

RESEARCH ARTICLE

Editorial Process: Submission:08/28/2021 Acceptance:09/12/2021

# ACAN, MDFI, and CHST1 as Candidate Genes in Gastric Cancer: A Comprehensive Insilco Analysis

Farzane Vafaeie<sup>1</sup>, Samira Nomiri<sup>2</sup>, Javad Ranjbaran<sup>2</sup>, Hossein Safarpour<sup>3\*</sup>

## Abstract

**Background:** Gastric cancer (GC) is a complex disorder with an inadequate response to treatment. Although many efforts have been made to clarify the development of GC, the exact etiology and molecular mechanisms of this malignancy remain unclear. This study was designed to identify and characterize essential associated genes with GC to construct a prognostic model. **Methods:** In this Insilco study, the gene expression microarray dataset GSE122401 was downloaded from the Gene Expression Omnibus (GEO). The raw data were processed and quantile-normalized with the edgeR package of R ver.3.5.3. The module-trait relationship and hub-genes associated with GC were analyzed with Weighted Gene Co-expression Network Analysis (WGCNA). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed by Cluepedia and Enrichr Database. Finally, hub-genes were screened and validated by GEPIA online database. **Results:** According to the WGCNA results, the blue module was found to be strongly correlated with the GC ( $r=0.91$ ,  $p\text{-value}=1e-57$ ). DEGs analysis was performed by edgeR package of R and indicated a total of 47 genes as hub-genes. Verifying the hub-genes expression using GEPIA online database showed a significantly increased level of ACAN gene expression in primary cancer cell line compared to metastatic cell line. On the other hand, the expression of MDFI and CHST1 genes in primary cell lines were lower compared to metastatic cancer cell lines. **Conclusions:** This study provides a framework of the co-expression gene modules ACAN, MDFI, and CHST1 as hub-genes. These hub-genes might offer candidate biomarkers to targeted therapy against GC. Further experiment validation and animal models are needed to reveal the exact mechanism of the above-mentioned genes in the pathogenesis and prognoses of GC.

**Keywords:** WGCNA- Hub-genes- biomarkers- systems biology

*Asian Pac J Cancer Prev*, **23** (2), 683-694

## Introduction

Gastric cancer (GC), also known as stomach adenocarcinoma, is the third leading cause of cancer-related death worldwide. However, Geographical diversity, socio-cultural and economic entities play a role in its prevalence and incidence (Olnes and Martinson, 2021). GC is a heterogeneous, multifactorial, and complex disorder involving various genetic and epigenetic alterations. Environmental factors, genetic background, infectious agents, and dietary habits affect the development of GC. High GC mortality is mainly due to late detection and poor response to treatment. The heterogeneous nature of this disease is one of the reasons that may explain the poor clinical results. A comprehensive understanding of GC molecular classification and its biology may have implications for patient diagnosis and treatment (Tokumaru et al., 2021). Prevention and personal treatment are considered to be the best selection to reduce mortality. Recognizing predictive biomarkers and therapeutic

targets of GC in early-stage and developing appropriate therapies is a prospective method for diagnosing GC subtypes and improving the prognosis of patients with advanced gastric cancer (Zhou et al., 2020). Current approaches in transcriptomic studies, such as microarray-based monitoring, provide novel insights into identifying new markers as their diagnostic or therapeutic target. A microarray technique may provide information about new candidate genes, offers a comprehensive molecular landscape of GC and novel molecular classification that describe molecular tumor heterogeneity, and allows new targeted therapies (Businello et al., 2021). Due to the great variety of data generated by microarray, new approaches are necessary to obtain significant correlations from highly multivariate datasets (Wang et al., 2021). Weighted Gene Co-expression Network Analysis (WGCNA) is an effective statistical method for exploring the co-expression and connections between modules, summarize clusters utilizing the module eigengene or the intramodular hub-genes, and detecting candidate biomarkers or therapeutic

<sup>1</sup>Department of Biology, Faculty of science, Ferdowsi University of Mashhad, Mashhad, Iran. <sup>2</sup>Department of Biochemistry, Faculty of Medicine, Birjand University of medical sciences, Birjand, Iran. <sup>3</sup>Cellular & Molecular Research Center, Birjand University of Medical Sciences, Birjand, Iran. \*For Correspondence: safarpour701@yahoo.com

goals by using correlation networks (Derakhshani et al., 2020; Qin et al., 2021). Moreover, WGCNA is an evolutionary, systematic, biology-based method used to distinguish intrinsic transcriptome organization and calculation of topological properties (Langfelder and Horvath, 2008). This method has been commonly used in cancer biology to identify candidate biomarkers or therapeutic targets successfully. In GC, Qi Zhou et al. have identified the prognostic and predictive model of 10 genes for diffuse-type GC (Zhou et al., 2020). Recently Zhen-Hu Ren et al. identified a total of 275 lncRNAs that were found to be dysregulated during the tumor progression of GC by using WGCNA (Ren et al., 2020). In this investigation, we analyzed the dataset of the GC samples obtained from online databases to identify potential biomarkers or therapeutic targets of GC at the transcriptomic level.

## Materials and Methods

The workflow for the overall procedure is shown in Figure 1.

### *Microarray data and identification of DEGs*

The gene expression dataset GSE122401 was downloaded from the Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) database, which is based on Illumina HiSeq 2500 GPL16791 platform. Notably, this dataset provides a raw read count of 160 human GC samples and normal adjacent specimens (Mun et al., 2019). Next, to investigate the differentially expressed genes (DEGs) between the GC and the adjacent normal tissues, the edgeR package of R ver.3.5.3 was used. Of note, the studied genes considered to be DEGs, if the following conditions were met:  $FDR \leq 1.00e-10$  and  $|\text{Log}_2\text{FC}| \geq 1.5$ . In the next step, functional enrichment analysis of DEGs was performed using Cluepedia tools of Cytoscape ver. 3.6.1. The enriched ontological terms and pathways with a threshold of Benjamin-adjusted p-value  $< 0.05$  have been selected.

### *Constructing Gene Co-Expression Network and Identification of Significant Modules*

We used WGCNA for reconstructing the co-expression network of GC genes. In summary, using the Pearson Correlation Coefficient test, the gene expression matrixes were transformed into matrixes consisting of the similarities of the pairwise mRNAs, followed by a conversion to adjacency matrixes. Then, a minimum possible  $\beta$ -value was considered based on the represented scale-free gene co-expression topological algorithm. The adjacency matrix met the requirements of scale-free topology (Langfelder and Horvath, 2008). Subsequently, Topological Overlap Matrix (TOM) and dissimilarity TOM (dissTOM) have been established using TOM similarity and dissimilarity modules for the next step. Lastly, the module identification was performed by dynamic tree cut, while the minimum module size was set to 30. Modules with high similarity scores have been combined with a threshold value for each dataset.

### *Construction of Module-Trait Relationships and Functional Enrichment Analysis*

In order to identify significantly related modules to the evaluated clinical features, the expression profiles of each module were summarized as the eigenvector correlated to the first principal component of the expression matrix using the Module Eigengene (ME). The module eigengene was measured as the first central component of a given module, which may be considered as a representative of gene expression levels among these samples in a module. Moreover, the values of gene significance (GS) were used to calculate the association of individual genes with the GC. In addition, the Module Membership (MM) was defined as the ME correlation as well as the gene expression profile for each module. If the GS and MM are strongly associated, it can be stated that the most important (central) elements in the modules are also closely related to the trait. Thereafter, they can also be used to create a network and identify the hub-genes. Based on that, Genes with MM and GS values over 0.75 were considered the module's signature genes for GC. Next, module genes were analyzed by the GO and KEGG pathway using Cluepedia tools of Cytoscape ver. 3.6.1. The enriched ontological terms and pathways with a threshold of Benjamin-adjusted p-value  $< 0.05$  have been selected.

### *Validation of Selected hub-genes on External Databases*

In order to analyze the effect of the selected hub-genes on overall survival (OS) in the patients with GC, the GEPIA online database (<http://gepia.cancer-pku.cn>) was used. Additionally, the expression level of the filtered genes in the control and GC groups were compared using the GEPIA database (Tang et al.). For initial validation of selected hub-genes expression in primary tumors, normal solid tissue, recurrent tumor, and metastatic tumors, data were extracted from the TCGA-STAD database using the UCSC Cancer Browser (<https://xenabrowser.net/>) (accessed on 18 May 2021)) (Goldman et al., 2020). Then, the expression of selected hub-genes in primary and metastatic cells was obtained from the Cancer Cell Line Encyclopedia (CCLE) datasets using the Xena browser (Garraway, 2005).

## Results

### *DEG Identification and Pathway Enrichment Analyses*

Quantile normalization was conducted to reduce the effect of technical noise of the gene expression. A total of 15 outliers was found by clustering the samples in the dataset, and then 145 samples were included in this study (Figure 2). A total of 880 genes, including 466 up-regulated and 414 down-regulated genes, were selected as DEGs based on  $FDR \leq 1.00e-10$  and  $|\text{Log}_2\text{FC}| \geq 1.5$ . Correspondingly, anatomical structure morphogenesis, serine-type endopeptidase activity, collagen extracellular matrix organization are known as the most important biological functions and pathways for DEGs (Figure 3).

### *Identification of clinically significant modules*

A total of 4,000 genes were included in WGCNA based on their differences in expression values. At this

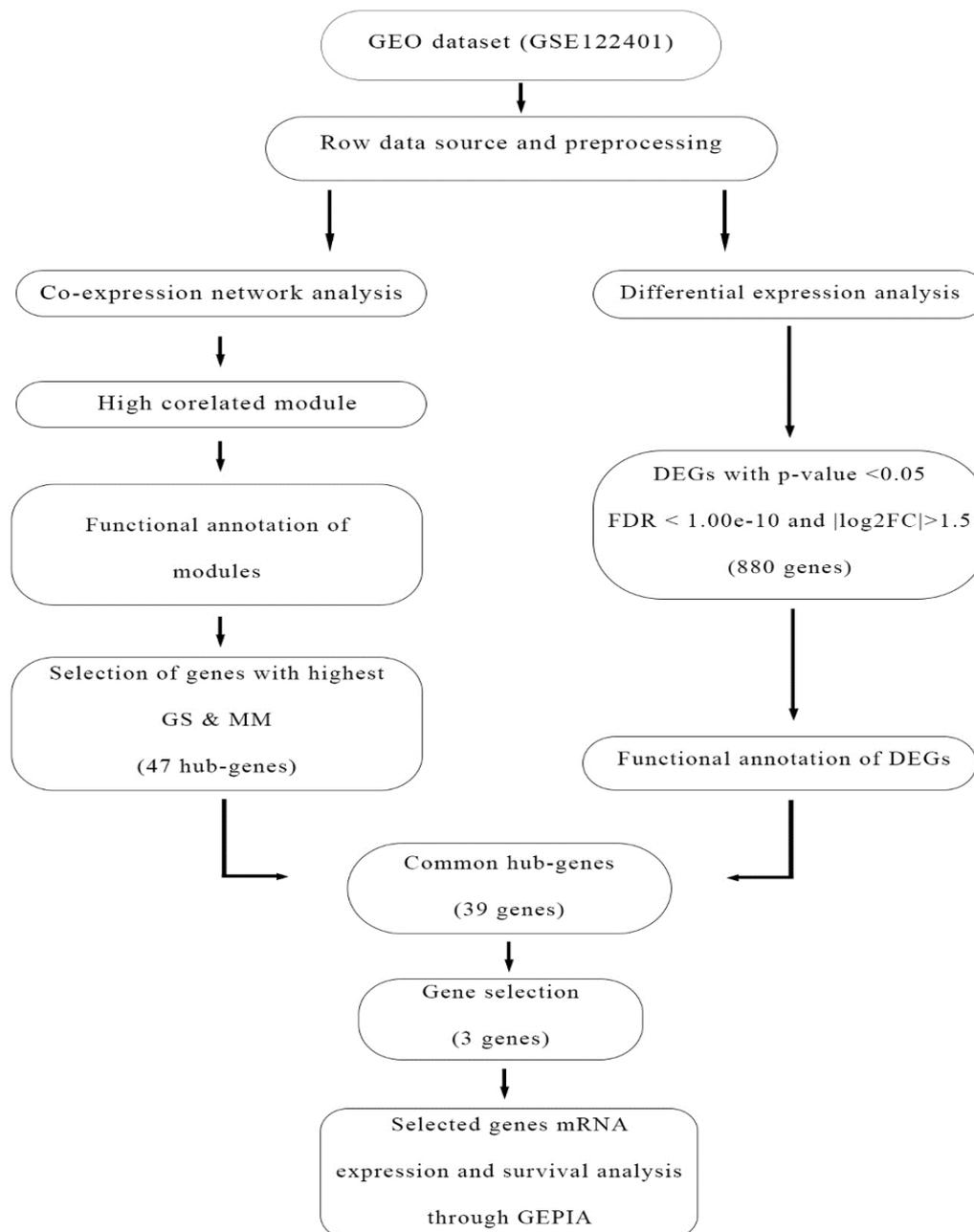


Figure 1. Workflow of Data Preparation, Processing, and Analysis in This Study. The analyses were performed in two separate manners. In the first one, in order to find the most differentiated expressed genes, DEG analysis was performed. In parallel, WGCNA was done to find the genes with the highest values of ‘gene significance’ and ‘module membership’, which present the weight of genes in the network. Finally, the similar genes between these two lists were selected as hub-genes. This procedure guarantees that each gene has the highest differential expression as well as the highest communication with other genes and also a correlation with disease state.

stage, we computed the strength of the Pearson correlation between each gene pair, in order to obtain an adjacency matrix by raising the matrix to a soft threshold using the WGCNA pick soft-threshold function. Correspondingly, this function offers a scale-free topology fit index that reaches values above 0.6 for low power ( $< 30$ ), implying that the network topology is scale-free. In this regard,  $\beta=5$  was chosen as a soft-threshold power. Subsequently, a weighted co-expression network of the GC patients and adjacent normal samples were then reconstructed (Figure 4a). As a result, hierarchical clustering identified 16 modules observed in dendrogram branches of different

colors (Figure 4b). The number of genes in each module ranged from 9 (grey) to 675 (red) (Table 1).

#### Enrichment analysis and hub-genes detection

In order to represent the relationship between modules and the occurrence of GC and the module-module correlation accurately, the Eigengene was determined for each module. Accordingly, the blue module was found to be strongly correlated with the GC ( $r=0.91$ ,  $p\text{-value}=1e-57$ ) (Table 1 and Figure 5a). The important biological process relevant to the selected modules was visualized using Cluepedia, as shown in Figure 5b.

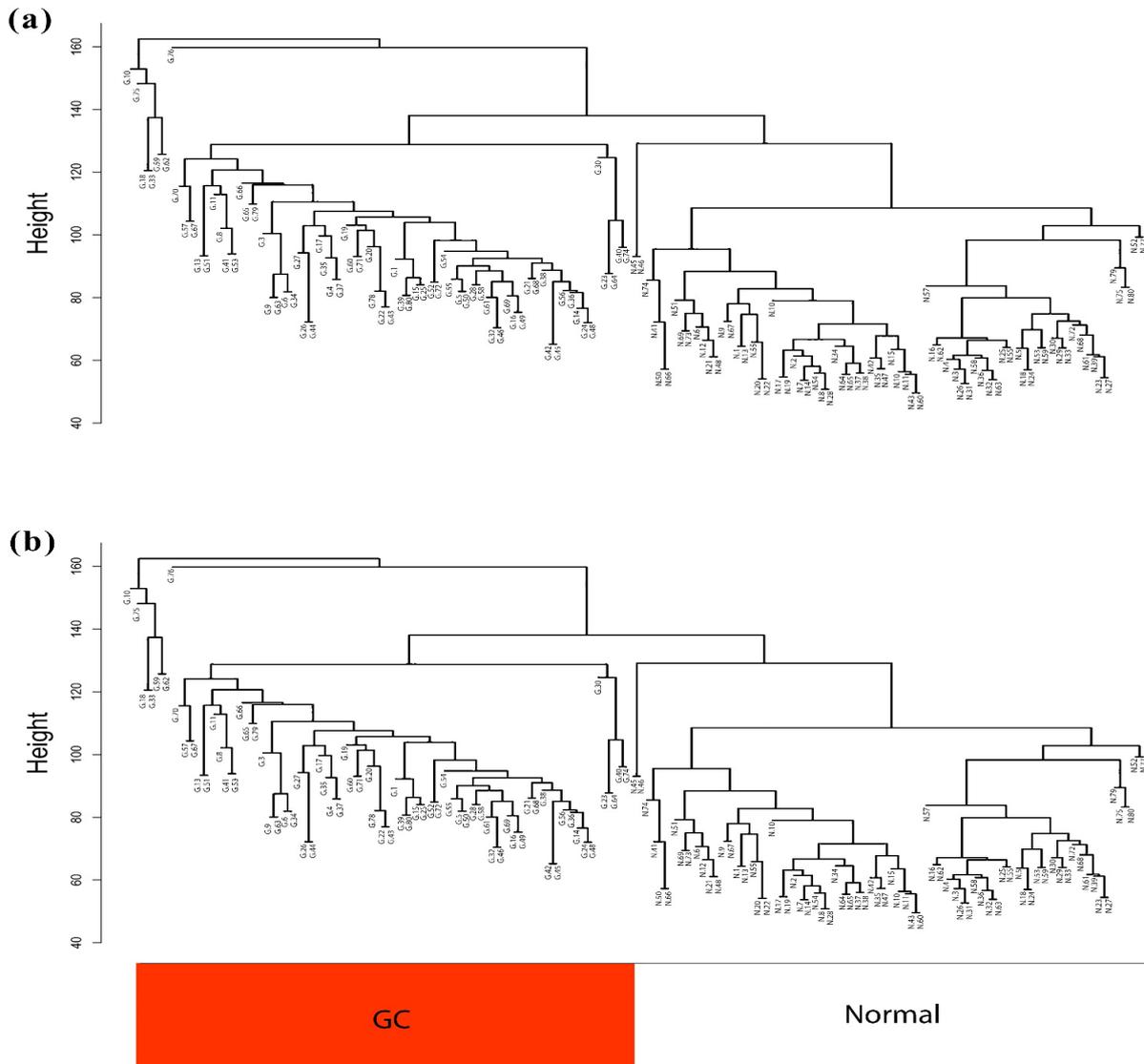


Figure 2. Sample Clustering to Detect Outliers. (a), Before removing outliers; (b), After removing one outlier

Regulation of hormone level and extracellular matrix organization is the most critical pathways associated with the blue module. The association between the characteristics (MM and GS) of the selected modules

(Figure 6a) has consequently led to the discovery of interest-bearing hub-genes that were strongly correlated with GC pathogenesis. The genes with the maximum MM and GS scores in the blue module were compared with

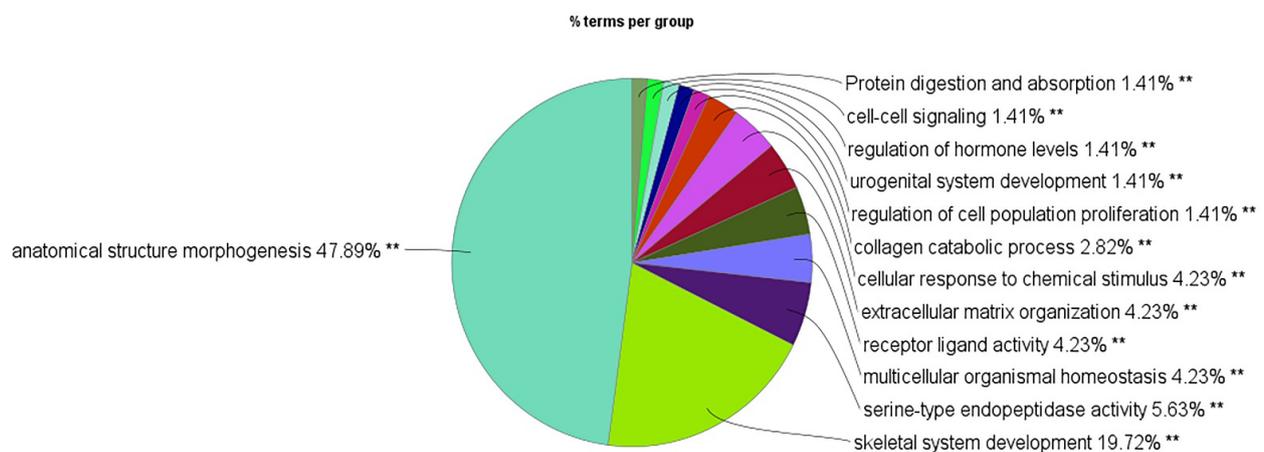


Figure 3. DEGs Enrichment Analysis

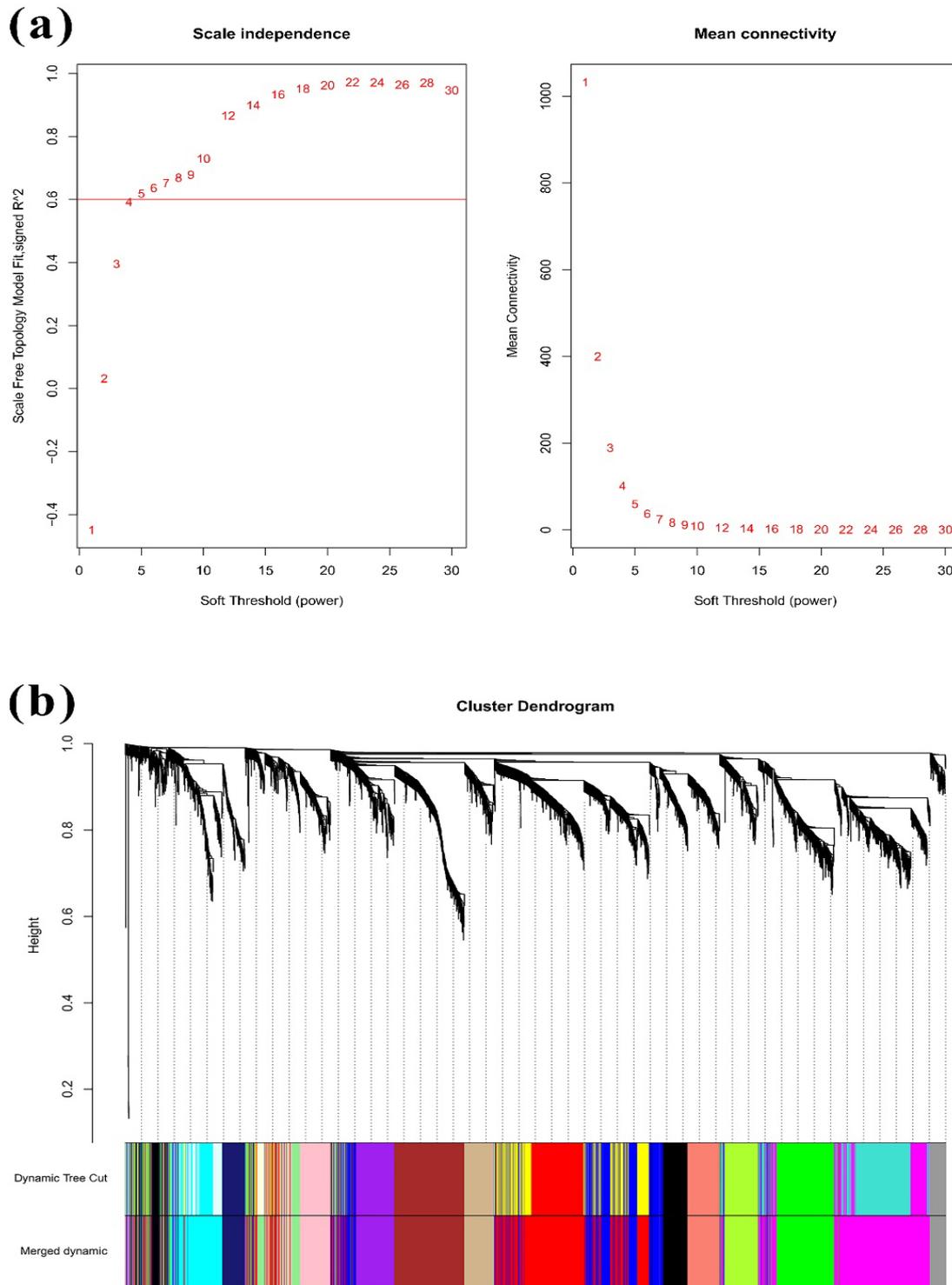


Figure 4. Identification of WGCNA Modules. (a), Selection of the soft-thresholding powers; (b), Cluster dendrogram and module assignment from WGCNA. The branches correspond to highly interconnected groups of genes. Colors in the horizontal bar represent the modules.

the DEG list, and similar genes were considered as the proposed hub-genes (Figure 6b). Finally, the co-expression network of each module has been reconstructed using GeneMANIA and Cytoscape software (Figure 6c).

#### Characterization of hub-genes Expression by External Databases

In order to verify the hub-genes expression, clinicopathological data of patients with primary GC

were obtained from the UCSC Xena browser (<https://xenabrowser.net/> (accessed on 18 May 2021)) (Garraway, 2005). As shown in Figure 7, ACAN, MDFI, and CHST1 expression were significantly different in solid tissue normal (N=35) and primary (N=415). It was also revealed that up-regulation of CHST1 gene has a worse overall survival in GC patients (Figure 8) (Tang et al.). Investigation of GC tissue from human protein atlas demonstrated that these genes expression was up-regulated

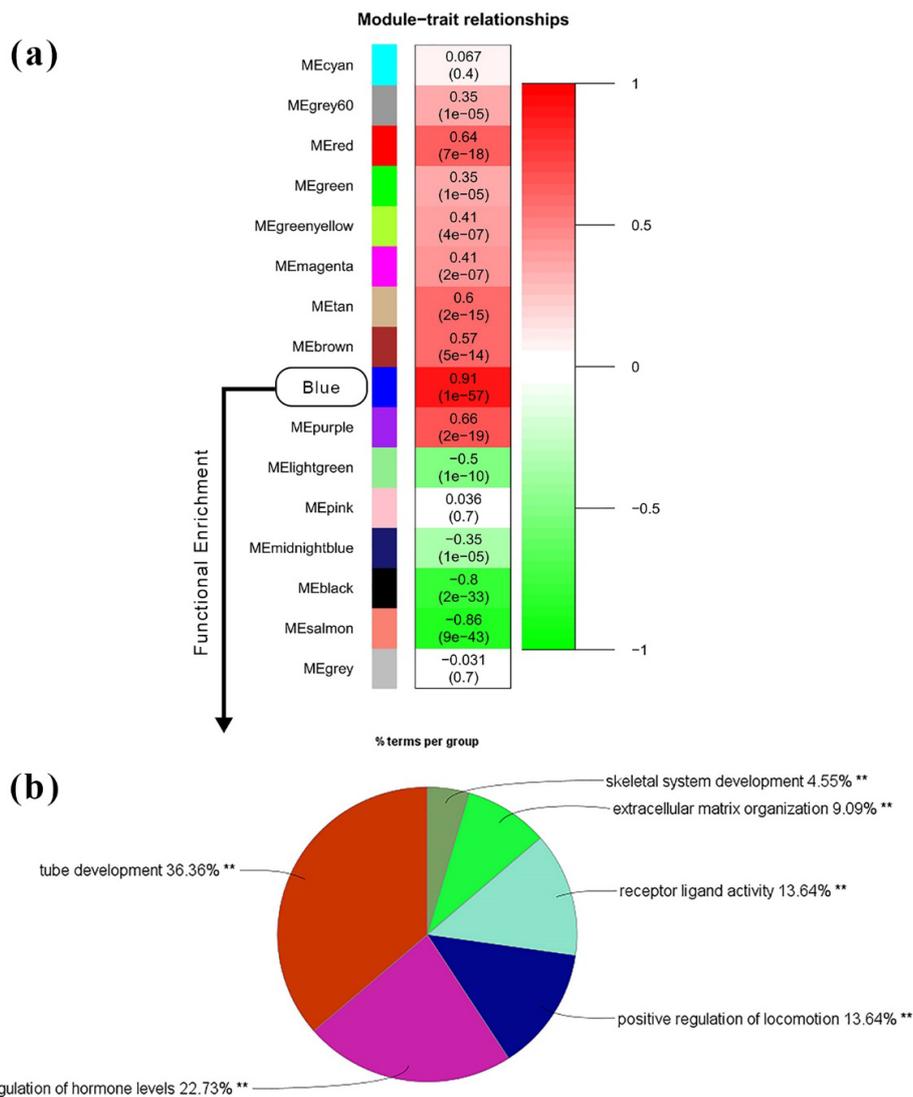


Figure 5. Module-Trait Association Analysis. (a), Module-trait relationship. Each row corresponds to a module Eigengene, and the column corresponds to GC status. The numbers in each cell represent the corresponding correlation and p-value; (b), Functional enrichment analysis of the blue module.

in the primary tumor (Figure 9) (Berglund et al., 2008).

Inconsistent with the results mentioned above, based on the CCLE database (<https://portals.broadinstitute.org/ccle>), we showed that level of ACAN expression in primary cell lines (FU97, SH10TC, and 2313287) was higher compared to metastatic (RERFGC1B, GSS, and SNU620) cell lines (Barretina et al., 2012). However, the expression of MDFI and CHST1 genes in primary cell lines (FU97, SKGT2, and SH10TC) and (SH10TC, FU97, and SNU520) was lower compared to metastatic (NUGC2, SNU216, and SNU1) and (LMSU, HUTU80, and SNU216) cell lines respectively (Figure 10).

Finally using of Enrichr database we analyzed the biological pathways and processes of selected hub-genes. As shown in Figure 11 and Table 2, the keratan sulfate biosynthetic process (GO:0018146) and cytoplasmic sequestering of transcription factors (GO:0042994) were most significantly associated with these hub-genes (Xie et al., 2021).

## Discussion

GC is one of the most common cancers and a significant health problem worldwide due to its frequency and substantial heterogeneity. In this study, we aimed to consider the expression profile of the key genes involved in GC. In this regard, the gene expression dataset GSE122401 was obtained from the GEO database, and reconstructing the co-expression network of GC genes was carried out by WGCNA. Our data indicated that the blue module could be strongly correlated with the GC ( $r=0.84$ ,  $p\text{-value}=4e-40$ ). Functional enrichment analysis showed that serine-type endopeptidase activity, collagen catabolic process, receptor-ligand activity, and extracellular matrix organization are known as the most important biological functions and pathways for DEGs and blue module. The result showed a significantly increased level of ACAN expression in primary compared to metastatic cell lines. It was also revealed that the expression of MDFI and CHST1 genes in the primary was lower than in metastatic samples.

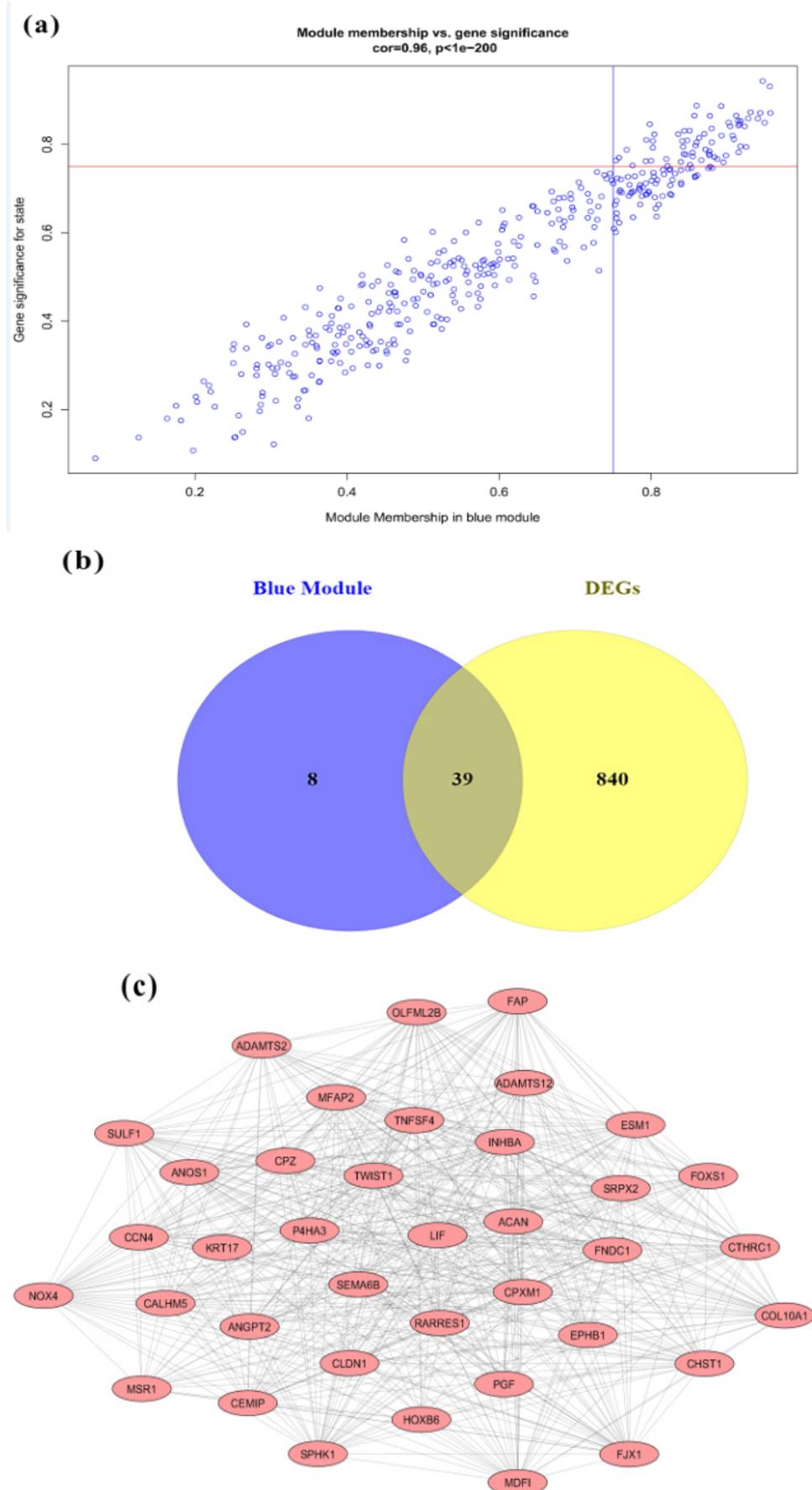


Figure 6. Hub-Genes Detection. (a), Blue module features of GS and MM, which were significantly correlated with GC status (adjacent normal vs. GC). Each point represents an individual gene within each module, which was plotted by GS on the y-axis and MM on the x-axis; (b), Evaluation of similarity between DEGs and hub-genes lists using a Venn diagram; (c), 39 genes that were similar in both lists, were imported to Gene MANIA to construct a co-expression network.

**Cohort: TCGA stomach cancer (STAD) (n=450)**

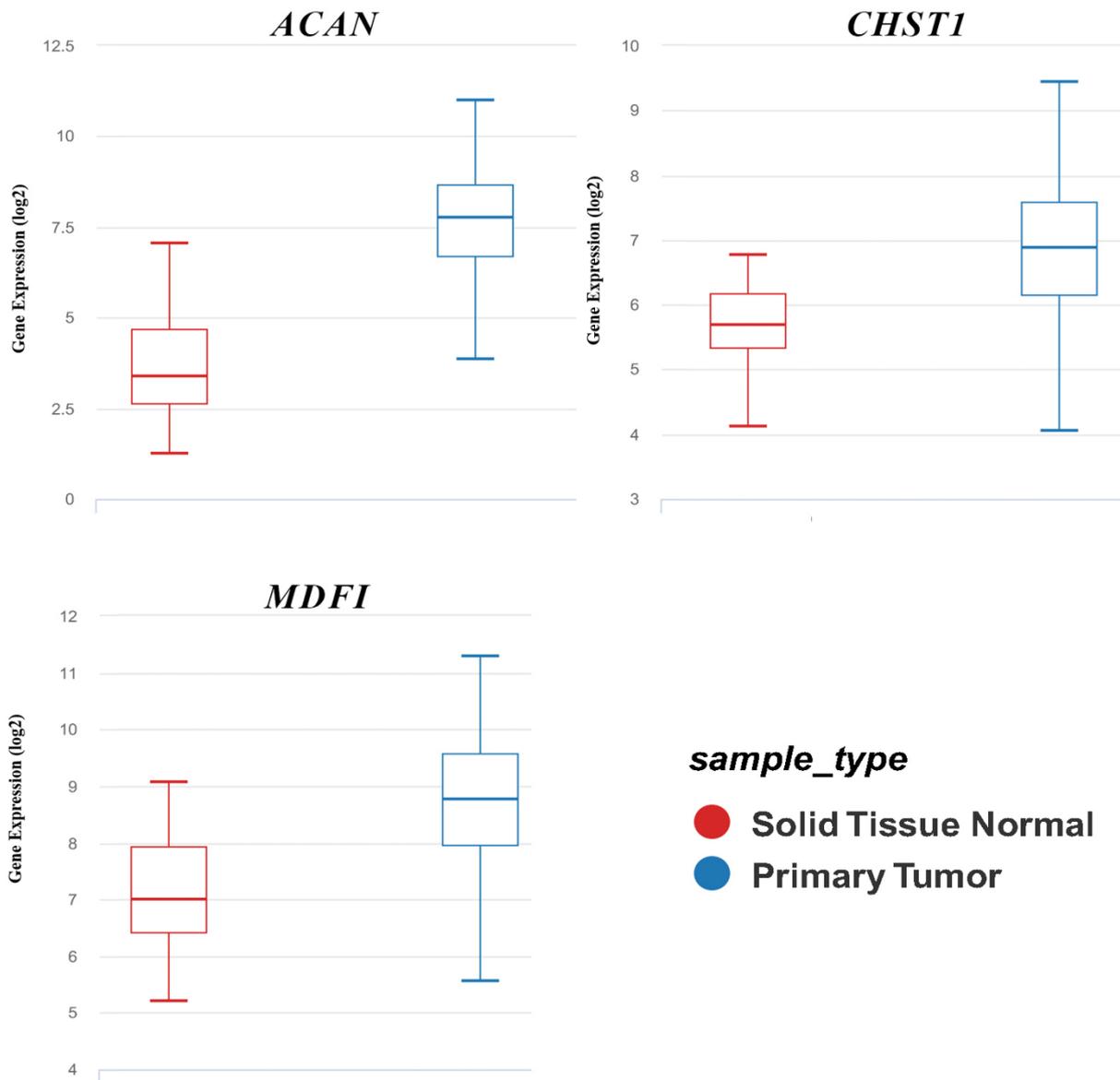


Figure 7. Characterization of *ACAN*, *MDFI*, and *CHST1* Expression Using External Databases. Differential expression in Gastric (n=415) and normal (n=35) samples. <https://xenabrowser.net/> (accessed on 18 May 2021)

*ACAN* is a protein-coding gene and a member of the aggrecan/versican proteoglycan family that exists in the extracellular matrix of various tissues. Different protein isoforms have been observed in *ACAN* gene due to multiple alternatively spliced transcript variants. Many studies indicated that mutations in *ACAN* gene might be involved in skeletal dysplasia, including short stature and multiple intervertebral disc herniation, however, investigating its role in cancer requires further research (Cikach et al., 2018). Vizeacoumar et al. researched to identify the possible novel biomarkers for the prevention and management of Gastro-esophageal cancers. Comparison between normal stomach tissue and stomach tumors showed the leading five genes that were overexpressed in stomach cancer. According to their

analysis, *ACAN* gene showed significant upregulation in all stages of cancer (Vizeacoumar et al., 2021). Similar research was performed to investigate the interaction of miRNAs and target genes in Stomach adenocarcinoma (STAD). Based on their result, differential expression of the mature form of miRNAs and target genes, including *ACAN* may play an important role in STAD development (Liu et al., 2018). Further studies have been done on the effect of the *ACAN* gene on other cancers. CpG island methylator phenotype (CIMP) study in lung adenocarcinomas indicated that of the six CIMP markers, *ACAN* gene was significantly associated with patient survival (Koh et al., 2016).

According to DEGs and WGCNA analyses, *MDFI* was selected as the hub-gene in GC. MyoD family inhibitor 1

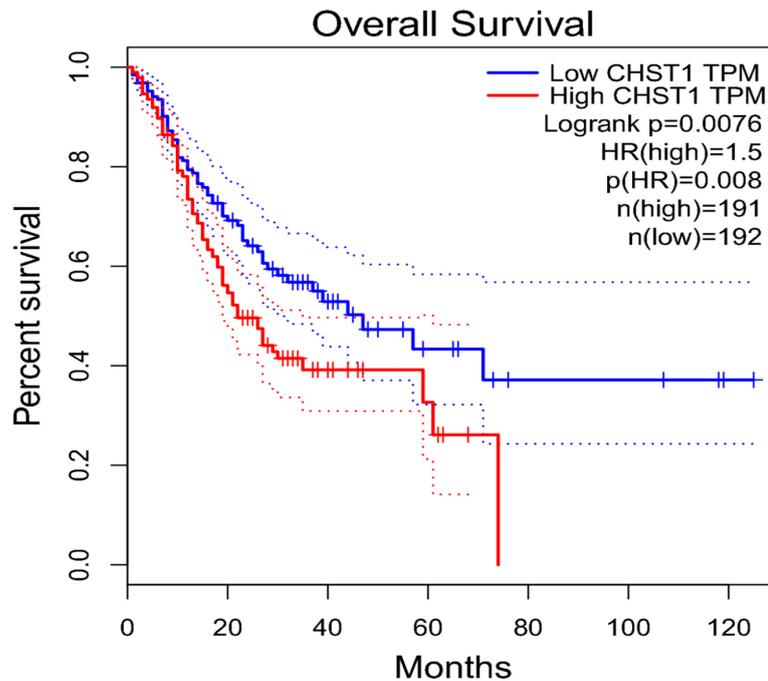


Figure 8. The Overall Survival Rate of High and Low-Risk Patients based on the Expression Level of *CHST1* Gene in the GEPIA Database.

(MDFI) inhibits Wnt/beta-catenin signaling by interacting with the axin complex and lymphocyte enhancer factor (LEF) (Fodde and Brabletz, 2007). Abnormal Wnt/ $\beta$ -catenin signaling has been reported in various cancers, including gastric and colorectal cancer (Cheng et al., 2019). MDFI is a tumor suppressor gene involved in tumorigenesis and a negative regulator of myogenic regulatory factors (Snider et al., 2001). Loss of MDFI, a key negative regulator of the Wnt pathway, has been linked to human breast cancer and myeloid neoplasm (Cigognini et al., 2007). A study of MDFI and the MyoD family inhibitor domain-containing protein (MDFIC) expression in different cancers discovered that MDFI gene was up- and MDFIC downregulated in colorectal, ovarian, and prostate cancer. While overexpression of MDFI and MDFIC were observed in the brain, gastric

and pancreatic tumors. This implicated that MDFI might perform tumor-promoting activities (Sui et al., 2020). Further investigation in colorectal cancer also showed that DNA methylation of MDFI gene could be considered as an epigenetic biomarker in colorectal cancer (Li et al., 2017).

Additionally, the dry-lab results suggested a significant association between the Carbohydrate Sulfotransferase 1 (CHST1) and GC. CHST genes encode Golgi-localized enzymes (sulfotransferase and keratan sulfotransferase) that install sulfate at defined locations in glycans (Jung et al., 2021). There are few pieces of literature and imperial evidence on the role of CHST1 in cancer. A similar analysis by Qian Chen et al. was performed on FPKM RNA-seq data from the TCGA database and clinical data for patients with stomach adenocarcinoma. According to

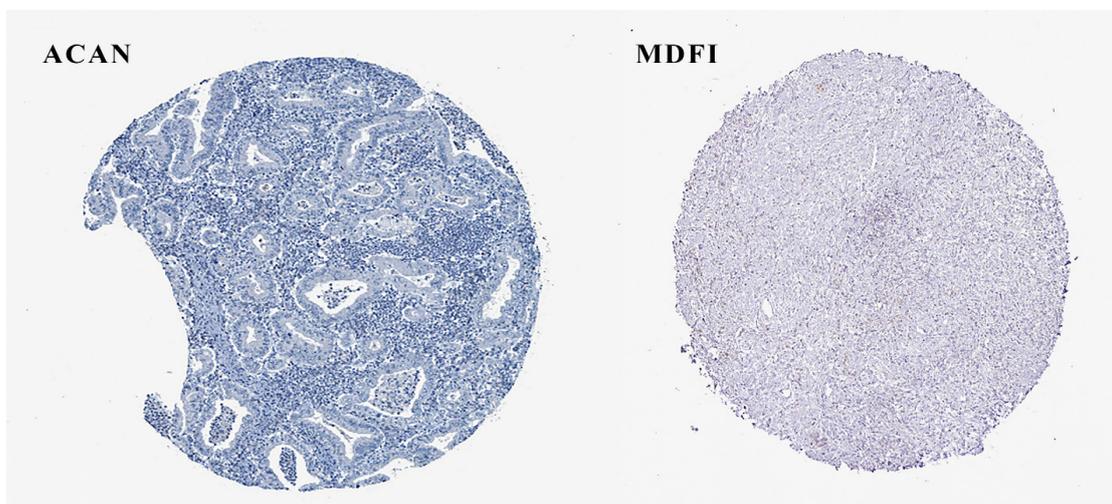


Figure 9. Immunohistochemical Staining of ACAN and MDFI Proteins in GC Patients. <https://www.proteinatlas.org>.

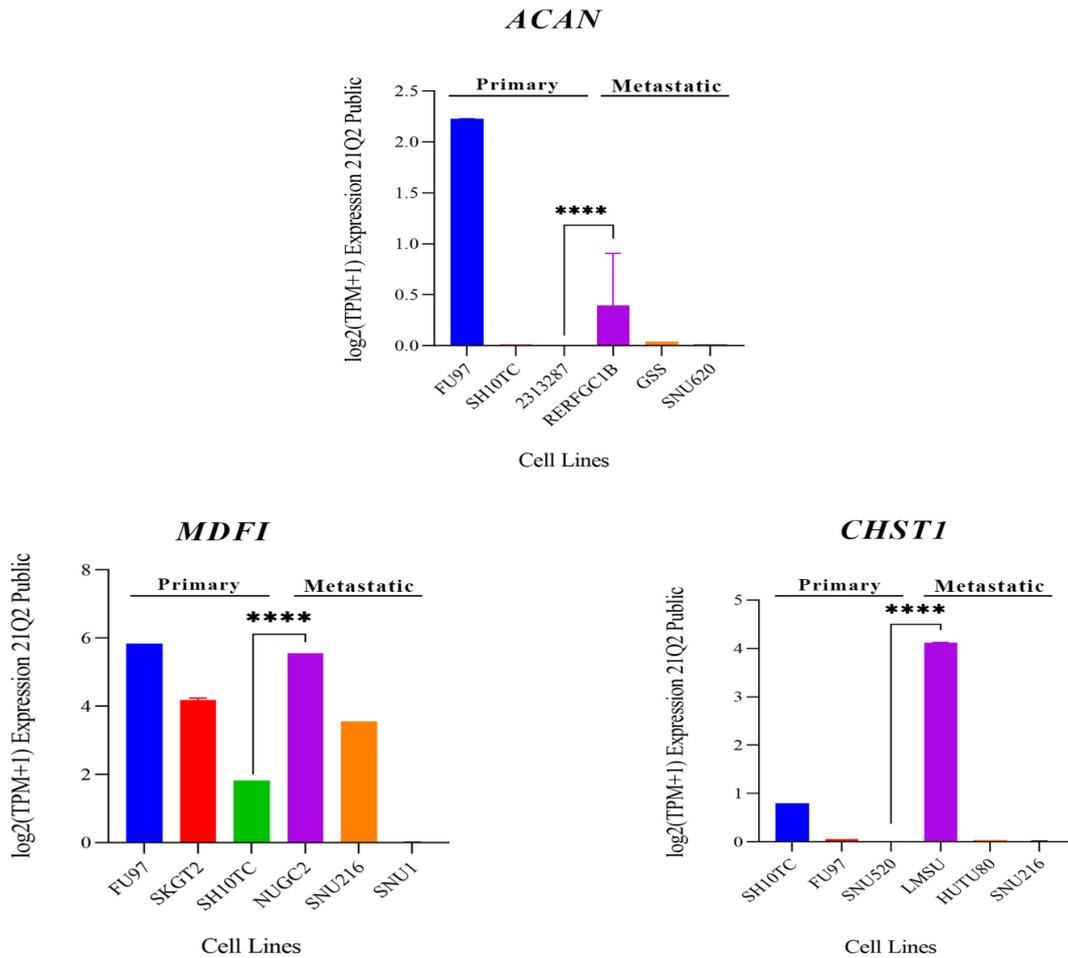


Figure 10. The Expression Levels of *ACAN*, *MDFI*, and *CHST1* in Primary and Metastatic Cell Lines. <https://portals.broadinstitute.org/ccle>.

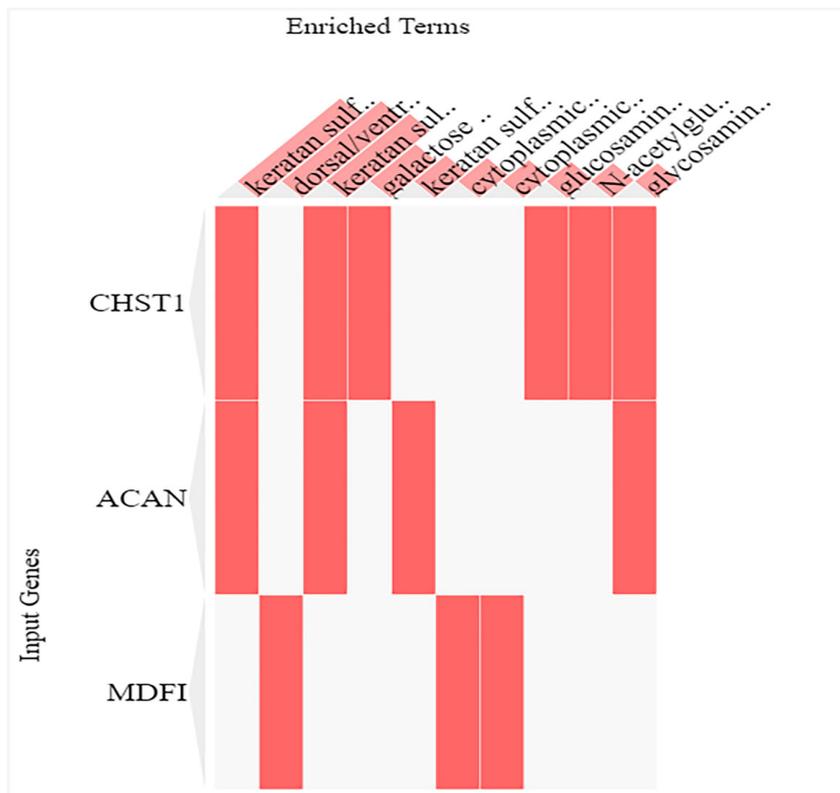


Figure 11. The Clustergram Plot Depicted the Biological Processes in which *ACAN*, *MDFI*, and *CHST1* were Involved.

Table 1. Supplement and Supporting Data (SSD)

Abbreviation list	
Abbreviation	Definition
GC	Gastric Cancer
GEO	Gene Expression Omnibus
WGCNA	Weighted Gene Co-expression Network Analysis
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Gene and Genome
GEPIA	Gene Expression Profiling Interaction Analysis
DEG	Differentially Expressed Gene
ACAN	Aggrecan
MDFI	MyoD family inhibitor
CHST1	Carbohydrate Sulfotransferase 1
TOM	Topological Overlap Matrix
diss TOM	dissimilarity Topological Overlap Matrix
ME	Module Eigengene
MM	Module Membership
OS	Overall Survival
CCLL	Cancer Cell Line Encyclopedia
TCGA-STAD	The Cancer Genome Atlas Stomach Adenocarcinoma
UCSC	University of California, Santa Cruz

their results, the expression of CHST1 in gastric cancer tissues was higher than that in normal tissues (Chen et al., 2020). In addition, the expression profiles of lncRNA, miRNA, and mRNA in gastric cancer and para cancer tissues indicated that CHST1 gene and miR-183-5p could be used as prognostic factors of gastric cancer. Their result showed CHST1 had a negative correlation while miR-183-5p had a positive correlation with the survival of patients (Li et al., 2020). Studies on other members of carbohydrate sulfotransferase (CS) demonstrated that CS members could have oncogenic or tumor suppressor activity. Overexpression of CHST3 and CHST11 have been correlated to BC and the development of metastasis. On the contrary, downregulation of CHST10 and CHST14 has been related to melanoma and late colon cancer stages, respectively (Terkelsen et al., 2020).

Our work presented a significantly increased ACAN, MDFI, and CHST1 expression in gastric cancer samples compared to controls. Notably, The Cancer Cell Line Encyclopedia (CCLE) model showed a high level of ACAN expression in primary (FU97, SH10TC, and 2313287) compared to metastatic (RERFGC1B, GSS, and SNU620) cell lines. However, the expression of MDFI and CHST1 genes in a primary cancer cell lines (FU97, SKGT2, and SH10TC) and (SH10TC, FU97, and SNU520) was lower compared to metastatic (NUGC2, SNU216, and SNU1) and (LMSU, HUTU80, and SNU216) cell lines.

However, these results still have some limitations including small sample size and lack of verification test; thus, a large clinical sample size will be needed in future studies. Moreover, further experimental studies require to investigate the functional role of genes identified in dry-lab and confirm their expression at the protein level.

In the present study, we applied bioinformatics methods to screen out the genes involved in the progression of GC.

In summary, WGCNA and co-expression analysis were performed to identify highly correlated genes with GC. Three hub-genes were identified, and survival analysis was further performed. Subsequently, we applied the TCGA and GTEx GC samples to verify the hub-genes expression. Therefore, our findings may enhance the fundamental knowledge of the molecular mechanisms underlying this disease. We revealed that the diagnosed GC-associated genes would be a new reference to construct a predictive model based on hub-genes for gastric cancer and offer new ideas for individual treatment.

## Author Contribution Statement

Not applicable.

## Acknowledgements

Not applicable.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- Barretina J, Caponigro G, Stransky N, et al (2012). 22 The Cancer Cell Line Encyclopedia-Using Preclinical Models to Predict Anticancer Drug Sensitivity. *Eur J Cancer*, S5-S6.
- Berglund L, Björling E, Oksvold P, et al (2008). A genecentric Human Protein Atlas for expression profiles based on antibodies. *Mol Cell Proteom*, 7, 2019-27.
- Businello G, Galuppini F, Fassan M (2021). The impact of recent next generation sequencing and the need for a new classification in gastric cancer. *Best Pract Res Clin Gastroenterol*, 101730.
- Chen Q, Hu L, Chen K (2020). Construction of a nomogram based on a hypoxia-related lncRNA signature to improve the prediction of gastric cancer prognosis. *Front genet*, 11, 1293.
- Cheng X, Xu X, Chen D, et al (2019). Therapeutic potential of targeting the Wnt/ $\beta$ -catenin signaling pathway in colorectal cancer. *Biomed Pharmacother*, 110, 473-81.
- Cigognini D, Corneo G, Fermo E, et al (2007). HIC gene, a candidate suppressor gene within a minimal region of loss at 7q31.1 in myeloid neoplasms. *Leuk Res*, 31, 477-82.
- Cikach FS, Koch CD, Mead TJ, et al (2018). Massive aggrecan and versican accumulation in thoracic aortic aneurysm and dissection. *JCI Insight*, 3.
- Derakhshani A, Mollaei H, Parsamanesh N, et al (2020). Gene co-expression network analysis for identifying modules and functionally enriched pathways in vitiligo disease: A systems biology study. *Iran J Allergy Asthma Immunol*, 1-12.
- Fodde R, Brabletz T (2007). Wnt/ $\beta$ -catenin signaling in cancer stemness and malignant behavior. *Curr Opin Cell Biol*, 19, 150-8.
- Goldman MJ, Craft B, Hastie M, et al (2020). Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol*, 38, 675-8.
- Jung J, Enterina JR, Bui DT, et al (2021). Carbohydrate sulfation as a mechanism for fine-tuning Siglec ligands. bioRxiv.
- Koh YW, Chun S-M, Park Y-S, et al (2016). Association between the CpG island methylator phenotype and its prognostic significance in primary pulmonary adenocarcinoma. *Tumor*

- Biol*, **37**, 10675-84.
- Langfelder P, Horvath S (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform*, **9**, 1-13.
- Li J, Chen C, Bi X, et al (2017). DNA methylation of CMTM3, SSTR2, and MDFI genes in colorectal cancer. *Gene*, **630**, 1-7.
- Li R, Jiang W, Jin S, et al (2020). Construction and analysis of competitive endogenous RNA regulatory network related to gastric cancer. *Zhonghua zhong liu za zhi* [Chinese journal of oncology], **42**, 115-21.
- Liu J, Liu F, Shi Y, et al (2018). Identification of key mi RNAs and genes associated with stomach adenocarcinoma from The Cancer Genome Atlas database. *FEBS Open Bio*, **8**, 279-94.
- Mun D-G, Bhin J, Kim S, et al (2019). Proteogenomic characterization of human early-onset gastric cancer. *Cancer Cell*, **35**, 111-24. e10.
- Olnes MJ, Martinson HA (2021). Recent advances in immune therapies for gastric cancer. *Cancer Gene Ther*, **2021**, 1-11.
- Qin M, Liang Z, Qin H, et al (2021). Novel Prognostic Biomarkers in Gastric Cancer: CGB5, MKNK2, and PAPP2. *Front Oncol*, **11**, 2236.
- Ren Z-H, Shang G-P, Wu K, et al (2020). WGCNA co-expression network analysis reveals ILF3-AS1 functions as a CeRNA to regulate PTBP1 expression by sponging miR-29a in gastric cancer. *Front genet*, **11**, 39.
- Snider L, Thirlwell H, Miller JR, et al (2001). Inhibition of Tcf3 binding by I-mfa domain proteins. *Mol Cell Biol*, **21**, 1866-73.
- Sui Y, Li X, Oh S, et al (2020). Opposite roles of the JMJD1A interaction partners MDFI and MDFIC in colorectal cancer. *Sci Rep*, **10**, 1-11.
- Tang Z, Li C, Kang B, Zhang ZJNAR, et al (2017). GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*, **45**, W98-W102.
- Terkelsen T, Russo F, Gromov P, et al (2020). Secreted breast tumor interstitial fluid microRNAs and their target genes are associated with triple-negative breast cancer, tumor grade, and immune infiltration. *Breast Cancer Res*, **22**, 1-36.
- Tokumaru Y, Oshi M, Huyser MR, et al (2021). Low expression of miR-29a is associated with aggressive biology and worse survival in gastric cancer. *Sci Rep*, **11**, 1-11.
- Vizeacoumar FS, Guo H, Dwernychuk L, et al (2021). Mining the plasma-proteome associated genes in patients with gastro-esophageal cancers for biomarker discovery. *Sci Rep*, **11**, 1-13.
- Wang B, Zhang Y, Qing T, et al (2021). Comprehensive analysis of metastatic gastric cancer tumour cells using single-cell RNA-seq. *Sci Rep*, **11**, 1-10.
- Xie Z, Bailey A, Kuleshov MV, et al (2021). Gene Set Knowledge Discovery with Enrichr. *Curr Protoc*, **1**, e90.
- Zhou Q, Zhou L-Q, Li S-H, et al (2020). Identification of subtype-specific genes signature by WGCNA for prognostic prediction in diffuse type gastric cancer. *Aging (Albany NY)*, **12**, 17418.



This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.