Original Research *Highlight article*

Predicting functional circular RNA-based competitive endogenous RNA network in gastric carcinoma using novel bioinformatics analysis

Zirui Zhu[1], Rui Huang[2] and Baojun Huang[1]

[1]Department of Surgical Oncology and General Surgery, Key Laboratory of Precision Diagnosis and Treatment of Gastrointestinal Tumors, Ministry of Education, The First Affiliated Hospital of China Medical University, Shenyang, Liaoning 110001, China; [2]Department of Clinical Medicine of Year 2017, Da Lian Medical University, Dalian, Liaoning 116044, China
Corresponding author: Baojun Huang. Email: 905317555@qq.com

Impact statement

Gastric cancer (GC) remains one of the most prevalent types of malignancies worldwide and the third reported primary factor to account for tumor-related deaths in 2018. This article illuminated the underlying circular RNAs-based molecular mechanism in GC by developing a new way of thinking and combining multiple bioinformatics methods. Competing endogenous RNA mechanism has become the spotlight of intense research during these recent years while most studies only focused on one regulatory network. A series of novel inclusion criteria were selected in our research, and finally, two ceRNA regulatory networks were built to unveil the detailed mechanism in GC. Ultimately, specific circular RNAs were located and can be developed as reliable biologically targets to participate in early screening, identification, and therapy of GC to improve clinical outcomes.

Abstract

Gastric cancer (GC) remains one of the most prevalent types of malignancies worldwide, and also one of the most reported lethal tumor-related diseases. Circular RNAs (circRNAs) have been certified to be trapped in multiple aspects of GC pathogenesis. Yet, the mechanism of this regulation is mostly undefined. This research is designed to discover the vital circRNA-microRNA (miRNA)-messenger RNA (mRNA) regulatory network in GC. Expression profiles with diverse levels including circRNAs, miRNAs, and mRNAs were all determined using microarray public datasets from Gene Expression Ominous (GEO). The differential circRNAs expressions were recognized against the published robust rank aggregation algorithm. Besides, a circRNA-based competitive endogenous RNA (ceRNA) interaction network was visualized via Cytoscape software (version 3.8.0). Functional and pathway enrichment analysis associated with differentially expressed targeted mRNAs were conducted using Cytoscape and an online bioinformatics database. Furthermore, an interconnected protein–protein interaction association network which consisted of 51 mRNAs was predicted, and hub genes were screened using STRING and CytoHubba. Then, several hub genes were chosen to explore their expression associated with survival rate and clinical stage in GEPIA and Kaplan-Meier Plotter databases. Finally, a carefully designed circRNA-related ceRNA regulatory subnetwork including four circRNAs, six miRNAs, and eight key hub genes was structured using the online bioinformatics tool.

Keywords: Bioinformatics, GC, gene, medicine/oncology/pathology, signaling, regulation

***Experimental Biology and Medicine* 2022; 247: 131–144. DOI: 10.1177/15353702211048757**

Introduction

Gastric cancer (GC) remains one of the most prevalent types of malignancies worldwide and the third reported primary factor to account for tumor-related deaths in 2018.[1] Although substantial progress has been made in personalizing treatment for GC, clinical outcomes of GC are still not desirable, with a five-year overall survival (OS) <30%.[2] The low OS is due primarily to an unsatisfactory early diagnostic rate, and nearly one-third of sufferers with GC come through fatally regional and distant recurrence or metastasis after experiencing traditional surgical or non-surgical intervention.[3] It is therefore pressing to develop reliable biologically targets for early screening, identification, and therapy of GC to improve clinical outcomes.

International Human Genome Sequencing Consortium has proclaimed that less than 2% of the overall human genome encodes for protein-coding genes, yet the enormous parts of recorded transcripts cover non-protein-coding RNAs such as microRNA (miRNA), long noncoding RNA (lncRNA), or circular RNA (circRNA).[4] CircRNA, a newly noticed non-coding RNA that typified a continuous closed circular loop fabric, represents plentiful striking features and functions compared to linear RNAs. Due to the circle shorts of the 3′-poly A tail and the 5′-cap, circRNAs are more structurally stable in specific tissues and organs than their certain counterparts as well as strongly resistant to the degradation by exonuclease RNase R.[5,6] Above all, numerous studies have declared that circRNAs play momentous roles in tumorigenesis and advanced carcinoma ascribed to their generous biological purpose on tumor cells migration, proliferation, invasion, and metastasis.[7–9] CircRNAs have been piecemeal researched nearly 50 years from initial discovery in 1971[10]; however, they have been widely thought to occur as the unexpected outcome of spicing errors, and their biological mechanisms leading to GC remain chiefly uncharted. Thanks to the speedy forwards in high-throughput RNA sequencing (RNA-Seq) combined with bioinformatics technology in recent years, circRNAs, for the most part, have been explored and a handful of rudimentary functions of circRNAs also uncovered, for instance, serving as protein scaffolds in the congregated of multi-subunit protein complexes,[11] confiscating certain proteins from their intrinsic sub-cellular localization,[12] forming RNA-protein complexes by incorporating with RNA-related proteins to manage transcriptional gene regulation,[13] and modulating the relative expression of their parental genes.[14] Moreover, it is worth mentioning that some of them can even be translated into proteins.[15] At present, absorbing miRNAs as miRNA response element (MRE) is one of the best-representative characteristics among circRNAs, also known as the ceRNA mechanism.[16–18] For instance, circular RNA circCCDC9 checked the growth of gastric carcinoma via binding miR-6792-3p.[19] Another research found that augmented expression of circNRIP1 obviously promoted the progression of GC by handling the miR-149-5p/AKT1/mTOR axis.[20] Further experiments showed that circLMTK2 promoted GC-related proliferation and metastasis by sponging miR-150-5p; meanwhile, a high circLMTK2 expression level corresponded to a high risk of lymph node metastasis, the advanced TNM clinical stage, and an unsatisfactory prognostic outcome in GC patients.[21]

Despite many exciting features and functions in circRNA biology have been clarified, how regulatory networks control circRNA molecular mechanism is still mostly ill-defined. In this research, we downloaded and analyzed the expression profiles with diverse levels including circRNAs, miRNAs, and messenger RNA (mRNAs) in GC-based tissues along with abutting healthy gastric mucosa, and plasma specimens of GC patients and healthy controls via manifold training sets of the GEO public database. Afterward, we fabricated a regulatory network consisting of 5 circRNAs, 36 miRNAs, 125 mRNAs by spotting the circRNA-sponged miRNAs together with possible targeted mRNAs, and then we investigated the relevant miRNAs and mRNAs in this regulatory network by utilizing several online prediction websites, various packages in R, the Gene Ontology (GO) annotation, and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. Moreover, a protein–protein interaction (PPI) interaction network was structured and hub genes were sought from this interconnection modules, and subsequently, the association between hub genes and the clinical prognosis was further assessed. A regulatory module of circRNAs-miRNAs-hub genes subnetwork was finally structured to systematically comprehend the potential functional circRNAs-based mechanism in the pathogenesis and advance of GC.

Materials and methods

Data collection

The public microarray datasets applied in our research were acquired from the GEO databank (https://www.ncbi.nlm.nih.gov/gds/). Five expression profiles of circRNAs were attained among GSE78092 (containing three couples of GC tissue and adjacent healthy mucosa tissue), GSE83521 (containing six couples of GC tissue and adjacent normal mucosa tissue), GSE89143 (containing three couples of GC tissue and matched precancerous tissue), and GSE93541 (containing three couples of GC plasma samples and healthy controls). Expression profiles of miRNAs were retrieved from GSE23739 (containing 40 couples of GC tissue and adjacent healthy mucosa tissue), GSE54397 (containing 16 couples of GC tissue and adjacent healthy mucosa tissue regardless of the *Helicobacter pylori* infection), GSE78091 (containing three couples of GC tissue and adjacent healthy mucosa tissue), and GSE93415 (containing 20 couples of GC tissue and adjacent healthy mucosa tissue). In addition, expression profiles of mRNAs were extracted from GSE54129 (containing 111 cancerous mucosae and 21 healthy noncancerous mucosae of GC tissues). No informed consent nor ethical committee approval was demanded in the current research owning to the public-free accessibility of datasets from the GEO domain training sets.

Recognition of differentially expressed RNAs

The downloaded microarray datasets including series of matrix files and platform documents were both performed resorting to the latest *R* language software and correlative annotation packages. Then circRNA ID terms were transformed into internationally recognized normal denominations (circRNA symbols). The identification of differentially expressed circRNAs (DEcircRNAs) was carried out by means of the Bioconductor Limma R package in each dataset. We further merged and then ranked DEcircRNAs refer to the robust rank aggregation algorithm.[22] Besides, differentially expressed miRNAs (DEmiRNAs) were discovered via GEO2R in datasets GSE23729, GSE54397, GSE78091, and GSE93415, as the absolute value of log2 Fold Change (Log2FC) is smaller than 1 in the midst of most miRNAs;

therefore, the inclusion criterion that False Discovery Rate (FDR) values <0.05 and |log2FC|>0.05 was rationally supposed to be statistically significant. In addition, differentially expressed mRNAs (DEmRNAs) were also sorted within the threshold that FDR values <0.05 and |log2FC|>1 by using GEO2R in dataset GSE54129.

Prediction of miRNA-sponged locus

According to the consequence of the robust rank aggregation algorithm, we chose the top five of DEcircRNAs called IDEcircRNAs. MiRNA binding sites (MREs) of IDEcircRNAs were forecasted via The Cancer-Specific CircRNA Disease (CSCD) (http://gb.whu.edu.cn/CSCD/) and The Circular RNA Interactome (CircInteractome) (https://circinteractome.nia.nih.gov/) (we denominated them as CPmiRNAs), and after that, we computed the shared junction of CPmiRNAs and DEmiRNAs, associated with the inclusion criteria were the convergence between CSCD (and/or CircInteractome) and more than two GSE datasets, which we named ICPDEmiRNAs.

Prediction of mRNA target genes

Perl script was applied to predict respectively ICPDEmiRNAs targeted genes based on the three target gene prediction online systems, comprising miRDB (http://www.mirdb.org/), miRTarBase (http://mirtarbase.mbc.nctu.edu.tw/), and TargetScan (http://www.targetscan.org/).[23–25] Not all outcomes were included, mutually owned target genes predicted within three databases were recognized as mRNAs candidates, and we called TmRNAs. Similarly, we took the shared sections of them together with the above-mentioned DEmRNAs, and then we obtained some of the final potentially functional mRNAs, also called FmRNAs.

Visualization of the ceRNA regulatory mechanism

Following the prediction of all targeted RNAs, a circRNA-based ceRNA network was built utilizing the amalgamation of circRNA-miRNA couples plus miRNA-mRNA couples in compliance with the noble ceRNA components interaction mechanism and the dynamic swings of the triples. Finally, the hypothesis network was beautifully portrayed by Cytoscape software.

Functional GO and KEGG analysis

In order to explore the primary potentially functional enrichment traits in GC, FmRNAs in the midst of ceRNA triples were appraised via the GO terms and the KEGG pathways analysis through http://www.bioinformatics.com.cn, an online platform for data analysis and visualization, along with the *P* values <0.05 as the inclusion measures. DAVID (http://david.abcc.ncifcrf.gov/) was also implemented to compute the statistical enrichment of target genes in the KEGG pathways. What's more, the network of the KEGG signaling pathways together with pathway-relevant mRNAs was structured by means of Cytoscape and its plugin GlueGO[26] as well as Cluepedia[27] app.

Construction of PPI network

The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING, http://string-db.org) analysis tool was put into use for building a particular PPI group based on the FmRNAs identified. Visualization was rendered using Cytoscape, and the top 10 nodes (hub genes) ranked by Degree algorithm was chosen in plugin CytoHubba. Also, their expression and correlation with survival rate and clinical stage were summarized in Kaplan-Meier Plotter (https://kmplot.com/analysis/) as well as GEPIA (http://gepia.cancer-pku.cn/).

Results

Identification of DEcircRNAs, DEmiRNAs, and DEmRNAs

Differentially expressed circRNAs profiles in GC tissue along with plasma and negative controls were assessed among the four GEO series containing GSE78092, GSE83521, GSE89143, and GSE93541, and their the fundamental components are demonstrated in Table 1. Totally, 112 DEcircRNAs (23 up-regulated and 89 down-regulated) were screened in the dataset GSE78092; 73 DEcircRNAs (43 up-regulated and 30 down-regulated) were screened in the dataset GSE83521; 53 DEcircRNAs (8 up-regulated and 45 down-regulated) were screened in the dataset GSE89143; 306 DEcircRNAs (146 up-regulated and 160 down-regulated) were screened in the dataset GSE93541. We further integrated and then ranked the DEcircRNAs from the four different datasets, and totally, five circRNAs, one up-regulated and four down-regulated, were demonstrated in the top rank using heatmap ($P < 0.05$) (Figure 1), which were known as hsa_ circ_0067934, hsa_circ_0007763, hsa_circ_0007991, hsa_circ_0048607, and hsa_circ_0064557, and the elementary characteristics of the five DEcircRNAs from CSCD are given in Table 2. Besides, their constructional features were explored, which was exhibited in Figure 2. What's more, the differential expression profiles of miRNAs in GC tissue along with adjacent control counterpart was assessed in four microarrays datasets (GSE23739, GSE54397, GSE78091 and GSE93415), and their fundamental details were displayed in Table 3. The whole of 52 DEmiRNAs, 49 up-regulated and 3 down-regulated, were screened in the dataset GSE23739. The whole of 368 DEmiRNAs, 153 up-regulated and 215 down-regulated, were screened in the dataset GSE54387. The whole of 795 DEmiRNAs, 643 up-regulated and 152

Table 1. Basic information of the 4 circRNA microarray datasets from GEO.

Dataset	Platform	Sample volume (T/N)	Country	Author	Year
GSE78092	GPL21485	3/3	China	Huang	2016
GSE83521	GPL19978	6/6	China	Zhang	2017
GSE89143	GPL19978	3/3	China	Guo	2017
GSE93541	GPL19978	3/3	China	Guo	2017

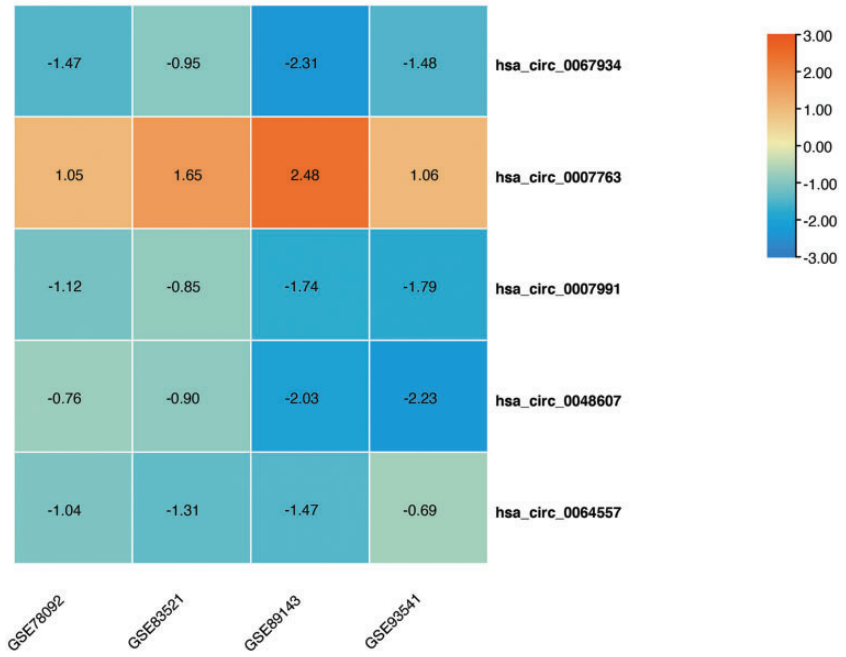

Figure 1. Heatmap of the five differentially co-expressed circRNAs among the four GEO datasets. (A color version of this figure is available in the online journal.)

Table 2. Basic characteristics of the five DEcircRNAs.

circRNA symbol	Position	Strand	Genomic length	Best transcript	Gene symbol	Regulation
hsa_circ_0007763	chr16:3526221-3529600	+	3379	NM_001083601	NAA60	up
hsa_circ_0007991	chr1:21329205-21415706	–	86501	NM_001198801	EIF4G3	down
hsa_circ_0048607	chr19:4408899-4409756	+	857	NM_005483	CHAF1A	down
hsa_circ_0064557	chr3:18456602-18462483	–	5881	NM_001195470	SATB1	down
hsa_circ_0067934	chr3:170013698-170015181	+	1483	NM_002740	PRKCI	down

down-regulated, were screened in the dataset GSE78091. The whole of 629 DEmiRNAs, 264 up-regulated and 365 down-regulated, were screened in the dataset GSE23739. Similarly, the same analysis was performed on GSE54129, and entirely, 4345 differentially expressed mRNAs were sorted, with 482 up-regulated and 3863 down-regulated mRNAs.

Looking for the interaction among circRNA, miRNA, and mRNA

According to our inclusion principle, totally, 36 ICPDEmiRNAs were obtained from the varied intersection of CPmiRNAs and DEmiRNAs and then visualized using the TBtools software (Figure 3).[28] In the case of target gene prediction, 640 TmRNAs of 36 ICPDEmiRNAs were forecasted after removing the repeated items. At last, 125 finally mRNAs (FmRNAs) were screened via examination of the overleaped intersection from TmRNAs and previously acquired 4345 DEmRNAs.

Visualization of the ceRNA regulatory mechanism

ceRNA network of the molecular level in GC was established for a deeper understanding of the mutual relationship among circRNAs, miRNAs, and mRNAs. In light of previous data results, 5 IDEcircRNAs including 1 up-regulated and 4 down-regulated objects, 36 ICPDEmiRNAs including 22 up-regulated and 14 down-regulated objects, 125 FmRNAs including 18 up-regulated and 129 down-regulated objects and their relationships were described in the ceRNA network and their relative size was computed using the Enrichment Degree in Cytoscape (Figure 4).

GO and KEGG analysis of FmRNAs

Next, 125 FmRNAs sifted were applied to further investigate the potentially functional circRNAs, containing GO annotation and KEGG pathways. All the GO annotation outcomes within BP, CC, and MF were ranked respectively according to the enrichment scores (-log(*P*-value)) and top 10 of those represented in Figure 5. In the analysis of biological process (BP), intestinal epithelial cell development, intestinal epithelial cell differentiation, and vesicle targeting, rough ER to cis-Golgi were the top three most significantly enriched categories. In the analysis of cellular component (CC), complex of collagen trimers, extracellular matrix component, and cyclin-dependent protein kinase holoenzyme complex were the top three most significantly enriched categories. In the analysis of molecular function (MF), mRNA 3′-UTR AU-rich region binding, AU-rich

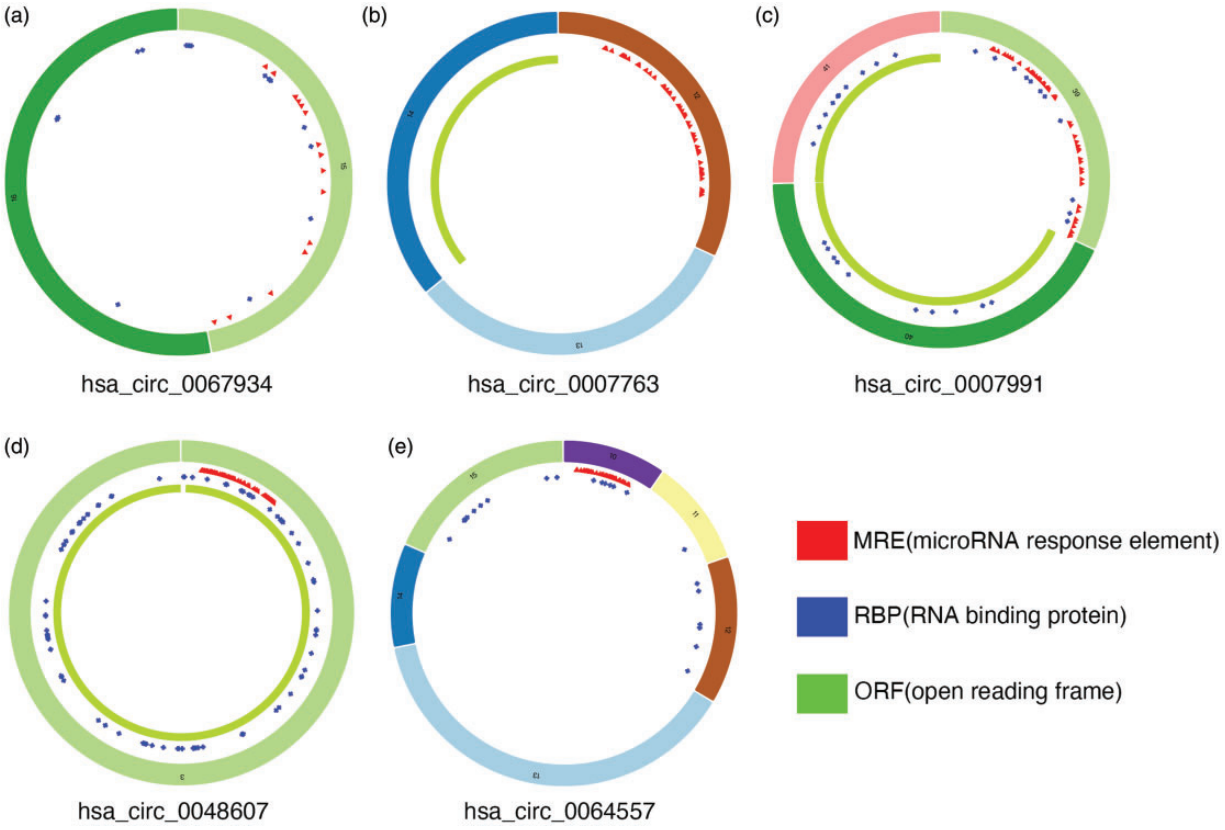

Figure 2. The circular genome structure of five circRNAs. (a) hsa_circ_0067934. (b) hsa_circ_0007763. (c) hsa_circ_0007991. (d) hsa_circ_0048607. (e) hsa_circ_0064557. (A color version of this figure is available in the online journal.)

Table 3. Basic information of the 4 miRNA microarray datasets from GEO.

Dataset	Platform	Sample Volume (T/N)	Country	Author	Year
GSE23739	GPL7731	40/40	Switzerland	Oh H	2011
GSE54397	GPL15159	16/16	South Korea	Chang H	2014
GSE78091	GPL21439	3/3	China	Huang	2016
GSE93415	GPL19071	20/20	Poland	Sierzega M	2017

element binding, and transcription coactivator binding were the top three most significantly enriched categories. Besides, to better understand the mutual relationships of the three main categories of GO terms and the co-expressed genes, GO Chord with enrichment scores >15 in three terms was plotted (Figure 6). Then the KEGG pathways were further ranked using the enrichment scores algorithm and the top 10 most KEGG-enriched pathways for this network module were revealed in Figure 7, in which the p53 signaling pathway was significantly enriched with the smallest *P* value. In addition, the network consisted of the most significantly enriched modules, together with their correlative target mRNAs indicated that CCND1, CASP3, COL4A1, COL1A2, and SMAD4 were all cross-talk genes that involved no less than four crucial signaling pathways (Figure 8).

Identification of hub genes

The STRING tool and Cytoscape were put into use so as to additionally investigate the hub genes among 125 FmRNAs, the result showed that 51 genes were co-related, and then the highly interconnected genes were clustered into three modules (Figure 9). The top 10 nodes ordered by Degree algorithm were used in Plugin CytoHubba to represent the most momentous gene modules (Hub genes), and then the network diagram of top 10 hub genes was exhibited in Figure 10, and the 8 of 10 hub genes were highly inter-correlated we called key hub genes. Also, we found that COL1A2 (Figure 11), COL4A1 (Figure 12), COL5A2 (Figure 13), and FBN1 (Figure 14) were both highly differentially expressed not only in GEPIA but in our present PPI network results. Besides, we discovered that over-expressed COL4A1 and COL5A2 had a reduced OS rate both in GEPIA as well as Kaplan-Meier Plotter website ($P < 0.05$) (Figures 15 to 18). Moreover, the up-regulated expression of COL1A2 (Figure 19), COL5A2 (Figure 20), and FBN1 (Figure 21) was associated with an increased clinical stage in GEPIA ($P < 0.05$). Ultimately, a circRNA-based ceRNA interaction sub-module was structured using an alluvial plot based on the consequence of sifted hub genes, as shown in Figure 22, four circRNAs, six miRNAs, and eight key hub genes were included as the final ceRNA network.

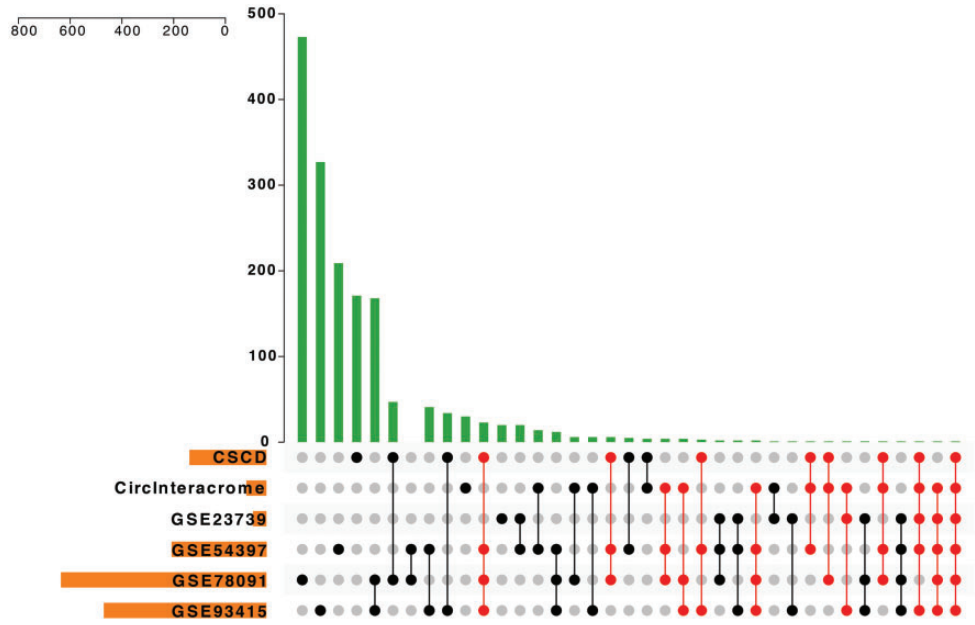

Figure 3. The various intersection of CPmiRNAs and DEmiRNAs. (A color version of this figure is available in the online journal.)

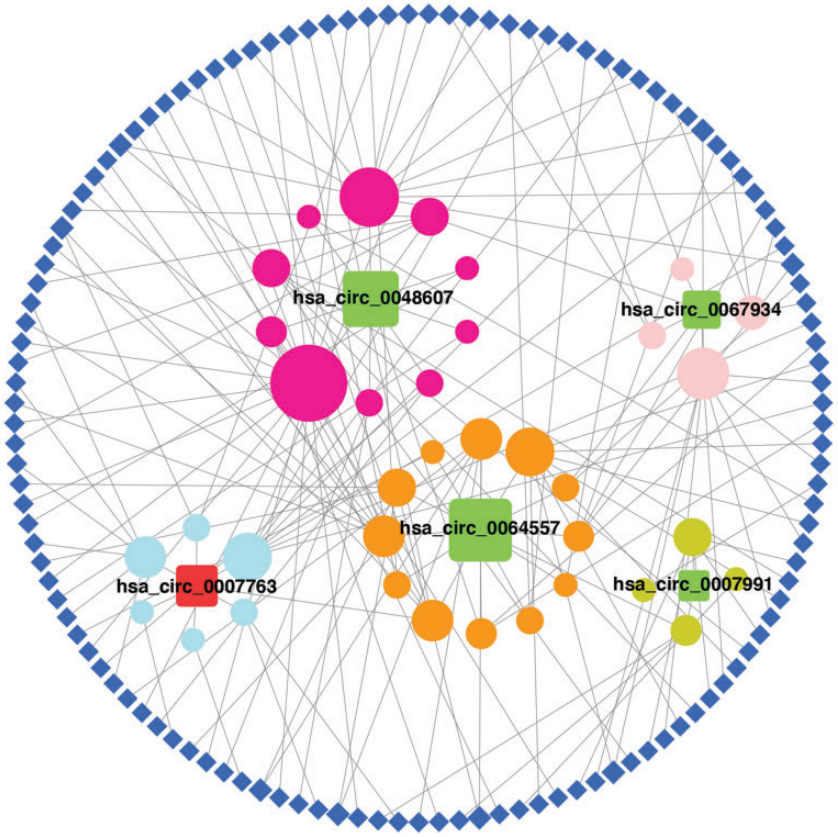

Figure 4. The circRNA-based ceRNA network. Squares refer to circRNAs, in which the nodes spotlighted in red and green respectively indicate up-regulated and down-regulated expression. Rotundities refer to miRNAs. Diamonds refer to mRNAs. The relative size of them was computed using the Enrichment Degree in Cytoscape. (A color version of this figure is available in the online journal.)

Discussion

Despite of many signs of treatment progress, GC remains one of the deadliest solid tumors with a complicated molecular mechanism that is still not well understood. In the last few years, intensive attempts have been made to elucidate the underlying molecular mechanisms in gastric cancer, but mainly focusing on long non-coding RNA or microRNAs.[29,30] CircRNAs, as a sort of novel non-protein

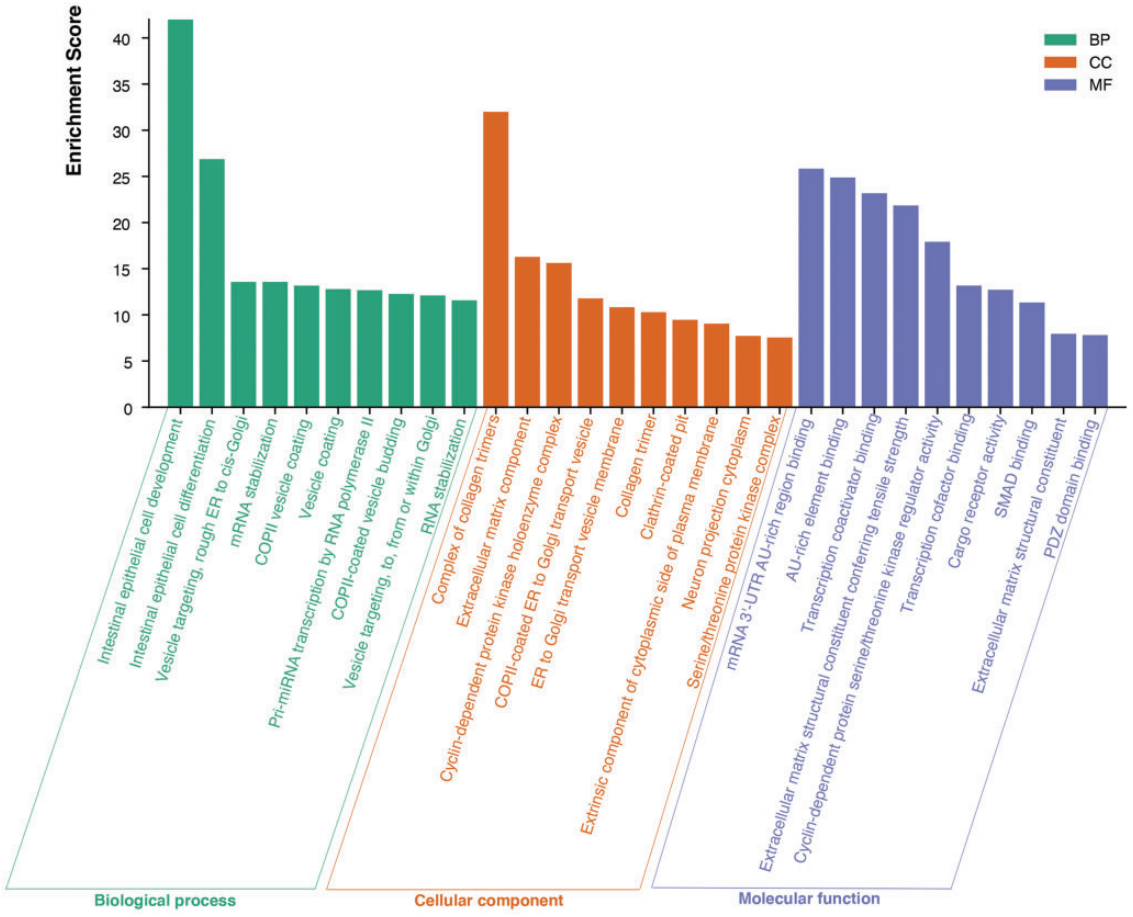

Figure 5. The Top 10 of enrichment scores in GO annotation. (A color version of this figure is available in the online journal.)

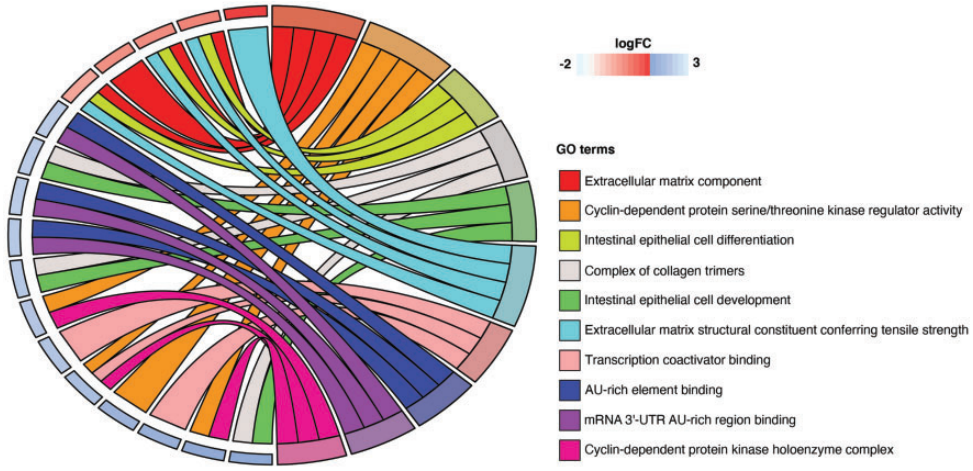

Figure 6. GO Chord with enrichment scores in three terms >15. The genes are associated with their target terms through colored ribbons. Genes are ranked in accordance with the log Fold Change (logFC), red and blue refer to the up- and down-regulated genes respectively. (A color version of this figure is available in the online journal.)

coding RNAs, have become one of the hottest topics in biomedicine and medicine owing to their highly developmental specificity and stability, and plentiful papers have also uncovered that circRNAs have great implications for diverse human diseases, including malignant tumors.[5,31,32] Several studies have found how circRNAs involve in multiple biological processes, which are considered to participate in tumorigenesis and tumor progression.[8,33] Other evidence has also proved that circRNAs can be served as the miRNA-sponged locus by competitive binding to miRNAs and consequently regulate their respective targeted mRNAs, hence influencing the complex biological and molecular behaviors of carcinomas. Nevertheless, the precise role of circRNAs in GC is still

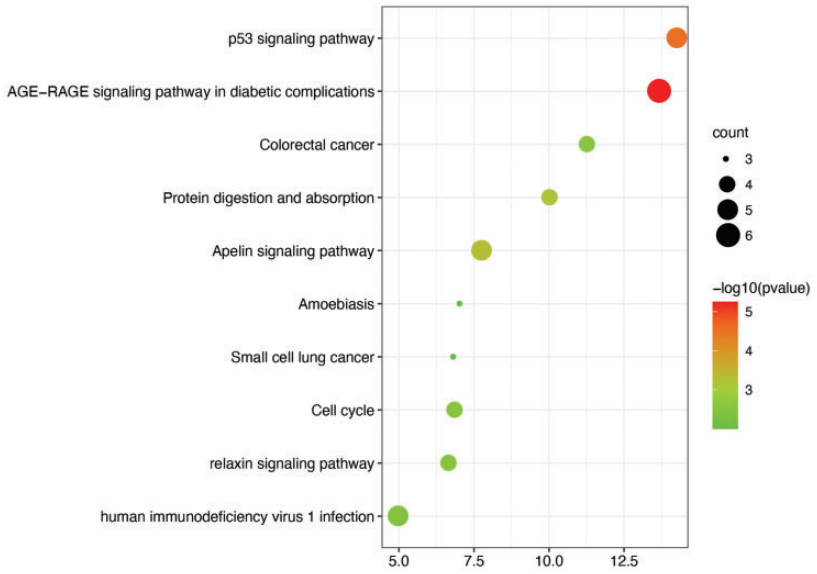

Figure 7. The Top 10 most KEGG-enriched pathway for the ceRNA network module. (A color version of this figure is available in the online journal.)

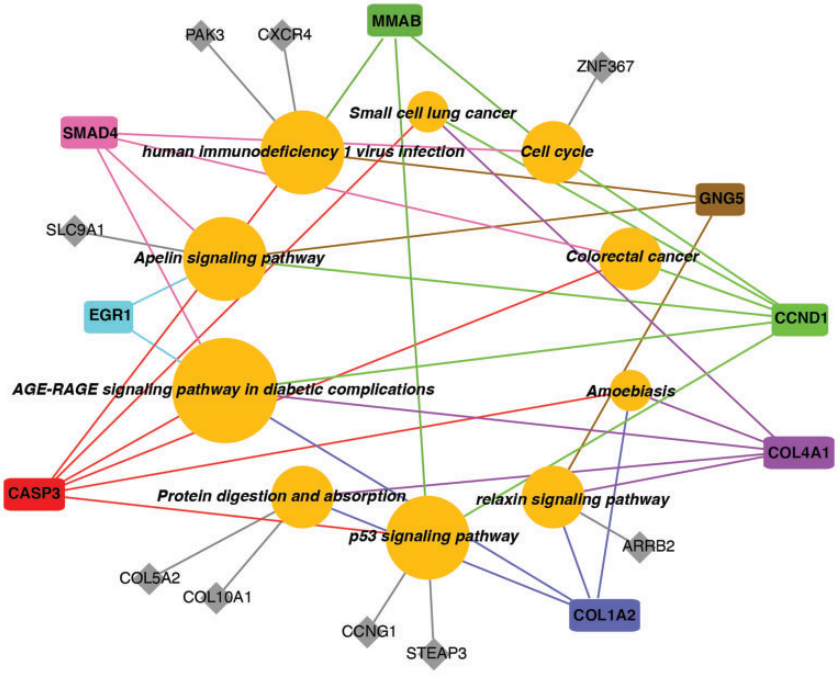

Figure 8. The network consisted of the most significantly KEGG-enriched pathways and their correlative target mRNAs. (A color version of this figure is available in the online journal.)

obscure. In this current study, a handful of fresh circRNAs were discovered which regulated the downstream of target mRNAs employing sponge-absorbed miRNAs. We further developed the functional annotations within these genes so as to promote a greater depth of understanding of circRNAs.

In our study, after multiple intersections of differentially expressive circRNAs, miRNAs, and mRNA, a circRNA-based ceRNA interaction network was explored by means of bioinformatics and extended a PPI regulatory module including 51 interconnected mRNAs. Moreover, the top 10 hub genes we exhibited within the PPI modules indicated that eight key hub genes were highly interrelated, including CCND1, COL1A2, SMAD4, CASP3, HIST2H2BE, COL4A1, COL5A2, and FBN1. Finally, a ceRNA regulatory alluvial plot was constructed according to the eight key hub genes.

Additionally, functional profiling analysis, containing GO annotations and KEGG pathways, was carried out based upon the selective miRNAs and the targeted FmRNAs. Then GO Chord with enrichment scores >15 in three respective terms was built for representing the connections between the existing GO terms and crucial genes. The result of cancer-related KEGG signaling pathway

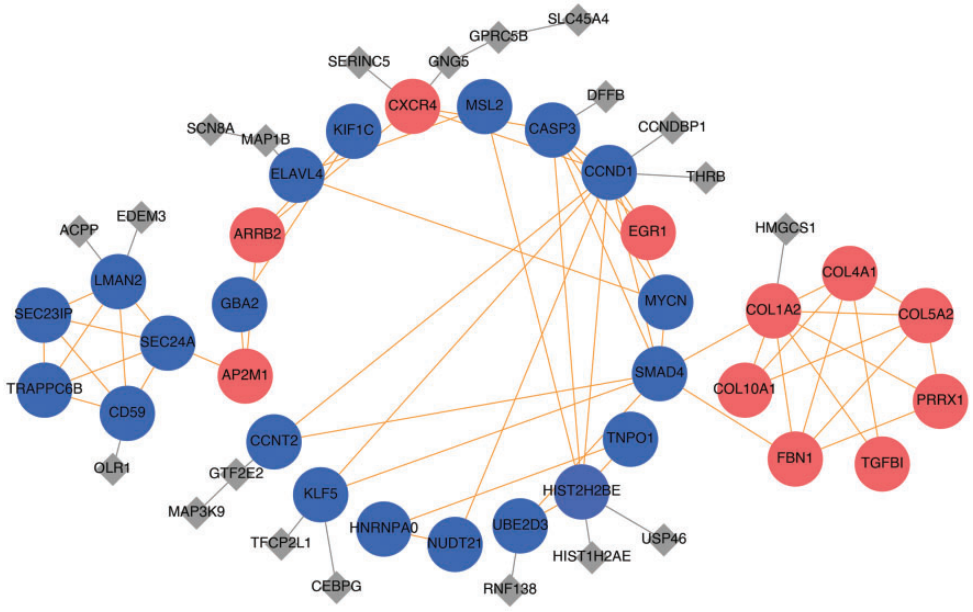

Figure 9. PPI network of 51 interconnected mRNAs. (A color version of this figure is available in the online journal.)

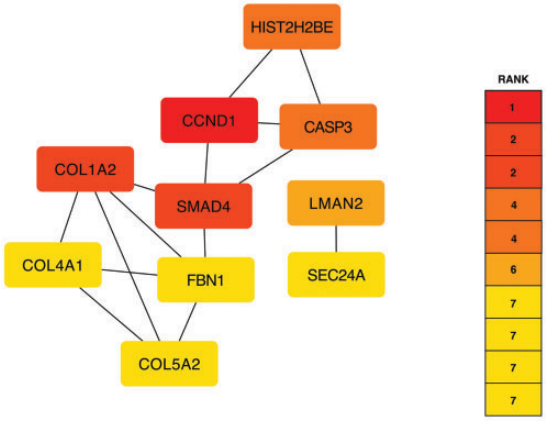

Figure 10. PPI network of top 10 hub genes. (A color version of this figure is available in the online journal.)

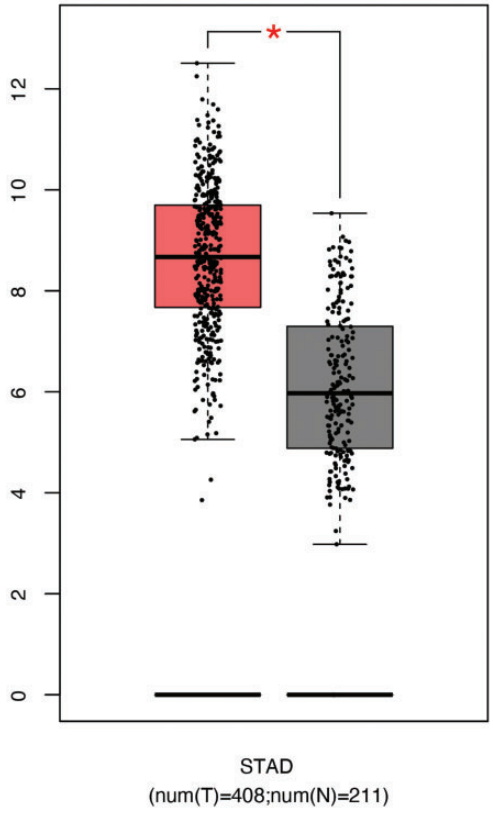

Figure 11. Co-overexpressing COL1A2 both in GEPIA and our PPI network result. (A color version of this figure is available in the online journal.)

analysis suggested that the p53 signaling pathway (CCND1/CASP3/CCNG1/STEAP3/MMAB) was significantly enriched with the smallest *P* value. P53 protein, considered as a historically well-known cancer suppressor gene, is the core of the p53 signaling pathway, and mutation in p53 gene frequently associate with many human types of tumors, including gastric cancer.[34–36] Activation of p53 as a transcription factor may affect DNA repair, cellular apoptosis, cell cycle arrest, and metastasis.[37] In gastric cancer, Fu et al. found that curcumin induced proliferation, autophagy, and apoptosis via mobilizing the p53 signaling pathway as well as blocking the PI3K pathway.[38] Besides, Liu et al. found that the over-expressed lncRNA GAS5 restrained the invasion and migration of GC cells via modulating the p53 signaling pathway.[39] As displayed in Figure 8, CCND1, CASP3, COL4A1, COL1A2, and SMAD4 were all momentous cross-talk genes that involved no less than four crucial signaling pathways, indicating that they might as bridges linking a number of interactions. In the meantime, the expression of COL4A1 and COL1A2 in the KEGG signaling pathway was in accordance with the GEPIA database outcomes. Their clinical stages and overall survivals combined with their respective level of expression were also excavated. We could conclude that high expression of COL4A1 was linked with a poor prognosis

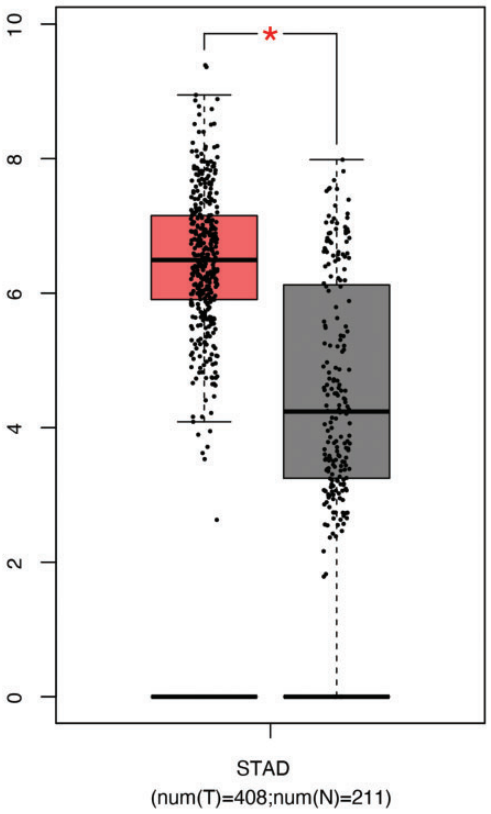

Figure 12. Co-overexpressing COL4A1 both in GEPIA and our PPI network result. (A color version of this figure is available in the online journal.)

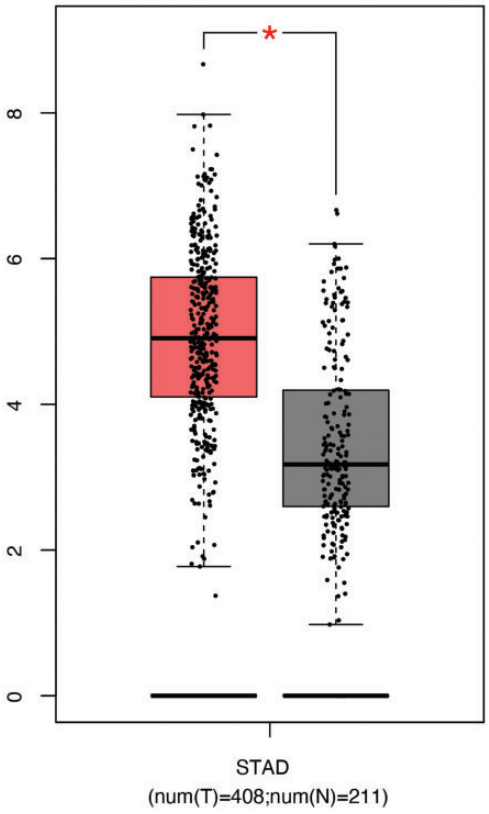

Figure 13. Co-overexpressing COL5A2 both in GEPIA and our PPI network result. (A color version of this figure is available in the online journal.)

both in GEPIA together with Kaplan-Meier Plotter online database, while the expression of COL1A2 was a significantly positive correlation with the clinical stage. Above all, COL4A1 and COL1A2 are a common family of proteins that strengthen and hold different types of tissues in humans, containing cartilage, bone, tendon, dermis, and the white sclera.[40] Collagen alpha-1(IV) chain (COL4A1) encodes a type IV collagen alpha protein, its mutations have been primarily reported in porencephaly, cerebrovascular and neuromuscular disease, and renal and ocular defects.[41,42] On top of this, the high expression of COL4A1 promotes proliferation, invasion, and migration in hepatocellular carcinoma (HCC) via activation of FAK-Src signaling.[43] Collagen alpha-2(I) chain is a protein encoded by the COL1A2 gene, which is the bulk significantly rich form of collagen in humans. At the same time, numerous studies have revealed that COL1A2 is closely related to cancers. For instance, COL1A2 belongs to TBX3 target gene that exerts an influence on the migration of chondrosarcoma and fibrosarcoma cells through the AKT1/TBX3/COL1A2 axis.[44] Apart from this, frequent 7q21-22 genomic region amplification and its intronic miR-25 is able to suppress a massive number of genes that are co-expressed with COL1A2, played a vital role in the angiogenesis and epithelial to mesenchymal transition (EMT), which are the representative characteristics in diffuse type of gastric carcinoma.[45] Additionally, it is found that COL1A2, COL6A3, and THBS2 were all highly expressed in stomach cancer, and suppressing COL1A2,

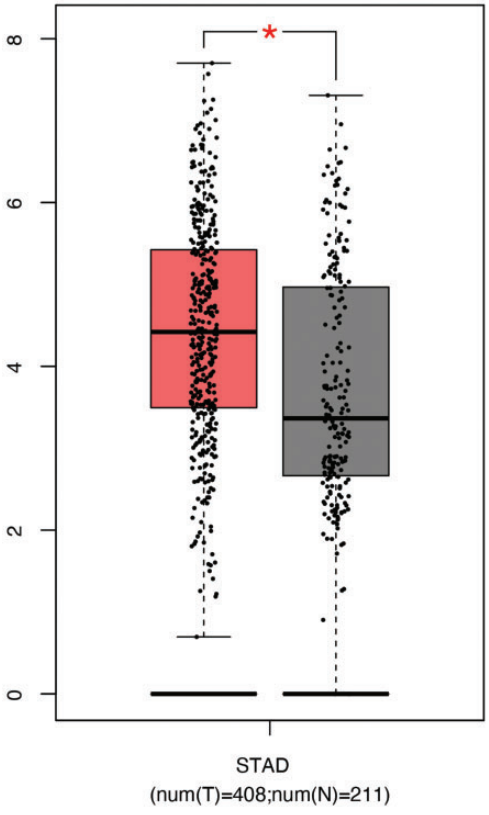

Figure 14. Co-overexpressing FBN1 both in GEPIA and our PPI network result. (A color version of this figure is available in the online journal.)

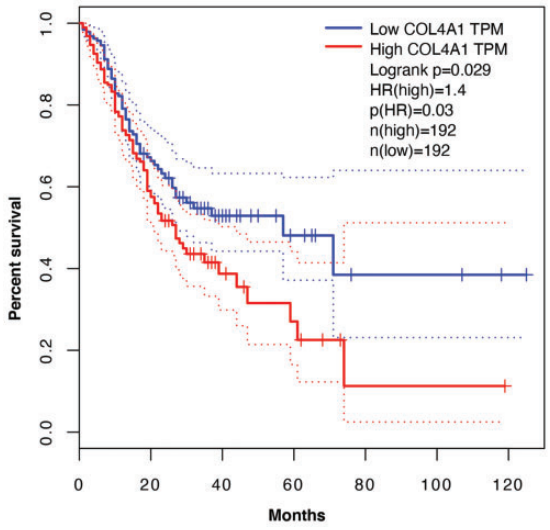

Figure 15. Overall survival of COL4A1 in GEPIA. (A color version of this figure is available in the online journal.)

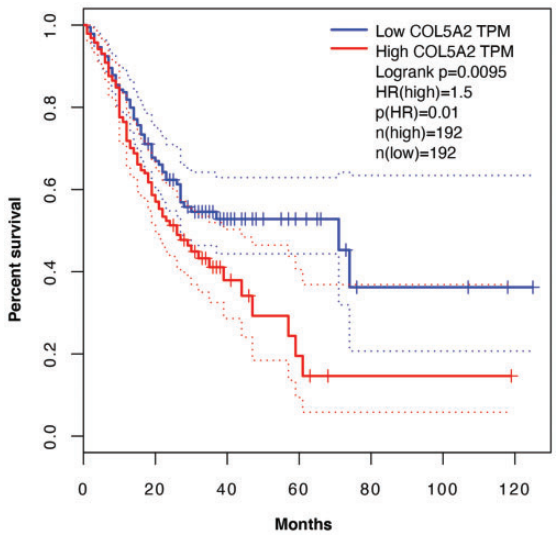

Figure 16. Overall survival of COL5A2 in GEPIA. (A color version of this figure is available in the online journal.)

COL6A3, and THBS2 exhibited inhibited proliferation, invasion, and migration via the PI3K-Akt signaling pathway.[46] Nevertheless, the relationship between circRNA and COL4A1 as well as COL1A2 still remains unclarified. By a combination of the most significantly enriched pathways in KEGG and the final ceRNA alluvial plot, two significantly central circRNA-based ceRNA modules were sifted, inclusive of hsa_circ_0048607/hsa-mir-4747-5p/COL1A2 and hsa_circ_0007763/hsa-mir-767-5p/COL4A1. Finally, this research is still at the stage of the preclinical hypothesis since all data analyzed were generating based upon the public domain training sets, the limitations of this study are the same as those that are intrinsic to the database itself; thus, selection bias, limited sample size, and mixed quality of chips should all deserve consideration. Beyond that, the limitations of

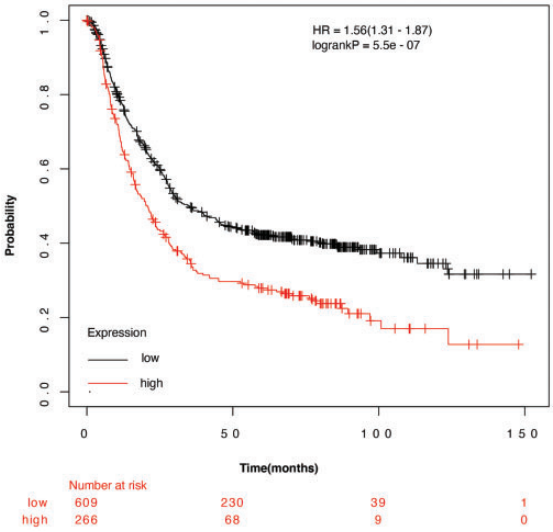

Figure 17. Overall survival of COL4A1 in Kaplan-Meier Plotter. (A color version of this figure is available in the online journal.)

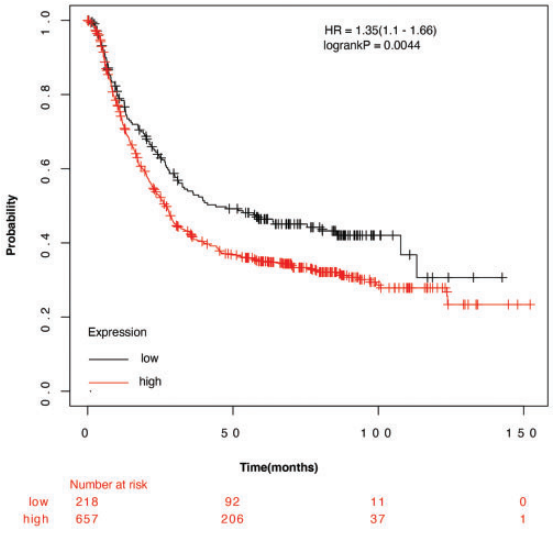

Figure 18. Overall survival of COL5A2 in Kaplan-Meier Plotter. (A color version of this figure is available in the online journal.)

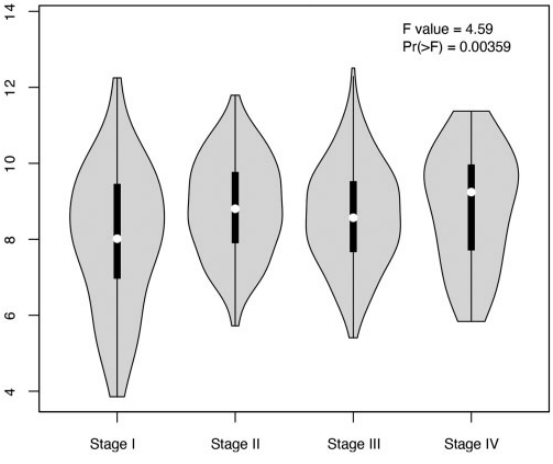

Figure 19. Association between COL1A2 expression and clinical stage of GC.

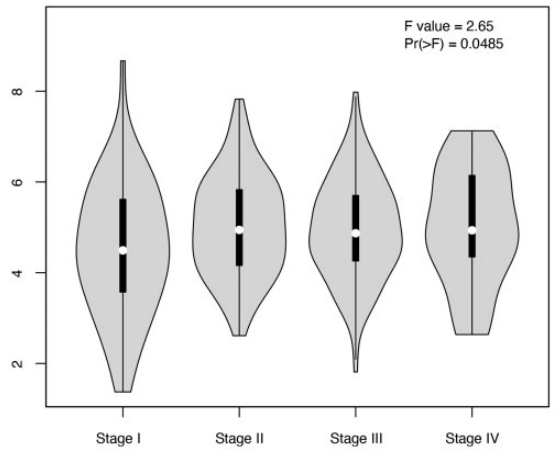

Figure 20. Association between COL5A2 expression and clinical stage of GC.

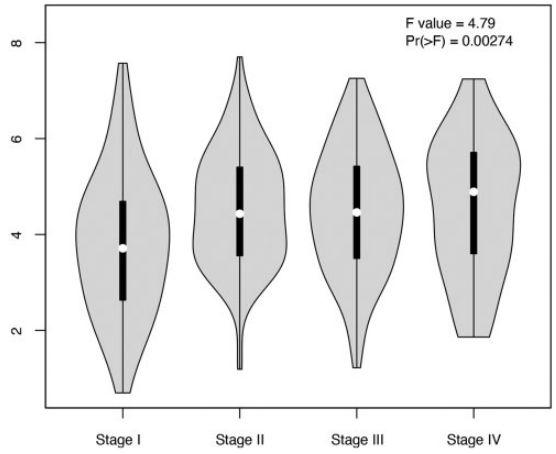

Figure 21. Association between FBN1 expression and clinical stage of GC.

multiplicity control also give more or less influence to our predictive model. More importantly, further exploration including clinical and experimental data as well as prospective evaluation was needed to verify our proposed GC-associated ceRNA crosstalk mechanism in future studies.

AUTHORS' CONTRIBUTIONS

ZZ conceived this current research and carried out data analysis; RH participated in data collection and prepared the figures. ZZ and RH drafted the primary manuscript. ZZ and BH reviewed and modified the manuscript. All authors have read and approved the final manuscript.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Zirui Zhu https://orcid.org/0000-0002-0939-3440

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;**68**:394–424
2. Allemani C, Weir HK, Carreira H, Harewood R, Spika D, Wang XS, Bannon F, Ahn JV, Johnson CJ, Bonaventure A, Marcos-Gragera R, Stiller C, Azevedo e Silva G, Chen WQ, Ogunbiyi OJ, Rachet B,

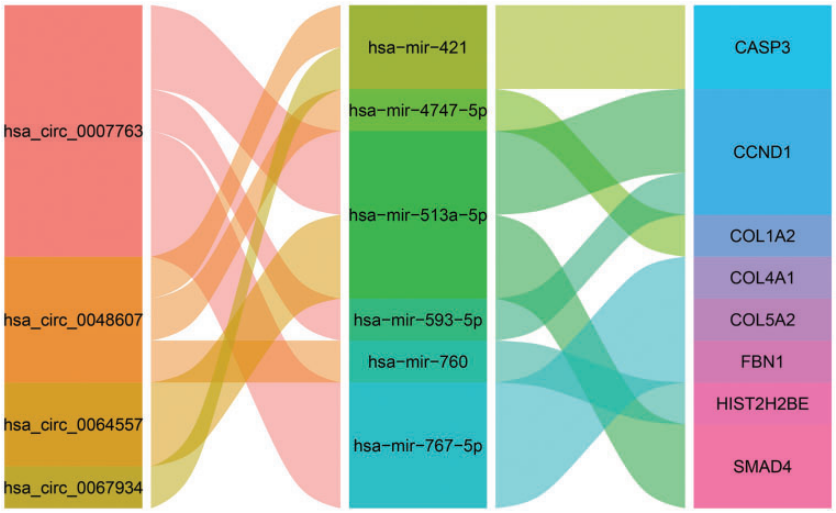

Figure 22. CeRNA alluvial plot constructed after screening the key eight hub genes. (A color version of this figure is available in the online journal.)

Soeberg MJ, You H, Matsuda T, Bielska-Lasota M, Storm H, Tucker TC, Coleman MP. Global surveillance of cancer survival 1995-2009: analysis of individual data for 25,676,887 patients from 279 population-based registries in 67 countries (CONCORD-2). *Lancet (Lancet)* 2015;**385**:977–1010

3. Spolverato G, Ejaz A, Kim Y, Squires MH, Poultsides GA, Fields RC, Schmidt C, Weber SM, Votanopoulos K, Maithel SK, Pawlik TM. Rates and patterns of recurrence after curative intent resection for gastric cancer: a United States multi-institutional analysis. *J Am Coll Surg* 2014;**219**:664–75
4. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakrabortty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Falconnet E, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena H, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Luo OJ, Park E, Persaud K, Preall JB, Ribeca P, Risk B, Robyr D, Sammeth M, Schaffer L, See LH, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Ruan X, Hayashizaki Y, Harrow J, Gerstein M, Hubbard T, Reymond A, Antonarakis SE, Hannon G, Giddings MC, Ruan Y, Wold B, Carninci P, Guigó R, Gingeras TR. Landscape of transcription in human cells. *Nature* 2012;**489**:101–8
5. Kristensen LS, Andersen MS, Stagsted LVW, Ebbesen KK, Hansen TB, Kjems J. The biogenesis, biology and characterization of circular RNAs. *Nat Rev Genet* 2019;**20**:675–91
6. Rybak-Wolf A, Stottmeister C, Glažar P, Jens M, Pino N, Giusti S, Hanan M, Behm M, Bartok O, Ashwal-Fluss R, Herzog M, Schreyer L, Papavasileiou P, Ivanov A, Öhman M, Refojo D, Kadener S, Rajewsky N. Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. *Mol Cell* 2015;**58**:870–85
7. Guarnerio J, Bezzi M, Jeong JC, Paffenholz SV, Berry K, Naldini MM, Lo-Coco F, Tay Y, Beck AH, Pandolfi PP. Oncogenic role of fusion-circRNAs derived from cancer-associated chromosomal translocations. *Cell* 2016;**165**:289–302
8. Guarnerio J, Zhang Y, Cheloni G, Panella R, Mae Katon J, Simpson M, Matsumoto A, Papa A, Loretelli C, Petri A, Kauppinen S, Garbutt C, Nielsen GP, Deshpande V, Castillo-Martin M, Cordon-Cardo C, Dimitrios S, Clohessy JG, Batish M, Pandolfi PP. Intragenic antagonistic roles of protein and circRNA in tumorigenesis. *Cell Res* 2019;**29**:628–40
9. Goodall GJ, Wickramasinghe VO. RNA in cancer. *Nat Rev Cancer* 2021;**21**:22–36
10. Sanger HL, Klotz G, Riesner D, Gross HJ, Kleinschmidt AK. Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc Natl Acad Sci U S A* 1976;**73**:3852–6
11. Wu N, Yuan Z, Du KY, Fang L, Lyu J, Zhang C, He A, Eshaghi E, Zeng K, Ma J, Du WW, Yang BB. Translation of yes-associated protein (Yap) was antagonized by its circular RNA via suppressing the assembly of the translation initiation machinery. *Cell Death Differ* 2019;**26**:2758–73
12. Armakola M, Higgins MJ, Figley MD, Barmada SJ, Scarborough EA, Diaz Z, Fang X, Shorter J, Krogan NJ, Finkbeiner S, Farese RV Jr, Gitler AD. Inhibition of RNA lariat debranching enzyme suppresses TDP-43 toxicity in ALS disease models. *Nat Genet* 2012;**44**:1302–9
13. Du WW, Yang W, Liu E, Yang Z, Dhaliwal P, Yang BB. Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2. *Nucleic Acids Res* 2016;**44**:2846–58
14. Li Z, Huang C, Bao C, Chen L, Lin M, Wang X, Zhong G, Yu B, Hu W, Dai L, Zhu P, Chang Z, Wu Q, Zhao Y, Jia Y, Xu P, Liu H, Shan G. Exon-intron circular RNAs regulate transcription in the nucleus. *Nat Struct Mol Biol* 2015;**22**:256–64
15. Legnini I, Di Timoteo G, Rossi F, Morlando M, Briganti F, Sthandier O, Fatica A, Santini T, Andronache A, Wade M, Laneve P, Rajewsky N, Bozzoni I. Circ-ZNF609 is a circular RNA that can be translated and functions in myogenesis. *Mol Cell* 2017;**66**:22–37.e9
16. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 2011;**146**:353–8
17. Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA crosstalk and competition. *Nature* 2014;**505**:344–52
18. Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J. Natural RNA circles function as efficient microRNA sponges. *Nature* 2013;**495**:384–8
19. Luo Z, Rong Z, Zhang J, Zhu Z, Yu Z, Li T, Fu Z, Qiu Z, Huang C. Circular RNA circCCDC9 acts as a miR-6792-3p sponge to suppress the progression of gastric cancer through regulating CAV1 expression. *Mol Cancer* 2020;**19**:86
20. Zhang X, Wang S, Wang H, Cao J, Huang X, Chen Z, Xu P, Sun G, Xu J, Lv J, Xu Z. Circular RNA circNRIP1 acts as a microRNA-149-5p sponge to promote gastric cancer progression via the AKT1/mTOR pathway. *Mol Cancer* 2019;**18**:20
21. Wang S, Tang D, Wang W, Yang Y, Wu X, Wang L, Wang D. circLMTK2 acts as a sponge of miR-150-5p and promotes proliferation and metastasis in gastric cancer. *Mol Cancer* 2019;**18**:162
22. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and Meta-analysis. *Bioinformatics* 2012;**28**:573–80
23. Fromm B, Billipp T, Peck LE, Johansen M, Tarver JE, King BL, Newcomb JM, Sempere LF, Flatmark K, Hovig E, Peterson KJ. A uniform system for the annotation of vertebrate microRNA genes and the evolution of the human microRNAome. *Annu Rev Genet* 2015;**49**:213–42
24. Huang HY, Lin YC, Li J, Huang KY, Shrestha S, Hong HC, Tang Y, Chen YG, Jin CN, Yu Y, Xu JT, Li YM, Cai XX, Zhou ZY, Chen XH, Pei YY, Hu L, Su JJ, Cui SD, Wang F, Xie YY, Ding SY, Luo MF, Chou CH, Chang NW, Chen KW, Cheng YH, Wan XH, Hsu WL, Lee TY, Wei FX, Huang HD. miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res* 2020;**48**:D148–54
25. Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res* 2020;**48**:D127–31
26. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pagès F, Trajanoski Z, Galon J. ClueGO: a cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 2009;**25**:1091–3
27. Bindea G, Galon J, Mlecnik B. CluePedia cytoscape plugin: pathway insights using integrated experimental and in silico data. *Bioinformatics* 2013;**29**:661–3
28. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant* 2020;**13**:1194–202
29. Zhu X, Tian X, Yu C, Shen C, Yan T, Hong J, Wang Z, Fang JY, Chen H. A long non-coding RNA signature to improve prognosis prediction of gastric cancer. *Mol Cancer* 2016;**15**:60
30. Alessandrini L, Manchi M, De Re V, Dolcetti R, Canzonieri V. Proposed molecular and miRNA classification of gastric cancer. *Int J Mol Sci* 2018;**19**:1683
31. Kristensen LS, Hansen TB, Venø MT, Kjems J. Circular RNAs in cancer: opportunities and challenges in the field. *Oncogene* 2018;**37**:555–65
32. Vo JN, Cieslik M, Zhang Y, Shukla S, Xiao L, Zhang Y, Wu YM, Dhanasekaran SM, Engelke CG, Cao X, Robinson DR, Nesvizhskii AI, Chinnaiyan AM. The landscape of circular RNA in cancer. *Cell* 2019;**176**:869–81.e13
33. Yang F, Hu A, Li D, Wang J, Guo Y, Liu Y, Li H, Chen Y, Wang X, Huang K, Zheng L, Tong Q. Circ-HuR suppresses HuR expression and gastric cancer progression by inhibiting CNBP transactivation. *Mol Cancer* 2019;**18**:158
34. Levine AJ. p53: 800 million years of evolution and 40 years of discovery. *Nat Rev Cancer* 2020;**20**:471–80
35. Joerger AC, Fersht AR. The p53 pathway: Origins, inactivation in cancer, and emerging therapeutic approaches. *Annu Rev Biochem* 2016;**85**:375–404
36. Sethi N, Kikuchi O, McFarland J, Zhang Y, Chung M, Kafker N, Islam M, Lampson B, Chakraborty A, Kaelin WG Jr, Bass AJ. Mutant p53 induces a hypoxia transcriptional program in gastric and esophageal adenocarcinoma. *JCI Insight* 2019;**4**:e128439
37. Harris SL, Levine AJ. The p53 pathway: positive and negative feedback loops. *Oncogene* 2005;**24**:2899–908

38. Fu H, Wang C, Yang D, Wei Z, Xu J, Hu Z, Zhang Y, Wang W, Yan R, Cai Q. Curcumin regulates proliferation, autophagy, and apoptosis in gastric cancer cells by affecting PI3K and P53 signaling. *J Cell Physiol* 2018;**233**:4634–42
39. Liu Y, Yin L, Chen C, Zhang X, Wang S. Long non-coding RNA GAS5 inhibits migration and invasion in gastric cancer via interacting with p53 protein. *Dig Liver Dis* 2020;**52**:331–8
40. Gelse K, Pöschl E, Aigner T. Collagens–structure, function, and biosynthesis. *Adv Drug Deliv Rev* 2003;**55**:1531–46
41. Labelle-Dumais C, Schuitema V, Hayashi G, Hoff K, Gong W, Dao DQ, Ullian EM, Oishi P, Margeta M, Gould DB. COL4A1 mutations cause neuromuscular disease with tissue-specific mechanistic heterogeneity. *Am J Hum Genet* 2019;**104**:847–60
42. Zagaglia S, Selch C, Nisevic JR, Mei D, Michalak Z, Hernandez-Hernandez L, Krithika S, Vezyroglou K, Varadkar SM, Pepler A, Biskup S, Leão M, Gärtner J, Merkenschlager A, Jaksch M, Møller RS, Gardella E, Kristiansen BS, Hansen LK, Vari MS, Helbig KL, Desai S, Smith-Hicks CL, Hino-Fukuyo N, Talvik T, Laugesaar R, Ilves P, Õunap K, Körber I, Hartlieb T, Kudernatsch M, Winkler P, Schimmel M, Hasse A, Knuf M, Heinemeyer J, Makowski C, Ghedia S, Subramanian GM, Striano P, Thomas RH, Micallef C, Thom M, Werring DJ, Kluger GJ, Cross JH, Guerrini R, Balestrini S, Sisodiya SM. Neurologic phenotypes associated with COL4A1/2 mutations: Expanding the spectrum of disease. *Neurology* 2018;**91**:e2078–88
43. Wang T, Jin H, Hu J, Li X, Ruan H, Xu H, Wei L, Dong W, Teng F, Gu J, Qin W, Luo X, Hao Y. COL4A1 promotes the growth and metastasis of hepatocellular carcinoma cells by activating FAK-Src signaling. *J Exp Clin Cancer Res* 2020;**39**:148
44. Omar R, Cooper A, Maranyane HM, Zerbini L, Prince S. COL1A2 is a TBX3 target that mediates its impact on fibrosarcoma and chondrosarcoma cell migration. *Cancer Lett* 2019;**459**:227–39
45. Tamilzhalagan S, Rathinam D, Ganesan K. Amplified 7q21-22 gene MCM7 and its intronic miR-25 suppress COL1A2 associated genes to sustain intestinal gastric cancer features. *Mol Carcinog* 2017;**56**:1590–602
46. Ao R, Guan L, Wang Y, Wang JN. Silencing of COL1A2, COL6A3, and THBS2 inhibits gastric cancer cell proliferation, migration, and invasion while promoting apoptosis through the PI3k-Akt signaling pathway. *J Cell Biochem* 2018;**119**:4420–34

(Received June 11, 2021, Accepted September 7, 2021)