6E,F,G,H), first progression (FP) (Figure 6I,J,K,L) and post progression survival (PPS) (Figure 6M,N,O,P). Next, the ROC analysis of four key genes was performed to evaluate the diagnostic value of four key genes for OS, as showed in Figure 7, all the AUC indicated a moderate diagnostic...
Figure 1 Standardization of gene expression. (A) The standardization of GSE26899 data, (B) the standardization of GSE54129 data, and (C) the standardization of GSE63089 data. (D) the standardization of GSE79973 data. The green bar represents the data before normalization, and the red bar represents the normalized data.
Figure 2 Venn plot of the DEGs between the integrated four GEO datasets and the TCGA dataset. DEGs, differentially expressed genes; GEO, Gene Expression Omnibus; TCGA, The Cancer Genome Atlas.

Figure 3 GO enrichment analysis of overlapped genes into three functional groups: molecular function, biological processes, and cell composition. GO, gene ontology.
| Factors                | Univariate Cox-regression analysis | Multivariate Cox-regression analysis |
|------------------------|------------------------------------|--------------------------------------|
|                        | HR   | 95%CI   | P-value | HR     | 95%CI    | P-value   |
| Age                    | 1.028 | 1.0091−1.0473 | 0.0036 | 1.0347 | 1.0148−1.055 | 0.0006    |
| Gender                 | 1.4923 | 0.9935−2.2415 | 0.0538 |       |          |           |
| Grade                  | 1.4189 | 0.9873−2.0394 | 0.0587 |       |          |           |
| Tumor size (cm)        | 1.2535 | 0.9969−1.576 | 0.0531 |       |          |           |
| Metastasis             | 2.0319 | 1.0891−3.7906 | 0.0259 | 2.6655 | 1.3957−5.0906 | 0.003     |
| Lymph node             | 1.2851 | 1.0880−1.5178 | 0.0031 | 1.306  | 1.1026−1.547 | 0.002     |
| CTHRC1 (high/low)      | 1.6083 | 1.1130−2.324 | 0.0113 | 1.6199 | 1.1143−2.3548 | 0.0115    |

| Factors                | Univariate Cox-regression analysis | Multivariate Cox-regression analysis |
|------------------------|------------------------------------|--------------------------------------|
|                        | HR   | 95%CI   | P-value | HR     | 95%CI    | P-value   |
| Age                    | 1.028 | 1.0091−1.0473 | 0.0036 | 1.0345 | 1.0144−1.0549 | 0.0007    |
| Gender                 | 1.4923 | 0.9935−2.2415 | 0.0538 |       |          |           |
| Grade                  | 1.4189 | 0.9873−2.0394 | 0.0587 |       |          |           |
| Tumor size (cm)        | 1.2535 | 0.9969−1.576 | 0.0531 |       |          |           |
| Metastasis             | 2.0319 | 1.0891−3.7906 | 0.0259 | 2.6205 | 1.3733−5.0004 | 0.0035    |
| Lymph node             | 1.2851 | 1.0880−1.5178 | 0.0031 | 1.3356 | 1.128−1.5814 | 0.0008    |
| SERPINE1 (high/low)    | 1.7616 | 1.2142−2.5557 | 0.029  | 1.5497 | 1.0682−2.2483 | 0.021     |

| Factors                | Univariate Cox-regression analysis | Multivariate Cox-regression analysis |
|------------------------|------------------------------------|--------------------------------------|
|                        | HR   | 95%CI   | P-value | HR     | 95%CI    | P-value   |
| Age                    | 1.028 | 1.0091−1.0473 | 0.0036 | 1.0381 | 1.0181−1.0585 | 0.0002    |
| Gender                 | 1.4923 | 0.9935−2.2415 | 0.0538 |       |          |           |
| Grade                  | 1.4189 | 0.9873−2.0394 | 0.0587 |       |          |           |
| Tumor size (cm)        | 1.2535 | 0.9969−1.576 | 0.0531 |       |          |           |
| Metastasis             | 2.0319 | 1.0891−3.7906 | 0.0259 | 2.4464 | 1.2849−4.6577 | 0.0065    |
| Lymph node             | 1.2851 | 1.0880−1.5178 | 0.0031 | 1.3091 | 1.1048−1.5511 | 0.0019    |
| UPK1B (high/low)       | 1.8657 | 1.286−2.7068 | 0.001  | 1.8243 | 1.2555−2.6509 | 0.0016    |

| Factors                | Univariate Cox-regression analysis | Multivariate Cox-regression analysis |
|------------------------|------------------------------------|--------------------------------------|
|                        | HR   | 95%CI   | P-value | HR     | 95%CI    | P-value   |
| Age                    | 1.028 | 1.0091−1.0473 | 0.0036 | 1.0378 | 1.0176−1.0585 | 0.0002    |
| Gender                 | 1.4923 | 0.9935−2.2415 | 0.0538 |       |          |           |
| Grade                  | 1.4189 | 0.9873−2.0394 | 0.0587 |       |          |           |
| Tumor size (cm)        | 1.2535 | 0.9969−1.576 | 0.0531 |       |          |           |
| Metastasis             | 2.0319 | 1.0891−3.7906 | 0.0259 | 2.9    | 1.5128−5.593 | 0.0013    |
| Lymph node             | 1.2851 | 1.0880−1.5178 | 0.0031 | 1.3103 | 1.106−1.5523 | 0.0018    |
| VCAN (high/low)        | 1.7418 | 1.2017−2.5246 | 0.0034 | 1.8005 | 1.2372−2.6205 | 0.0021    |

Figure 4 Univariate and multivariate analysis of clinicopathologic characteristics and key genes for OS. (A) CTHRC1 (B) SERPINE1 (C) UPK1B (D) VCAN. OS, overall survival; CTHRC1, collagen triple helix repeat containing 1; SERPINE1, plasminogen activator inhibitor type 1; UPK1B, uroplakin Ib; VCAN, Verscan.
**Figure 5** Comparison of the key gene expression. The different expression level in normal tissue and GC (A-D), (A) CTHRC1, (B) SERPINE1, (C) UPK1B, (D) VCAN. The different expression level in paracancerous tissue and GC (E-H), (E) CTHRC1, (F) SERPINE1, (G) UPK1B, (H) VCAN. CTHRC1, collagen triple helix repeat containing 1; SERPINE1, plasminogen activator inhibitor type 1; UPK1B, uroplakin Ib; VCAN, Verscan.
Figure 6 The performance analysis using OncoLnc (A-D) and Kaplan-Meier Plotter (E-H) OS, (I-L) FP, (M-P) PPS. (A,E,I,M) CTHRC1, (B,F,J,N) SERPINE1, (C,G,K,O) UPK1B, (D,H,L,P) VCAN. CTHRC1, collagen triple helix repeat containing 1; SERPINE1, plasminogen activator inhibitor type 1; UPK1B, uroplakin Ib; VCAN, Verscan; GC, gastric cancer.
value (CTHRC1: 0.772, SERPINE1: 0.702, UPK1B: 0.691, VCAN: 0.759). Furthermore, patients with high expression level have poorer DFS than the patients with low expression level (P<0.05, Figure 8A,B,C,D). The ROC curve for DFS demonstrated that CTHRC1, SERPINE1, UPK1B and VCAN were specific and sensitive than any clinical characteristics, including age, gender, grade, tumor size, lymph node and metastasis (Figure 8E,F,G,H). In addition, univariate and multivariate-Cox analysis for DFS displayed four key genes were all powerful and independent factors for DFS (Figure 9). Finally, correlation analysis between TMN and expression of key genes was analyzed by performing Mann-Whitney-Wilcoxon Test based on TCGA data, it revealed that gene expression was associated with tumor stage, including CTHRC1, SERPINE1, VCAN. Meantime, UPK1B expression was associated with distant metastasis (Figure 10).

**Biological processes and signaling pathway analysis for the co-expressed genes associated with GC**

We identified the co-expressed genes associated with key genes in GC. In addition, the biological processes and signaling pathway analysis of key genes in GC were investigated. These co-expressed genes were involved in a variety of biological processes, such as endodermal cell
Figure 8 The performance analysis for DFS in GC. The ROC curve for DFS in GC (A-D), Kaplan-Meier plotter for DFS in GC (E-H). (A,E) CTHRC1, (B,F) SERPINE1, (C,G) UPK1B, (D,H) VCAN. CTHRC1, collagen triple helix repeat containing 1; SERPINE1, plasminogen activator inhibitor type 1; UPK1B, uroplakin Ib; VCAN, Verscan; GC, gastric cancer.
### Figure 9

Univariate and multivariate analysis of clinicopathologic characteristics and key genes for DFS.

#### Table A

| Factors                  | Univariate Cox-regression analysis | Multivariate Cox-regression analysis |
|--------------------------|-----------------------------------|--------------------------------------|
|                          | HR  | 95%CI                  | P-value | HR  | 95%CI                  | P-value |
| Age                      | 0.9876 | 0.9687−1.0068            | 0.2049  | 2.0844 | 1.2767−3.4032          | 0.0033  |
| Gender                   | 2.154 | 1.3217−3.5103            | 0.0021  | 2.0844 | 1.2767−3.4032          | 0.0033  |
| Grade                    | 1.3957 | 0.9248−2.1066            | 0.1124  | 2.154 | 1.3217−3.5103          | 0.0021  |
| Tumor size (cm)          | 1.1034 | 0.8613−1.4136            | 0.4362  | 1.3957 | 0.9248−2.1066          | 0.1124  |
| Metastasis               | 1.4179 | 0.6182−3.2525            | 0.4097  | 1.1034 | 0.8613−1.4136          | 0.4362  |
| Lymph node               | 1.3449 | 1.1145−1.6228            | 0.002   | 1.3449 | 1.1145−1.6228          | 0.002   |
| CTHRC1 (high/low)        | 1.4842 | 1.2726−2.2648            | 0.0047  | 1.3449 | 1.1145−1.6228          | 0.002   |

#### Table B

| Factors                  | Univariate Cox-regression analysis | Multivariate Cox-regression analysis |
|--------------------------|-----------------------------------|--------------------------------------|
|                          | HR  | 95%CI                  | P-value | HR  | 95%CI                  | P-value |
| Age                      | 0.9876 | 0.9687−1.0068            | 0.2049  | 2.0911 | 1.23−3.2817            | 0.0053  |
| Gender                   | 2.154 | 1.3217−3.5103            | 0.0021  | 2.154 | 1.3217−3.5103          | 0.0021  |
| Grade                    | 1.3957 | 0.9248−2.1066            | 0.1124  | 1.3957 | 0.9248−2.1066          | 0.1124  |
| Tumor size (cm)          | 1.1034 | 0.8613−1.4136            | 0.4362  | 1.1034 | 0.8613−1.4136          | 0.4362  |
| Metastasis               | 1.4179 | 0.6182−3.2525            | 0.4097  | 1.4179 | 0.6182−3.2525          | 0.4097  |
| Lymph node               | 1.3449 | 1.1145−1.6228            | 0.002   | 1.3449 | 1.1145−1.6228          | 0.002   |
| SERPINE1 (high/low)      | 2.1124 | 1.3793−2.335             | 0.0066  | 2.1124 | 1.3793−2.335           | 0.0066  |

#### Table C

| Factors                  | Univariate Cox-regression analysis | Multivariate Cox-regression analysis |
|--------------------------|-----------------------------------|--------------------------------------|
|                          | HR  | 95%CI                  | P-value | HR  | 95%CI                  | P-value |
| Age                      | 0.9876 | 0.9687−1.0068            | 0.2049  | 2.1743 | 1.3335−3.5451          | 0.0018  |
| Gender                   | 2.154 | 1.3217−3.5103            | 0.0021  | 2.154 | 1.3217−3.5103          | 0.0021  |
| Grade                    | 1.3957 | 0.9248−2.1066            | 0.1124  | 1.3957 | 0.9248−2.1066          | 0.1124  |
| Tumor size (cm)          | 1.1034 | 0.8613−1.4136            | 0.4362  | 1.1034 | 0.8613−1.4136          | 0.4362  |
| Metastasis               | 1.4179 | 0.6182−3.2525            | 0.4097  | 1.4179 | 0.6182−3.2525          | 0.4097  |
| Lymph node               | 1.3449 | 1.1145−1.6228            | 0.002   | 1.3449 | 1.1145−1.6228          | 0.002   |
| UPK1B (high/low)         | 1.7957 | 1.1783−2.7366            | 0.0065  | 1.7957 | 1.1783−2.7366          | 0.0065  |

#### Table D

| Factors                  | Univariate Cox-regression analysis | Multivariate Cox-regression analysis |
|--------------------------|-----------------------------------|--------------------------------------|
|                          | HR  | 95%CI                  | P-value | HR  | 95%CI                  | P-value |
| Age                      | 0.9876 | 0.9687−1.0068            | 0.2049  | 2.0358 | 1.245−3.329            | 0.0046  |
| Gender                   | 2.154 | 1.3217−3.5103            | 0.0021  | 2.154 | 1.3217−3.5103          | 0.0021  |
| Grade                    | 1.3957 | 0.9248−2.1066            | 0.1124  | 1.3957 | 0.9248−2.1066          | 0.1124  |
| Tumor size (cm)          | 1.1034 | 0.8613−1.4136            | 0.4362  | 1.1034 | 0.8613−1.4136          | 0.4362  |
| Metastasis               | 1.4179 | 0.6182−3.2525            | 0.4097  | 1.4179 | 0.6182−3.2525          | 0.4097  |
| Lymph node               | 1.3449 | 1.1145−1.6228            | 0.002   | 1.3449 | 1.1145−1.6228          | 0.002   |
| VCAN (high/low)          | 1.6653 | 1.0981−2.5256            | 0.0164  | 1.6653 | 1.0981−2.5256          | 0.0164  |

DFS, disease free survival; CTHRC1, collagen triple helix repeat containing 1; SERPINE1, plasminogen activator inhibitor type 1; UPK1B, uroplakin IIs; VCAN, Verscan.
differentiation, endoderm development, and extracellular matrix organization for CTHRC1 (Figure 11A), regulation of angiogenesis, positive regulation of leukocyte chemotaxis cellular, and regulation of vasculature development for SERPINE1 (Figure 11B), extracellular matrix organization, collagen fibril organization, collagen metabolic process, and endodermal cell differentiation for UPK1B, extracellular matrix organization, and cellular response to transforming growth factor beta stimulus for VCAN. These co-expressed genes were involved in a variety of biological processes, such as ECM-receptor interaction, AGE-RAGE signaling pathway in diabetic complications, PI3K-Akt signaling pathway for CTHRC1 (Figure 12A), such as NF-kappa B signaling pathway, PI3K-Akt signaling pathway, Toll-like receptor signaling pathway for SERPINE1 (Figure 12B), such as ECM-receptor interaction, PI3K-Akt signaling pathway, relaxin signaling pathway for UPK1B (Figure 12C), such as ECM-receptor interaction, PI3K-Akt signaling pathway for VCAN (Figure 12D).

GSEA identifies prognostic genes-related signaling pathway

In order to further explore the mechanism of prognostic genes in patients with GC, we conducted GSEA between low and high expression group to identify the significant pathways (FDR < 0.05, NOM P value < 0.05). For CTHRC1, some significant pathways which were active in the high-expression group, including KEGG_ECM_RECEPTOR_INTERACTION, KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION, KEGG_TGF_BETA_SIGNALING_PATHWAY, KEGG_PATHWAYS_IN_CANCER, KEGG_FOCAL_ADHESION. Several significant pathways which were active in the low-risk group, including KEGG_PROPAANOATE_METABOLISM, KEGG_CITRATE_CYCLE_TCA_CYCLE, KEGG_BETA_ALANINE_METABOLISM, KEGG_LONG_TERM_POTENTIATION, KEGG_LINOLEIC_ACID_METABOLISM (Figure 13A). The
Figure 11 Potential biological processes for the key genes in GC. GC, gastric cancer.
most significant pathways were presented for SERPINE1 (Figure 13B), UPK1B (Figure 13C) and VCAN (Figure 13D).

**Discussion**

GC is one most common malignant cancer worldwide and it is very difficult to treat the advanced-stage SC. Although the formation, progression and underlying mechanisms for GC have been revealed from some basic and clinical studies, the incidence and mortality of GC is still very high worldwide (10). Therefore, it is necessary to identify novel prognostic and therapeutic target for GC.

UPK1B is a structural protein on the surface of urothelial cells1, which was considered as the entirely specific for urothelium, recent studies have indicated that UPK1B also expressed in other tissues, including bladder, brain, eye, kidney, lung, stomach (11). UPK1B may promote the occurrence and development of cancer (12,13), UPK1B could promote the proliferation, invasion and metastasis in bladder cancer (14,15). Su et al. (16) showed abnormal expression of UPK1B in various types of cancers. However, the role of UPK1B in GC has not been reported. In this study, the different expression level between normal and cancer is of significance, high gene expression was significantly associated with a shorter OS, UPK1B is significant diagnostic factor in GC. The expression level was associated with distant metastasis, UPK1B may participate in the biological processes (extracellular matrix organization, collagen fibril organization, collagen metabolic process, endodermal cell differentiation) through ECM-receptor interaction, PI3K-Akt signaling pathway, relaxin signaling pathway to promote the metastasis in GC.
Figure 13 Enrichment plots from GSEA for (A) CTHRC1, (B) SERPINE1, (C) UPK1B, (D) VCAN. GSEA, gene set enrichment analysis; CTHRC1, collagen triple helix repeat containing 1; SERPINE1, plasminogen activator inhibitor type 1; UPK1B, uroplakin Ib; VCAN, Verscan.

The CTHRC1 gene belongs to chromosome 8q22.3, which encoded a protein to participate in the vascularity and bone formation and so on (17). The expression level was different between normal tissue and tumor tissue for some types of tumors, including breast cancer (18), cervical cancer (19), colorectal cancer (20), liver cancer (21) and GC (22), the aberrant expression level was associated with poor OS and progression-free survival and it was the independent prognostic marker in GC in GC, which was consistent with our result. Recently, Ding et al. (23) have reported that HIF-1α/CXCR4 signaling may be involved in the migration and invasion in GC, however, the underlying molecular mechanism for CTHRC1 promoting the occurrence and development of GC is not very clear. In this study, we identify several signaling pathway which may be involved in the occurrence and development of GC.

SERPINE1 gene encodes plasminogen activator inhibitor type 1, which participated in inhibiting tissue plasminogen activator and uridylyl phosphate adenosine, the aberrant expression in many types of cancer and SERPINE1 could be an independent risk factor for various types of cancers, including head and neck cancer (24,25), esophageal cancer (26), bladder cancer (27), melanoma (28). Li et al. (29) indicated SERPINE1 is a poor prognosis for GC, and SERPINE1 could promote tumour cell proliferation, migration, and invasion by regulating EMT. However, SERPINE1 still remains largely unknown in GC. In our study, SERPINE1 was an significant diagnostic factor in GC, and we found
the expression level of SERPINE1 was associated with depth of invasion, the potential signal pathways may participated in the biological process including NF-kappa B signaling pathway, PI3K-Akt signaling pathway, Toll-like receptor signaling pathway.

VCAN is a chondroitin sulfate proteoglycan, a member of the aggregating chondroitin sulfate PGs family, which is an important component of ECM (30). Verscan expression often occurs in the context of tissue remodeling, angiogenesis, including: follicular growth (31), inflammation (32), wound healing (33) or atherosclerotic lesions (34), and environmental significance around progressive tumors (35). It has been previously reported that tumor stromal cells play an important role in tumor formation and tumor progression (36), and VCAN is expressed and secreted by tumor stromal cells. Yeung et al. indicated that CAF-specific VCAN was up-regulated by TGF-β signal to promote tumorigenesis and invasion in ovarian cancer (37). The level of VCAN increased in many patients with malignant tumors includes colon cancer (38), rectal cancer (39), melanoma (40), odontogenic cancer (41), and ovarian cancer (42). In vitro and in vivo research, it has shown that VCAN can promote the proliferation, metastasis and invasion of cancer cells (43-45), with playing an important role in the formation of extracellular matrices that support tumor growth and metastasis. Shen et al. (46) reported that VCAN expression can be used as a prognostic indicator for GC patients, VCAN expression is higher in cancer tissues than in adjacent tissues, and could promote proliferation and invasion in GC cells. However, few literatures mentioned VCAN associated signaling pathways that promote the development of GC. We identified some signaling pathways that may be involved in the development of GC. This regulatory mechanism needs to be further elucidated.

Conclusions

In conclusion, by integrating four GEO and TCGA gene expression profile datasets, we identified four key genes (CTHRC1, SERPINE1, VCAN, UPK1B) which might as the novel potential prognostic molecular markers for GC. The four key genes have high prognostic performance, and could considered as independent prognostic factors for OS and DFS in GC. The four key genes act as oncogene to promote the development of GC, CTHRC1 participated in endodermal cell differentiation, extracellular matrix organization, SERPINE1 participated in regulation of angiogenesis, positive regulation of leukocyte chemotaxis cellular, regulation of vasculature development, UPK1B participated in extracellular matrix organization, collagen fibril organization, collagen metabolic process, endodermal cell differentiation, VCAN participated in extracellular matrix organization, cellular response to transforming growth factor beta stimulus. The study would provide some novel genes for the future prognosis prediction and potential molecular targeting therapy for GC.

However, further biological experiments should be performed to validate our results.

Acknowledgments

We are grateful to the reviewers for their constructive comments which led to improvements in this manuscript. In addition, we thank Bin Zhao (Official Wechat Account: Bio_Med2017) of Xiamen University for suggestions on the manuscripts.

Funding: This study was supported by Xiamen Scientific and Technological Plan (No. 3502Z20194005).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/tcr-20-211). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license).

References

1. Bertuccio P, Chatenoud L, Levi F, et al. Recent patterns
2. Ferro A, Peleteiro B, Malvezzi M, et al. Worldwide trends in gastric cancer mortality (1980-2011), with predictions to 2015, and incidence by subtype. Eur J Cancer 2014;50:1330-44.

3. Peleteiro B, Severo M, La Vecchia C, et al. Model-based patterns in stomach cancer mortality worldwide. Eur J Cancer Prev 2014;23:524-31.

4. Chen W. Cancer statistics: updated cancer burden in China. Chin J Cancer Res 2015;27:1.

5. Figueiredo C, Camargo MC, Leite M, et al. Pathogenesis of Gastric Cancer: Genetics and Molecular Classification. Curr Top Microbiol Immunol 2017;400:277-304.

6. He J, Jin Y, Chen Y, et al. Downregulation of ALDOB is associated with poor prognosis of patients with gastric cancer. Onco Targets Ther 2016;9:6099-109.

7. Oh SC, Sohn BH, Cheong JH, et al. Clinical and genomic landscape of gastric cancer with a mesenchymal phenotype. Nat Commun 2018;9:1777.

8. Hippo Y, Taniguchi H, Tsutsumi S, et al. Global gene expression analysis of gastric cancer by oligonucleotide microarrays. Cancer Res 2002;62:233-40.

9. Zhang X, Ni Z, Duan Z, et al. Overexpression of E2F mRNAs associated with gastric cancer progression identified by the transcription factor and miRNA co-regulatory network analysis. PLoS One 2015;10:e0116979.

10. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin 2019;69:7-34.

11. Yu J, Lin JH, Wu XR, et al. Uroplakins Ia and Ib, two major differentiation products of bladder epithelium, belong to a family of four transmembrane domain (4TM) proteins. J Cell Biol 1994;125:171-82.

12. Webb GC, Finch JL, Cowled PA. Assignment of the Uroplakin 1b (Upk1b) gene to mouse chromosome 16 bands B5-C2 by in situ hybridization. Cytogeten Cell Genet 1999;84:37-8.

13. Finch JL, Webb GC, Evdokiou A, et al. Chromosomal localization of the human urothelial "tetraspan" gene, UPK1B, to 3q13.3-q21 and detection of a TaqI polymorphism. Genomics 1997;40:501-3.

14. Osman I, Kang M, Lee A, et al. Detection of circulating cancer cells expressing uroplakins and epidermal growth factor receptor in bladder cancer patients. Int J Cancer 2004;111:934-9.

15. Olsburgh J, Harnden P, Weeks R, et al. Uroplakin gene expression in normal human tissues and locally advanced bladder cancer. J Pathol 2003;199:41-9.

16. Su J, Zhang Y, Su H, et al. A recurrence model for laryngeal cancer based on SVM and gene function clustering. Acta Otolaryngol 2017;137:557-62.

17. Takeshita S, Fumoto T, Matsuoka K, et al. Osteoclast-secreted CTHRC1 in the coupling of bone resorption to formation. J Clin Invest 2013;123:3914-24.

18. Lai YH, Chen J, Wang XP, et al. Collagen triple helix repeat containing-1 negatively regulated by microRNA-30c promotes cell proliferation and metastasis and indicates poor prognosis in breast cancer. J Exp Clin Cancer Res 2017;36:92.

19. Li N, Chen L, Liu C, et al. Elevated CTHRC1 expression is an indicator for poor prognosis and lymph node metastasis in cervical squamous cell carcinoma. Hum Pathol 2019;85:235-41.

20. Ni S, Ren F, Xu M, et al. CTHRC1 overexpression predicts poor survival and enhances epithelial-mesenchymal transition in colorectal cancer. Cancer Med 2018;7:5643-54.

21. Zhou H, Su L, Liu C, et al. CTHRC1 May Serve As A Prognostic Biomarker For Hepatocellular Carcinoma. Oncotarget 2016;7:57351-66.

22. Ding X, Huang R, Zhong Y, et al. CTHRC1 promotes gastric cancer metastasis via HIF-1alpha/CXCR4 signaling pathway. Biomed Pharmacother 2020;123:109742.

23. Pavon MA, Arroyo-Solera I, Cespedes MV, et al. uPA/uPAR and SERPINE1 in head and neck cancer: role in tumor resistance, metastasis, prognosis and therapy. Oncotarget 2016;7:57351-66.

24. Pavon MA, Arroyo-Solera I, Tellez-Gabriel M, et al. Enhanced cell migration and apoptosis resistance may underlie the association between high SERPINE1 expression and poor outcome in head and neck carcinoma patients. Oncotarget 2015;6:29016-33.

25. Klimczak-Bittner AA, Kordek R, Bitner J, et al. Expression of MMP9, SERPINE1 and miR-134 as prognostic factors in esophageal cancer. Oncol Lett 2016;12:4133-8.

26. Zhang G, Gomes-Giacoia E, Dai Y, et al. Validation and clinicopathologic associations of a urine-based bladder cancer biomarker signature. J Urol 2014;192:200.
poor prognosis of gastric adenocarcinoma revealed by microarray and bioinformatics. Sci Rep 2019;9:7827.

30. Schwartz NB, Pirok EW, 3rd, Mensch JR, Jr., et al. Domain organization, genomic structure, evolution, and regulation of expression of the aggrecan gene family. Prog Nucleic Acid Res Mol Biol 1999;62:177-225.

31. Russell DL, Ochsner SA, Hsieh M, et al. Hormone-regulated expression and localization of versican in the rodent ovary. Endocrinology 2003;144:1020-31.

32. Hakkinen L, Westermanck J, Kahari VM, et al. Human granulation-tissue fibroblasts show enhanced proteoglycan gene expression and altered response to TGF-beta 1. J Dent Res 1996;75:1767-78.

33. Cattaruzza S, Schiappacassi M, Ljungberg-Rose A, et al. Distribution of PG-M/versican variants in human tissues and de novo expression of isoform V3 upon endothelial cell activation, migration, and neoangiogenesis in vitro. J Biol Chem 2002;277:47626-35.

34. Evanko SP, Raines EW, Ross R, et al. Proteoglycan distribution in lesions of atherosclerosis depends on lesion severity, structural characteristics, and the proximity of platelet-derived growth factor and transforming growth factor-beta. Am J Pathol 1998;152:533-46.

35. Crowther M, Brown NJ, Bishop ET, et al. Microenvironmental influence on macrophage regulation of angiogenesis in wounds and malignant tumors. J Leukoc Biol 2001;70:478-90.

36. Calon A, Lonardo E, Berenguer-Llergo A, et al. Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. Nat Genet 2015;47:320-9.

37. Yeung TL, Leung CS, Wong KK, et al. TGF-beta modulates ovarian cancer invasion by upregulating CAF-derived versican in the tumor microenvironment. Cancer Res 2013;73:5016-28.

38. Theocharis AD. Human colon adenocarcinoma is associated with specific post-translational modifications of versican and decorin. Biochim Biophys Acta 2002;1588:165-72.

39. Tsara ME, Theocharis AD, Theocharis DA. Compositional and structural alterations of proteoglycans in human rectum carcinoma with special reference to versican and decorin. Anticancer Res 2002;22:2893-8.

40. Touab M, Villena J, Barranco C, et al. Versican is differentially expressed in human melanoma and may play a role in tumor development. Am J Pathol 2002;160:549-57.

41. Ito Y, Abiko Y, Tanaka Y, et al. Immunohistochemical localization of large chondroitin sulfate proteoglycan in odontogenic tumor. Med Electron Microsc 2002;35:173-7.

42. Voutilainen K, Anttila M, Sillanpaa S, et al. Versican in epithelial ovarian cancer: relation to hyaluronan, clinicopathologic factors and prognosis. Int J Cancer 2003;107:359-64.

43. Ang LC, Zhang Y, Cao L, et al. Versican enhances locomotion of astrocytoma cells and reduces cell adhesion through its G1 domain. J Neuropathol Exp Neurol 1999;58:597-605.

44. Zhang Y, Cao L, Yang BL, et al. The G3 domain of versican enhances cell proliferation via epidermal growth factor-like motifs. J Biol Chem 1998;273:21342-51.

45. Yang BL, Zhang Y, Cao L, et al. Cell adhesion and proliferation mediated through the G1 domain of versican. J Cell Biochem 1999;72:210-20.

46. Shen XH, Lin WR, Xu MD, et al. Prognostic significance of Versican expression in gastric adenocarcinoma. Oncogenesis 2015;4:e178.

Cite this article as: Zhu Z, Xu J, Li L, Ye W, Chen B, Zeng J, Huang Z. Comprehensive analysis reveals CTHRC1, SERPINE1, VCAN and UPK1B as the novel prognostic markers in gastric cancer. Transl Cancer Res 2020;9(7):4093-4110. doi: 10.21037/tcr-20-211
The table lists the 23 survival-related genes identified by gene set enrichment analysis that were significantly associated with survival (P < 0.05). The genes are ranked by their statistical significance, with the top 10 genes listed first. The table includes the gene name, description, and P-value. The Kaplan-Meier analysis was used to perform the survival analysis.