Research Paper

Identification of a competing endogenous RNA axis related to gastric cancer

Fuqiang Zu[1], Hezhou Han[2], Weiwei Sheng[3], Jian Sun[3], Hui Zang[1], Yu Liang[1], Qingfeng Liu[1]

[1]Department of General Surgery, The People's Hospital of China Medical University, Shenyang 110016, Liaoning, China
[2]Central Laboratory of General Surgery, Shengjing Hospital of China Medical University, Shenyang 110004, Liaoning, China
[3]Department of Gastrointestinal and Hernia Surgery, First Affiliated Hospital of China Medical University, Shenyang 110001, Liaoning, China

Correspondence to: Qingfeng Liu; **email:** qfliu@cmu.edu.cn, https://orcid.org/0000-0002-6693-2869

ABSTRACT

Competing endogenous RNA (ceRNA) pathways play pivotal roles in the formation and progression of gastric cancer (GC). Employing multi-omics analysis, we sought to identify a ceRNA network associated with GC progression. We analyzed 3 Gene Expression Omnibus datasets as well as data from The Cancer Genome Atlas to identify genes that were differentially expressed in GC tissues. A total of 84 upregulated genes and 106 downregulated genes were found. Enrichment analysis indicated that some pathways were strongly linked with tumor formation and progression. We also screened hub genes to establish a lncRNA-miRNA-mRNA network. We ultimately identified 8 hub genes, 6 key miRNAs and 4 key lncRNAs that interact within a common ceRNA network. Correlation analysis and in vitro experiments were conducted to verify the regulatory effect of the ceRNA network in GC. A knockdown assay confirmed that the *DLGAP1-AS1/miR-203a-3p/THBS2* axis is a ceRNA network involved in GC progression. In this study, we elucidated the role of the *DLGAP1-AS1/miR-203a-3p/THBS2* ceRNA network in the progression of GC. These molecules maybe evaluated as therapeutic targets and prognostic biomarkers for GC.

INTRODUCTION

Although there has been a decline in the incidence rate, gastric cancer remains the fourth most common malignancy and the second most deadly neoplasm globally. Over a million newly diagnosed cases annually, resulting in 783,000 deaths in 2018 [1]. Owing to a lack of specific symptoms in the early stage, the majority of patients with GC are diagnosed at advanced stages, which reduces the chances for a radical resection [2].

Non-coding RNA (ncRNA) does not directly code functional protein, but it is abundant in the human genome. Accumulating evidence has indicated that

ncRNAs, including microRNA (miRNA), long noncoding RNA (lncRNA) and circular RNA (circRNA), play an important role in oncogenesis and tumor progression of various cancers such as GC [3–5]. ncRNA can combine with protein-coding mRNAs and regulate gene expression at the transcriptional and post-transcriptional levels [6]. Moreover, multiple RNAs can interact with each other via miRNA response elements and assemble as a competing endogenous RNA (ceRNA) network [7]. In this network, lncRNA can act as a "sponge" to absorb and bind miRNA, thereby weakening its binding ability to mRNA. Emerging data have suggested that ceRNA networks play a pivotal role in progression and metastasis in breast cancer, ovarian cancer and GC [8–10].

In our work, we determined which genes are differentially expressed (DEGs) in GC tissues compared to normal tissues. Datasets for analysis were obtained from the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases. Protein-protein interaction (PPI) networks were constructed by a string database and we distinguished the top 10 hub genes according to their degrees score. Hub genes refer to genes with high connectivity. These genes play a crucial role in the biological function of another gene that is a functional target. According to the ceRNA hypothesis, lncRNA can diminish miRNA activity via adsorptive action. Therefore, a qualified candidate lncRNA should be negatively linked with miRNA expression and positively correlated with the mRNA level at the same time [7, 11]. Following this hypothesis, we predicted gene-related upstream miRNA via the miRecords database, and miRNA-linked upstream lncRNA through the miRNet dataset. The prognostic properties of the chosen RNAs were assessed both in bioinformatics databases and using qRT-PCR methods. Ultimately, a promising ceRNA regulatory network related to the progression of GC was successfully identified.

RESULTS

DEG identification

Three GEO (GSE54129, GSE29272, GSE13911) and one TCGA dataset were enrolled in a training group comprising663 cancer samples and 223 normal cases. Dataset GSE54129 contained111 GC samples and 21 non-tumor samples; dataset GSE29272 contained 134 GC samples and 134 non-tumor samples; dataset GSE13911 contained 38 GC tissues and 31 non-tumor samples; and the TGCA dataset contained 380 GC tissues and 37 non-tumor samples. DEGs in the training group were identified and are displayed in the volcano plot in Figure 1 with thresholds of |log2FC| >1 and a P value < 0.05 (Figure 1A–1D). As depicted in Venn plots (Figure 1E, 1F), we integrated commonly expressed genes that intersected in the training group, and we successfully identified 84 upregulated and 106 downregulated DEGs. The details of these DEGs, which were chosen for the following analysis, can be found in Supplementary Table 1 (Supplementary Table 1).

Enrichment analysis of DEGs

To detect potential biological functions of upregulated and downregulated DEGs, gene ontology (GO) enrichment analysis and KEGG pathway analysis were conducted. As depicted in Figure 2A, 2B, upregulated genes were mainly enriched in the extracellular matrix organization and cell adhesion process (Figure 2A).

Additionally, several cancer-related pathways were detected in the KEGG pathway analysis, such as the *PI3K-AKT* signaling pathway and cytokine receptor interaction pathway (Figure 2B). Downregulated genes were mainly enriched in the oxidation-reduction process and xenobiotic metabolic process (Figure 2C). Genes associated mainly with the KEGG pathway were those involved in cytochrome P450 metabolism and other metabolic pathways (Figure 2D). GO enrichment analysis of the molecular function and cellular component of the DEGs are depicted in Supplementary Figure 1. In generally, the results from functional enrichment analysis were tightly linked with GC.

Screening and validation of hub genes

To understand the interaction between the upregulated DEG group and the downregulated DEG group separately, a PPI network was constructed. We calculated the node degree of the PPI network using cytoHubba tools from Cytoscape software and classified the top 20 hub genes into two groups (Figure 2E, 2F). Next, to improve our result reliability, we verified the top 10 hub genes in the validation group. The GSE27342 dataset contained 80 GC samples and 80 normal samples; theGSE37023 dataset contained112 tumor samples and 39 normal samples; and theGSE65801 dataset contained 32 cancer samples and 32 normal samples. All the hub genes were involved in the validation group (Supplementary Figure 2A, 2B). Subsequently, GEPIA and Kaplan Meier (KM) plot database was performed to assess each hub gene's expression and its relationship to prognosis. In the upregulated hub gene group, 6 genes (*COL1A1, COL1A2, TIMP1, SPP1*, *BGN, and THBS2*) were not only significantly upregulated in GC but also strongly correlated with poor GC prognosis (Figure 3A, 3C and 3E). As for downregulated hub genes, two genes (*SST, TFF2*) were expressed at a low level in GC and were correlated with a favorable GC prognosis (Figure 3B, 3D and 3F). Other candidate hub genes, which were chosen for the following analysis, are shown in Supplementary Figure 2C–2F).

## Identification and validation of upstream miRNA

Based on the results from candidate hub genes, we identified the upstream miRNA of those genes through the miRNA-target interactions database, miRecords. Only miRNAs that appeared at least 3 times were considered as candidate miRNAs. A total of 136 miRNAs were predicted to regulate 6 hub genes. Only 13 miRNAs that were identified as candidate miRNAs (Supplementary Table 2). Additionally, the starBase platform was applied to validate the expression role of candidate miRNAs, and the KM plot database was used

to verify the prognostic value of candidate miRNAs. As shown in Supplementary Figures 3 and 4, we confirmed that 6 miRNAs are related to upregulated hub genes (*hsa-miR-203a-3p, hsa-miR-204-5p, hsa-miR-26a-5p, hsa-miR-339-5p, hsa-miR-1225-3p* and *hsa-miR-378a*). Not only was the expression of these genes downregulated in GC, but they were also linked with poor prognosis in the disease. However, for downregulated hub genes, we found only one miRNA (*hsa-mir-9-5p*) that was upregulated in GC and associated with a favorable prognosis. All qualified miRNAs were selected for further tests.

Prediction and validation of upstream lncRNA

To predict upstream lncRNA of candidate miRNAs, we used the miRNet database. According to the prediction in Supplementary Table 3, we identified 139 lncRNAs
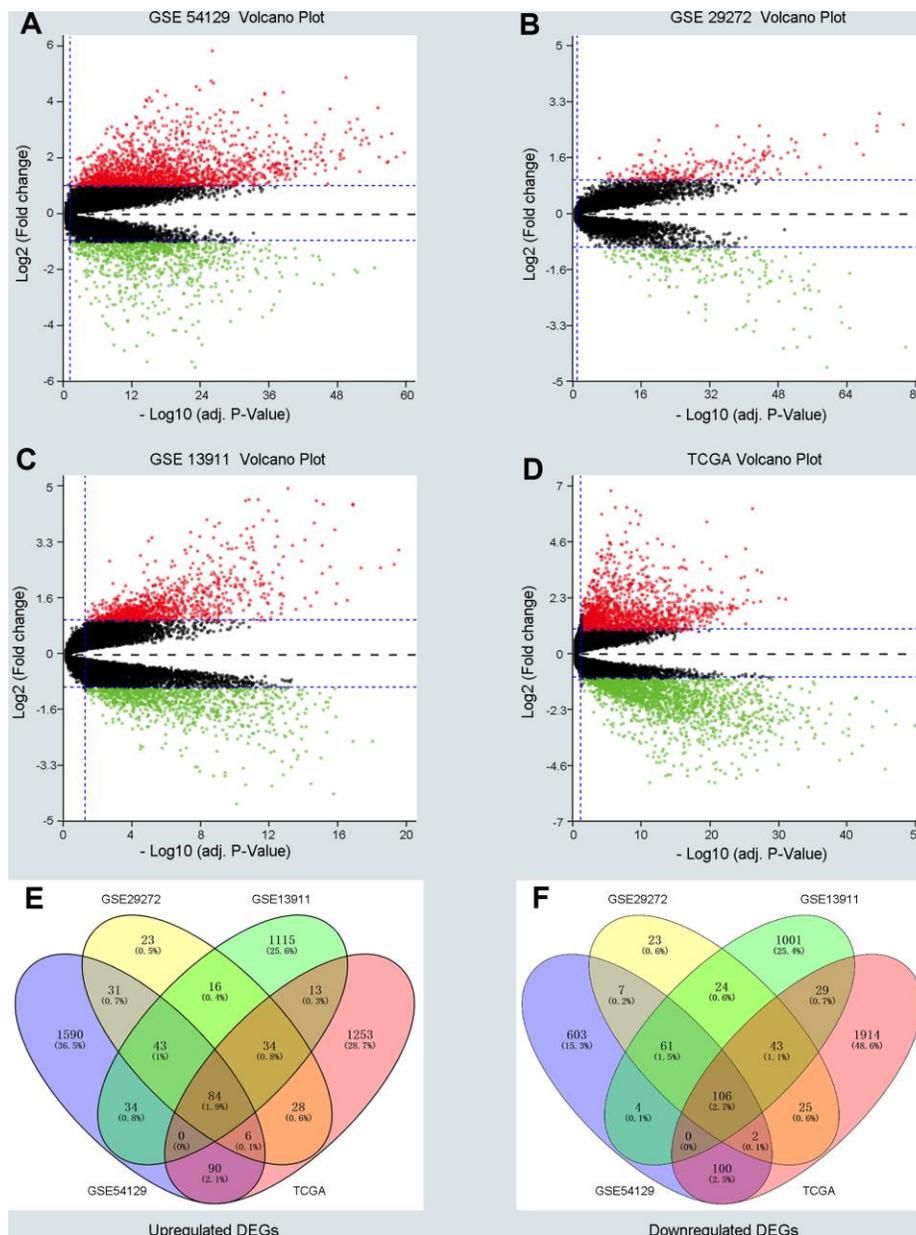
Figure 1. Screening differentially expressed genes (DEGs) between gastric cancer (GC) and normal samples in three GEO datasets and TCGA database. (**A–D**) The volcano plots of DEGs in GSE54129, GSE29272, GSE13911 and TCGA datasets with thresholds of |log2FC| > 1, adjust P value < 0.05. The red dots and green dots represent the upregulated and downregulated DEGs separately. The black dots mean no significantly different genes. (**E**, **F**) The intersection of upregulated DEGs and downregulated DEGs in four datasets, respectively.

for 6 downregulated miRNAs and 36 lncRNAs for *miR-9-5p*. Next, GEPIA and the KM plotter database were used to evaluate the expression role and prognostic value of predicted lncRNAs. According to the ceRNA hypothesis motioned above, we screened out 4 eligible lncRNAs (*DLGAP1-AS1, PVT1, RECQL4 and HCG18*) associated with downregulated miRNAs that were both significantly upregulated in GC and correlated with very poor survival (Figure 4A–4D). However, there was no qualified lncRNA that met the expression and prognostic criteria of miR-9-5p. Finally, we identified the eligible lncRNAs, miRNAs and mRNAs that not only satisfied the standards for expression and prognosis but also complied with the ceRNA network hypothesis.
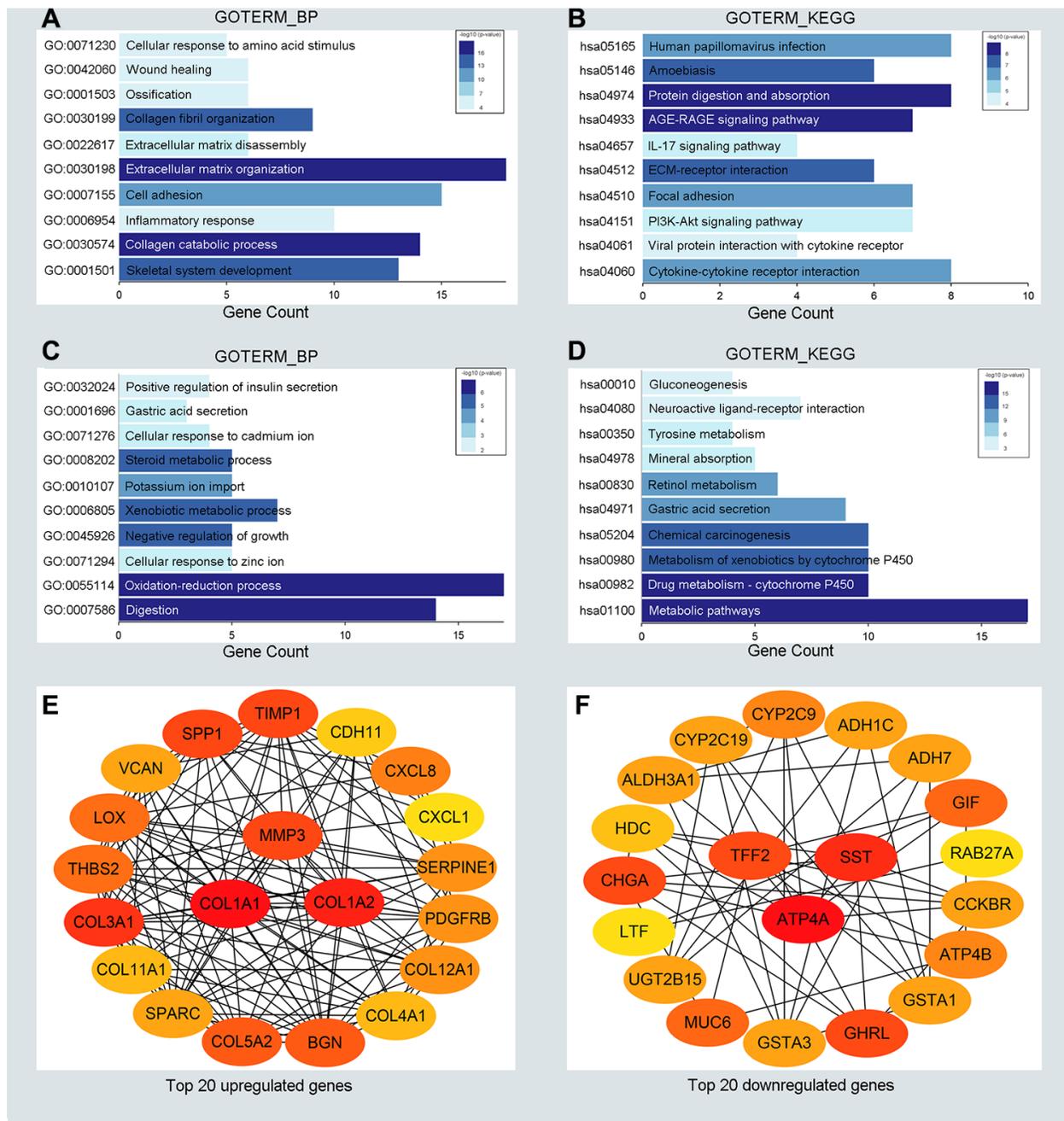

Figure 2. Functional enrichment analysis for the significant DEGs and identification of hub genes. (**A**, **C**) The top ten enriched biological processes of the significantly upregulated DEGs and downregulated DEGs, respectively. (**B**, **D**) The top ten enriched KEGG pathways of the significantly upregulated DEGs and downregulated DEGs, respectively. (**E**, **F**) The top 20 hub genes of the significantly upregulated DEGs and downregulated DEGs separately.

Construction and verification of the ceRNA network

We constructed a lncRNA-miRNA-mRNA ceRNA network from our previous results. As presented in Figure 4E, there were a total of 8lncRNA-mRNA groups, 5lncRNA-miRNA groups and 5miRNA-mRNA groups. Eligible miRNA has an opposite interaction with mRNA and lncRNA, whereas lncRNA has a positive co-expression relationship with mRNA. We used the starBase platform performed to characterize the co-expression relationships among lncRNA-miRNA and miRNA-mRNA. Ultimately, we established the

DLGAP1-AS1/miR-203a-3p/THBS2 ceRNA pathway, which was not only significantly associated with the prognosis of GC patients but also played pivotal roles in the progression of GC (Supplementary Figure 5A–5C). Other co-expression networks are depicted in Supplementary Figure 5D–5Q.

To further evaluate the reliability of our result, we verified the ceRNA network using an in vitro assay. First, the relative expression level of *DLGAP1-AS1* was quantified in 4 GC cell lines (MKN-45, AGS, HGC-27 and MGC-803), as well as and normal gastric cells
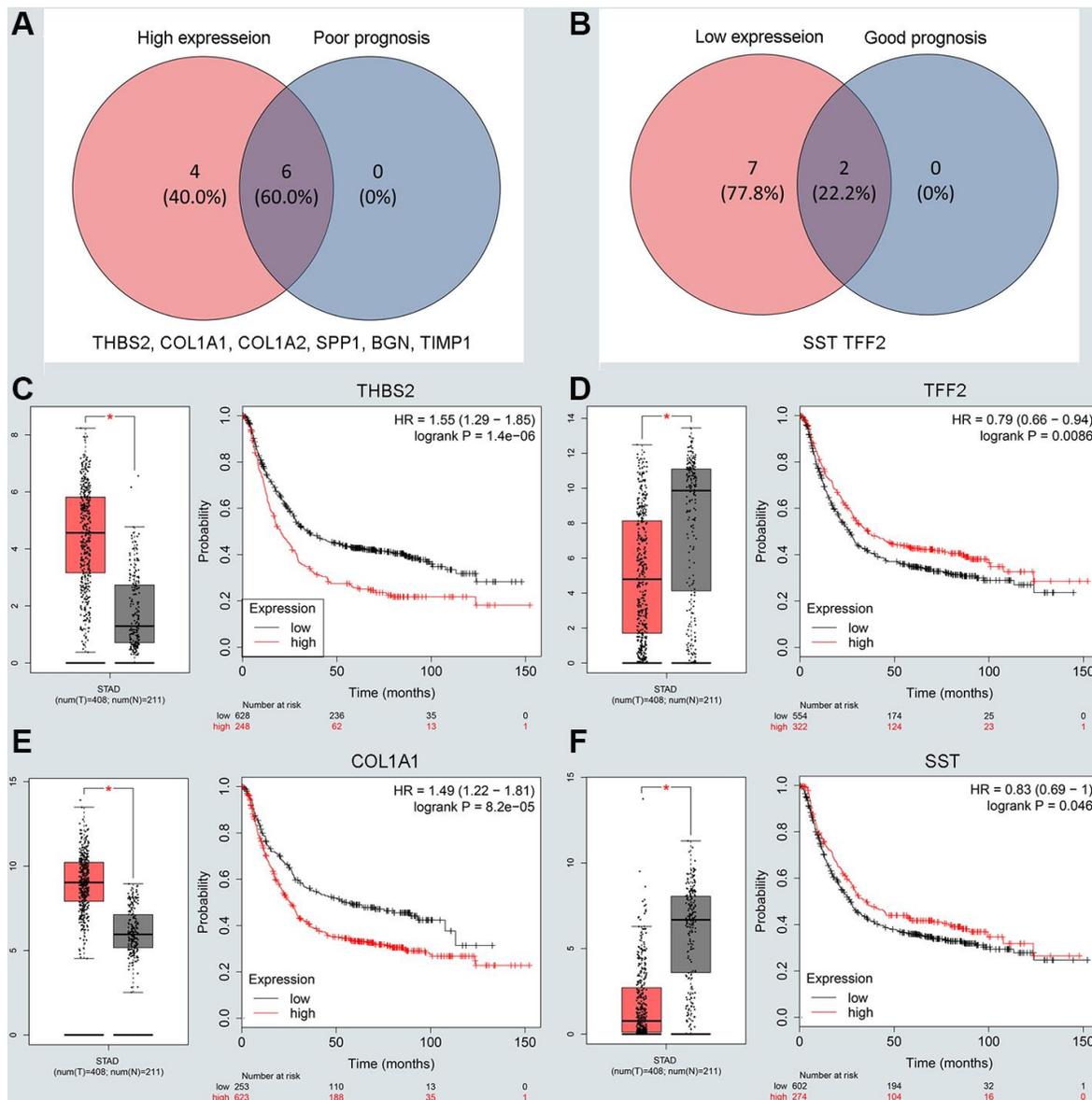
Figure 3. Screening and validating the expression roles and prognosis values of key genes in GC. (**A**) Screening the key genes with high expression and dismal prognosis values in upregulated hub genes. (**B**) Screening the key genes with low expression and good prognosis values in downregulated hub genes. (**C**–**F**) Validating expression roles and prognosis values of key genes in hub genes using GEPIA and Kaplan–Meier plotter databases.

(GES-1). Our results indicated that MKN-45 and AGS express higher levels of *DLGAP1-AS1* than other cell lines (Supplementary Figure 6A). Therefore, we designed siRNA assay to knockdown the expression of *DLGAP1-AS1* in MKN-45 and AGS cell lines separately and observed the change in relative expression of *DLGAP1-AS1, miR-203a-3p, and THBS2*. The knockdown efficiency was measured by qRT-PCR (Supplementary Figure 6B, 6C). As presented in Figure 5A–5F, the expression levels of *DLGAP1-AS1 and THBS2* were significantly reduced, while the relative expression of *miR-203a-3p* was enhanced after

DLGAP1-AS1 knockdown both in MKN-45 and AGS cell lines. A CCK-8 assay revealed that silencing *DLGAP1-AS1* could significantly inhibit gastric cancer cell proliferation (Figure 5G, 5H). And the raw data of our results are shown in Supplementary Table 4. Furthermore, a migration and invasion assay indicated that *DLGAP1-AS1* knockdown also plays an inhibitor role in GC migration and invasion (Supplementary Figure 7A–7D). The *DLGAP1-AS1/miR-203a-3p/THBS2* ceRNA pathway and its potential roles in the progression of GC is illustrated in schematic representations in Figure 6. Altogether, we have described a ceRNA
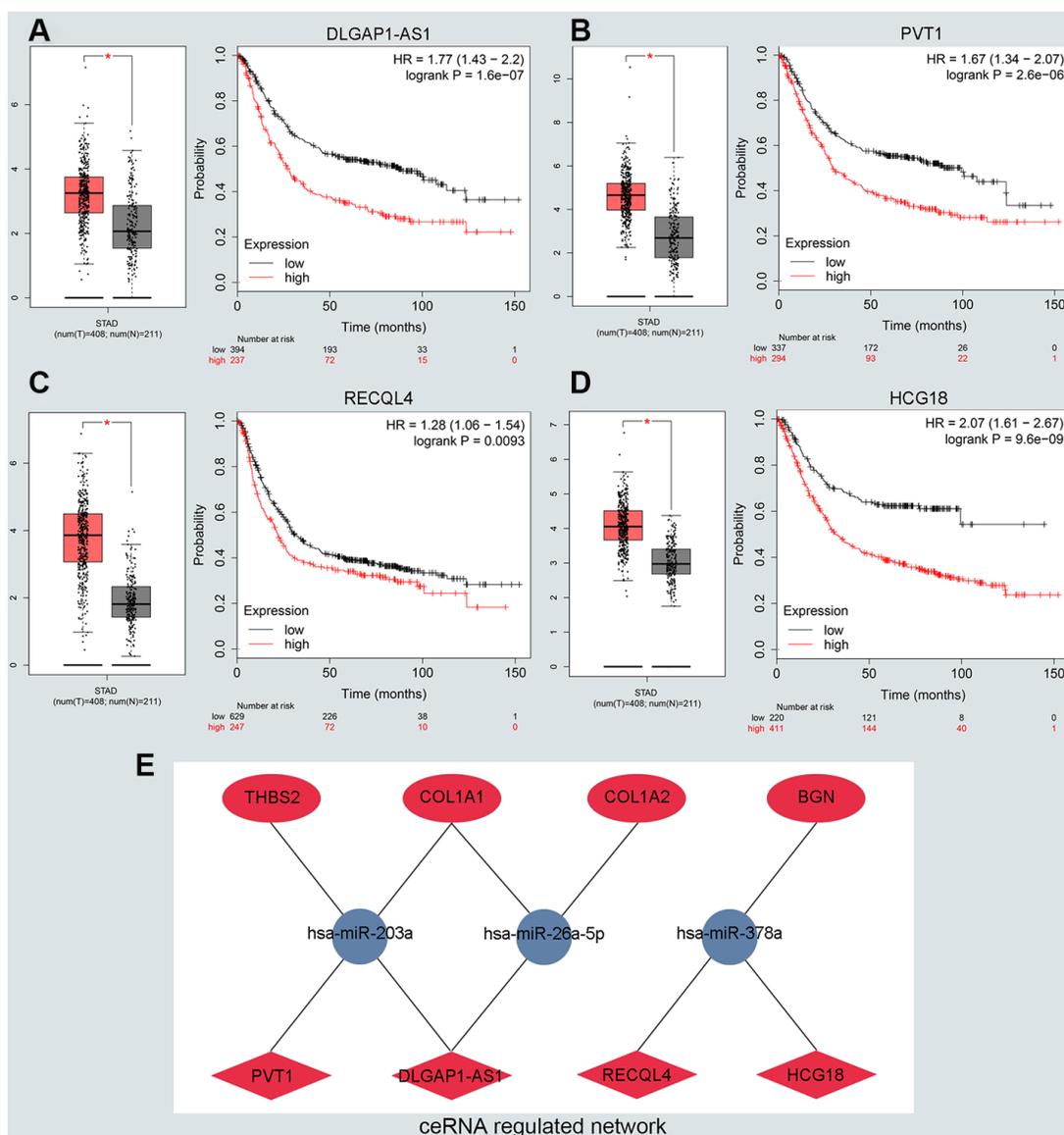

Figure 4. Identifying the key long noncoding RNA and constructing the ceRNA network in GC. (**A**–**D**) Validating the expression and prognostic value of four key lncRNAs using GEPIA and Kaplan–Meier plotter databases. (**E**) The potential mRNA-miRNA-lncRNA regulatory network related to GC prognosis. The ellipse, round and diamond shape represents lncRNAs, miRNA and mRNA respectively. Red and blue represent the ups and downs expression, respectively.

network that could shed light on the oncogenesis of GC and may contain future diagnostic markers and therapeutic targets.

DISCUSSION

Some evidence has indicated that lncRNAs play regulatory roles in the oncogenesis and tumor progression of various cancers [12]. Recent research has suggested that lncRNAs interact with miRNA and regulate downstream mRNA expression [7]. In GC, lncRNAs, miRNA and mRNA function as a unit, not simply through one-to-one interactions. For instance, Huang et al. reported that lncRNA *IGF2-AS* was involved in encouraging tumor growth and invasion in GC through the *IGF2-AS/miR-503/SHOX*2 ceRNA
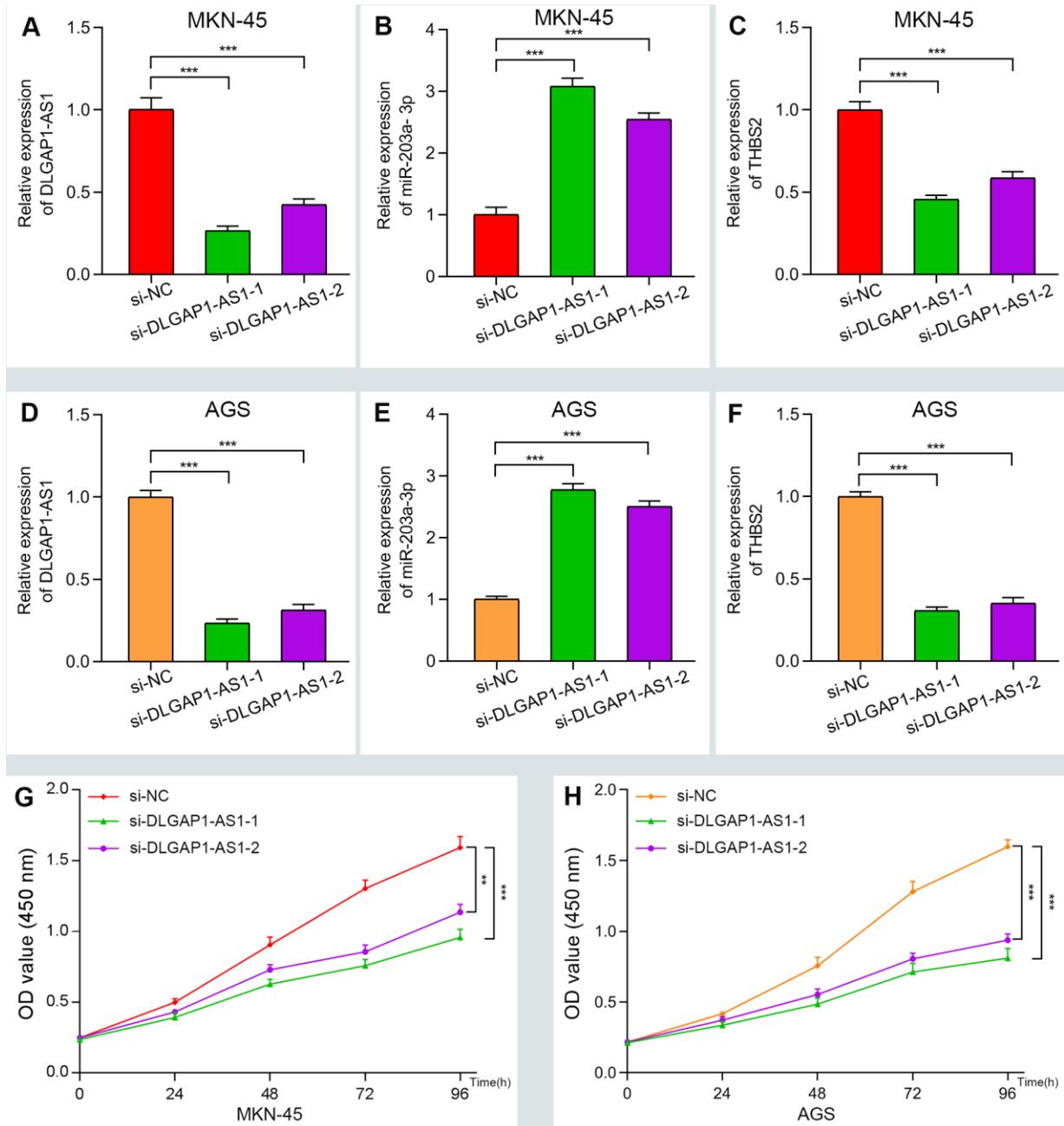
Figure 5. Verifying the ceRNA network through knockdown and CCK-8 assay. (**A**–**F**) The relative expression of DLGAP1-AS1 and THBS2 were significantly reduced after silencing DLGAP1-AS1, whereas the relative expression of miR-203a-3p was significantly increased in MKN-45 and AGC cell lines. (**G**–**H**) DLGAP1-AS1 knockdown efficiently suppressed MKN-45 and AGC cell proliferation, respectively. (*P < 0.05, **P < 0.01, ***P < 0.001)

network [13]. Xiao et al. indicated that *TRPM2-AS* functioned in a ceRNA network and negatively regulated *miR-612* expression, thereby leading to GC progression and radio resistance by disrupting the expression of IGF2BP1 and FOXM1 [14]. Wei et al. demonstrated that the *CTC-497E21.4/miR-22-3p/NET1* ceRNA network played positive roles in GC progression via the RhoA signaling pathway [15].

In the present study, we successfully identified a promising *DLGAP1-AS1/miR-203a-3p/THBS2*ce RNA network involved in GC progression through integrated bioinformatic analysis and verification assays. First, we detected 84 upregulated DEGs and 106 downregulated DEGs that were commonly expressed in the training group. GO term and KEGG pathway functional analysis revealed that those DEGs were significantly enriched in the cancer-related process. Then, those DEGs were visualized in the PPI network and selected as hub genes according to their node degrees, which were calculated by the cytoHubba tool. The expression role and survival value of the top 10 hub genes were validated using GEPIA and KM plotter databases. Eight qualified genes met the criteria of expression validation and survival analyses. Notably, their oncogenic roles were also detected in GC progression. For example, the high expression of *THBS2* and *COL1A2* facilitated GC proliferation and invasion but inhibited tumor apoptosis through the PI3K-Akt signaling pathway [16]. Overexpression of *COL1A1* was increased cancer invasion and was

directly regulated by *let-7i* miRNA [17]. BGN induced phosphorylation of FAK and Paxillin in GC metastasis, thereby activating the FAK signaling pathway [18].

Additionally, the upstream miRNAs of hub genes were predicted and validated by the relational databases mentioned above. Seven qualified miRNAs were selected as key miRNAs. Some of the miRNAs were involved in the development of GC. For instance, Yang et al. demonstrate that *miR-203a* served as a tumor suppressor and was able to inhibit the proliferation of GC cells by direct bonding with *E2F3* [19]. Li et al. showed that *miR-26a-5p* could impede proliferation and the invasion of GC cells by targeting *COL10A1* [20] Zhang et al. demonstrated that *miR-204a-5p* functioned as an oncogene by targeting *USP47* and *RAB22A* in stomach cancer [21].

Next, we identified the upstream lncRNAs of key miRNAs. Four lncRNAs conformed to expression and prognostic standards. *DLGAP1-AS1* facilitated HCC proliferation and progression via the *miR-486-5p/H3F3B* axis [22]. The elevated expression of *PVT1*has been significantly correlated with poor prognosis in multiple types of cancer [23, 24], including GC [25]. Some interactions within the ceRNA network have been previously identified. These include the *DLGAP-1-AS1/miR-26a-5p* axis in GC and the *PVT1/miR-203a* axis in multiple myeloma [26, 27]. Finally, we established a novel *DLGAP1-AS1/*
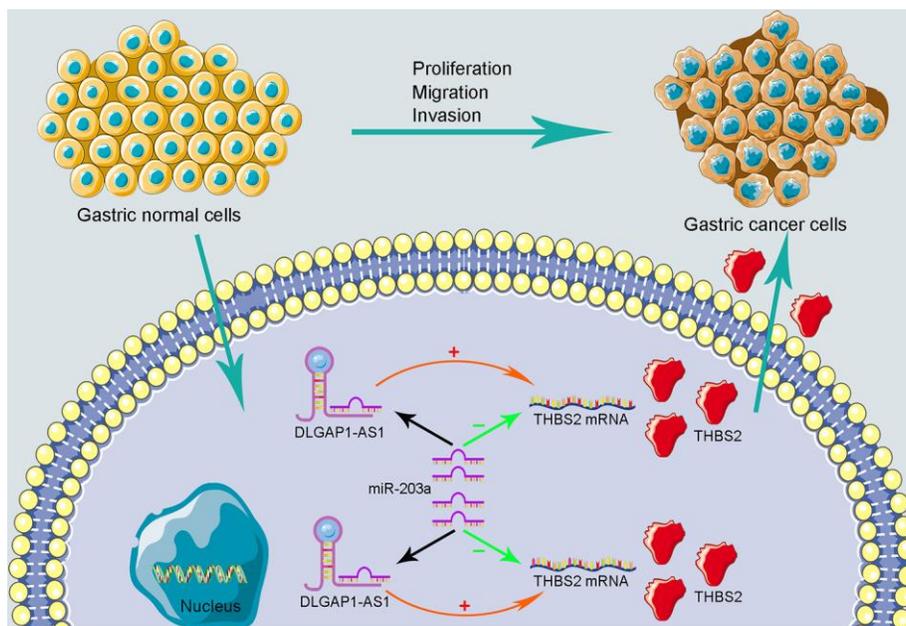

Figure 6. Schematic representations of DLGAP1-AS1/miR-203a-3p/THBS2 ceRNA pathway and its potential roles in the progression of GC.

miRNA-203a-3p/THBS2 ceRNA network that satisfies the conditions of the ceRNA hypothesis.

Bioinformatic analysis had previously been performed on a ceRNA network linked with GC [28, 29]. However, few studies concentrated on the prognostic value of the ceRNA axis using multi-omics analysis combined with experimental tests. Furthermore, to our knowledge, this is the first study of GC that constructed a ceRNA network in the order of the mRNA-miRNA-lncRNA pattern. There are some limitations to our study. First, we did not stratify the samples based on their clinical characteristics, such as sex. A recent study has indicated that incorporating sex as a biological variable is reveals new information about cancer mechanisms [30]. Second, we verified the expression and prognostic value using online databases rather than the date from clinical samples. To offset this limitation, we constructed the ceRNA axis through comprehensive analysis and validated the data under the same conditions.

In summary, by multi-omics analysis and experimental verification, we successfully constructed a *DLGAP1-AS1/miR-203a-3p/THBS2* ceRNA regulatory network in which all RNAs are correlated with the prognosis of GC patients. In addition to the prognostic value of this network, it also provides some key clues for future molecular mechanism explorations.

MATERIALS AND METHODS

Data selection

To identify compressive gene expression patterns in GC samples versus normal samples, we obtained the mRNA microarray profiles from GEO (www.ncbi.nlm.nih.gov/geo/) datasets. Only datasets consisting of at least 20 samples of both GC and normal tissues were collected. Eventually, 6 GEO datasets were selected for subsequent analyses. To increase the reliability of our research, GSE54129, GSE29272 and GSE13911 were chosen as a training group, whereas GSE27342, GSE37023 and GSE65801 were selected as a validation group. Furthermore, the TCGA database was integrated into the training group to enhance the reliability of our results.

Identification of DEGs and functional annotation analysis

Raw RNA-seq data were downloaded from the UCSC TCGA website. These datasets comprised 380 GC samples and 37 normal samples (https://xena.ucsc.edu/public/) [31]. The raw data were annotated by relational platforms and standardized by the method of $\log_2(x+1)$. Datasets were normalized using the "normalize between array" function of the LIMMA package from R

Software (version 3.6.1) [32]. The package was used to screen differential mRNAs with thresholds of $|\log2FC| > 1$ and a P value < 0.05. The commonly expressed mRNAs in the training group were defined as the DEGs and divided into upregulated DEGs and downregulated DEGs. These data were visualized through Venn diagrams using VENNY 2.1.0 (https://bioinfogp.cnb.csic.es/tools/venny/index.html) [33].

To elucidate the potential functions of the DEGs, we performed the GO functional enrichment analysis and KEGG pathway analysis via DAVIDv6.8 (https://david.ncifcrf.gov/) [34] and KOBAS 3.0 (http://kobas.cbi.pku.edu.cn/) software [35]. The top 10 enriched GO terms and KEGG pathways were visualized by the "ggplot2" package with cut-off criteria of $P < 0.05$ [36].

Hub genes identification and validation

The STRING v11.0 database (https://string-db.org/) [37] revealed a PPI network for DEGs with a combined confidence score ≥0.4. Next, we utilized cytoHubba, an app in Cytoscape software (Version 3.7.2) [38], to identify the top 20 hub genes according to their connection degree within the PPI network. We validated the expression levels of hub genes using the GEPIA database to analyze 408 GC samples and 211 normal controls from TCGA and Genotype-Tissue Expression GTEx data (http://gepia.cancer-pku.cn/index.html) [39]. The threshold value was set as $|\log FC| > 1$ and $P < 0.01$. Ultimately, the prognostic roles of key genes expressed in 436 GC samples were evaluated through the KM plotter (http://kmplot.com) [40]. The hazard ratio with a 95% confidence interval and log-rank P value was generated online. $P < 0.05$ was viewed as a statistically significant.

Prediction of upstream miRNA and lncRNA

To obtain comprehensive and reliable prediction results, we used the miRecords database that integrates the 11miRNA target prediction tools to predict the upstream miRNAs of hub genes [41]. The tools were DIANA-micro T, MicroInspector, miRanda, MirTarget2, miTarget, NBmiRTar, PicTar, PITA, RNA22, RNAhybrid and TargetScan. Only miRNAs that appeared at least 3 times were considered as validation miRNAs. The miRNet, was employed to find potential lncRNAs that would bind to validation miRNAs (https://www.mirnet.ca/miRNet/) [42]. GEPIA and KM plot databases verified the effectiveness of potential lncRNAs.

Correlation analysis and experimental verification

We used the starBase platform, a tool that evaluates potential RNA-RNA interactions [43], to validate

predicted miRNA and lncRNA. We also assessed the entire interrelation among the lncRNA, miRNA and genes via the starBase platform.

We validated the regulatory role of the ceRNA network in vitro assays. Normal gastric cells (GES-1) and Human GC cell lines (MKN-45, AGS, HGC-27 and MGC-803) were obtained from the cell bank of the Chinese Scientific Academy (Shanghai, China). All cells were incubated in RPMI-1640 (Gibco, Life Technologies, CA, USA), supplemented with 10% fetal bovine serum (Certified, US origin) and incubated at 37°C and 5% CO_2.

Three siRNAs (*si-DLGAP1-AS1*#1: 5'-GCU AUA UGU CUG GUA AAC AGA-3', *si-DLGAP1-AS1*#2: 5'-CAG AAU AAA UAG UAC UUG AGC-3' and *si-DLGAP1-AS1*#3: 5'-GCU GCU AUA UGU CUG GUA AAC-3') and negative control siRNA were purchased from Riobio Company (Guangzhou, China) and separately transfected into MKN-45 and AGS cells using Lipofectamine 3000 (Invitrogen, USA). Then, the qRT-PCR assay was performed to evaluate knockdown efficiency and detect the relative expression of *DLGAP1-AS1, miR-203a-3p* and *THBS2* through the SYBR-green method. Primer sequences were also synthesized by Riobio as shown as follows: *DLGAP1-AS1*:5'-GGG GCA GGA GTA AAG TGG AC-3' (forward), 5'-CCA GAC ATA TAG CAG CCG GG-3' (reverse); *miR-203a-3p*: 5'-CAC CAT AAA GAC AGG AAC CTG-3' (forward), 5'-GGA GGT GCC ATC AAT ACC TGC-3' (reverse); *THBS2*: 5'-TTA TGG CGT TGC ATC AGT-3' (forward), 5'-GTG GTG CAG AGG AGA TGT GT-3' (reverse); and *GAPDH*:5'-GAT TTG GTC GTA TTG GGC GC-3' (forward), 5'-GCG CCC AAT ACG ACC AAA TC-3' (reverse). The relative expression levels were quantified using the $2^{-\Delta\Delta Ct}$ method.

Cell viability detection was measured using the Cell Counting Kit-8 (CCK-8) assay according to instructions (Dojindo, Kumamoto, Japan). In brief, we incubated transfected cells in 96-well plates and then measured their viability with the CCK-8 kit at 0, 24, 48, 72 and 96 hours. The absorbance was detected at 450 nm with a microplate reader (BioTek, VT, USA). The result was repeated 3 times.

Wound healing, migration and invasion assays were performed. For the wound healing assay, the movement of cells was measured in a scrape, which was made with a sterile 200 μL pipette tip, and the spread of the wound closure was observed after 24 hours. The scratched areas at 0 and 24 hours were photographed at 100x magnification using an inverted microscope (Nikon DS-

Ri2, Japan). For migration assay, transwell chambers with 8-μm porous membranes (Corning, NY, USA) were used. 80,000 cells were seeded onto the upper chamber with 200 μL serum-free medium, and 800 μL medium containing 10% FBS was added to the bottom chambers as a chemo-attractant. Following 24 hours of incubation, cells that did not migrate to the bottom chambers were removed from the top side with a cotton swab. The invasion assay was performed in a similar procedure except the upper chambers were coated with Matrigel (BD Biosciences), which was diluted at 1:10 with serum-free medium. Cells traversing the membranes were fixed and stained with a Hematoxylin-Eosin Staining Kit (Solarbio, Beijing, China) according to the manufacturer's instruction. Cell images were captured using a microscope (Nikon E800) at 200x magnification and 5 random fields per insert were counted. Results were presented as cells migrated per field.

Statistical analysis

Most of the statistical analyses was conducted using the R Software and the other bioinformatic tools mentioned above. GraphPad Prism 7.0 (IBM, New York City, NY, USA) software was also utilized to analyze data. The two-tailed Student's *t*-test was applied to analyze the relative expression levels of mRNA, miRNA and lncRNA. Correlations between RNA expression were evaluated through Pearson correlation analysis. A *P* value<0.05 was considered statistically significant, which is presented in figures according to *P< 0.05, **P< 0.01 and ***P< 0.001.

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.
https://doi.org/10.3322/caac.21492 PMID:30207593

2. Kim SG, Park CM, Lee NR, Kim J, Lyu DH, Park SH, Choi IJ, Lee WS, Park SJ, Kim JJ, Kim JH, Lim CH, Cho JY, et al. Long-Term Clinical Outcomes of Endoscopic Submucosal Dissection in Patients with Early Gastric Cancer: A Prospective Multicenter Cohort Study. Gut Liver. 2018; 12:402–410.
https://doi.org/10.5009/gnl17414 PMID:29588436

3. Lee YS, Dutta A. MicroRNAs in cancer. Annu Rev Pathol. 2009; 4:199–227.
https://doi.org/10.1146/annurev.pathol.4.110807.092222 PMID:18817506

4. Huarte M. The emerging role of lncRNAs in cancer. Nat Med. 2015; 21:1253–61.
https://doi.org/10.1038/nm.3981 PMID:26540387

5. Wang Y, Mo Y, Gong Z, Yang X, Yang M, Zhang S, Xiong F, Xiang B, Zhou M, Liao Q, Zhang W, Li X, Li X, et al. Circular RNAs in human cancer. Mol Cancer. 2017; 16:25.
https://doi.org/10.1186/s12943-017-0598-7 PMID:28143578

6. Beermann J, Piccoli MT, Viereck J, Thum T. Non-coding RNAs in development and disease: background, mechanisms, and therapeutic approaches. Physiol Rev. 2016; 96:1297–325.
https://doi.org/10.1152/physrev.00041.2015 PMID:27535639

7. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the rosetta stone of a hidden RNA language? Cell. 2011; 146:353–58.
https://doi.org/10.1016/j.cell.2011.07.014 PMID:21802130

8. Zhong G, Lou W, Yao M, Du C, Wei H, Fu P. Identification of novel mRNA-miRNA-lncRNA competing endogenous RNA network associated with prognosis of breast cancer. Epigenomics. 2019; 11:1501–18.
https://doi.org/10.2217/epi-2019-0209 PMID:31502865

9. Lou W, Ding B, Zhong G, Du C, Fan W, Fu P. Dysregulation of pseudogene/lncRNA-hsa-miR-363-3p-SPOCK2 pathway fuels stage progression of ovarian cancer. Aging (Albany NY). 2019; 11:11416–39.
https://doi.org/10.18632/aging.102538 PMID:31794425

10. Qi M, Yu B, Yu H, Li F. Integrated analysis of a ceRNA network reveals potential prognostic lncRNAs in gastric cancer. Cancer Med. 2020; 9:1798–817.
https://doi.org/10.1002/cam4.2760 PMID:31923354

11. Thomson DW, Dinger ME. Endogenous microRNA sponges: evidence and controversy. Nat Rev Genet. 2016; 17:272–83.
https://doi.org/10.1038/nrg.2016.20 PMID:27040487

12. Guo LL, Song CH, Wang P, Dai LP, Zhang JY, Wang KJ. Competing endogenous RNA networks and gastric cancer. World J Gastroenterol. 2015; 21:11680–87.
https://doi.org/10.3748/wjg.v21.i41.11680 PMID:26556995

13. Huang J, Chen YX, Zhang B. IGF2-AS affects the prognosis and metastasis of gastric adenocarcinoma via acting as a ceRNA of miR-503 to regulate SHOX2. Gastric Cancer. 2020; 23:23–38.
https://doi.org/10.1007/s10120-019-00976-2 PMID:31183590

14. Xiao J, Lin L, Luo D, Shi L, Chen W, Fan H, Li Z, Ma X, Ni P, Yang L, Xu Z. Long noncoding RNA TRPM2-AS acts as a microRNA sponge of miR-612 to promote gastric cancer progression and radioresistance. Oncogenesis. 2020; 9:29.
https://doi.org/10.1038/s41389-020-0215-2 PMID:32123162

15. Zong W, Feng W, Jiang Y, Cao Y, Ke Y, Shi X, Ju S, Cong H, Wang X, Cui M, Jing R. LncRNA CTC-497E21.4 promotes the progression of gastric cancer via modulating miR-22/NET1 axis through RhoA signaling pathway. Gastric Cancer. 2020; 23:228–40.
https://doi.org/10.1007/s10120-019-00998-w PMID:31451992

16. 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.
https://doi.org/10.1002/jcb.26524 PMID:29143985

17. Shi Y, Duan Z, Zhang X, Zhang X, Wang G, Li F. Down-regulation of the let-7i facilitates gastric cancer invasion and metastasis by targeting COL1A1. Protein Cell. 2019; 10:143–48.
https://doi.org/10.1007/s13238-018-0550-7 PMID:29858755

18. Hu L, Duan YT, Li JF, Su LP, Yan M, Zhu ZG, Liu BY, Yang QM. Biglycan enhances gastric cancer invasion by activating FAK signaling pathway. Oncotarget. 2014; 5:1885–96.
https://doi.org/10.18632/oncotarget.1871 PMID:24681892

19. Yang H, Wang L, Tang X, Bai W. miR-203a suppresses cell proliferation by targeting E2F transcription factor 3 in human gastric cancer. Oncol Lett. 2017; 14:7687–90. https://doi.org/10.3892/ol.2017.7199 PMID:29344215

20. Li HH, Wang JD, Wang W, Wang HF, Lv JQ. Effect of miR-26a-5p on gastric cancer cell proliferation, migration and invasion by targeting COL10A1. Eur Rev Med Pharmacol Sci. 2020; 24:1186–94. https://doi.org/10.26355/eurrev_202002_20170 PMID:32096148

21. Zhang B, Yin Y, Hu Y, Zhang J, Bian Z, Song M, Hua D, Huang Z. MicroRNA-204-5p inhibits gastric cancer cell proliferation by downregulating USP47 and RAB22A. Med Oncol. 2015; 32:331. https://doi.org/10.1007/s12032-014-0331-y PMID:25429829

22. Peng X, Wei F, Hu X. Long noncoding RNA DLGAP1-AS1 promotes cell proliferation in hepatocellular carcinoma via sequestering miR-486-5p. J Cell Biochem. 2020; 121:1953–62. https://doi.org/10.1002/jcb.29430 PMID:31633236

23. Liu HT, Fang L, Cheng YX, Sun Q. LncRNA PVT1 regulates prostate cancer cell growth by inducing the methylation of miR-146a. Cancer Med. 2016; 5:3512–19. https://doi.org/10.1002/cam4.900 PMID:27794184

24. Wan L, Sun M, Liu GJ, Wei CC, Zhang EB, Kong R, Xu TP, Huang MD, Wang ZX. Long noncoding RNA PVT1 promotes non-small cell lung cancer cell proliferation through epigenetically regulating LATS2 expression. Mol Cancer Ther. 2016; 15:1082–94. https://doi.org/10.1158/1535-7163.MCT-15-0707 PMID:26908628

25. Kong R, Zhang EB, Yin DD, You LH, Xu TP, Chen WM, Xia R, Wan L, Sun M, Wang ZX, De W, Zhang ZH. Long noncoding RNA PVT1 indicates a poor prognosis of gastric cancer and promotes cell proliferation through epigenetically regulating p15 and p16. Mol Cancer. 2015; 14:82. https://doi.org/10.1186/s12943-015-0355-8 PMID:25890171

26. Lin Y, Jian Z, Jin H, Wei X, Zou X, Guan R, Huang J. Long non-coding RNA DLGAP1-AS1 facilitates tumorigenesis and epithelial-mesenchymal transition in hepatocellular carcinoma via the feedback loop of miR-26a/b-5p/IL-6/JAK2/STAT3 and Wnt/β-catenin pathway. Cell Death Dis. 2020; 11:34. https://doi.org/10.1038/s41419-019-2188-7 PMID:31949128

27. Yang M, Zhang L, Wang X, Zhou Y, Wu S. Down-regulation of miR-203a by lncRNA PVT1 in multiple myeloma promotes cell proliferation. Arch Med Sci. 2018; 14:1333–39. https://doi.org/10.5114/aoms.2018.73975 PMID:30393487

28. Pan H, Guo C, Pan J, Guo D, Song S, Zhou Y, Xu D. Construction of a competitive endogenous RNA network and identification of potential regulatory axis in gastric cancer. Front Oncol. 2019; 9:912. https://doi.org/10.3389/fonc.2019.00912 PMID:31637209

29. Zhang QN, Zhu HL, Xia MT, Liao J, Huang XT, Xiao JW, Yuan C. A panel of collagen genes are associated with prognosis of patients with gastric cancer and regulated by microRNA-29c-3p: an integrated bioinformatics analysis and experimental validation. Cancer Manag Res. 2019; 11:4757–72. https://doi.org/10.2147/CMAR.S198331 PMID:31213898

30. Wilson MA, Buetow KH. Novel mechanisms of cancer emerge when accounting for sex as a biological variable. Cancer Res. 2020; 80:27–29. https://doi.org/10.1158/0008-5472.CAN-19-2634 PMID:31722998

31. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome Res. 2002; 12:996–1006. https://doi.org/10.1101/gr.229102 PMID:12045153

32. Smyth GK, Michaud J, Scott HS. Use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics. 2005; 21:2067–75. https://doi.org/10.1093/bioinformatics/bti270 PMID:15657102

33. Oliveros JC. (2007-2015) Venny. An interactive tool for comparing lists with Venn's diagrams. https://bioinfogp.cnb.csic.es/tools/venny/index.html.

34. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009; 4:44–57. https://doi.org/10.1038/nprot.2008.211 PMID:19131956

35. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY, Wei L. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res. 2011; 39:W316–22. https://doi.org/10.1093/nar/gkr483 PMID:21715386

36. Maag JL. Gganatogram: an R package for modular visualisation of anatograms and tissues based on ggplot2. F1000Res. 2018; 7:1576. https://doi.org/10.12688/f1000research.16409.2 PMID:30467523

37. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S,

Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019; 47:D607–13.
https://doi.org/10.1093/nar/gky1131
PMID:30476243

38. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol. 2014 (Suppl 4); 8:S11.
https://doi.org/10.1186/1752-0509-8-S4-S11
PMID:25521941

39. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 2017; 45:W98–102.
https://doi.org/10.1093/nar/gkx247 PMID:28407145

40. Gyorffy B, Lánczky A, Szállási Z. Implementing an online tool for genome-wide validation of survival-associated biomarkers in ovarian-cancer using microarray data from 1287 patients. Endocr Relat Cancer. 2012; 19:197–208.
https://doi.org/10.1530/ERC-11-0329
PMID:22277193

41. Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T. miRecords: an integrated resource for microRNA-target interactions. Nucleic Acids Res. 2009; 37:D105–10.
https://doi.org/10.1093/nar/gkn851 PMID:18996891

42. Fan Y, Siklenka K, Arora SK, Ribeiro P, Kimmins S, Xia J. miRNet - dissecting miRNA-target interactions and functional associations through network-based visual analysis. Nucleic Acids Res. 2016; 44:W135–41.
https://doi.org/10.1093/nar/gkw288
PMID:27105848

43. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-seq data. Nucleic Acids Res. 2014; 42:D92–97.
https://doi.org/10.1093/nar/gkt1248
PMID:24297251
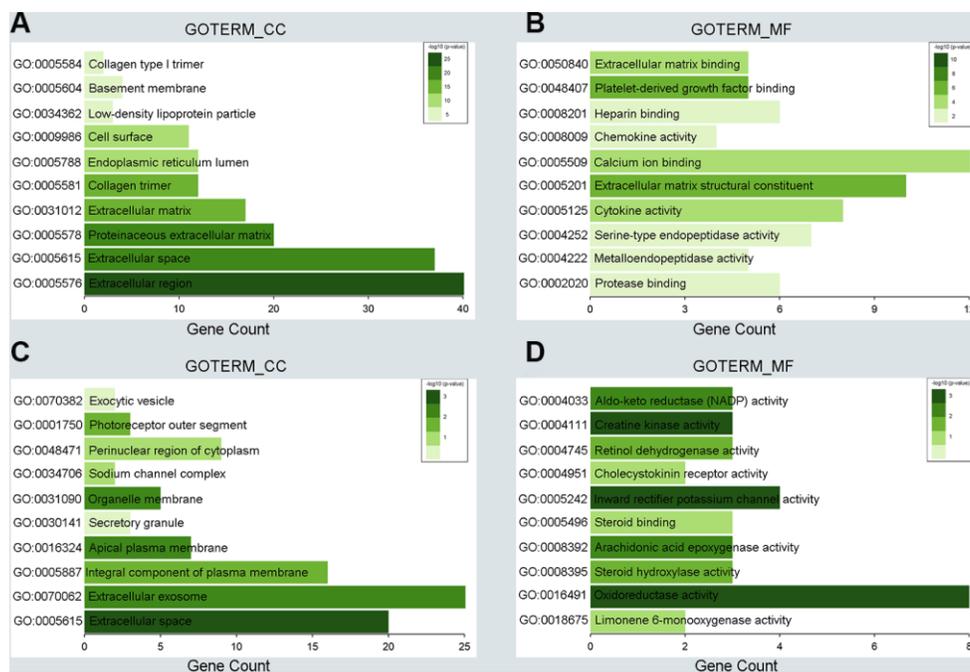
Supplementary Figure 1. Functional enrichment analysis for the significant DEGs. (**A**, **C**) The top ten enriched cellular components of the significantly upregulated DEGs and downregulated DEGs respectively. (**B**, **D**) The top ten enriched molecular function of the significantly upregulated DEGs and downregulated DEGs respectively.
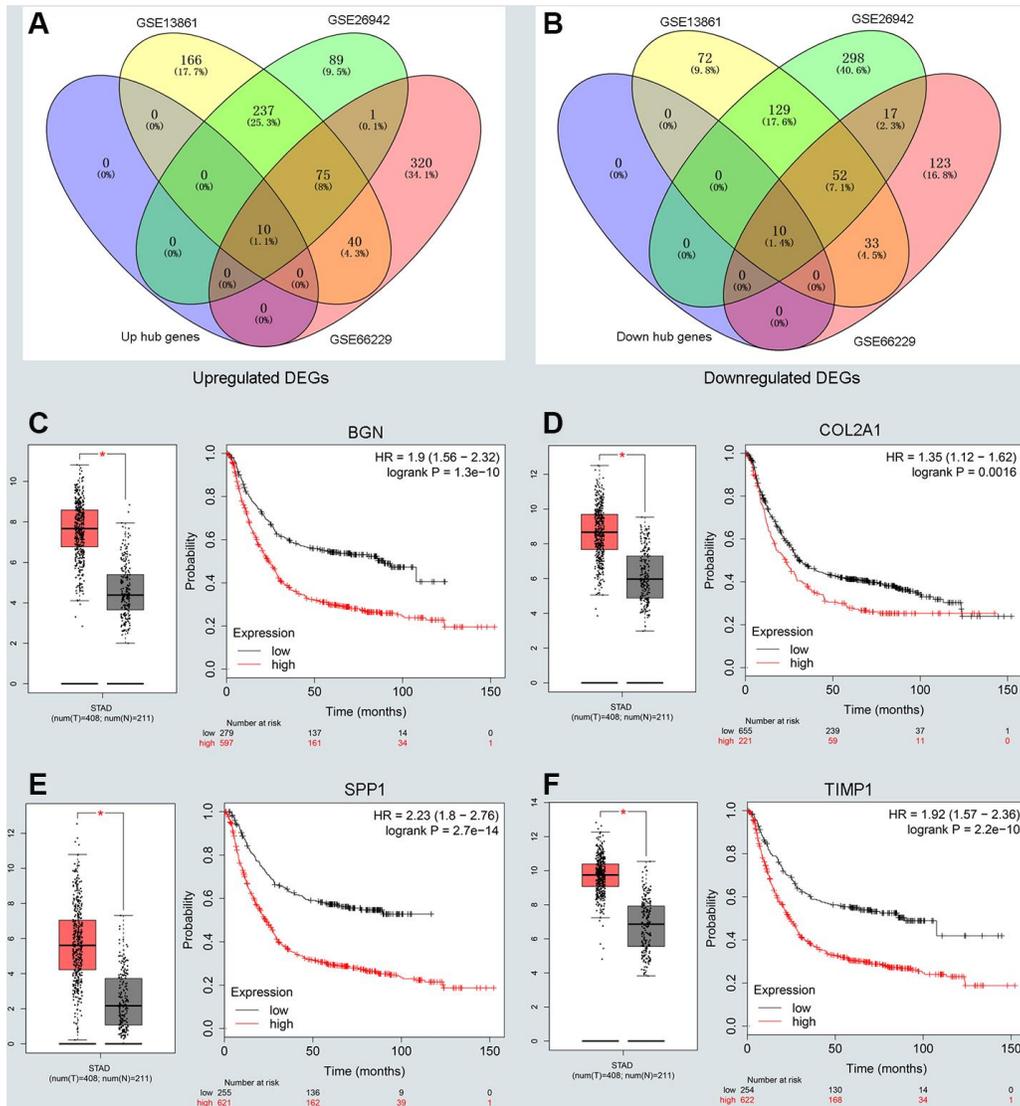
Supplementary Figure 2. Verifying the distribution of hub genes in the validation group and identifying key genes in GC. (**A**, **B**) The intersection of the top ten key hub genes in the validation group (GSE27342, GSE37023, and GSE65801). (**C**–**F**) Validating expression roles and prognosis values of key genes in hub genes using GEPIA and Kaplan–Meier plotter databases.
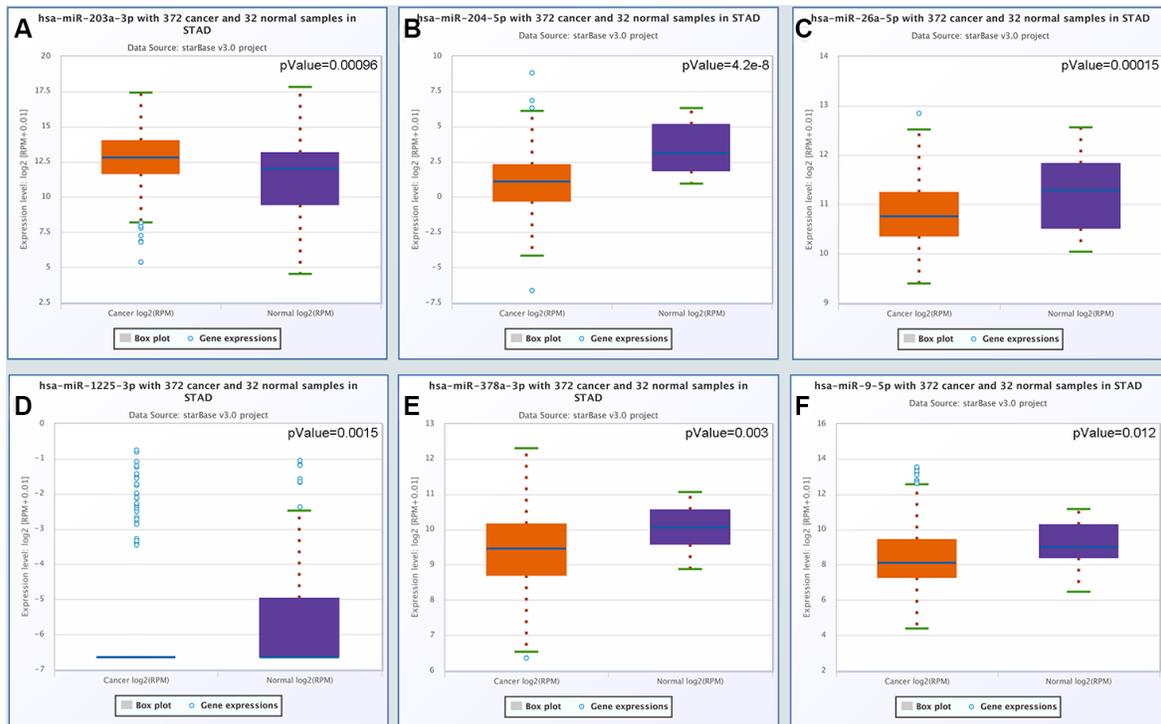
Supplementary Figure 3. Screening and validating the expression roles of key miRNAs in GC. (**A**–**F**) Validating expression roles of key miRNAs using GEPIA databases.
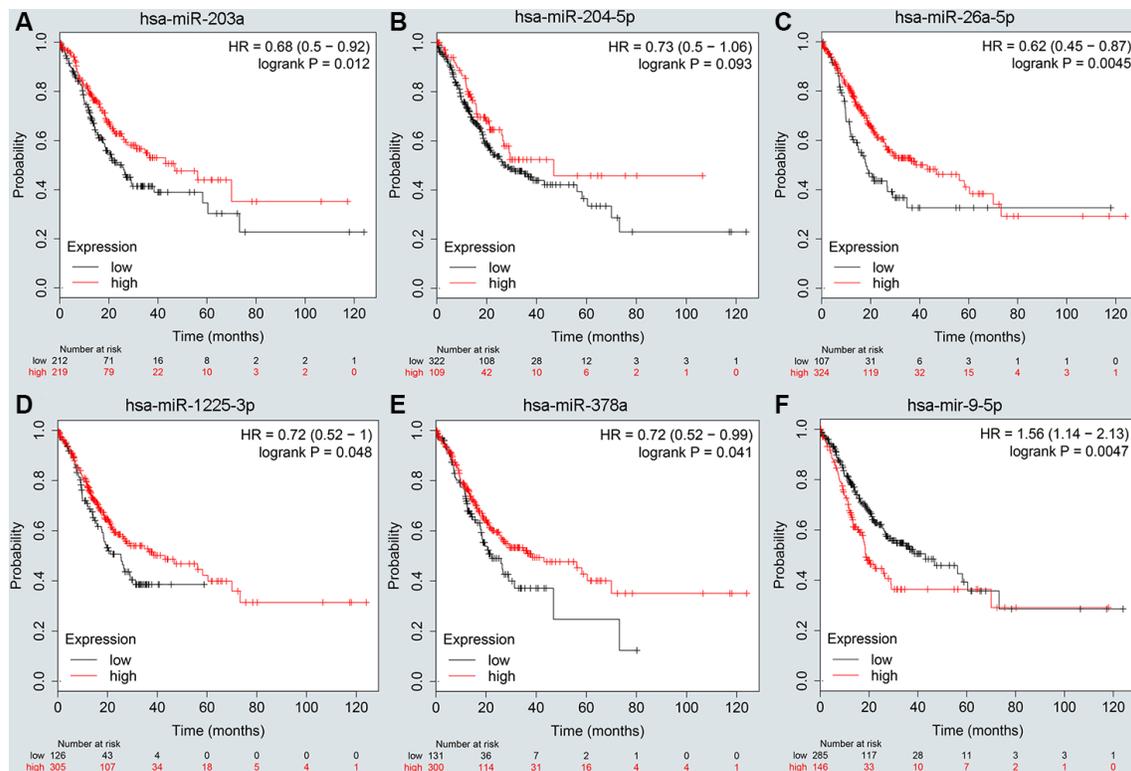
Supplementary Figure 4. Screening and validating the prognostic values of key miRNAs in GC. (**A**–**F**) Validating prognosis values of key miRNAs using Kaplan–Meier plotter databases.
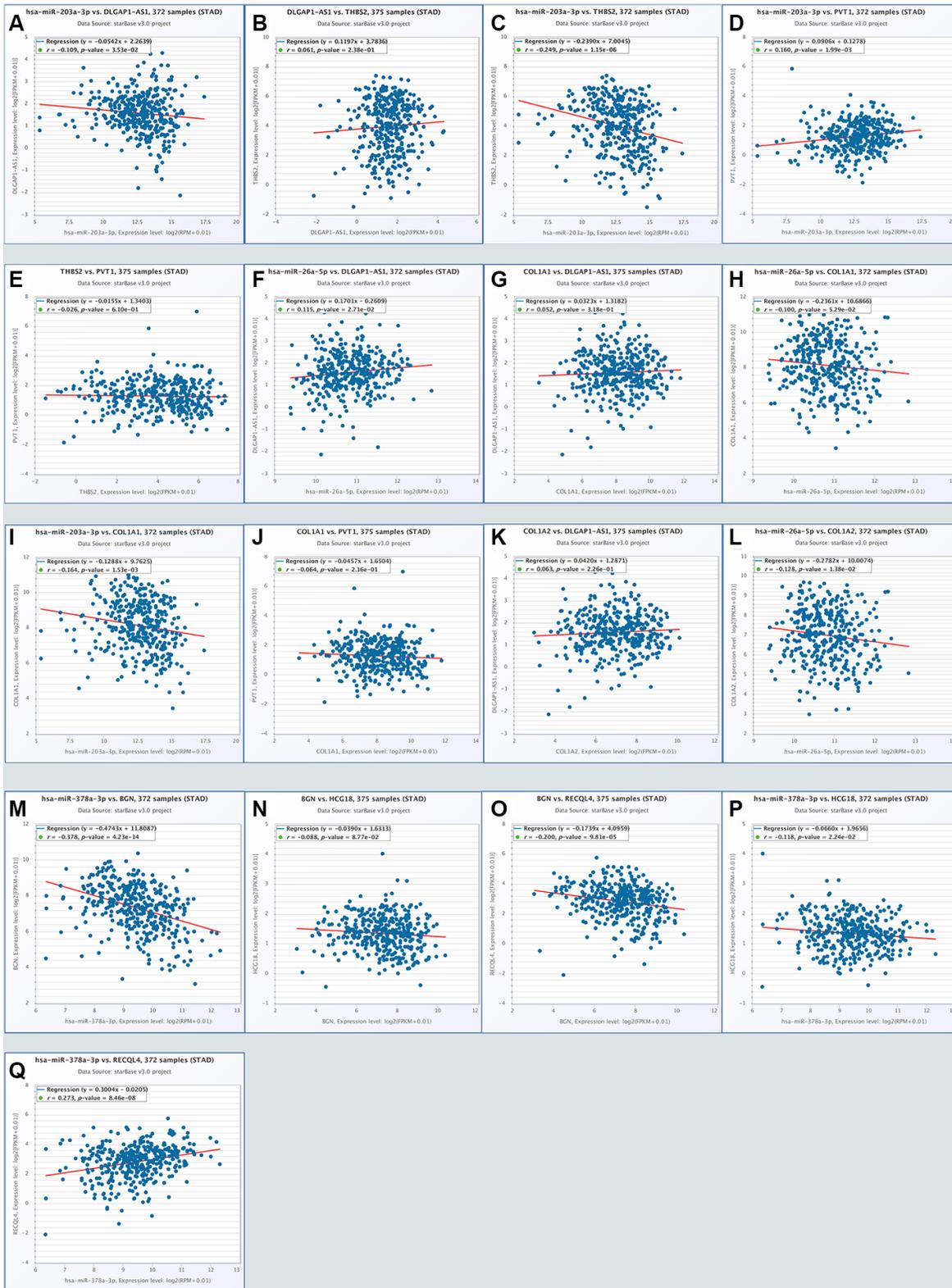
Supplementary Figure 5. Identifying the qualified ceRNA network through correlation analysis. Only DLGAP1-AS1/ miR-203a-3p/THBS2 ceRNA axis met the correlation analysis according to the ceRNA hypothesis (**A**–**C**). Other ceRNA networks failed the criteria that lncRNAs positively associated with mRNAs while miRNAs negative related to lncRNAs and mRNAs (**D**–**Q**).
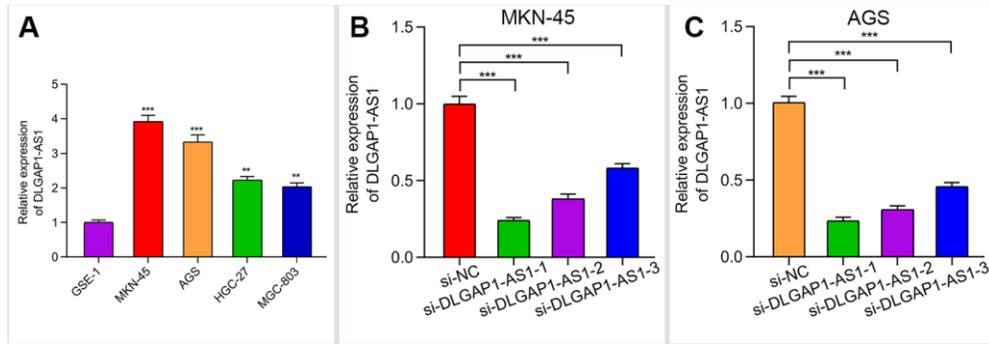
Supplementary Figure 6. Detecting the expressed of DLGAP1-AS1 in GC cell lines and assessing knockdown efficiency. (**A**) The expression of DLGAP1-AS1 in GC cell lines (MKN-45, AGS, HGC-27, and MGC-803) and a normal cell line GSE-1 detected by RT-qPCR. (**B**, **C**) Knockdown efficiency of DLGAP1-AS1 in MKN-45 and AGS cells.
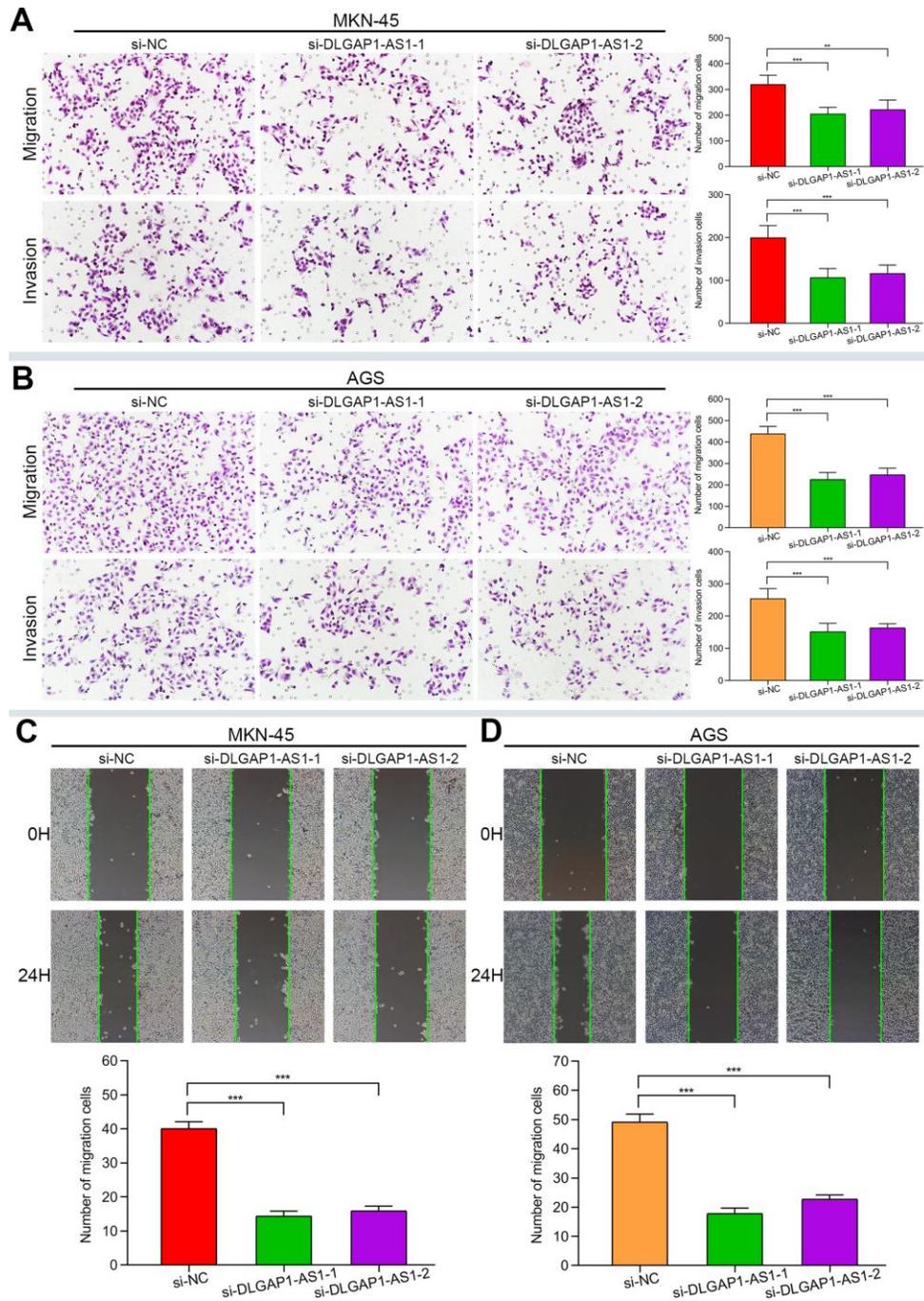
Supplementary Figure 7. Cell migration and invasion assays in MKN-45 and AGS cells. (**A**) Cell migration and invasion assays in si-NC, si-DLGAP1-AS1-1 and si-DLGAP1-AS1-2 transfected MKN-45 cells. (**B**) Cell migration and invasion assays in si-NC, si-DLGAP1-AS1-1 and si-DLGAP1-AS1-2 transfected AGS cells. (**C**) Wound-healing assay assays in si-NC, si-DLGAP1-AS1-1 and si-DLGAP1-AS1-2 transfected MKN-45 cells. (**D**) Wound-healing assay assays in si-NC, si-DLGAP1-AS1-1 and si-DLGAP1-AS1-2 transfected MKN-45 cells. Data are presented as means of three different experiments. Bars indicate ± SE. * P<0.05, **P<0.01, and ***P<0.001 compared with control groups.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 2 to 4.

Supplementary Table 1. The commonly upregulated or downregulated genes in GSE54129, GSE29272 and GSE13911 datasets and TCGA databases.

Upregulated genes	Downregulated genes
APOE	ACER2
ASPN	ADH1C
BGN	ADH7
CDH11	ADHFE1
CDH3	ADRB2
CEMIP	AKR1B10
CHI3L1	AKR1B15
CLDN1	AKR1C1
CLEC5A	AKR7A3
COL10A1	ALDH3A1
COL11A1	ALDH6A1
COL12A1	ALDOB
COL1A1	AMPD1
COL1A2	ANXA10
COL3A1	ATP4A
COL4A1	ATP4B
COL5A2	B3GNT6
COL8A1	BHLHA15
COMP	CA2
CPXM1	CAPN9
CST1	CCKAR
CST2	CCKBR
CTHRC1	CHGA
CXCL1	CKB
CXCL10	CKM
CXCL8	CKMT2
CXCL9	CLIC6
DTL	CPB1
ECT2	CTSE
F2RL2	CYP2C18
FAP	CYP2C19
FKBP10	CYP2C9
FNDC1	DPT
HOXA10	DUOX1
HOXA13	EPN3
HOXC10	ESRRG
HOXC6	ETNPPL
IFITM1	FAM46C
IGF2BP3	FBP2
INHBA	FCGBP
ITGBL1	FMO5
KLK6	GATA5
LAIR2	GCNT4
LIF	GHRL
LIPG	GIF
LOX	GKN1
LY6E	GKN2
MAGEA6	GPER1
MFAP2	GRIA4
MMP11	GSTA1
MMP12	GSTA3
MMP3	GUCA2B
MMP7	HDC
MSR1	HOMER2
NOX4	IGJ
OLFM4	IRX3
OLFML2B	KCNE2

OLR1	KCNJ13
P4HA3	KCNJ15
PDGFRB	KCNJ16
PLA2G2A	KLF4
PLA2G7	KRT20
PLAU	LDHD
PMEPA1	LTF
PRRX1	MAMDC2
RARRES1	MAOA
SALL4	MT1E
SERPINE1	MT1G
SERPINH1	MT1H
SFRP4	MT1M
SNX10	MT1X
SPARC	MUC6
SPP1	MYRIP
SULF1	NKX6-2
THBS2	NR0B2
THY1	PDIA2
TIMP1	PGA4
TNFAIP6	PGC
TNFRSF11B	PIGR
TNFSF4	RAB27A
TREM1	RDH12
TREM2	REG3A
VCAN	SCGB2A1
WNT2	SCNN1B
	SCNN1G
	SH3GL2
	SIDT2
	SLC26A7
	SLC5A5
	SMPD3
	SOX21
	SPTSSB
	SST
	SSTR1
	SULT2A1
	SYTL5
	TCN1
	TFF1
	TFF2
	TMED6
	TNFRSF17
	UGT2B15
	UPK1B
	VSIG1
	VSIG2
	VSTM2A

Supplementary Table 2. The mRNA-miRNA pairs predicted by the miRecords database.

Supplementary Table 3. The miRNA-lncRNA pairs predicted by the miRNet database.

Supplementary Table 4. The raw data of Figure 5.